COMPUTATIONAL AND SYNTHETIC APPROACHES IN THE DESIGN AND
DEVELOPMENT OF CHEMICAL PROBES FOR ESTROGEN RECEPTOR FUNCTION

BY

CHRISTOPHER G. MAYNE

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

      Professor John A. Katzenellenbogen, Chair
      Professor Paul J. Hergenrother
      Professor Ann M. Nardulli
      Assistant Professor Martin D. Burke

ABSTRACT

A member of the nuclear receptor superfamily, the estrogen receptor (ER) is a ligand-regulated transcription factor responsible for the regulation of hundreds of genes. Consequently, ERs are involved in numerous disease states, including cellular proliferation, post-menopausal symptoms, inflammation, and neurodegeneration, all of which represent potential opportunities for endocrine therapies. Sustained efforts in structural biology have led to the deposition of many high resolution x-ray crystal structures of ligand-receptor complexes into the PDB, and provide valuable insight into the key ligand-receptor interactions determining binding affinity and, in some cases, specific macroscopic structural attributes that directly affect ER function. As described herein, we have leveraged selected PDB structures, supplemented by additional unpublished structures obtained from collaborators, to design and develop chemical probes of ER function.

The underlying mechanisms driving ligand affinity and selectivity remain a key focus in understanding ER function. Correspondingly, we describe the development of imidazo[1,2-a]pyridine ligands to probe the importance of the core scaffold structure in ligand binding affinity. Computational analysis of ligand and receptor structures has led to the identification of the interaction between their respective dipole moments as an important receptor-ligand interaction. We also set out to discover novel ER scaffolds through a virtual screening approach and follow-up synthetic efforts to identify and further investigate a thiadiazole scaffold bearing an extended alkyl substituent. Docking structures suggest an intriguing binding mode probing the presence of a putative second binding volume reminiscent of that observed for the high affinity and highly selective glucocorticoid receptor (GR) ligand, deacylcortivazol (DAC).

The diverse biological roles of ERs also provide opportunities to probe receptor function through alternative mechanisms. We report the use of recent crystal structures to design novel modifications of known ER ligands in probing the molecular basis for receptor crosstalk between ER and NF-κB, and the resulting effect on inflammatory pathways. These ligand modifications are centered on destabilizing helix 12 by disrupting the position of a single histidine residue within the binding pocket, and have been shown to effect antagonist activity on both classical ER pathways and the expression of IL-6, the latter being representative of inflammatory responses, *in vivo*. Previous work in our labs has also demonstrated that the assessment of ER-dependent targets can provide a better indication of ER function *in vivo* than assaying ER itself. To this end, we have developed new scoring functions for evaluating docked structures of fluorinated analogues of Tanaproget for the progesterone receptor (PR), whose expression is tightly controlled by ER. These functions have been applied in the design and selection of new synthetic targets for use in imaging ER-positive tumors via positron emission tomography (PET).

ACKNOWLEDGEMENTS

The pursuit of a Ph.D. is not a simple process, and many people deserve acknowledgment for their contributions to my graduate career. First and foremost, I am forever indebted to my research advisor, Professor John A. Katzenellenbogen, for taking me into his laboratory and offering his guidance for the past several years. I am most grateful for his patience and willingness to let me follow my own curiosities in lab, to find dead ends and new ideas, and to figure out which was which. A great deal of thanks goes out to all of my labmates including Andy LaFrate, John Comninos, Jill Gunther, Davis Oldham, and Dr. Sung-Hoon Kim for their invaluable discussions, ideas, and support. I would like to express a special degree of gratitude for Terry Moore and the many conversations, afternoon coffee breaks, late-night dry erase board sessions, and his genuine friendship. Finally, the completion of this degree would not have been possible without the steadfast support of my Ultimate teammates, friends, and family.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

I. Estrogen Receptor Background

The estrogen receptor (ER) belongs to a superfamily of 48 known nuclear receptors (NRs), 24 of which have known endogenous ligands. NRs consist of three modular domains: an N-terminal transactivation domain, a central DNA binding domain (DBD), and a C-terminal ligand binding domain (LBD).[1] The general three-dimensional structure of NRs is conserved through the receptor superfamily and is well illustrated by the recently solved crystal structure of a RXR/PPAR heterodimer bound to all of the individual elements required for transcriptional competency, which include a ligand, coactivator binding protein, and DNA response element (Fig. 1.1).[2] While each of these elements has been targeted in efforts to modulate NR-mediated activity,[3] a long-standing interest in the Katzenellenbogen laboratory has been the modulation of ER activity by targeting the ligand binding pocket. The specific monomeric complex for ER bound to its endogenous ligand, estradiol (E2), is shown in Figure 1.2 and provides more ER-
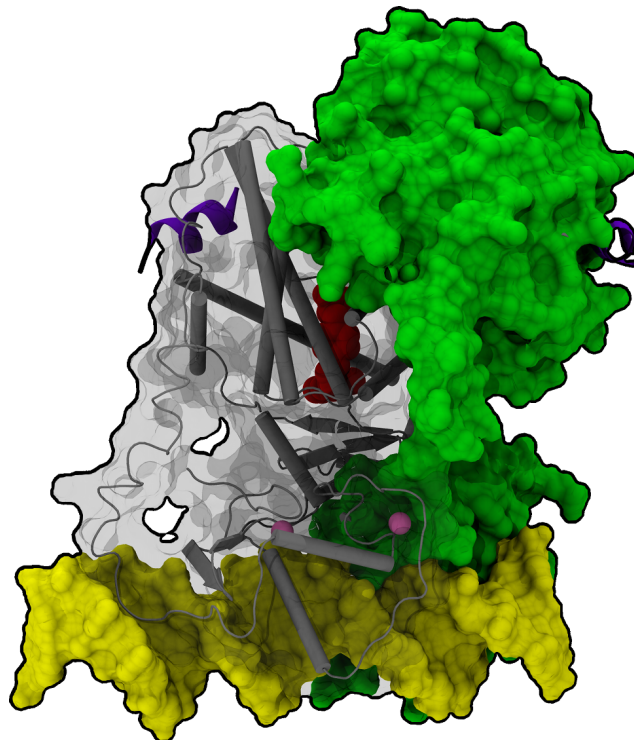

Figure 1.1. Crystal structure illustrating the important elements for the transcriptional competency of a NR. The dimeric receptor structure is comprised of RXR (green) and PPAR (grey) monomers bound to a DNA response element (yellow). The translucent molecular surface of the PPAR monomer reveals the positioning of Rosiglitazone (red) in the ligand binding pocket and the zinc atoms (pink) located within the zinc fingers responsible for binding DNA. A short peptide bearing an LXXLL motif (purple) is bound into each of the coactivator binding grooves, where full-length coactivator proteins bind. Figure prepared from pdb structure 3DZY.[2]
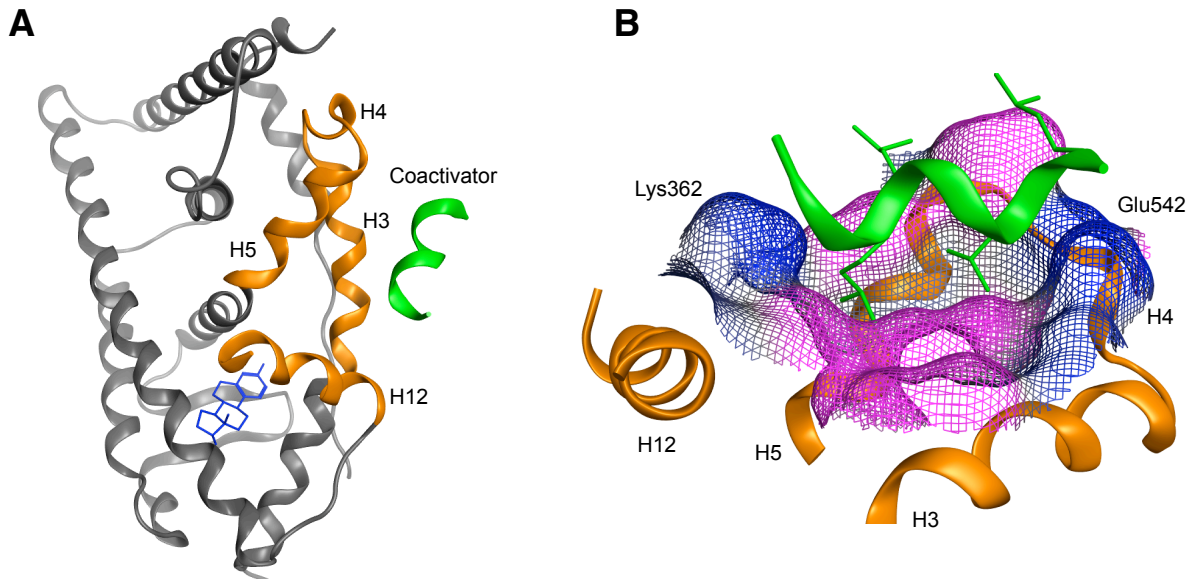
Figure 1.2. *Panel A.* Upon the binding of an agonist (e.g., estradiol [blue]) within the ligand binding pocket of ER, the conformation of H12 is stabilized in a position over the face of the pocket, forming a hydrophobic region between helices 3, 4, 5 and 12 (orange). *Panel B.* Coactivator proteins (green) bearing an LXXLL motif bind to the surface of the protein (wireframe), such that two leucine residues are buried into the hydrophobic core (pink) of the groove, and charged residues Lys362 and Glu542 (blue) form a charge-clamp with the intrinsic dipole of the helix. Figure prepared from pdb structure 1GWR.[4]

specific structural details.[4] The bound ligand is completely surrounded by the receptor in a pocket which is formed by 11 helices and capped by a 12th helix (H12) that serves as the 'lid' for the binding pocket. This receptor conformation positions H12 in a manner that exhibits a hydrophobic groove above the binding pocket that recruits associated coactivator proteins bearing an LXXLL motif (L = Leu, X = any amino acid) and initiates a biochemical sequence that activates transcriptional machinery. In this manner, ERs regulate hundreds of genes and are involved in numerous physiological and disease states, including cellular proliferation,[5,6] post-menopausal symptoms,[7,8] inflammation,[9-12] and neurodegeneration,[13] all of which represent potential opportunities for the expanded development of endocrine therapies.

ERs play a diverse role across a spectrum of biological functions, and an important aim in endocrine therapies has been that of designing ligands that exhibit tissue-specific activity. For instance, antagonist ligands are desired for treatment of estrogen-dependent cancer by slowing cell growth, while agonist ligands are sought to stimulate bone and cardiovascular tissues. However, agonism versus antagonism of a ligand is not fully tissue specific, and unwanted side-effects can occur. The pharmacological profiles of the so-called "selective estrogen receptor modulators (SERMs) tamoxifen and raloxifene (Table 1.1) are classic examples demonstrating ligand tissue-specificity.[14-16] While tamoxifen acts as an antagonist in breast tissue, as an anti-cancer therapy it is plagued by agonist activity that can stimulate new cancers in uterine tissue. Raloxifene, however, acts as an antagonist in breast and uterine

tissue, but as an agonist in bone leading to its use as both an anti-cancer agent and in hormone replacement therapies. This tissue selectivity stands in stark contrast to that of E2, which acts as an agonist in each of these tissues, and ICI 182,780 (fulvestrant), which is known as a "pure antagonist."[17,18] Several factors influence the pharmacological profile of SERMs.[19] The structural attributes of the ligand itself impacts both agonism/antagonism and a selectivity for one of two different ER subtypes (discussed in detail below). Furthermore, every receptor-ligand complex functions in an environment that includes a constellation of effectors, such as coactivators, corepressors, and DNA response elements that constitute a cellular and tissue-specific context.[20,21] Combined, these three component parts result in a model of "tripartite pharmacology"[22-24] for ERs within which synthetic ligands are currently under development.

Table 1.1. Ligand-Dependent Pharmacological Profiles of ER .

Ligand	Breast	Uterine	Bone
Estradiol	+	+	+
Tamoxifen	-	+	+
Raloxifene	-	-	+
ICI 182,780	-	-	-

The literature regarding ER is quite expansive, and several reviews have recently been published covering biological functions,[25-27] molecular actions,[28,29] and the state-of-the-art in ligand development.[30-32] Much of the work presented in the following chapters is founded or otherwise inspired by structural underpinnings of ER function, and are greatly aided by a firm understanding of several important structural aspects of the ligand binding pocket of ER.

II. Structural Attributes of the Estrogen Receptor

A. Topology of the Ligand Binding Pocket

The first x-ray crystal structure of a ligand-bound ER was solved by Brzozowski et al. in 1997,[33] and was only preceded by structures of TR[34] and RXR[35,36] within the NR superfamily. While the ER structure carried with it a broader applicability to NRs as a superfamily, it provided for the first time a direct description of receptor-ligand interactions in ER explaining observed trends in binding affinity. From this model, the topology of the pocket is typically defined in terms of its relationship to the bound E2 ligand, and consists of A-ring and D-ring ends, a central core, and the B-ring sub-pocket (Fig. 1.3.A). Important hydrogen bonding contacts are present at both ends of the pocket. The A-ring of E2 participates in a hydrogen bonding network between Glu353, Arg394, and a highly ordered water molecule. This network of interactions is essential for high affinity to ER. The 17β-hydroxyl of E2 also forms a hydrogen bond to His524
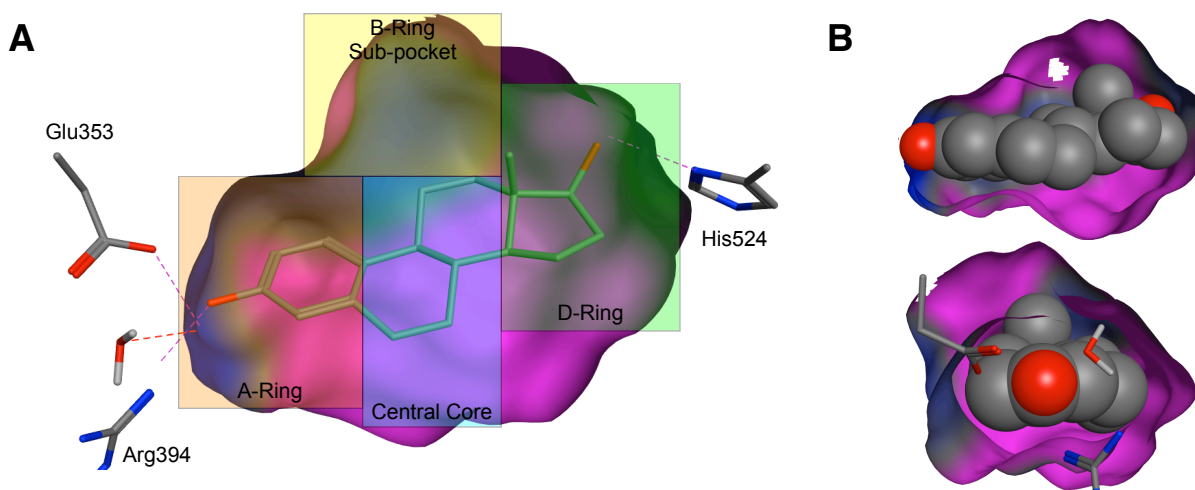
Figure 1.3. The topology of the ER ligand binding pocket is often defined relative to the endogenous ligand, E2. These attributes include hydrogen bonding contacts at the A and D-ring ends of the pocket, a central lipophilic core (magenta = lipophilic, blue = hydrophilic), and additional volume above the B-ring (*Panel A*). The volume of the pocket is significantly larger than is required for E2, with the majority of the unoccupied space above and below the steroidal scaffold (*Panel B*). Figure prepared from pdb structure 1ERE.[33]

at the D-ring end of the pocket; however, this contact is less essential to achieving high affinity. The central core of the binding pocket, in contrast, is largely lipophilic and interacts with the ligand through non-specific van der Waals (vdW) interactions. Finally, the B-ring sub-pocket is frequently accessed to engender subtype selectivity and is the site of receptor remodeling to induce the antagonist structure, as described in more detail in the following sections.

The total volume of the binding pocket is significantly larger than the volume occupied by the E2 ligand, 450 Å^3 and 245 Å^3, respectively (Fig. 1.3.B).[33] The majority of this unoccupied space exists above and below the plane of the steroidal scaffold. Ligands designed with a larger three-dimensional cross section aimed at filling this space, such as the bicyclononane ligand shown in Figure 1.4, demonstrate significantly higher binding affinity than that of E2.[37]
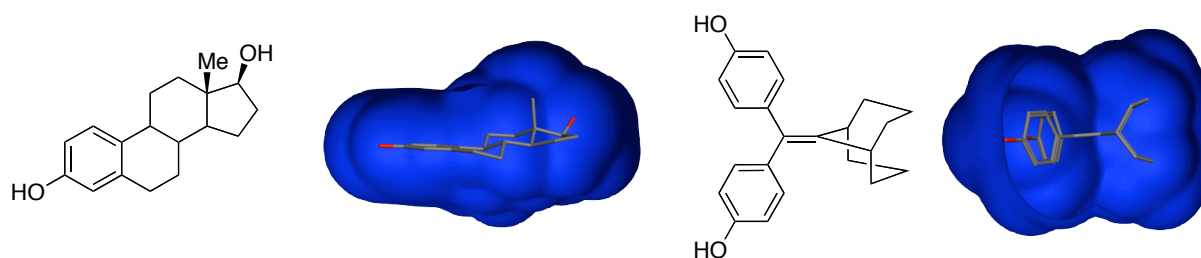
Figure 1.4. A bicyclononane scaffold was employed to yield a larger three dimensional van der Waals surface (blue), yielding a 5-fold increase in binding affinity relative to E2.

The bulk electronic properties of the ligand binding pocket are noticeably lacking from discussions of pocket structure in the literature; however, we have recently acquired quantum mechanical and semi-empirical derived data indicating that the dipole moment within the local

environment of the binding volume is rather large (~36 Debye). The alignment of receptor and ligand dipoles appears to have significant implications on binding affinity, and by extension, overall ligand design, and is described in further detail in Chapter 3.

B. Structural Models of Agonism vs. Antagonism

Pharmacophores for agonism versus antagonism have long been established based on medicinal chemistry; however, the specific molecular mechanisms remained poorly understood until the crystal structures of ER bound to E2 and raloxifene were solved in 1997,[33] and supplemented by the publication of many additional structures since that time. The structures of agonist-bound receptor-ligand complexes highlight that the ligand is completely surrounded by the receptor, forming a closed ligand binding pocket and the hydrophobic coactivator binding groove depicted in Figure 1.2. Antagonist-bound ER structures, in contrast, generally contain a sidechain that is extended through the B-ring sub-pocket, introducing a significant conformational change that opens a channel out of the binding pocket, and precludes the binding of these coactivators through one of two mechanisms representing classical antagonism. The first of these mechanisms is demonstrated by raloxifene-bound ERα,[33] as shown in Figure 1.5.A. The bulky basic sidechain of the ligand extends into the position occupied by H12 in agonist-bound structures. Accordingly, H12 undergoes a major shift so that it binds in the coactivator binding groove itself, with two leucine residues within H12 mimicking the LXXLL motif of coactivator binding proteins. Alternatively, the pure antagonist ICI 164,384
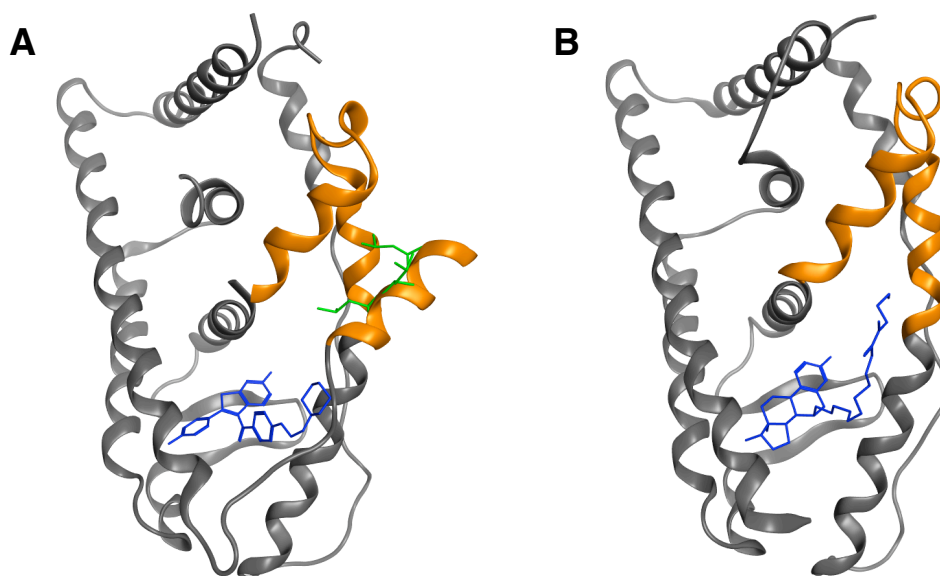
Figure 1.5. *Panel A.* The bulky basic sidechain of raloxifene (blue) exits the ligand binding pocket and shift H12 into the coactivator binding groove, where leucine residues (green) mimic the LXXLL motifs of coactivator proteins. *Panel B.* The longer sidechain of ICI 164,384 (blue) extends out of the ligand binding pocket and binds directly into the coactivator binding groove. The position of H12 is disordered, and is not present in the structure. Figure prepared from pdb structures 1ERR[33] and 1HJ1.[39]

(an analogue of ICI 182,780)[17,38] bears a substantially longer sidechain that is capable of extending out of the ligand binding pocket and binding directly within the coactivator binding groove (Fig. 1.5.B).[39] In this structure, H12 is disordered and is not observed.

Additional structural data has also resulted in the several hypotheses regarding non-classical modes of antagonism. Wang *et al.* have reported a crystal structure of ER that contains hydroxytamoxifen within the ligand binding pocket and a second hydroxytamoxifen molecule bound directly in the coactivator binding groove.[40] The binding mode of this second ligand molecule buries the unsubstituted phenyl group deep within the lipophilic coactivator binding groove; however, the remainder of the molecule does not appear to form any other specific interactions (e.g., hydrogen bonds) with the receptor surface. Based on the presented data, we believe that this binding mode is unlikely at physiologically relevant ligand concentrations, and not a true mode of antagonism.[3,41] More convincing, however, is recent structural data suggesting antagonism through the destabilization of the last two turns in helix 11 by disrupting the position of His524 within the ligand binding pocket. Termed "passive antagonism", this phenomenon has only recently been observed for a limited number of ligands including tetrahydrochrysenes (in ERβ only),[42] oxabicycloheptenes, pyrazolopyrimidines,[43] and indazoles.[44]

While the described structural attributes broadly define agonism vs. antagonism for ERs, it is important to note that the position of H12 is not accurately described in the binary context of "on" or "off." More recently, it has been suggested that H12 can adopt intermediate positions and exists as a dynamic ensemble of receptor conformations that blurs the distinction between agonist and antagonist activity.[42,44-46]

C. Receptor Plasticity

The extensive ligand-induced remodeling of the receptor structure illustrated by antagonist conformations of ER indicates a degree of receptor flexibility. Not all facets of the ligand binding pocket are flexible, and modifications of the A-ring mimetics, or increasing the steric bulk of certain internal substituents result in significant reductions in binding affinity. While introduction of new substituents that challenge binding constraints often initially form unfavorable steric interactions, when properly positioned, these interactions can cause a remodeling of the receptor structure to open new pockets or channels with the opportunity to increase receptor-ligand interactions yielding an overall increase in binding affinity and/or specificity. The latter is particularly important due to the conserved structural attributes across NRs that can lead to cross reactivity. For these reasons, ER is referred to as having "plasticity." Receptor plasticity, however, is difficult to model by computational techniques[47] and therefore,
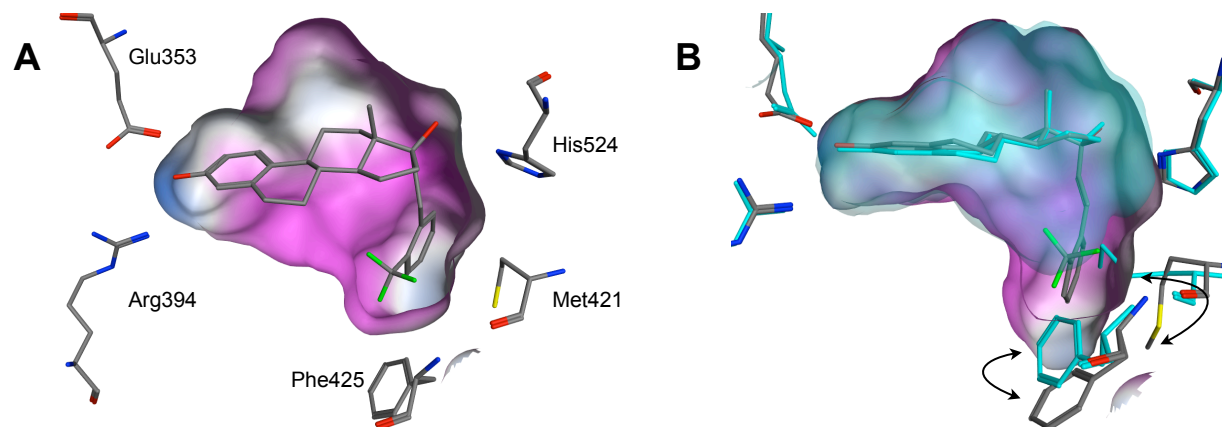
Figure 1.6. *Panel A.* The incorporation of a 17α-trifluromethylphenylvinyl substituent on E2 causes a reorganization of the receptor to open an extended pocket below the D-ring of the steroidal scaffold. *Panel B.* Overlaying the binding pocket observed for E2 (cyan) highlights the rotation of Met421 and Phe425 residues to create the extended pocket (black arrows). Figure prepared from pdb structures 2P15[48] and 1GWR.[4]

investigation is heavily reliant on experimental approaches, namely synthetic efforts through medicinal chemistry and structural biology.

In this manner, a second area of receptor flexibility has recently been demonstrated in ER below the D-ring of the E2 steroidal scaffold. Ligands bearing an extended 17α-phenylvinyl substituent causes residues Met421 and Phe425 to rotate out of position and open a potentially solvent accessible sub-pocket (Fig. 1.6).[48] The incorporation of an *ortho*-trifluoromethyl group in the aryl linker is proposed to form positive interactions with several methionine residues that significantly stabilize receptor remodeling and result in an increase in binding affinity of 3-fold over parent compounds.

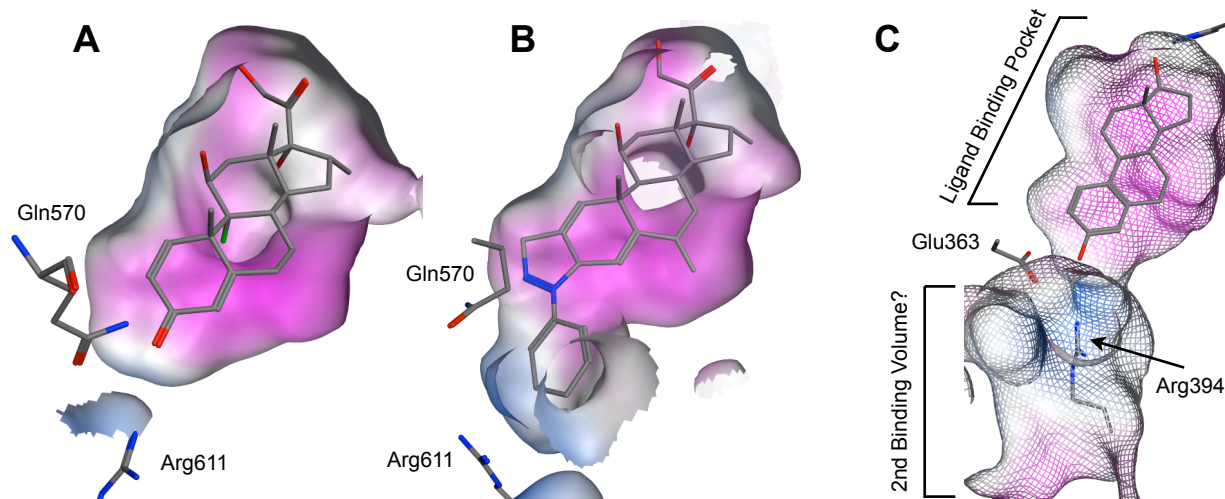**Figure 1.7.** *Panel A.* Crystal structure of Dexamethasone bound to glucocorticoid receptor (GR). *Panel B.* The *N*-phenylpyrazole of deacylcortivazol (DAC) causes Gln570 and Arg611 to adopt new positions double the binding volume. *Panel C.* Residues Glu363 and Arg394 block access to a potential extended binding volume in ER. Figure prepared from pdb structures 1M2Z,[49] 3BQD,[50] and 1GWR.[4]

7

An important example of receptor plasticity was recently disclosed for the glucocorticoid receptor (GR) that holds potential implications for ERs as well. Deacylcortivazol (DAC) is a well established high affinity steroidal ligand for GR that contradicts the steric sensitivity associated with NR pharmacophores regarding substitutions on the A-ring. The molecular mechanisms explaining the increased affinity were only recently elucidated via a crystal structure indicating that the *N*-phenyl pyrazole causes residues Gln570 and Arg611 to rotate out of position and allow access to the adjoining solvent channel.[49,50] This structural insight has been key to the design of novel, non-steroidal, high affinity, high selectivity ligands for use as anti-inflammatory therapeutics.[51] A similar structural motif is observed in ER, and has been postulated as a putative second binding volume based on modeling studies[52], in conjunction with non-competitive binding of a tetrahydrochrysene ketone and E2.[53] This hypothesis is the foundation for the synthesis of thiadiazole-based ligands identified by virtual screening as described in Chapter 5.
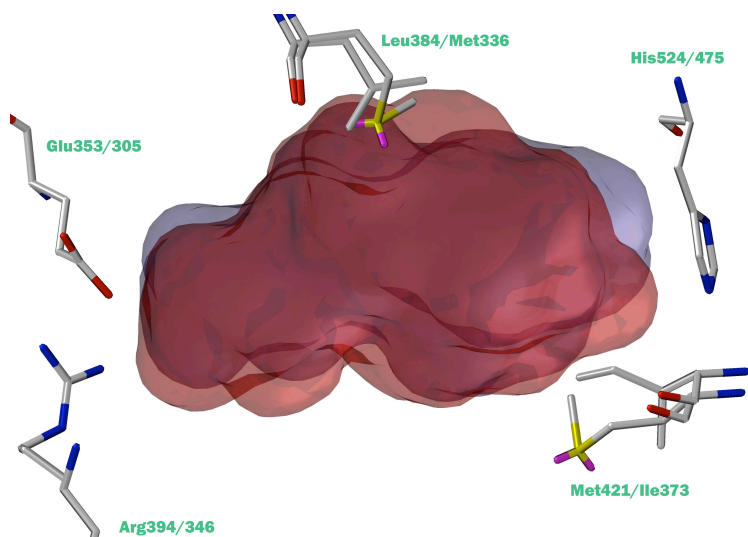
Figure 1.8. The ligand binding pocket of ERα (red, 490 Å^3) is slightly larger than that of ERβ (blue, 390 Å^3), and differs by only two amino acids: Leu384 to Met336 and Met421 to Ile373. Figure prepared from pdb structures 1ERE[33] and 1X7B.[57]

D. Subtype Selectivity: ERα vs. ERβ

The estrogen receptor and all structures described up until this point have been for what is now known as ERα. In 1996, Kuiper et al.[54,55] discovered a second subtype of ER, termed ERβ. This new receptor subtype represents a completely different gene product from ERα, and contains only 59% amino acid sequence identity within the ligand binding domain. The interior of the ligand binding pocket, however, remains largely unchanged, containing only two conservative amino acid residue changes: Leu384 to Met336, and Met421 to Ile373, and a slightly smaller binding volume for ERβ (390 Å^3) relative to ERα (490 Å^3).[56,57]
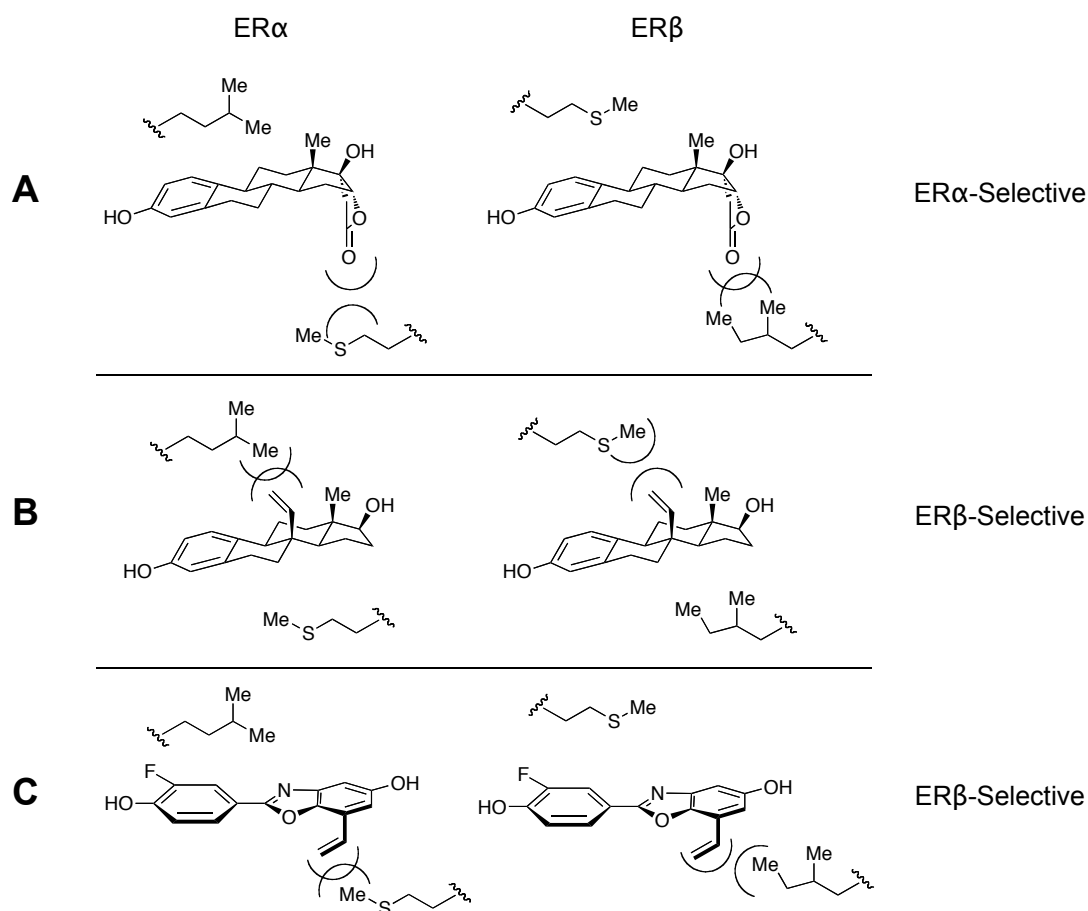
Figure 1.9. The unfavorable steric interaction of the lactone substituent with Ile373 of ERβ engenders subtype selectivity favoring ERα for α-lactone estradiol (A). Similarly, ERβ-selectivity can be achieved through unfavorable interactions with Leu384 (B) or Met421 (C) of ERα.

While subtle, these differences have provided adequate opportunity for the discovery and design of subtype-selective ligands.[31] Many ligands leverage the differential placement of the branched aliphatic residues (Leu, Ile) versus the linear nature of the methionine residue. The introduction of substituents that initiate unfavorable steric interactions yield a reduction in binding affinity, thereby favoring binding to the opposing receptor subtype. By this mechanism, ERα-selective ligands typically project a substituent towards the Ile373 residue of ERβ, such as the α-lactone-estradiol shown in Figure 1.9.A.[58] Similarly, ERβ-selectivity can be engendered by directing a substituent towards the Leu384 residue in ERα (Fig. 1.9.B)[59] or, alternatively, positioning a substituent that fits within the confines defined by Ile373 of ERβ but clashes with the extended sidechain of Met421 in ERα (Fig. 1.9.C).[57]

Subtype selectivity can also be influenced by establishing a positive interaction with either of these differential residue pairs. This proposed mechanism explains the ERβ-selectivity

Diarylpropionitrile (DPN)   propylpyrazole triol (PPT)
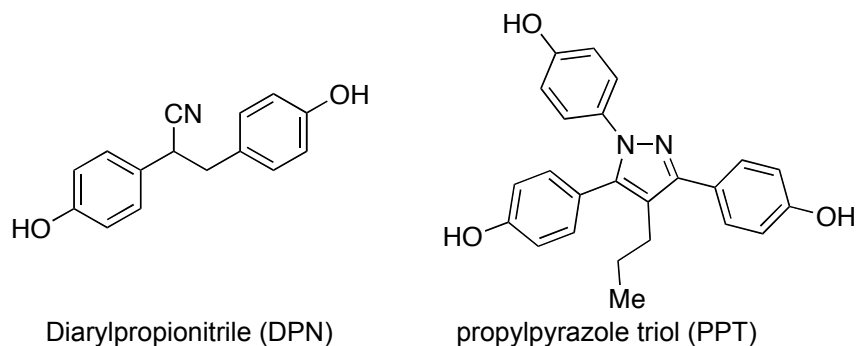
Figure 1.10. Structures of ERβ-selective DPN and ERα-selective PPT ligands.

of DPN (Fig. 1.10),[60] and is supported by quantum mechanical calculations that invoke a polarized interaction between the π-system of the nitrile substituent and the sulfur atom of the Met336 sidechain.[57] Potentially, the most straightforward approach to engendering selectivity in favor of ERα-selectivity is by taking advantage of the larger binding volume of ERα through size exclusion principles. Ligands, such as PPT (Fig. 1.10)[61] that include an internal aromatic substituent that fits well within ERα, however, exceed the confines of the narrower pocket in ERβ. While the reverse philosophy may work to engender a degree of ERβ-selectivity based on reduced vdW contacts of smaller ligands in a larger pocket, this approach lacks a clearly defined example.
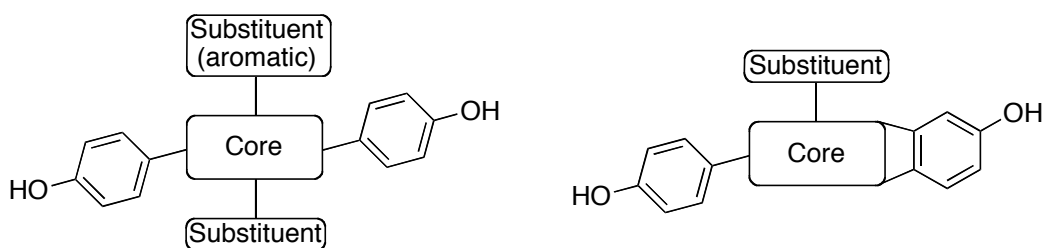
Figure 1.11. Pharmacophores for non-steroidal ERα (left) or ERβ (right) selective ligands.

III. Non-steroidal Ligand Pharmacophores

The elucidation of structural attributes of NRs, and more specifically ER, has resulted in a more complete understanding of ER pharmacophores. A strong emphasis has been placed on non-steroidal ligands due to their synthetic tractability and superior selectivity compared to many steroidal scaffolds. As described in Figure 1.11, these compounds contain a core scaffold bearing several substituents.[61,62] All structures contain a phenolic moiety to mimic the A-ring of E2, and this is required for high binding affinities. ERα selective ligands typically contain two additional aryl substituents (often phenolic) and a fourth smaller (generally alkyl) internal substituent. The phenolic substituents typically serve as a hydrogen bond donor to interact with

His524 at the D-ring end of the pocket, and the internal aryl group is directed into the B-ring sub-pocket. The alkyl substituent generally serves to fill the remainder of the available binding volume, forming vdW interactions with the lipophilic central core region.

The ERβ selective ligand pharmacophore is similar, with slight modifications based on the structural differences described above. While the A-ring pendant phenol remains in place, the D-ring phenol is often fused to the central core, reducing overall ligand size while retaining a hydrogen bond donor to interact with His524. This is typically achieved through a 5,6- or 6,6-fused bicyclic core. The reduced width of the binding pocket precludes the incorporation of the internal aryl substituent; however, the smaller internal substituent remains and plays an important role in establishing binding affinity and selectivity.

While these pharmacophores provide a guide for many considerations in ligand design, there are a number of aspects that require continued investigation and development. For instance, it remains unclear whether the core scaffold plays a direct role in influencing ligand binding affinity, or is simply responsible for displaying each substituent in the necessary spatial location.[61] Furthermore, many pharmacophores can be converted from agonist to antagonist structures by attaching an extended B-ring substituent similar to raloxifene or ICI 164,384; however, the specific mechanisms controlling the tissue specificity of agonism vs. antagonism for SERMs have not been fully elucidated.[20]

IV. Conclusions

The complex role played by the two ERs in an array of biological functions provides many opportunities for the further development of endocrine therapeutics. As synthetic chemists, however, our power to modulate the activity of ERs within the context of a tripartite pharmacology model is currently limited to manipulation of protein structure through ligand design. The deposition of hundreds of ER and ER-related crystal structures into the PDB since Brzozowski's first structures in 1997 has greatly improved our understanding of receptor-ligand interactions, multiple modes of agonism vs. antagonism, subtype selectivity, and receptor plasticity. In many cases, the impact of each structural attribute on ER pharmacology is clear directly from the structure itself; however, others still fail to provide a solid understanding of how subtle changes in the ligand binding pocket specifically affect the receptor dynamics underlying observed phenotypes.

Within this context, we have embarked on several projects described in the following chapters that utilize a mixture of organic synthesis and computational techniques to probe unresolved aspects of receptor-ligand complexes. In Chapter 2, we further investigate how the nature of the core scaffold structure affects binding affinity by replacing the indazole scaffold of known high affinity, ERβ-selective ligands[62] with the geometrically equivalent imidazo[1,2-

a]pyridine constitutional isomer. Subsequently, in Chapter 3 we disclose our initial results in structure-aided design of ER ligands that enforce an alternate conformation of His524 correlated to ER-dependent inhibition of inflammatory pathways mediated by NF-κB.[43] Switching gears slightly in Chapter 4, we describe the use of computational methods to elucidate the molecular mechanisms behind observed trends in binding affinity for fluorinated analogs of the non-steroidal progesterone receptor (PR) ligand, Tanaproget.[63] Finally, in Chapter 5 we report on our use of virtual screening to identify a thiadiazole core scaffold and followup synthetic efforts in probing the presence and importance of a second putative binding volume in the ligand binding pocket of ER.

V. References

(1) Gronemeyer, H.; Gustafsson, J.-Å.; Laudet, V. *Nat. Rev. Drug Discov.* **2004**, *3*, 950.
(2) Chandra, V.; Huang, P.; Hamuro, Y.; Raghuram, S.; Wang, Y.; Burris, T. P.; Rastinejad, F. *Nature* **2008**, 350.
(3) Moore, T. W.; Mayne, C. G.; Katzenellenbogen, J. A. *Mol. Endocrinol.* **2010**, *24*, 683.
(4) Warnmark, A.; Treuter, E.; Gustafsson, J.-Å.; Hubbard, R. E.; Brzozowski, A. M.; Pike, A. C. W. *J. Biol. Chem.* **2002**, *277*, 21862.
(5) Vogel, C. L. *Anti-Cancer Drugs* **2003**, *14*, 265.
(6) Ariazi, E. A.; Ariazi, J. L.; Cordera, F.; Jordan, V. C. *Curr. Top. Med. Chem.* **2006**, *6*, 181.
(7) Clemett, D.; Spencer, C. M. *Drugs* **2000**, *60*.
(8) Wellington, K.; Plosker, G. L. *Disease Management & Health Outcomes* **2003**, *11*, 673.
(9) Chadwick, C. C.; Chippari, S.; Matelan, E.; Borges-Marcucci, L.; Eckert, A. M.; Keith, J. C.; Albert, L. M.; Leathurby, Y.; Harris, H. A.; Bhat, R. A.; Ashwell, M.; Trybulski, E.; Winneker, R. C.; Adelman, S. J.; Steffan, R. J.; Harnish, D. C. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 2543.
(10) Evans, M. J.; Harris, H. A.; Miller, C. P.; Karathanasis, S. K.; Adelman, S. J. *Endocrinology* **2002**, *143*, 3785.
(11) Tyree, C. M.; Zou, A.; Allegretto, E. A. *J. Steroid Biochem. Mol. Biol.* **2002**, *80*, 291.
(12) Caulin-Glaser, T.; Watson, C. A.; Pardi, R.; Bender, J. R. *J. Clin. Invest.* **1996**, *98*, 36.
(13) Saijo, K.; Collier, J. G.; Li, A. C.; Katzenellenbogen, J. A.; Glass, C. K., An ADIOL-ERβCTBP transrepression pathway negatively regulates microglia-mediated inflammation, *Cell*, submitted for publication, **2010**.
(14) Sato, M.; Glasebrook, A.; Bryant, H. *J. Bone Miner. Metab.* **1994**, *12*, S9.
(15) Katzenellenbogen, B. S.; Katzenellenbogen, J. A. *Science* **2002**, *295*, 2380.
(16) Shelly, W.; Draper, M. W.; Krishnan, V.; Wong, M.; Jaffe, R. B. *Obstet. Gynecol. Surv.* **2008**, *63*, 163.
(17) Wakeling, A. E.; Dukes, M.; Bowler, J. *Cancer Res.* **1991**, *51*, 3867.
(18) Howell, A.; Osborne, C. K.; Morris, C.; Wakeling, A. E. *Cancer* **2000**, *89*, 817.
(19) Frasor, J.; Stossi, F.; Danes, J. M.; Komm, B.; Lyttle, C. R.; Katzenellenbogen, B. S. *Cancer Res.* **2004**, *64*, 1522.
(20) Shang, Y.; Brown, M. *Science* **2002**, *295*, 2465.
(21) McKenna, N. J.; O'Malley, B. W. *Nat. Med.* **2000**, *6*, 960.
(22) Katzenellenbogen, J. A.; O'Malley, B. W.; Katzenellenbogen, B. S. *Mol. Endocrinol.* **1996**, *10*, 119.
(23) Katzenellenbogen, B. S.; Katzenellenbogen, J. A. *Breast Cancer Res* **2000**, *2*, 335.
(24) Katzenellenbogen, B. S.; Montano, M. M.; Ediger, T. R.; Sun, J.; Ekena, K.; Lazennec, G.; Martini, P. G. V.; McInerney, E. M.; Delage-Mourroux, R.; Weis, K.; Katzenellenbogen, J. A. *Recent Prog. Horm. Res.* **2000**, *55*, 163.

(25) Heldring, N.; Pike, A.; Andersson, S.; Matthews, J.; Cheng, G.; Hartman, J.; Tujague, M.; Ström, A.; Treuter, E.; Warner, M.; Gustafsson, J.-Å. *Physiol. Rev.* **2007**, *87*, 905.
(26) Gustafsson, J.-Å. *Trends Pharmacol. Sci.* **2003**, *24*, 479.
(27) Nilsson, S.; Gustafsson, J.-Å. *Breast Cancer Res.* **2000**, *2*, 360.
(28) McDonnell, D. P.; Connor, C. E.; Wijayaratne, A.; Chang, C.-y.; Norris, J. D. *Recent Progress in Hormone Research* **2002**, *57*, 295.
(29) Lonard, D. M.; Smith, C. L. *Steroids* **2002**, *67*, 15.
(30) Meegan, M. J.; Lloyd, D. G. *Curr. Med. Chem.* **2003**, *10*, 181.
(31) Minutolo, F.; Macchia, M.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A. *Med. Res. Rev.* **2009**, n/a.
(32) Wallace, O. B.; Richardson, T. I.; Dodge, J. A. *Curr. Top. Med. Chem.* **2003**, *3*, 1663.
(33) Brzozowski, A. M.; Pike, A. C. W.; Dauter, Z.; Hubbard, R. E.; Bonn, T.; Engstrom, O.; Ohman, L.; Greene, G. L.; Gustafsson, J.-A.; Carlquist, M. *Nature* **1997**, *389*, 753.
(34) Wagner, R. L.; Apriletti, J. W.; McGrath, M. E.; West, B. L.; Baxter, J. D.; Fletterick, R. J. *Nature* **1995**, *378*, 690.
(35) Bourguet, W.; Ruff, M.; Chambon, P.; Gronemeyer, H.; Moras, D. *Nature* **1995**, *375*, 377.
(36) Renaud, J.-P.; Rochel, N.; Ruff, M.; Vivat, V.; Chambon, P.; Gronemeyer, H.; Moras, D. *Nature* **1995**, *378*, 681.
(37) Muthyala, R. S.; Carlson, K. E.; Katzenellenbogen, J. A. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 4485.
(38) Bowler, J.; Lilley, T. J.; Pittam, J. D.; Wakeling, A. E. *Steroids* **1989**, *54*, 71.
(39) Pike, A. C. W.; Brzozowski, A. M.; Walton, J.; Hubbard, R. E.; Thorsell, A.-G.; Li, Y.-L.; Gustafsson, J.-Å.; Carlquist, M. *Structure* **2001**, *9*, 145.
(40) Wang, Y.; Chirgadze, N. Y.; Briggs, S. L.; Khan, S.; Jensen, E. V.; Burris, T. P. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 9908.
(41) Moore, T. W.; Katzenellenbogen, J. A. In *Annu. Rep. Med. Chem.*; John, E. M., Ed.; Academic Press: 2009; Vol. Volume 44, p 443.
(42) Shiau, A. K.; Barstad, D.; Radek, J. T.; Meyers, M. J.; Nettles, K. W.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A.; Agard, D. A.; Greene, G. L. *Nat. Struct. Mol. Biol.* **2002**, *9*, 359.
(43) Nettles, K. W.; Bruning, J. B.; Gil, G.; Nowak, J.; Sharma, S. K.; Hahm, J. B.; Kulp, K.; Hochberg, R. B.; Zhou, H.; Katzenellenbogen, J. A.; Katzenellenbogen, B. S.; Kim, Y.; Joachimiak, A.; Greene, G. L. *Nat. Chem. Biol.* **2008**, *4*, 241.
(44) Bruning, J. B.; Parent, A. A.; Gil, G.; Zhao, M.; Nowak, J.; Pace, M. C.; Smith, C. L.; Afonine, P. V.; Adams, P. D.; Katzenellenbogen, J. A.; Nettles, K. W. *Nat. Chem. Biol.* **2010**, *6*, 837.
(45) Heldring, N.; Pawson, T.; McDonnell, D.; Treuter, E.; Gustafsson, J.-Å.; Pike, A. C. W. *J. Biol. Chem.* **2007**, *282*, 10449.
(46) McDonnell, D. P.; Clemm, D. L.; Hermann, T.; Goldman, M. E.; Pike, J. W. *Mol. Endocrinol.* **1995**, *9*, 659.
(47) Spyrakis, F.; Cozzini, P. *Curr. Med. Chem.* **2009**, *16*, 2987.
(48) Nettles, K. W.; Bruning, J. B.; Gil, G.; O'Neill, E. E.; Nowak, J.; Hughs, A.; Kim, Y.; DeSombre, E. R.; Dilis, R.; Hanson, R. N.; Joachimiak, A.; Greene, G. L. *EMBO Rep.* **2007**, *8*, 563.
(49) Bledsoe, R. K.; Montana, V. G.; Stanley, T. B.; Delves, C. J.; Apolito, C. J.; McKee, D. D.; Consler, T. G.; Parks, D. J.; Stewart, E. L.; Willson, T. M.; Lambert, M. H.; Moore, J. T.; Pearce, K. H.; Xu, H. E. *Cell* **2002**, *110*, 93.
(50) Suino-Powell, K.; Xu, Y.; Zhang, C.; Tao, Y.-g.; Tolbert, W. D.; Simons, S. S., Jr.; Xu, H. E. *Mol. Cell. Biol.* **2008**, *28*, 1915.
(51) Biggadike, K.; Bledsoe, R. K.; Coe, D. M.; Cooper, T. W. J.; House, D.; Iannone, M. A.; Macdonald, S. J. F.; Madauss, K. P.; McLay, I. M.; Shipley, T. J.; Taylor, S. J.; Tran, T. B.; Uings, I. J.; Weller, V.; Williams, S. P. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 18114.
(52) van Hoorn, W. P. *J. Med. Chem.* **2002**, *45*, 584.
(53) Tyulmenkov, V. V.; Klinge, C. M. *Arch. Biochem. Biophys.* **2000**, *381*, 135.

(54) Kuiper, G. G. J. M.; Enmark, E.; Pelto-Huikko, M.; Nilsson, S.; Gustafsson, J.-A. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 5925.

(55) Mosselman, S.; Polman, J.; Dijkema, R. *FEBS Lett.* **1996**, *392*, 49.

(56) Pike, A. C. W.; Brzozowski, A. M.; Hubbard, R. E.; Bonn, T.; Thorsell, A.-G.; Engstrom, O.; Ljunggren, J.; Gustafsson, J.-A.; Carlquist, M. *EMBO J.* **1999**, *18*, 4608.

(57) Manas, E. S.; Xu, Z. B.; Unwalla, R. J.; Somers, W. S. *Structure* **2004**, *12*, 2197.

(58) Mueller, G.; Kollenkirchen, U.; Kosemund, D.; Fritzemeier, K.-H., Preparation of 19-nor-17α-pregna-1,3,5(10)-trien-17β-ols containing a 21,16α-lactone ring with a preference for the estrogen alpha receptor. WO 2002026763.

(59) Peters, O.; Hillisch, A.; Thieme, I.; Elger, W.; Hegele-Hartung, C.; Kollenkirchen, U.; Fritzemeier, K.-H.; Patchev, V., Preparation of 8β-hydrocarbyl-substituted estratrienes for use as selective estrogens. WO 2001077139.

(60) Meyers, M. J.; Sun, J.; Carlson, K. E.; Marriner, G. A.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A. *J. Med. Chem.* **2001**, *44*, 4230.

(61) Fink, B. E.; Mortensen, D. S.; Stauffer, S. R.; Aron, Z. D.; Katzenellenbogen, J. A. *Chem. Biol.* **1999**, *6*, 205.

(62) De Angelis, M.; Stossi, F.; Carlson, K. A.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A. *J. Med. Chem.* **2005**, *48*, 1132.

(63) Fensome, A.; Bender, R.; Chopra, R.; Cohen, J.; Collins, M. A.; Hudak, V.; Malakian, K.; Lockhead, S.; Olland, A.; Svenson, K.; Terefenko, E. A.; Unwalla, R. J.; Wilhelm, J. M.; Wolfrom, S.; Zhu, Y.; Zhang, Z.; Zhang, P.; Winneker, R. C.; Wrobel, J. *J. Med. Chem.* **2005**, *48*, 5092.

THE DESIGN AND SYNTHESIS OF IMIDAZO[1,2-*a*]PYRIDINES TO INVESTIGATE THE ROLE OF THE CORE SCAFFOLD IN SELECTIVE ESTROGEN RECEPTOR MODULATORS

I. Introduction

The discovery of the ERβ subtype in 1996[1,2] served to reenergize research efforts in understanding the complex pharmacology of ERs. Much of this enthusiasm was driven by the differential tissue distributions of ERα and ERβ,[3] which many believe is the key to harnessing many "good" effects of ER stimulation (through ERβ) without activating cellular proliferation pathways (through ERα). Consequently, several high affinity, highly ERβ-selective compounds have been reported in the literature,[4] the design of which is discussed more extensively in Chapter 1. Very recently the high affinity ERβ-selective chloroindazole (Fig. 2.1), synthesized in our labs by Dr. Meri De Angelis,[5] has demonstrated unique potent neuroprotective activity in models of neurodegenerative diseases that is not observed for other ERβ-selective ligands such as DPN[6] and ERB-041.[7,8]
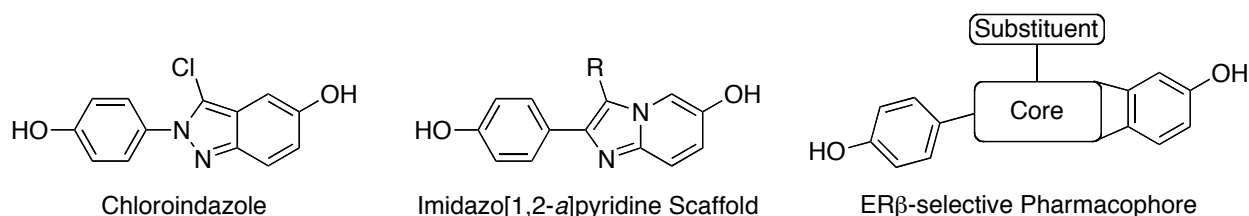
Figure 2.1. Using the chloroindazole as a reference, the indazole scaffold will be replaced by an imidazo[1,2-*a*]pyridine. Both of these structures adhere to the general pharmacophore of non-steroidal ERβ-selective ligands.

This chloroindazole compound originates from a series of ligands based on the indazole core scaffold and substituted in accordance with the pharmacophore for ERβ-selective ligands (Fig. 2.1; see also Chapter 1).[5] Varying the internal substituent profoundly affected both ER affinity and selectivity in *in vitro* binding assays. Incorporating a nitrile at this position yielded the highest binding affinity (RBA=69, β/α=18), while a chloro-substituent maximized affinity and ERβ-selectivity (RBA=32, β/α=107). Crystal structures of both the ethyl- (1.89Å, pdb accession code: 2QAB) and chloro-substituted (2.39Å, pdb accession code: 2QGW) indazoles bound to ERα[9] indicate a binding mode in which the pendant phenol serves as the A-ring mimetic, engaging the hydrogen bonding network between Glu353, Arg394, and a water molecule, while the heterocyclic hydroxyl group hydrogen bonds to His524 at the opposite end of the pocket. The internal substituent is directed towards the B-ring pocket. From this structural data, the basis behind the observed ERβ-selectivity is non-obvious; however, a similar binding orientation

in the ERβ pocket would position the core directly beneath the sidechain of the Met336 residue, implying a positive interaction similar to that proposed for ERβ-selectivity of DPN.[6,10,11] The presence of such an interaction would provide evidence towards clarifying the role of the scaffold core in affecting ligand binding affinity. Using this ERβ-selective indazole core as a reference, we set out to investigate the nitrogen isomer imidazo[1,2-a]pyridine as a scaffold replacement in order to probe how the identity of the heterocyclic core scaffold specifically affects ligand binding to ER and neuroprotective activity *in vivo*.

Many derivatives of imidazo[1,2-a]pyridines have previously been investigated as sedatives (e.g., Ambien®),[12] anti-malarial agents,[13] anti-implantation agents,[14] and the treatment of neurological disorders,[15,16] ulcers,[17] and anxiety disorders[18] among many other uses. While many derivatives of imidazo[1,2-a]pyridines exist in the literature, we are aware of only one report in which this chemotype has been directly utilized as an ER ligand.[19] This example, bearing 2-(4-hydroxyphenyl) and 3-(4-bromophenyl) substituents, bound at micromolar concentrations for ERβ and sub-micromolar for ERα, demonstrating modest affinity but ~30-fold ERα-selectivity.

In line with the established pharmacophore model (Fig. 2.1) and using the indazoles as a guide, we planned *p*-hydroxyphenyl as the A-ring mimetic, a second hydroxyl substituent at the 6- or 7-position to interact with His475, and a series of internal substituents including halo, alkyl, nitrile, and phenyl. Retrosynthetically, we envisioned the imidazo[1,2-a]pyridines to be formed through a condensation of a 2-aminopyridine with an α-bromoacetophenone. We imagined that the hydroxyl group on the heterocyclic ring would be installed via Ullmann-type coupling or S_NAr, which can be installed either before or after the condensation. Alkyl substituents at the 3-position would be installed by pre-functionalizing the α-bromoacetophenone via Friedel-Crafts acylation. We envisioned all other substituents would be introduced by taking advantage of the inherent reactivity of the parent imidazopyridine scaffold in selective electrophilic aromatic substitution reactions, and subsequent Pd-mediated coupling of the 3-halo intermediates.
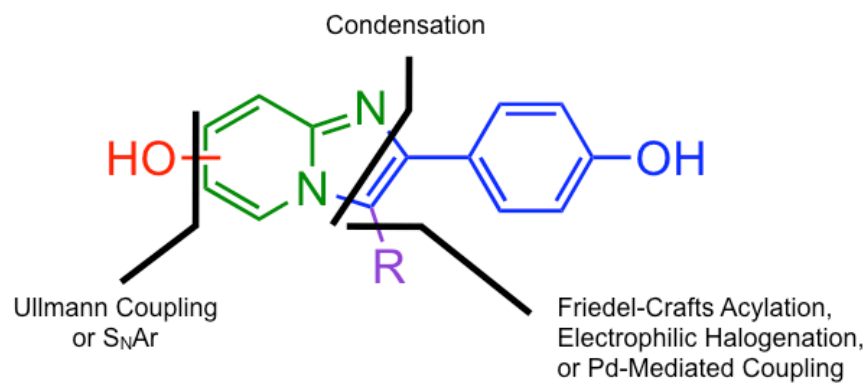
Figure 2.2. Retrosynthetic Analysis of Imidazo[1,2-a]pyridine Core and Planned Substitutions

16

II. Chemistry

A. Construction of Imidazo[1,2-a]pyridine Core

The unsubstituted parent imidazo[1,2-a]pyridine core was synthesized through the condensation of 2-aminopyridine and 2-bromoacetophenone according to a known procedure, as shown in Scheme 2.1.[20-22] Construction of the heterocylic core in this manner allowed for the easy installation of a 2-(4-methoxyphenol) substituent necessitated by our pharmacophore from commercially available 2-bromo-4'-methoxyacetophenone. Furthermore, the simple modification of the α-bromoketone to include an α-alkyl substituent represents a straightforward method for the installation of 3-alkyl substituents upon forming the core scaffold. The necessary ketones were either commercially available or could be synthesized via a Friedel-Crafts acylation of anisole[23] using the appropriate acyl chloride to yield intermediates **2a-d**. The subsequent bromination of each ketone was initially conducted by the addition of Br$_2$ in methylene chloride (DCM) followed by aqueous NaHCO$_3$ to furnish the desired α-bromoketones **3a,b** in moderate to excellent yield.[24] This reaction, however, was accompanied by the bromination of the electron-rich aryl ring, giving a byproduct that was not separable from the desired product at this stage. Accordingly, all subsequent brominations were conducted by

Scheme 2.1. Synthesis of Imidazo[1,2-a]pyridine Scaffold

1, 71%
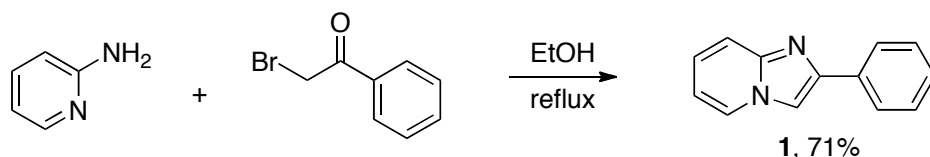
Scheme 2.2. Synthesis of 3-Alkyl Imidazo[1,2-a]pyridines

2a, R = Me, 97%
2b, R = Et, 100%
2c, R = n-Pr, 81%
2d, R = i-Pr, 100%

3a, R = Me, 67%
3b, R = Et, 97%
3c, R = n-Pr, 93%
3d, R = i-Pr, 90%

4a, R = H, 71%
4b, R = Me, 28%
4c, R = Et, 60%
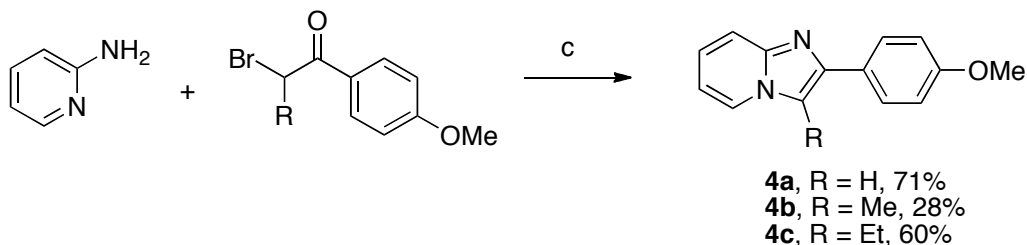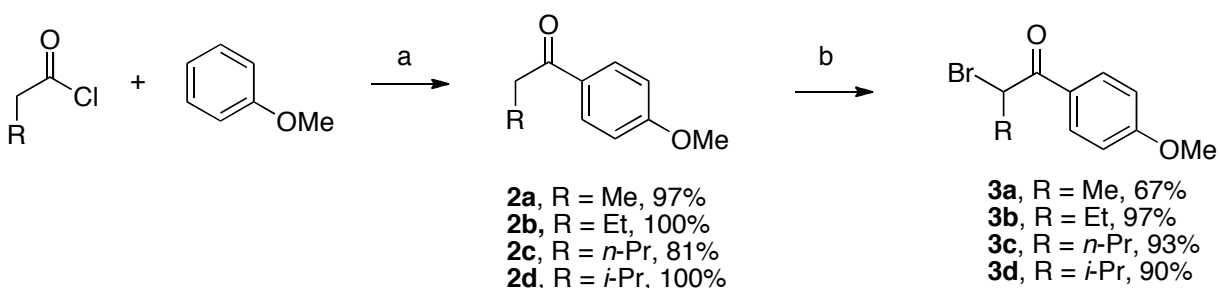
(a) AlCl$_3$, DCM. (b) CuBr$_2$, EtOAc/CHCl$_3$, reflux or 1) Br$_2$, 2) aq. NaHCO$_3$. (c) EtOH, reflux.

adding a solution of ketone dissolved in EtOAc into a solution of CuBr$_2$ refluxing in a mixture of EtOAc/CHCl$_3$.[25-27]   Under these conditions, **3a-d** were synthesized in good to excellent yields without any detection of the over-brominated product.   Each α-bromoketone was then condensed with 2-aminopyridine to yield **4a-c** in low to moderate yields.   The desired product was easily removed from any over-brominated impurities at this stage, where needed.

B. Installation of Protected 6-OH

The protected 6-OH functional group was installed as an alkyl ether via Ullmann-type coupling between the corresponding alcohol and aryl halide.   This reaction could be implemented at a number of steps highlighted by paths A, B, and C in Scheme 2.3, each stemming from 2-amino-5-iodopyridine (**5**) as a common precursor.   This compound was synthesized through the iodination of 2-aminopyridine with periodic acid and iodine in a mixture of refluxing aqueous acetic and sulfuric acids.[28]   The reaction was quenched with saturated aqueous sodium thiosulfate upon observation of 2-amino-3,5-diiodopyridine formation by TLC, achieving **5** in good yield.   Allowing further reaction time resulted in increased recovery of the diiodo-product and decreased yields of the desired product. Following path A, **5** was transferred to a pressure tube charged with CuI, 1,10-phenanthroline, and Cs$_2$CO$_3$ suspended in anhydrous methanol.[29] The tube was sealed using a teflon screw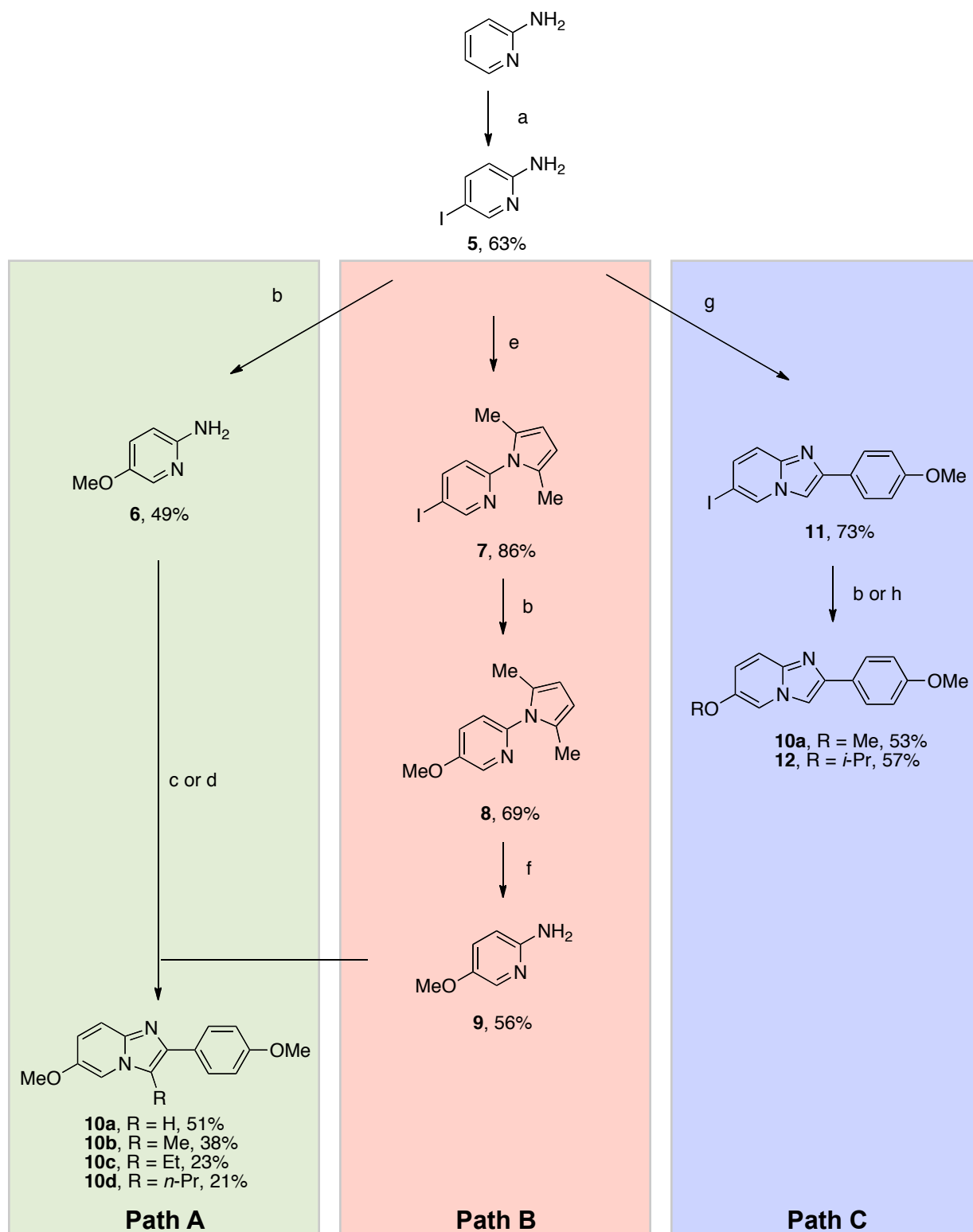 cap and heated to 110 °C for ~24 hours behind a weighted blast shield. The crude product was filtered and isolated using column flash chromatography to yield **6**. This material, however was substantially contaminated with a 5-H reduction side product that could not be separated at this stage. An alternative route, path B, utilized a pyrrole protection scheme[30-34] to facilitate the purification of the Ullmann coupling product.[34] Via this pathway, **5** was protected through a condensation with acetonyl acetone with a catalytic amount of *p*-TsOH in refluxing toluene or THF under dehydrating conditions using either a Dean-Stark apparatus or adding molecular sieves (4Å), respectively, to afford **7**.   The previously described Ullmann methoxylation conditions were used to install the 5-methoxy substituent.   The 5-H reduction product was removed using dry column vacuum chromatography (DCVC)[35,36] followed by recrystallization from hexanes to yield **8**.   The 2,5-dimethylpyrrole was removed by refluxing with hydroxylamine in an ethanol-water mixture to afford **9**. In the final step, paths A and B converge, whereby 2-amino-5-methoxypyridine (**6, 9**) was condensed with a series of α-bromoketones to yield the imidazo[1,2-*a*]pyridines **10a-d**.

Paths A and B represented divergent syntheses for imidazo[1,2-*a*]pyridines bearing alkyl substituents at the 3-position; however, we developed a more efficient method for the synthesis of imidazo[1,2-*a*]pyridine bearing the protected 6-OH, but lacking a 3-substituent.   Following path C, condensing **5** directly with 2-bromo-4'-methoxypyridine gave the fully formed core **11** directly in good yield.   This intermediate contained no acidic hydrogens or sensitive functional

Scheme 2.3. Synthesis of 6-Alkoxyimidazo[1,2-*a*]pyridines

5, 63%

6, 49%

7, 86%

11, 73%

8, 69%

10a, R = Me, 53%
12, R = *i*-Pr, 57%

9, 56%

10a, R = H, 51%
10b, R = Me, 38%
10c, R = Et, 23%
10d, R = *n*-Pr, 21%

Path A

Path B

Path C

(a) H_5IO_6, I_2, H_2SO_4, AcOH, H_2O, reflux. (b) CuI, 1,10-phenanthroline, Cs_2CO_3, MeOH, sealed tube, 110°C. (c) α-bromoketone, EtOH, reflux. (d) α-bromoketone, Et_3N, *t*-BuOH. (e) acetonyl acetone, THF + mol. sieves or toluene + Dean-Stark, reflux. (f) $NH_2OH \cdot HCl$, *aq.* KOH, EtOH/H_2O, reflux. (g) 2-bromo-4'-methoxyacetophenone, EtOH, reflux. (h) CuI, 1,10-phenanthroline, Cs_2CO_3, *i*-PrOH, sealed tube, 110°C.

groups and thus was established as a target for Ullmann-type alkoxylation. Carrying out this reaction under optimized reaction conditions, **10a** and **12** were achieved in good yield with <10% reduction product as determined by ^{1}H-NMR. This material was further purified by flash column chromatography followed by careful recrystallization from EtOAc/hexanes to yield **10a** containing ~3% 5-H impurity, or **12** containing no detectable amount of 5-H impurity by ^{1}H-NMR.

C. Installation of Protected 7-OH

Initially, we intended to install a protected 7-OH through an analogous reaction scheme proceeding through a 2-amino-4-iodopyridine intermediate, which was primed for Ullmann-type coupling. Literature procedures for synthesizing this intermediate installed the iodide through displacement of chloride from 4-chloro-2-picolinic methyl ester.[37-39] We recognized this as proceeding through an addition-elimination mechanism facilitated by the electron poor nature of the picolinic ester, and found that the chloride could also be displaced using alkoxide nucleophiles (Scheme 2.4). Unfortunately, the subsequent Curtius rearrangement used to install the 2-amino functionality appeared to be negatively affected by this replacement, as we never successfully achieved 2-amino-4-methoxypyridine through this scheme.

Scheme 2.4. Attempted Synthesis of 2-Amino-4-methoxypyridine via Curtius Rearrangement

13, 70% **14**, 41%

(a) 1) SOCl$_3$ (neat), Δ, 2) MeOH. (b) MeONa/MeOH.
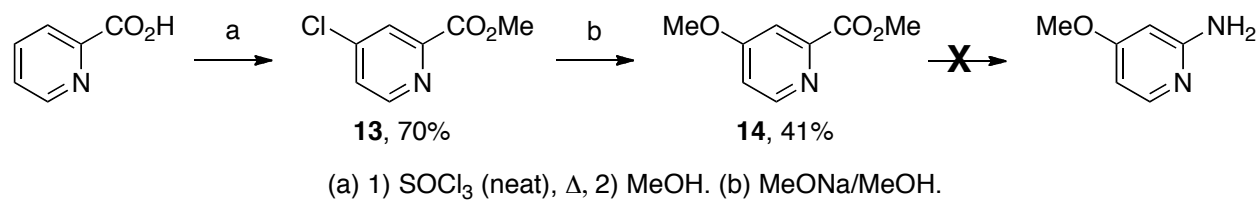
Scheme 2.5. Synthesis of 7-Alkoxyimidazo[1,2-*a*]pyridines

15, 87% **16**, 100%. **17**, 73%

18a, R = H, 64%
18b, R = Et, 56%

(a) acetonyl acetone, THF + mol. sieves or toluene + Dean-Stark, reflux. (b) MeONa/MeOH, reflux.
(c) NH$_2$OH•HCl, *aq.* KOH, EtOH/H$_2$O, reflux. (d) EtOH, reflux.
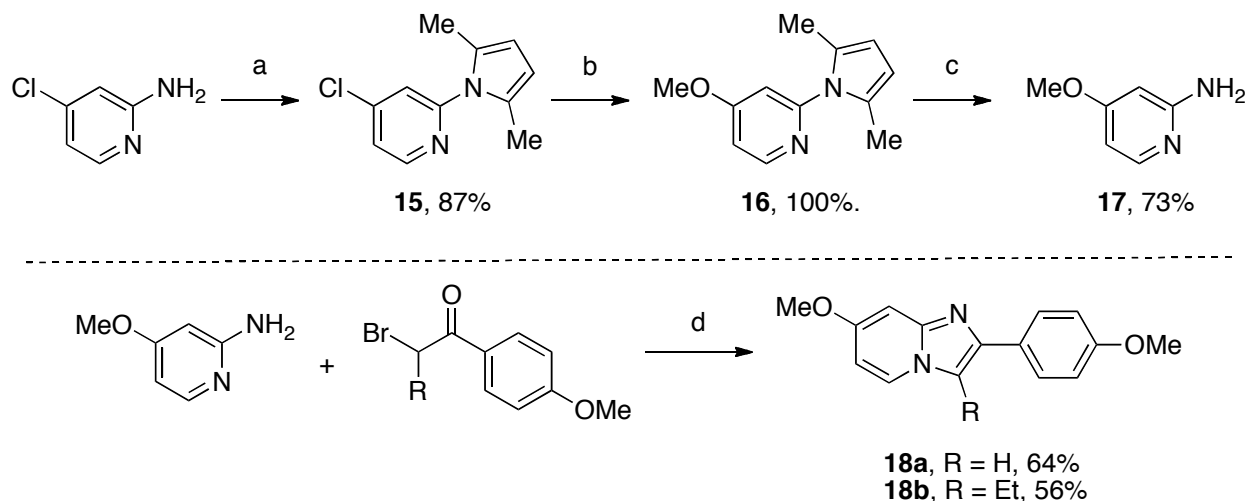
While these investigations were ongoing, 2-amino-4-chloropyridine became commercially at reasonable cost. We promptly discovered that by utilizing the 2,5-dimethylpyrrole protection scheme to yield intermediate **15**, we could take advantage of the electron poor position of the 4-chloro-substituent and access the same type of S$_N$Ar reactivity described above. **15** was treated with a freshly prepared solution of MeONa/MeOH to furnish **16** without further purification. The pyrrole protecting group was then removed to yield 2-amino-4-methoxypyridine (**17**) in excellent overall yield. Finally, **17** was condensed with a series of α-bromoketones to yield the imidazo[1,2-*a*]pyridines **18a,b**.

D. Functionalization of the 3-Position

Alkyl substituents at the 3-position were primarily installed by pre-functionalizing the α-bromoketones as described in Schemes 2.2 and 2.5. Additional substitution patterns, however, were accessed by taking advantage of the inherent reactivity of the imidazopyridine core.[22,40-42] The scaffold was easily halogenated in acetonitrile with the corresponding *N*-halosuccinimide[43-45] to give **21**-**23** in excellent yield. The substantial cost of NIS, however, precluded the scaled synthesis of **23**, which also served as a common intermediate for products derived from transition metal-catalyzed couplings. As an alternative procedure, **12** was stirred with iodine and potassium hydroxide in ACN[42] until the starting material was consumed, as monitored by TLC. The reaction was then quenched with aqueous sodium thiosulfate to give **23**.

Aryl iodides **23a-c** served as advanced common intermediates from which a variety of substitutions were accessible via transition metal catalyzed-coupling, namely palladium catalysis. To install a 3-cyano group, **23b**, Pd(dba)$_2$Cl$_2$, dppf, Cs$_2$CO$_3$, Zn(CN)$_2$, and zinc dust were suspended in DMA and heated to 130°C to give **24** after isolation and purification.[5,46-48] Both of the 3-phenyl and 3-cyclopropyl substituents were installed via Suzuki-Miyaura (SM) coupling[44,49,50] of the corresponding commercially available boronic acids in a suspension of Pd(dppf)Cl$_2$·CHCl$_3$ and Cs$_2$CO$_3$ in anhydrous THF, followed by heating to 130 °C in a sealed tube to yield **25** and **26**. This reaction required significant optimization which included screening the base, solvent, Pd-source, aryl halide, and temperature to arrive at optimal conditions favoring product formation and minimizing the 3-H reduction side product.

Additional carbon-carbon bond forming reactions at the 3-position could also be effected through specialized electrophilic reagents. Heating the parent core **12** with a dibenzothiophenium trifluoromethylating reagent[51-55] in anhydrous THF for 4 hours gave the desired product **27**. Attempts to optimize this reaction by changing solvent, temperature, or stoichiometry, adding a base, or employing metal-halogen exchange of a 3-haloimidazopyridine analog failed to achieve yields above 40%. The core was also formylated at the 3-position via

Scheme 2.6. Synthesis of Functionalized Imidazo[1,2-a]pyridines

19a, R=6-OMe, 57%
19b, R=6-O*i*-Pr, 73%

X	R = H	6-OMe	6-O*i*-Pr	7-OMe
Cl	**21a**, 86%	**21b**, 92%	**21c**, 66%	**21d**, 77%
Br	**22a**, 86%	**22b**, 81%	**22c**, 66	**22d**, 59%
I	**23a**, 94	**23b**, 95%	**23c**, 63%	**23d**, 54%

27a, R=6-OMe, 31%
27b, R=7-OMe, 41%

20a, R=6-OMe, 84%
20b, R=6-O*i*-Pr, 59%

24, R=6-OMe, 57%

25a, R'=Ph, R=6-OMe, 89%
25b, R'=Ph, R=6-O*i*-Pr, 78%
26a, R'=*c*-Pr, R=6-OMe, 82%
26b, R'=*c*-Pr, R=6-O*i*-Pr, 62%
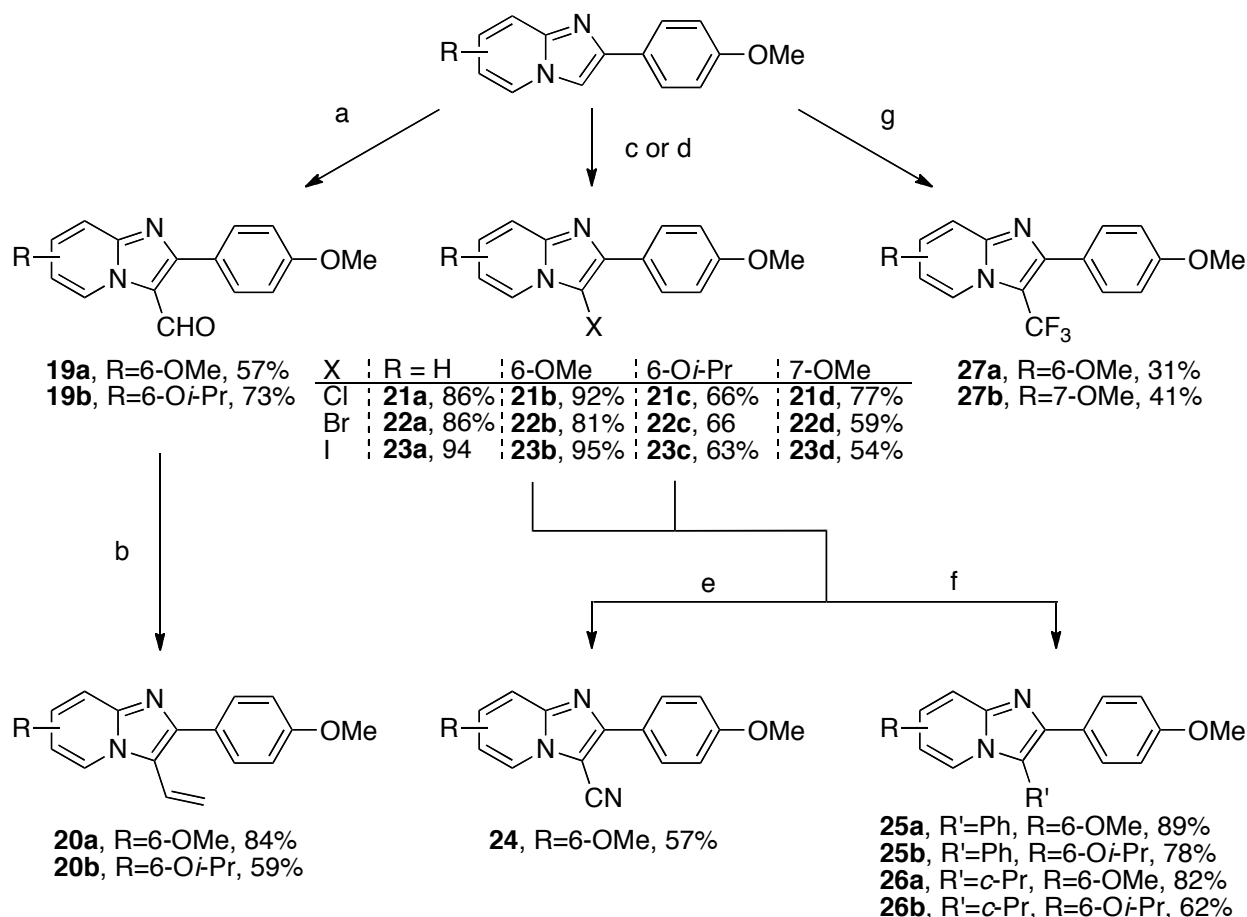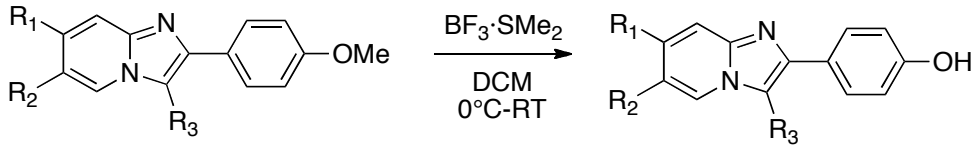
(a) POCl$_3$, DMF, 100°C. (b) Ph$_3$PMeBr, *n*-BuLi, THF, -78°C to RT. (c) NXS, ACN. (d) I$_2$, KOH, ACN. (e) Pd$_2$(dba)$_3$, dppf, Zn(CN)$_2$, Zn, DMA, 130°C. (f) Pd(dppf)Cl$_2$•CHCl$_3$, boronic acid, Cs$_2$CO$_3$, THF, sealed tube, 130°C. (g) 5-(Trifluoromethyl)dibenzothiophenium trifluoromethanesulfonate, THF, 60°C.

the Vilsmeier-Haack reaction.[56]   Thionyl chloride was heated in DMF for one hour followed by the addition of **12** dissolved in DMF and further heating for 2.5 hrs. Aqueous workup gave **19** after isolation and purification.   This intermediate was further manipulated to generate the 3-vinyl substituent by Wittig reaction, furnishing **20** in moderate yield.

E. Deprotection

The final step in all cases required the cleavage of each alkyl ether to yield the desired mono- or bis-hydroxyl product, and a number of reagents were employed to effect this transformation. The majority of compounds in Table 2.1 were deprotected using large excess of BF$_3$·SMe$_2$ (>50 equiv/ether) in CH$_2$Cl$_2$ overnight.[5] Quenching such a large excess of Lewis acid greatly complicated the aqueous workup and extraction of the desired product using common organic solvents.   In order to mitigate the amounts of Lewis acid required, we employed two equivalents of the more reactive reagent BBr$_3$.   In most cases, these conditions yielded

Table 2.1. Deprotection of Imidazo[1,2-*a*]pyridines

(reaction scheme: starting material with R_1, R_2, R_3 substituents and OMe group; reagents $BF_3 \cdot SMe_2$, DCM, 0°C-RT; giving product with R_1, R_2, R_3 and OH group)

Starting Material			Product				
Cmpd	R_1	R_2	R_1	R_2	R_3	Cmpd	Yield
4a	H	H	H	H	H	**28a**	50%
4b	H	H	H	H	Me	**28b**	100%
4c	H	H	H	H	Et	**28c**	70%
21a	H	H	H	H	Cl	**28d**	100%
22a	H	H	H	H	Br	**28e**	75%
23a	H	H	H	H	I	**28f**	48%
10a	H	OMe	H	OH	H	**29a**	94%
10b	H	OMe	H	OH	Me	**29b**	31%
10c	H	OMe	H	OH	Et	**29c**	83%
10d	H	OMe	H	OH	*n*-Pr	**29d**	90%
26b	H	O*i*-Pr	H	OH	*c*-Pr	**29e**	28%
21b	H	OMe	H	OH	Cl	**29f**	88%
22b	H	OMe	H	OH	Br	**29g**	27%
23b	H	OMe	H	OH	I	**29h**	43%
25b	H	O*i*-Pr	H	OH	Ph	**29i**	62%
24[a]	H	OMe	H	OH	CN	**29j**	36%
27a[b]	H	OMe	H	OH	CF₃	**29k**	16%
18a	OMe	H	OH	H	H	**30a**	93%
18b	OMe	H	OH	H	Et	**30b**	27%

[a] deprotected in neat pyridine·HCl at 220°C, [b] deprotected with BBr₃ in DCM at -78°C-RT

decomposition products; however, **27a** was successfully deprotected to give **29k**. Alternatively, compound **24** was deprotected using a pyridine·HCl melt at 220°C for 2 hours[57] to give **29j** in moderate yields. The deprotection of **25a** and **26a** proved difficult under each of the aforementioned conditions. Based on test reactions, we suspected that the methyl ether at the 6-position of the imidazopyridine was proving to be significantly more difficult to cleave than the methyl ether on the phenolic ring. Modifying our protection scheme at this position from a methyl ether to an isopropyl ether allowed for the use of significantly fewer equivalents of Lewis acid to effect cleavage (2.5-10 equiv $BF_3 \cdot SMe_2$/ether). During this optimization, we also found that buffering the aqueous solution to pH ~7.55 using KH_2PO_4/K_2HPO_4 enhanced the efficiency of product recovery through extraction with EtOAc. Despite multiple attempts using a variety of conditions, we were unable to isolate and purify the deprotected products for compounds **19**, **20**, **21-23a**, which we believe to be unstable based on physical observations.

F. Alternative Syntheses

The generation of the side product resulting from the reductive cleavage of the aryl-halide bond in the Ullmann-type coupling proved to be a substantial stumbling block in Scheme 2.3. Consequently, we sought out methods for installing the heterocyclic hydroxyl functionality that accessed a different reaction mechanism to avoid these impurities. The following strategies accomplish this task by changing the synthetic disconnect, such that the 6- and 7-hydroxyl substituents could easily be incorporated from commercially available *m*- or *p*-pyridinol.

In this fashion, we synthesized 2-amino-5-benzyloxypyridine from *m*-pyridinol via nucleophilic amination of an activated *N*-oxide.[58] The hydroxyl substituent of *m*-pyridinol was alkylated using benzyl bromide and NaH in DMF to give intermediate **31**. The pyridinyl nitrogen was subsequently oxidized to the *N*-oxide with *m*-CPBA in DCM, followed by treatment with tosyl chloride and *t*-butyl amine in toluene affording protected product **32**. Using the same methods, 2-amino-4-methoxypyridine (**17**) was synthesized from commercially available 4-methoxypyridine-*N*-oxide to furnish the isomer with the oxygen in the alternate position.

Scheme 2.7. Alternative Synthesis of Substituted 2-Aminopyridines
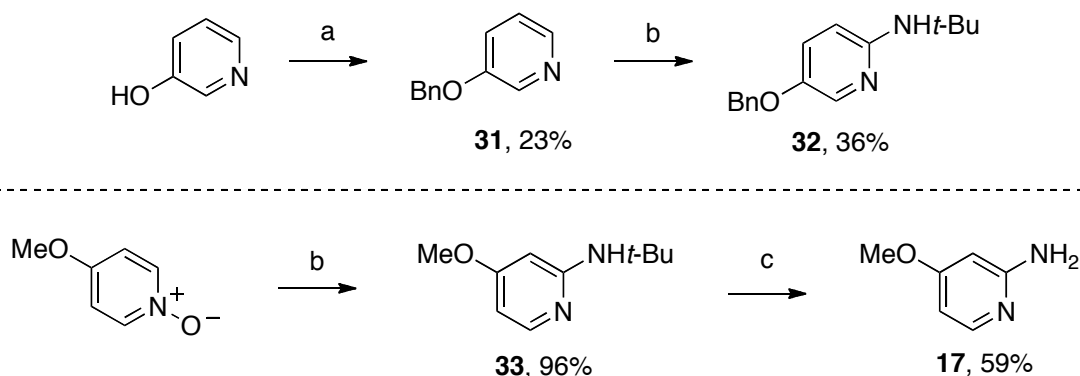
(a) NaH, BnBr, DMF. (b) 1) *m*-CPBA, DCM; 2) TsCl, *t*-BuNH$_2$, THF. (c) TFA, MeOH.

Literature precedent exists for an alternative route in the construction of the imidazo[1,2-*a*]pyridine scaffold through an intramolecular cyclization of a substituted phenacyl bromide *O*-methyloxime.[59] Following this scheme, the synthesis of **4a** proceeds through oxime formation by heating 2-bromo-4'-methoxyacetophenone with methoxyamine hydrochloride in methanol[60,61] to give intermediate **34** after isolation and purification. **34** was then reacted in neat pyridine, which after initial dissolution quickly yielded a light pink precipitate that was filtered and washed to afford **35** in near quantitative yield. We subsequently screened numerous conditions to facilitate the cyclization of this intermediate, and it was found that only NaH in anhydrous DMF yielded the desired product, albeit in only 14% yield. Based on the low yield of the cyclization reaction and uncertainty regarding the reactivity and regioselectivity of cyclization upon exchanging pyridine for *m*-pyridinol, this synthetic route was not pursued further.

Scheme 2.8. Alternative Construction of Imidazo[1,2-*a*]pyridine Core Scaffold
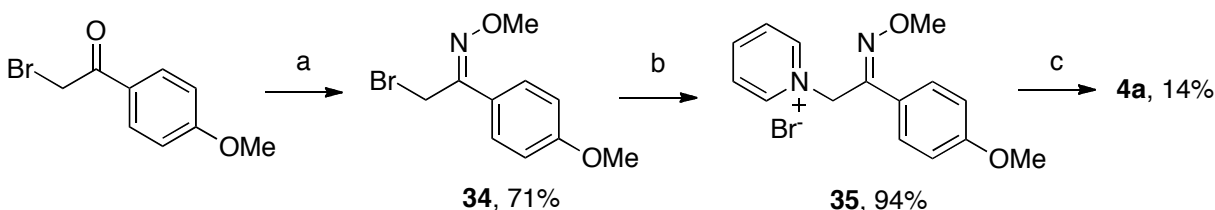
(a) NH$_2$OMe•HCl, KOH, MeOH, reflux. (b) neat pyridine. (c) NaH, DMF, 90°C.

III. *in vitro* Binding Affinity

The relative binding affinity (RBA) of each successfully deprotected imidazopyridine was determined by radiometric competitive binding assay, using 17β-[^{3}H]estradiol as the tracer.[62,63] This assay was performed by Kathryn Carlson, and the resulting data for both ERα and ERβ are given in Table 2.2 as a percentage relative to estradiol (E2=100%). These data can be easily converted to IC$_{50}$ or K$_i$ values, as described elsewhere.[4,64]

Table 2.2. Relative Binding Affinity of Imidazo[1,2-*a*]pyridines
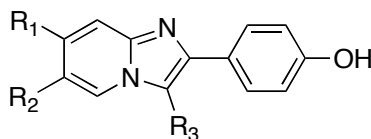
Entry	Compound	R$_1$	R$_2$	R$_3$	ERα	ERβ	β/α
1	**28a**	H	H	H	0.001	0.001	1
2	**28b**	H	H	Me	0.002	<0.001	<0.5
3	**28c**	H	H	Et	0.016	0.016	1.0
4	**28d**	H	H	Cl	0.009	0.014	1.60
5	**28e**	H	H	Br	0.015	0.012	0.80
6	**28f**	H	H	I	0.022	0.019	0.86
7	**29a**	H	OH	H	<0.002	0.006	>3
8	**29b**	H	OH	Me	0.017	0.347	20
9	**29c**	H	OH	Et	0.031	0.920	30
10	**29d**	H	OH	*n*-Pr	0.017	0.313	18
11	**29e**	H	OH	*c*-Pr	0.015	0.324	22
12	**29f**	H	OH	Cl	0.021	0.866	41
13	**29g**	H	OH	Br	0.006	0.099	17
14	**29h**	H	OH	I	0.016	0.058	3.6
15	**29i**	H	OH	Ph	0.054	1.93	36
16	**29j**	H	OH	CN	0.111	3.44	31.0
17	**29k**	H	OH	CF$_3$	0.893	13.51	15.1
18	**30a**	OH	H	H	0.006	0.043	7
19	**30b**	OH	H	Et	0.039	0.025	0.64

IV. Structure-Based Modeling

Structural models of each receptor-ligand pair were constructed *in silico* by docking each compound into the ligand binding pocket of both the ERα and ERβ receptor subtypes. Several high resolution crystal structures for each receptor subtype were available in the PDB; we selected ERα complexed with a chloroindazole (2.39 Å resolution, accession code 2QGW[9]) and ERβ complexed with a benzofuran (2.4 Å resolution, accession code 1U9E[11]) due to the geometric similarities of the ligands as 5,6-fused bicyclic cores containing heteroatoms. All docked ligands were constructed using the Molecular Operating Environment (MOE)[65] and docked into the protein structure using AutoDock Tools (ADT)[66] and AutoDock Vina (ADV).[67] Acceptable low energy poses were selected using the lowest energy ADV score, and in the case of ERβ structures, additional low energy poses were manually selected by visual analysis. The selected poses were further minimized using a multistep protocol and analyzed using a battery of commercial scoring functions and custom-built numerical analyses.

Visual analysis of the lowest energy poses as determined by using ADV's built-in scoring function yielded contrasting results for each of the receptor subtypes. Overlaying the output for each ligand docked in ERα (Fig. 2.3A) suggests that all ligands bind in the same orientation and adopt strikingly similar poses. In each pose, the pendant phenol is directed towards a hydrogen bonding network between Glu353, Arg394, and a crystallographic water, while the heterocyclic hydroxyl group is positioned within hydrogen bonding distance of His524, a known hydrogen

Figure 2.3. The lowest energy poses for ligands docked into ERα (**A**), all demonstrate identical binding orientation. The same ligands docked into ERβ (**B**), however, suggested different binding orientations based on the identity of the 3-substituent. The five lowest energy docking poses for imidazopyridine **29k** bearing a 3-CF$_3$ (**C**), for example, yields two opposing binding orientations within 0.1 kcal/mol as scored by the docking routine.

bond acceptor, and the 3-substituent is directed towards the B-ring pocket of the binding volume. These binding poses closely mirror the positioning of the chloroindazole present in the initial crystal structure. The corresponding lowest energy poses for ligands docked into ERβ, however, show multiple binding orientations (Fig. 2.3B) within the binding volume. Upon further analysis of all docking output for ERβ, we inferred that multiple binding orientations could not be eliminated based on ADV scoring.   The docking output for 2-hydroxyphenyl-3-trifluoromethylimidazo[1,2-*a*]pyridin-6-ol (**29k**) serves as a particularly illustrative example. An overlay of the five lowest energy poses is shown in Figure 3C, which represents two unique binding orientations with ADV calculated binding affinities within 0.1 kcal/mol.   To avoid misidentifying the correct binding orientation in ERβ, we selected all unique low energy poses for followup analysis.   Based on the relatively low binding affinities observed for all imidazopyridines in ERα (Table 2.2), we chose to focus on identifying a numerical scoring function that correlates to the observed RBA for ligands in ERβ.

The inability to predict the correct binding orientation in ERβ with any certainty obscures the identification of a clear molecular basis for binding affinity or subtype-selectivity based on visual analysis alone.   Accordingly, we sought out a quantitative method for analyzing all potential binding modes for each receptor-ligand complex, with the goal of identifying a metric or series of metrics, that would correlate to the observed RBA data. In addition to the ADV scoring output, we re-scored the raw docking output (poses prior to multistep minimization) using 8 additional scoring algorithms implemented in FRED[68-73] (Table 2.3)   We also applied methods for analyzing receptor-ligand contacts that we had previously developed to assess the binding affinity for ligands docked into the progesterone receptor (see Chapter 4).[74]   Focusing on receptor-ligand interaction energies, approximated binding energies, and scoring hydrogen bonding contacts, these methods were applied to both the raw docking output and the minimized structures (Tables 2.4 and 2.5). Unfortunately, in all cases we were unable to predict even the relative order of binding affinity based on these data.

V. Ligand-Based Modeling
The difficulties encountered during our attempts at utilizing receptor-based methods described above, speaks to the complexity of the molecular system under investigation and the variety of interactions that drive ligand binding. In order to apply a more rigorous computational analysis, we drastically simplified the molecular system to consider only the ligand itself so as to identify the major differences in electronic properties. A geometry optimization was performed for each ligand using closed-shell Restricted Hartree-Fock (RHF) gas phase calculations at the 6-31G* level of theory, followed by a single point energy calculation using DFT B3LYP

Table 2.3. ADV Scoring and FRED Re-Scoring of Manually Selected Docked Poses

Compound	R	pose #	ADV score	Consensus	Shapegauss	PLP	C-gauss2	C-gauss3	OE Chem	Screen	Zapbind
29a	H	2	-8.5	295	-368.0	-57.7	-40.9	-76.9	-38.2	-132.4	-13.2
29b	Me	1	-9	170	-389.6	-65.1	-44.2	-82.8	-41.1	-140.7	-15.7
29c	Et	1	-8.9	122	-406.9	-63.5	-44.8	-85.4	-42.7	-142.0	-15.9
29c	Et	8	-8.8	252	-383.6	-59.6	-44.7	-84.2	-37.8	-114.7	-14.6
29d	n-Pr	1	-8.6	129	-422.3	-62.9	-45.7	-88.5	-37.0	-134.0	-22.4
29e	c-Pr	1	-8.6	163	-413.7	-61.9	-44.6	-85.1	-37.5	-129.1	-24.0
29f	Cl	1	-9	192	-390.5	-65.0	-44.1	-83.0	-39.0	-139.2	-14.5
29g	Br	1	-8.9	205	-387.0	-62.7	-43.1	-83.3	-38.2	-141.7	-15.9
29g	Br	4	-8.9	308	-369.6	-58.5	-43.0	-82.3	-32.9	-119.8	-13.0
29h	I	1	-9	168	-400.9	-63.9	-44.3	-84.4	-38.3	-142.1	-14.7
29h	I	4	-8.9	310	-380.6	-58.6	-43.4	-81.0	-32.9	-116.0	-12.8
29i	Ph	1	-9.5	51	-476.8	-64.1	-48.2	-94.4	-44.1	-137.2	-20.1
29j	CN	1	-9.1	202	-401.0	-60.9	-44.0	-80.2	-40.1	-141.3	-15.3
29j	CN	4	-9.1	259	-392.2	-56.5	-44.9	-83.2	-35.0	-124.9	-14.0
29k	CF_3	1	-8.8	205	-398.7	-62.7	-44.3	-85.4	-32.6	-134.0	-17.8
29k	CF_3	4	-8.7	205	-397.8	-55.4	-46.0	-85.2	-35.9	-123.6	-19.8

28

Table 2.4. Interaction and Binding Energies of Receptor-Ligand Complex

Compound	R	pose #	Raw Docking Poses[a]		Minimized Docking Poses[b]	
			E_int^b	E_bin^c	E_int^b	E_bin^c
29a	H	2	-8.4	-215.9	-36.5	-252.9
29b	Me	1	7.6	-186.1	-35.9	-249.5
29c	Et	1	2.0	-187.4	-38.9	-252.7
29c	Et	8	18.6	-184.7	-55.3	-268.4
29d	*n*-Pr	1	-8.3	-203.7	-44.4	-253.9
29e	*c*-Pr	1	-4.3	-169.2	-41.9	-260.5
29f	Cl	1	-11.5	-216.3	-60.2	-273.6
29g	Br	1	-11.2	-204.8	-38.1	-251.1
29g	Br	4	-6.0	-215.0	-56.5	-268.8
29h	I	1	-9.2	-197.8	-58.2	-268.7
29h	I	4	-2.0	-207.5	-55.4	-269.1
29i	Ph	1	-0.1	-207.6	-54.1	-264.6
29j	CN	1	-10.1	-211.6	-61.3	-275.1
29j	CN	4	-5.3	-215.8	-56.8	-270.4
29k	CF_3	1	8.3	-198.5	-51.1	-261.8
29k	CF_3	4	-36.5	-244.2	-59.1	-272.1

[a] Raw docking poses were merged back into the receptor structure and rescored. [b] Merged docking poses were minimized using a multistep minimization routine prior to rescoring. [c] Interaction Energy = E_{RL} - (E_R + E_L). [d] Binding Energy = E_{RL} - (E_{Rmin} + E_{Lmin}).

Table 2.5. Hydrogen Bonding Scores for each Receptor-Ligand Complex

Compound	R	pose #	Raw Docking Poses[a]					Minimized Docking Poses[b]					
			GLU305	ARG346	H_2O	HIS524	a_mean[c]	GLU305	ARG346	H_2O	HIS524	a_mean[c]	g_mean[d]
29a	H	2	0.1268	0.2165	0.0303	---	0.1245	0.0176	0.2692	0.0185	---	0.1018	0.0444
29b	Me	1	0.0354	0.0236	---	---	0.0295	0.0176	0.1191	0.0132	---	0.0500	0.0303
29c	Et	1	0.0105	---	---	0.0238	0.0171		0.0510	0.0247	0.0048	0.0269	---
29c	Et	8	0.1525	0.2014	0.0288	---	0.1276	0.7814	0.6820	0.0745	---	0.5126	0.3411
29d	Pr	1	---	---	---	0.1083	0.1083	---	---	---	0.2765	0.2765	---
29e	c-Pr	1	0.1268	0.1669	0.0222	---	0.1053	---	---	---	0.3589	0.3589	---
29f	Cl	1	0.1888	0.2014	0.0239	---	0.1380	0.6406	0.4944	0.0341	---	0.3897	0.2210
29g	Br	1	0.1307	0.2110	0.0294	---	0.1237	0.0181	0.1959	0.0532	---	0.0891	0.0573
29g	Br	4	0.0121	---	---	0.0205	0.0163	0.7878	0.6737	0.0913	---	0.5176	0.3646
29h	I	1	0.0105	---	---	0.0209	0.0157	0.6624	0.4948	0.0322	---	0.3964	0.2193
29h	I	4	0.0945	0.2013	0.0340	---	0.1099	0.7698	0.6914	0.0717	---	0.5110	0.3367
29i	Ph	1	---	---	---	0.0460	0.0460	0.7277	0.6175	0.0564	---	0.4672	0.2937
29j	CN	1	0.0149	---	---	0.0183	0.0166	0.6628	0.5184	0.0392	---	0.4068	0.2379
29j	CN	4	0.0869	0.2142	0.0352	---	0.1121	0.7765	0.6360	0.0785	---	0.4970	0.3384
29k	CF_3	1	---	---	---	0.0419	0.0419	0.0084	0.1109	0.0671	---	0.0621	0.0398
29k	CF_3	4	0.1391	0.0429	---	0.0005	0.0608	0.6026	0.4557	0.0225	---	0.3602	0.1835

[a] Raw docking poses were merged back into the receptor structure and rescored. [b] Merged docking poses were minimized using a multistep minimization routine prior to rescoring. [c] Arithmetic Mean. [d] Geometric Mean.

functionals at the 6-311+G** level of theory.[75]   The calculations were performed in both GAMESS[76] and Spartan'08,[77] and the output was further analyzed as described in the following sections.

*A. Proton Affinity/Relative Basicity*

Simple pK_a approximations (see Section VI, Fig. 2.9) backed by experimental observations indicates that the 1,2-configuration of nitrogen atoms in the indazole structures yields a much less basic heterocyclic core relative to the imidazopyridine core. To further probe this difference in a more rigorous fashion, we calculated two measures of proton affinity based on the quantum mechanical calculations described above.

The first of these is a calculation of the energy of proton transfer in an isodesmic reaction between each compound and imidazole (Scheme 2.9), which will be denoted as E(PT). [75] The energy associated with proton transfer, E(PT), was calculated from eq. **(2.1)** and is given in Table 2.6.   Favorable reaction energies, negative values for E(PT), indicate that the compound is more basic than imidazole, leading to proton transfer; the reverse is true for positive values of the E(PT) calculation. The experimental pK_a for imidazole in water is 6.993,[78] conveniently close to neutral. For all cases involving the imidazopyridine cores, the protonated state is favored, while two out of 8 indazole compounds disfavored protonation. In all cases, protonation of the indazole compound was significantly less favored relative to the correseponding imidazopyridine analogue.

An important limitation of the E(PT) calculation is that the indazole core is less chemically similar to imidazole compared to imidazo[1,2-*a*]pyridine, and therefore, potentially introduces error into the calculation as the reaction is no longer isodesmic. In order to address this shortcoming, we additionally calculated the gas phase proton affinity, P(A),[79,80] for each of the compounds listed in Table 2.6. This calculation follows the form given in Scheme 2.10, in which the change in energy for the protonation reaction is calculated directly from eq. **(2.2)**, and is also given in Table 2.6. While the P(A) is an absolute energy (e.g., not relative to a reference compound), an additional calculation to define the P(A) relative to imidazole allows a direct comparison to the E(PT) calculations.   Unsurprisingly, these data also indicate that the imidazopyridine core is substantially more basic than the indazole analogues. Furthermore, the relative P(A) values are in good agreement with the E(PT) calculations.

B. Electrostatic Potential

The electrostatic potential for selected imidazopyridines and the corresponding indazoles was projected on the 0.02 au density isosurface, as shown in Figures 2.4 and 2.5. This visual technique is generally suitable for visualizing the charge distribution for each

Scheme 2.9. Isodesmic Proton Transfer Between an Imidazo[1,2-*a*]pyridine and Imidazole
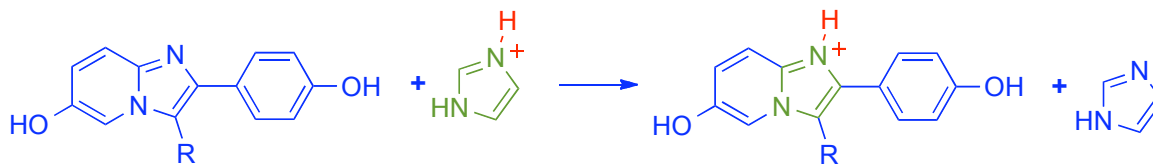
$$E(PT) = (E_{cmpdH^+} + E_B) - (E_{cmpd} + E_{BH^+}) \qquad \textbf{(2.1)}$$

Scheme 2.10. General Reaction Scheme for Proton Affinity
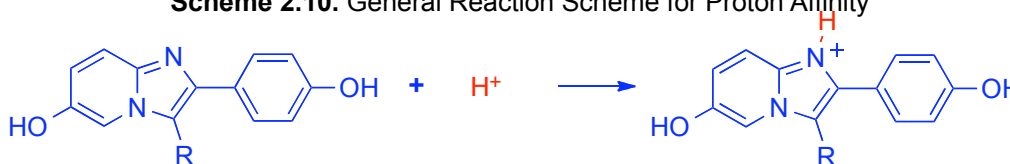
$$P(A) = -\Delta H = -\Delta E_{ele} - \Delta ZPE + (5/2)RT \qquad \textbf{(2.2)}$$

Table 2.6. Energy of Proton Transfer (E[PT]), Proton Affinity (P[A]), and pKa

R	Imidazo[1,2-*a*]pyridine				Indazole			
	E(PT)[a]	P(A)[b]	Rel. P(A)[c]	pKa[d]	E(PT)[a]	P(A)[b]	Rel. P(A)[c]	pKa[d]
H	-12.9	-238.1	-13.3	6.11	-3.7	-229.6	-4.8	0.95
Me	-16.2	-241.5	-16.7	6.63	-9.2	-235.5	-10.8	1.44
Et	-17.1	-242.3	-17.5	6.54	-10.3	-236.3	-11.5	1.39
Pr	-17.5	-242.7	-17.9	6.52	-10.8	-236.8	-12.0	1.37
c-Pr	-17.2	-242.5	-17.7	6.42	-9.4	-235.4	-10.6	1.31
Cl	-9.1	-234.5	-9.7	4.87	-2.5	-228.5	-3.7	0.54
CN	-2.1	-227.6	-2.8	3.73	8.1	-217.9	6.9	0.22
CF$_3$	-6.8	-232.0	-7.3	4.75	8.9	-217.2	7.6	0.49

[a] Relative basicity calculated according to Eq. **(2.1)** using imidazole as the reference base. [b] Proton affinity calculated according to Eq. **(2.2)**. [c] Relative to imidazole. [d] pKa calculated using MarvinSketch

compound.   Within each scaffold type, the nature of the 3-substituent appears to play a significant role in shifting the charge distribution relative to the parent compound (i.e., H). In the case of alkyl substituents (e.g., Et), electron density is donated from the substituent to the heterocyclic core where it is mostly denoted by larger electron density on N1.   Electron withdrawing substituents (e.g., Cl, CF_3) indeed appear to reduce the charge associated with N1 and increase the charge localized to the substituent.   Comparing the electrostatic potential between the two scaffolds also indicates a difference in charge distribution.   In the imidazopyridine series, the majority of the charge density is located in the pendant phenol, while it is more evenly distributed but generally more localized to the heterocyle core in the indazole series.

C. HOMO and LUMO

A similar analysis to the electrostatic potential was performed for the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) for each set of ligands. The effect of electron donating substituents at the 3-position appears to have minimal effect on the size of the HOMO; however, the electron withdrawing groups serve to reduce the size of the HOMO on the heterocycle, particularly on N1, by an appreciable amount. Between the two heterocycle core types the HOMOs differ extensively. The orbitals associated with the pendant phenol are substantially larger for the imidazopyridines series relative to the indazoles. The reverse is true for the orbitals located on the heterocyclic core, and the arrangement and shape of the orbitals differ as well. The arrangement of the LUMO within each series failed to yield any generally insightful characteristics.   Between each heterocycle series, the LUMO associated with the pendant phenol was noticeably larger for each indazole relative to the imidazopyridines.   The size, shape, and arrangement were similar for the LUMOs of the heterocyclic core.

D. Dipole Moment

The dipole moment was calculated for all compounds and a representative selection is shown in Figures 2.5 and 2.6. This quantity is particularly convenient as it simplifies the various elements of charge distribution typically visualized as surfaces (e.g., electrostatic potentials, HOMO, LUMOs) into a discrete vector quantity. It is clear from visual analysis that the direction of the vector is heavily dependent on the identity of the 3-substituent. Electron donating groups (EDGs) at the 3-position direct the vector away (e.g., Me, Et, Pr, *c*-Pr, *i*-Pr), while electron withdrawing groups (EWGs) cause the vector to align more along the principal axis of the structure (e.g., Cl, Br, CF_3), or even back towards the 3-position (e.g., CN; data not shown). These observations are in line with the previously discussed effect of substituent identity on

33

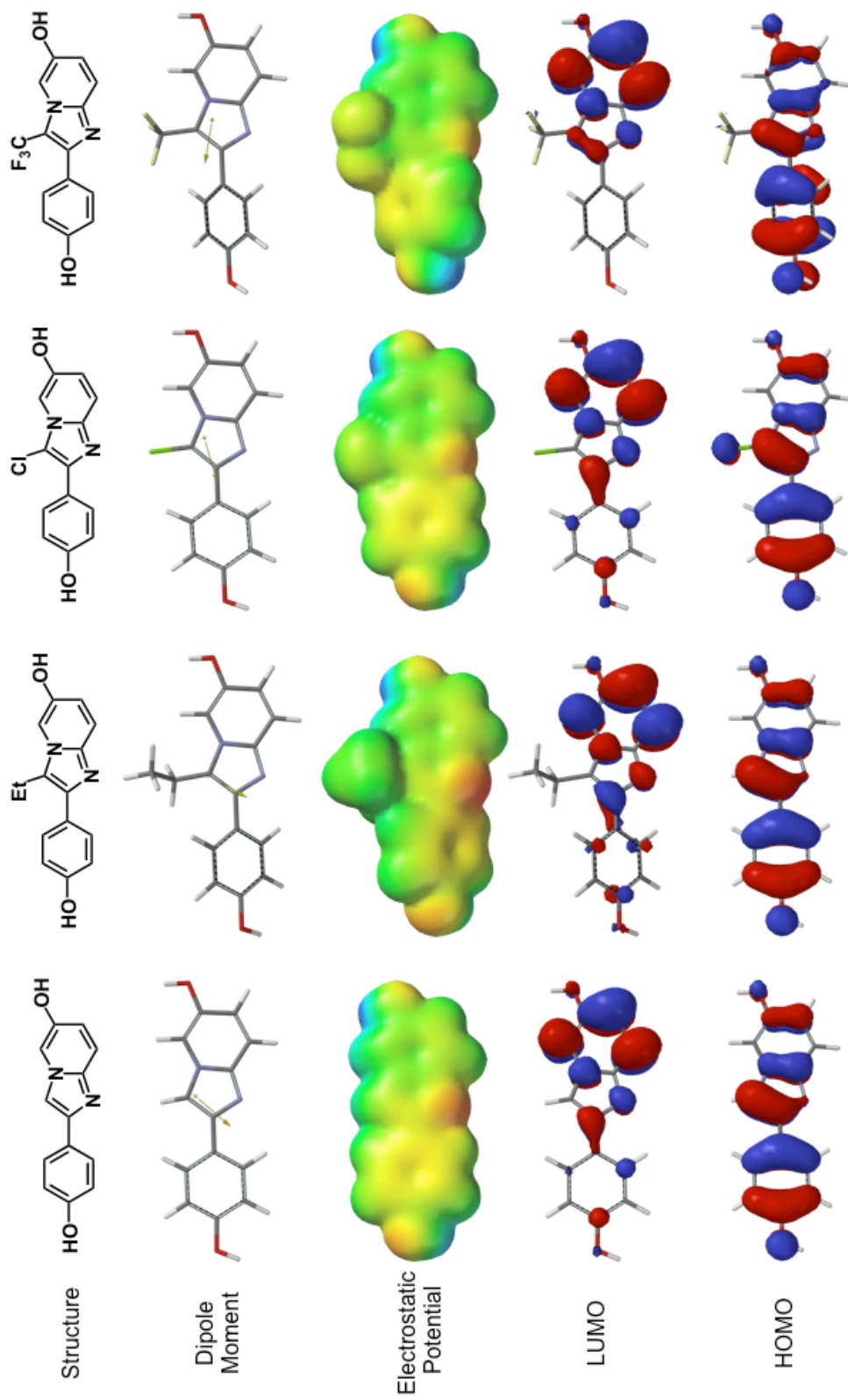**Figure 2.4.** Structure, Dipole Moment, Electrostatic Potential, HOMOs, and LUMOs for Representative Imidazo[1,2-a]pyridines.
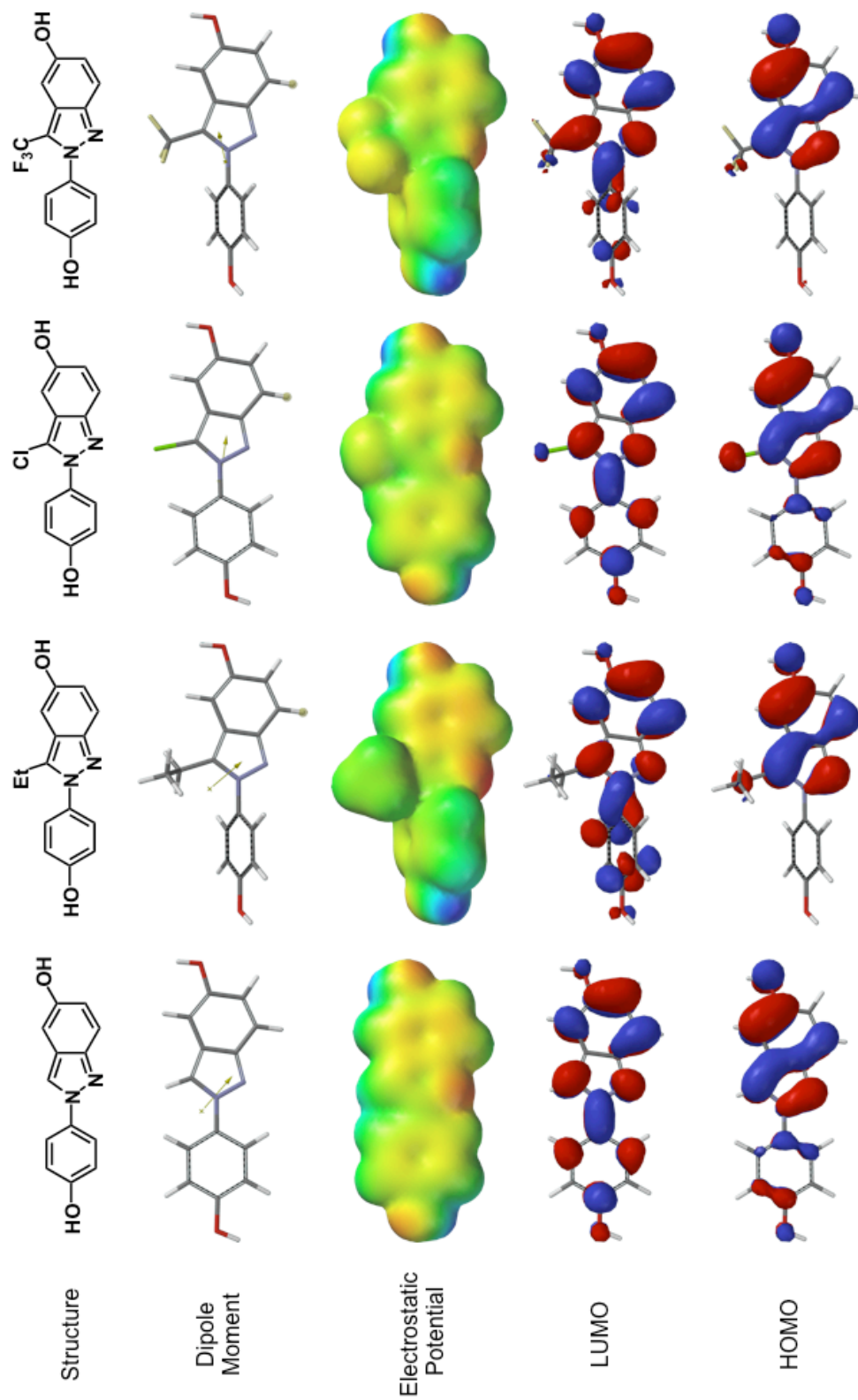
34

Figure 2.5. Structure, Dipole Moment, Electrostatic Potential, HOMOs, and LUMOs for Representative Indazoles.

35

proton affinity and charge distribution. Visual comparison of the dipoles for imidazo[1,2-a]pyridines relative to the analogous indazoles, indicates that the dipole is roughly reflected across a vertical plan perpendicular to the principal axis of the ligand. The magnitude of each dipole moment ranged from 2.08-3.52 Debye for the imidazo[1,2-a]pyridines, and 0.98-4.79 Debye for the corresponding indazole analogs.

Motivated by the opposing dipoles between scaffolds, we calculated the local dipole moment of the ligand binding pocket. All residues within 4.5 Å of the benzofuran ligand were extracted from the 2QGW crystal structure,[11] and the backbone carboxyls were capped as amides. Based on the number of heavy atoms involved (190 heavy atoms, 244 light atoms), rigorous *ab initio* calculations (e.g., DFT, RHF-MP2) were prohibitively costly. Alternatively, we used three different semi-empirical methods (AM1, PM3, MNDO) for single point energy calculations, from which the dipole moment was taken. All three methods were in agreement regarding the direction of the dipole, as shown in Figure 2.6, and in the magnitude of the vector, measured at 36.97, 37.80, and 35.44 Debye for AM1, PM3, and MNDO, respectively.
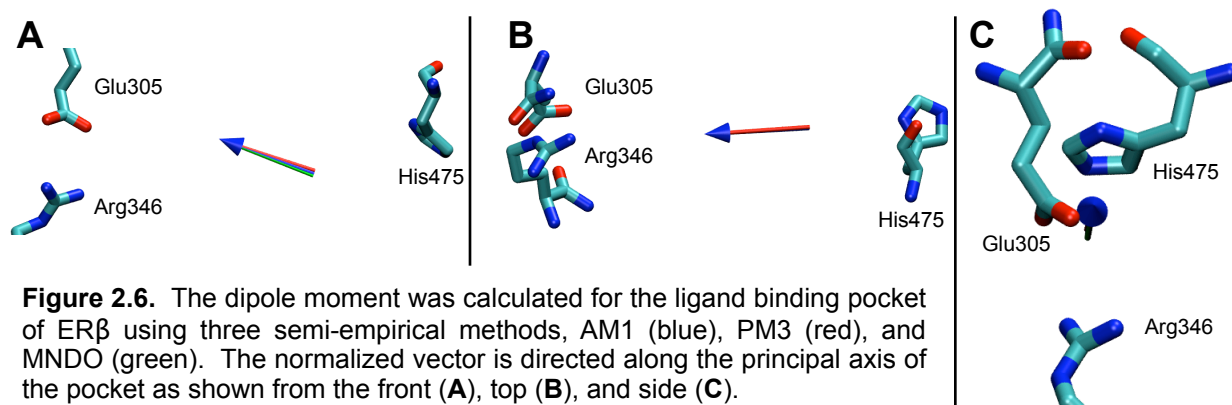

Figure 2.6. The dipole moment was calculated for the ligand binding pocket of ERβ using three semi-empirical methods, AM1 (blue), PM3 (red), and MNDO (green). The normalized vector is directed along the principal axis of the pocket as shown from the front (**A**), top (**B**), and side (**C**).

VI. Discussion

A. Chemical Synthesis

With the exception of a few examples involving specialized substitution patterns,[22,59,81,82] imidazo[1,2-a]pyridines are typically synthesized through the condensation of a 2-aminopyridine with an α-bromoketone,[20,21] and this route generally proceeds in moderate to good yields. The reaction generates an equivalent of HBr, which necessitated an aqueous workup under basic conditions (e.g., saturated *aq.* NaHCO$_3$, 1M KOH) in order to effect efficient extraction into organic solvents. The increasing concentration of HBr as the reaction proceeds does not appear to affect the overall yield of the isolated condensation product, as the addition of a base (e.g., Et$_3$N, Na$_2$CO$_3$) to the reaction mixture did not result in markedly increased yields. Substitution of the 2-aminopyridine (4-Cl, 4-OMe, 5-Br, 5-I, 5-OMe) was generally well tolerated; however, the reaction was sensitive to substitution at the 2-position of the α-bromoketone, as evidenced by reduced yields with increasing steric bulk as indicated in Scheme 2.3. The most

drastic case was the failure of the condensation reaction when R=*i*-Pr, even under forcing conditions such as extended reaction times, elevated temperatures, and the addition of silver-based additives to facilitate displacement of the bromide.

The installation of the hydroxyl functional group on the heterocyclic core proved challenging. Based on the binding data for the indazole analogues,[5] we initially focused on positioning the hydroxyl at the 6-position via an Ullmann-type alkoxylation. A known drawback to Ullmann-type coupling is the generation of a reduction side product in which the aryl-halide bond is cleaved.[83-89] In preliminary attempts at the coupling of methanol with either 2-amino-5-iodopyridine (**5**) or 6-iodo-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (**11**) through paths A and C, this side product was observed in quantities up to 20% under conventional conditions (i.e., CuCl, MeONa/MeOH), and was inseparable from the desired product using common chromatographic methods. Optimization of the reaction conditions,[86,87,90-92] including implementing conditions developed by Buchwald,[29,93] and using freshly distilled or commercially available anhydrous methanol, were capable of reducing the percentage of side product to ~5-10%, but did eliminate it completely (Path C, R=6-OMe).

Based on previous work by Lützen and coworkers,[94] we found that the use of a 2,5-dimethylpyrrole protection scheme was necessary to facilitate purification and was successfully leveraged to remove all traces of the 5-H impurity via path B. Cleavage of the pyrrole to unveil the desired intermediate **9**, however, proved operationally difficult due to the greatly increased polarity engendered by the newly installed methoxy-substituent. Consequently, we considered a number of alternative protection schemes with more favorable deprotection strategies. Benzyl protecting groups (i.e., Bn, PMB) could not be hydrogenolyzed, as the benzyl-nitrogen bond is substantially more difficult to cleave than the analogous benzyl-oxygen bond. Furthermore, conditions to effect cleavage of PMB groups through a quinone-methide intermediate using Lewis acids, would also cleave the newly installed methyl-aryl ether. Carbamate protecting groups (e.g., Boc, or di-Boc)[95,96] were found to cleave under the Ullmann coupling conditions. Finally, we were unable to form either benzophenone imine or benzyl imine,[33] even in the presence of silyl transfer reagents (e.g., BSA), presumably due to the reduced nucleophilicity of the exocyclic amine based on resonance delocalization of charge into the π-system of the pyridinyl ring.

While we were able to generate enough material to proceed with the synthesis of compounds **10a-d** using the chemistries in paths A and B, we remained unconvinced that these were the most efficient methods for the scaled synthesis of advanced common intermediates **10a** and **12**. In order to optimize path C, we screened alternative coupling partners that would serve as a protected hydroxyl substituent and potentially change the R_f of the product enough to allow a complete separation from the reduction impurity using chromatographic methods.

Benzyl alcohol (BnOH), *p*-methoxybenzyl alcohol (PMBOH), *t*-butyl alcohol (*t*-BuOH), monomethylethylene glycol, and Wang resin were among the alcohols tested. Coupling to BnOH and PMBOH afforded the desired product, but it was not separable from reduction impurities. Careful analysis of the PMBOH reaction mixture led to the identification of *p*-methoxybenzaldehyde as an additional side product, suggesting that reductive cleavage of the aryl-halide bond was accompanied by oxidation of the alcohol.[84,89] Substituting in *t*-BuOH,[85] which would not be susceptible to this type of oxidative process, however, resulted only in the recovery of starting material, presumably due to the high steric hindrance of the alcohol. Attempts to couple monomethylethyleneglycol,[85] a relatively high polarity alcohol, or Wang resin, a solid phase reagent bearing a benzylic alcohol,[97] both also failed under the given reaction conditions.

In order to alleviate problems during the deprotection step, we investigated isopropanol as a coupling partner[85] (Path C, R=6-O*i*-Pr), and found that it drastically changed the chromatographic properties of the desired product. Installation of the 6-O*i*-Pr both changed the order of elution for the products, such that the desired product now eluted before the reduction product and offered a partial separation of these products using the acetone/DCM solvent system optimized for 6-OMe products. More importantly, the isopropyl group greatly increased the product solubility in EtOAc relative to the methyl group; this allowed us to switch to an EtOAc/hexanes solvent system and yielded increased separation. These changes, however, came with the drawback that the 6-O*i*-Pr compounds were less crystalline than the 6-OMe analogs, slightly complicating the recrystallization process.

In contrast, the installation of the hydroxyl substituent in the 7-position of the imidazo[1,2-*a*]pyridine core was accessible through S_NAr of the 2-amino-4-chloropyridine[37-39] precursor, as outlined in Scheme 2.5. While this reaction was not susceptible to the reduction processes that plagued our synthetic strategies described above,[97-99] it hinged on taking advantage of the electron poor nature of pyridine ring systems. As such, we screened several protection strategies for the exocyclic nitrogen, and found that only the 2,5-dimethylpyrrole protecting group facilitated the desired reactivity. We suspect that the success of this protection strategy is founded on the electron withdrawing effects of the conjugated pyrrole ring system, which is twisted out of plane from the pyridine π-system by the steric encumbrance of the 2,5-dimethyl substituents. This hypothesis is backed by analysis of quantum mechanical DFT calculations of geometry optimized structures (Fig. 2.7, Table 2.7). These data show that alone neither rotation of the exocyclic nitrogen out of plane, nor the electron withdrawing effect of the pyrrole, are sufficient for reducing the charge observed at C4. The combination of these two effects through the dimethyl pyrrole protection scheme, however, yields a positive charge at C4
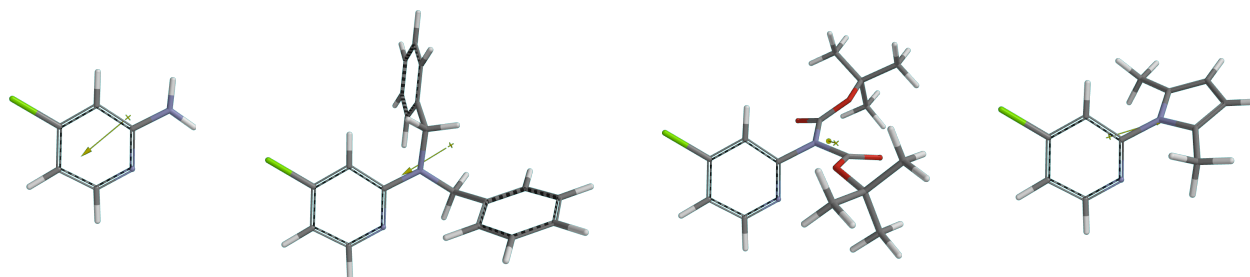
Figure 2.7. QM-Optimized Geometry and Dipole Moment of Masked 2-Amino-4-Chloropyridines

Table 2.7. Quantitative Analysis of Geometry Optimized Masked 2-Amino-4-Chloropyridines.[a]

| Protection Scheme	Mulliken Charge on C4	Dihedral Angle (degrees)	Dipole Moment (Debye)
None	0.0042	0	1.88
Dibenzyl	-0.214	11.17	1.74
Di-Boc	-0.088	67.32	2.14
Dimethyl Pyrrole	0.125	59.85	1.92
Pyrrole	-0.098	0	0.92

[a] All values calculated in Spartan'08. Geometry Optimized at RHF/6-31G*; Single Point Energy at DFT/B3LYP/6-311+G**

and completely reverses the direction of the dipole moment relative to the unprotected 2-amino-4-chloropyridine.

Attempts to extend this methodology to the synthesis of 2-amino-5-methoxypyridine were universally met with failure. By placing the halogen at the 5-position, there no longer exists a resonance structure placing the negative charge that builds during the addition step on the pyridinyl nitrogen, as is observed for the 4-position. However, such a structure exists for 5-bromo-2-nitropyridine in which the accumulating charge can be delocalized through the nitro group via stable resonance structures. Initial attempts to displace the bromide revealed that in all cases we displaced the 2-nitro group instead (Scheme 2.11).   Subsequent reactions employing silver-based additives to reverse the selectivity by promoting the displacement of the bromide through the formation of insoluble silver bromide salts were met with failure.

Functionalization of the imidazopyridine core proceeded smoothly and selectively at the 3-position via eletrophilic aromatic substitution. The selectivity of this reaction proved superior than what was previously observed for the corresponding indazole analogs,[5,100] and can be

Scheme 2.11. Reaction of 5-Bromo-2-nitropyridine with Alkoxide Nucleophiles
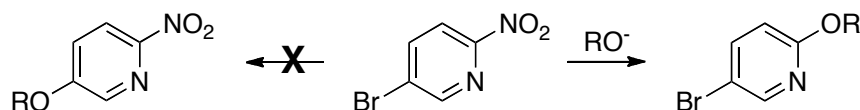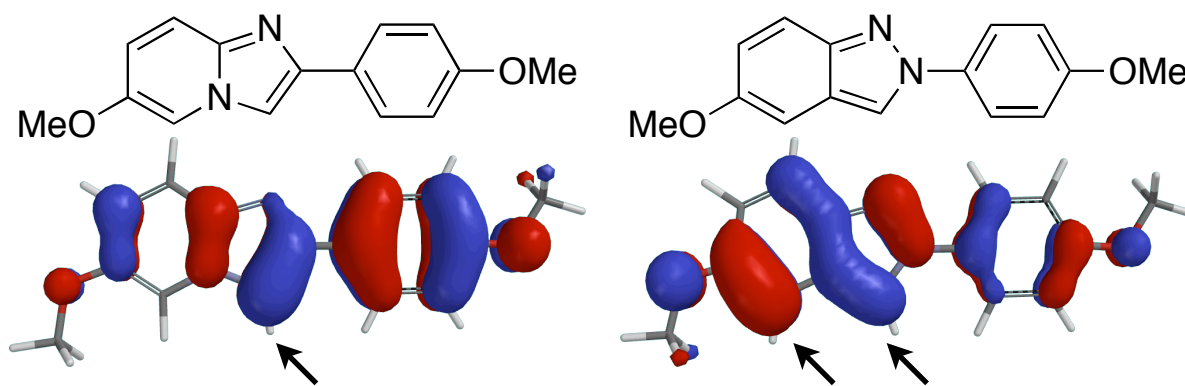
Figure 2.8. Highest Occupied Molecular Orbitals (HOMO) for the imidazopyridine (left) and indazole (right) scaffolds. The large coefficient for the orbital localized on C3 in the imidazopyridine leads to enhanced selectivity for electrophilic aromatic substitution. Delocalization of the corresponding orbital for the indazole scaffold results in reduced regioselectivity, yielding competing substitution at C3 and C4.[100]

justified by the observed charge buildup at C3 through quantum mechanical DFT calculations, as shown in Figure 2.8.

Our initial investigations into the further manipulation of aryl iodides **23a-d** to install additional functionality via Pd-catalyzed reaction manifolds were met with minimal success. Common conditions for SM couplings failed to yield product, while more rigorous microwave-catalyzed conditions[39] yielded 50% of the 3-H reduction product in the synthesis of **23a** (R=Ph). Notably, the installation of the 3-nitrile proceeded in good yield via a similar reaction mechanism, leading us to pursue experiments aimed at optimizing SM reaction conditions. Starting from boronic acid, Pd, and base in dioxane, we found that the reaction only proceeded to give the desired product at 130 °C; however, product formation was accompanied by a significant amount of 3-H reduction product. The strict elimination of water by replacing reagent grade dioxane and bench-stored K_3PO_4 with anhydrous THF from a solvent delivery system and Cs_2CO_3 stored in a desiccator, resulted in the significant diminution in observed 3-H reduction product. The identity of the halogen appeared to greatly affect the rate of reaction, as judged by TLC, such that I>Br>>Cl. The palladium source did not appear to greatly affect reaction course, as both Pd(dppf)Cl$_2$ and Pd$_2$dba$_3$ yielded comparable results. While phenylboronic acid reacted with good conversion, the coupling of cyclopropylboronic acid was further optimized by adding an additional equivalent of boronic acid after 30 minutes of reaction time. The related isopropylboronic acid,[44] however, failed to give the desired product under all conditions tested. We made several attempts at installing the 3-vinyl substituent using analogous methodologies. [101] The vinylboronic anhydride pyridine complex, a stable vinyl boronic acid equivalent,[102] gave the desired product; however, significant quantities of the 3-H reduction side product formed and was not separable at this stage. Unfortunately the required elevated temperature precluded the use of other technologies such as the slow release of MIDA-protected vinylboronic acids.[103]

This product was accessible through more traditional chemistries, as described in Scheme 2.6, and thus was not pursued further by other transition-metal coupling methods (e.g., Stille coupling).[45,104-106]

The successful deprotection, isolation, and purification of each imidazopyridine was heavily dependent on the substituent at the 3-position. This resulted in the use of a variety of reagents to affect cleavage of the alkyl ethers, as described in section II.E. Efficient extraction of the deprotected product from the quenched reaction mixture proved to be extremely pH sensitive.   Based on the approximation of pK_a using MarvinSketch,[107,108] we designed a potassium buffer to enable efficient extraction that maximized product recovery and was generally applicable (Fig. 2.9). The final purification using flash column chromatography was particularly important for compounds originating from path A (Scheme 2.3), as the reduction product impurities were only separable after deprotection. The separation of final products was greatly improved by using a highly controlled gradient elution on a CombiFlash® MPLC instrument.   Despite substantial effort, we were unable to develop conditions for the recrystallization of the free base or HCl salts.
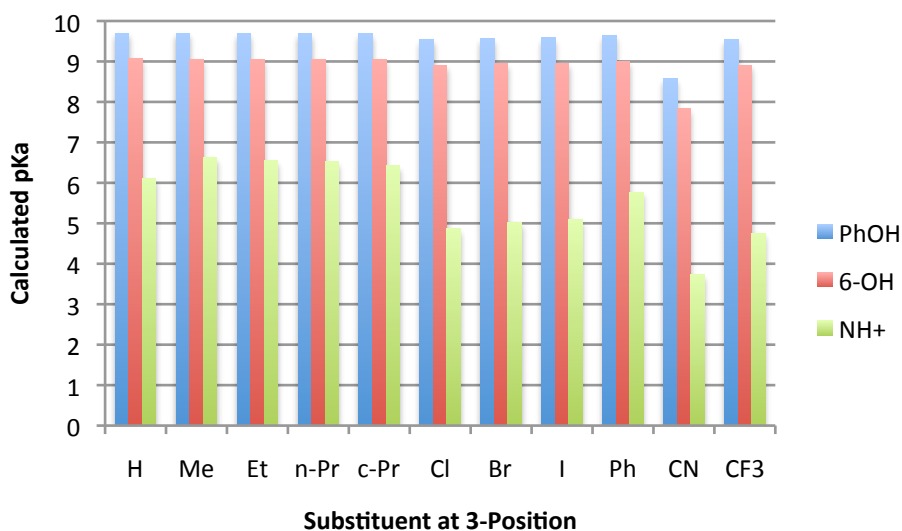
Figure 2.9. Calculated pK_a for Substituted Imidazo[1,2-*a*]pyridinols

B. Biological Data

From the relative binding affinity (RBA) data given in Table 2.2, a number of trends become apparent. As highlighted by entries **29f-k**, all compounds lacking a hydroxyl group on the heterocyclic core demonstrated extremely low affinities for either ER subtype.   The incorporation of a hydroxyl substituent at either 6- or 7-position alone is not sufficient for increased binding affinity (Table 2.2, entries 7, 13, 14, 18, 19), as only in cases bearing a 6-OH and a suitable 3-substituent (Table 2.2, entries 9, 12, 16, 17) was increased binding observed.

While comparing Me to Et shows approximately 3-fold increased affinity, extending to *n*-propyl or cyclopropyl leads to reduced affinities comparable to Me. Entry 12 bearing the 3-Cl demonstrates good affinity and high selectivity for the ERβ-subtype; however, moving down the group to 3-Br and 3-I significantly reduces both affinity and selectivity. Other electron withdrawing substituents, such as 3-Ph, 3-CN, and 3-CF$_3$ show increased affinity with moderate ERβ-selectivity. Entry 17 where R=CF$_3$ represents the highest affinity imidazopyridine synthesized herein, with and RBA of 13.5%, corresponding to a K$_i$=2.9 nM.

Comparing these data to the RBA data for the corresponding indazole,[5] the imidazopyridines generally bind with significantly reduced affinity and moderately reduced β-selectivity relative to the corresponding indazole analog. The lone exception to this is entry 4 in Table 2.8, where 3-Ph, which binds with an RBA of 1.93%, 4-fold higher than the indazole analog. Curiously, this compound is also the only example that we are aware of for a ligand bearing an internal aromatic substituent that retains ERβ-selectivity. Ligands of this structure are generally quite selective for ERα.

Table 2.8. Comparison of Relative Binding Affinity Across Scaffold Structure

Entry	Cmpd	R	Imidazopyridines			Indazoles		
			ERα	ERβ	β/α	ERα	ERβ	β/α
1	**29f**	Cl	0.021	0.866	41.2	0.30	32.1	107
2	**29g**	Br	0.006	0.099	16.5	0.18	18.4	102
3	**29h**	I	0.016	0.058	3.63	0.17	8.5	50
4	**29i**	Ph	0.054	1.93	35.7	0.04	0.46	12
5	**29j**	CN	0.111	3.44	31.0	1.4	30.1	22
6	**29k**	CF$_3$	0.893	13.51	15.1	3.9	69	18

C. Molecular Modeling

Our attempts to identify a molecular basis to explain the observed trends in binding affinity through the quantitation of each receptor-ligand complex proved disappointing. We are primarily concerned with the selection of the crystal structure used as a guide for each receptor

subtype, which was carefully selected based on the geometric similarity to the heterocyclic core of our designed imidazopyridine ligands. Overlaying the binding volumes obtained from crystal structures of ERα bound to three different classes of ligands suggests that the shape of the pocket remains relatively consistent (Fig. 2.10A). The ERβ overlay, in contrast, emphasizes that the shape of the pocket is heavily dependent on structural characteristics of the bound ligand (Fig. 2.10B). This apparent flexibility of the ERβ receptor reinforces the concept that the selection of the reference crystal structure used as a guide can heavily influence the resulting docking poses. The multistage minimization routine is designed to minimize this dependence for small changes in binding volume (e.g., those observed in the ERα overlay); however, it would be difficult to extend this methodology to account for the extensive changes in pocket shape observed for ERβ.[109]

Our secondary concerns are centered on the use of molecular mechanics force fields to analyze these systems, as implemented in ADV docking and MOE energy minimization routines. As discussed in Chapter 1, one of the major differences in the ligand binding pocket between ERα and ERβ is the Leu384→Met336. In ERβ, this methionine resides over the face of the ligand, and we suspect that the ligand interaction with the methionine sidechain is responsible for high affinity binding to ERβ, and by extension engenders subtype selectivity. Based on the QM data justifying the ERβ-selectivity of DPN,[6,10,11] we hypothesized that the primary basis of this interaction appears to be a polarized interaction of the methionine sidechain and the π-system of the heterocyclic core. Molecular mechanics force fields, however, implement parameterized point charges and were not explicitly developed to handle interactions founded on atom polarization. The inability to take this important interaction into account potentially undermines all of our attempts at scoring the receptor-ligand complexes using molecular mechanics-based approaches. Accordingly, we refocused our efforts on ligand-based approaches in which we could circumvent the limitations of MMFFs through the use of quantum mechanical calculations.

Comparing the observed RBA values from Table 2.8 with the analogous indazole structures highlights that with only one exception (entry 4, R=Ph) the imidazopyridines yielded significantly lower RBA values than the corresponding indazole compound. These compounds are geometrically identical, merely differing in the position of a single nitrogen atom; therefore, we conclude that the remarkable erosion of binding affinity is due wholly to changes in the electronics within the ligand. We hypothesized that upon identifying the changes in electronic structure, these data could ultimately be projected onto the receptor structure, elucidate the binding orientation of the ligand, and ascertain the consequences of changing the position of the nitrogen atom within the context of the protein structure.
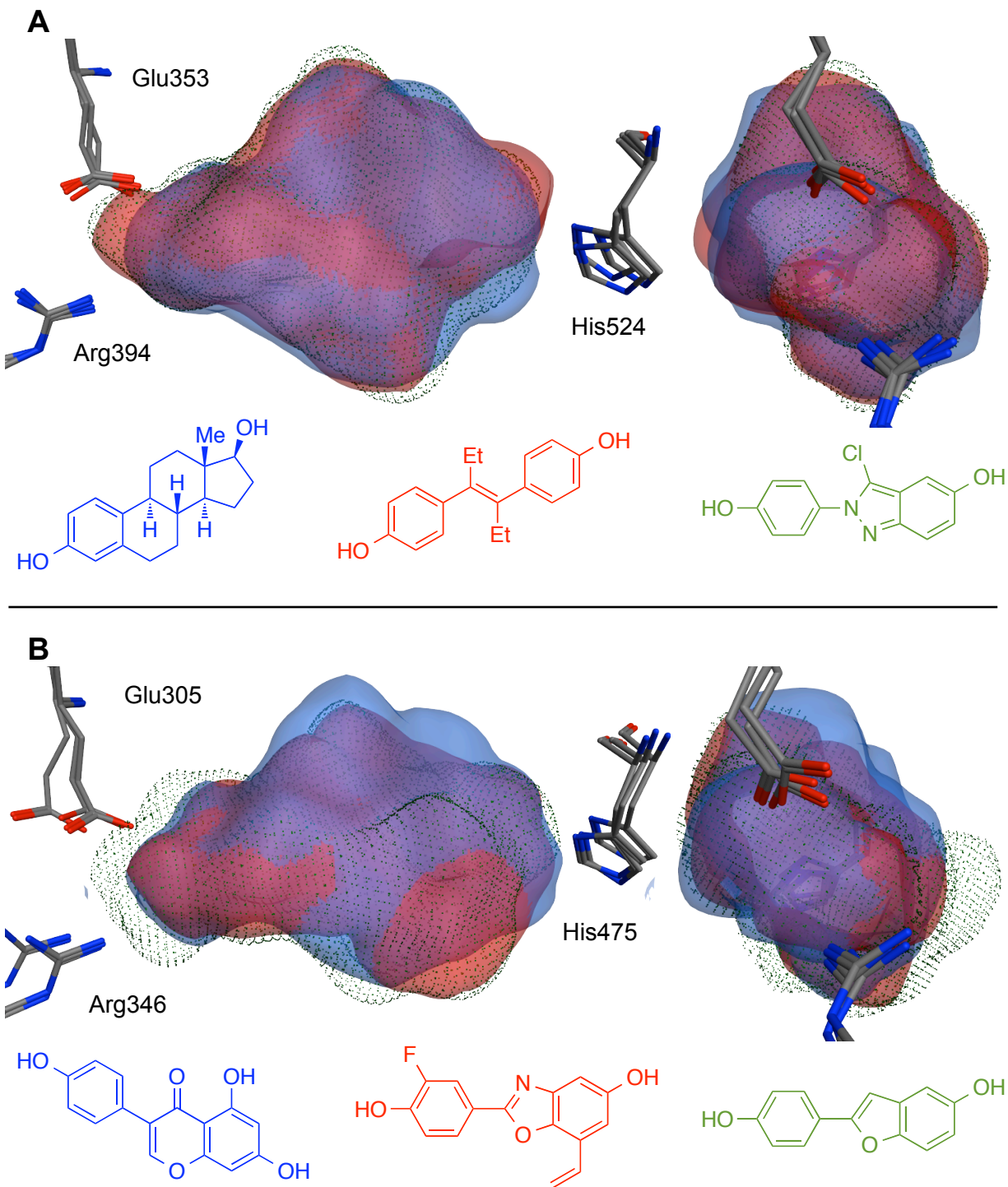
Figure 2.10. **A)** The crystal structures of ERα complexed with E2 (blue, pdb code: 1ERE), diethylstilbestrol (red, pdb code: 3ERD), and chloroindazole (green, pdb code: 2QGW), demonstrate minimal variation in the volume of the ligand binding pocket as shown from the front and side views. **B)** Similar views of crystal structures of ERβ complexed with genistein (blue, pdb code: 1X7J), ERB-041 (red, pdb code: 1X7B), and WAY-397 (green, pdb code: 1U9E), however, indicate substantial variation in the shape and size of the ligand binding pocket.

Expanding upon the pK_a approximations in Figure 2.9, the E(PT) and P(A) data (Table 2.6) yield a more rigorous metric for assessing the impact of 3-substitution and scaffold structure on the relative basicity of each compound. Compounds bearing EDGs are substantially more basic than compounds bearing EWGs. This is due to increased localization of charge at N1, as observed from projection of the electrostatic potential over the vdW surface and the angle of the dipole moment (Figs. 2.4, 2.5). Comparing the imidazopyridines to the corresponding indazole analogs indicates that the indazole core is substantially less basic, and charge density is more evenly distributed throughout the heterocyclic core, as denoted by increased yellow coloration in the electrostatic potential surface compared to the green coloration in the imidazopyridines. The horizontally flipped dipole moment of the indazoles, relative to the imidazopyridines, also indicates that the charge density is shifted away from the N1 position as well. The ramifications of these properties were readily apparent in the pH-senstitivity of the deprotected compounds that complicated isolation procedures, and by extension, may play an important role affecting binding energy by modulating the ligand desolvation penalty that must be overcome during a binding event.

Within the data presented above, there are a number of examples in which compounds with higher charge localization (stronger basicity) yield higher RBAs than compounds with lower charge localization (e.g., R=CN vs. R=CF$_3$). Furthermore, smaller magnitude in dipole moment does not necessarily correlate to increased affinity. An effective way of evaluating this latter correlation while removing steric variables includes plotting the dipole moment of each imidazopyridine against the dipole moment for the analogous indazole compound, as shown in Figure 2.11. All points above the solid line represent compounds in which the calculated dipole moment is larger for the indazole analog, while the reverse is true for points below the line. These data fail to cluster based on RBA trends, indicating the magnitude of the dipole moment alone does not appear to be a significant determinant of binding affinity.

The opposing directions of the dipole moments calculated for imidazopyridine and indazole core scaffolds (Figs. 2.4, 2.5) is a striking trend. Based on these data and their loose correlation to the observed trends in RBA, we hypothesized that the ligand binding pocket has an intrinsic dipole moment with a significant component directing from the from the His475 end to the Glu305/Arg346 end of the pocket. Assuming that the ligand binds in the generally accepted orientation, the dipole moment of the indazole scaffold generally opposes the local dipole of the binding pocket, representing a positive electrostatic interaction leading to reduced binding energies (higher binding affinity). The dipole moment of the imidazopyridines, in contrast, would align with the dipole of the binding pocket, incurring an electrostatic repulsion and increased binding energies (lower binding affinity). Semi-empirical calculations of the isolated ligand binding pocket, in fact, support this hypothesis. The calculated dipole moment is
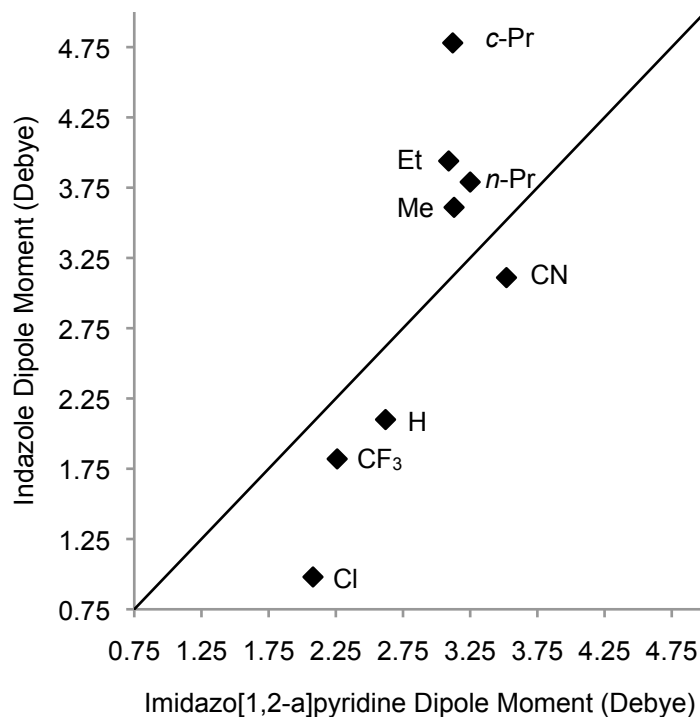
Figure 2.11. Comparing the Magnitude of Dipole Moment For Substituted Imidazopyridine and Indazole Scaffolds.

projected onto the ERβ ligand binding pocket shown in Figure 12. Overlaying the 3-CF$_3$ ligand for each scaffold demonstrates the opposing relationship of the dipole moments for the indazole and receptor (Fig. 2.12A), while the corresponding dipoles are aligned for the imidazopyridine (Fig. 2.12B). We hypothesize that these interactions may also explain the unique reversal in binding trends for the 3-phenyl analogs. In this case, we envision that the relatively large phenyl substituent is unable to properly fit within the reduced pocket volume of ERβ, inducing the ligand to bind in a "horizontally flipped" orientation, in which the pendant phenol is directed towards the His524 residue. Consequently, the dipole moment of the ligand would also be reversed, yielding a positive dipole-dipole interaction for the imidazopyridine, and a repulsive interaction for the analogous indazole.

VII. Conclusions.

We set out to clarify the role of the ligand core scaffold in determining binding affinity and selectivity in ERs by through a combination of synthetic and computational approaches. Using a series of previously synthesized indazoles as a guide, we designed a structurally conservative change in the core scaffold by implementing the isomeric imidazo[1,2-*a*]pyridine as a scaffold replacement. Several compounds were then synthesized and tested for *in vitro* binding affinity to ERs. The resulting data demonstrated both reduced affinity and subtype selectivity relative to
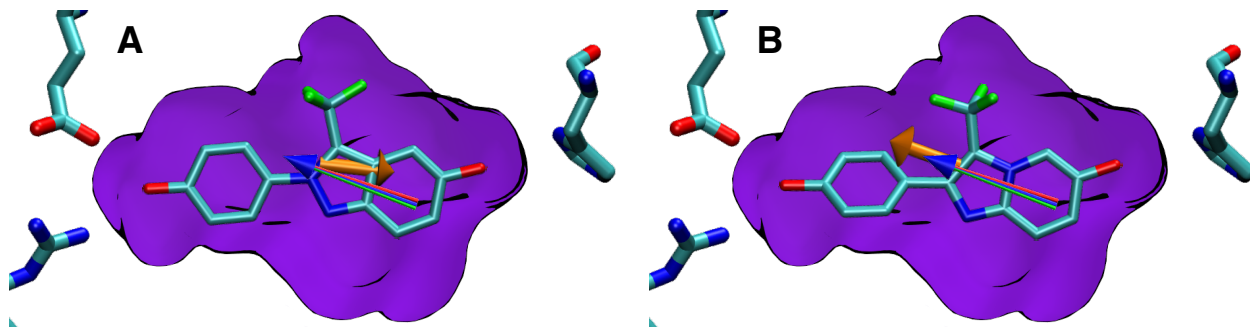
Figure 2.12. The dipole moment of the ligand binding pocket (multicolored arrow) and ligand (orange arrow) are superimposed over the ligand binding volume (purple surface). Assuming the standard binding orientation, the dipole moment of the indazole ligand generally opposes the dipole vector of the binding pocket (**A**), forming a positive interaction. The dipole moment of the imidazopyridine (**B**), in contrast, is well aligned with the dipole of the binding pocket, inducing an electrostatic repulsion that would negatively affect binding affinity.

the analogous indazole compounds. Further investigations using a variety of molecular modeling-based approaches highlighted striking differences in the electronic properties for each ligand, most notably the differences in relative basicity and opposing dipole moment between the imidazopyridine and indazole scaffolds. Followup calculations on the ligand binding pocket indicated the presence of a strong dipole moment within the local environment of the ligand. From these data, we propose that the electronic properties of the core scaffold play a direct role in influencing binding affinity through a strong dipole-dipole interaction between the ligand and the net dipole of the ligand binding pocket. Other contributing factors include the influence of charge localization on protonation state and desolvation penalty, in addition to more traditional factors such matched hydrogen bonding contacts, lipophilicity profile, and steric constraints.

VIII. Experimental.

General Considerations. All chemical reagents were purchased from commercial suppliers without further purification. Anhydrous DMF was obtained from Aldrich in a Sureseal™ bottle stored in a secondary container with desiccant; all other anhydrous solvents were obtained from a solvent dispensing system unless otherwise stated. All glassware was oven or flame-dried and cooled under vacuum or in a dry box. All reactions were conducted under argon. NMR spectra were obtained on Varian Oxford instruments and worked up using ACD, Inc. 1D-NMR processing software. The chemical shifts are reported in ppm and referenced to the solvent peak. Electron Impact (EI) mass spectra were obtained on a 70-VSE mass spectrometer with an ionization energy of 70 eV. Electrospray Ionization (ESI) mass spectra were obtained on a Q-Tof mass spectrometer. Melting point ranges were measured using a Thomas Hoover capillary melting point apparatus.

General procedure for the acylation of anisole. A round bottomed flask charged with AlCl₃ stirring in anhydrous DCM under argon was cooled in an ice bath. Anisole and the corresponding acyl chloride were in sequentially injected into the reaction mixture, which was slowly warmed to ambient temperature as the reaction was allowed proceed for 3-14 hours. The reaction mixture was then poured into ice water and exhaustively extracted with DCM, the organic layer dried over MgSO₄, vacuum filtered, and concentrated by rotary evaporation. The crude material was used in subsequent reactions unless otherwise specified.

General procedure for the α-bromination of ketones. A round bottomed flask affixed with a reflux condenser was charged with CuBr₂ stirring in EOAc and heated to reflux for one hour, yielding a green heterogeneous solution. A solution of ketone dissolved in CHCl₃ was added to the reaction mixture, which was held at reflux until solution changed color from green to amber (usually within 1-3 hours). The reaction mixture was then filtered through celite, washed with EtOAc, and purified by flash chromatography over silica gel (25% EtOAc in hexanes).

General procedure for the formation of imidazo[1,2-*a*]pyridine core. The corresponding 2-aminopyridine and 2-bromoacetophenone were dissolved in ethanol and heated to reflux while stirring until the reaction was complete by TLC (5-20% acetone in DCM), usually after 8-12 hours. The reaction products were isolated using a standard workup consisting of concentration by rotary evaporation, re-suspension in DCM and 1M KOH (aq) or saturated NaHCO₃ (aq), exhaustive extraction with DCM, drying over MgSO₄, vacuum filtration, and concentration by rotary evaporation. The products were purified by flash chromatography over silica gel (5-20% acetone in DCM).

General procedure for the halogenation of imidazo[1,2-*a*]pyridines. The corresponding imidazo[1,2-*a*]pyridine and *N*-halosuccinimide were dissolved in acetonitrile and stirred at room temperature until the reaction was complete by TLC (5-20% acetone in DCM), usually after 30 minutes. The reaction products were concentrated by rotary evaporation and purified by flash chromatography over silica gel (5-20% acetone in DCM).

General procedure for Ullmann-type alkoxylation of aryl iodides. A pressure tube or heavy walled round bottomed flask was charged with aryl iodide, CuI, 1,10-phenanthroline, Cs₂CO₃, and the corresponding alcohol. The reaction vessel was sealed using a teflon screw-cap, covered in aluminum foil, and heated to 110 °C behind a weighted blast shield for ~24 hours. The reaction flask was then allowed to cool to ambient temperature, and the reaction

mixture filtered through celite, washed with the appropriate organic solvent (R=MeO: DCM, R=*i*-PrO:EtOAc), and concentrated by rotary evaporation. The resulting solid was transferred to a separatory funnel containing a brine solution and exhaustively extracted with organic solvent. The organic layer was dried over MgSO$_4$, vacuum filtered, and concentrated by rotary evaporation. The desired product was purified by flash chromatography over silica gel using the specified solvent system.

General procedure for Suzuki-Miyaura (SM) coupling of aryl iodides. A pressure tube was charged with aryl iodide, boronic acid, Pd(dppf)·CH$_2$Cl$_2$, and anhydrous THF and sealed with a teflon screw cap. The reaction vessel was heated to 130°C behind a weighted blast shield for one hour. The reaction mixture was allowed to cool to ambient temperature, transferred to a separatory funnel, suspended in distilled water, and exhaustively extracted with DCM or EtOAC. the organic phase was dried over MgSO$_4$, vacuum filtered, and concentrated by rotary evaporation. The desired product was isolated by flash chromatography over silica gel as specified.

General procedure for the cleavage of aryl-alkyl ethers. The compound bearing an aryl-alkyl ether was placed in a dry round bottomed flask charged with DCM and cooled in an ice bath. Boron triflouride as the dimethylsulfide complex was injected via syringe and the reaction mixture was slowly warmed to ambient temperature as the reaction was allowed proceed for ~24 hours. The reaction mixture was re-cooled in an ice bath, quenched with methanol, stirred for 30 minutes, and concentrated by streaming nitrogen bubbled through bleach to contain thiol vapors. The crude material was re-suspended in EtOAc and water, which was buffered to ~7.6 using a K$_2$HPO$_4$/KH$_2$PO$_4$ (aq) solution and allowed to stir until the aqueous layer became colorless. The products were isolated by exhaustive extraction with EtOAc, dried over MgSO$_4$, vacuum filtered, and concentrated by rotary evaporation. The desired product was purified by flash chromatography over silica gel (10-20% IPA in DCM) or MPLC using a CombiFlash® instrument (gradient elution, 0-20% IPA in DCM).

2-Phenylimidazo[1,2-*a*]pyridine (1). Following the general procedure for the formation of imidazo[1,2-*a*]pyridine core from 2-aminopyridine (5.04g g, 53.6 mmol) and 2-bromoacetophenone (10.61 g, 53.3 mmol), **1** was isolated as a beige solid (7.37 g, 71.2%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.08 (dt, J = 6.75, 1.02 Hz, 1H), 7.95 (dt, J = 8.09, 1.63 Hz, 2H), 7.83 (s, 1H), 7.62 (dd, J = 9.11, 0.75 Hz, 1H), 7.42 (m, 2H), 7.32 (tt, J = 7.29, 1.29 Hz, 1H), 7.14 (ddd, J = 9.11, 6.75, 1.29 Hz, 1H), 6.74 (td, J = 6.75, 1.07 Hz, 1H); ^{13}C-NMR (CDCl$_3$, 125 MHz)

δ 145.77, 145.65, 133.74, 128.69, 127.93, 126.00, 125.55, 124.59, 117.52, 112.36, 108.07; HRMS (EI) calcd for $C_{13}H_{10}N_2$ M^+ 194.084398, found 194.084433.

4'-methoxypropiophenone (2a). Following the general procedure for the acylation of anisole from $AlCl_3$ (1.68 g, 12 mmol), propionyl chloride (0.9mL, 10 mmol), and anisole (1.0 mL, 10 mmol), to yield **2a** as a cloudy white oil (1.64 g, 96.6%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.96 (AA'XX', 2 H), 6.94 (AA'XX', 2 H), 3.87 (s, 3 H), 2.96 (q, J=7.29 Hz, 2 H), 1.22 (t, J=7.29 Hz, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 199.52, 163.25, 130.19, 129.96, 113.62, 55.41, 31.37, 8.38; HRMS (ESI) calc'd for $C_{10}H_{13}O_2$ [M+H]$^+$ 165.0916, found 165.0912.

4'methoxybutyrophenone (2b). Following the general procedure for the acylation of anisole from $AlCl_3$ (1.64 g, 12 mmol), butyryl chloride (1.0 mL, 10 mmol), and anisole (1.0 mL, 10 mmol), to yield **2b** as a cloudy oil (2.82 g, quant.). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.95 (AA'XX', 2 H), 6.94 (AA'XX', 2 H), 3.87 (s, 3 H), 2.90 (t, J=7.29 Hz, 2 H), 1.76 (sxt, J=7.42 Hz, 2 H), 1.00 (t, J=7.50 Hz, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 199.11, 163.26, 130.28, 130.15, 129.41, 120.60, 113.82, 113.61, 55.40, 40.16, 17.95; HRMS (ESI) calc'd for $C_{11}H_{15}O_2$ [M+H]$^+$ 179.1072, found 179.1072.

4'-methoxyvalerophenone (2c). Following the general procedure for the acylation of anisole from $AlCl_3$ (1.7 g, 12.7 mmol) valeryl chloride (1.2mL, 10 mmol), and anisole (1.0 mL, 10 mmol), to yield **2c** (1.56 g, 81.2%). ^{1}H-NMR (CDCl$_3$, 400 MHz) δ 7.95 (AA'XX', 2 H), 6.94 (AA'XX', 2 H), 3.87 (s, 2 H), 2.92 (m, 2 H), 1.71 (dt, J=15.08, 7.48 Hz, 1 H), 1.41 (dq, J=14.98, 7.46 Hz, 2 H), 0.95 (t, J=7.32 Hz, 2 H); ^{13}C-NMR (CDCl$_3$, 100 MHz) δ 199.21, 163.25, 130.29, 130.17, 113.62, 77.00, 55.43, 37.99 , 26.73, 22.52, 13.94; HRMS (ESI) calc'd for $C_{12}H_{17}O_2$ [M+H]$^+$ 193.1229, found 193.1225.

4'-methoxyisovalerophenone (2d). Following the general procedure for the acylation of anisole from $AlCl_3$ (1.6 g, 12 mmol), isovaleryl chloride (1.2mL, 10 mmol), and anisole (1.0 mL, 10 mmol), to yield **2d** as a colorless oil (1.97 g, quant.). ^{1}H-NMR (CDCl$_3$, 400 MHz) δ 7.94 (AA'XX', 2 H), 6.93 (AA'XX', 2 H), 3.87 (s, 3 H), 2.78 (d, J=7.08 Hz, 2 H), 2.29 (m, 1 H), 0.99 (d, J=6.84 Hz, 6 H); ^{13}C-NMR (CDCl$_3$, 100 MHz) δ 198.86, 163.22, 130.33, 113.58, 77.76, 55.40, 47.14, 25.34, 22.77.

2-Bromo-4'-methoxyphenylpropiophenone (3a). *Method A.* 4'-methoxypropiophenone (827 mg, 5 mmol) was dissolved in methylene chloride (2 mL) and 1,4-

dioxane (4 mL). A solution of Br_2 (800 mg, 5mmol) dissolved in methylene chloride (2 mL) was added drop-wise with stirring at 0 °C. Upon completion of Br_2 addition, the reaction was allowed to warm to room temperature and stir for one hour. Aqueous saturated sodium bicarbonate solution (4 mL) was added and stirred for an additional one hour. The reaction was quenched with sodium thiosulfate, exhaustively extracted with methylene chloride, dried over magnesium sulfate, vacuum filtered, and concentrated by rotary evaporation. The product solidified upon cooling to 0 °C, and was further purified by trituration from hexanes to yield **3a** as a white solid (821 mg, 67%). *Method B.* Following the general procedure for α-bromination of ketones from **2a** (1.50 g, 9.1 mmol) and $CuBr_2$ (3.59 g, 16.2 mmol), yielded **3a** as light green oil (1.34 g, 87.5%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.99 (AA'XX', 2H), 6.92 (AA'XX', 2H), 5.24 (q, J = 6.65 Hz), 1H), 3.85 (s, 1H), 1.86 (d, J = 6.43 Hz, 3H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 191.93, 163.86, 131.23, 126.70, 113.88, 55.46, 41.38, 20.14; HRMS (ESI) calc'd for $C_{10}H_{11}O_2NaBr$ [M+Na]$^+$ 264.9840, found 264.9850.

2-Bromo-4'-methoxyphenylbutyrophenone (3b). Following the same procedure for **3a** *Method A*, the bromination of 4'-methoxyphenylbutryophenone (723 mg, 4 mmol) yielded **3b** as a white solid (1.01 g, 97%). *Method B.* Following the general procedure for α-bromination of ketones from **2a** (2.82 g, 10.0 mmol) and $CuBr_2$ (3.61 g, 16.7 mmol), to yield **3b** as light yellow oil (2.43 g, 94.7%). ^{1}H-NMR (CDCl$_3$, 400 MHz) δ 8.00 (AA'XX', 2H), 6.95 (AA'XX', 2H), 5.03 (dd, J = 7.81, 6.35 Hz, 1H), 2.21 (m, J = 7.3, 0.7 Hz, 1H), 2.12 (m, J = 7.3 Hz, 1H), 1.06 (t, J = 7.32 Hz, 3H); ^{13}C-NMR (CDCl$_3$, 100 MHz), δ 191.84, 163.88, 131.19, 127.26, 113.92, 55.52, 48.98, 26.98, 12.20; HRMS (ESI) calc'd for $C_{11}H_{13}O_2NaBr$ [M+Na]$^+$ 278.9997, found 278.9996.

2-Bromo-4'-methoxyvalerophenone (3c). Following the general procedure for α-bromination of a ketone from **2c** (2.06 g, 10.0 mmol) and $CuBr_2$ (3.95 g, 17.3 mmol) to yield **3c** as light green oil (2.70 g, 93.2%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.01 (AA'XX', 2 H), 6.96 (AA'XX', 2 H), 5.13 (dd, J=7.81, 6.35 Hz, 1 H), 3.89 (s, 3 H), 2.14 (m, 2 H), 1.56 (m, 1 H), 1.44 (m, 1 H), 0.99 (t, J=7.45 Hz, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 191.89, 163.90, 131.19, 127.31, 113.95, 77.00, 55.53, 47.00, 35.57, 20.79 , 13.58; HRMS (ESI) calc'd for $C_{12}H_{16}O_2Br$ [M+H]$^+$ 271.0334, found 271.0328.

2-Bromo-4'-methoxyisovalerophenone (3d). Following the general procedure (method B) for α-bromination of ketones from **2d** (1.97, 10.0 mmol) and $CuBr_2$ (3.71 g, 16.7 mmol) to yield **3d** as light yellow-green solid (2.42 g, 89.7%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.00 (AA'XX', 2 H), 6.97 (AA'XX', 2 H), 4.91 (d, J=8.79 Hz, 1 H), 3.89 (s, 3 H), 2.48 (m, 1 H), 1.22 (d,

J=6.65 Hz, 3 H), 1.02 (d, J=6.65 Hz, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 192.19, 163.87, 131.10, 113.96, 55.74, 55.52, 31.14, 20.80, 20.38; HRMS (ESI) calc'd for C$_{12}$H$_{15}$O$_2$NaBr [M+Na]$^+$ 293.0153, found 293.0163.

2-(4-Methoxyphenyl)imidazo[1,2-*a*]pyridine (4a). *Method A.* Following the general procedure for the formation of the imidazo[1,2-*a*]pyridine core from 2-aminopyridine (1.63 g, 17 mmol) and 2-bromo-4'-methoxyacetophenone (3.94 g, 17 mmol), **4a** was isolated as a yellow-green solid (2.72 g, 71%). *Method B.* Sodium hydride (60% in a mineral oil dispersion, 48 mg, 1.2 mmol) was added to a solution of **38** (332 mg, 1 mmol) in anhydrous DMF (3 mL). A short path distillation head was attached and the solution stirred at 90°C for six hours. The reaction products were isolated by concentration by rotary evaporation, re-suspension in DCM and 1M KOH (aq), exhaustive extraction with DCM, drying over MgSO$_4$, vacuum filtration, concentration by rotary evaporation, and flash chromatography over silica gel (25% hexanes in EtOAc) to yield **4a** as a pale yellow solid (30.1 mg, 14%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.03 (dt, J = 6.75, 1.23 Hz, 1H), 7.86 (AA'XX', 2H), 7.72 (s, 1H), 7.58 (ddd, J = 9.11, 1.82, 1.07 Hz, 1H), 7.11 (ddd, J = 9.11, 6.75, 1.29 Hz, 1H), 6.95 (AA'XX', 2H), 6.70 (td, J = 6.70, 1.18 Hz, 1H), 3.82 (s, 3H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 159.49, 145.66, 145.55, 127.21, 126.45, 125.40, 124.35, 117.20, 114.05, 112.13, 107.16, 55.22; HRMS (ESI): calc'd for C$_{14}$H$_{12}$N$_2$O [M+H]$^+$ 225.1028, found 225.1035.

3-Methyl-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (4b). Following the general procedure for the formation of the imidazo[1,2-*a*]pyridine core from 2-aminopyridine (282 mg, 3 mmol) and **3a** (609 mg, 2.5 mmol), **4b** was isolated as a yellow solid (168 mg, 28%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.85 (dt, J = 6.86, 1.18 Hz, 1H), 7.72 (AA'XX', 2H), 7.60 (dt, J = 9.06, 1.05 Hz, 1H), 7.13 (ddd, J = 9.00, 6.75, 1.18 Hz, 1H), 7.00 (AA'XX', 2H), 6.80 (td, J = 6.75, 1.07 Hz, 1H), 3.84 (s, 3H), 2.59 (s, 3H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 159.00, 144.22, 142.28, 129.46, 127.45, 123.23, 122.70, 117.19, 115.12, 113.92, 111.81, 55.24, 9.56; HRMS (ESI): calc'd for C$_{15}$H$_{14}$N$_2$O [M+H]$^+$ 239.1184, found 239.1189.

3-Ethyl-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (4c). Following the general procedure for the formation of the imidazo[1,2-*a*]pyridine core from 2-aminopyridine (223.7 mg, 2.2 mmol) and **3b** (513.2 mg, 2 mmol), **4c** was isolated as a yellow solid (300 mg, 60%). ^{1}H-NMR (CDCl$_3$, 400 MHz) δ 7.93 (dt, J = 6.84, 1.10 Hz, 1H), 7.71 (m, 2H), 7.64 (dt, J = 9.03, 1.10 Hz, 1H), 7.15 (ddd, J = 9.03, 6.71, 1.34 Hz, 1H), 7.00 (m, 2H), 6.82 (td, J = 6.84, 1.22 Hz, 1H), 3.84 (s, 3H), 3.07 (q, J = 7.57 Hz, 2H), 1.33 (t, J = 7.57 Hz, 3H); ^{13}C-NMR (CDCl$_3$, 100 MHz) δ

159.11, 144.05, 141.39, 129.30, 127.14, 123.61, 122.75, 121.06, 117.31, 113.97, 112.02, 55.26, 17.02, 12.19.

2-Amino-5-iodopyridine (5). 2-Aminopyridine (7.58 g, 80 mmol) was added to a solution of periodic acid (3.67 g, 16 mmol), and iodine (3.13 g, 32) stirring in distilled water (16 mL). Glacial acetic acid (80 mL) and concentrated sulfuric acid (2.4 mL) were slowly added, and the reaction heated to 100°C until 2-amino-3,5-diiodopyridine was observed by TLC (R_f 0.55, 1:1 hexanes:ethyl acetate), usually within one hour. The reaction was then quenched with sodium thiosulfate, cooled to 0°C, and neutralized with concentrated aqueous potassium hydroxide. The products were then isolated by exhaustive extraction with EtOAc, dried over magnesium sulfate, vacuum filtered and concentrated by rotary evaporation. The products were purified by flash chromatography over silica gel (1:1 hexanes:EtOAc) to yield **5** as a beige crystalline solid (11.2 g, 63%). Alternatively, upon neutralization of the quenched reaction mixture, the precipitate was vacuum filtered, thoroughly washed with distilled water, dried overnight under reduced pressure, and recrystallized from toluene to yield **5** directly. ^{1}H-NMR (CDCl$_3$, 400 MHz) δ 8.21 (dd, J = 2.32, 0.73 Hz, 1H), 7.61 (dd, J = 8.67, 2.32 Hz, 1H), 6.34 (dd, J = 8.67, 0.73 Hz, 1H), 4.49 (s, 2H); ^{13}C-NMR (CDCl$_3$, 100 MHz), δ 157.23, 153.77, 145.29, 110.84, 77.89; HRMS (ESI) calcd for C$_5$H$_4$IN$_2$ [M+H]$^+$ 220.9576, found 220.9571.

2-Amino-5-methoxypyridine (6). Following the general procedure for Ullmann-type alkoxylation of **5** (5.05 g, 22.9 mmol) and distilled MeOH (~25 mL) yielded **6** as a red oil (2.06 g, 49.8%) after purification by flash chromatography over silica gel (15% IPA in DCM). This product contained 27% 5-H impurity. ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.76 (d, J=2.93 Hz, 1 H), 7.07 (dd, J=8.79, 2.93 Hz, 1 H), 6.47 (m, 1 H), 4.31 (br. s., 1 H), 3.76 (s, 2 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 149.40, 147.89, 137.67, 125.56, 113.85, 56.23; LRMS (ESI) C$_6$H$_8$N$_2$O [M+H]$^+$ found 125.0.

2-(2,5-dimethyl-pyrrol-1-yl)-5-iodopyridine (7). Acetonyl acetone (1.8 mL, 16 mmol) and **5** (1.77 g, 8 mmol) were added to molecular sieves (4 Å) and *p*-toluenesulfonic acid (62 mg, 0.35 mmol) stirring in anhydrous toluene (10 mL). The reaction was heated to 130°C under argon for four hours. The products were filtered, washed with EtOAc, concentrated by rotary evaporation, and subsequently purified by flash chromatography over silica gel (20% EtOAc in hexanes) to yield **7** as a yellow solid (2.06 g, 86%). ^{1}H-NMR (CDCl$_3$, 400 MHz) δ 8.79 (dd, J = 2.38, 0.67 Hz, 1H), 8.10 (dd, J = 8.30, 2.32 Hz, 1H), 7.02 (dd, J = 8.30, 0.61 Hz, 1H), 5.90 (s,

2H), 2.13 (s, 6H); ^{13}C-NMR (CDCl$_3$, 100 MHz) δ 155.37, 151.11, 146.08, 128.51, 123.47, 107.39, 90.91, 13.20; HRMS (ESI) calc'd for C$_{11}$H$_{10}$IN$_2$ [M+H]$^+$ 299.0045, found 299.0058.

2-(2,5-Dimethyl-pyrrol-1-yl)-5-methoxypyridine (8). **7** (pyrrole protected) (15.02 g, 50 mmol) and distilled MeOH (~100 mL) were coupled following the general procedure for Ullmann-type alkoxylation of aryl iodides. The desired product was isolated by dry column vacuum chromatography (DCVC, 0-80% EtOAc in hexanes), and purified by recrystalliation from hexanes to yield **8** as beige needles (7.94 g, 77.9%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.29 (d, J=3.17 Hz, 1 H), 7.34 (dd, J=8.67, 3.05 Hz, 1 H), 7.17 (d, J=8.55 Hz, 1 H), 5.89 (s, 2 H), 3.93 (s, 3 H), 2.10 (s, 6 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 154.72, 144.94, 136.10, 128.65, 122.56, 122.38, 106.29, 55.87, 12.91; HRMS (ESI) calc'd for C$_{12}$H$_{15}$N$_2$ [M+H]$^+$ 203.1184, found 203.1183.

2-Amino-5-methoxypyridine hydrochloride (9). Hydroxylamine hydrochloride (10.9 g, 157 mmol), potassium hydroxide (5.9 g, 105 mmol), and **8** (7.68 g, 38.0 mmol) were dissolved in EtOH (133 mL) and water (67 mL) and stirred at 110°C for 10 hours behind a weighted blast shield. The reaction was cooled to ambient temperature and the EtOAc removed by rotary evaporation. The mixture was then transferred to a liquid-liquid continuous extractor, the pH adjusted to ~8 with a saturated NaHCO$_3$ (aq) solution, and extracted with DCM for 10 hours. The organic layer was concentrated by rotary evaporation and the desired product isolated by flash chromatography over silica get (25% hexanes in EtOAc followed by 20% IPA in DCM). This material was taken up in diethyl ether, cooled to 0°C, and ~100 mL 2M HCl in ether was added by cannula transfer with rapid stirring. The resulting purple precipitate was filtered, washed with diethyl ether, and recrystallized from EtOH to yield **9** (4.10 g, 67.2%) as a beige solid. ^{1}H-NMR (DMSO-d$_6$, 500 MHz) δ 8.01 (m, 1 H), 7.72 (dd, J=9.52, 2.93 Hz, 1 H), 7.64 (d, J=2.69 Hz, 1 H), 7.06 (dd, J=9.52, 0.49 Hz, 1 H), 3.72 (s, 3 H); ^{13}C-NMR (DMSO-d$_6$, 125 MHz) δ 150.76, 147.10, 136.08, 116.97, 114.73, 56.58.

6-Methoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (10a). *Method A.* Following the general procedure for the formation of the imidazo[1,2-*a*]pyridine core from **6** (1.0 g, 8 mmol) and 2-bromo-4'-methoxyacetophenone (1.87 g, 8.2 mmol) yielded **10a** (1.0 g, 51%) as a red solid. *Method B.* Following a modified procedure for the formation of the imidazo[1,2-*a*]pyridine core, **9** (2.9 g, 18 mmol) and 2-bromo-4'-methoxyacetophenone (5.0 g, 21.8 mmol) were dissolved in *t*-BuOH (100 mL) and Et3N (3 mL) to yield **10a** (1.6 g, 35%) as a red solid. Method C. Following the general procedure for the Ullman-type alkoxylation of aryl iodides from **11**

(6.49 mg, 18.6 mmol), 1,10-phenanthroline (678.9 mg, 3.7 mmol), copper (I) iodide (381 mg, 1.9 mmol), Cs_2CO_3 (12.2 g, 37.2 mmol) to which freshly distilled methanol (~100-150 mL) yielded **10a** (2.49 g, 52.7%) as a brick-red crystalline solid. ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.84 (AA'XX', 2 H), 7.68 (d, J=0.49 Hz, 1 H), 7.59 (dd, J=2.32, 0.61 Hz, 1 H), 7.48 (dt, J=9.77, 0.73 Hz, 1 H), 6.95 (AA'XX', 2 H), 3.84 (s, 3 H), 3.79 (s, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 159.32, 149.08, 145.57, 142.71, 126.94, 126.68, 119.38, 117.35, 114.04, 108.30, 107.36, 56.08, 55.24; HRMS (ESI): calc'd for $C_{15}H_{15}N_2O_2$ [M+H]$^+$ 255.1164, found 255.1126.

3-Methyl-6-methoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (10b). Following the general procedure for the formation of the imidazo[1,2-*a*]pyridine core from **6** (450.0 mg, 2.5 mmol) and **3a** (693.3 mg, 2.5 mmol), **10b** (256.1 mg, 38.2%) was isolated as a light yellow solid. This product contained ~25% 6-H impurity. ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.71 (AA'XX', 2 H), 7.53 (dd, J=9.64, 0.61 Hz, 1 H), 7.35 (d, J=1.95 Hz, 1 H), 7.00 (AA'XX', 2 H), 6.97 (dd, J=9.64, 2.32 Hz, 1 H), 3.87 (s, 3 H), 3.86 (s, 3 H), 2.58 (s, 3 H); ^{13}C (CDCl$_3$, 125 MHz) δ 158.91, 149.15, 142.31, 141.31, 129.29, 127.52, 118.22, 117.37, 116.15, 113.92, 104.78, 56.24, 55.27, 9.86; HRMS (ESI): calc'd for $C_{16}H_{17}N_2O_2$ [M+H]$^+$ 269.1290, found 269.1292.

3-Ethyl-6-methoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (10c). Following the general procedure for the formation of the imidazo[1,2-*a*]pyridine core from **6** (723.6 mg, 4.0 mmol) and **3b** (1.30 g, 4.1 mmol), **10c** as a yellow-green solid (459.0 mg, 22.5%). This product contains ~50% 6-H impurity. ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.70 (AA'XX', 2 H), 7.53 (dd, J=9.77, 0.49 Hz, 1 H), 7.43 (d, J=1.95 Hz, 1 H), 7.00 (AA'XX', 2 H), 6.97 (dd, J=9.64, 2.32 Hz, 1 H), 3.87 (s, 3 H), 3.86 (s, 3 H), 3.08 (q, J=7.57, 2 H), 1.36 (t, J=7.57 Hz, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 158.96, 149.06, 141.96, 141.44, 129.12, 127.69, 122.00, 121.04, 118.14, 117.66, 113.93, 56.31, 55.26, 17.24, 11.96; HRMS (ESI): calc'd for $C_{17}H_{19}N_2O_2$ [M+H]$^+$ 283.1447, found 283.1449.

6-Methoxy-2-(4-methoxyphenyl)-3-*n*-propylimidazo[1,2-*a*]pyridine (10d). Following a modified version of the general procedure for the formation of the imidazo[1,2-*a*]pyridine core, **9** (520.2 mg, 3.24 mmol) and **3c** (1.36 g, 4.9 mmol) were condensed in *t*-BuOH (6 mL) with triethyl amine (450 µL) to yield **10d** as a red oil (205.3 mg, 21.4%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.70 (AA'XX', 2 H), 7.53 (dd, J=9.77, 0.49 Hz, 1 H), 7.42 (d, J=1.95 Hz, 1 H), 7.00 (AA'XX', 2 H), 6.97 (dd, J=9.64, 2.32 Hz, 1 H), 3.87 (s, 3 H), 3.86 (s, 3 H), 3.00 (t, J=7.81 Hz, 2 H), 1.76 (sext., J=8.06 Hz, 2 H), 1.05 (t, J=7.32 Hz, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 158.94, 149.01,

142.38, 141.42, 129.17, 127.72, 120.92, 118.10, 117.60, 113.91, 105.17, 56.34, 55.27, 25.98, 20.86, 14.15; HRMS (ESI): calc'd for $C_{18}H_{21}N_2O_2$ [M+H]$^+$ 297.1603, found 297.1597.

6-Iodo-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (11). Following the general procedure for the formation of the imidazo[1,2-*a*]pyridine core from **5** (16.55 g, 75.2 mmol) and 2-bromo-4'-methoxyacetophenone (17.12 g, 74.8 mmol), **11** (15.3g, 58.4%) was isolated as a beige crystaline solid after recrystallization (EtOAc/hexanes). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.35 (dd, J=1.59, 0.85 Hz, 1 H), 7.86 (AA'XX', 2 H), 7.70 (d, J=0.49 Hz, 1 H), 7.41 (m, 1 H), 7.31 (dd, J=9.40, 1.59 Hz, 1 H), 6.97 (AA'XX', 2 H), 3.86 (s, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 159.86, 146.29, 144.14, 132.31, 130.27, 127.40, 125.89, 118.28, 114.22, 106.87, 74.75, 55.32; HRMS (ESI) calc'd for $C_{14}H_{12}N_2OI$ [M+H]$^+$ 350.9994, found 350.9992.

6-isopropoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (12). Following the general procedure for the Ullmann-alkoxylation of aryl iodides from **11** (10.6 g, 30 mmol), CuI (1.2 g, 6 mmol), 1,10-phenanthroline (2.2 g, 12 mmol), and Cs$_2$CO$_3$ (19.5 g, 60 mmol), yielded crude material that was isolated by flash chromatography over silica gel (50% EtOAc in hexanes), and purified by recrystallization (EtOAc/hexanes) to give **12** (4.8 g, 57%) as a fluffy white solid. ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.82 (AA'XX', 2 H), 7.63 (s, 1 H), 7.62 (dd, J=2.32, 0.61 Hz, 1 H), 7.48 (d, J=9.52 Hz, 1 H), 6.93 (AA'XX', 2 H), 6.90 (d, J=2.20 Hz, 1 H), 4.34 (spt, J=6.06 Hz, 1 H), 3.81 (s, 3 H), 1.32 (d, J=6.10 Hz, 6 H); 13C-NMR (CDCl3, 125 mHz) δ 159.29, 146.58, 145.57, 142.75, 126.93, 126.60, 120.56, 117.13, 113.98, 111.30, 107.95, 72.11, 55.15, 21.77; HRMS (ESI) calc'd for $C_{17}H_{19}N_2O_2$ [M+H]$^+$ 283.1447, found 283.1435.

4-Chloro-pyridine-2-carboxylic acid methyl ester (13). Thionyl chloride (18 mL, 250 mmol) was placed in a round bottomed flask charged with argon. Picolinic acid (6.14 g, 50 mmol) was then added, followed by the slow addition of anhydrous DMF (0.8 mL, 10 mmol). The mixture was heated to 70°C and stirred for 44 hours. The reaction was quenched by the slow addition of anhydrous MeOH (20 mL) and Et$_3$N (5 mL), stirring for thirty minutes, and concentration by rotary evaporation. This process was repeated until the evolution of HCl (g) was no longer observed upon addition of MeOH/Et$_3$N. The products were purified by flash chromatography over silica gel (25% hexanes in EtOAc) to yield **13** as a green-yellow solid (6.02 g, 70%). ^{1}H NMR (CDCl$_3$, 400 MHz) δ 8.63 (dd, J = 5.13, 0.61 Hz, 1H), 8.12 (dd, J = 2.08, 0.61 Hz, 1H), 7.48 (dd, J = 5.25, 2.08 Hz, 1H), 4.00 (s, 3H); ^{13}C NMR (CDCl$_3$, 100 MHz) δ 164.56, 150.56, 149.11, 145.34, 127.08, 125.58, 53.16; HRMS (ESI) calc'd for $C_7H_6ClNO_2$ [M+H]$^+$ 172.0165, found 172.0168.

4-Methoxy-pyridine-2-carboxylic acid methyl ester (14). Sodium metal (720 mg, 31 mmol) was added to anhydrous MeOH (30 mL) stirring at 0°C under argon. The solution was allowed to warm to room temperature with continuous stirring until all of the sodium was dissolved. Subsequently, **13** was added.(2.57 g, 15 mmol) and heated to 90°C for 12 hours. The solvent was removed by rotary evaporation, and the products isolated by basic aqueous extraction with DCM, dried over MgSO$_4$, vacuum filtered, and concentrated by a second rotary evaporation to yield **14** as a light-yellow solid (1.02 g, 41%). ^{1}H NMR (CDCl$_3$, 400 MHz) δ 8.53 (dd, J = 5.62, 0.49 Hz, 1H), 7.67 (dd, J = 2.69, 0.49 Hz, 1H), 6.97 (dd, J = 5.74, 2.56 Hz, 1H), 3.99 (s, 3H), 3.91 (s, 3H); ^{13}C NMR (CDCl$_3$, 100 MHz) δ 166.47, 165.71, 150.87, 149.49, 113.08, 111.09, 55.50, 52.96; HRMS (ESI) calc'd for C$_8$H$_9$NO$_3$ [M+H]$^+$ 168.0661, found 168.0654.

4-Chloro-2-(2,5-dimethyl-pyrrol-1-yl)-pyridine (15). A round bottomed flask charged with 2-amino-4-chloropyridine(1.1g, 8.4 mmol), acetonyl acetone (2.0 mL, 16.5 mmol), molecular sieves (4Å), a catalytic amount of *p*-tosylic acid, and anhydrous toluene (30mL), was heated to reflux for 14 hours. The reaction mixture was cooled, concentrated by rotary evaporation, transferred to a separatory funnel, resuspended in saturated NaHCO$_3$ (aq), and exhaustively extracted with EtOAc. The resulting organic phase was dried over MgSO$_4$, vacuum filtered, and concentrated by rotary evaporation. The product was isolated by flash chromatography over silica gel (20% EtOAc in hexanes) to yield **15** as a light yellow-green oil (1.50 g, 86.5%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.52 (d, J=5.37 Hz, 1 H), 7.33 (dd, J=5.37, 1.71 Hz, 1 H), 7.25 (d, J=1.95 Hz, 1 H), 5.92 (s, 2 H), 2.16 (s, 6 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 153.10, 149.85, 145.14, 128.59, 122.53, 122.06, 107.47, 13.19; HRMS (ESI): calc'd for C$_{11}$H$_{12}$ClN$_2$ [M+H]$^+$ 207.0689, found 207.0682.

2-(2,5-Dimethyl-pyrrol-1-yl)-4-methoxypyridine (16). Potassium metal (1.13 g, 28.9 mmol) was carefully added to a round bottomed flask charged with anhydrous MeOH (30 mL) stirring at 0°C under argon. A solution of **15** (1.23 g, 6.0 mmol) dissolved in anhydrous MeOH (2mL) was slowly injected into the reaction vessel, followed by the addition of 18-crown-6-ether (159.4 mg, 0.6 mmol). The mixture was heated to 90°C for 14 hours, cooled, and concentrated by rotary evaporation. The crude product was transferred to a separatory funnel, resuspended in saturated NaHCO$_3$ (aq), and exhaustively extracted with DCM. The resulting organic phase was dried over MgSO$_4$, vacuum filtered, and concentrated by rotary evaporation. The product was isolated by flash chromatography over silica gel (25% EtOAc in hexanes) to yield **16** as an

amber oil (1.3g, 100%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.42 (d, J=5.86 Hz, 1 H), 6.85 (dd, J=5.86, 2.44 Hz, 1 H), 6.75 (d, J=2.44 Hz, 1 H), 5.90 (s, 2 H), 3.90 (s, 3 H), 2.15 (s, 6 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 166.99, 153.54, 150.07, 128.48, 108.84, 108.00 , 106.69, 55.45, 13.08; HRMS (ESI): calc'd for C$_{12}$H$_{15}$N$_2$O [M+H]$^+$ 203.1184, found 203.1184.

2-Amino-4-methoxypyridine (17). **16** (3.5 g, 17.4 mmol) was placed in a round bottomed flask and dissolved in an EtOH (60mL)/H$_2$O (30mL) mixture, to which NH$_2$OH·HCl (4.90g, 70.5 mmol) and KOH (2.63 g, 47.0 mmol) were added. The reaction mixture was refluxed for 18 hrs, cooled, and then concentrated by rotary evaporation. The resulting solid was resuspended in saturated NaHCO$_3$ (aq), transferred to a liquid-liquid continuous extractor, and extracted with DCM for 18 hours. The organic phase collected and concentrated by rotary evaporation. The product was isolated by flash chromatography over silica gel (20% IPA in DCM) to yield **17** as light beige solid (1.6g, 73.1%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.88 (d, J=5.86 Hz, 1 H), 6.25 (dd, J=5.86, 2.20 Hz, 1 H), 5.96 (d, J=2.20 Hz, 1 H), 4.51 (br. s., 2 H), 3.76 (s, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 167.20, 160.14, 149.14, 102.44, 92.31, 54.84; HRMS (ESI): calc'd for C$_6$H$_9$N$_2$O [M+H]$^+$ 125.0715, found 125.0721.

7-Methoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (18a). Following the general procedure for the formation of the imidazo[1,2-*a*]pyridine core from **17** (1.45 g, 11.7 mmol) and 2-bromo-4'-methoxyacetophenone (2.76 g, 12.0 mmol), **18a** was isolated as a white fluffy solid (1.88 g, 63.5%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.86 (dd, J=7.32, 0.73 Hz, 1 H) ,7.84 (AA'XX', 2 H), 7.58 (d, J=0.49 Hz, 1 H), 6.96 (AA'XX', 2 H), 6.89 (d, J=2.44 Hz, 1 H), 6.46 (dd, J=7.45, 2.56 Hz, 1 H), 3.84 (s, 3 H), 3.84 (s, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 159.27, 157.69, 146.98, 145.32, 126.93, 126.64, 125.76   114.00, 107.07, 105.82, 94.56, 55.37, 55.24; HRMS (ESI): calc'd for C$_{15}$H$_{15}$N$_2$O$_2$ [M+H]$^+$ 255.1134, found 255.1125.

3-Ethyl-7-methoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (18b). Following the general procedure for the formation of imidazo[1,2-*a*]pyridine core from **17** (185.3 mg, 1.5 mmol) and **3b** (442.6 mg, 1.65 mmol), to give **18b** (237.6 mg, 56.4%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.75 (d, J=7.32 Hz, 1 H), 7.70 (AA'XX', 2 H), 6.99 (AA'XX', 2 H), 6.90 (d, J=2.44 Hz, 1 H), 6.53 (dd, J=7.45, 2.56 Hz, 1 H), 3.85 (s, 6 H), 3.03 (q, J=7.57 Hz, 2 H), 1.32 (t, J=7.57 Hz, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 158.84, 157.11, 145.49, 140.85, 128.98, 127.70, 123.23, 119.79, 113.87, 106.70, 94.71, 55.36, 55.23, 16.97, 12.57; HRMS (ESI): calc'd for C$_{17}$H$_{19}$N$_2$O$_2$ [M+H]$^+$ 283.1447, found 283.1448.

6-Methoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine-3-carboxyaldehyde (19a). Phosphorus oxychloride (2.2 mL, 24 mmol) was injected into a round bottomed flask charged with anhydrous DMF (3.5 mL, 44 mmol) stirring at 0 °C under argon. The mixture was subsequently heated to 90°C for one hour, followed by the addition of a solution of **10a** (1.0 g, 4.0 mmol) dissolved in anhydrous DMF, and an additional 2.5 hours of heating. The solution was cooled to ambient temperature, poured over ice, and neutralized with 3M KOH (aq). The resulting mixture was transferred to a separatory funnel, saturated $NaHCO_3$ (aq) added, and exhaustively extracted with DCM. The organic phase was dried over $MgSO_4$, vacuum filtered, and concentrated by rotary evaporation. The crude material was recrystallized (EtOAc/hexanes) to give **19a** (645.1 mg, 56.7%) as strong dark yellow needles. ^{1}H-NMR ($CDCl_3$, 500 MHz) δ 10.02 (s, 1 H), 9.29 (dd, J=1.95, 0.49, 1 H), 7.76 (AA'XX', 2 H), 7.67 (dd, J=9.77, 0.49 Hz, 1 H), 7.33 (dd, J=9.52, 2.44 Hz, 1 H), 7.04 (AA'XX', 2 H), 3.92 (s, 3 H), 3.88 (s, 3 H); ^{13}C-NMR ($CDCl_3$, 125 MHz) δ 179.41, 160.95, 157.54, 151.07, 144.25, 130.95, 124.85 , 124.66 , 121.12, 117.03, 114.34, 110.84, 56.38, 55.38; HRMS (ESI): calc'd for $C_{16}H_{15}N_2O_3$ [M+H][+] 283.1083, found 283.1079.

6-isopropoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine-3-carboxaldehyde (19b). Following the same procedure described for **19a** from phosphorus oxychloride (1.7 mL, 18 mmol), anhydrous DMF (2.5 mL, 33 mmol), and a solution of **12** (857 mg, 3 mmol) dissolved in anhydrous DMF, **19b** (687.7 mg, 73.0%) was isolated as a fluffy white solid after isolation by flash chromatography over silica gel (50% EtOAc in hexanes) and purification by recrystallization (EtOAc/hexanes). ^{1}H-NMR ($CDCl_3$, 500 MHz) δ 10.02 (s, 1 H), 9.34 (d, J=2.44 Hz, 1 H), 7.76 (AA'XX', 2 H), 7.68 (d, J=9.77 Hz, 1 H), 7.32 (dd, J=9.52, 2.44 Hz, 1 H), 7.05 (AA'XX', 2 H), 4.59 (spt, J=6.06 Hz, 1 H), 3.89 (s, 3 H), 1.41 (d, J=6.10 Hz, 6 H); ^{13}C-NMR ($CDCl_3$, 125 MHz) δ 179.32, 160.97, 157.54, 149.04, 144.17, 130.97, 125.88, 124.82, 120.93, 116.98, 114.35, 113.80, 72.26, 55.39, 21.75; HRMS (ESI): calc'd for $C_{18}H_{19}N_2O_2$ [M+H][+] 311.1396, found 311.1382.

6-Methoxy-2-(4-methoxyphenyl)-3-vinylimidazo[1,2-a]pyridine (20a). To a solution of methyl triphenylphosphonium bromide (1.08 g, 3.0 mmol) in anhydrous THF (5 mL) cooled to -78 °C, *n*-BuLi (1.75mL, 1.6M in hexanes, 2.8 mmol) was added, stirred for 10 minutes and then allowed to warm to ambient temperature for an additional 30 minutes. Upon warming the solution changed from a white heterogeneous suspension to homogeneous yellow solution. The mixture was re-cooled to -78 °C, and a solution of **19a** (285.7 mg, 1.0 mmol) dissolved in THF (5 mL) was added. The reaction mixture was stirred at -78 °C for 20 minutes, allowed to

warm to ambient temperature for two hours, re-cooled to 0 °C, and quenched with distilled water.   The mixture was transferred to a separatory funnel, saturated $NaHCO_3$ (aq) added, added and exhaustively extracted with DCM.   The organic phase was dried over $MgSO_4$, vacuum filtered, and concentrated by rotary evaporation.   Crude product was isolated by flash chromatography over silica gel (15% acetone in DCM), and recrystallized (EtOAc/hexanes) to yield 20a (239.1 mg, 84.3%) as light green needles. ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.90 (d, J=2.20 Hz, 1 H), 7.74 (AA'XX', 2 H), 7.55 (dd, J=9.77, 0.49, 1 H), 7.02 (dd, J=9.77, 2.44 Hz, 1 H), 6.99 (AA'XX', 2 H), 6.93 (dd, J=18.31, 11.96 Hz, 1 H), 5.69 (dd, J=18.07, 0.73 Hz, 1 H), 5.53 (dd, J=11.96, 0.73 Hz, 1 H), 3.86 (s, 3 H), 3.85 (s, 3 H); HRMS (ESI): calc'd for $C_{17}H_{17}N_2O_2$ [M+H]$^+$ 281.1290, found 281.1287.

6-Isopropoxy-2-(4-methoxyphenyl)-3-vinylimidazo[1,2-*a*]pyridine (20b).   Following the same procedure for **20a** from methyl triphenylphosphonium bromide (1.07g, 3 mmol), *n*-BuLi (1.75mL, 1.6M in hexanes, 2.8 mmol), and **19b** (306.5 mg, 1 mmol) the crude product was isolated by flash chromatography (50% EtOAc in hexanes) to yield a yellow-green oil, which solidified upon standing, but was not amenable to recrystallization.   This material was further purified by MPLC using a CombiFlash® instrument (gradient elution, 0-50% EtOAc in hexanes) to yield **X** (178.4 mg, 58.6%) as a light beige solid. ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.00 (dd, J=2.20, 0.49 Hz, 1 H), 7.74 (AA'XX', 2 H), 7.60 (d, J=9.52 Hz, 1 H), 7.05 (dd, J=9.64, 2.32 Hz, 1 H), 6.99 (AA'XX', 2 H), 6.92 (dd, J=18.07, 11.96 Hz, 1 H), 5.68 (dd, J=18.07, 0.73 Hz, 1 H), 5.53 (dd, J=12.09, 0.61 Hz, 1 H), 4.46 (spt, J=6.06 Hz, 1 H), 3.86 (s, 3 H), 1.38 (d, J=6.10 Hz, 6 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) 159.45, 147.07, 144.70, 142.37, 130.03, 126.76, 124.63, 120.70, 118.76, 117.44, 114.99, 113.93, 110.81, 72.51, 55.33, 21.94; HRMS (ESI): calc'd for $C_{19}H_{21}N_2O_2$ [M+H]$^+$ 309.1603, found 309.1605.

3-Chloro-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (21a).   Following the general procedure for the halogentation of imidazo[1,2-*a*]pyridines from **4a** (453 mg, 2 mmol) and *N*-chlorosuccinimide (294 mg, 2.2 mmol), **21a** was isolated as a pale yellow solid (450 mg, 86%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.07 (AA'XX', 2H), 8.06 (dt, J = 6.86, 1.29, 1H), 7.59 (dt, J = 9.06, 1.04 Hz, 1H), 7.20 (ddd, J = 9.00, 6.86, 1.29 Hz, 1H), 6.99 (AA'XX', 2H), 6.87 (td, J = 6.86, 1.07 Hz, 1H), 3.84 (s, 3H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 159.59, 143.56, 139.67, 128.71, 125.11, 124.58, 122.52, 117.34, 112.64, 104.72, 55.23; HRMS (ESI); calc'd for $C_{14}H_{11}N_2OCl$ [M+H]$^+$ 259.0638, found 259.0635.

3-Bromo-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (22a). Following the general procedure for the halogentation of imidazo[1,2-*a*]pyridines from **4a** (452 mg, 2 mmol) and *N*-bromosuccinimide (398 mg, 2.2 mmol), **22a** was isolated as a grey solid (523 mg, 86%). ^{1}H - NMR (CDCl$_3$, 500 MHz) δ 8.13 (dt, J = 6.86, 1.07 Hz, 1H), 8.06 (AA'XX', 2H), 7.60 (dt, J = 9.22, 1.07 Hz, 1H), 7.22 (ddd, J = 9.11, 6.75, 1.29 Hz, 1H), 7.00 (AA'XX', 2H), 6.89 (td, J = 6.81, 1.18 Hz, 1H), 3.84 (s, 3H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 159.67, 145.34, 142.54, 129.13, 125.42, 124.88, 123.82, 117.35, 113.86, 112.83, 90.85, 55.26; HRMS (ESI): calc'd for C$_{14}$H$_{11}$N$_2$OBr [M +H]$^+$ 303.0133, found 303.0141.

3-Iodo-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (23a). Following the general procedure for the halogentation of imidazo[1,2-*a*]pyridines from **4a** (455 mg, 2 mmol) and *N*-iodosuccinimide (540 mg, 2.4 mmol), **23a** was isolated as a yellow-green solid (668 mg, 94%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.20 (dt, J = 6.86, 1.07 Hz, 1H), 8.00 (AA'XX', 2H), 7.58 (dt, J = 9.00, 1.07 Hz, 1H), 7.23 (ddd, J = 9.00, 6.86, 1.29 Hz, 1H), 7.00 (AA'XX', 2H), 6.90 (td, J = 6.86, 1.07 Hz, 1H), 3.24 (s, 3H); ^{13}C NMR (CDCl$_3$, 125 MHz) δ 159.73, 148.06, 129.75, 126.41, 126.07, 125.38, 117.36, 113.75, 112.98, 105.25, 55.28, 36.36; HRMS (ESI): calc'd for C$_{14}$H$_{11}$N$_2$OI [M+H]$^+$ 350.9994, found 350.0087.

3-Chloro-6-methoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (21b). Following the general procedure for the halogenation of imidazo[1,2-*a*]pyridines from **4a** (252.0 mg, 1.0 mmol) and *N*-chlorosuccinimide (141.6 mg, 1.1 mmol), **21b** was isolated as a light purple solid (262.1 mg, 91.6%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.05 (AA'XX', 2 H), 7.56 (d, J=2.44 Hz, 1 H), 7.50 (dd, J=9.64, 0.61 Hz, 1 H), 7.01 (AA'XX', 3 H), 3.87 (s, 3 H), 3.86 (s, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 159.42, 149.66, 140.61, 139.66, 128.45, 125.36, 119.79, 117.62, 113.88, 105.44, 104.25, 56.17, 55.26; HRMS (ESI): calc'd for C$_{15}$H$_{14}$N$_2$O$_2$Cl [M+H]$^+$ 289.0744, found 289.0737.

3-Bromo-6-methoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (22b). Following the general procedure for the halogenation of imidazo[1,2-*a*]pyridines from **4a** (127.4 mg, 0.5 mmol) and *N*-bromosuccinimide (98.7 mg, 0.55 mmol), **22b** was isolated as a while solid (135.6 mg, 81.2%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.05 (AA'XX', 2 H), 7.65 (dd, J=2.36, 0.43 Hz, 1 H), 7.51 (dd, J=9.65, 0.86 Hz, 1 H), 7.02 (AA'XX', 3 H), 3.89 (s, 3 H), 3.87 (s, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 159.50, 149.79, 142.47, 142.36, 128.89, 125.64, 120.05, 117.62, 113.82, 105.65, 91.68, 56.20, 55.27; HRMS (ESI): calc'd for C$_{15}$H$_{14}$N$_2$O$_2$Br [M+H]$^+$ 333.0234, found 333.0248.

3-Iodo-6-methoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (23b). *Method A.* Following the general procedure for the halogenation of imidazo[1,2-*a*]pyridines from **10a** (502.6 mg, 2.0 mmol) and *N*-iodosuccinimide (507.9 mg, 2.3 mmol), **23b** was isolated as a while solid (710.0 mg, 94.5%). *Method 2.* Potassium hydroxide (695 mg, 12 mmol) and acetonitrile (30 mL) were placed in a round bottomed flask and sonicated until the KOH pellets were partially solubilized. **4a** (1.53 g, 6 mmol) and iodine (1.68g, 6.6 mmol) were then added and stirred at ambient temperature. Small portions of additional iodine were added every 10 minutes until the reaction was complete by TLC (15% acetone in DCM). The reaction mixture was poured into a solution of saturated sodium thiosulfate, transferred to a separatory funnel, and exhaustively extracted with DCM. The organic phase was dried over $MgSO_4$, vacuum filtered, and concentrated by rotary evaporation. The crude product was isolated by flash chromatography over silica gel (15% acetone in DCM), and purified by recrystallization (EtOAc/hexanes) to yield **23b** (1.70 g, 74.3%) as a beige fluffy crystalline solid. ^{1}H-NMR ($CDCl_3$, 500 MHz) δ 7.99 (m, 2 H), 7.73 (d, J=2.44 Hz, 1 H), 7.51 (dd, J=9.64, 0.61 Hz, 1 H), 7.03 (m, 2 H), 3.91 (s, 3 H), 3.88 (s, 3 H); ^{13}C-NMR (CDl_3, 125 MHz) δ , 159.56, 149.86, 147.55, 144.82, 129.51, 126.05, 120.47, 117.45, 113.69, 108.51, 59.83, 56.23, 55.25; HRMS (ESI): calc'd for $C_{15}H_{14}N_2O_2I$ [M+H]$^+$ 381.0100, found 381.0097.

3-Chloror-6-isopropoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (21c). Following the general procedure for the halogenation of imidazo[1,2-*a*]pyridines from **12** (568 mg, 2.0 mmol) and *N*-chlorosuccinimide (272 mg, 2.0 mmol), **21c** (420 mg, 65.8%) was isolated as lavender crystals after recrystallization (EtOAc/hexanes). ^{1}H-NMR ($CDCl_3$, 500MHz) δ 8.06 (AA'XX', 2 H), 7.65 (dd, J=2.20, 0.49 Hz, 1 H), 7.55 (dd, J=9.64, 0.61 Hz, 1 H), 7.02 (AA'XX', 2 H), 4.51 (spt, J=6.06 Hz, 1 H), 3.87 (s, 2 H), 1.40 (d, J=6.10 Hz, 4 H); ^{13}C-NMR ($CDCl_3$, 125 MHz) δ ppm 159.83, 147.70, 140.84, 139.87, 128.83, 125.39, 121.37, 117.74, 114.23, 108.29, 105.43, 72.56, 55.53, 22.10; HRMS (ESI): calc'd for $C_{17}H_{18}N_2O_2Cl$ [M+H]$^+$ 317.1057, found 317.1057.

3-Bromo-6-isopropoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (22c). Following the general procedure for the halogenation of imidazo[1,2-*a*]pyridines from **12** (565 mg, 2.0 mmol) and *N*-bromosuccinimide (361 mg, 2.0 mmol) freshly recrystallized from acetic acid, **22c** (480 mg, 66.4%) was isolated as a fluffy white solid after recrystallization (EtOAc/hexanes). ^{1}H-NMR ($CDCl_3$, 500MHz) δ 8.05 (AA'XX', 2 H), 7.73 (d, J=1.95 Hz, 1 H), 7.55 (d, J=9.77 Hz, 1 H), 7.05 (dd, J=9.64, 2.32 Hz, 1 H), 7.01 (AA'XX', 2 H), 4.51 (spt, J=6.02 Hz, 1 H), 3.87 (s, 2 H),

1.40 (d, J=6.10 Hz, 4 H); ^{13}C-NMR (CDCl$_3$, 125MHz) δ 159.63, 147.52, 142.40, 142.28, 128.98, 125.39, 121.31, 117.43, 113.89, 109.53, 91.35, 72.35, 55.28, 21.86; HRMS (ESI): calc'd for C$_{17}$H$_{18}$N$_2$O$_2$Br [M+H]$^+$ 361.0552, found 361.0552.

3-Iodo-6-isopropoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (23c). *Method A.* Following the general procedure for the halogenation of imidazo[1,2-*a*]pyridines from **12** (2.80 g, 9.92 mmol) and *N*-iodosuccinimide (2.54 g, 11.2 mmol), **23c** (2.340 g, 57.7%) was isolated as a yellow-brown solid after recrystallization (EtOAc/Hexanes). *Method B.* Potassium hydroxide (1.0 g, 17.7 mmol) and acetonitrile (100 mL) were placed in a round bottomed flask and sonicated until the KOH pellets were partially solubilized. **12** (2.54g, 9.0 mmol) and iodine (2.59g, 10.2mmol) were then added and stirred at room temperature. Small portions of additional iodine were added every 30 minutes until the reaction was complete by TLC (50% EtOAc in hexanes). The reaction mixture was poured into a solution of saturated sodium thiosulfate, transferred to a separatory funnel, and exhaustively extracted with EtOAc. The organic phase was dried over MgSO$_4$, vacuum filtered, and concentrated by rotary evaporation. The crude product was passed through a silica plug with a 1:1 mixture of EtOAc in hexanes, and further isolated by MPLC using a CombiFlash® instrument (gradient elution, 25-75% EtOAc in hexanes). This isolated material was further purified by recrystallization (EtOAc/hexanes) to yield **23c** (2.33 g, 63.3%) as light pink crystals. ^{1}H-NMR (CDCl$_3$, 500MHz) δ 7.99 (AA'XX', 2 H), 7.81 (d, J=1.95 Hz, 1 H), 7.54 (d, J=9.52 Hz, 1 H), 7.06 (dd, J=9.77, 2.20 Hz, 1 H), 7.02 (AA'XX', 1 H), 4.52 (spt, J=6.02, 1 H), 3.87 (s, 3 H), 1.41 (d, J=6.10 Hz, 6 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 159.71, 147.68, 147.60, 144.91, 132.30, 129.63, 126.00, 121.67, 117.39, 113.79, 112.53, 72.45, 21.88; HRMS (ESI): calc'd for C$_{17}$H$_{18}$IN$_2$O$_3$ [M+H]$^+$ 409.0413, found 409.0399.

3-Chloro-7-methoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (21d). Following the general method for the halogenation of imidazo[1,2-*a*]pyridines from **18a** (205.4, 0.8 mmol) and *N*-chlorosuccinimide (116.2 mg, 0.9 mmol), **21d** was isolated as a light purple solid (180.3 mg, 77.4%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.06 (AA'XX', 2 H), 7.91 (dd, J=7.40, 0.54 Hz, 1 H), 7.01 (AA'XX', 2 H), 6.89 (d, J=2.36 Hz, 1 H), 6.63 (dd, J=7.50, 2.57 Hz, 1 H), 3.88 (s, 3 H), 3.87 (s, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 159.40, 158.00, 153.03, 144.98, 138.95, 128.42 , 125.33, 123.01, 113.89, 107.70, 94.70, 55.58, 55.26.

3-Bromo-7-methoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (22d). Following the general method for the halogenation of imidazo[1,2-*a*]pyridines from **18a** (148.8, 0.6 mmol) and *N*-bromosuccinimide (121.1 mg, 0.7 mmol), **22d** was isolated as an off-white solid (114.6

mg, 58.8%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.06 (AA'XX', 2 H), 7.97 (dd, J=7.57, 0.49 Hz, 1 H), 7.01 (AA'XX', 2 H), 6.92 (d, J=2.44 Hz, 1 H), 6.63 (dd, J=7.57, 2.44 Hz, 1 H), 3.88 (s, 3 H), 3.87 (s, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 159.48, 158.20, 146.59, 141.94, 128.83, 125.58 , 124.24 , 113.82, 107.79, 94.71, 88.92, 55.62, 55.28; HRMS (ESI): calc'd for C$_{15}$H$_{14}$N$_2$O$_2$Br [M+H]$^+$ 333.0239, found 333.0237.

3-Iodo-7-methoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (23d). Following the general method for the halogenation of imidazo[1,2-*a*]pyridines from **18a** (383.1 mg, 1.5 mmol) and *N*-iodosuccinimide (345.2 mg, 1.5 mmol), **23d** was isolated as a yellow crystalline solid (307.4 mg, 53.7%) after recrystallization (EtOAc/hexanes). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.01 (AA'XX', 2 H) 7.01 (AA'XX', 2 H), 6.90 (d, J=2.57 Hz, 1 H), 6.63 (dd, J=7.50, 2.36 Hz, 1 H), 3.88 (s, 3 H), 3.87 (s, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 159.72, 158.73, 149.03, 147.66, 129.52, 128.54, 126.75, 126.22, 113.80, 107.84, 94.88, 55.69, 55.30; HRMS (ESI): calc'd for C$_{15}$H$_{14}$N$_2$O$_2$I [M+H]$^+$ 381.0100, found 381.0101.

3-Cyano-6-methoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (24). To a pear shaped round bottomed flask **23b** (1.14 g, 3 mmol), dppf (334.8 mg, 0.6 mmol), Pd$_2$dba$_3$ (157.7 mg, 0.3 mmol), zinc dust (46.6 mg, 0.72 mmol), and zinc(II) cyanide (258.0 mg, 2.2 mmol) were added. The flask was evacuated and backfilled with argon (x3), dimethylacetamide (DMA, 12 mL) was injected, and the mixture was heated with rapid stirring to 130 °C for three hours. Upon cooling, the reaction mixture was transferred to a separatory funnel, suspended in saturated NaHCO$_3$ (aq), and exhaustively extracted with DCM. The resulting organic phase was dried over MgSO$_4$, vacuum filtered, and concentrated by rotary evaporation. The crude product was isolated by flash chromatography over silica gel (10% acetone in DCM) and further purified recrystallized (EtOAc/hexanes) following a hot filtration to yield **24** (475.6 mg, 56.8%) as fluffy light yellow crystalline solid. ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.10 (m, 2 H), 7.79 (d, J=2.36 Hz, 1 H), 7.60 (d, J=9.86 Hz, 1 H), 7.19 (dd, J=9.65, 2.36 Hz, 1 H), 7.01 (m, 2 H), 3.89 (s, 3 H), 3.87 (s, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 160.87, 152.72, 150.70, 143.52, 128.43, 123.92, 123.56, 117.88, 114.27, 113.36, 107.28, 93.80, 56.40, 55.34; HRMS (ESI): calc'd for C$_{16}$H$_{14}$N$_3$O$_2$I [M+H]$^+$ 280.1088, found 280.1088.

6-Methoxy-2-(4-methoxyphenyl)-3-phenylimidazo[1,2-*a*]pyridine (25a). Following the general procedure for SM coupling of aryl iodides from **23b** (195.0 mg, 0.5 mmol), phenylboronic acid (133.5 mg, 1.0 mmol), Pd(dppf)·CH$_2$Cl$_2$ (39.1 mg, 0.05 mmol), Cs$_2$CO$_3$ (500.1 mg, 1.5 mmol), and anhydrous THF (1.0 mL) the desired product was isolated by flash

chromatography over silica gel (20% acetone in DCM) to yield **25a** (150.4 mg, 88.8%) as off-white needles after recrystallization (DCM/hexanes). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.56 (m, 5 H), 7.48 (m, 4 H), 7.01 (dd, J=9.52, 2.44 Hz, 1 H), 6.82 (AA'XX', 2 H), 3.80 (s, 3 H), 3.72 (s, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 159.00, 149.23, 142.28, 141.77, 130.58, 130.18, 129.59, 129.07, 128.79, 126.72, 121.15, 119.54, 117.45, 113.68, 105.17, 56.20, 55.17; HRMS (ESI): calc'd for C$_{21}$H$_{19}$N$_2$O$_2$I [M+H]$^+$ 331.1447, found 331.1446.

6-Isopropoxy-2-(4-methoxyphenyl)-3-phenylimidazo[1,2-*a*]pyridine (25b). Following the general procedure for SM coupling of aryl iodides from **23c** (304.4mg, 0.75 mmol), phenylboronic acid (187.6 mg, 1.5 mmol), Pd(dppf)·CH$_2$Cl$_2$ (62.0 mg, 0.075 mmol), Cs$_2$CO$_3$ (758 mg, 2.3 mmol), and anhydrous THF (2mL), the desired product was isolated by flash chromatography over silica gel (50% EtOAc in hexanes) to yield **25b** (208.1 mg, 77.9%) as a beige solid after recrystallization (EtOAc/hexanes). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.63 (d, J=9.52 Hz, 1 H), 7.57 (AA'XX', 2 H), 7.50 (m, 6 H), 7.04 (dd, J=9.64, 2.32 Hz, 1 H), 4.33 (spt, J=6.06 Hz, 1 H), 3.80 (s, 3 H), 1.30 (d, J=6.10 Hz, 6 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 159.09, 146.81, 142.08, 141.70, 130.60, 129.94, 129.55, 129.18, 129.11, 128.83, 126.39, 120.78, 117.17, 113.72, 109.56, 72.35, 55.13, 21.79; HRMS (ESI): calc'd for C$_{23}$H$_{23}$N$_2$O$_2$ [M+H]$^+$ 359.1760, found 359.1760.

3-Cyclopropyl-6-methoxy-2-(4-methoxyphenyl)imidazo[1,2-a]pyridine (26a). Following a modification of the general procedure for SM coupling of aryl iodides from **23b** (381.2 mg, 1.0 mmol), cyclopropylboronic acid (172.0 mg, 2.0 mmol), Pd(dppf)·CH$_2$Cl$_2$ (84.9 mg, 0.01 mmol), and Cs$_2$CO$_3$ (996.5 mg, 3.0 mmol), a second equivalent of boronic acid was added after 30 minutes. The desired product was isolated by flash chromatography over silica gel (20% acetone in DCM) to yield **26a** (241.1 mg, 82.0%) as a brick red solid. This material could be further purified by dissolving in THF and adding conc. HCl dropwise. The resulting precipitate was filtered, washed with additional THF, dried, and redissolved in DCM. This solution was transferred to separatory funnel to which saturated NaHCO3 (aq) was added and extracted with DCM. The organic layer was dried over MgSO4, vacuum filtered, and concentrated by rotary evaporation. The obtained solid was recrystallized (DCM/hexanes) to yield purple rhomboidal crystals. ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.90 (AA'XX', 2 H), 7.78 (dd, J= 1.71, 0.73 Hz, 1 H), 7.52 (dd, J= 9.77, 0.73 Hz, 1 H), 7.00 (dd, J=7.32, 2.44 Hz, 1H), 6.99 (AA'XX', 2 H), 3.89 (s, 3 H), 3.86 (s, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 158.97, 148.92, 143.17, 140.88, 129.53, 126.98, 120.87, 118.73, 117.21, 113.49, 105.90, 56.28, 55.23, 7.67, 4.30; HRMS (ESI): calc'd for C$_{18}$H$_{19}$N$_2$O$_2$ [M+H]$^+$ 295.1447, found 295.1445.

3-Cyclopropyl-6-isopropoxy-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (26b). Following a modification of the general procedure for SM coupling of aryl iodides from **23c** (609.9 mg, 1.5 mmol), cyclopropylboronic acid (253.3mg, 3.0 mmol), Pd(dppf)·CHCl$_2$ (119.2 mg, 0.15 mmol), and Cs$_2$CO$_3$ (1.47 g, 4.5 mmol), a second equivalent of boronic acid was added after 30 minutes. The desired product was isolated by flash chromatography over silica gel (gradient elution, 50%-75% EtOAc in hexanes) to yield **26b** (299.4 mg, 62.2%) as a brown crystalline solid after recrystallization (EtOAc/hexanes). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.91 (AA'XX', 2 H), 7.88 (d, J=2.44 Hz, 1 H), 7.55 (d, J=9.52 Hz, 1 H), 7.02 (dd, J=9.64, 2.32 Hz, 1 H), 6.99 (AA'XX', 2 H), 4.49 (spt, J=6.10 Hz, 1 H), 3.87 (s, 2 H), 1.97 (tt, J=7.97, 5.22 Hz, 1 H), 1.40 (d, J=5.86 Hz, 4 H), 1.16 (m, 1 H), 0.52 (m, 1 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 159.01, 146.39, 143.14, 140.93, 129.58, 126.76, 120.54, 120.10, 117.02, 113.51, 110.27, 72.46, 55.24, 22.90, 7.69, 4.25; HRMS (ESI): calc'd for C$_{20}$H$_{23}$N$_2$O$_2$ [M+H]$^+$ 323.1760, found 323.1750.

6-Methoxy-2-(4-methoxyphenyl)-3-trifluoromethylimidazo[1,2-*a*]pyridine (27a). 10a (568.3 mg, 2.2 mmol) and 5-(trifluoromethyl)dibenzothiophenium trifluoromethanesulfonate (750.9 mg, 1.9 mmol) were suspended in anhydrous THF yielding a heterogenous orange solution and stirred at ambient temperature for 45 minutes. The reaction mixture was subsequently heated to 60°C for three hours, whereby the solution became homogeneous upon heating. The reaction was allowed to cool, transferred to a separatory funnel, suspended in saturated NaHCO$_3$ (aq), and exhaustively extracted with DCM. The resulting organic phase was dried over MgSO$_4$, vacuum filtered, and concentrated by rotary evaporation. The desired product was isolated by flash chromatography over silica gel (5% acetone in DCM) to yield **27a** (183.8 mg, 30.6%) as an amber solid. ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.78 (d, J=2.2 Hz, 1H), 7.62 (AA'XX', 2H), 7.60 (d, J=10 Hz, 2H), 7.16 (dd, J=9.6, 2.3 Hz, 1H), 6.98 (AA'XX', 2H), 3.88 (s, 3H), 3.87 (s, 3H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 160.16, 150.06, 147.26, 142.78, 130.73, 125.06, 123.02, 122.15, 120.89, 117.77, 113.64, 107.49, 56.29, 55.28; ^{19}F-NMR (CDCl$_3$, 470 MHz) δ -58.40; HRMS (ESI): calc'd for C$_{16}$H$_{14}$N$_2$O$_2$F$_3$ [M+H]$^+$ 323.1007, found 323.1002.

6-methoxy-2-(4-methoxyphenyl)-3-trifluoromethyl[1,2-*a*]pyridine (27b). Following the same procedure as for **28a** from **X** (253.1 mg, 1 mmol) and 5-(trifluoromethyl)dibenzothiophenium trifluoromethanesulfonate (411.2 mg, 1 mmol), and THF (2mL), the desired product was isolated by flash chromatography over silica gel (10% acetone in DCM) to yield **27b** (131.0 mg, 40.8%) as a light red solid. ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.10 (d, J=7.72 Hz, 1 H), 7.64 (AA'XX', 2 H), 6.99 (m, 3 H), 6.69 (dd, J=7.50, 2.57 Hz, 1 H), 3.91 (s, 4

H), 3.87 (s, 3 H); ^{13}C-NMR (CDCl$_3$, 500 MHz) δ 160.35, 159.40, 147.79, 147.62, 130.78, 125.85, 125.13 , 123.20, 121.07, 113.73, 108.77, 95.08, 55.73 , 55.31.

4-Imidazo[1,2-*a*]pyridine-2-ylphenol (28a).   Following the general procedure for the cleavage of aryl-alkyl ethers from **4a** (457 mg, 2 mmol), **28a** was isolated as a brown solid (216 mg, 50%). ^{1}H-NMR (DMSO-d$_6$, 500 MHz) δ 9.57 (s, 1H), 8.46 (dt, J = 6.71, 1.16 Hz), 8.19 (s, 1H), 7.76 (AA'XX', 2H), 7.51 (dd, J = 9.03, 0.49 Hz, 1H), 7.18 (ddd, J = 9.03, 6.84, 1.22, 1H), 6.83 (td, J = 6.59, 1.22 Hz), 6.82 (AA'XX', 2H); ^{13}C-NMR (DMSO-d$_6$, 125 MHz) δ 157.27, 144.90, 144.65, 126.95, 126.60, 124.94, 116.29, 115.47, 111.92, 107.54; HRMS (ESI): calc'd for C$_{13}$H$_{10}$N$_2$O [M+H]$^+$ 211.0871, found 211.0869.

4-(3-Chloroimidazo[1,2-*a*]pyridine-2-yl)phenol (28d).   Following the general procedure for the cleavage of aryl-alkyl ethers from **21a** (322 mg, 1.25 mmol), **28d** was isolated as a fine grey powder (316 mg, 104%). ^{1}H-NMR (DMSO-d$_6$, 500 MHz) δ 9.77 (s, 1H), 8.34 (dt, J = 6.86, 0.96 Hz, 1H), 7.90 (AA'XX', 2H), 7.63 (dt, J = 9.11, 0.91 Hz, 1H), 7.37 (ddd, J = 9.00, 6.86, 1.29 Hz, 1H), 7.08 (td, J = 6.86, 1.07 Hz, 1H), 6.89 (AA'XX', 2H); ^{13}C-NMR (DMSO-d$_6$, 125 MHz) δ 157.76, 142.75, 138.61, 128.42, 125.72, 123.37, 122.80, 116.53, 114.52, 113.38, 103.55. HRMS (ESI): calc'd for C$_{13}$H$_9$N$_2$OCl [M+H]$^+$ 245.0482, found 245.0476.

4-(3-Bromoimidazo[1,2-*a*]pyridine-2-yl)phenol (28e).   Following the general procedure for the cleavage of aryl-alkyl ethers from **22a** (382 mg, 1.25 mmol), **28e** was isolated as a grey powder (273 mg, 75%). ^{1}H-NMR (DMSO-d$_6$, 500 MHz) δ 9.74 (s, 1H), 8.34 (d, J = 6.43 Hz, 1H), 7.92 (AA'XX', 2H), 7.62 (d, J = 8.58 Hz, 1H), 7.35 (t, J = 7.50 Hz, 1H), 7.06 (t, J = 6.54 Hz, 1H), 6.89 (AA'XX', 2H); ^{13}C-NMR (DMSO-d$_6$, 125 MHz) δ 157.74, 128.77, 125.64, 124.46, 123.51, 116.70, 115.41, 113.41, 109.35; HRMS (ESI): calc'd for C$_{13}$H$_9$N$_2$OBr [M+H]$^+$ 288.9976, found 288.9976.

4-(3-Iodoimidazo[1,2-*a*]pyridine-2-yl)phenol (28f).   Following the general procedure for the cleavage of aryl-alkyl ethers from **23a** (206 mg, 0.57 mmol), **28f** was obtained and further purified by flash chromatography over silica gel (25% hexanes in EtOAc) to yield a white crystaline powder (94.8 mg, 48%). ^{1}H-NMR (DMSO-d$_6$, 500 MHz) δ 9.69 (s, 1H), 8.37 (dt, J = 6.86, 0.96 Hz, 1H), 7.88 (AA'XX', 2H), 7.37(dt, J = 8.95, 0.99 Hz, 1H), 7.32 (ddd, J = 8.95, 6.81, 1.18 Hz, 1H), 7.03 (td, J = 6.81, 1.18 Hz, 1H), 6.87 (AA'XX', 2H); ^{13}C-NMR (DMSO-d$_6$, 125 MHz) δ 157.58, 147.18, 146.97, 129.35, 126.87, 125.68, 124.55, 116.63, 115.17, 113.27, 48.62; HRMS (ESI): calcd for C$_{13}$H$_9$N$_2$OI [M+H]$^+$ 336.9838, found 336.9835.

4-(3-Methylimidazo[1,2-*a*]pyridin-2-yl)phenol (28b). Following the general procedure for the cleavage of aryl-alkyl ethers from **4b** (116 mg, 0.53 mmol), **28b** was isolated as a light brown solid (115 mg, 106%). ^{1}H-NMR (DMSO-d_6, 500 MHz) δ 9.60 (s, 1H), 8.28 (d, J = 6.43 Hz, 1H), 7.61 (AA'XX', 2H), 7.53 (d, J = 8.79 Hz, 1H), 7.21 (t, J = 7.61 Hz, 1H), 6.92 (t, J = 6.75 Hz, 1H), 6.87 (AA'XX', 2H), 2.58 (s, 3H); ^{13}C-NMR (DMSO-d_6, 125 MHz) δ 156.91, 129.15, 124.25, 123.73, 116.10, 115.42, 111.81, 9.41.   HRMS (ESI) calc'd for $C_{14}H_{12}N_2O$ [M+H]$^+$ 225.1028, found 225.1024.

4-(3-Ethylimidazo[1,2-*a*]pyridine-2-yl)phenol (28c). Following the general procedure for the cleavage of aryl-alkyl ethers from **4c** (252 mg, 1.2 mmol) **28c** was isolated as a beige solid (197 mg, 69.6%). ^{1}H-NMR (DMSO-d_6, 500 MHz) δ 9.61 (s, 1H), 8.39 (d, J = 6.65 Hz, 1H), 7.58 (m, 2H), 7.55 (AA'XX', 1H), 7.25 (ddd, J = 8.79, 6.86, 0.64 Hz, 1H), 6.94 (t, J = 6.75 Hz, 1H), 6.87 (AA'XX', 2H), 3.07 (q, J = 7.43 Hz, 2H), 1.24 (t, J = 7.50, 3H); ^{13}C-NMR (DMSO-d_6, 125 MHz) δ 157.02, 142.87, 140.15, 128.92, 125.06, 124.24, 120.81, 116.11, 115.46, 112.12, 59.76, 16.38, 12.19; HRMS (ESI): calc'd for $C_{15}H_{14}N_2O$ [M+H]$^+$ 239.1184, found 239.1188.

2-(4-hydroxyphenyl)imidazo[1,2-*a*]pyridin-6-ol (29a).   Following the general procedure for the cleavage of aryl-alkyl ethers from **10a** (254.2 mg, 1.0 mmol) and BF$_3$·SMe$_2$ (4.2 mL, 40 mmol), the desired product was isolated by flash chromatography over silica gel (20% IPA in DCM) to give **29a** as a purple solid (212.9 mg, 94.1%). ^{1}H NMR (DMSO-d_6, 400 MHz) δ 9.50 (s, 2H), 8.07 (s, 1H), 7.92 (dd, J = 2.20, 0.73 Hz, 1H), 7.70 (m, 2H), 7.38 (d, J = 9.52 Hz, 1H), 6.90 (dd, J = 9.52, 2.20 Hz, 1H), 6.80 (m, 2H); ^{13}C NMR (DMSO-d_6, 100 MHz) δ 156.98, 145.39, 144.65, 141.69, 126.62, 125.36, 119.38, 116.36, 115.42, 109.89, 107.90; HRMS (ESI): calcd for $C_{13}H_{10}N_2O_2$ [M + H]$^+$ 227.0821, found 227.0830.

2-(4-Hydroxyphenyl)-3-methylimidazo[1,2-*a*]pyridin-6-ol (29b). Following the general procedure for the cleavage of aryl-alkyl ethers from **10b** (256.8 mg, 0.72 mmol) and BF$_3$·SMe$_2$ (6.0 mL, 57 mmol), the desired product was isolated by flash chromatography over silica gel (10% MeOH in DCM) to give **29b** as brown solid (49.8 mg, 30.9%). ^{1}H-NMR (CD$_3$OD, 500 MHz) δ 7.85 (d, J=1.71 Hz, 1 H), 7.59 (d, J=9.52 Hz, 1 H), 7.49 (AA'XX', 2 H), 7.41 (dd, J=9.52, 2.20 Hz 1 H), 6.94 (AA'XX', 2 H), 4.96 (br. s., 2 H) 2.56 (s, 3 H); ^{13}C-NMR (CD$_3$OD, 125 MHz) δ 160.08, 150.29, 137.89, 136.73, 131.06, 125.55, 121.56, 118.84, 117.12, 113.90, 110.57, 9.15; HRMS (ESI): calc'd for $C_{14}H_{13}N_2O_2$ [M+H]$^+$ 241.0977, found 241.0971.

3-Ethyl-2-(4-hydroxyphenyl)imidazo[1,2-*a*]pyridin-6-ol (29c). Following the general procedure for the cleavage of aryl-alkyl ethers from **10c** (518.2 mg, 1.8 mmol) and $BF_3 \cdot SMe_2$ (8.0 mL, 76 mmol), the desired product was isolated by flash chromatography over silica gel (10% MeOH in DCM) to give **29c** as brown solid (193.3 mg, 82.9%). ^{1}H-NMR (CD_3OD, 500 MHz) δ 7.87 (d, J=17.82 Hz, 1 H), 7.53 (d, J=9.77 Hz, 1 H), 7.48 (AA'XX', 2 H), 7.30 (dd, J=9.52, 2.20 Hz, 1 H), 6.93 (AA'XX', 2 H), 4.95 (br. s., 2 H), 3.04 (q, J=7.57 Hz, 2 H), 1.32 (t, J=7.57 Hz, 3 H); ^{13}C-NMR (CD_3OD, 125 MHz) δ 159.72, 149.66, 139.20, 130.83, 124.33, 123.86, 123.14, 116.99, 115.02, 109.96, 79.59, 17.55, 12.11; HRMS (ESI): calc'd for $C_{15}H_{15}N_2O_2$ [M+H]$^+$ 255.1134, found 255.1130.

2-(4-Hydroxyphenyl)-3-*n*-propylimidazo{1,2-*a*]pyridin-6-ol (29d). Following the general procedure for the cleavage of aryl-alkyl ethers from **10d** (197.9 mg, 0.67 mmol) and $BF_3 \cdot SMe_2$ (5.0 mL, 48 mmol), the desired product was isolated by flash chromatography over silica gel (10% MeOH in DCM) to give **29d** as green-yellow solid (166.8 mg, 90.1%). ^{1}H-NMR (CD_3OD, 500 MHz) δ 8.06 (dd, J=2.20, 0.73 Hz, 1 H), 7.73 (dd, J=9.77, 0.73 Hz, 1 H), 7.60 (dd, J=9.64, 2.08 Hz, 1 H), 7.50 (AA'XX', 2 H), 6.99 (AA'XX', 2 H), 4.89 (s, 2 H), 3.06 (t, J=7.81 Hz, 2 H), 1.76 (sxt, J=7.57 Hz, 2 H), 1.02 (t, J=7.45 Hz, 2 H); ^{13}C-NMR (CD_3OD, 125 MHz) δ 160.87, 151.25, 136.32, 134.42, 131.21, 127.69, 123.62, 119.34, 117.45, 113.08, 111.41, 25.51, 21.45, 14.16; HRMS (ESI): calc'd for $C_{16}H_{17}N_2O_2$ [M+H]$^+$ 269.1290, found 269.1299.

3-Cyclopropyl-2-(4-hydroxyphenyl)imidazo[1,2-*a*]pyridin-6-ol (29e). Following the general procedure for the cleavage of aryl-alkyl ethers from **26b** (159.8 mg, 0.5 mmol) and $BF_3 \cdot SMe_2$ (1.05 mL 10 mmol), the desired product was purified by MPLC using a CombiFlash® instrument (gradient elution, 0-20% IPA in DCM) to yield **29e** (37.4 mg, 28.3%) as a brown solid. ^{1}H-NMR (DMSO-d_6, 500 MHz) δ 9.54 (s, 1 H), 9.48 (s, 1 H), 7.87 (dd, J=2.20, 0.73 Hz, 1 H), 7.73 (AA'XX', 2 H), 7.40 (dd, J=9.52, 0.73 Hz, 1 H), 6.95 (dd, J=9.52, 2.20 Hz, 1 H), 6.82 (AA'XX', 2 H), 2.02 (tt, J=7.84, 5.22 Hz, 1 H), 1.13 (m, 2 H), 0.34 (m, 2 H); ^{13}C-NMR (DMSO-d_6, 125 MHz) δ 156.55, 145.22, 142.43, 139.82, 129.07, 125.60, 119.48, 118.79, 116.43 , 114.81, 107.60, 7.61, 4.12; HRMS (ESI): calc'd for $C_{16}H_{15}N_2O_2$ [M+H]$^+$ 267.1143, found 267.1135.

3-Chloro-2-(4-hydroxyphenyl)imidazo[1,2-*a*]pyridin-6-ol (29f). Following the general procedure for the cleavage of aryl-alkyl ethers from **21b** (288 mg, 1.0 mmol) and $BF_3 \cdot SMe_2$ (8.5 mL 80 mmol), the crude product was obtained as an oil and was triturated from hexanes to give **29f** (229.6 mg, 88.0%) as a grey solid. ^{1}H-NMR (CD_3OD, 500 MHz) δ 8.07 (dd, J=2.14, 0.64 Hz, 1 H), 7.79 (dd, J=9.65, 0.64 Hz, 1 H), 7.73 (AA'XX', 2 H), 7.68 (dd, J=9.65, 2.14 Hz, 1 H),

6.97 (AA'XX', 2 H), 4.95 (br. s., 2 H); ^{13}C-NMR (CD$_3$OD, 500 MHz) δ 161.51, 152.07, 135.93, 133.37, 130.55, 128.92, 117.41, 117.30, 113.48, 110.51, 109.47, 49.15; HRMS (ESI): calc'd for C$_{13}$H$_{10}$N$_2$O$_2$Cl [M+H]$^+$ 261.0431, found 261.0422.

3-Bromo-2-(4-hydroxyphenyl)imidazo[1,2-a]pyridin-6-ol (29g). Following the general procedure for the cleavage of aryl-alkyl ethers from **22b** (103.2 mg, 0.31 mmol) and BF$_3$·SMe$_2$ (1.3 mL, 12.4 mmol), the desired product was isolated by flash chromatography over silica gel (15% IPA in DCM) to yield **29g** (25.4 mg, 26.9%) as dark brown solid. ^{1}H-NMR (CD$_3$OD, 500 MHz) δ 7.88 (m, 1 H) 7.67 (AA'XX', 2 H), 7.37 (d, J=9.52 Hz, 1 H), 7.03 (dd, J=9.64, 2.32 Hz, 1 H), 6.83 (AA'XX', 2 H), 4.96 (br. s., 2 H); ^{13}C-NMR (CD$_3$OD, 125 MHz) δ 158.93, 148.09, 146.08, 143.30, 128.41, 126.03, 122.10, 116.69, 116.50, 111.69, 109.56; HRMS (ESI): calc'd for C$_{13}$H$_{10}$N$_2$O$_2$Br [M+H]$^+$ 304.9926, found 304.9918.

2-(4-Hydroxyphenyl)-3-iodoimidazo[1,2-*a*]pyridin-6-ol (29h). Following the general procedure for the cleavage of aryl-alkyl ethers from **23b** (193.7 mg, 0.5 mmol) and BF$_3$·SMe$_2$ (2.1 mL, 20 mmol), the desired product was isolated by flash chromatography over silica gel (15% IPA in DCM) to yield **29h** (76.9 mg, 42.9%) as a purple glassy solid. ^{1}H-NMR (CD$_3$OD, 500 MHz) δ 7.40 (m, 1 H), 6.98 (AA'XX', 2 H), 6.90 (d, J=9.52 Hz, 1 H), 6.71 (dd, J=9.64, 2.08 Hz, 1 H), 6.23 (AA'XX', 2 H), 4.39 (br. s., 2 H); ^{13}C-NMR (CD$_3$OD, 125 MHz) δ 160.30, 150.08, 141.01, 139.42, 128.65, 125.87, 121.30, 117.11, 114.06, 113.11, 110.32; HRMS (ESI): calc'd for C$_{13}$H$_{10}$N$_2$O$_2$I [M+H]$^+$ 352.9787, found 352.9795.

2-(4-Hydroxyphenyl)-3-phenylimidazo[1,2-*a*]pyridin-6-ol (29i). Following the general procedure for the cleavage of aryl-alkyl ethers from **25b** (92.2 mg, 0.26 mmol) and BF$_3$·SMe$_2$ (0.53 mL 5 mmol). The crude material was purified by purified by MPLC using a CombiFlash® instrument (gradient elution, 0-20% IPA in DCM) to yield **29i** (47.8 mg, 61.5%) as light brown solid. ^{1}H-NMR (DMSO-d$_6$, 500 MHz) δ 9.45 (s, 1 H), 9.44 (s, 1 H), 7.58 (m, 2 H), 7.49 (m, 5 H), 7.36 (AA'XX', 2 H), 6.97 (dd, J=9.64, 2.32 Hz, 1 H), 6.65 (AA'XX', 2 H); ^{13}C-NMR (DMSO-d$_6$, 125 MHz) δ 156.72, 145.75, 141.55, 140.71, 130.53, 130.15, 129.57, 128.71, 128.48, 125.31, 119.67, 119.49, 116.81, 115.01, 106.34; HRMS (ESI): calc'd for C$_{19}$H$_{15}$N$_2$O$_2$ [M+H]$^+$ 303.1134, found 303.1128.

3-Cyano-2-(4-hydroxyphenyl)imidazo[1,2-*a*]pyridin-6-ol (29j). **24** (286.3 mg, 1.0 mmol) was placed in a pear shaped round bottomed flask followed by enough pyridine·HCl to completely cover the starting material. The solid mixture was then heated to 220 °C resulting in

the formation of a melt which was stirred for two hours. The reaction mixture solidified upon cooling, and was resuspended in distilled water with sonication. The solution was transferred to a separatory funnel and exhaustively extracted with EtOAc. The organic phase was dried over $MgSO_4$, vacuum filtered, and concentrated by rotary evacuation. The desired product was isolated by flash chromatography over silica gel (100% EtOAc) to yield **29j** (92.6 mg, 36.0%) as light beige solid. ^{1}H-NMR (DMSO-d_6, 500 MHz) δ 10.20 (s, 1 H), 9.96 (s, 1 H), 7.93 (dd, J=2.14, 0.64 Hz, 1 H), 7.89 (AA'XX', 2 H), 7.67 (d, J=9.65 Hz, 1 H), 7.27 (dd, J=9.54, 2.25 Hz, 1 H), 6.92 (AA'XX', 2 H); ^{13}C-NMR (DMSO-d_6, 125 MHz) δ 158.95, 151.77, 147.83, 142.85, 128.05, 123.80, 122.47, 117.52, 115.84, 113.48, 109.97, 91.80; HRMS (ESI): calc'd for $C_{14}H_{10}N_3O_2$ [M+H]$^+$ 252.0773, found 252.0771.

2-(4-Hydroxyphenyl)-3-trifluoromethylimidazo[1,2-*a*]pyridin-6-ol (29k). **27a** (183.8 mg, 0.57 mmol) was redissolved in anhydrous DCM (4 mL), transfered to a round bottomed flask, and cooled to 0°C in an ice bath under an atmosphere of argon. BBr$_3$ (3.0 mL of 1.0 M solution) was added slowly, and the reaction was allowed to warm to ambient temperature with stirring overnight. The reaction mixture was re-cooled to 0°C, quenched with MeOH (10 mL), and concentrated by streaming N$_2$. The resulting solid was transfered to a separatory funnel, to which distilled water and saturated NaHCO$_3$ (aq.) was added and exhaustively extracted with EtOAc. The organic layer was dried over MgSO$_4$, vacuum filtered, and concentrated by rotary evaporation. The desired product was isolated by flash chromatography over silica gel (25% hexanes in EtOAc) to yield **29k** (27.3 mg, 16.2%) as an orange solid. ^{1}H-NMR (DMSO-d_6, 500 MHz) δ 10.04 (s, 1 H), 9.74 (s, 1 H), 7.90 (dd, J=1.39, 0.75 Hz, 1 H), 7.65 (d, J=9.86 Hz, 1 H), 7.42 (m, 2 H), 7.24 (dd, J=9.65, 2.14 Hz, 1 H), 6.85 (m, 2 H); 13C-NMR (DMSO-d_6, 125 MHz) δ 158.02, 147.20, 142.33, 130.47, 123.65, 123.24, 122.53, 121.11, 117.60, 114.99, 109.07; HRMS (ESI): calc'd for $C_{14}H_{10}N_2O_2F_3$ [M+H]$^+$ 295.0694, found 295.0688.

2-(4-hydroxyphenyl)imidazo[1,2-a]pyrdin-7-ol (30a). Following the general procedure for the cleavage of aryl-alkyl ethers from **18a** (95.8 mg, 0.38 mmol) and BF$_3$·SMe$_2$ (3.2 mL, 30.4 mmol), **30a** (79.3 mg, 93.0%) was isolated as light beige. ^{1}H-NMR (CD$_3$OD, 500 MHz) δ 8.39 (d, J=7.32 Hz, 1 H), 7.96 (s, 1 H), 7.53 (AA'XX', 2 H), 6.90 (dd, J=7.32, 2.44 Hz, 1 H), 6.86 (m, 3 H), 5.00 (br. s., 2 H); ^{13}C-NMR (CD$_3$OD, 125 MHz) δ , 164.98, 160.70, 144.61, 136.95, 131.10, 128.48 , 119.17, 117.29, 112.32, 108.91, 94.07; HRMS (ESI): calc'd for $C_{13}H_{11}N_2O_2$ [M+H]$^+$ 227.0821, found 227.0810.

3-Ethyl-2-(4-hydroxyphenyl)imidazo[1,2-*a*]pyrin-7-ol (30b).   Following the general procedure for the cleavage of aryl-alkyl ethers from **18b** (237.6 mg, 0.84 mmol) and BF$_3$·SMe$_2$ (7.0 mL, 80 mmol), to yield **30b** (56.7 mg, 26.5%). ^{1}H-NMR (CD$_3$OD, 500 MHz) δ 8.52 (d, J=7.32 Hz, 1 H), 7.47 (AA'XX', 2 H), 7.05 (dd, J=7.32, 2.20 Hz, 1 H), 6.97 (AA'XX', 2 H), 6.96 (m, 1 H), 4.94 (br. s., 2 H) ,3.07 (q, J=7.57 Hz, 2 H), 1.35 (t, J=7.57 Hz, 3 H).

3-Benzyloxypyridine (31).   To a round bottomed flask charged with NaH (250.9 mg, 6 mmol, 60% by wt. in mineral oil) stirring in anhydrous DMF (10 mL) and cooled to 0°C under argon, 3-hydroxypyridine (578.4 mg, 6 mmol) was added and stirred for 15 minutes.   Benzyl bromide (0.6 mL, 5 mmol) was slowly injected and the reaction was allowed to warm to ambient temperature and proceed overnight (~20 hrs).   The reaction mixture was transferred to a separatory funnel, distilled water added, and exhaustively extracted with EtOAc.   The organic layer was dried over MgSO$_4$, vacuum filtered, and concentrated by rotary evaporation.   The desired product was purified by flash chromatography over silica gel (25% hexanes in EtOAc) to yield **31** (212.1 mg, 22.7%) as a yellow oil. ^{1}H-NMR (CDCl$_3$, 400 MHz) δ 8.41 (dd, J=2.81, 0.61 Hz, 1 H), 8.24 (dd, J=4.39, 1.46 Hz, 1 H), 7.41 (m, 5 H), 7.24 (m, 2 H), 5.12 (s, 1 H).

5-benzyloxy-*N*-*t*-butylpyridin-2-amine (32).   To a round bottomed flask charged with **31** (400 mg, 2.2 mmol) dissolved in DCM, *m*-CPBA (497 mg, 2.2 mmol, 77% max. by wt.) was added and stirred overnight.   The solvent was removed by rotary evaporation and dried under reduced pressure.   The crude material was dissolved in anhydrous toluene (10 mL) to which TsCl (839 mg, 4.3 mmol) was added, cooled to 0°C, stirred or 15 minutes, and followed by the addition of *t*-butyl amine (2.4 mL, 22 mmol).   The reaction was allowed to proceed for 4 hours, transferred to a separatory funnel, saturated NaHCO$_3$ (aq) was added, and exhaustively extracted with DCM.   The organic layer was dried over MgSO$_4$, vacuum filtered, and concentrated by rotary evaporation.   The desired product was purified by flash chromatography over silica gel (25% EtOAc in hexanes) to yield **32** (201.1 mg, 36.3%) as a green oil.   1H-NMR (CDCl3,) δ 7.72 (dd, J=5.1, 1.5 Hz, 1H), 7.41 (m, 4H), 7.36 (m, 1H), 6.85 (dd, J=7.7, 1.5 Hz, 1H), 6.44 (dd, J=7.7, 1.5 Hz, 1H), 5.05 (s, 2H), 4.94 (s, 1H), 1.49 (s, 9H).

5-methoxy-*N*-*t*-butylpyridin-2-amine (33). A round bottomed flask was charged with 4-methoxypyridine-*N*-oxide monohydrate (2.00 g, 16 mmol), *t*-butyl amine (8.4 mL, 80 mmol), and anhydrous toluene (140 mL). The reaction flask was cooled to 0°C and TsCl (6.11 g, 32 mmol) was added in slowly in small portions over 5 minutes. The reaction was stirred maintaining 0°C for 4 hours, quenched with distilled water, transferred to a separatory funnel, and exhaustively extracted with DCM.   The organic layer was dried over MgSO$_4$, vacuum filtered and

concentrated by rotary evaporation. The desired product was purified by flash chromatography over silica gel (25% EtOAc in hexanes) to yield **34** (2.78 g, 96.4%) as an orange oil. ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.91 (d, J=6 Hz, 1H), 6.16 (dd, J=5.9, 2.3 Hz, 1H), 5.95 (d, J=2.1 Hz, 1H), 3.76 (s, 3H), 1.41 (s, 9H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 166.61, 159.92, 149.27, 129.43, 126.96, 100.32, 92.59, 54.78, 50.56, 30.11, 29.50.

2-Bromo-1-(4-methoxyphenyl)-ethanone *O*-methyloxime (34).   2-Bromo-4'-methoxyacetophenone (921 mg, 4 mmol), methoxyamine hydrochloride (515 mg, 6 mmol), and sodium carbonate (635 mg, 6 mmol) were dissolved in MeOH (10 mL) and stirred at room temperature for two hours. The reaction mixture was transferred to a separatory funnel, distill water added, and exhaustively extracted with DCM. The organic layer was dried over MgSO$_4$, vacuum filtered, and concentrated by rotary evaporation. The desired product was purified by flash chromatography over silica gel to yield **34** (741 mg, 71%) as a light yellow oil. ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.65 (AA'XX', 2H), 6.91 (AA'XX', 2H), 4.33 (s, 2H), 4.06 (s, 3H), 3.83 (s, 3H); ^{13}C-NMR (CDCl$_3$, 100 MHz) δ 160.77, 152.19, 127.39, 125.83, 114.02, 62.61, 55.31, 18.06; HRMS (ESI): calc'd for C$_{10}$H$_{12}$BrNO$_2$ [M+H]$^+$ 258.0130, found 258.0117.

1-[2-Methoxyimino-2-(4-methoxyphenyl)-ethyl]-pyridinium bromide (35).   34 (550 mg, 2.1 mmol) was reacted with neat pyridine (1 mL) and stirred at ambient temperature for one hour. The excess pyridine was removed by concentration under reduced pressure to yield **35** (674 mg, 94%) as a white solid. ^{1}H-NMR (DMSO-d$_6$, 500 MHz) δ 8.99 (dd, J = 6.47, 1.10 Hz, 2H), 8.58 (tt, J = 7.81, 1.34 Hz, 1H), 8.11 (m, 2H), 7.68 (m, 2H), 6.96 (m, 2H), 6.01 (s, 2H), 3.94 (s, 3H), 3.76 (s, 3H); ^{13}C-NMR (DMSO-d$_6$, 125 MHz) δ 160.84, 150.24, 146.27, 145.12, 128.34, 128.28, 124.13, 114.32, 62.60, 55.36, 54.18; HRMS (ESI) calc'd for C$_{15}$H$_{17}$N$_2$O$_2$ [M-Br]$^+$ 257.1290, found 257.1289.

MOLECULAR MODELING

Ligand Docking.

The structures for ERα (accession code: 2QGW) and ERβ (accession code: 1U9E) were obtained from the PDB databank and prepared using the Molecular Operating Environment (MOE). Any excess chains were removed to yield the monomeric structure, to which explicit hydrogens were added, followed by minimization to a termination gradient of 0.5 using MMFF94x charges. The coactivator peptide, ligand, and all waters were deleted, with the exception of the single water molecule involved in the hydrogen bonding network within the

binding pocket.   The resulting structure was further processed using AutoDock Tools (ADT) to compute Gasteiger charges and assign AD4 atom types.

All docked compounds were constructed in MOE and minimized to a termination gradient of 0.5 using MMFF94x charges.   The ligands were prepared for docking using AutoDock Tools to assign AD4 atom types, calculate Gasteiger charges, and set all rotatable bonds as active torsions.   Each ligand was docked into the receptor using AutoDock Vina (ADV). The grid box was centered on the crystal structure ligand, and measured 20 Å by 20 Å by 20 Å.   The exhaustiveness parameter was set to 100 (default=8, linear scale); all other default settings were used.

The top ten poses for each ligand were visually inspected in MOE. Unreasonable poses were discarded, and the lowest energy conformation for each unique pose was selected for further use. The receptor structure and each selected pose were merged into a single file, all hydrogens were explicitly added, MMFF94x partial charges added, and the protein-ligand complex was minimized. This minimization was conducted in four stages: stage 1 minimized the hydrogen atoms that were added to the structure in the previous step; stage 2 minimized the ligand; stage 3 minimized any residues within 4.5 Å of the ligand; and stage 4 minimized both the ligand and any residues within 4.5 Å. The raw docking output, receptor-ligand complex prior to the multistage minimization (denoted "merged structure"), and the minimized structure (denoted as "worked-up structure") were used in subsequent calculations and analyses.

Docking Rescore

The raw docking output file for each selected pose was re-scored using FRED.   The prepared structure was used as the receptor input, and the following scoring algorithms were used: plp, chemgauss2, chemgauss3, shapegauss, oechemscore, screenscore, zapbind, and consensus.

Interaction and Binding Energies

Each of the merged and worked-up structures was opened in MOE and the total energy of the receptor-ligand complex was calculated (E_{RL}). The ligand was deleted from the structure and the receptor energy was calculated (E_R). The pocket residues (within 4.5A of the ligand) and associated water molecule were minimized to a termination gradient of 0.01, and the minimized receptor energy was calculated (E_{Rmin}). The worked up structure was reloaded, the receptor was deleted, and the ligand energy was calculated (E_L).   Using these values, the minimized ligand energy (E_{Lmin}) was calculated following a minimization of the ligand to a termination gradient of 0.01.   The interaction and binding energies were calculated from equations **(1)** and **(2)**, respectively.

$$E_{interaction} = E_{RL} - (E_R + E_L) \qquad\qquad (2.1)$$
$$E_{binding} = E_{RL} - (E_{Rmin} + E_{Lmin}) \qquad\qquad (2.2)$$

Hydrogen Bonding Score

The merged and worked-up structures for each selected pose was opened in MOE and the hydrogen bonds formed between the ligand and protein were scored using the "dock_HydrogenBonds" module found in the standard scientific vector language (SVL) library for MOE. This module forms the basis for the "ligand interactions" feature accessible from MOE's graphical user interface. The mean score was subsequently calculated as the geometric mean of the individual scores.

Protonation Energetics

Relative Basicity, E(PT).   The neutral species for each compound was constructed in Spartan'08 and minimized using the toolbar minimization function. The optimal geometry in the gas phase was calculated at the Hartree-Fock level of theory using the 6-31G* basis set. Setting the calculation to start from the MMFF geometry reduced the number of failed optimization calculations; however, was not necessary in all cases. The single point energy was subsequently calculated at the DFT level of theory using B3LYP functionals and the 6-311+G** basis set, starting from the current geometry. The protonated species of each compound was generated by modifying N1 to adopt a square-planar geometry using the inorganic toolkit and adjusting the adjacent carbon bonds as necessary. The geometry optimization and single point energy calculations were performed as previously described, noting that the charge must be adjusted to +1. The relative basicity was calculated relative to imidazole from equation (**3**)

$$E(PT) = (E_{compound-H}{}^+ + E_{imidazole}) - (E_{compound} + E_{imidazole-H}{}^+) \qquad (2.3)$$

Proton Affinity, P(A).   The neutral and protonated species for each compound were constructed in MOE, minimized to a termination gradient of 0.01, and saved as pdb files. The atoms and atom coordinates were extracted using Open Babel conversion to the text format, and used to generate a GAMESS input file for calculating the optimized geometry in the gas phase at the Restricted Hartree-Fock level of theory using the 6-31G* basis set.   The atom coordinates were parsed from the geometry optimization output file and used to generate a new GAMESS input file to calculate a single point energy at the DFT level of theory using B3LYP functionals and the 6-311+G** basis set.   Finally, the total energy and zero point energy were parsed from the output file, and used to calculate the P(A) for each compound from equation (**4**).

$$P(A) = -(E_{H+} - E_{neutral}) - (ZPE_{H+} - ZPE_{neutral}) + (5/2)RT \qquad \textbf{(2.4)}$$

Calculation of Electrostatic Potential, HOMO and LUMO, and Dipole Moment.

The vector or surface for each metric was taken from the single point energy calculated in Spartan '08, as described in section **V** (B, C, D). The dipole moment is projected onto the structure, originating on the center of mass. The electrostatic potential is projected onto a fixed electron density isosurface at 0.002 au. The HOMO and LUMO are shown at the 0.032 au surface.

pK$_a$ Prediction

Each compound was drawn separately in the Marvin Sketch editor, and the pK$_a$ predicted using the pK$_a$ module accessible from the "Tools/Protonation" menu.   The optimal pH for obtaining the maximal concentration of the neutral species was taken from the distribution chart option, available from the "pK$_a$ options" window.

Dipole Moment of ERβ Ligand Binding Pocket.

The structure for ERβ (PDB accession code: 1UE9) was opened in MOE and all hydrogens were explicitly added. All residues containing an atom within 4.5 Å of the ligand were extracted from the structure. Any disrupted peptide linkages were capped as amides, and the protonation state of each side chain was checked (i.e. Glu305 = -1, Arg346 = +1). The resulting structure was loaded in Spartan'08, and the gas phase single point energy calculation conducted at the MMFF and semi-empirical level of theory based on the current geometry. In the latter case, the calculation was performed using the AM1 and PM3 methods.

IX. References.

(1)   Kuiper, G. G. J. M.; Enmark, E.; Pelto-Huikko, M.; Nilsson, S.; Gustafsson, J.-A. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 5925.
(2)   Mosselman, S.; Polman, J.; Dijkema, R. *FEBS Lett.* **1996**, *392*, 49.
(3)   Nilsson, S.; Gustafsson, J.-Å. *Breast Cancer Res.* **2000**, *2*, 360.
(4)   Minutolo, F.; Macchia, M.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A. *Med. Res. Rev.* **2009**, n/a.
(5)   De Angelis, M.; Stossi, F.; Carlson, K. A.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A. *J. Med. Chem.* **2005**, *48*, 1132.
(6)   Meyers, M. J.; Sun, J.; Carlson, K. E.; Marriner, G. A.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A. *J. Med. Chem.* **2001**, *44*, 4230.
(7)   Harris, H. A.; Albert, L. M.; Leathurby, Y.; Malamas, M. S.; Mewshaw, R. E.; Miller, C. P.; Kharode, Y. P.; Marzolf, J.; Komm, B. S.; Winneker, R. C.; Frail, D. E.; Henderson, R. A.; Zhu, Y.; Keith, J. C., Jr. *Endocrinology* **2003**, *144*, 4241.
(8)   Saijo, K.; Collier, J. G.; Li, A. C.; Katzenellenbogen, J. A.; Glass, C. K. *Cell* **2010**, *submitted*.

(9) Nettles, K. W.; Bruning, J. B.; Gil, G.; Nowak, J.; Sharma, S. K.; Hahm, J. B.; Kulp, K.; Hochberg, R. B.; Zhou, H.; Katzenellenbogen, J. A.; Katzenellenbogen, B. S.; Kim, Y.; Joachimiak, A.; Greene, G. L. *Nat. Chem. Biol.* **2008**, *4*, 241.

(10) Sun, J.; Baudry, J.; Katzenellenbogen, J. A.; Katzenellenbogen, B. S. *Mol. Endocrinol.* **2003**, *17*, 247.

(11) Manas, E. S.; Unwalla, R. J.; Xu, Z. B.; Malamas, M. S.; Miller, C. P.; Harris, H. A.; Hsiao, C.; Akopian, T.; Hum, W.-T.; Malakian, K.; Wolfrom, S.; Bapat, A.; Bhat, R. A.; Stahl, M. L.; Somers, W. S.; Alvarez, J. C. *J. Am. Chem. Soc.* **2004**, *126*, 15106.

(12) Harrison, T. S.; Keating, G. M. *CNS Drugs* **2005**, *19*, 65.

(13) Srivastava, P.; Pandey, V. C.; Misra, A. P.; Gupta, P.; Raj, K.; Bhaduri, A. P. *Biorg. Med. Chem.* **1998**, *6*, 181.

(14) Raj, K.; Vishnoi, S. P.; Shoeb, A.; Gupta, D. N.; Keshri, G.; Kamboj, V. P., **1997**.

(15) Shah, C.; McAtee, L.; Breitenbucher, J. G.; Rudolph, D.; Li, X.; Lovenberg Timothy, W.; Mazur, C.; Wilson Sandy, J.; Carruthers Nicholas, I. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 3309.

(16) Chai, W.; Breitenbucher, J. G.; Kwok, A.; Li, X.; Wong, V.; Carruthers Nicholas, I.; Lovenberg Timothy, W.; Mazur, C.; Wilson Sandy, J.; Axe Frank, U.; Jones Todd, K. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1767.

(17) Kaminski, J. J.; Bristol, J. A.; Puchalski, C.; Lovey, R. G.; Elliott, A. J.; Guzik, H.; Solomon, D. M.; Conn, D. J.; Domalski, M. S.; et al. *J. Med. Chem.* **1985**, *28*, 876.

(18) Goodacre, S. C.; Street, L. J.; Hallett, D. J.; Crawforth, J. M.; Kelly, S.; Owens, A. P.; Blackaby, W. P.; Lewis, R. T.; Stanley, J.; Smith, A. J.; Ferris, P.; Sohal, B.; Cook, S. M.; Pike, A.; Brown, N.; Wafford, K. A.; Marshall, G.; Castro, J. L.; Atack, J. R. *J. Med. Chem.* **2006**, *49*, 35.

(19) Patel, H. S.; Linn, J. A.; Drewry, D. H.; Hillesheim, D. A.; Zuercher, W. J.; Hoekstra, W. J. *Tetrahedron Lett.* **2003**, *44*, 4077.

(20) Tschitschibabin, A. E. *Ber. Dtsch. Chem. Ges.* **1925**, *58*.

(21) Elliott, A. J.; Guzik, H.; Soler, J. R. *J. Heterocycl. Chem.* **1982**, *19*, 1437.

(22) Hajos, G.; Riedl, Z. *Science of Synthesis* **2002**, *12*.

(23) Gore, P. H. *Chem. Rev.* **1955**, *55*, 229.

(24) Meakins, G. D.; Musk, S. R. R.; Robertson, C. A.; Woodhouse, L. S. *J. Chem. Soc., Perkin Trans. 1* **1989**, 643.

(25) King, L. C.; Ostrum, G. K. *J. Org. Chem.* **1964**, *29*, 3459.

(26) Bauer, D. P.; Macomber, R. S. *J. Org. Chem.* **1975**, *40*, 1990.

(27) Doifode, K. B.; Marathey, M. G. *J. Org. Chem.* **1964**, *29*, 2025.

(28) Hama, Y.; Nobuhara, Y.; Aso, Y.; Otsubo, T.; Ogura, F. *Bull. Chem. Soc. Jpn.* **1988**, *61*, 1683.

(29) Wolter, M.; Nordmann, G.; Job, G. E.; Buchwald, S. L. *Org. Lett.* **2002**, *4*, 973.

(30) Breukelman, S. P.; Meakins, G. D.; Tirel, M. D. *J. Chem. Soc., Chem. Commun.* **1982**, 800.

(31) Bruekelman, S. P.; Leach, S. E.; Meakins, G. D.; Tirel, M. D. *J. Chem. Soc., Perkin Trans. 1* **1984**, 2801.

(32) Macor, J. E.; Chenard, B. L.; Post, R. J. *J. Org. Chem.* **1994**, *59*, 7496.

(33) Ragan, J. A.; Makowski, T. W.; Castaldi, M. J.; Hill, P. D. *Synthesis* **1998**, 1599.

(34) Ragan, J. A.; Jones, B. P.; Castaldi, M. J.; Hill, P. D.; Makowski, T. W. *Organic Syntheses* **2002**, *78*, 63.

(35) Harwood, L. M. *Aldrichimica Acta* **1985**, *18*, 25.

(36) Pedersen, D. S.; Rosenbohm, C. *Synthesis* **2001**, *2001*, 2431.

(37) Lombardino, J. G. *J. Med. Chem.* **1981**, *24*, 39.

(38) Lohse, O. *Synth. Commun.* **1996**, *26*, 2017.

(39) Wu, Z.; Fraley, M. E.; Bilodeau, M. T.; Kaufman, M. L.; Tasber, E. S.; Balitza, A. E.; Hartman, G. D.; Coll, K. E.; Rickert, K.; Shipman, J.; Shi, B.; Sepp-Lorenzino, L.; Thomas, K. A. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 909.

(40) Paudler, W. W.; Blewitt, H. L. *J. Org. Chem.* **1965**, *30*, 4081.

(41) Almirante, L.; Polo, L.; Mugnaini, A.; Provinciali, E.; Rugarli, P.; Biancotti, A.; Gamba, A.; Murmann, W. *J. Med. Chem.* **1965**, *8*, 305.

(42) Hand, E. S.; Paudler, W. W. *J. Org. Chem.* **1976**, *41*, 3549.

(43) Paolini, J. P.; Robins, R. K. *J. Org. Chem.* **1965**, *30*, 4085.

(44) Enguehard, C.; Renou, J.-L.; Collot, V.; Hervet, M.; Rault, S.; Gueiffier, A. *J. Org. Chem.* **2000**, *65*, 6572.

(45) Koubachi, J.; El Kazzouli, S.; Berteina-Raboin, S.; Mouaddib, A.; Guillaumet, G. *Synthesis* **2008**, *2008*, 2537.

(46) Tschaen, D. M.; Desmond, R.; King, A. O.; Fortin, M. C.; Pipik, B.; King, S.; Verhoeven, T. R. *Synth. Commun.* **1994**, *24*, 887.

(47) Maligres, P. E.; Waters, M. S.; Fleitz, F.; Askin, D. *Tetrahedron Lett.* **1999**, *40*, 8193.

(48) Jin, F.; Confalone, P. N. *Tetrahedron Lett.* **2000**, *41*, 3271.

(49) Kotha, S.; Lahiri, K.; Kashinath, D. *Tetrahedron* **2002**, *58*, 9633.

(50) Chemler, S. R.; Trauner, D.; Danishefsky, S. J. *Angew. Chem. Int. Ed.* **2001**, *40*, 4544.

(51) Umemoto, T. *Chem. Rev.* **1996**, *96*, 1757.

(52) Ma, J.-A.; Cahard, D. *J. Fluorine Chem.* **2007**, *128*, 975.

(53) Umemoto, T.; Sumi, I. *Tetrahedron Lett.* **1990**, *31*, 3579.

(54) Umemoto, T.; Ishihara, S. *J. Am. Chem. Soc.* **1993**, *115*, 2156.

(55) Magnier, E.; Blazejewski, J.-C.; Tordeux, M.; Wakselman, C. *Angew. Chem. Int. Ed.* **2006**, *45*, 1279.

(56) Jones, G.; Stanforth, S. P. *The Vilsmeier Reaction of Fully Conjugated Carbocycles and Heterocycles*; John Wiley & Sons, Inc., 2004.

(57) Zhou, H.-B.; Comninos, J. S.; Stossi, F.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A. *J. Med. Chem.* **2005**, *48*, 7261.

(58) Yin, J.; Xiang, B.; Huffman, M. A.; Raab, C. E.; Davies, I. W. *J. Org. Chem.* **2007**, *72*, 4554.

(59) Artyomov, V. A.; Shestopalov, A. M.; Litvinov, V. P. *Synthesis* **1996**, 927.

(60) Goda, H.; Ihara, H.; Hirayama, C.; Sato, M. *Tetrahedron Lett.* **1994**, *35*, 1565.

(61) Li, C.; Zhang, H.; Cui, Y.; Zhang, S.; Zhao, Z.; Choi, M. C. K.; Chan, A. S. C. *Synth. Commun.* **2003**, *33*, 543.

(62) Katzenellenbogen, J. A.; Johnson, H. J., Jr.; Myers, H. N. *Biochemistry* **1973**, *12*, 4085.

(63) Carlson, K. E.; Choi, I.; Gee, A.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A. *Biochemistry* **1997**, *36*, 14897.

(64) De Angelis, M.; Stossi, F.; Waibel, M.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A. *Biorg. Med. Chem.* **2005**, *13*, 6529.

(65) Molecular Operating Environment (MOE), Chemical Computing Group, http://www.chemcomp.com/.

(66) Sanner, M. F. *J. Mol. Graphics Model.* **1999**, *17*, 57.

(67) Trott, O.; Olson, A. J. *J Comput Chem* **2010**, *31*, 455.

(68) FRED, version 2.2.5, OpenEye Scientific Software, Inc., Santa Fe, NM, USA, www.eyesopen.com, 2010.

(69) Grant, J. A.; Pickup, B. T.; Nicholls, A. *J. Comput. Chem.* **2001**, *22*, 608.

(70) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. *Biopolymers* **2003**, *68*, 76.

(71) Verkhivker, G. M.; Bouzida, D.; Gehlhaar, D. K.; Rejto, P. A.; Arthurs, S.; Colson, A. B.; Freer, S. T.; Larson, V.; Luty, B. A.; Marrone, T.; Rose, P. W. *J. Comput. Aided Mol. Des.* **2000**, *14*, 731.

(72) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *J. Comput. Aided Mol. Des.* **1997**, *11*, 425.

(73) Stahl, M.; Rarey, M. *J. Med. Chem.* **2001**, *44*, 1035.

(74) Zhou, H.-B.; Lee, J. H.; Mayne, C. G.; Carlson, K. E.; Katzenellenbogen, J. A. *J. Med. Chem.* **2010**, *53*, 3349.

(75) Hehre, W. J. *A Guide to Molecular Mechanics and Quantum Chemical Calculations*; Wavefunction, Inc.: Irvine, CA, 2003.

(76) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

(77) Spartan'08, Wavefunction, Inc., Irvine, CA, www.wavefun.com.

(78) *CRC Handbook of Chemistry and Physics*; 91st ed.; Haynes, W. M., Ed.; CRC Press, 2010.

(79) Deakyne, C. A. *Int. J. Mass spectrom.* **2003**, *227*, 601.

(80) Dittrich, M. *Introduction to QM Simulations*, Beckman Institute for Advance Science and Techanology, University of Illinois at Urbana-Champaign, 2003, http://www.ks.uiuc.edu/~markus/sum_notes_revised/qm_tutorial.html.

(81) Alcaide, B.; Plumet, J.; Sierra, M. A.; Vicent, C. *J. Org. Chem.* **1989**, *54*, 5763.

(82) Hulme, C.; Lee, Y.-S. *Mol. Divers.* **2008**, *12*, 1.

(83) Bacon, R. G. R.; Stewart, O. J. *J. Chem. Soc. (C)* **1969**, 301.

(84) Bacon, R. G. R.; Rennison, S. C. *J. Chem. Soc. (C)* **1969**, 308.

(85) Bacon, R. G. R.; Rennison, S. C. *J. Chem. Soc. (C)* **1969**, 312.

(86) Aalten, H. L.; van Koten, G.; Grove, D. M.; Kuilman, T.; Piekstra, O. G.; Hulshof, L. A.; Sheldon, R. A. *Tetrahedron* **1989**, *45*, 5565.

(87) Moroz, A. A.; Shvartsberg, M. S. *Russ. Chem. Rev.* **1974**, *43*, 679.

(88) Peeters, L. D.; Jacobs, S. G.; Eevers, W.; Geise, H. J. *Tetrahedron* **1994**, *50*, 11533.

(89) Keegstra, M. A.; Peters, T. H. A.; Brandsma, L. *Tetrahedron* **1992**, *48*, 3633.

(90) Evano, G.; Blanchard, N.; Toumi, M. *Chem. Rev.* **2008**, *108*, 3054.

(91) Fagan, P. J.; Hauptman, E.; Shapiro, R.; Casalnuovo, A. *J. Am. Chem. Soc.* **2000**, *122*, 5043.

(92) Mann, G.; Hartwig, J. F. *J. Org. Chem.* **1997**, *62*, 5413.

(93) Altman, R. A.; Shafir, A.; Choi, A.; Lichtor, P. A.; Buchwald, S. L. *J. Org. Chem.* **2007**, *73*, 284.

(94) Lützen, A.; Hapke, M.; Staats, H.; Bunzen, J. *Eur. J. Org. Chem.* **2003**, 3948.

(95) Moraczewski, A. L.; Banaszynski, L. A.; From, A. M.; White, C. E.; Smith, B. D. *J. Org. Chem.* **1998**, *63*, 7258.

(96) Nantermet, P. G.; Barrow, J. C.; Lindsley, S. R.; Young, M.; Mao, S.-S.; Carroll, S.; Bailey, C.; Bosserman, M.; Colussi, D.; McMasters, D. R.; Vacca, J. P.; Selnick, H. G. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2141.

(97) Woiwode, T. F.; Rose, C.; Wandless, T. J. *J. Org. Chem.* **1998**, *63*, 9594.

(98) Bullock, K. M.; Burton, D.; Corona, J.; Diederich, A.; Glover, B.; Harvey, K.; Mitchell, M. B.; Trone, M. D.; Yule, R.; Zhang, Y.; Toczko, J. F. *Org. Process Res. Dev.* **2008**, *13*, 303.

(99) Li, F.; Meng, Q.; Chen, H.; Li, Z.; Wang, Q.; Tao, F. *Synthesis* **2005**, *2005*, 1305.

(100) De Angelis, M., unpublished data,

(101) Denmark, S. E.; Butler, C. R. *Chem. Commun.* **2009**, 20.

(102) Kerins, F.; O'Shea, D. F. *J. Org. Chem.* **2002**, *67*, 4968.

(103) Knapp, D. M.; Gillis, E. P.; Burke, M. D. *J. Am. Chem. Soc.* **2009**, *131*, 6961.

(104) Enguehard-Gueiffier, C.; Croix, C.; Hervet, M.; Kazock, J.-Y.; Gueiffier, A.; Abarbri, M. *Helv. Chim. Acta* **2007**, *90*, 2349.

(105) Kawai, Y.; Satoh, S.; Yamasaki, H.; Kayakiri, N.; Yoshihara, K.; Oku, T., *PCT Int. Appl.*, WO 9634866, **1996**.

(106) Badger, A.; Bender, P.; Esser, K.; Griswold, D.; Nabil, H.; Lee, J.; Votta, B.; Simon, P., *PCT Int. Appl.*, WO 9100092, **1991**.

(107) Marvin was used for drawing, displaying, and characterizing chemical structures and reactions, Marvin 5.3.2, 2010, ChemAxon, http://www.chemaxon.com.

(108) Calculator Plugins were used for structure property prediction and calculation, Marvin 5.3.2, 2010, ChemAxon, http://www.chemaxon.com.

(109) B-Rao, C.; Subramanian, J.; Sharma, S. D. *Drug Discov. Today* **2009**, *14*, 394.

CHAPTER 3

MODIFYING KNOWN ESTROGEN RECEPTOR LIGANDS FOR
NOVEL ANTI-INFLAMMATORY ACTIVITY

I. Introduction

A relatively new area in ER research involves controlling crosstalk between ERs and NF-κB. NF-κB is itself a transcription factor that when inactive is bound by IκB in the cytoplasm, masking a nuclear localization sequence. Upon activation by cytokines and other cellular stress, IκB is degraded, followed by translocation of NF-κB to the nucleus where it stimulates an inflammatory response, including the recruitment of macrophages which are known to promote aggressive tumor growth, metastasis, and resistance towards chemotherapeutics. It has recently been found that ER agonists suppress NF-κB inflammatory pathways, although the specific interactions involved in this regulation are unknown.[1,2] This receptor crosstalk represents an opportunity to develop a new class of breast cancer therapeutics that act as antagonists on classical ER transcriptional pathways but retain agonistic activity associated with the suppression of NF-κB.
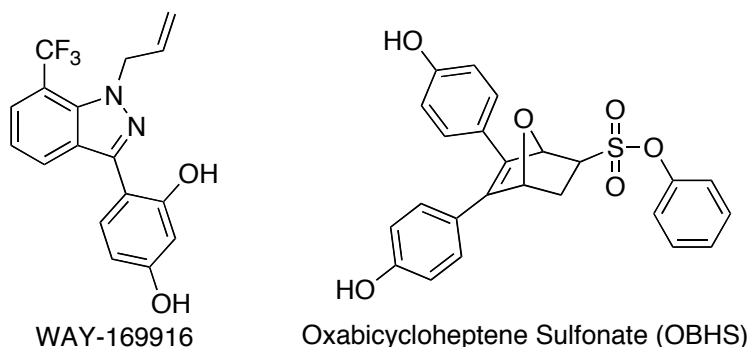
Figure 3.1. Compounds WAY-169916, developed by Wyeth, and OBHS, developed in our labs, show minimal activation of classical ER transcriptional pathways but retain agonist-like activity in ER-modulated suppression of NF-κB.

To date, only a few ligands exhibit this highly desired pharmacological profile: the indazole WAY-169916 first published by Wyeth in 2004,[3,4] and a unique oxabicycloheptene (OBH) scaffold more recently discovered in our labs in collaboration with Kendall Nettles, of the Scripps Research Institute, Florida.[5,6] In crystal structures of ER bound to ligands exhibiting a more standard pharmacology, His524 in the binding pocket serves as a hydrogen bond acceptor that is responsible for increased ligand affinities for both agonists and traditional antagonists. Careful analysis of crystal structures for both WAY-169916 and OBHS, however, indicates that His524 is significantly displaced from the binding pocket.[6-8] Recent MD simulations have identified His524 as the keystone in a hydrogen bonding network responsible for stabilizing the

Figure 3.2. *Panel A.* The extended phenyl sulfonate of OBHS displaces His524 (magenta) from the binding pocket, causing two turns of helix 11 to unwind (red cylinder). Compare to the genestei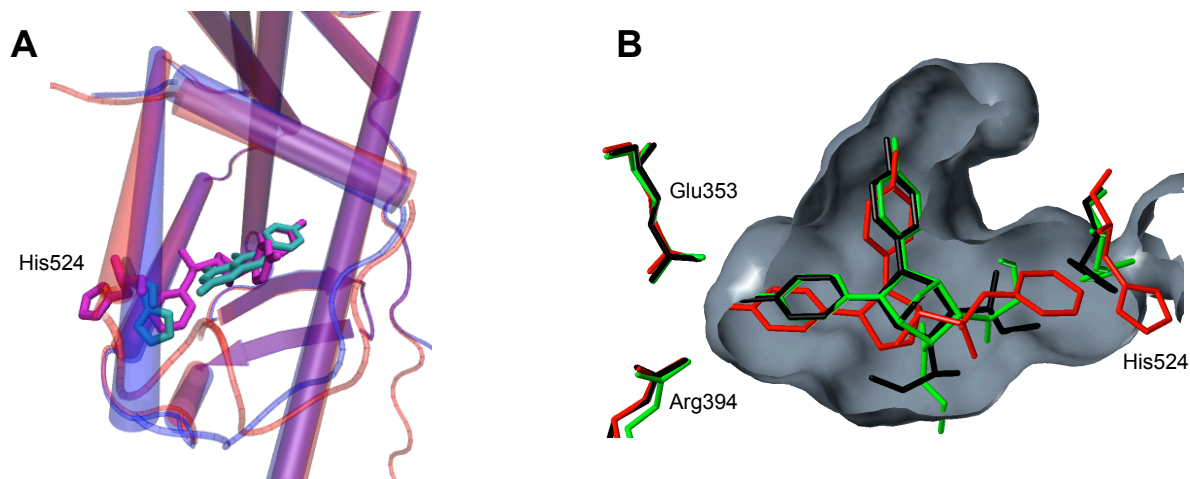n-bound structure (blue). *Panel B.* Changing the size of the extended substituent of the OBH scaffold modulates the position of His524, and correlates to activity. *Panel C.* Ligand effects on the positioning of His524 and overall receptor structure were also characterized by calculating the RMSD of each Cα atom in a series of ligand-bound ERα structures relative to the estradiol bound receptor structure. All active compounds give values >0.4 for His524, while the values for control structures are <0.15.
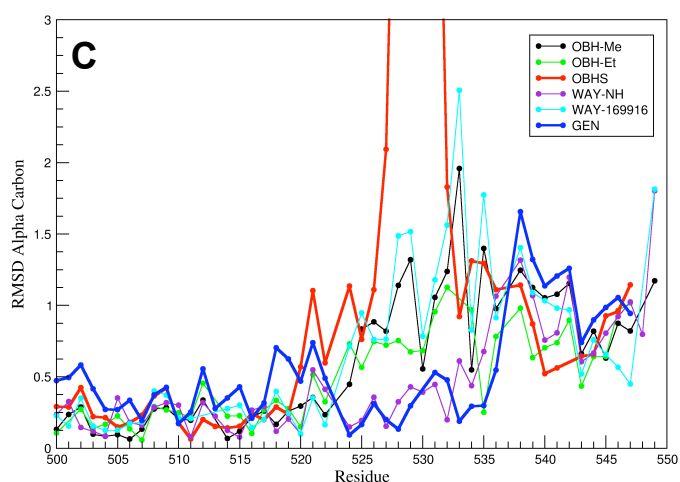
ends of helices 3 and 11,[9] which are essential to positioning helix 12 in the agonist conformation. The RMSD profiles of several ligands when compared to an E2-bound receptor highlight the broader effects of displacing His524 (Fig. 3.2). All compounds that exhibit the desired pharmacological profile have an RMSD >0.5 Å at position 524 while the control compounds are all below 0.25 Å. Our working hypothesis is that the displacement of His524 destabilizes the end of helix 11, specifically residues 525-532, with OBHS (red) being the most striking example (compared to genistein [blue]). Overlaying several crystal structures of ER-bound OBH ligands shows that a large displacement of His524 causes two turns of helix 11 to unravel, giving the large increase in RMSD. Furthermore, this increasing degree of His524 displacement is well correlated to the desired pharmacology within the OBH and WAY ligand series.[6]

Efforts are currently underway in our labs and with collaborators to probe the SAR of the OBH scaffolds; however, synthetic access to these structures is non-trivial. Accordingly, we have also set out to design a series of ligands based on known ER scaffolds with established

synthetic routes to serve as functional analogs of OBHS and generate an expanded SAR. These efforts are founded on the information gleaned from the crystal structures described above, in which we plan to modify the established ER pharmacophores by placing an extended substituent towards the D-ring end of the binding pocket. Herein, we describe the design, synthesis, and preliminary biological analysis of such ligands through iterative *in silico* design supported by synthetic chemistry and biological testing.

II. Structure-Based Design

Several classes of ligands were designed to bind ERs in a similar mode to that observed for OBHS. These ligands were based on the core scaffolds of known ER ligands, and are shown in Figure 3.3.[10-14] To determine the most appropriate substitution patterns, multiple ligands of each structure were utilized, modifying substitution position and linking atom between the scaffold core and extended substituent. All proposed ligands were constructed and minimized using the Molecular Operating Environment (MOE),[15] prepared using AutoDock Tools (ADT),[16] and docked into ERα binding volume established by OBHS using AutoDock Vina (ADV).[17] The lowest energy poses were selected using ADV's internal scoring algorithm, merged back into the receptor structure, and minimized in Sybyl.[18] Each ligand-receptor pair was visually analyzed with particular attention to steric constraints, matched hydrogen bonding pairs, and lipophilicity profile. The most promising substitution patterns, shown in Figure 3.4, were further pursued via synthetic methods. Ligands based on triarylethylenes and benzothiophenes were pursued by Dr. Terry More, Dr. Markéta Lebl-Rinnová, and Davis Oldham in our labs, while the modification of steroidal scaffolds is described herein.

III. Chemistry
A. Synthesis of 16β-benzyl-17β-estradiol

Following conditions reported in the literature,[19] estrone (E1) was condensed with benzaldehyde to give aldol product **1** (Scheme 3.1). The ketone at the 17-position was reduced with sodium borohydride in THF to yield the 17β-hydroxyl (**2**), followed by palladium-catalyzed hydrogenation of the exocyclic double bond under elevated pressure (30 psi) of hydrogen gas in a Parr-shaker to furnish 16β-benzyl-17β-estradiol (**3**). Both reductions proceeded stereoselectively under substrate control without any detection of alternate stereoisomers.

B. Reductive Amination of Estrone

The phenol of E1 was protected as the methyl ether from sodium hydride, methyl iodide, and DMF to yield **4** (Scheme 3.2). This intermediate was reacted with aniline in formic acid at 160 °C overnight via the Leukart-Wallach reaction to give the reductive amination products **5**

Ligand Precedent

Designed Ligands

ERα: 166   ERβ: 269

R = O, OH, NHPh

R' = H/H, Ph, Ph

ERα: 56.2   ERβ: 4.68

X = O, N
R = H, F, Me

ERα: 698   ERβ: 898

X = O, N
R = H, F, Me
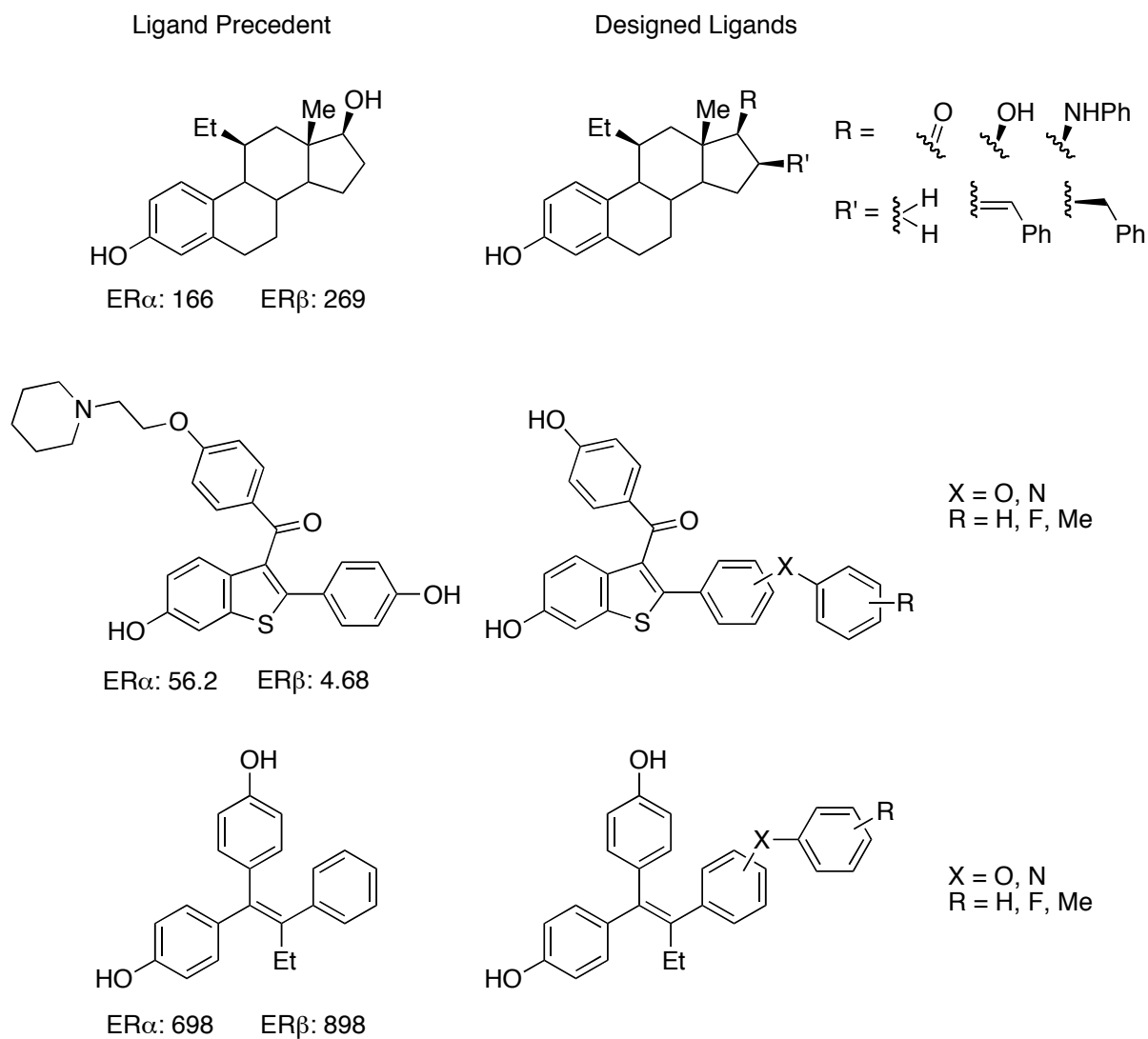
Figure 3.3. The scaffolds of three known ER ligands were redesigned to bear substituents that extend towards the D-ring end of the binding pocket in order to displace His524. The steroidal scaffold is based on a known modification of estradiol that increases affinity and selectivity for ERβ. The latter two structures designed herein, but synthesized by others, are based on raloxifene and tamoxifen scaffolds used in hormone replacement therapy and the treatment of breast cancers, respectively.
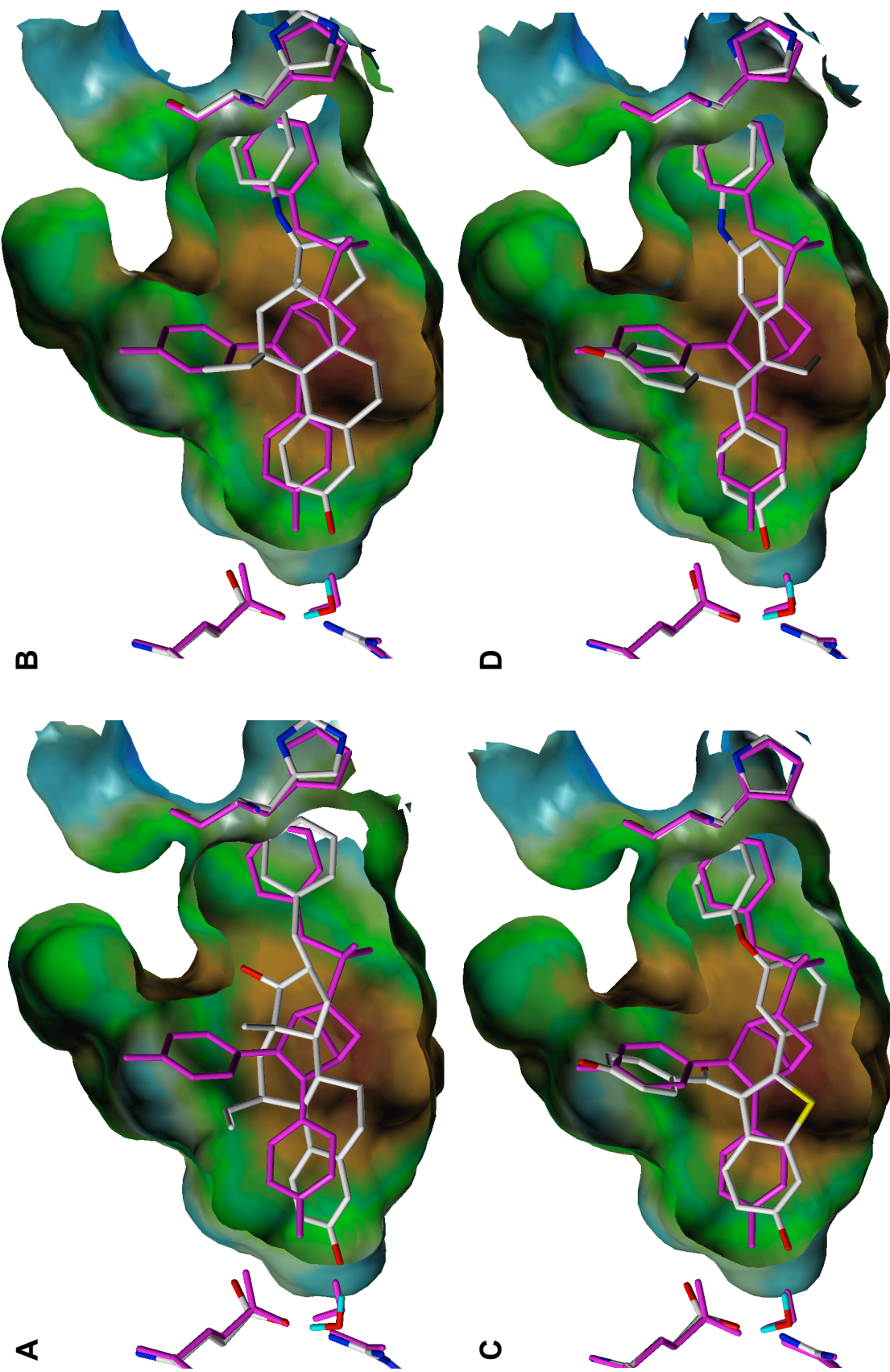
Figure 3.4. Representative docked poses for the estradiol (**A**, **B**), benzothiophene (**C**), and triarylethylene (**D**) scaffolds are shown. The docked ligand is colored by atom type and the pocket is colored by lipophilicity (brown=lipophilic, blue=hydrophilic). Each pose is overlaid with the ERα-OBHS crystal (magenta) structure to facilitate structural comparisons.

and **6**,[20] which were separable by standard chromatographic methods. The *N*-formyl group of **5** was subsequently reduced with sodium borohydride in THF to yield the corresponding *N*-methyl product (**8**). **6** and **8** were deprotected by methyl ether cleavage under Lewis acidic conditions using boron tribromide in DCM to yield **7** and **9**.

Scheme 3.1. Synthesis of 16β-benzyl-17β-Estradiol

a) benzaldehyde, NaOH, EtOH. b) NaBH₄, EtOH/THF. c) Pd/C, H₂ (30 psi), EtOH/THF
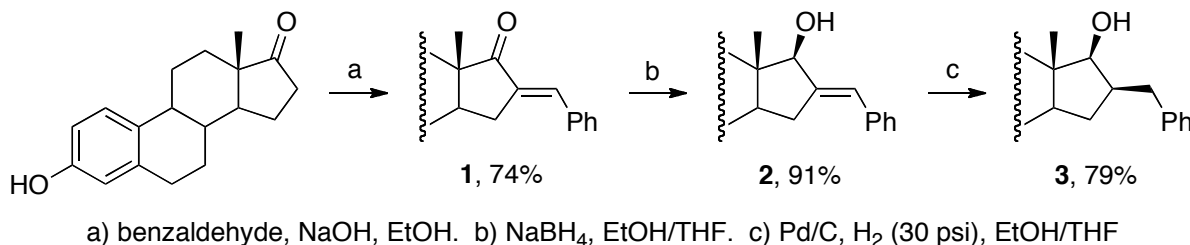
Scheme 3.2. Reductive Amination of Estrone

a) MeI, NaH, DMF. b) aniline, formic acid, reflux. c) BBr₃, DCM. d) NaBH₄, THF.
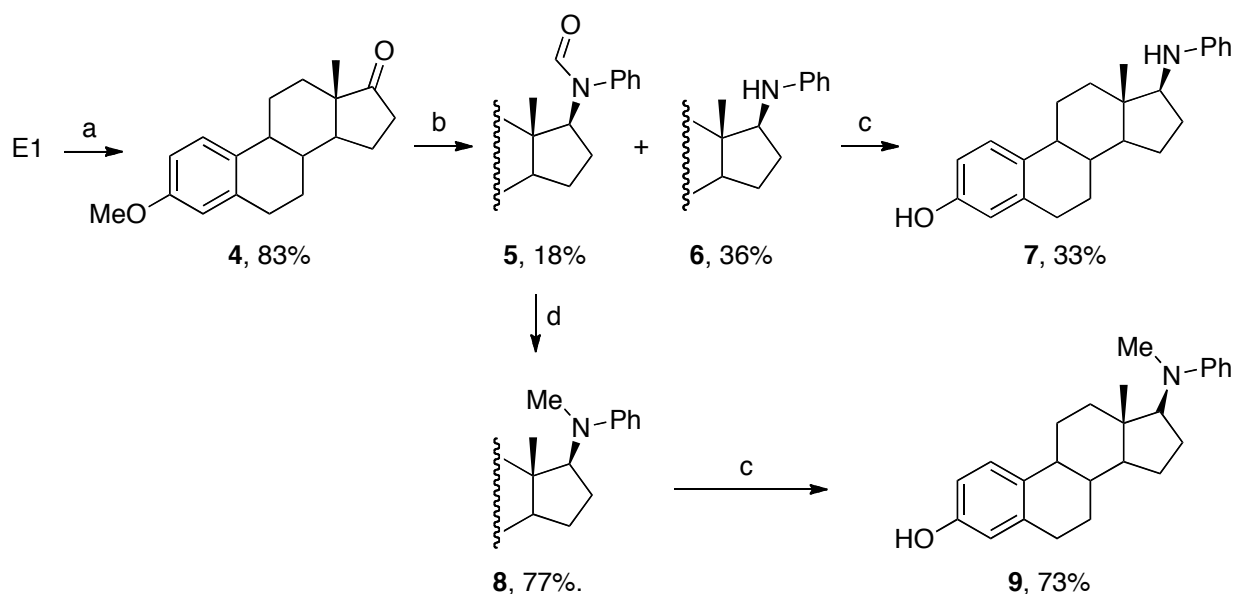
IV. Biological Assay

The relative binding affinities (RBA) of all deprotected compounds were determined by radiometric competitive binding assay, using 17β-[³H]estradiol as the tracer.[21,22] This assay was performed by Kathryn Carlson, and the resulting data for both ERα and ERβ are given in Tables 3.1 and 3.2 as a percentage relative to estradiol (E2=100%). These RBA values can be easily converted to IC₅₀ or Kᵢ values, as described elsewhere.[23,24] All compounds with the exception of the benzothiophenes were carried into cell-based assays by our collaborator Kendall Nettles. In these preliminary single point assays, ERE activity was established using a luciferase reporter gene construct in HepG2 cells, and anti-inflammatory activity was determined by measuring IL-6 protein levels (by ELISA) in MCF7 cells treated with TNFα (Table 3.3).

Table 3.1. Relative Binding Affinity of Steriodal Scaffolds

compound	ERα	ERβ	β/α
1	1.98	0.198	0.10
2	0.368	2.12	5.76
3	1.40	8.22	5.87
7	1.09	0.670	0.615
9	0.394	0.695	1.76

Table 3.2. Relative Binding Affinity of Benzothiophene and Triarylethylene Scaffolds

Scaffold I   **Scaffold II**   **Scaffold III**
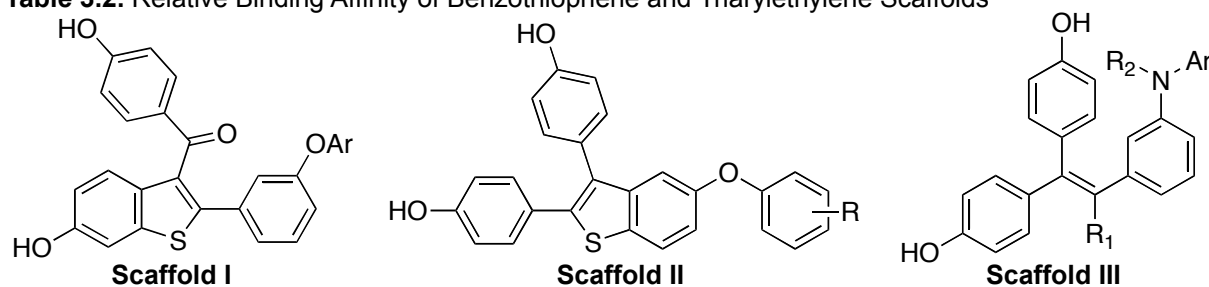
compound	Scaffold		Ar		ERα	ERβ	β/α
10a[a]	I		Ph		0.004	0.005	1
10b[a]	I		4-PhF		0.004	0.005	1
10c[a]	I		3-PhMe		0.007	0.009	1
11a[b]	II		Ph		3.68	1.04	0.283
11b[b]	II		3-PhF		4.17	2.75	0.659
11c[b]	II		4-PhF		1.51	0.387	0.256
12a[a]	III	R_1 = Et	R_2 = H	Ar = H	50.6	21.1	0.417
12b[a]	III	Et	H	4-PhF	110	18.4	0.17
12c[a]	III	Et	H	3-PhMe	89.4	29.1	0.326
12d[a]	III	Et	H	1-naphthyl	20.3	18.7	0.921
12e[c]	III	Me	H	Ph	73.7	15.7	0.213
12f[c]	III	Me	Me	Ph	46.3	29.6	0.639
12g[c]	III	Me	H	4-PhF	96.3	14.5	0.151
12h[c]	III	Me	Me	4-PhF	34.8	28.3	0.813
12h[c]	III	CN	H	4-PhF	68.7	19.4	0.282
12i[c]	III	H	H	4-PhF	123	33.4	0.272

[a] Synthesized by Dr. Terry Moore, [b] Synthesized by Davis Oldham, [c] Synthesized by Dr. Markéta Lebl-Rinnová

Table 3.3. Transcriptional Readout of IL-6 and ERE Activities

Compound	IL-6 ELISA	+/-	ERE-luc	+/-
1	21895	4485	24376	508
2	20626	5496	45044	5876
3	19736	2491	42715	1795
12a	42552	19799	20300	5649
12b	66642	7092	12117	2147
12c	70626	9173	6145	2416
12d	43742	8283	20277	2168
12e	70773	36359	20275	2314
12f	85977	10070	34474	8296
12g	101430	4827	20424	2994
12h	79013	11226	27664	2597
12h	149951	42290	22531	1267
12i	137040	19737	56709	2825
7	21152	4279	26476	4793
9	18092	3750	23178	4773
E2	21320	2310	114373	8527
genestein	37511	3255	159843	12128
ICI	65129	9422	5040	112
OBHS	26631	6142	44620	10049
triarylethylene cntrl 1	19033	2435	28747	3025
triarylethylene cntrl 2	44754	5139	39366	3395
pyrazolopyrimidine 1	8715	1752	24344	2183
pyrazolopyrimidine 2	26702	4484	17834	1962

V. Discussion

A. Structure-Based Design

The crystal structures of the WAY and oxabicycloheptene (OBH) ligands complexed with ERα provide a compelling structural basis for the observed novel biological activity of these ligands.[6] The key feature of each structure is the displacement of His524 from the standard position observed for conventional ER ligands such as E2 and genistein (Fig. 3.2). With the exception of the steroidal analogs, each of the designed ligand scaffolds is a planar aromatic structure that generally adheres to the established pharmacophore for high affinity ER ligands (see Chapter 1). These structures bear a pendant phenol that mimics the A-ring of estradiol, which participates in a hydrogen bonding network between Glu353, Arg394, and a water molecule that is crucial for high affinity binding to ERs. The core scaffold also bears an internal substituent that projects into the B-ring pocket, for which large aromatic substituents engender ERα-selectivity, while smaller substituents (e.g., chloro, 11β-ethyl) yield ERβ-selectivity.[11,25] The focal point of our designed ligands is the extension of the structure towards the D-ring end of the pocket. In contrast to conventional ligands that form a hydrogen bond with His524, our designed ligands contain a substituent that is specifically introduced to clash with His524, causing the residue to rotate out of position in a manner similar to that observed in the OBHS-bound crystal structure.[8]

Visual analysis of the docked poses for each designed chemotype strongly suggests positioning the D-ring substituent in the *meta*-position for benzothiophene (Fig. 3.4C) and triarylethylene scaffolds (Fig. 3.4D). The wider angle of diaryl ether (Ar-O-Ar) or diaryl amine (Ar-NH-Ar) heteroatom linkages is favored over the more acute angle observed for benzyl (Ar-C-Ar) linkages. The preferred substituent positioning and linkage type for steroidal-based ligands was less clear from the docking poses, although substitution through the 16β- and 17β-positions appeared reasonable (Fig. 3.4B).

Following known substitution patterns for established ER ligands, the internal substitution for non-steroidal ligands was comprised of an aromatic ring attached either via a direct linkage to the scaffold, or through a single carbon linker such as a methylene or carbonyl group. For 5,6-fused bicyclic scaffolds such as the benzothiophene ligands, attachment through such a linker appeared to be favored based on steric constraints in which the rotational flexibility allowed the aryl ring to follow the natural progression of the extended pocket. This was not the case for the triarylethylene scaffold, which was well positioned to access the channel through a direct connection. The positioning of internal aromatic substituents on steroidal scaffolds is less well established; however, it is known that incorporating an 11β-chloromethyl or 11β-ethyl substituent increases affinity and engenders ERβ-selectivity.[10,11] Further substituting the ligand
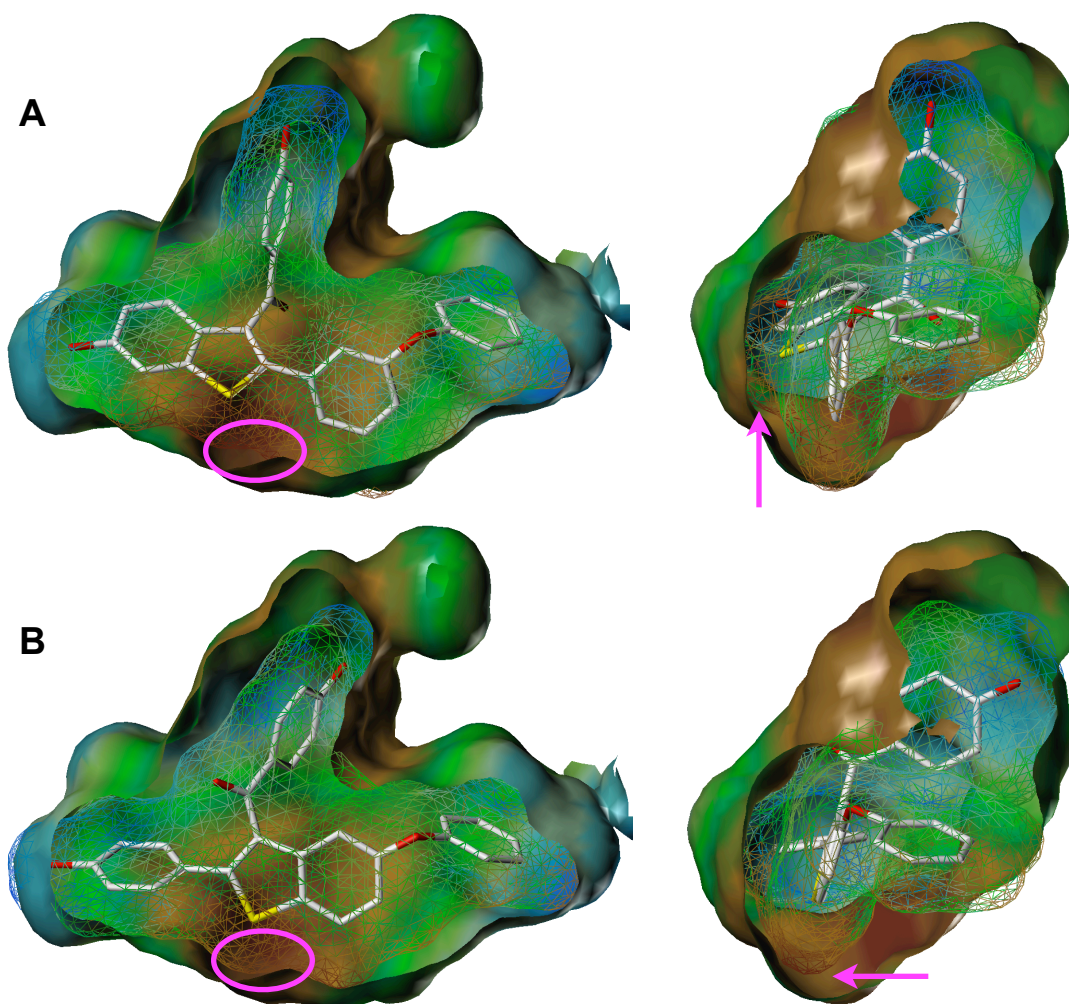
Figure 3.5. Comparison of the docked poses for the benzothiophene (**A**) and reversed-benzothiophene (**B**) scaffolds show a subtle difference in the positioning of the sulfur atom relative to a shallow sub-pocket of increased lipophilicity on the receptor surface (magenta circle). The ligand surface is represented by wire framework, while the receptor surface is opaque; both are mapped by lipophilicity. The fused framework of the benzothiophene projects the sulfur atom towards the front of the pocket, as highlighted when looking down the pocket from His524 towards Glu353. The core of the reversed-benzothiophene, in contrast, rotates relative to the pendant phenol to position the sulfur atom within the lipophilic sub-pocket.

at either 16β- or 17β-positions did not appear to negatively affect the positioning of this 11β-ethyl substituent.

Lipophilic mapping of the receptor surface reiterates that the core of the ligand binding pocket is strongly lipophilic. These properties are best matched by the benzothiophene scaffold, in which the internal sulfur atom engenders substantial lipophilicity to the surrounding hydrocarbon scaffold. Our initial designs for this chemotype strictly followed the binding mode observed in the crystal structure of raloxifene bound to ER,[26,27] in which the pendant phenol is directed towards the D-ring end of the ligand binding pocket (Fig. 3.5A). The docked poses, however, showed a slight offset between a subtle pocket of strong lipophilicity in the receptor

surface, and the positioning of the sulfur atom. Redesigning the ligand to bind in a "reversed" mode by attaching the phenoxy/*N*-phenylamine substituent at the 5-position of the core structure effectively repositioned the sulfur atom directly in line with the observed lipophilic pocket in the receptor surface, while retaining hydrogen bonding contacts and access to the B-ring pocket (Fig. 3.5B). The A-ring and D-ring ends of the pocket, in contrast, appear significantly more hydrophilic due to the hydrogen bonding contacts and access to bulk solvent, respectively (Figs. 3.4, 3.5). As described above, all designed ligands contained a phenolic moiety to participate in the known hydrogen bonding networks; however, due to the inability to effectively incorporate receptor flexibility into ligand docking routines, we were unable to sufficiently probe the affects of various substitution patterns on the extended D-ring substituent (e.g., H, F, Me, OH), or the ability to extend into bulk solvent in this direction.

B. Chemical Synthesis

The synthesis of 16β-benzyl-17β-estradiol closely followed the published route to similar compounds.[19] The aldol condensation was performed in straight forward manner, followed by two subsequent reductions. Both the sodium borohydride reduction of the 17-ketone and the hydrogenation of the exocyclic olefin each proceed in a stereoselective manner, whereby reduction takes place from the α-face of the steroid skeleton due to the steric bulk of the axial 18β-methyl substituent. All steps in this sequence were performed without the use of protecting groups, and each intermediate was easily purified due to the highly crystalline nature of the steroidal scaffold.

Reductive amination of the 17-ketone was performed using the Leukart-Wallach reaction,[20] in which formic acid serves as the hydride donor to capture the *in situ* formed imine, again, from the α-face. The crude product was a mixture of the *N*-formyl and fully hydrolyzed amines, and protection of the A-ring phenol as the corresponding methyl ether greatly aided in the separation of the product mixture. The *N*-formyl component was further reduced with sodium borohydride to yield the *N*-methyl amine, and each compound was subsequently deprotected under Lewis acidic conditions.

C. Biological Evaluation

All compounds were initially tested in *in vitro* radiometric competitive binding assays to gauge their ability to bind the receptor, followed by *in vivo* experiments carried out by our collaborator Kendall Nettles to probe their affect on ER and inflammatory pathways in tumor cells. Our initial synthetic efforts focused on the benzothiophene scaffold based on the raloxifene structure. Initial RBA data for this class, however, yielded RBA values between
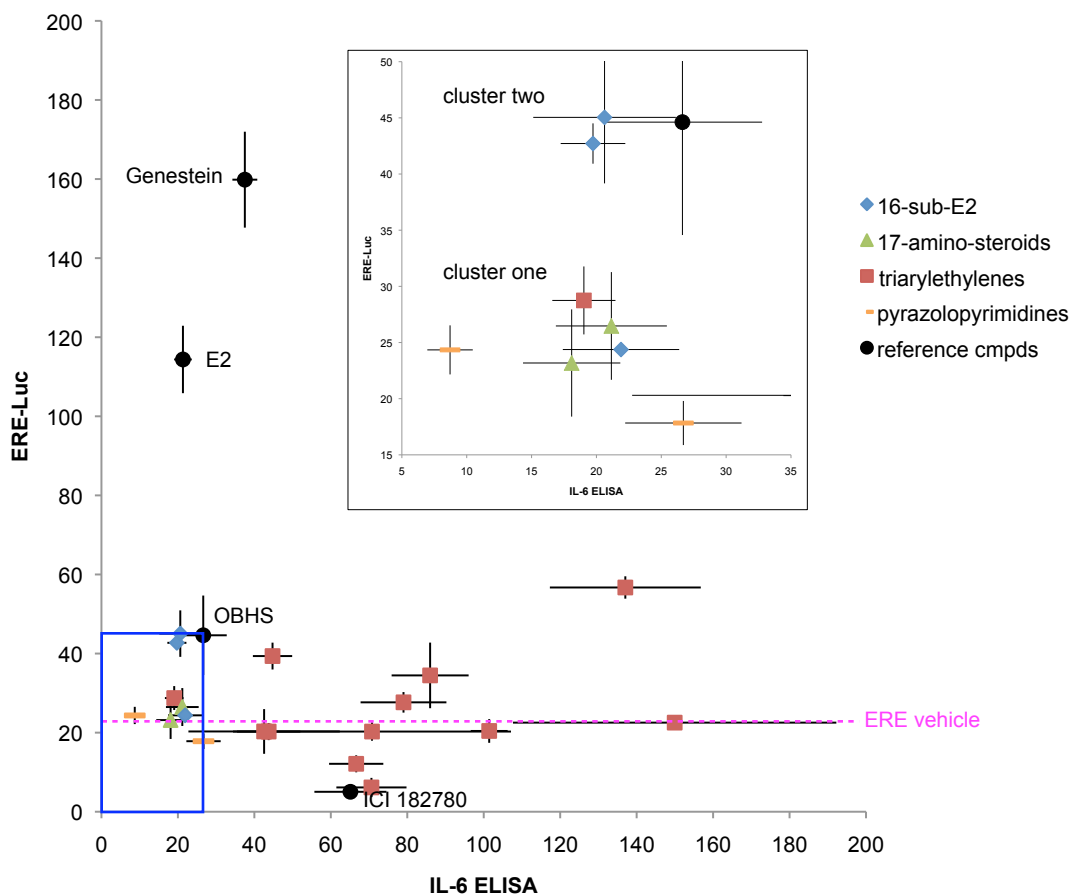
Figure 3.6. The *in vivo* activity of each compound in ERE and IL-6 transcriptional assays are plotted in a two-dimensional analysis. All 17-amino (green triangles) and 16-substituted (blue diamonds) estradiol analogs were found within the desired activity window (blue square), along with one triarylethylene (red squares). These compounds cluster into two groups (inset). Several well-characterized ER ligands (black circles) and two active pyrazolopyrimidines (orange rectangles) are also included for reference.
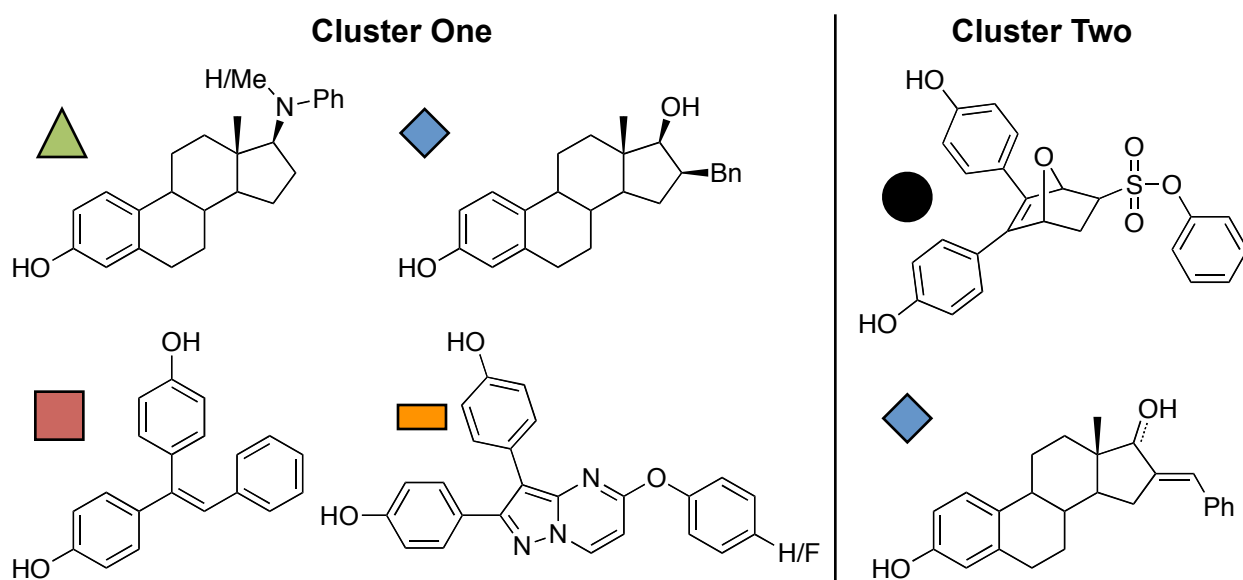
Figure 3.7. Structures of compounds located within the desired activity window, organized by cluster.

0.004-0.007 and 0.005-0.009 for ERα and ERβ, respectively. From this data it is evident that the extended D-ring substituent greatly reduced binding affinity relative to the parent scaffold. Accordingly, the second generation ligands were specifically designed from scaffolds that have previously demonstrated high affinities for ERs. This class included steroidal structures based on estradiol, triarylethylenes based on tamoxifen, and the reversed-orientation benzothiophenes carefully redesigned to optimize the position of the sulfur atom within the binding pocket based on lipophilic mapping. These compounds too demonstrated reduced binding affinities relative to analogous ligands lacking the extended D-ring substituent; however, measured RBA reached biologically interesting levels. Reversing the binding orientation of the benzothiophene scaffold and exchanging the benzoyl sidechain with an aryl substituent increased the RBA by 920-fold for ERα and 208-fold for ERβ. 16β-Substitution of estradiol was favored over replacing the 17β-hydroxyl with an aniline; however, both compound sets yielded RBA values ranging from 0.198 to 8.22 (K_i = 6.1-75 nM) for ERβ. Most notably, all triarylethylene compounds bound with RBA values >34.8 (K_i = 0.57 nM) and >14.5 (K_i = 3.4 nM) for ERα and ERβ, respectively.

With the exception of the benzothiophenes, each of these compounds has been tested for *in vivo* activity. These data, shown in Figure 3.6, are represented in a two-dimensional analysis comparing IL-6 activity on the X-axis, and activation of ERE transcriptional pathways on the Y-axis. The desired phenotype is located in the lower left corner of the plot, in which low IL-6 activity demonstrates suppression of inflammatory pathways, presumably through NF-κB, without activating conventional ER pathways (e.g., cell growth). The data points are colored by scaffold type, and several well-characterized ER ligands were included as reference structures. Using the OBHS datapoint as a guide, two clusters of compounds appeared within the desired activity window. Cluster one consists of five compounds (Fig. 3.7) that demonstrate little ERE activity above vehicle, and IL-6 activity below that observed for OBHS. Included in this cluster are two pyrazolopyrimidines designed and synthesized by Davis Oldham that represent 5,6-fused bicycles similar to the indazoles presented herein. Cluster two, which includes OBHS, demonstrates slightly higher ERE activity with similar IL-6 activity. While all five of the synthesized steroidal ligands were located within the desired activity window, only a single triarylethylene was located within the same range. Curiously, this compound was synthesized as a control lacking an extended D-ring substituent all together (Fig. 3.7). Each of the compounds found within the desired range will be further tested *in vivo* to confirm efficacy and establish potency.

VI. Conclusions

Recent data from structural biology provide compelling evidence that the displacement of His524 from the ligand binding pocket of ERs is well correlated with the suppression of inflammatory pathways without activating classical ER pathways (e.g., cell growth). Based on these data, we set out to design, synthesize, and test a series of steroidal and non-steroidal ligands to achieve this highly desired pharmacology. Using an iterative process, we successfully leveraged the available structural models to modify four scaffolds from known high affinity ER ligands *in silico* to determine the appropriate substitution patterns that mimic the desired structural effects. A series of compounds were subsequently synthesized for each scaffold type, and assayed for binding affinity *in vitro*, and activity on ERE and IL-6 transcriptional pathways. In this manner, our ligand design has been refined to yield several lead structures within the desired activity window suitable for follow-up efforts expanding upon the observed structure activity relationships.

VII. Experimental

A. Chemistry

General Considerations. All chemical reagents were purchased from commercial suppliers without further purification. Anhydrous DMF was obtained from Aldrich in a Sureseal™ bottle stored in a secondary container with desiccant; all other anhydrous solvents were obtained from a solvent dispensing system unless otherwise stated. All glassware was oven or flame-dried and cooled under vacuum or in a dry box. All reactions were conducted under argon. NMR spectra were obtained on Varian Oxford instruments and worked up using ACD, Inc. 1D-NMR processing software. The chemical shifts are reported in ppm and referenced to the solvent peak. Electron Impact (EI) mass spectra were obtained on a 70-VSE mass spectrometer with an ionization energy of 70 eV. Electrospray Ionization (ESI) mass spectra were obtained on a Q-Tof mass spectrometer. Melting point ranges were measured using a Thomas Hoover capillary melting point apparatus.

16-Phenylmethylene-estrone (1). Estrone (411.9 mg, 1.5 mmol), NaOH (305 mg, 7.5 mmol), and benzaldehyde (152 μL, 1.5 mmol) were dissolved in EtOH (15 mL) and stirred at ambient temperature for three hours. The reaction was quench by slow addition of glacial acetic acid until the reaction mixture became acidic, along with the precipitation of a white solid. The mixture was the filtered and washed with cold EtOH to yield **12** (402.2 mg, 74%) as a white solid. A portion of this material was further purified by recrystallization (EtOH/water) prior to biological assay. ^{1}H-NMR (DMSO-d$_6$, 500 MHz) δ 9.06 (br. s., 1 H), 7.65 (m, 2 H), 7.44 (m, 3 H), 7.33 (s, 1 H), 7.05 (d, J=8.55 Hz, 1 H), 6.53 (dd, J=8.30, 2.69 Hz, 1 H), 6.47 (d, J=2.44 Hz, 1

H), 2.88 (ddd, J=16.24, 6.47, 1.46 Hz, 1 H), 2.78 (m, 2 H), 2.58 (ddd, J=15.93, 12.76, 2.81 Hz, 1 H), 2.34 (m, 1 H), 2.19 (m, 1 H), 1.99 (m, 1 H), 1.87 (m, 1 H), 1.60 (m, 1 H), 1.41 (m, 4 H), 1.06 (t, J=6.96 Hz, 1 H), 0.90 (s, 3 H); ^{13}C-NMR (DMSO-d_6, 125 MHz) δ 208.45, 155.05, 137.06, 136.36, 135.10, 131.91, 130.30, 129.90, 129.36, 128.80, 125.99, 114.98, 112.82, 47.74, 47.20, 43.41, 37.55, 31.36, 29.02, 28.62, 26.28, 25.58, 14.26; HRMS (ESI): calc'd for $C_{25}H_{27}O_2$ [M+H]$^+$ 359.2011, found 359.2018.

16-Phenylmethylene-17β-estradiol (2). **1** (348.7 mg, 1 mmol) was added to a round bottomed flask and resuspended in a mixture EtOH (20 mL) and THF (20 mL). The reaction flask was then cooled to 0 °C in an ice bath, and NaBH$_4$ (100 mg, 2.6 mmol) was slowly added. The mixture was slowly warmed to ambient temperature as the reaction was allowed to proceed overnight. The THF was removed by rotary evaporation and the remaining aqueous mixture was transferred to a separatory funnel and exhaustively extracted with EtOAc. The organic layer was dried over MgSO$_4$, vacuum filtered, and concentrated by rotary evaporation to yield **2** (318.3 mg, 91%) as a white solid. A portion of this material was further purified by recrystallization (EtOH/water) prior to biological assay. ^{1}H-NMR (DMSO-d_6, 500 MHz) δ 9.00 (s, 1 H), 7.36 (m, 4 H), 7.18 (m, 1 H), 7.06 (d, J=8.55 Hz, 1 H), 6.52 (dd, J=8.42, 2.56 Hz, 1 H), 6.45 (d, J=2.44 Hz, 1 H), 6.41 (d, J=2.44 Hz, 1 H), 5.15 (d, J=6.10 Hz, 1 H), 3.98 (m, 1 H), 2.71 (m, 3 H), 2.29 (m, 1 H), 2.17 (m, 2 H), 1.91 (m, 2 H), 1.34 (m, 5 H), 0.60 (s, 3 H); ^{13}C-NMR (DMSO-d_6, 125 MHz) δ 154.94, 146.65, 137.88, 137.14, 130.37, 128.40, 127.87, 125.96, 121.71, 114.94, 112.72, 83.46, 46.97, 43.55, 42.73, 38.04, 36.07, 30.28, 29.11, 27.05, 26.06, 14.10, 11.39; HRMS (ESI): calc'd for $C_{25}H_{29}O_2$ [M+H]$^+$ 361.2168, found 361.2163.

16β-Benzyl-17β-estradiol (3). **2** (232.6 mg, 0.6 mmol) was placed in a hydrogenation flask and resuspended in a mixture of EtOH (25 mL) and THF (25 mL). A catalytic amount of Pd/C was added and the flask was pressurized to 30 psi with hydrogen gas and shaken overnight in a Parr-shaker hydrogenation apparatus. The reaction mixture was subsequently filtered through celite and washed with THF. The crude product was obtained by rotary evaporation and further purified by flash chromatography over silica gel (33% EtOAc in hexanes) to yield **3** (182.9 mg, 78.2%) as a white solid. A portion of this material was further purified by recrystallization (EtOH/water) prior to biological assay. ^{1}H-NMR (DMSO-d_6, 500 MHz) δ 8.97 (s, 1 H), 7.24 (m, 2 H), 7.18 (m, 2 H), 7.13 (m, 1 H), 7.03 (d, J=8.55 Hz, 1 H), 6.50 (dd, J=8.42, 2.56 Hz, 1 H), 6.41 (d, J=2.44 Hz, 1 H), 4.67 (d, J=4.88 Hz, 1 H), 3.67 (dd, J=9.52, 4.88 Hz, 1 H), 3.03 (dd, J=13.18, 3.66 Hz, 1 H), 2.66 (m, 2 H), 2.31 (m, 3 H), 2.06 (td, J=10.99, 3.66 Hz, 1 H), 1.88 (m, 1 H), 1.66 (m, 1 H), 1.52 (m, 1 H), 1.24 (m, 4 H), 1.00 (m, 2 H), 0.76 (s,

3 H); ^{13}C-NMR (DMSO-d$_6$, 125 MHz) δ 154.88, 142.57, 137.11, 130.44, 128.61, 128.10, 125.99, 125.31, 114.89, 112.68, 80.37, 48.09, 43.87, 43.57, 41.68, 38.11, 37.62, 37.46, 31.74, 29.15, 27.10, 26.03, 12.75; HRMS (ESI): calc'd for C$_{25}$H$_{31}$O$_2$ [M+H]$^+$ 363.2324, found 363.2328.

3-Methoxyestra-1,3,5(10)-trien-17-one (4). Estrone (2.7 g, 10 mmol) was placed into a round bottomed flask and dissolved in DMF (20 mL). The reaction vessel was cooled to 0 °C in an ice bath, and mineral oil dispersion of NaH (60% by wt., 562 mg, 14 mmol, 60%) and MeI (685 μL, 11 mmol) were added sequentially. The reaction was allowed to proceed for two hours an slowly warmed to ambient temperature, upon which the reaction was quenched with MeOH. The resulting mixture was poured into distilled water yielding a precipitated what was filtered, washed with additional water, and dried under reduced pressure. The crude material was purified by recrystallization (DCM/MeOH) to yield **4** (2.4 g, 83.1%) as dense white flakes. ^{1}H-NMR (CDCl$_3$, 500 MHz,) δ 7.22 (d, J=8.58 Hz, 1 H), 6.74 (dd, J=8.58, 2.79 Hz, 1 H), 6.66 (d, J=2.57 Hz, 1 H), 3.79 (s, 4 H), 2.92 (m, 2 H), 2.52 (dd, J=19.08, 8.79 Hz, 1 H), 2.41 (m, 1 H), 2.27 (td, J=10.45, 4.18 Hz, 1 H), 2.16 (m, 1 H), 2.04 (m, 3 H), 1.96 (m, 1 H), 1.54 (m, 6 H), 0.92 (s, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 220.92, 157.51, 137.69, 131.95, 126.29, 113.80, 111.50, 55.15, 50.33 , 47.96, 43.91, 38.31, 35.83, 31.52, 29.63, 26.50, 25.88, 21.54, 13.81; HRMS (ESI): calc'd for C$_{19}$H$_{25}$O$_2$ [M+H]$^+$ 285.1855, found 285.1854.

***N*-Formyl-3-methoxy-*N*-phenylestra-1,3,5(10)-trien-17-amine (5) and 3-methoxy-*N*-phenylestra-1,3,5(10)-trien-17β-amine (6).** **4** (285.5 mg, 1 mmol) was placed into a pear-shaped round bottomed flask to which aniline (0.5 mL, 5 mmol) and formic acid (0.2 mL, 5 mmol) were added. The reaction mixture was heated to 160 °C under an argon atmosphere for 24 hours. After cooling to ambient temperature, the mixture was transferred to a separatory funnel, to which a brine solution was added, and exhaustively extracted with EtOAc. The organic layer was dried over MgSO$_4$, vacuum filtered, and concentrated by rotary evaporation. Two different product spots were observed by TLC analysis and isolated by flash chromatography over silica gel with a gradient elution (25-50% EtOAc in hexanes). The resulting material was recrystallized (MeOH/water) to yield **5** (66.5 mg, 18%) and **6** (130.0 mg, 35.8%) as white solids. **5**. major isomer: ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.30 (m, 1 H), 7.36 (m, 3 H), 7.22 (m, 3H), 6.71 (m, 1 H), 6.63 (m, 1 H), 4.51 (t, J=9.9 Hz, 1 H), 3.78 (s, 3 H), 2.86 (m, 2 H), 2.06 (m, 7 H), 1.68 (m, 1 H), 1.40 (m, 4 H), 1.22 (m, 1 H), 0.68 (s, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 164.56, 157.36, 140.57, 137.81, 132.59, 130.07, 129.03, 128.53, 126.28, 113.71, 111.37, 64.76, 55.16, 51.14, 45.74, 43.74, 38.52, 37.92, 29.71, 27.18, 26.17, 24.58, 22.86, 13.60; **5**. minor isomer: ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 8.57 (s, 1 H), 7.36 (m, 3 H), 7.22 (m, 3 H),

6.71 (m, 1 H), 6.63 (m, 1 H), 3.82 (t, J=9.9 Hz, 1 H), 3.78 (s, 3 H), 2.86 (m, 2 H), 2.06 (m, 7 H), 1.68 (m, 1 H), 1.40 (m, 4 H), 1.22 (m, 1 H), 0.81 (s, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 163.34, 157.46, 140.01, 137.73, 132.09, 129.01, 127.88, 127.47, 126.18, 113.77, 111.37, 70.62, 55.16, 51.32, 44.46, 43.69, 38.61, 37.51, 29.64, 27.13, 26.14, 25.63, 22.55, 12.40; LRMS (ESI) [M+H]$^+$ found 390.3. **6**. ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.19 (m, 3 H), 6.72 (m, 5 H), 3.81 (s, 3 H), 3.52 (m, 1 H), 2.91 (m, 2 H), 2.29 (m, 3 H), 1.95 (m, 2 H), 1.82 (m, 1 H), 1.43 (m, 7 H), 0.83 (s, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 157.38, 148.26, 137.90, 132.58, 129.13, 126.28, 116.92, 113.75, 113.36, 111.39, 63.72, 55.16, 51.84, 43.88, 43.75, 38.98, 38.16, 29.97, 29.80, 27.35, 26.38, 23.33, 11.99; LRMS (ESI) [M+H]$^+$ found 362.4.

3-Hydroxy-*N*-phenylestra-1,3,5(10)-trien-17β-amine (7). A round bottomed flask charged with **6** (77.2 mg, 0.2 mmol) dissolved in DCM (4 mL) was cooled in a ice/NaCl bath. BBr$_3$ (0.5 mL of 1M solution in DCM) was added dropwise and the reaction mixture was allowed to warm to ambient temperature. After 2.5 hours the reaction was judge to be complete by TLC, re-cooled to 0 °C in an ice bath, and quenched with MeOH, and concentrated by rotary evaporation. The crude material was isolated by flash chromatography over silica gel (25% EtOAc in hexanes) and purified by recrystallization (MeOH/water) to yield **7** (24.2 mg, 32.6%) as white needles. ^{1}H-NMR (DMSO-d$_6$, 500 MHz) δ 8.99 (s, 1 H), 7.01 (m, 3 H), 6.65 (d, J=7.72 Hz, 2 H), 6.50 (dd, J=8.36, 2.57 Hz, 1 H), 6.44 (m, 2 H), 5.28 (d, J=8.36 Hz, 1 H), 3.42 (q, J=8.65 Hz, 1 H), 2.72 (m, 2 H), 2.21 (m, 1 H), 2.09 (m, 2 H), 1.80 (m, 1 H), 1.70 (m, 2 H), 1.42 (m, 2 H), 1.28 (m, 5 H)0.73 (s, 3 H); ^{13}C-NMR (DMSO-d$_6$, 125 MHz) δ 154.90, 149.23, 137.11, 130.33, 128.70, 126.06, 116.41, 114.92, 112.69, 112.40, 62.58, 51.17, 43.76, 43.43, 38.86, 37.98, 29.18, 28.77, 27.05, 26.17, 22.93, 12.05; HRMS (ESI): calc'd for C$_{24}$H$_{30}$NO [M+H]$^+$ 348.2327, found 348.2321.

3-Methoxy-*N*-methyl-*N*-phenylestra-1,3,5(10)-trien-17α-amine (8) A solution of **5** (57.0 mg, 0.15 mmol) dissolved in THF (1 mL) was slowly added to a round bottomed flask charged with LAH (26 mg, 0.6 mmol) in THF (1 mL). The reaction mixture was heated to 70 °C for 3 hours, after which the reaction was cooled to 0 °C, quenched with water, transferred to a separatory funnel, a solution of Rochelle salt was added, and the mixture was exhaustively extracted with DCM. The organic layer was dried over MgSO$_4$, vacuum filtered, and concentrated by rotary evaporation. The desired product was isolated by flash chromatography over silica gel (25% EtOAc in hexanes) to yield **8** (42.4 mg,77.3%) as a white solid. ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.26 (m, 3 H), 6.96 (d, J=8.36 Hz, 2 H), 6.76 (m, 2 H), 6.68 (d, J=2.36 Hz, 1 H), 3.85 (t, J=1.00 Hz, 1 H), 3.81 (s, 3 H), 2.91 (s, 3 H), 2.91 (m, 2 H), 2.30 (m, 2 H), 2.03 (m, 1

H), 1.91 (m, 4 H), 1.48 (m, 6 H), 0.88 (s, 3 H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 157.38, 151.96, 137.90, 132.70, 128.84, 126.26, 116.80, 114.11, 113.74, 111.38, 69.33, 55.17, 51.09, 45.55, 43.96, 38.56, 38.47, 35.60, 29.82, 27.36, 26.33, 24.01, 23.23, 13.40.

3-Hydroxyoxy-*N*-methyl-*N*-phenylestra-1,3,5(10)-trien-17α-amine (9). Following the same procedure as for **7** in the cleavage of aryl-methyl ethers from **8** (46.0 mg, 0.12 mmol) and BBr$_3$ (0.25 mL of a 1M solution in DCM) yielded **9** (32.4 mg, 73.2%) as light yellow needles. ^{1}H-NMR (DMSO-d$_6$, 500 MHz) δ 8.99 (s, 1 H), 7.15 (dd, J=8.55, 7.32 Hz, 2 H), 7.02 (d, J=8.55 Hz, 1 H), 6.87 (d, J=8.06 Hz, 2 H), 6.62 (t, J=7.20 Hz, 1 H), 6.49 (dd, J=8.30, 2.69 Hz, 1 H), 6.43 (d, J=2.69 Hz, 1 H), 3.84 (m, 1 H), 2.80 (s, 3 H), 2.72 (m, 2 H), 2.20 (m, 1 H), 2.12 (m, 1 H), 1.99 (m, 1 H), 1.81 (m, 1 H), 1.75 (m, 2 H), 1.63 (m, 1 H), 1.52 (td, J=12.57, 3.66 Hz, 1 H), 1.29 (m, 5 H), 0.78 (s, 3 H); ^{13}C-NMR (DMSO-d$_6$, 125 MHz) δ 154.91, 151.46, 137.06, 130.33, 128.70, 125.99, 116.23, 114.90, 113.60, 112.68, 68.11, 50.24, 45.25, 43.43, 38.30, 37.96, 35.09, 29.15, 27.01, 26.00, 23.49, 22.85, 13.32; HRMS (ESI): calc'd for C$_{25}$H$_{32}$NO [M+H]$^+$ 362.2484, found 362.2479.

B. Molecular Modeling

The structure for ERα complexed to OBHS was obtained from Kendall Nettles and prepared using Sybyl 8.1.1. Any excess chains were removed to yield the monomeric structure, to which explicit hydrogens were added, partial charges calculated, and minimized with MMFF94 to a termination gradient of 0.5. The coactivator peptide and all water molecules were deleted with the exception of the ordered water participating in the hydrogen bonding network between Glu353, Arg394, and the ligand (OBHS). The resulting structure was further processed using AutoDock Tools (ADT) to compute Gasteiger charges and assign AD4 atom types.

All docked compounds were constructed in Sybyl, cleaned up using Concord, and minimized with MMFF94 to a termination gradient of 0.05. The ligands were further prepared for docking with ADT to calculate Gasteiger charges, assign AD4 atom types, and set all rotatable bonds as active torsions. Each ligand was then docked into the prepared receptor structure using AutoDock Vina (ADV). The grid box was centered on the position of OBHS and measured 22 Å by 20 Å by 26 Å, and the exhaustiveness parameter was set to 100 (default=8, linear scale). All other default settings were used.

The lowest energy pose as determined by the internal ADV scoring algorithm was selected for further use. The docked pose was merged back into the receptor structure in Sybyl, all hydrogens were explicitly added, and the receptor ligand complex was re-minimized to a termination gradient of 0.5. Each receptor-ligand pair was overlaid with the initial OBHS-ERα

crystal structure, and analyzed visually. Particular attention was paid to the geometric fit of the ligand within the binding volume, matched hydrogen bonding contacts, and positioning of the extended D-ring substituent relative to the phenyl sulfonate of OBHS. Suitable ligands were further analyzed by generating a molCAD surface of ligand and the receptor surface within the binding volume, which was subsequently mapped by lipophilic potential.

VIII. References

(1) Kurebayashi, S.; Miyashita, Y.; Hirose, T.; Kasayama, S.; Akira, S.; Kishimoto, T. *J. Steroid Biochem. Mol. Biol.* **1997**, *60*, 11.
(2) Kalaitzidis, D.; Gilmore, T. D. *Trends Endocrinol. Metab.* **2005**, *16*, 46.
(3) Steffan, R. J.; Matelan, E.; Ashwell, M. A.; Moore, W. J.; Solvibile, W. R.; Trybulski, E.; Chadwick, C. C.; Chippari, S.; Kenney, T.; Eckert, A.; Borges-Marcucci, L.; Keith, J. C.; Xu, Z.; Mosyak, L.; Harnish, D. C. *J. Med. Chem.* **2004**, *47*, 6435.
(4) Chadwick, C. C.; Chippari, S.; Matelan, E.; Borges-Marcucci, L.; Eckert, A. M.; Keith, J. C.; Albert, L. M.; Leathurby, Y.; Harris, H. A.; Bhat, R. A.; Ashwell, M.; Trybulski, E.; Winneker, R. C.; Adelman, S. J.; Steffan, R. J.; Harnish, D. C. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 2543.
(5) Zhou, H.-B.; Comninos, J. S.; Stossi, F.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A. *J. Med. Chem.* **2005**, *48*, 7261.
(6) Nettles, K. W.; Bruning, J. B.; Gil, G.; Nowak, J.; Sharma, S. K.; Hahm, J. B.; Kulp, K.; Hochberg, R. B.; Zhou, H.; Katzenellenbogen, J. A.; Katzenellenbogen, B. S.; Kim, Y.; Joachimiak, A.; Greene, G. L. *Nat. Chem. Biol.* **2008**, *4*, 241.
(7) Bruning, J. B.; Parent, A. A.; Gil, G.; Zhao, M.; Nowak, J.; Pace, M. C.; Smith, C. L.; Afonine, P. V.; Adams, P. D.; Katzenellenbogen, J. A.; Nettles, K. W. *Nat. Chem. Biol.* **2010**, *6*, 837.
(8) Nettles, K. W., unpublished data.
(9) Celik, L.; Lund, J. D. D.; Schiøtt, B. *Biochemistry* **2007**, *46*, 1743.
(10) Baran, J. S.; Langford, D. D.; Laos, I.; Liang, C. D. *Tetrahedron* **1977**, *33*, 609.
(11) Bindal, R. D.; Carlson, K. E.; Reiner, G. C. A.; Katzenellenbogen, J. A. *J. Steroid Biochem.* **1987**, *28*, 361.
(12) Black, L. J.; Goode, R. L. *Life Sci.* **1980**, *26*, 1453.
(13) Stoessel, S.; Leclercq, G. *J. Steroid Biochem.* **1986**, *25*, 677.
(14) Gyling, M.; Leclercq, G. *J. Steroid Biochem.* **1988**, *29*, 1.
(15) Molecular Operating Environment (MOE), Chemical Computing Group, http://www.chemcomp.com/.
(16) Sanner, M. F. *J. Mol. Graphics Model.* **1999**, *17*, 57.
(17) Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, *31*, 455.
(18) Sybyl 8.1.1, Tripos International, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
(19) Allan, G. M.; Lawrence, H. R.; Cornet, J.; Bubert, C.; Fischer, D. S.; Vicker, N.; Smith, A.; Tutill, H. J.; Purohit, A.; Day, J. M.; Mahon, M. F.; Reed, M. J.; Potter, B. V. L. *J. Med. Chem.* **2006**, *49*, 1325.
(20) Johnson, R. A.; Bundy, G. L.; Youngdale, G. A.; Morton, D. R.; Wallach, D. P., Cyclic Hydrocarbons with an Aminoalkyl Sidechain. U.S. Patent 4,917,826, **1990**.
(21) Katzenellenbogen, J. A.; Johnson, H. J., Jr.; Myers, H. N. *Biochemistry* **1973**, *12*, 4085.
(22) Carlson, K. E.; Choi, I.; Gee, A.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A. *Biochemistry* **1997**, *36*, 14897.
(23) De Angelis, M.; Stossi, F.; Waibel, M.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A. *Biorg. Med. Chem.* **2005**, *13*, 6529.
(24) Minutolo, F.; Macchia, M.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A. *Med. Res. Rev.* **2009**, n/a.

(25) De Angelis, M.; Stossi, F.; Carlson, K. A.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A. *J. Med. Chem.* **2005**, *48*, 1132.

(26) Brzozowski, A. M.; Pike, A. C. W.; Dauter, Z.; Hubbard, R. E.; Bonn, T.; Engstrom, O.; Ohman, L.; Greene, G. L.; Gustafsson, J.-Å.; Carlquist, M. *Nature* **1997**, *389*, 753.

(27) Pike, A. C. W.; Brzozowski, A. M.; Hubbard, R. E.; Bonn, T.; Thorsell, A.-G.; Engstrom, O.; Ljunggren, J.; Gustafsson, J.-A.; Carlquist, M. *EMBO J* **1999**, *18*, 4608.

CHAPTER 4

COMPUTATIONAL METHODS IN EVALUATING THE BINDING AFFINITY OF FLUORINATED ANALOGUES OF TANAPROGET FOR THE PROGESTERONE RECEPTOR[†]

I. Introduction

The progesterone receptor (PR) is another member of the nuclear receptor superfamily and is primarily found in female reproductive tissues where it is bound by the endogenous steroidal ligand, progesterone. This ligand-receptor pair is commonly associated with a variety of biological functions related to pregnancy; however, elevated PR levels are often found in ER-positive breast cancer tumor cells as a result of ER-dependent expression. This phenomenon has led to the investigation of the use of PR as a target in breast cancer therapy,[1-5] and as an indicator of responsiveness to endocrine therapies.[6-8] The research described herein will focus on this latter application.

An ER-positive tumor status is a well known requirement for efficacy in endocrine therapies; however, only 50% of such patients will respond to treatment. One explanation for this outcome is that while ER is present in these tumors, ER-pathways are non-functional. Since PR expression is ER-dependent, the presence of elevated PR levels in tumor cells should indicate properly functioning ER-mediated transcriptional pathways that would be susceptible to standard endocrine therapies.[9,10] This theory has been investigated by a number of clinical studies with conflicting results,[11-15] and it has been suggested that the practical difficulties in measuring PR levels may obfuscate these data, leading to significant mis-categorization of PR status.[16]

An initial approach to addressing these issues may be to magnify the effect of ER-dependent expression of PR by taking advantage of a phenomenon known as "clinical flare reaction."[16-18] During the initial stages of endocrine therapy, tamoxifen acts as an agonist on ER, frequently leading to a temporary worsening of disease symptoms. These tamoxifen-induced flares are strongly correlated to a positive response to endocrine therapies; however, it is difficult to differentiate between worsening of symptoms due to a flare vs. disease progression, or identify the sub-clinical symptoms of mild flares. As early as 1980, Baulieu[19] suggested using "tamoxifen as a hormonal challenge test" in which the patient is treated with tamoxifen for a period of time,[7] followed by analysis of PR levels as a method to quantitate cellular response. This method, however, necessitates a tumor biopsy and does not completely address the practical difficulties in measuring PR levels.
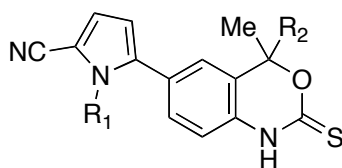

Aiming to circumvent the problems associated with directly measuring PR levels in tumor biopsies, the Welch laboratory at Washington University in St. Louis investigated a method for non-invasive quantitative assessment of flares using positron emission tomography (PET). Rather than focusing on PR levels, these studies assessed the change in the metabolic state of tumor cells using 2-[^{18}F]fluoro-2-deoxyglucose (FDG) as a chemical probe.[20-23] The underlying assumption of this methodology is that the stimulation of ER pathways responsible for flares during initial stages of tamoxifen treatment would be accompanied by increased metabolism, which is readily observed by monitoring tumor glucose uptake.   Accordingly, FDG was administered prior to and 7-10 days after treatment with tamoxifen, and visualized by PET scan. It was found that quantitative analysis of the change in local concentrations of FDG was predictive for response to endocrine therapy (tamoxifen) in a statistically significant manner. While this method circumvents the need for tumor biopsy and provides a rapid assessment of therapeutic response, it hinges the readily available but otherwise general FDG as the chemical probe for changes in metabolism. We have subsequently suggested that the use of a probe specifically targeted for an ER-gene product (e.g., PR) would yield better results,[23-25] and therefore, we selected a newly disclosed synthetic PR agonist, Tanaproget,[26] to serve this purpose.

All PR agonists currently approved for use in oral contraceptives are steroidal analogs of progesterone, which suffer from a number of side effects due to low selectivity for PR. Tanaproget, specifically developed by Wyeth as a non-steroidal PR agonist,[27] boasts significantly higher affinity and better selectivity for PR, and is currently in Phase II clinical trials. In addition to reporting activity data for various analogs, the researchers at Wyeth were able to obtain a crystal structure of Tanaproget bound to PR that explains the observed trends in activity, primarily through increased hydrogen bonding.[28] We have leveraged this information to introduce a ^{18}F-radiolabel into several analogs of Tanaproget using rational design, which were subsequently synthesized by Dr. Haibing Zhou and Dr. Jae Hak Lee in our labs.[29]

The binding affinities for the Tanaproget analogs shown in Table 4.1 show a strong dependence on the nature and position of the alkyl or fluoroalkyl substituent on the pyrrole-benzoxazin-2-thione ligand core. Thus, while the pyrrole N-methyl group of **1** is well tolerated (**1**, RBA = 151), the larger fluoroalkyl substituents are not (**3** = 18.5, **4** = 0.99), indicating that the ligand-binding pocket of PR has very limited tolerance for more than a methyl group. By contrast, when fluoroethyl or propyl substituents replace one of the C4 methyl groups of the benzoxazin-2-thione unit, high affinity compounds are produced, with three compounds (**5**, **6**, **8**) having equivalent or higher binding affinity than that of compound **1** itself and indicating that this region of the PR ligand binding pocket has good bulk tolerance, as suggested by prior work.[26]

Intriguingly, there appears to be some "reciprocity" between the substituent on the pyrrole

Table 4.1. Relative Binding Affinities (RBAs) of Tanaproget Derivatives

compound	R₁	R₂	RBA[a] (R5020 = 100)
Tanaproget (**1**)	CH_3	CH_3	151±39
2	H	CH_3	600±8
3	CH_2CH_2F	CH_3	18.5±5.2
4	$CH_2CH_2CH_2F$	CH_3	0.99±0.28
5[b]	CH_3	CH_2CH_2F	151±13
6[b]	H	CH_2CH_2F	198±30
7[b]	CH_3	$CH_2CH_2CH_2F$	90.9±27
8[b]	H	$CH_2CH_2CH_2F$	189±40

[a] RBA values determined by competitive radiometric binding assay, using [³H]R5020 as a tracer.

[b] Compounds **5**, **6**, **7**, **8** were studied as racemates.

nitrogen and at the C4 position such that larger groups (fluoroethyl and fluoropropyl) are more readily accommodated at C4 when there is a smaller group (H) on the pyrrole nitrogen. Also, it is of note that while the pyrrole N-substituted compounds **3** and **4** are achiral, the high affinity Tanaproget analogs modified at C4 position of benzoxazin-2-thione are chiral, but thus far have only been studied as racemates. Thus, it is possible that one enantiomer will have higher binding affinity than the other.

To investigate further the structure-binding affinity correlations we have uncovered, including the apparent reciprocity between the pyrrole N-substituent and the C4-substituents, and to explore possible enantioselectivity, we set out to build structural models for each ligand using molecular modeling software. Each model was evaluated both visually and with a series of quantitative analyses to identify structural characteristics that modulate the observed RBA values.

II. Ligand Docking

Structural models of each receptor-ligand pair were constructed *in silico* by docking the ligand into the ligand binding pocket of PR. Conveniently, a 2 Å resolution crystal structure of PR complexed to **1** was available in the PDB (accession code 1ZUC),[28] and used as a guide for each docking. All ligands were constructed using Sybyl[30] and docked into the protein structure

using AutoDock Vina.[31]   Acceptable low energy poses were manually selected by visual analysis, further minimized using a multistep protocol, and finally analyzed using several numerical methods within the Molecular Operating Environment (MOE).[32]

III. Visual Analysis

Visual analysis of the N-substituted analogs yielded a relatively straightforward explanation for the reduced affinity with increasing substituent size. As depicted in Figure 4.1, the benzoxazine-core establishes a consistent position in the binding pocket while the pyrrole rotates away from Gln725 with increasing substituent size. This rotation greatly reduces the ability of the nitrile substituent to engage the known hydrogen bonding network between Gln725 and Arg766, an interaction that is essential for high affinity binding. Furthermore, the largest substituent, $CH_2CH_2CH_2F$ (compound **4**), begins to challenge the confines of the available binding volume while Me substituent easily fits entirely within the pocket.
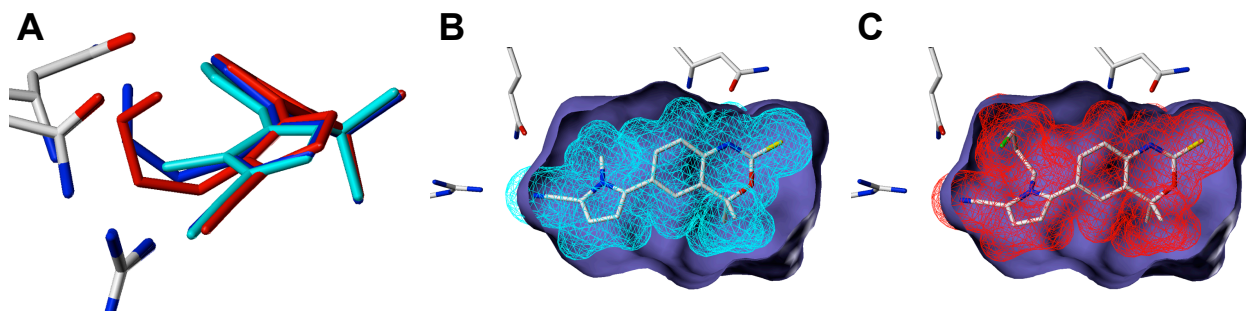
Figure 4.1. Docking Poses for N-substituted analogs of Tanaproget. (A) The steric bulk of the pyrrole N-substituent for **1** (cyan), **3** (blue), and **4** (red) induces an increasingly large dihedral angle, rotating the nitrile away from Gln725, and resulting in reduced hydrogen bonding. (B) The van der Waals surface of **1** (cyan lines) clearly fits within the binding volume (purple solid); however, (C) the surface of **4** (red lines) begins to challenge the confines of the pocket.

Shifting our focus to C4-substitution, we observed that while docked poses for the *N*-Me and *N*-H series of analogues show a conserved binding mode, the C4-substituent only subtly affects ligand conformation and placement of the benzoxazine-core. Overlaying the docked poses for compounds in the *N*-Me (Fig. 4.2) and *N*-H series (not shown) suggest that the gem-dimethyl substituted compounds adopt an intermediate position. As the size of the C4-substituent increases, the compound rotates to a degree related to the steric bulk and in the direction that reflects the stereochemical configuration.

The poses presented in Figure 4.2 imply that the stereochemical configuration at C4 should have a marked effect on the ligand-protein interactions. The *R*-configuration directs the substituent down in the binding pocket, where steric constraints force the core to rotate up and adopt a relatively small dihedral angle. As will be discussed in section **IV**.A., the shift to a smaller dihedral angle represents a higher energy ligand conformation. It is notable that the *N*-
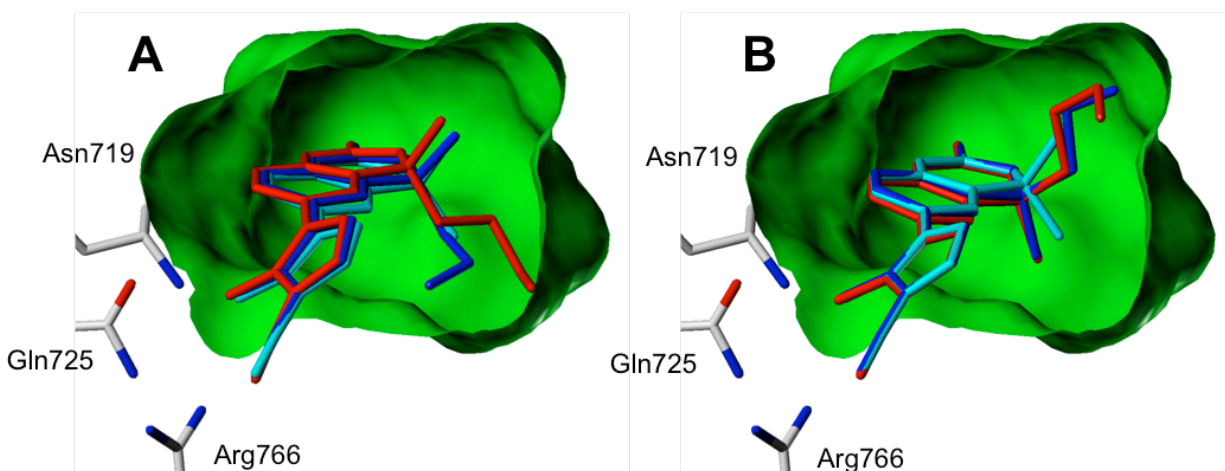
Figure 4.2. Effects of methyl (cyan), fluoroethyl (blue), and fluoropropyl (red) substitution at C4 of Tanaproget. As substituent bulk increases, the core of the *R*-enantiomer (A) is forced upward to adopt a higher energy ligand conformation, while the *S*-enantiomer (B) allows a downward rotation to a more energetically favorable dihedral angle.

H analogues follow a similar trend, albeit with a much smaller energy difference. The *S*-configuration, in contrast, directs the substituent up in the pocket, causing the core to rotate downward and adopt a more energetically favorable dihedral angle.

IV. Numerical Analysis

While visual analysis provided qualitative observations that explained the observed binding affinities for N-substituted analogs and suggests enantioselectivity for C4-substituted analogues, there is no obvious structural basis to differentiate the C4-substituted series of analogs. Therefore, we sought out a method to quantitate receptor-ligand interactions and identify a relationship that directly correlates to the observed RBA values. Such a method would allow us to further elaborate on structure-activity relationships established using visual analysis, and evaluate additional analogues prior to engaging synthetic efforts.

A. Torsion Energetics

Evaluating Figure 4.1, the benzoxazine-core of each ligand adopts a strikingly similar pose within the binding volume regardless of N-substitution, highlighting the rotation of the pendant pyrrole as the driving force behind reduced binding affinity within the *N*-R series, but also carries over to *N*-Me. Accordingly, we targeted the effects of substituent size on ligand conformation by analyzing the torsional energetics of the ligand independently of the protein. The contributing energy terms were calculated separately while rotating the pyrrole from 0° to 90° with respect to the benzoxazine-core as shown in Figure 4.3. For the unsubstituted pyrrole **2**, the torsional (tor), electronic (ele), and van der Waals (vdW) terms all significantly contribute
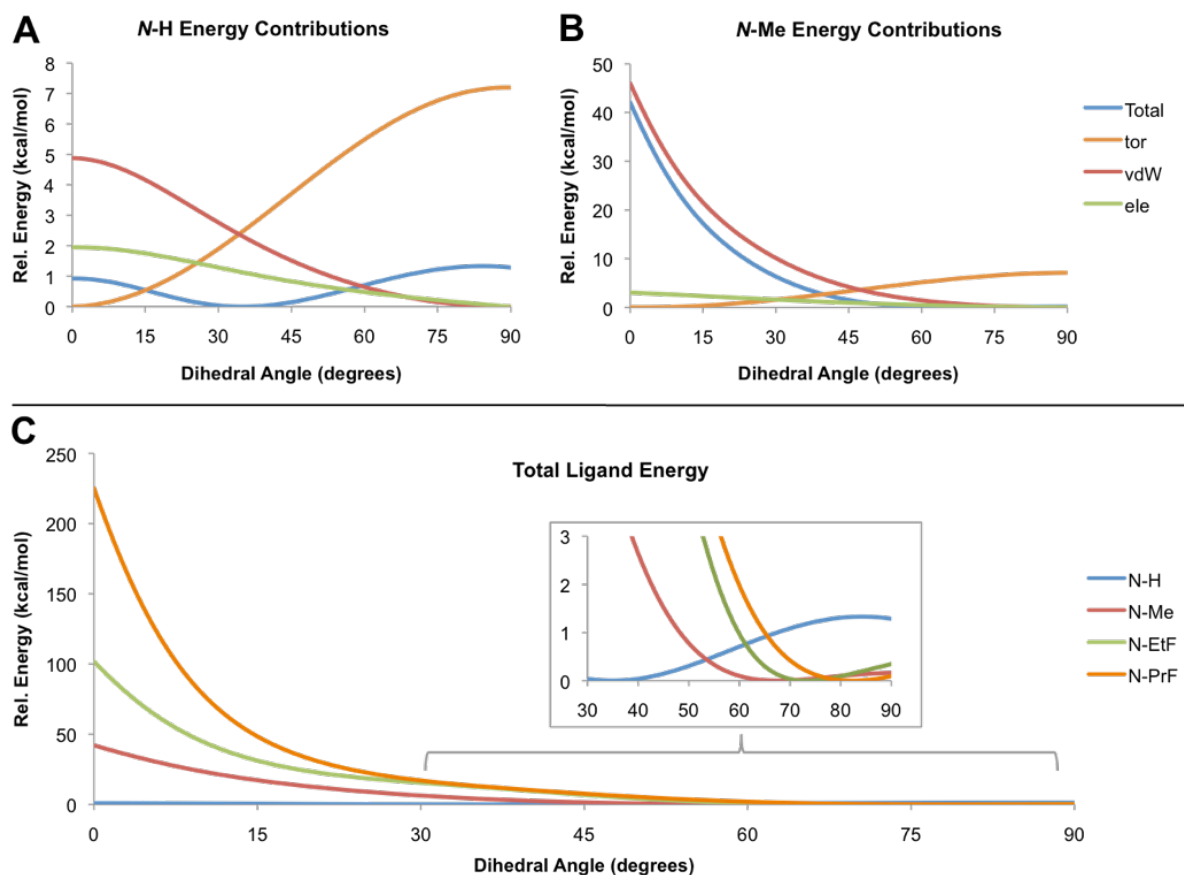
Figure 4.3. Ligand energy dependence on pyrrole-benzoxazin-2-thione core dihedral angle. At relatively small angles, several energy terms contribute to the change in total ligand energy when rotating around the bond connecting the *N*-H pyrrole and benzoxazin-2-thione core (A). N-Substitution at the pyrrole (B) changes these torsional dynamics by drastically increasing the magnitude of the vdW energy term. This effect is magnified with increasing substituent size (3C) and results in a shift to larger dihedral angles for the minimum energy conformations (C inset). (Note differences in the energy scales on the different panels).

to the total energy. For all cases of N-substitution (e.g., *N*-Me, Fig 4.3B.), however, the vdW term dominates the total ligand energy. Increasing the substituent size further magnifies this effect (Fig. 4.3C), and by extension, results in a shift in the minimum energy to larger dihedral angles (Fig. 4.3C-inset). This shift in dihedral angle mirrors the rotation of the pyrrole in the minimized docking poses for N-substituted analogues.

B. Interaction Energy

The torsional energetics of the ligand indicate a shift to larger dihedral angles with N-substitution; however, it is unclear what role the protein plays in accommodating or opposing the rotation of the pyrrole towards reaching a low energy ligand conformation. To this end, we probed the rotational freedom of the pyrrole within the binding pocket through the calculation of an "interaction energy" as the pyrrole was rotated relative to the core. The interaction energy,

defined as the energy difference between the receptor-ligand complex and the individual energy contributions of the receptor and ligand, represents the energy associated directly with the interaction between the receptor and ligand pair. It is important to note that this calculation does not take into account the energy associated with changes in receptor or ligand conformation due to a binding event, and therefore, is not a binding energy (see section **VI.**). Based on these calculations the width of the energy well represents the rotational freedom of the pyrrole within the context of the binding volume. Figure 4.4 shows the anticipated outcome in which the unsubstituted pyrrole enjoys a wide range of dihedral angles that form low energy contacts, while the N-substituted analogs display increasingly narrow energy wells corresponding to a tighter fit. Somewhat unexpectedly, the *N*-methyl pyrrole demonstrates a lower energy interaction compared to the unsubstituted analog, indicating the presence of a positive non-specific interaction between the *N*-methyl and receptor. Any energy gained from this positive interaction would serve to help offset the energy penalty resulting from forcing the ligand into a sub-optimal conformation in order to fit into the binding volume, as observed from the docking poses.

$$E_{int} = E_{RL} - (E_R + E_L)\qquad \textbf{(4.1)}$$
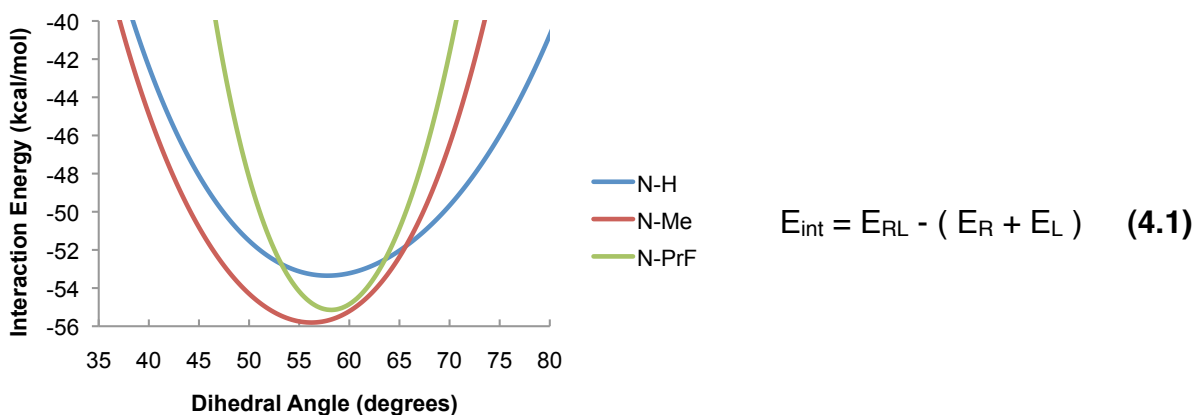
Figure 4.4. Calculating the interaction energy using Eq. **(4.1)** suggests that small substitutions (e.g., *N*-Me and *N*-EtF) have less rotational freedom but make beneficial contacts with the receptor compared to unsubstituted analogues.

C. Hydrogen Bond Scoring

From experimental SAR and PR-ligand crystal structures, it has been well established that hydrogen bonding with residues Gln725, Arg766, Asn719, and a crystallographic water is a requirement for tight binding.[26,28] Hydrogen bonding is heavily dependent on the distance and angle between the two contacts; therefore, the subtle shifts in the position of the core potentially attenuates these interactions in a measurable manner. Accordingly, the hydrogen bonding contacts were scored using the "ligand interactions" module in MOE and are given in Table 4.2. On the basis of the position and interdependence of the hydrogen bonding partners, we modeled RBA values as a function of the geometric mean of the calculated hydrogen bonding scores. Linearization of this function yielded a highly correlated relationship (Figure 5, slope =

Table 4.2. Calculated Hydrogen Bonding Scores[a] for Docked Poses of Tanaproget Analogues

Series	ligand	Gln725	Arg766	Asn719	H_2O	geometric mean	AVG	RBA
NR	**3**	0.57	0.42	0.38	0.27	0.396	0.396	18.5
	4	0.23	0.35	0.40	0.29	0.311	0.311	0.99
NMe	**1**	0.61	0.49	0.40	0.27	0.424	0.424	151
	R-**5**	0.57	0.50	0.45	0.24	0.419	0.425	151
	S-**5**	0.62	0.47	0.42	0.28	0.430		
	R-**7**	0.60	0.54	0.44	0.20	0.411	0.421	90.9
	S-**7**	0.62	0.49	0.42	0.27	0.431		
NH	**2**	0.68	0.45	0.43	0.30	0.446	0.446	600
	R-**6**	0.65	0.46	0.46	0.27	0.439	0.445	198
	S-**6**	0.69	0.44	0.44	0.31	0.451		
	R-**8**	0.66	0.49	0.44	0.25	0.434	0.440	189
	S-**8**	0.71	0.47	0.44	0.27	0.446		

[a] Calculated using the "ligand interactions" module in MOE.
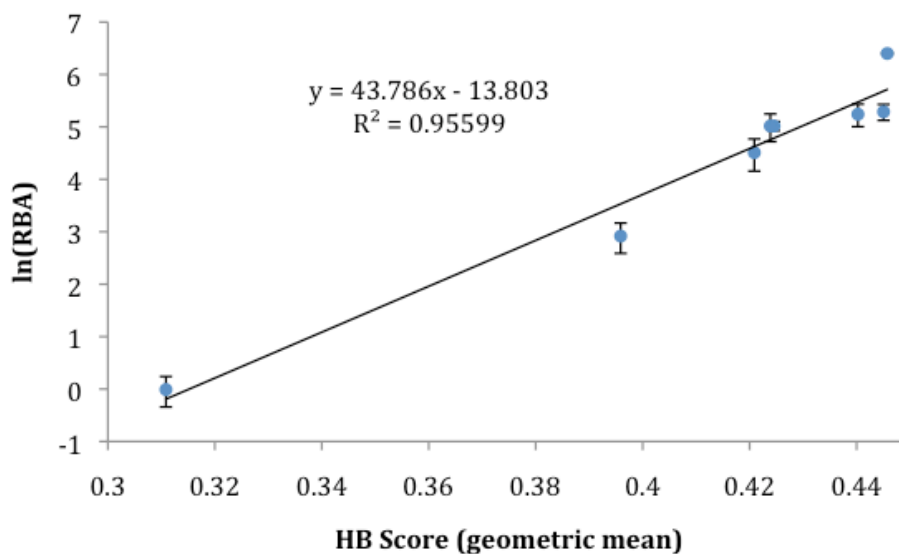
Figure 4.5. Linear Fit Relating the Calculated Hydrogen Bonding Score to RBA.

0.43, $R^2 = 0.96$), and the mean hydrogen bonding scores correctly predict the relative order of binding affinity across all three series, including extremely low (e.g., **3**, **4**) and high (e.g., **2**) affinity compounds.

V. A Unified Model for Assessing Binding Affinity

As described in the previous sections, we have successfully identified three methods for the analyzing structural models of receptor-ligand complexes. Together, these methods (visual analysis, torsional energetics, and hydrogen bonding) represent a cohesive model for assessing the molecular basis behind the observed trends in ligand binding affinity. The application of our model to understand the observed RBA values for compounds **3** and **4** is rather straightforward. The rotation of the pyrrole away from Gln725, as highlighted in Figure 4.1, corresponds to the erosion of the mean hydrogen bonding score (Table 4.2, series *N*-R) with increased substituent size. The difference between RBA values for the remainder of the compounds in Table 4.1 are much smaller, and the analysis correspondingly more subtle. We propose that the ligand binds in a pose generally conserved for all analogs; the ligand conformation strikes a delicate balance between positioning the core to accommodate substitution at C4, maximizing hydrogen bonding contacts and positive interactions between the pyrrole and receptor, and minimizing internal ligand energy.

The interplay between these interactions is best illustrated by the marked decrease in RBA value for compound **7**. The large substituent at C4 restricts the rotational freedom of the core, while *N*-Me substitution increases the internal vdW energy of the ligand, resulting in a high-energy conformation that precludes strong hydrogen bonding interactions. The reduction of C4-substituent size allowing the core to rotate to a more favorable position (ex. **5**), or the removal of the N-substituent reducing the energy associated with internal torsion (ex. **8**), results in increased hydrogen bonding scores that correspond well to the observed higher RBA values. This model also supports the prediction that the *S*-configuration of the C4-substitued systems will have higher affinity than the *R*-stereoisomer.

While our model is rather simplistic, we obtain with it a linear relationship between HB score and ln(RBA) having a correlation coefficient over 0.95 (Fig. 4.5). This relationship accurately predicts the relative order of binding affinity, and it gives predicted RBA values within an average factor of 1.4 (and always better than a factor of 2) relative to experimentally determined ones, comparable to the error in experimentally determined values (cf. Table 4.1). We suspect that the limitations in our model could be reduced through the development of a more complex model that includes a direct treatment of energy calculations, the evaluation of non-specific protein-ligand interactions, and multivariate analysis. Alternative methodologies such as molecular dynamics simulations[33] or quantum mechanical calculations[34] require substantially more

computational resources and the development of new parameters and methodologies. Our very simple approach of analyzing ligand binding affinity through the quantification of the hydrogen bonding contacts, however, is suitable for explaining the binding trends of the compounds we have studied (Table 4.1), and it generates a valuable predictive model for exploring enantioselectivity using pre-built modules and minimal computational resources.

VI. Alternative Approaches

The observation that the *N*-methyl pyrrole forms positive non-specific contacts with the binding pocket based on our binding interaction calculations may suggest that the MMFF94x forcefield is competent in assessing the energetics of matched or unmatched van der Waals contacts, at least for this relatively simple binary comparison (*N*-Me vs. *N*-H). These interactions in conjunction with the well defined electrostatic components of hydrogen bonding comprise the majority of protein-ligand interactions in PR. With this in mind, we extended the interaction energy calculation to the docked poses for all ligands in the hope that it would correlate to observed RBA values. This data, shown in Table 4.3, correctly predicted the relative order of RBA within *N*-R and *N*-Me series; however, failed to properly order the *N*-H series or

$$E_{bin} = E_{RL} - (E_{Rmin} + E_{Lmin}) \quad \textbf{(4.2)}$$

Table 4.3. Calculated Total, Interaction[a], and Binding[b] Energies for Docked Poses of Tanaproget

Series	ligand	E_{RL}	E_{int}	E_{bin}	AVG E_{RL}[d]	AVG E_{int}	AVG E_{bin}	RBA
NR	**3**	488.9	-55.0	-367.0	488.9	-55.0	-367.0	18.5
	4	498.7	-52.5	-355.6	498.7	-52.5	-355.6	0.99
NMe	**1**	486.0	-58.0	-371.1	486.0	-58.0	-371.1	151
	R-**5**	490.3	-60.6	-366.9	488.1	-60.9	-368.5	151
	S-**5**	485.9	-61.2	-370.1				
	R-**7**	500.5	-55.0	-359.6	496.7	-55.6	-361.4	90.9
	S-**7**	493.0	-56.1	-363.3				
NH	2	481.9	-55.4	-368.4	481.9	-55.4	-368.4	600
	R-**6**	485.8	-58.4	-366.7	484.0	-58.4	-367.1	198
	S-**6**	482.2	-58.5	-367.5				
	R-**8**	496.7	-54.4	-355.7	493.3	-53.5	-357.9	189
	S-**8**	490.0	-52.5	-360.2				

[a] Interaction Energy calculated from Eq. **(4.1)**. [b] Binding Energy calculated from Eq. **(4.2)**. [c] All energy calculations are given in kcal/mol. [d] ERL represents the total energy of the receptor-ligand complex.

allow cross-series comparisons.

While these calculations provide some insight into isolated interactions between highly related species, the interaction energy does not take into account the energy associated with the change in the receptor and ligand conformations between the unbound and bound states. This shortcoming is clearly born out in the failure of cross-series comparisons, in which the energy required to adopt the internal torsion angle of the docked ligand is substantially different for substituted vs. unsubstituted pyrrole species as highlighted by Figure 4.3. In order to account for these discrepancies, Eq. **(4.1)** was readily modified to approximate the binding energy by minimizing the receptor and protein individually prior to calculating their energies, as given in Eq. **(4.2)**. This methodology was applied to the docking pose for each ligand and given in Table 4.3. Performing slightly better than the interaction energy, this method was capable of predicting the relative order of RBA within all sereis; however, it also failed in cross-series comparisons.

VII. Proposed Ligands as Radiotherapeutics

Preliminary data from work by Dr. Ephraim Parent in our lab suggests that estrogens labeled with ^{76}Br/^{77}Br may be an effective ER-directed radiotherapy for breast cancers. One potential drawback to this type of targeting system is that ER is expressed in many different non-target tissues which could be damaged in addition to the tumor. PR, in contrast, is expressed in many fewer tissues yet can reach high expression levels in some breast tumors, potentially making it a better target for this type of therapy. Our collaborators in the Welch group have approached this strategy using fluoro furanyl norprogesterone (FFNP) analogs; however, these compounds demonstrated only low to moderate RBA values (0.81-65).[35,36]

We have designed a set of synthetically tractable compounds based on the Tanaproget core probing different labeling strategies for the incorporation of ^{76}Br/^{77}Br or ^{123}I/^{124}I. Each compound was then docked into the PR ligand binding pocket, post-processed, and evaluated using our unified model as described above. Visual analysis suggested that placing a bromide at position 1 strongly interfered with hydrogen bonding to Asn719 through steric shielding. Moving the bromine to positions 2 or 3, on the other hand, introduced steric bulk ortho to the pyrrole, causing it to rotate away from the optimal position required to engage the hydrogen bonding network between Gln725, Arg766, and the crystallographic water. Introducing the desired isotope into the *R*-position introduced more subtle effects to the positioning of the benzoxazine core. The trans-configuration of vinyl iodide yielded an extended system that challenged the confines of the binding pocket, while the cis-configuration positioned the iodide over the face of the heterocyclic core where there was more available space to occupy (Fig. 4.6A). Linear bromoalkyl substituents adopted conformations similar to the CH_2CH_2F (**6**) and

Table 4.4. Calculated Hydrogen Bonding Scores[a] of Potential Radiopharmaceuticals
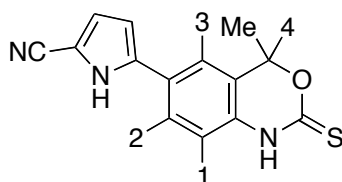
1	2	3	4	Gln725	Arg766	Asn719	H_2O	geometric mean	AVG mean
Br	H	H	CH₃	0.589	0.460	0.264	0.313	0.387	0.387
H	Br	H	CH₃	0.559	0.519	0.425	0.204	0.398	0.398
H	H	Br	CH₃	0.674	0.471	0.446	0.195	0.408	0.408
H	H	H	CH₂CH₂Br	0.616	0.424	0.409	0.279	0.415	0.417
H	H	H		0.740	0.510	0.362	0.224	0.418	
H	H	H	CH₂CH₂CH₂Br	0.627	0.440	0.454	0.279	0.432	**0.433**
H	H	H		0.757	0.476	0.418	0.236	0.434	
H	H	H	-CH=CHI (trans)	0.684	0.534	0.384	0.210	0.414	0.414
H	H	H		0.710	0.521	0.366	0.216	0.414	
H	H	H	-CH=CHI (cis)	0.716	0.580	0.444	0.167	0.419	**0.421**
H	H	H		0.682	0.550	0.439	0.196	0.424	

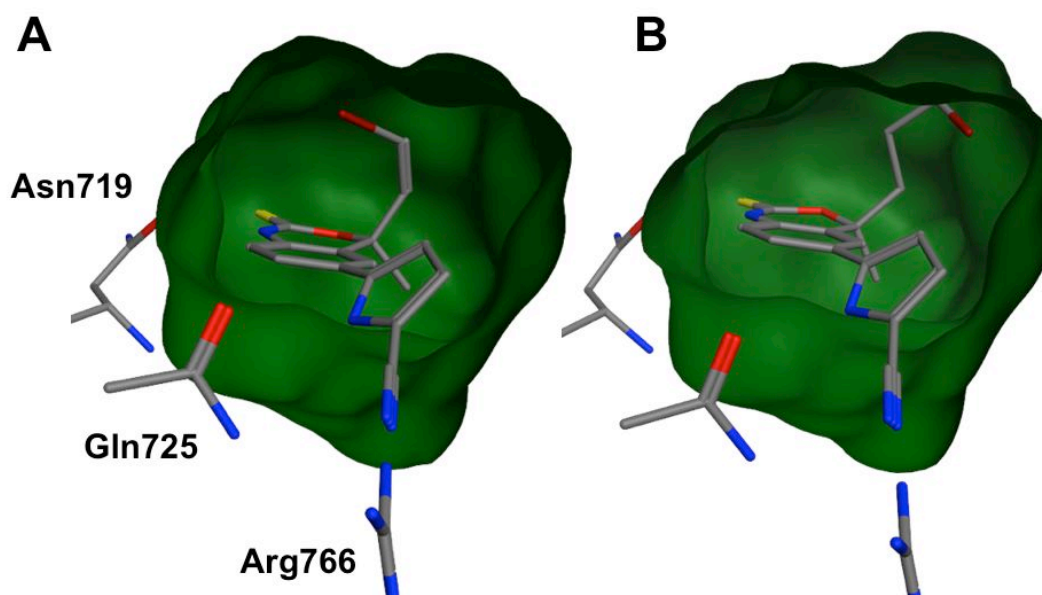[a] Calculated using the "ligand interactions" module in MOE.

Figure 4.6. Docked Poses of Two Potential Radiotherapeutics. (A) The *S*-cis-vinyl substituent at C4 is positions the vinyl iodide over the face of the benzoxazine-2-thione core, where there is available space. (B) The *S*-bromopropyl subtituent adopts a pose very similar to **5** (see Fig. 2B).

$CH_2CH_2CH_2F$ (**8**) conformations described in previous sections (Fig. 4.6B); however, the increased size of bromine relative to fluorine appeared to affect ideal positioning of the core . Each of these observations is also mirrored in the calculated hydrogen bonding scores given in Table 4.4. Based on these data, the synthesis of compounds bearing a cis-vinyliodide or bromopropyl substituent at the *R*-position are currently under consideration as initial synthetic targets.

VIII. Conclusions

We set out to probe the molecular basis for the binding trends observed for a series of analogues of Tanaproget, given in Table 1. Structural models were constructed *in silico* using ligand docking software followed by a multistage minimization protocol. The resulting receptor-ligand complexes were investigated both qualitatively and quantitatively using a variety of computational methods. From these data, we have been able to rationalize the distinct structure-binding affinity relationships in this series by molecular modeling of ligand internal energy and ligand-receptor hydrogen bonding score, and our model yields an intriguing explanation for the interaction between substituents at the two ends of the Tanaproget core as well as offering predictions of enantioselectivity. Finally, we have applied these methods to identify two compounds of interest as radiotherapeutics for the targeted treatment of breast cancer.

IX. Experimental

A. Procedure for docking Tanaproget (1) and analogs.

The PR structure was obtained from the PDB databank (1ZUC)[28] and prepared using the Molecular Operating Environment (MOE).[32] Explicit hydrogen atoms were added, partial charges were computed using the MMFF94x force field, and the receptor-ligand complex was minimized with a termination gradient of 0.5. All water molecules were then deleted except the single water molecule hydrogen bound to Gln725 and Arg766. Finally, the ligand was removed and the receptor structure was processed using AutoDock Tools[37] to define the AD4 atom types and calculate Gasteiger charges.

All compounds in Table 4.1 were constructed in Sybyl 8.1.1,[30] cleaned up using the built-in Concord module,[38] and minimized using the Powell method with a termination gradient of 0.5kcal/(mol*A), 100K maximum iterations, and MMFF94 force fields and charges. The ligands were prepared for docking using AutoDock Tools to assign AD4 atom types, calculate Gasteiger charges, and set all rotatable bonds as active torsions. Each ligand was docked into the receptor using AutoDock Vina.[31] The grid box was centered on the ligand in the original crystal structure, and measured 18 Å by 18 Å by 22 Å. To ensure that the proper binding conformation

was found, the exhaustiveness parameter was set to 100 (default=8, linear scale); all other default settings were used.

The top five poses for each ligand were visually inspected in MOE. Unreasonable poses were discarded and the lowest energy conformation of the remaining poses was selected for further use. The selected pose and receptor structure were merged into a single file, all hydrogen atoms were explicitly added, the partial charges were calculated using the MMFF94x force field, and the protein-ligand complex was minimized. The minimization was conducted in four stages: stage 1 minimized the hydrogen atoms that were added to the structure in the previous step; stage 2 minimized the ligand; stage 3 minimized any residues within 4.5 Å of the ligand; and stage 4 minimized both the ligand and any residues within 4.5 Å. This worked up structure was used for all other calculations and analyses.

B. Procedure for calculating the torsion energetics of N-substituted analogs.

The worked up structure was opened in MOE and all atoms deleted except for the ligand. The dihedral angle relating the pyrrole to the benzoxazin-2-thione core was rotated from 0-90 degrees, measuring each energy term in 0.5-degree increments. The relative value of each term was calculated by subtracting the minimum measured value.

C. Procedure for calculating the effect of N-substitution on interaction energy of the pyrrole.

Similar to the above process, the pyrrole was rotated with respect to the core by setting the dihedral angle from 0-90 degrees, measuring the interaction energy [Eq. **(1)**] in 0.5-degree increments. To ensure that any observed change in the interaction energy was due solely to a difference in the N-substitution, we started from the worked up pose of compound 5a, deleting atoms as necessary to obtain *N*-Me (**1**) and *N*-H (**2**). The partial charges were recalculated using MMFF94x force field prior to beginning the rotation.

D. Procedure for calculating the interaction and binding energies of the protein-ligand complex.

The worked up structure was opened in MOE and the total energy of the receptor-ligand complex was calculated (E_{RL}). The ligand was deleted from the structure and the receptor energy was calculated (E_R). The pocket residues (within 4.5A of the ligand) and associated water molecule were minimized to a termination gradient of 0.01, and the minimized receptor energy was calculated (E_{Rmin}). The worked up structure was reloaded, the receptor was deleted, and the ligand energy was calculated (E_L). Using these values, the minimized ligand energy (E_{Lmin}) was calculated following a minimization of the ligand to a termination gradient of 0.01. The interaction and binding energies were calculated from equations **(4.1)** and **(4.2)**, respectively.

E. Procedure for calculating the mean hydrogen bond score (Table 4.4).

     The hydrogen bonds formed between the ligand and protein were scored using the "dock_HydrogenBonds" module found in the standard scientific vector language (SVL) library for MOE. This module forms the basis for the "ligand interactions" feature accessible from MOE's graphical user interface. The mean score was subsequently calculated as the geometric mean of the individual scores.

X. References

(1)   Benagiano, G.; Bastianelli, C.; Farris, M. *Expert Opin. Pharmacother.* **2008**, *9*, 2487.
(2)   Lanari, C.; Molinolo, A. A. *Breast Cancer Res.* **2002**, *4*, 240.
(3)   Klijn, J. G.; Setyono-Han, B.; Foekens, J. A. *Steroids* **2000**, *65*, 825.
(4)   Santen, R.; Manni, A.; Harvey, H.; Redmond, C. *Endocrine Rev.* **1990**, *11*, 221.
(5)   Bai, C.; Flores, O.; Schmidt, A. *Expert Opin. Drug Discov.* **2007**, *2*, 725.
(6)   Namer, M.; Lalanne, C.; Baulieu, E. E. *Cancer Res.* **1980**, *40*, 1750.
(7)   Noguchi, S.; Miyauchi, K.; Nishizawa, Y.; Koyama, H. *Cancer* **1988**, *61*, 1345.
(8)   Howell, A.; Harland, R. N.; Barnes, D. M.; Baildam, A. D.; Wilkinson, M. J.; Hayward, E.; Swindell, R.; Sellwood, R. A. *Cancer Res.* **1987**, *47*, 300.
(9)   Cui, X.; Schiff, R.; Arpino, G.; Osborne, C. K.; Lee, A. V. *J. Clin. Oncol.* **2005**, *23*, 7721.
(10) Osborne, C. K.; Schiff, R.; Arpino, G.; Lee, A. S.; Hilsenbeck, V. G. *Breast* **2005**, *14*, 458.
(11) Rutqvist, L. E.; Cedermark, B.; Fornander, T.; Glas, U.; Johansson, H.; Nordenskjold, B.; Rotstein, S.; Skoog, L.; Somell, A.; Theve, T. *J. Clin. Oncol.* **1989**, *7*, 1474.
(12) Early Breast Cancer Trialist's Collaborative Group. *Lancet* **1998**, *351*, 1451.
(13) Fernö, M.; Stål, O.; Baldetrop, B.; Hatschek, T.; Källström, A.-C.; Malmström, P.; Nordenskjöld, B.; Rydén, S. *Breast Cancer Res. Treat.* **2000**, *59*, 69.
(14) Ellis, M. J.; Coop, A.; Singh, B.; Mauriac, L.; Llombert-Cussac, A.; Jänicke, F.; Miller, W. R.; Evans, D. B.; Dugan, M.; Brady, C.; Quebe-Fehling, E.; Borgs, M. *J. Clin. Oncol.* **2001**, *19*, 3808.
(15) Bardou, V. J.; Arpino, G.; Elledge, R. M.; Osborne, C. K.; Clark, G. M. *J. Clin. Oncol.* **2003**, *21*, 1973.
(16) Plotkin, D.; Lechner, J. J.; Jung, W. E.; Rosen, P. J. *JAMA* **1978**, *240*, 2644.
(17) Legha, S. S. *Ann. Intern. Med.* **1988**, *109*, 219.
(18) Vogel, C. L.; Schoenfelder, J.; Shemano, I.; Hayes, D. F.; Gams, R. A. *J. Clin. Oncol.* **1995**, *13*, 1123.
(19) Mortimer, J. E.; Dehdashti, F.; Siegel, B. A.; Katzenellenbogen, J. A.; Fracasso, P.; Welch, M. J. *Clin. Cancer. Res.* **1996**, *2*, 933.
(20) Dehdashti, F.; Flanagan, F. L.; Mortimer, J. E.; Katzenellenbogen, J. A.; Welch, M. J.; Siegel, B. A. *Eur. J. Nucl. Med.* **1999**, *26*, 51.
(21) Mortimer, J. E.; Dehdashti, F.; Siegel, B. A.; Trinkaus, K.; Katzenellenbogen, J. A.; Welch, M. J. *J. Clin. Oncol.* **2001**, *19*, 2797.
(22) Dehdashti, F.; Mortimer, J. E.; Trinkaus, K.; Naughton, M. J.; Ellis, M.; Katzenellenbogen, J. A.; Welch, M. J.; Siegel, B. A. *Breast Cancer Res. Treat.* **2009**, *113*, 509.
(23) Buckman, B. O.; Bonasera, T. A.; Kirschbaum, K. S.; Welch, M. J.; Katzenellenbogen, J. A. *J. Med. Chem.* **1995**, *38*, 328.
(24) Verhagen, A.; Studeny, M.; Luurtsema, G.; Visser, G. M.; De Goeij, C. C. J.; Sluyser, M.; Nieweg, O. E.; van der Ploeg, E.; Go, K. G.; Vaalburg, W. *Nucl. Med. Biol.* **1994**, *21*, 941.
(25) Kochanny, M. J.; VanBrocklin, H. F.; Kym, P. R.; Carlson, K. E.; O'Neil, J. P.; Bonasera, T. A.; Welch, M. J.; Katzenellenbogen, J. A. *J. Med. Chem.* **1993**, *36*, 1120.

(26) Fensome, A.; Bender, R.; Chopra, R.; Cohen, J.; Collins, M. A.; Hudak, V.; Malakian, K.; Lockhead, S.; Olland, A.; Svenson, K.; Terefenko, E. A.; Unwalla, R. J.; Wilhelm, J. M.; Wolfrom, S.; Zhu, Y.; Zhang, Z.; Zhang, P.; Winneker, R. C.; Wrobel, J. *J. Med. Chem.* **2005**, *48*, 5092.

(27) Bapst, J. L.; Ermer, J. C.; Ferron, G. M.; Foss, D.; Orczyk, G. P. *Contraception* **2006**, *74*, 414.

(28) Zhang, Z.; Olland, A. M.; Zhu, Y.; Cohen, J.; Berrodin, T.; Chippari, S.; Appavu, C.; Li, S.; Wilhem, J.; Chopra, R.; Fensome, A.; Zhang, P.; Wrobel, J.; Unwalla, R. J.; Lyttle, C. R.; Winneker, R. C. *J. Biol. Chem.* **2005**, *280*, 28468.

(29) Zhou, H.-B.; Lee, J. H.; Mayne, C. G.; Carlson, K. E.; Katzenellenbogen, J. A. *J. Med. Chem.* **2010**, *53*, 3349.

(30) Sybyl 8.1.1, Tripos International, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.

(31) Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, *31*, 455.

(32) Molecular Operating Environment (MOE), Chemical Computing Group, http://www.chemcomp.com/.

(33) Mordasini, T.; Curioni, A.; Bursi, R.; Andreoni, W. *ChemBioChem* **2003**, *4*, 155.

(34) Harada, T.; Yamagishi, K.; Nakano, T.; Kitaura, K.; Tokiwa, H. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **2008**, *377*, 607.

(35) Zhou, D.; Carlson, K. E.; Katzenellenbogen, J. A.; Welch, M. J. *J. Med. Chem.* **2006**, *49*, 4737.

(36) Zhou, D.; Sharp, T. L.; Fettig, N. M.; Lee, H.; Lewis, J. S.; Katzenellenbogen, J. A.; Welch, M. J. *Nucl. Med. Biol.* **2008**, *35*, 655.

(37) Sanner, M. F. *J. Mol. Graphics Model.* **1999**, *17*, 57.

(38) R.S. Pearlman, "Concord" distributed by Tripos International, St. Louis, Missouri, 63144, USA.

CHAPTER 5

PROBING THE PRESENCE AND IMPORTANCE OF A PUTATIVE
SECOND LIGAND BINDING VOLUME IN THE ESTROGEN RECEPTOR

I. Introduction

Long term interests in the development of synthetic non-steroidal ligands for ERs have led to the development of well-establish pharmacophores, as discussed in Chapter 1. These pharmacophores, however, are built upon relatively similar classes of heterocyclic cores, namely, 5-membered and 6,5-fused bicyclic heterocycles. This dearth of novel core structures represents a bottleneck to expanding overall ligand diversity, and impedes the development of SERMs with novel biological properties.

Many crystal structures for both ERα and ERβ exist in the PDB, and the availability of commercial virtual compound libraries lends itself to *in silico* screening for potentially useful ER ligands containing fundamentally new scaffolds. A virtual screen of the Maybridge HitFinder™ library containing ~2000 compounds was conducted by our group identified nineteen initial structures of interest. These compounds were subsequently purchased from Maybridge and tested for *in vitro* binding affinity to ER. The highest affinities were observed for compounds BTB-10314 and S-08084, as given in Figure 5.1. These RBA data may appear low relative to E_2; however, the corresponding K_i values for S-08084, for example, are 71 nM and 161 nM for ERα and ERβ, respectively. Initial S-08084 and BTB-10314 binding affinities in the mid to high nanomolar range provided hope for their potential as effective ER ligands, particularly for the development of sub-type selective ligands.

| | RBA | | RBA |
| BTB-10314 | ERα: 1.126
ERβ: 1.392 | S-08084 | ERα: 0.283
ERβ: 0.309 |
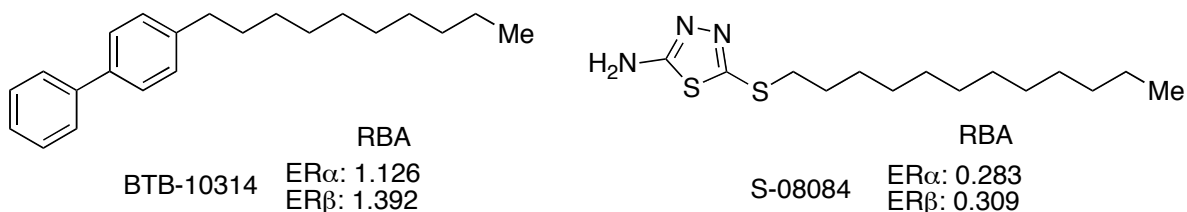
Figure 5.1. BTB-10314 and S-08084 were identified as potential SERMs from an *in silico* screening of the Maybridge HitFinder™ library consisting of ~2000 compounds. These compounds are of particular interested based on their novel structures and the presence of the extended alkyl sidechain, which may have implications as to the existence, importance, and function of a putative second binding pocket in ER.

The most intriguing property of these compounds, however, was the extended non-polar alkyl chain. Docking of S-08084 analog **1d** in ERβ suggested that the ligand might bind with the tetradecyl chain extending into a solvent channel (Figure 5.2) previously described as a possible second binding site within both ER subtypes. The presence of a second binding volume was proposed by Klinge[1] in response to fluorescence data demonstrating the non-competitive
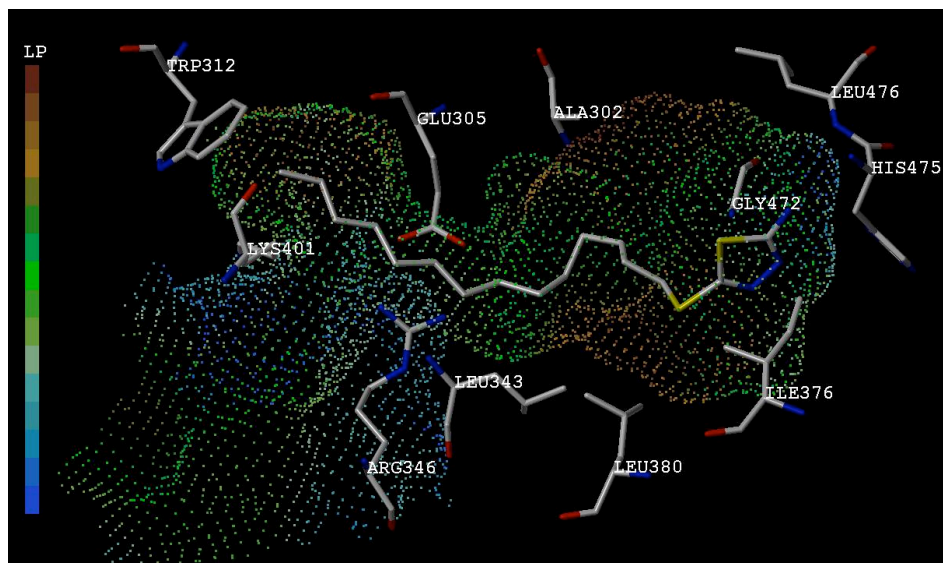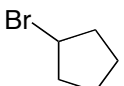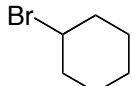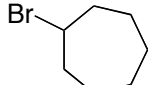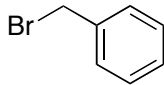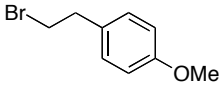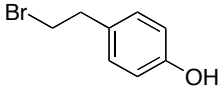
Figure 5.2. 5-Tetradecylsulfanyl-1,3,4-thiadiazol-2-ylamine (**1d**) docked into ERβ crystal structure using Sybyl (performed by John Comninos). The thiadiazole core binds in the ligand binding pocket while the tetradecyl alkyl sidechain protrudes through the channel created by Glu305 and Leu343 into the proposed second binding pocket.

binding of tetrahydrochrysene ketone (THCK) in the presence of E2. A subsequent modeling study by van Hoorn[2] identified the solvent channel described above as the potential binding volume; however, concluded that this site was too far from the ligand binding site to affect the conformation of helix 12, and therefore, could not be used to influence agonist versus antagonist behavior in either ER subtype. A similar binding mode has also been observed in related nuclear hormone receptors such as the glucocorticoid and vitamin D receptors. A recent crystal structure of deacylcortivazol (DAC) complexed with GR[3] highlights the flexibility of the receptor in this direction, expanding the pocket volume to nearly double the original size, and leading to the design of new non-steroidal high affinity GR ligands.[4] Accordingly, we set out to further investigate analogs of BTB-10314 and S-08084, as chemical probes for the existence, importance, and function of a putative second binding pocket within both ER subtypes. Analogues of BTB-10314 were pursued by John Comninos in our lab, and S-08084 are described herein.

II. Chemistry

Analogues of S-08084 were synthesized from 5-amino-1,3,4-thiadiazole-2-thiol and a series of alkyl halides in a suspension of NaH (60% wt. dispersion in mineral oil) in anhydrous DMF (Table 5.1). These reactions were initially performed according to method A, in which the UV-active thiadiazole represented the limiting reagent in order to facilitate reaction monitoring, and the desired product was isolated from remaining mineral oil via trituration from hexanes.

Table 5.1. S-Alkylation of 5-amino-1,3,4-thiadiazole-2-thiol using various halides
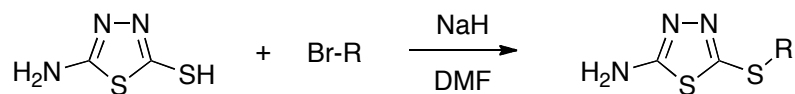
Compound	Halide	Method[1]	Rxn Time (hrs)	Yield
1a	$Br-(CH_2)_7CH_3$	a	1	22%
1b	$Br-(CH_2)_9CH_3$	a	1	16%
1c	$Br-(CH_2)_{11}CH_3$	a b	<1 2	31% 35%
1d	$Br-(CH_2)_{13}CH_3$	a b	<1 13	12% 77%
2a		b	4	74%[2]
2b		b	4.5	22%
2c		b	16	41%
3		b	3	68%
4a		b	5	54%
4b		b	3	21%

[1] a = alkyl halide used in excess, product isolated from mineral oil by trituration from hexanes; b = thiadiazole used in excess and removed during workup by KOH (aq) extraction, and the mineral oil was removed from NaH prior to reaction. [2] crude yield

Scheme 5.1. Preparation of 4-(2-bromoethyl)-phenol

5, 83%
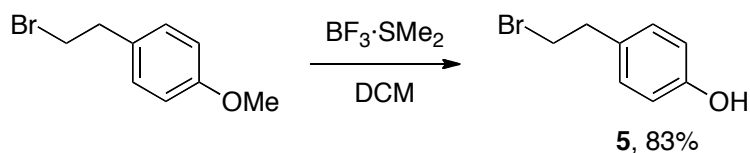
Under these conditions, extended chain alkyl halides generated emulsions during the aqueous workup, complicating the isolation and purification process. Optimization of the reaction conditions resulted in the development of method B, in which the NaH was purified from the mineral oil dispersion prior to the reaction, the thiadiazole was used in excess and removed during the workup by a KOH (aq) wash, and the reaction time was significantly extended in most cases due to difficulty in monitoring the reaction. Alkyl halide **5** was synthesized from commercially available 4-methoxyphenethyl bromide via cleavage of the aryl-methyl ether using the mild Lewis acid $BF_3 \cdot SMe_2$ in dichloromethane (DCM).

III. *in vitro* Biological Data

The *in vitro* binding affinity was determined by radiometric competitive binding assay relative to estradiol (E2) for products **1-4**. All compounds tested yielded RBA measurements at or below the sensitivity of the assay (RBA ~0.007). Subsequently, new samples of S-08084 and BTB-10314 were purchased from Maybridge and retested for binding affinity; both failed to bind significantly above the sensitivity of the assay, in contrast to initial data (Figure 5.1).

IV. Results and Discussion

In order to probe the importance and variability of the alkyl side chain and volume constraints of a second binding pocket, we prepared S-08084 along with a series of analogues bearing linear alkyl and cycloalkyl substitutents. The 5-amino-1,3,4-thiadiazole-2-thiol core has recently been substituted via S-alkylation using a biphasic system in which the desired product precipitated from water (the reaction solvent) under basic conditions, requiring only filtration and purification by recrystallization.[5] These reaction conditions, however, failed to yield product when alkylating with extended chains, presumably due to slow kinetics inherent in the minimal solubility of the alkyl halide at the solvent interface. The reaction conditions were then altered to a single phase reaction using DMF as a solvent, affording the desired products over a range of yields (Table 5.2, **1-4**).

Optimization of this reaction scheme was difficult primarily due to the inability to monitor the presence of alkyl halide by TLC with any significant sensitivity. Furthermore, as the alkyl chain lengths increased for an emulsion observed during the workup and correlated to reduced yields, most likely due to reduced solubility of longer alkyl chains in the reaction and workup solvents (DMF and EtOAc, respectively). Subsequent alteration of reaction conditions and workup, as denoted by the alkylation method in Table 2, eliminated the appearance of the emulsion and increased yield considerably as illustrated by **1d**.

With the exception of entry **1d**, the yields are lower than would be expected for a simple nucleophilic displacement of a halide. An increased yield with the overnight reaction time

suggests surprisingly slow kinetics and further research into the 5-amino-1,3,4-thiadiazole-2-thiol core clarified this situation. X-ray crystal,[6] FT-IR, and FT-RAMAN spectroscopy[7] studies have demonstrated that the predominant tautomer present at ambient temperatures is the thione form (Figure 5.3a). From the X-ray crystal structure, Downie estimated the C-SH bond at approximately 75% double bond character (Figure 5.3b); therefore, resonance stabilization of the negative charge through the heteroaromatic ring strongly decreases the nucleophilicity of the sulfur atom.   Evidence of this is also observed in the NMR of the 5-amino-1,3,4-thiadiazole-2-thiol starting material. As an example of a thiol attached to an aromatic system, the chemical shift of the thiol proton of benzenethiol is predicted at δ 2-4.[8] In contrast, the thiol peak of the 5-amino-1,3,4-thiadiazole-2-thiol is observed at δ 13.15 in DMSO, indicating a strongly deshielded proton.
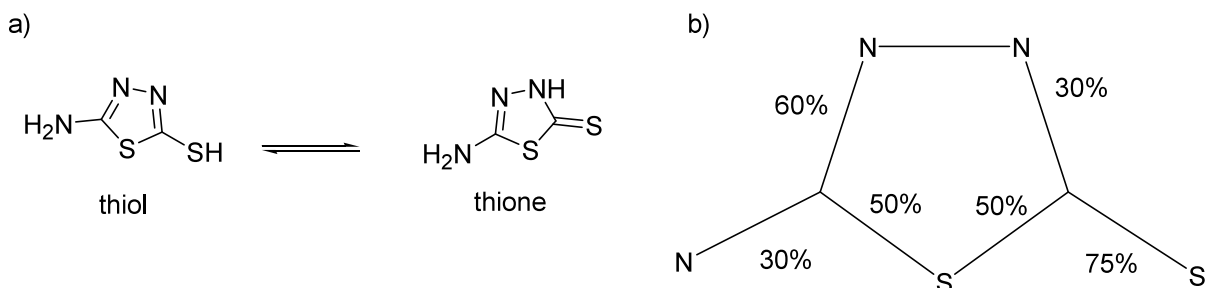
Figure 5.3. a) Tautomerization between the thiol and thione form; the thione form is predominate at ambient temperatures. b) Double bond character (%) as approximated from spectroscopic studies.[refs]

We subsequently selected 4-hydroxy phenethyl and benzyl substituents in an effort to improve binding affinity by mimicking the A-ring and 17β hydroxyl substituents of E2. Previous experience with these types of systems and protection schemes has demonstrated product decomposition due to elimination of the core as a leaving group from the benzyl carbon upon deprotection (Figure 5.4a).   In an attempt to circumvent this, we deprotected the 4-hydroxyl substituent prior to alkylating the thiadiazole core.   The deprotection of 4-methoxybenzyl bromide using $BF_3 \cdot SMe_2$ appeared complete by TLC; however, the product underwent decomposition during isolation, presumably due to the elimination of bromide ion and formation of quinone methide (Figure 5.4b) among other decomposition products.   In a subsequent attempt, the deprotection reaction was quenched and pre-formed thiadiazole anion was injected directly.   NMR analysis of the crude product suggested that the desired product was formed; however, attempts at isolation failed.   In order to isolate the 4-hydroxylbenzyl system as the cause of the observed decomposition, the 5-amino-1,3,4-thiadiazole-2-thiol core was alkylated with benzyl bromide (Table 5.1, **3**) and 4-methoxyphenethyl bromide (Table 5.1, **4a**); both products were stable and isolable. The presence of an additional carbon linking the thiadiazole

core and the 4-hydroxylphenyl substituent prevented the aforementioned quinone-type decomposition pathway. The deprotection of 4-methoxyphenethyl bromide yielded the desired phenol isolated in 90% yield. Subsequent S-alkylation of the thiadiazole core afforded the desired product **4b**.
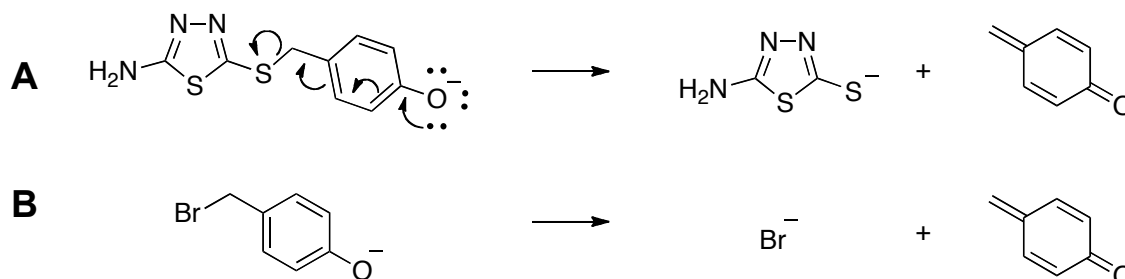
Figure 5.4. **A**) Proposed decomposition mechanism for the elimination of the thiadiazole core from the benzyl carbon upon deprotection of the 4-methoxy substituent. **B**) The analogous mechanism is proposed to explain the observed decomposition of 4-methoxybenzyl bromide upon isolation.

With a number of thiol alkylated products in hand, the next option was N-substitution, which would reverse the orientation of the core within the pocket, allowing hydrogen bonding from 2-SH versus 5-NH$_2$. Two methods for N-alkylation were attempted: imine formation followed by reductive amination, and amide formation followed by reduction to the amine. Despite a variety of aldehyde substrates and reaction conditions, the desired imine was never formed. Although surprising, these data are attributed to reduced nucleophilicity of the amine due to the presence of less reactive thiadiazole tautomers. Similar spectroscopic evidence places the C-NH$_2$ bond at 30% double bond character (Figure 5.3b).[6] After substantial attempts at refining reaction conditions, this scheme was eventually abandoned in favor of amide formation.

Examples of similar amides formed using the 5-amine-1,3,4-thiadiazole-2-thiol core existed in the literature.[9,10] Due to the low RBA values for all compounds synthesized by this point, we only chose one target, 5-(4-hydroxybenzyl)amino-1,3,4-thiadiazole-2-thiol, from *p*-anisoyl chloride (Scheme 5.2). The desired intermediate appears to have been obtained by

Scheme 5.2. Retrosynthetic analysis of 5-(4-hydroxybenzyl)amino-1,3,4-thiadiazole-2-thiol **5**.
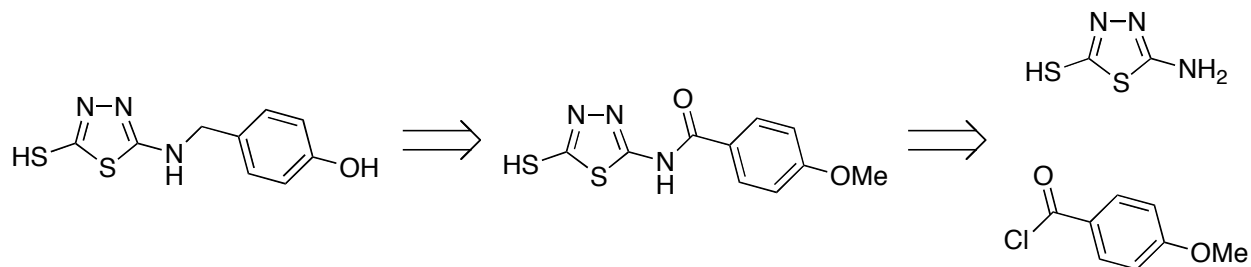
reacting with Et₃N in refluxing THF; however, it could not be isolated and purified despite several attempts. These failures were due to low solubility in solvents utilized for flash chromatography and recrystallization, and product decomposition prior to successful isolation. Optimization of the reaction conditions, workup, isolation, and purification were never fully explored as all compounds tested yielded RBA values below the sensitivity of the assay.

V. Conclusions

Computational screening of Maybridge's HitFinder™ virtual library yielded two compounds, S-08084 and BTB-10314 that appeared to have potential for the development as SERMs backed by favorable docking and initial binding affinities relative to estradiol (Figure 5.1). S-alkylation of the 2-thiol required significant optimization due to the hydrophobicity of the alkylating agents and the reduced nucleophilicity of the thiol. Furthermore, substitution of the amine via imine formation was never achieved using a variety of reaction conditions. Alternate N-substitution schemes via amide formation showed promise, but were not followed to completion due to poor binding data for all products synthesized. New samples of the original compounds were purchased from Maybridge in larger quantities and re-tested for binding affinity; both failed to bind significantly above the sensitivity of the binding assay, implying contamination or misidentification of the original samples.   As a result, 5-amino-1,3,4-thiadiazole-2-thiol is no longer of immediate interest as a scaffold in the development of SERMs.

VI. Experimental

General Considerations.   All chemical reagents were purchased from commercial suppliers without further purification.   Anhydrous DMF was obtained from Aldrich in a Sureseal™ bottle stored in a secondary container with desiccant; all other anhydrous solvents were obtained from a solvent dispensing system unless otherwise stated. All glassware was oven or flame-dried and cooled under vacuum or in a dry box. All reactions were conducted under argon. NMR spectra were obtained on Varian Oxford instruments and worked up using ACD, Inc. 1D-NMR processing software.   The chemical shifts are reported in ppm and referenced to the solvent peak. Electron Impact (EI) mass spectra were obtained on a 70-VSE mass spectrometer with an ionization energy of 70 eV.   Electrospray Ionization (ESI) mass spectra were obtained on a Q-Tof mass spectrometer.   Melting point ranges were measured using a Thomas Hoover capillary melting point apparatus.

General Procedure for S-Alkylation. *Method A*. 2-amino-1,3,4-thiadiazole-2-thiol (266 mg, 2.0 mmol) was dissolved in anhydrous DMF (3 mL). NaH (60% in mineral oil, 120 mg, 3.0 mmol) was added in a glove bag filled with N₂. The corresponding alkyl halide (2.4 mmol) was

added with stirring and heated to 60 °C under Ar until the reaction was complete by TLC (3:1 hexanes:EtOAc), usually within one hour. The reaction products were isolated using a standard workup consisting of (a) a quench in water, (b) exhaustive extraction with ethyl acetate, (c) drying over magnesium sulfate, (d) vacuum filtration, (e) re-suspension in hexanes, (f) a second vacuum filtration, and (g) concentration by rotary evaporation. The products were purified by recrystallization (EtOAc). *Method B*. NaH (60% in mineral oil, 80.0 mg, 2.0 mmol) was added to a 3-necked 25 mL round bottomed flask and dissolved in anhydrous THF (2.5-5.0 mL). While the solution was stirring under Ar the THF and mineral oil mixture was removed using a pipet with a cotton tip held under vacuum. The remaining solid was dissolved in 2 mL anhydrous DMF. 2-amino-1,3,4-thiadiazole-2-thiol (2.0 mmol, 266.4 mg) and the corresponding alkyl halide (1.8 mmol) were added subsequently; the reaction mixture was heated to 60 °C and stirred under Ar until done by TLC. The reaction products were isolated using a standard workup consisting of (a) a quench in 1 M KOH, (b) exhaustive extraction with EtOAc, (c) drying over MgSO$_4$, (d) vacuum filtration, and (e) concentration by rotary evaporation. The products were purified by recrystallization (EtOAc).

5-octylsulfanyl-1,3,4-thiadiazol-2-ylamine (1a). Following the general procedure for alkylation according to method A, **1a** was isolated as a beige solid (220.1 mg, 22.2%). ^{1}H-NMR (DMSO-d$_6$, 400 MHz) δ 7.26 (s, 2H), 3.00 (t, J = 7.29 Hz, 2H), 1.93 (s, 2H), 1.59 (m, 2H), 1.32 (m, 2H), 1.20 (m, 8H), 0.82 (t, J = 7.0 Hz, 3H); ^{13}C-NMR (DMSO-d$_6$, 100 MHz) δ 170.10, 150.90, 34.97, 31.90, 29.74, 29.28, 29.15, 28.56, 22.77, 14.56. HRMS (EI): calc'd for C$_{10}$H$_{19}$N$_3$S$_2$ [M+H]$^+$ 245.1020, found 245.1020; mp 113-114°C.

5-decylsulfanyl-1,3,4-thiadiazol-2-ylamine (1b). Following the general procedure for alkylation according to method A, **1b** was isolated as a beige solid (86.1 mg, 15.7%). ^{1}H-NMR (DMSO-d$_6$, 500 MHz) δ 7.24 (s, 2H), 3.00 (t, J = 7.18 Hz, 2H), 1.59 (m, 2H), 1.32 (m, 14H), 0.82 (t, J = 6.75 MHz, 3H); ^{13}C-NMR (DMSO-d$_6$, 125 MHz) δ 170.09, 150.87, 34.95, 32.00, 29.74, 29.65, 29.63, 29.41, 29.21, 28.57, 22.80, 14.57. HRMS (EI): calc'd for C$_{12}$H$_{23}$N$_3$S$_2$ [M+H]$^+$ 273.1333, found 273.1340; mp 113-114°C.

5-dodecylsulfanyl-1,3,4-thiadiazol-2-ylamine (1c). Following the general procedure for alkylation according to method A (186.7 mg, 31.3%) and method B (391.7 mg, 36.2%), **1c** was isolated as a yellow solid. ^{1}H-NMR (DMF-d7, 500 MHz) δ 7.38 (s, 2H), 3.09 (t, J = 7.29, 2H), 1.69 (m, 2H), 1.41 (m, 2H), 1.27 (m, 18H), 0.87 (t, J = 6.86 Hz, 3H); ^{13}C-NMR (DMF-d7, 125 MHz) δ 170.14, 151.21, 34.85, 32.04, 29.80, 29.78, 29.77, 29.73, 29.67, 29.49, 29.24, 28.60,

22.76, 13.95.   HRMS (EI): calc'd for $C_{14}H_{27}N_3S_2$ [M+H]$^+$ 301.1646, found 301.1636; mp 114-116°C.

5-tetradecylsulfanyl-1,3,4-thiadiazol-2-ylamine (1d). Following the general procedure for alkylation according to method A (70.5 mg, 11.9%) and method B (906.3 mg, 76.5%), **1d** was isolated as yellow-white solid. ^{1}H-NMR (DMSO-d$_6$, 500 MHz) δ 7.26 (s, 2H), 3.01 (t, J = 7.29 Hz, 2H), 1.59 (m, 2H), 1.33 (m, 2H), 1.21 (m, 20H), 0.83 (t, J = 6.85 Hz, 3H); ^{13}C-NMR (DMSO-d$_6$, 125 MHz) δ 169.44, 150.23, 34.30, 31.32, 29.08,29.07, 29.06, 29.04, 28.98, 28.93, 28.75, 28.49, 27.87, 22.12,13.95.   HRMS (EI): calc'd for $C_{16}H_{31}N_3S_2$ [M+H]$^+$ 329.1960, found 329.1951; mp 114-117°C.

5-cyclopentylsulfanyl-1,3,4-thiadiazol-2-ylamine (2a).   Following the general procedure for alkylation according to method B, **2a** was isolated as a yellow-white fibrous solid (267.6 mg, 73.6% crude). ^{1}H-NMR (DMSO-d$_6$, 500 MHz) δ 7.32 (s, 2H), 3.68 (m, 1H), 2.01 (m, 2), 1.61 (m, 6H); ^{13}C-NMR (DMSO-d$_6$, 125 MHz) δ 170.63, 150.42, 48.21, 33.71, 24.91. HRMS (EI): calc'd for $C_7H_{11}N_3S_2$ [M+H]$^+$ 201.0394, found 201.0389; mp 173-174°C.

5-cyclohexylsulfanyl-1,3,4-thiadiazol-2-ylamine (2b).   Following the general procedure for alkylation according to method B, **2b** was isolated as a white fibrous solid (86.3 mg, 21.6%). ^{1}H-NMR (DMSO-d$_6$, 500 MHz) δ 7.33 (s, 2H), 1.93 (m, 2H), 1.67 (m, 2H), 1.52 (m, 1H), 1.28 (m, 6H); ^{13}C-NMR (DMSO-d$_6$, 125 MHz) δ 170.94, 148.94, 48.38, 33.50, 25.91, 25.67. HRMS (EI): calcd for $C_8H_{13}N_3S_2$ [M+H]$^+$ 215.0551, found 215.0544; mp 178-179°C.

5-cycloheptylsulfanyl-1,3,4-thiadiazol-2-ylamine (2c).   Following the general procedure for the alkylation according to method B and scaling up by two-fold, **2c** was isolated as yellow-white solid (346.8 mg, 41.1%). ^{1}H-NMR (DMSO-d$_6$, 500 MHz) δ 7.32 (s, 2H), 3.48 (ddd, J = 13.45, 8.95, 4.39, 1H), 1.98 (m, 2H), 1.53 (m, 10H); ^{13}C NMR (DMSO-d$_6$, 125 MHz) δ 170.83, 149.92, 50.26, 34.95, 28.31, 25.74. HRMS (EI): calc'd for $C_9H_{15}N_3S_2$ [M+H]$^+$ 229.0707, found 229.0700; mp 155-157°C.

5-benzylsulfanyl-1,3,4-thiadiazol-2-ylamine (3). Following the general procedure for S-alkylation according to method B, **3** was isolated as a yellow solid (306.3 mg, 68.0%). ^{1}H-NMR (DMSO-d$_6$, 500 MHz) δ 7.34-7.23 (m, 4H), 7.27 (s, 2H), 7.25 (tt, J = 6.91 Hz, 1.72 Hz), 4.28 (s, 2H); ^{13}C-NMR (DMSO-d$_6$, 125 MHz) δ 169.94, 149.55, 137.14, 129.02, 128.52, 127.48, 38.49. HRMS (EI): calcd for $C_9H_9N_3S_2$ [M+H]$^+$ 223.0238, found 223.0244; mp 157-159°C.

5-(4-methoxyphenethyl)sulfanyl-1,3,4-thiadiazol-2-ylamine (4a). Following the general procedure for S-alkylation according to method B, **4a** was isolated as yellow-white solid (288.0 mg, 54.3%). ^{1}H-NMR (DMSO-d$_6$, 500 MHz) δ 7.28 (s, 2H), 7.14 (m, 1H), 6.84 (m, 1H), 3.70 (s, 3), 3.24 (t, J = 7.50 Hz, 2H), 2.85 (t, J = 7.50, 2H); ^{13}C-NMR (DMSO-d$_6$, 125 MHz) δ 169.53, 157.93, 150.26, 131.46, 129.64, 113.81, 54.84, 35.84, 34.21. HRMS (EI): calc'd for C$_{11}$H$_{13}$N$_3$OS$_2$ [M+H]$^+$ 267.0500, found 267.0502; mp 144-147°C.

4-(2-bromoethyl)-phenol (5). 4-methoxyphenethyl bromide (430 mg, 2.0 mmol) was added to a 50 mL 3-necked round bottomed flask and dissolved in 10 mL DCM. After purging with Ar, BF$_3$·SMe$_2$ (0.850 mL, 8.0 mmol) was injected and the solution stirred at room temperature overnight (~12 h). The reaction was quenched with MeOH, and the solvent removed by a stream of N$_2$. The remaining red oil was dissolved in EtOAc, washed with water, dried over MgSO$_4$, vacuum filtered, and concentrated by rotary evaporation. The product was purified by flash chromatography over silica gel (50% EtOAc in hexanes), and isolated as a white powder (362.5 mg, 82.9%). ^{1}H-NMR (CDCl$_3$, 500 MHz) δ 7.07 (AA'XX', 2H), 6.78 (AA'XX', 2H), 4.77 (s, 1H), 3.51 (t, J=7.6 Hz, 2H), 3.08 (J=7.6 Hz, 2H); ^{13}C-NMR (CDCl$_3$, 125 MHz) δ 154.41, 131.29, 129.93, 115.46, 38.56, 33.42.

5-(4-hydroxyphenethyl)-sulfanyl-1,3,4-thiadiazol-2-ylamine (4b). Following the general procedure for S-alkylation according to method B, **4b** was isolated as white flakey solid (90.9 mg, 21.2%). ^{1}H-NMR (DMSO-d$_6$, 500 MHz) δ 9.22 (s, 1H), 7.27 (s, 2H), 7.00 (m, 2H), 6.66 (m, 2H), 3.21 (t, J = 7.5 Hz, 2H), 2.79 (t, J = 7.5 Hz, 2H); ^{13}C-NMR (DMSO-d$_6$, 125 MHz) δ 170.10, 156.57, 150.91, 130.32, 130.18, 115.80, 36.55, 34.94. HRMS (ESI): C$_{10}$H$_{11}$N$_3$OS$_2$ [M+H]$^+$ calc'd 254.0422, found 254.0408.

VII. References
(1) Tyulmenkov, V. V.; Klinge, C. M. *Arch. Biochem. Biophys.* **2000**, *381*, 135.
(2) van Hoorn, W. P. *J. Med. Chem.* **2002**, *45*, 584.
(3) Suino-Powell, K.; Xu, Y.; Zhang, C.; Tao, Y.-g.; Tolbert, W. D.; Simons, S. S., Jr.; Xu, H. E. *Mol. Cell. Biol.* **2008**, *28*, 1915.
(4) Madauss, K. P.; Bledsoe, R. K.; McLay, I.; Stewart, E. L.; Uings, I. J.; Weingarten, G.; Williams, S. P. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 6097.
(5) Clerici, F.; Pocar, D.; Guido, M.; Loche, A.; Perlini, V.; Brufani, M. *J. Med. Chem.* **2001**, *44*, 931.
(6) Downie, T. C.; Harrison, W.; Raper, E. S.; Hepworth, M. A. *Acta Crystallogr. Sect. B: Struct. Sci.* **1972**, *28*, 1584.
(7) Edwards, H. G. M.; Johnson, A. F.; Lawson, E. E. *J. Mol. Struct.* **1995**, *351*, 51.
(8) Pretsch, E.; Buehlmann, P.; Affolter, C. *Structure Determination of Organic Compounds: Tables of Spectral Data, 3rd Completely Revised and Enlarged Edition*, 2000.

(9)   Almajan, G. L.; Innocenti, A.; Puccetti, L.; Manole, G.; Barbuceanu, S.; Saramet, I.; Scozzafava, A.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2347.

(10) Cho, N. S.; Kim, G. N.; Parkanyi, C. *J. Heterocycl. Chem.* **1993**, *30*, 397.

I. Docking Code
A. Templates
i. Filetree
    /docking_out
        /date-labelled subfolders
            /compound basenames
                raw docking output (pdbqt/pdb files)
    /etc
        LOGs
    /ligands
        various collected ligand files; general storage
    /ligands2dock
        pdbqt/pdb for ligands to be docked
    /runfiles
        prepared receptor pdbqt files
        configuration file(s) (eg. conf.txt)
    /src
        DockWkup.svl
        docking.bash
        moe2pdb.svl
        prepare_ligands.bash
        prepare_moe_min.bash

ii. Sample AutoDock Vina Configuration File
    receptor = ../runfiles/2QGW_empty.pdbqt

    center_x = 14.836
    center_y = 3.854
    center_z = 33.806

    size_x = 20
    size_y = 20
    size_z = 20

    exhaustiveness = 100

B. Bash Scripts
i. prepare_ligands.bash

```
# Written by Christopher Mayne, 02/10/2010
# converts all pdb files in ../ligands2dock to pdbqt files; shouldn't do anything to pdbqt files

cd ../ligands2dock

for f in `ls *.pdb`
do
        pythonsh /library/MGLTools/1.5.4/MGLToolsPckgs/AutoDockTools/Utilities24/
prepare_ligand4.py -l $f
done

cd ../src
```

ii. Docking.bash

```bash
#!/bin/bash

# Written by Christopher Mayne
# Date: 02/10/2010
# A better control file for docking using autodock vina

# Change Log
# 06/29/2010 - fix vina command line parameters to place ligand out and log file in $b folder
# 06/29/2010 - add the echo line to report the basename for each ligand during docking
# 06/29/2010 - add a timestamp at beginning and end of docking


###############################################################################
#                                                                             #
# Requires a specific filetree format:                                        #
#    ./src contains docking.bash (this file)                                  #
#    ./runfiles contains receptor_file.pdbqt and conf.txt                     #
#    ./ligands2dock contains all pdbqt files                                  #
#    ./docking_out will hold the output files                                 #
#                                                                             #
# The docked output will be held in ./docking_out folder, organized by date   #
# and basename                                                                #
#                                                                             #
# The output will also be post-processed to convert the pdbqt files to pdb    #
#                                                                             #
###############################################################################

# Beginning Timestamp
echo 'Script Start: '`date`

# Set docking run variables
date_folder=`date +%Y.%m.%d`

# Start a loop that docks each ligand with a pdbqt extension in the ligands2dock
# subdirectory

for f in `ls ../ligands2dock | grep .pdbqt`
do

# Find and set basename
    b=`basename $f .pdbqt`

# Create the subdirectory in the ../docking_out/<date> folder to place all docking output
    mkdir -p ../docking_out/$date_folder/$b

# Run Vina with common parameters
    echo 'Currently Docking: ' $b
    vina --config ../runfiles/conf.txt --ligand ../ligands2dock/$f --ligand_out ../docking_out/$date_folder/$b/out_ --log ../docking_out/$date_folder/$b/log.txt

# Post-process the pdbqt files
    cd ../docking_out/$date_folder/$b
    for outfile in `ls *.pdbqt`
```

```
        do
                pythonsh /library/MGLTools/1.5.4/MGLToolsPckgs/AutoDockTools/Utilities24/
        pdbqt_to_pdb.py -f $outfile
                done
                cd ../../../src

        done

        # End Timestamp
        echo 'Script End: '`date`
```

iii. prepare_moe_min.bash

```
    !/bin/bash

        # Written by Christopher Mayne
        # Date: 02/10/2010
        # Post Processing of Docking files for MOE Processing
        # Passed: requires the docking directory passed as argument
        # Returns: writes file with full path to lowest energy pdb files

        # Change the CWD to the appropriate docking_out subfolder
        cd ../docking_out/$1
        mkdir moe_wkup

        # Loop through each sub folder
        for f in `find * ! -name moe* -type d`
        do
                cd $f

                # Find the lowest energy output file
                pdb_file=`ls *1.pdb | awk '/out_(01|1)\.pdb/'`

                # Set a variable to the full path of the lowest energy pdb file
                dir_path=`pwd`
                full_file_path=$dir_path\/$pdb_file

                # Reset CWD
                cd ..

                # Write the path to a txt file in specific docking_out subfolder
                echo $full_file_path >> ./moe_wkup/moe_min_inputfile

        done
```

C. Scientific Vector Language (SVL) Programs
i. moe2pdb.svl

```
    carbocycle:src cmayne$ cat moe2pdb.svl
    // Written by Christopher G. Mayne
    // Created: 09/18/2010
    // Description: Converts .moe to pdb
    // Input: save directory, moe files
    // Output: .pdb files

    // ---Change Log---
```

```
//

global function moe2pdb [savedir, moefiles]
     local mfile, basename, savename;

     for mfile in moefiles loop

          Open mfile;
          View Atoms[];
          basename = ftail (fbase mfile);
          savename = twrite ['{}/{}.pdb', savedir, basename];
          SaveAs savename;
          Close[force:1];

     endloop

endfunction
```

ii. DockWkup.svl
```
     // Written by Christopher G. Mayne
     // Created: 11/02/2009
     // Description: Multistep minimization of docking poses
     // Input: receptor structure, file containing path to poses
     // Output: merged and minimized ligand-protein complex file (_wkup.moe)

     // ---Change Log---
     // 06/28/2010 - change 'third' to 'last' when selecting the ligand.  This will generalize the
     // selection in the event that there are not crystallographic waters left in, which occupied
     // the second chain in the development files.  Since the ligand file is opened after the
     // receptor, it will always occupy the last chain.




     // Import Built-in Functions
     function PartialCharge, MM;

     //--------------Proximal Residue Search Subroutine----------------------------------------
     // 11/2/0/2009 taken from ProxResFind.svl and converted to local function

     local function ProxResFind [atom_set, sel_atoms, radius]
        if not length sel_atoms then return 0; endif;

        const PACKET = 50;
        local r = max [0, radius];
        local mask = zero atom_set;

        local x;
        for x in split [x_id atom_set, PACKET] loop
           local key = prox_open [r, aPos atom_set[x], r/2];
           local idx = uniq second prox_find [ key, aPos sel_atoms, r/2 ];
           prox_close key;
           mask = put [mask, x[idx], 1];
        endloop
```

```
        // set the found atoms subtracting the initially selected atoms
        local found_atoms = diff [ atom_set | mask, sel_atoms ];
        // extend the atom set to include atoms of residues that contain an atom within radius
        local ext_found_atoms = cat oChildren ( uniq oParent found_atoms);
        // return the extended found atoms that are also present in the full atom set
        return join [ atom_set, ext_found_atoms ];

endfunction

//-----------------------------------------------------------------------------


//----------Main Code Block----------

global function DockWkup [receptor_file, ligand_poses]
        local orig_atoms, pose, pocket_res, pocket_res_notBB, basename, savename;
        local [q,pos] = [];

        for pose in ligand_poses loop

                Open receptor_file;
                Open pose;
                View Atoms[];

                orig_atoms = Atoms[];

                Add_H Atoms[];
                [q,pos] = PartialCharge [ Atoms[], 'MMFF94*' ];
                aSetCharge [ Atoms[], q ];

                write 'File Opened, Hydrogens Added, Charges Computed\n';

                // Stage 1: minimize added hydrogens
                write '\nStage 1: Minimizing added hydrogens\n\n';
                Beep[];
                aSetFixed [ orig_atoms, 1 ];
                MM[gtest:0.0001];
                aSetFixed [ Atoms[], 0 ];

                // Stage 2: minimize ligand
                write '\nStage 2: Minimizing Ligand\n\n';
                Beep[];
                aSetFixed [ diff [ Atoms[], cat oAtoms last Chains[] ], 1 ];
                MM[gtest:0.0001];
                aSetFixed [ Atoms[], 0 ];

                // Stage 3: minimize residues in pocket
                write '\nStage 3: Minimizing Residues in Pocket\n\n';
                Beep[];
                pocket_res = ProxResFind [ Atoms[], cat oAtoms last Chains[], 4.5 ];
                pocket_res_notBB = diff [ pocket_res, aBackbone pocket_res ];
                aSetFixed [ diff [ Atoms[], pocket_res_notBB ], 1 ];
                MM[gtest:0.0001];
```

131

```
                aSetFixed [ Atoms[], 0 ];

                // Stage 4: minimize ligand and residues in pocket
                write '\nStage 4: Minimizing Ligand + Residues in Pocket\n\n';
                Beep[];
                aSetFixed [ diff [ Atoms[], append [ cat oAtoms last Chains[], pocket_res_notBB ] ],
        1 ];

                MM[gtest: 0.0001];
                aSetFixed [ Atoms[], 0 ];

                basename = ftail ( fpath pose );
                savename = twrite [ '{}_wkup.moe', basename ];
                SaveAs savename;
                Close[force:1];

            endloop

        write 'DockWkup Module Complete\n';
        endfunction
```

iii. pyrrole_int.svl

```
        // Written by Christopher Mayne 12/3/09
        // File: pyrrole_int.svl
        // Desc: rotates pyrrole through an angle and calculates total energy and E_int for system

        // Preprocessing: add hydrogens, calc charges
        // Passed: atom keys to atoms in dihedral (in appropriate order), output filename
        // Returns: null
        // Output: output file with calculated values

        // Filetree
        //              ./src
        //                      ./pyrrole_int.svl
        function Potential;

        global function pyrrole_int [toratoms, filename]

                local dihang, E_tot, E_int, ligand_atoms;
                local Edata = [];
                local out_handle = fopenw filename;

                fwrite [ out_handle, '{}\n', filename ];
                fwrite [ out_handle, 'dihang\tE_tot\tE_int\n' ];

                ligand_atoms = cat oAtoms third Chains[];

                for dihang = 0, 90, 0.5 loop

                        aSetDihedralCWDeg [ toratoms, dihang ];
                        Edata = tr [
                                pot_strEnergy ligand_atoms, pot_angEnergy ligand_atoms,
        pot_stbEnergy ligand_atoms,
                                pot_oopEnergy ligand_atoms, pot_torEnergy ligand_atoms,
        pot_vdwEnergy ligand_atoms,
```

```
                              pot_eleEnergy ligand_atoms, pot_solEnergy ligand_atoms,
    pot_resEnergy ligand_atoms
                              ];

                      E_tot = add Edata(1);
                      E_int = add Edata(3);

                      fwrite [ out_handle, '{}\t{}\t{}\n', dihang, E_tot, E_int ];

              endloop

              fclose out_handle;

    endfunction
```

iv. recligE_MMFF94x.svl

```
    // Written by Christopher G. Mayne, 11/14/2009
    // Calculates and writes out interaction and binding energies for a set
    // receptor-ligand complexes
    // Passed: filenames to full structures (moe)
    // Returns: null
    // Output: ./recligE.out

    // Filetree
    //          ./src
    //                  ./recligE.svl
    //          ./docking_out
    //                  ./2010.02.10/full_structures/*.moe

    #set title 'Complex Energies'
    #set class 'MOE:user'

    function MM, Potential, PartialCharge;

    //--------------Proximity Selection Subroutine--------------------------------------

    local function ProxFind [atom_set, sel_atoms, radius]
       if not length sel_atoms then return 0; endif;

       const PACKET = 50;
       local r = max [0, radius];
       local mask = zero atom_set;

       local x;
       for x in split [x_id atom_set, PACKET] loop
               local key = prox_open [r, aPos atom_set[x], r/2];
               local idx = uniq second prox_find [ key, aPos sel_atoms, r/2 ];
               prox_close key;
               mask = put [mask, x[idx], 1];
       endloop

               return ( atom_set | mask );
    endfunction
```

```
//---------------------------------------------------------------------------

//-------------Main Function Code--------------------------------------------------
global function recligE_MMFF94x complex_files

            // Passed: array with paths to each pdb file to open
            // Returns: null
            // Outputs: writes energy values for calculating E(interaction) and E(binding)

            // Declare and set some variables
            local E_RL, E_R, E_L, E_Rm, E_Lm, E_int, E_bin, dX, W;
            local [q, pos] = [];
            local lig_name, pdbfile;

            // Open the output file for writing energy data, and write header
            local out_handle = fopenw './recligE.out';
            fwrite [ out_handle, 'Data Output from recligE_MMFF94x.svl, date/time: {}\n',
asctime[] ];
            fwrite [ out_handle, 'ligand\tE_int\tE_bin\tE_RL\tE_R\tE_Rm\tE_L\tE_Lm\n',
lig_name, E_int, E_bin, E_RL, E_R, E_Rm, E_L, E_Lm ];

            // Start a loop to run over each pdb file
            for pdbfile in complex_files loop

                    // Preprocess pdb file
                    lig_name = ftail (fbase pdbfile);
                    Open pdbfile;
                    View Atoms[];
                    Add_H Atoms[];
                    [q, pos] = PartialCharge [ Atoms[], 'MMFF94*' ];
                    aSetCharge [Atoms[], q];

                    // Define the pocket
                    local ligand_atoms = cat oAtoms third Chains[];
                    local pocket_atoms = diff [ (ProxFind [ Atoms[], ligand_atoms, 4.5 ]),
ligand_atoms ];

                    // Calculate Total Complex Energy
                    [E_RL,dX,W] = Potential[];

                    // delete out ligand and calc receptor/solvent energy
                    oDestroy third Chains[];
                    [E_R,dX,W] = Potential[];
                    aSetFixed [ (diff [ Atoms[], pocket_atoms ] ), 1 ];
                    MM[gtest:0.01];
                    aSetFixed [ Atoms[], 0 ];
                    [E_Rm,dX,W] = Potential[];

                    // Close the altered pdb file
                    Close[force:1];

                    // Reopen and preprocess the pdb file
                    Open pdbfile;
                    View Atoms[];
```

```
                Add_H Atoms[];
                [q, pos] = PartialCharge [ Atoms[], 'MMFF94*' ];
                aSetCharge [Atoms[], q];

                // delete receptor/solvent and calc ligand energy
                oDestroy first Chains[];
                oDestroy first Chains[];
                [E_L,dX,W] = Potential[];
                MM[gtest:0.01];
                [E_Lm,dX,W] = Potential[];

                // Close pdb file
                Close[force:1];

                // Calc interaction and binding energies
                E_int = E_RL - ( E_R + E_L );
                E_bin = E_RL - ( E_Rm + E_Lm );

                // Write out data to file
                fwrite [ out_handle, '{}\t{}\t{}\t{}\t{}\t{}\t{}\t{}\n', lig_name, E_int, E_bin,
        E_RL, E_R, E_Rm, E_L, E_Lm ];

                endloop

        // Close output file
        fclose out_handle;

    endfunction
    //------------End Main Function Code-------------------------------------------------------

v. HBscoring.svl
    // Written by Christopher G. Mayne, 02/09/2009
    // Calculates hydrogen bonding scores for receptor-ligand complex
    // Passed: array containing the .moe files to score
    // Returns: null
    // Output: dock_wkup_metric_HBscore.csv file

    // Filetree
    //          ./src
    //                  ./HBscoring.svl
    //          ./docking_out
    //                  ./2010.02.10/full_structures/*.moe

    function Potential, PartialCharge, dock_HydrogenBonds;

    global function HBscoring moe_files
                local file, basename, ligand, protein, energy, overall;
                local [ q, pos ] = [];
                local HBscores = [];
                local out_handle = fopenw 'dock_wkup_metric_HBscores.csv';

                fwrite [ out_handle, 'Ligand\tHB1\tHB2\tHB3\tHBh2o\n' ];

                for file in moe_files loop
```

135

```
                    basename = ftail (fbase file);
                    Open file;
                    [ q, pos ] = PartialCharge [ Atoms[], 'MMFF94*' ];
                    aSetCharge [ Atoms[], q ];

                    ligand = cat oAtoms third Chains[];
                    protein = diff [ Atoms[], ligand ];

                    energy = first Potential[dX:0];
                    HBscores = third dock_HydrogenBonds [ protein, ligand ];
                    overall = energy * mul HBscores;

                    fwrite [ out_handle, '{}\t{}\t{}\t{}\t{}\n', basename, third HBscores, first
        HBscores, last HBscores, second HBscores];
                    Close[force:1];

              endloop

              fclose out_handle;

       endfunction
```

II. GAMESS Code
A. Templates
<u>i. Filetree</u>
    General Filetree
    /data
        /autogamessltd
            /compound basenames
                /theory or basis set
                    inp, dat, log files
    /etc
        general LOG file
        paffinity LOG file
        /autogamessltd_logs
            log files for each autogamessltd/AGltd_DFTB3LYP run
    /pdbs
        general storage for pdb files
    /pdbs2calc
        pdb files for compounds to be calculated
    /src
        AGltd_DFTB3LYP.bash
        autogamessltd.bash
        cpunum.pl
        getcoord.pl
        getmem.pl
        paffinity.bash
        parse_DFTB3LYP.bash
        parse_ltd.bash
    /templates
        /H+
            inp templates for geometry opt., MP2 and DFT single point energy
        /neutral

inp templates for geometry opt., MP2 and DFT single point energy

ii. Neutral Geometry Optimization inp File Header
```
! Minimization RHF/6-31G*)
$CONTRL SCFTYP=RHF RUNTYP=OPTIMIZE EXETYP=RUN
NZVAR=1 $END
$ZMAT DLC=.TRUE. AUTO=.TRUE. $END
$BASIS NGAUSS=6 GBASIS=N31 NDFUNC=1 $END
$STATPT OPTTOL=0.0005 NSTEP=100 HSSEND=.TRUE. $END
$SCF DIRSCF=.TRUE. $END
$SYSTEM MWORDS=60 $END
```

iii. Neutral MP2 Single Point Energy inp File Header
```
! Single Point Energy MP2/6-311+G**
$CONTRL SCFTYP=RHF MPLEVL=2 RUNTYP=ENERGY EXETYP=CHECK
NZVAR=1 $END
$ZMAT DLC=.TRUE. AUTO=.TRUE. $END
$BASIS NGAUSS=6 GBASIS=N311 NDFUNC=1 NPFUNC=1
DIFFS=.TRUE. $END
$STATPT OPTTOL=0.0005 NSTEP=50 $END
$SCF DIRSCF=.TRUE. $END
$SYSTEM PARALL=.TRUE. MWORDS=20 MEMDDI=0 $END

$DATA
cartesian coordinates from 6-31G* geometry optimization output
C1
```

iv. Neutral DFT Single Point Energy inp File Header
```
! Single Point Energy DFTB3LYP/6-311+G**
$CONTRL SCFTYP=RHF DFTTYP=B3LYP RUNTYP=ENERGY EXETYP=RUN
NZVAR=1 $END
$ZMAT DLC=.TRUE. AUTO=.TRUE. $END
$BASIS NGAUSS=6 GBASIS=N311 NDFUNC=1 NPFUNC=1
DIFFS=.TRUE. $END
$STATPT OPTTOL=0.0005 NSTEP=50 $END
$SCF DIRSCF=.TRUE. $END
$SYSTEM MWORDS=125 $END

$DATA
cartesian coordinates from 6-31G* geometry optimization output
C1
```

v. Protonated Geometry Optimization inp File Header
```
! Minimization RHF/6-31G*)
$CONTRL SCFTYP=RHF RUNTYP=OPTIMIZE EXETYP=RUN
NZVAR=1 ICHARG=1 $END
$ZMAT DLC=.TRUE. AUTO=.TRUE. $END
$BASIS NGAUSS=6 GBASIS=N31 NDFUNC=1 $END
$STATPT OPTTOL=0.0005 NSTEP=100 HSSEND=.TRUE. $END
$SCF DIRSCF=.TRUE. $END
$SYSTEM MWORDS=60 $END
```

vi. Protonated MP2 Single Point Energy inp File Header
```
! Single Point Energy MP2/6-311+G**
```

```
 $CONTRL SCFTYP=RHF MPLEVL=2 RUNTYP=ENERGY EXETYP=CHECK
 NZVAR=1 ICHARG=1 $END
 $ZMAT DLC=.TRUE. AUTO=.TRUE. $END
 $BASIS NGAUSS=6 GBASIS=N311 NDFUNC=1 NPFUNC=1
 DIFFS=.TRUE. $END
 $STATPT OPTTOL=0.0005 NSTEP=50 $END
 $SCF DIRSCF=.TRUE. $END
 $SYSTEM PARALL=.TRUE. MWORDS=20 MEMDDI=0 $END

 $DATA
cartesian coordinates from 6-31G* geometry optimization output
 C1
```

<u>vii. Protonated DFT Single Point Energy inp File Header</u>
```
Single Point Energy DFTB3LYP/6-311+G**
 $CONTRL SCFTYP=RHF DFTTYP=B3LYP RUNTYP=ENERGY EXETYP=RUN
 NZVAR=1 ICHARG=1 $END
 $ZMAT DLC=.TRUE. AUTO=.TRUE. $END
 $BASIS NGAUSS=6 GBASIS=N311 NDFUNC=1 NPFUNC=1
 DIFFS=.TRUE. $END
 $STATPT OPTTOL=0.0005 NSTEP=50 $END
 $SCF DIRSCF=.TRUE. $END
 $SYSTEM MWORDS=125 $END

 $DATA
cartesian coordinates from 6-31G* geometry optimization output
 C1
```

B. Bash Scripts
<u>i. autogamessltd.bash</u>
```bash
#!/bin/bash
# Written by Christopher Mayne, 8/26/2010
# Filename: autogamessltd.bash
# Desc: Based on go2.bash (original autogamess), this file launches GAMESS to calculate
#      1) optimized geometry using 6-31G* basis set
#      2) single point energy calculation using MP2/6-311+G** from the 6-31G* geometry

# CHANGE LOG

# CUSTOM FUNCTIONS
#---------------------------
# Termination Status Function
#---------------------------
# pass gamess log file as argument (e.g. termstatus 6-31Gdp-MP2.log)

function termstatus {
    status=`awk '/EXECUTION OF GAMESS TERMINATED NORMALLY/' $1`

if (test -n "$status")
then
    echo 'OPTIMIAZATION COMPLETE'
    # continue
else
    echo OPTIMIZATION FAILED!
```

```
            echo Moving on to next molecule
            # reset the CWD back to scr
            cd ../../../../src
            # break out of loop for this compound
            continue
    fi

}

#----------------
# PREAMBLE
#----------------

# set variable for GAMESS version number; required by rungms script
verno=Mar252010R2

# write LOG header
echo Running AUTOGAMESSLTD
echo Description:
echo '1) optimize geometry using 6-31G* basis set'
echo '2) single point energy calculation using MP2/6-311+G**'
echo
echo Script Started: `date`
echo Input Molecules:
echo `ls ../pdbs2calc/ | grep .pdb`
echo 'Total: ' `ls ../pdbs2calc/ | grep .pdb | wc -l`

#---------------
#---------------
# MAIN CODE BLOCK
#---------------
#---------------

# control loop
for f in `ls ../pdbs2calc/ | grep .pdb`
do

#--------
# PREPARATION PER PDB
#--------
# set basename, ID in LOG
b=`basename $f .pdb`
echo
echo
echo "Current Molecule: $b"
echo --------------------

# move to data folder, check for existing data
cd ../data/autogamessltd/
if (test -d $b)
then
        echo Previous data detected...deleting.
        rm -r $b
fi
```

```
# create subfolder in data folder; change cwd
mkdir -p ./$b
cd ./$b

# determine if protonated
if [[ "$b" =~ H\+$ ]]
then
      charge=H+
else
      charge=neutral
fi


#---------------------------
# 6-31G* (6-31Gd) CALCULATION
#---------------------------

echo
echo ENTERING 6-31G* SUBROUTINE

# create 6-31Gd subfolder
echo Create 6-31Gd subfolder
mkdir -p ./6-31Gd
cd ./6-31Gd

# convert pdb to gamin
echo Converting pdb to gamin
babel -ipdb ../../../../pdbs2calc/$b.pdb -ogamin ./$b.gamin

# burn the gamin header
echo Processing gamin
pdblines=`wc -l $b.gamin | awk '{print $1}'`
coordlines=`expr $pdblines - 2`
tail -$coordlines $b.gamin > coords.footer

# generate inp file
echo Generating inp file
cat ../../../../templates/$charge/Geo6-31Gd.header coords.footer > $b-Geo6-31Gd.inp

# run the 6-31Gd Geometry Optimization
echo Running GAMESS Calc...
rungms $b-Geo6-31Gd.inp $verno 16 >& $b-Geo6-31Gd.log
echo ...GAMESS calc DONE!

# check that calculation terminated normally
echo Check termination status:
termstatus $b-Geo6-31Gd.log

# clean up the directory
echo Cleaning up directory, resetting variables
rm $b.gamin coords.footer
unset pdblines coordlines
```

```
# move up filetree into pdb-specific subfolder (/data/autogamessltd/$b)
cd ..

#--------------------------------------------
# MP2/6-311+G** (= MP2/6-311+Gdp) CALCULATION
#--------------------------------------------

echo
echo ENTERING MP2/6-311+G** SUBROUTINE
echo Calculation from 6-31G* geometry

# make subfolder and move into it
echo Create MP2-6-311\+Gdp subfolder
mkdir ./MP2-6-311\+Gdp
cd ./MP2-6-311\+Gdp

# extract final coords from 6-31Gdp calculation
echo Extract coords from previous calc
perl ../../../../src/getcoord.pl ../6-31Gd/$b-Geo6-31Gd.log > ./coords.xyz

# generate the check inp file
echo Generate the check inp file
cat ../../../../templates/$charge/EnerMP2-6-311\+Gdp.header ./coords.xyz > $b-
EnerMP2-6-311\+Gdp-crash.inp
echo ' $END' >> $b-EnerMP2-6-311\+Gdp-crash.inp

# crash the first check
echo Intentionally crash GAMESS check...
rungms $b-EnerMP2-6-311\+Gdp-crash.inp $verno 1 >& crash.log
echo ...Intentionally crashed check complete

# extract OK MWORDS and MEMDDI value
echo Extracting memory values
memvals1=`awk '$1 ~ /CORES/ && $2 ~ /MEMDDI/ {getline; mwords=$4; memddi=$2} END
{printf ("MWORDS=%i MEMDDI=%i", mwords, memddi)}' crash.log`
echo Found memory values are: $memvals1

# make some changes the inp file
echo Fixing the inp file
sed -e "s/MWORDS=20 MEMDDI=0/$memvals1/" $b-EnerMP2-6-311\+Gdp-crash.inp > $b-
EnerMP2-6-311\+Gdp-check.inp

# trash the dat file, and rerun the calc
echo Resetting CWD and running GAMESS check...
rm *.dat
rungms $b-EnerMP2-6-311\+Gdp-check.inp $verno 1 >& check.log
echo ...GAMESS check complete

# determine max number of utilizable cores
ncpu=`perl ../../../../src/cpunum.pl ./check.log`
echo Maximum number of utilizable cores: $ncpu

# extract MWORDS and MEMDDI values
echo Extracting memory values
```

```
memvals2=`perl ../../../../src/getmem.pl ./check.log`
echo Found Memory Values are: $memvals2

# make some changes to the inp file
echo Generating RUN inp file
sed -e "s/EXETYP=CHECK/EXETYP=RUN/" -e "s/$memval1/$memvals2/" -e "s/
PARALL=.TRUE.//" $b-EnerMP2-6-311\+Gdp-check.inp > $b-EnerMP2-6-311\+Gdp.inp

# trash the dat file, and run the calc
echo Resetting CWD and Running GAMESS Calc...
rm *.dat
rungms $b-EnerMP2-6-311\+Gdp.inp $verno $ncpu >& $b-EnerMP2-6-311\+Gdp.log
echo ...GAMESS calc DONE!

# check that calculation terminated normally
echo Check termination status
termstatus $b-EnerMP2-6-311\+Gdp.log

# clean up directory
echo Cleaning directory, Resetting variables
rm *.xyz crash.log $b-EnerMP2-6-311\+Gdp-crash.inp $b-EnerMP2-6-311\+Gdp-check.inp
unset memvals1 memvals2

# tidy up some things before moving onto the next molecule
# move up filetree (back to src)
cd ../../../../src

# remove pdb from pdbs2calc
echo Removing molecule from pdbs2calc
rm ../pdbs2calc/$f

# rest for 10 minutes
sleep 600

done

echo Script Ending: `date`
```

ii. AGltd_DFTB3LYP.bash

```
#!/bin/bash
# Written by Christopher Mayne, 09/02/2010
# Filename: AGltd_DFTB3LYP.bash
# Desc: Based on autogamessltd.bash, this file
#     1) pull coordinates of optimized geometries (RHF/6-31G*)
#     2) single point energy calculation using DFT/B3LYP/6-311+G** from 6-31G* geometry


# CHANGE LOG

# 08/02/2010: NOTE: unlike the MP2 calcs, DFT calcs don't require MEMDDI resources to
#                be set.  This allows me to bypass all of the complicated crash/check/run steps
#                in the autogamessltd.bash sequence.  Hopefully by setting the MWORDS
#                value to 1GB (125 MWORDS) I shouldn't have any memory related problems.
#                This make my current organization a little strange, but functional.
```

```
# CUSTOM FUNCTIONS
#---------------------------
# Termination Status Function
#---------------------------
# pass gamess log file as argument (e.g. termstatus 6-31Gdp-MP2.log)

function termstatus {
     status=`awk '/EXECUTION OF GAMESS TERMINATED NORMALLY/' $1`

if (test -n "$status")
then
     echo 'OPTIMIZATION COMPLETE'
     # continue
else
     echo OPTIMIZATION FAILED!
     echo Moving on to next molecule
     # reset the CWD back to scr
     cd ../../../../src
     # break out of loop for this compound
     continue
fi

}


#----------------
# PREAMBLE
#----------------

# set variable for GAMESS version number; required by rungms script
verno=Mar252010R2

# write LOG header
echo Running AGLTD_DFTB3LYP
echo Description:
echo '1) single point energy calculation using MP2/6-311+G** from 6-31G* geomtery'
echo
echo Script Started: `date`
echo Input Molecules:
echo `ls ../pdbs2calc/ | grep .pdb`
echo 'Total: ' `ls ../pdbs2calc/ | grep .pdb | wc -l`


#---------------
#---------------
# MAIN CODE BLOCK
#---------------
#---------------

# control loop
for f in `ls ../pdbs2calc/ | grep .pdb`
do
```

```
#--------
# PREPARATION PER PDB
#--------
# set basename, ID in LOG
b=`basename $f .pdb`
echo
echo
echo "Current Molecule: $b"
echo --------------------

# move to data folder, check for existing cmpd and geometry data
cd ../data/autogamessltd
echo Looking for data dependencies...
if (test -d $b)
then
      echo ...Previous compound data FOUND
else
      echo ...Previous compound data NOT FOUND
      echo Moving on to next molecule
      # reset the CWD back to scr
      cd ../../src
      # break out of loop for this compound
      continue
fi

if (test -a $b/6-31Gd/$b-Geo6-31Gd.log)
then
      echo ...Previous geometry optimization log FOUND
else
      echo ...Previous geometry optimization log NOT FOUND
      echo Moving on to next molecule
      # reset the CWD back to scr
      cd ../../src
      # break out of loop for this compound
      continue
fi

echo Testing previous geometry optimization termination...
ptermstatus=`termstatus $b/6-31Gd/$b-Geo6-31Gd.log`
if [ "$ptermstatus" == "OPTIMIZATION COMPLETE" ]
then
      echo ...Previous geometry optimization log OK
elif [ "$ptermstatus" == "OPTIMIZATION FAILED!" ]
then
      echo ...Previous geometry optimization log FAILED
      echo Moving on to next molecule
      # reset the CWD back to scr
      cd ../../src
      # break out of loop for this compound
      continue
else
      echo ...Unidentified problem with previous geometry optimization log
      echo Moving on to next molecule
```

```
            # reset the CWD back to scr
            cd ../../src
            # break out of loop for this compound
            continue
fi

# remove any previous DFTB3LYP data
if (test -d ./$b/DFTB3LYP)
then
            echo Previous data detected...deleting
            rm -r ./$b/DFTB3LYP
else
            echo No previous DFTB3LYP data detected...continuing
fi

# determine if protonated
if [[ "$b" =~ H\+$ ]]
then
            charge=H+
else
            charge=neutral
fi
echo Detected charge state: $charge

#-------------------------------
# DFT B3LYP/6-311+G** CALCULATION
#-------------------------------

echo
echo 'ENTERING DFT B3LYP/6-311+G** SUBROUTINE'
echo 'Calculation from 6-31G* geometry'

# create the DFTB3LYP subfolder
mkdir -p ./$b/DFTB3LYP
cd ./$b/DFTB3LYP

# extract final coords from 6-31Gdp calculation
echo Extracting coords from 6-31Gdp geometry optimization log
perl ../../../src/getcoord.pl ../6-31Gd/$b-Geo6-31Gd.log > ./coords.xyz

# generate the check inp file
echo Generate the check inp file
cat ../../../templates/$charge/EnerDFTB3LYP-6-311\+Gdp.header ./coords.xyz > $b-
EnerDFTB3LYP-6-311\+Gdp.inp
echo ' $END' >> $b-EnerDFTB3LYP-6-311\+Gdp.inp

# run the DFT calc
echo Running GAMESS calc...
rungms $b-EnerDFTB3LYP-6-311\+Gdp.inp $verno 16 >& $b-EnerDFTB3LYP-6-311\
+Gdp.log
echo ...GAMESS calc done!

# check that calculation terminated normally
echo Check termination status
```

```
        termstatus $b-EnerDFTB3LYP-6-311\+Gdp.log

        # clean up directory
        echo Cleaning directory, Resetting variables
        rm *.xyz

        # tidy up some things before moving onto the next molecule
        # move up filetree (back to src)
        cd ../../../../src

        # remove pdb from pdbs2calc
        echo Removing molecule from pdbs2calc
        rm ../pdbs2calc/$f

        # rest for 5 minutes
        sleep 300

        done

        echo Script Ending: `date`
```

iii. parse_ltd.bash

```bash
        #!/bin/bash

        # move to the autogamessltd data folder
        cd ../data/autogamessltd

        # write output header
        printf "cmpd\tNeutral ZPE\tNeutral Total E\tNeutral Dipole\tH+ ZPE\t H+ Total E\tH+ Dipole
        \n"

        # start control loop
        for f in `ls -d */ | awk '/H\+/'`
        do
              b=`basename $f H+`
        #     echo $b

              # move into the neutral subdirectory, pull out ZPE from Geo, pull out Total E from
        Energy calc
              cd ./${b}
              cd ./6-31Gd
              zpeN=`awk '/[\-0123456789\.]+ KCAL\/MOL +[\-0123456789\.]+ KJ\/MOL/ {print $1}' $
        {b}\-Geo6\-31Gd.log`
              cd ../MP2-6-311\+Gdp
              tenergyN=`awk '/TOTAL ENERGY =/ {print $4}' ${b}\-EnerMP2\-6\-311\+Gdp.log`
              dipoleN=`awk '/ +DX +DY +DZ +/ {getline; print $4}' ${b}\-EnerMP2\-6\-311\+Gdp.log`

              # move into the H+ subdirectory, pull out ZPE from Geo, pull out Total E from Energy
        calc
              cd ../../${b}H\+
              cd ./6-31Gd
              zpeH=`awk '/[\-0123456789\.]+ KCAL\/MOL +[\-0123456789\.]+ KJ\/MOL/ {print $1}' $
        {b}H\+-Geo6\-31Gd.log`
              cd ../MP2-6-311\+Gdp
```

```bash
        tenergyH=`awk '/TOTAL ENERGY =/ {print $4}' ${b}H\+\-EnerMP2\-6\-311\+Gdp.log`
        dipoleH=`awk '/ +DX +DY +DZ +/ {getline; print $4}' ${b}H\+\-EnerMP2\-6\-311\
+Gdp.log`

        # reset the cwd
        cd ../../

        # convert total energy from hartrees to kcal/mol
        tenergyNkcal=`echo "$tenergyN * 627.53" | bc`
        tenergyHkcal=`echo "$tenergyH * 627.53" | bc`

        # output the parsed values
        printf "%s\t%f\t%f\t%f\t%f\t%f\n" $b $zpeN $tenergyNkcal $dipoleN $zpeH
$tenergyHkcal $dipoleH

    done
```

iv. parse_DFTB3LYP.bash

```bash
    #!/bin/bash
    # Written by Christopher Mayne, 09/02/2010
    # Filename: parse_DFTB3LYP.bash
    # Desc: pulls out DFT/B3LYP/6-311+G** Total Energy values

    # move into the data subfolder
    cd ../data/autogamessltd

    # write output header
    printf "cmpd\tNeutral Total E\tNeutral Dipole\tH+ Total E\tH+ Dipole\n"

    # start control loop
    for f in `ls -d */ | awk '/H\+/'`
    do
        # set the basename
        b=`basename $f H+`

        # move into neutral DFT directory
        cd ./$b/DFTB3LYP
        tenergyN=`awk '/TOTAL ENERGY =/ {print $4}' ${b}\-EnerDFTB3LYP\-6\-311\+Gdp.log`
        dipoleN=`awk '/ +DX +DY +DZ +/ {getline; print $4}' ${b}\-EnerDFTB3LYP\-6\-311\
+Gdp.log`


        # move into the H+ DFT directory
        cd ../../${b}H\+/DFTB3LYP
        tenergyH=`awk '/TOTAL ENERGY =/ {print $4}' ${b}H\+\-EnerDFTB3LYP\-6\-311\
+Gdp.log`
        dipoleH=`awk '/ +DX +DY +DZ +/ {getline; print $4}' ${b}H\+\-EnerDFTB3LYP\-6\-311\
+Gdp.log`

        # reset the cwd
        cd ../../

        # convert the total energy from hartrees to kcal/mol
        tenergyNkcal=`echo "$tenergyN * 627.53" | bc`
```

```
            tenergyHkcal=`echo "$tenergyH * 627.53" | bc`

            # output the parsed values
            printf "%s\t%f\t%f\t%f\t%f\n" $b $tenergyNkcal $dipoleN $tenergyHkcal $dipoleH

    done
```

<u>v. paffinity.bash</u>
```
    #!/bin/bash

    # move to the autogamessltd data folder
    cd ../data/autogamessltd

    # write output header
    printf "cmpd\tNeutral ZPE\tNeutral Total E\t H+ ZPE\t H+ Total E\n"

    # start control loop
    for f in `ls -d */ | awk '/H\+/'`
    do
        b=`basename $f H+`
#       echo $b

        # move into the neutral subdirectory, pull out ZPE from Geo, pull out Total E from
Energy calc
        cd ./${b}
        cd ./6-31Gd
        zpeN=`awk '/[\-0123456789\.]+ KCAL\/MOL +[\-0123456789\.]+ KJ\/MOL/ {print $1}' $
{b}\-Geo6\-31Gd.log`
        cd ../MP2-6-311\+Gdp
        tenergyN=`awk '/TOTAL ENERGY =/ {print $4}' ${b}\-EnerMP2\-6\-311\+Gdp.log`

        # move into the H+ subdirectory, pull out ZPE from Geo, pull out Total E from Energy
calc
        cd ../../${b}H\+
        cd ./6-31Gd
        zpeH=`awk '/[\-0123456789\.]+ KCAL\/MOL +[\-0123456789\.]+ KJ\/MOL/ {print $1}' $
{b}H\+-Geo6\-31Gd.log`
        cd ../MP2-6-311\+Gdp
        tenergyH=`awk '/TOTAL ENERGY =/ {print $4}' ${b}H\+\-EnerMp2\-6\-311\+Gdp.log`

        # reset the cwd
        cd ../../

        # convert total energy from hartrees to kcal/mol
        tenergyNkcal=`echo "$tenergyN * 627.53" | bc`
        tenergyHkcal=`echo "$tenergyH * 627.53" | bc`

        # output the parsed values
        printf "%s\t%f\t%f\t%f\t%f\n" $b $zpeN $tenergyNkcal $zpeH $tenergyHkcal

    done
```

C. Perl Helper Programs
<u>i cpunum.pl</u>

```perl
#!/usr/bin/env perl

# Read from stdin (or perl will read from a filename given on the command line)
while(<>)
{
    # burn line unless this pattern matches
    next unless /DISTRIBUTED\s+--REPLICATED--/;

    # read some more lines from stdin
    while(<>)
    {
        # search by regex; anything in () is stored in variable
        /^\s+([0-9]+)\s+([0-9]+)\s+[0-9]+.\s+([0-9]+)/;

        $ncpus=$1;

        if ($1==16) {
            print "16";
            last;
        }

        if (/^\s*$/) {
        print $ncpus;
            last;
        }

    }
}
```

ii. getmem.pl

```perl
#!/usr/bin/env perl

# Read from stdin (or perl will read from a filename given on the command line)
while(<>)
{
    # burn line unless this pattern matches
    next unless /DISTRIBUTED\s+--REPLICATED--/;

    # read some more lines from stdin
    while(<>)
    {
        # search by regex; anything in () is stored in variable
        /^\s+([0-9]+)\s+([0-9]+)\s+[0-9]+.\s+([0-9]+)/;

        $mwords=$3;
        $memddi=$2;

        if ($1==16) {
            print "MWORDS=$mwords MEMDDI=$memddi";
            last;
        }

        if (/^\s*$/) {
        print "MWORDS=$mwords MEMDDI=$memddi";
```

```
                last;
            }

        }
    }
```

iii. getcoord.pl

```perl
#!/usr/bin/env perl

# Read from stdin (or perl will read from a filename given on the command line)
while(<>)
{
    # burn line unless this pattern matches
    next unless /EQUILIBRIUM GEOMETRY LOCATED/;

    # read some more lines from stdin
    while(<>)
    {
        # Example of using parens to capture matched regex sections
        #/^\s([A-Z]+)\s+([0-9.]+)/;
        #print "Atom $1, charge $2\n";

        # Print line to stdout if this pattern matches
        print if /^\s[A-Z]+\s+[0-9.]+/;

        # Break from this inner while if the line is blank
        last if /^\s*$/;
    }
}
```