



Pharmacophore modeling strategies for the development of novel nonsteroidal inhibitors of human aromatase (CYP19)

Yagmur Muftuoglu^a, Gabriela Mustata^{b,*}

^a Departments of Biophysics and Chemical Biomolecular Engineering, Johns Hopkins University, Baltimore, MD 21218, United States

^b Department of Computational Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15213, United States

ARTICLE INFO

Article history:

Received 2 November 2009

Revised 27 March 2010

Accepted 31 March 2010

Available online 8 April 2010

Keywords:

Aromatase

CYP19

Pharmacophore model

Structure-based drug design

Ligand-based drug design

ABSTRACT

The present study utilizes for the first time the structural information of aromatase, an important pharmacological target in anti-breast cancer therapy, to extract the pharmacophoric features important for interactions between the enzyme and its substrate, androstenedione. A ligand-based pharmacophore model developed from the most comprehensive list of nonsteroidal aromatase inhibitors (AIs) is described and explained, as well. This study demonstrates that the ligand-based pharmacophore model contributes to efficacy while the structure-based model contributes to specificity. It is also shown that a 'merged' model (i.e., a merged structure-based and ligand-based model) can successfully identify known AIs and differentiate between active and inactive inhibitors. Therefore, this model can be effectively used to identify the next generation of highly specific and less toxic aromatase inhibitors for breast cancer treatment.

© 2010 Elsevier Ltd. All rights reserved.

The present study utilizes for the first time the structural information of aromatase, an important pharmacological target in anti-breast cancer therapy, to extract the pharmacophoric features important for interactions between the enzyme and its substrate, androstenedione. A ligand-based pharmacophore model developed from the most comprehensive list of nonsteroidal aromatase inhibitors (AIs) is described and explained, as well. This study demonstrates that the ligand-based pharmacophore model contributes to efficacy while the structure-based model contributes to specificity. It is also shown that a 'merged' model (i.e., a merged structure-based and ligand-based model) can successfully identify known AIs and differentiate between active and inactive inhibitors. Therefore, this model can be effectively used to identify the next generation of highly specific and less toxic aromatase inhibitors for breast cancer treatment.

Excluding cancers of the skin, breast cancer is the most frequently diagnosed cancer in women¹ and ranks second as a cause of tumor-related death after lung cancer. Currently, it is predicted that one in eight American women will develop invasive breast cancer some time during her life. Approximately two-thirds of breast cancer tumors are hormone-dependent and require estrogens to grow.² One approach in treating hormone-dependent cancer involves interfering with endogenous hormone production. Aromatase, also known as estrogen synthase, has always been considered the most promising target for the endocrine treatment of

breast cancer³ because, through inhibition of the aromatase enzyme, estrogen production is decreased, and tumor growth is stopped or reduced.

Aromatase is a multienzymatic complex that is mostly expressed in the ovaries of premenopausal women, in the placenta of pregnant women, and, additionally, in peripheral adipose tissue, breast tissue, and the brain.⁴ The enzyme is overexpressed in or near breast cancer tissue and is responsible for local estrogen production and proliferation of breast cancer tumors.^{5,6} It is located in the endoplasmic reticulum of cells and is composed of a cytochrome P450 heme protein (CYP19), which carries out the aromatization reaction, and a NADPH-cytochrome P450 reductase, a flavoprotein required for the electron transfer from NADPH to the cytochrome P450 enzyme.^{7,8} Aromatase catalyzes the synthesis of estrogens via the aromatization of the A ring of androgen precursors, namely androstenedione and testosterone.

Considerable research efforts over the past decades have been devoted to the study of this enzyme and to the development of potent and selective agents able to interfere with its action. Several classes of steroidal and nonsteroidal aromatase inhibitors (AIs) have been developed,^{2,7,9–20} and, on the basis of their inhibition mechanism and chemical origin, these molecules are divided into two classes: steroidal (type I) and nonsteroidal (type II).^{2,6}

Steroidal AIs are derivatives or analogues of the preferred androgenic substrates and inhibit aromatase irreversibly. They can be further divided into competitive inhibitors and mechanism-based inhibitors. Competitive inhibitors bind non-covalently to aromatase in a manner similar to that of the natural substrate

* Corresponding author. Tel.: +1 412 648 7799; fax: +1 412 648 3163.

E-mail address: gmustata@pitt.edu (G. Mustata).

and block it from being enzymatically modified by aromatase. Mechanism-based inhibitors, also known as suicide inhibitors, bind to the enzyme and are converted into a reactive intermediate that covalently binds the enzyme and permanently inactivates it. This process often destabilizes the enzyme and also increases its rate of degradation by the intracellular proteasome.^{6,21}

Nonsteroidal inhibitors, on the other hand, bind entirely non-covalently and contain a heteroatom that coordinates with the iron atom of the heme group to block the active site and reversibly inhibit the enzyme.^{2,6} These types of AIs are further divided into categories based on the order in which they were discovered or synthesized: first-, second-, and third-generation AIs. Currently, the third-generation of triazole-derived AIs are approved as front-line therapy for early and even advanced cases of breast cancer in postmenopausal women.^{21,22}

Nevertheless, for both steroidal and nonsteroidal AIs, important side effects—ranging from mild to severe, short-term to long-term—have been suggested or reported. For example, steroidal AIs often give way to androgenic side effects where other related systems are disturbed due to lack of inhibitor specificity.²³ Also, prolonged estrogen deprivation can lead to bone loss, osteoporosis, reproductive problems, or even other types of cancers.^{21,24} Undoubtedly, for some women, the benefits of existent AIs outweigh the associated side effects. However, for many women, quality-of-life issues are serious enough to cause them to discontinue their use of the prescribed AIs. Therefore, more selective and less toxic CYP19 inhibitors are needed, especially since mutations in intratumoral aromatase, which can cause changes in its stability, efficiency, or sensitivity to different classes of AIs, can vary from patient to patient.⁶

Until fairly recently, no structural information on the enzyme was available, except several homology models^{25–29} that proved to be valuable for understanding the binding determinants of several classes of inhibitors.^{15,18,22,29–32} However, the low sequence identity between members of different P450 families has limited the success of these models in structure-based virtual screening. The recent determination of the crystal structure (X-ray) of aromatase complexed with androstenedione [PDB code: 3EQM]³³ reveals the molecular basis for the enzyme's androgenic specificity and unique catalytic mechanism.

Androgens, the preferred substrates of aromatase, are believed to enter the enzyme's catalytic cleft and active site through an access channel open to the outside.³³ The catalytic cleft of the aromatase enzyme encompasses a volume of approximately 400 Å³, which is considerably smaller than the volume of about 530 Å³ of the binding pockets in CYP3A4 and CYP2D6, the two CYP450 enzymes with highest sequence identities to human aromatase.³³ Due in part to this smaller volume and also to the unique locations of the catalytically important residues, the catalytic cleft of aromatase is very specific to its androgenic substrates.

Since there has been no recent report on developing inhibitors using the newly published aromatase structure [PDB code: 3EQM],³³ the present study provides a hypothetical picture of the primary structural and chemical features responsible for activity and is expected to provide useful knowledge for developing the next generation of inhibitors targeted to human aromatase.

Using a comprehensive database of nonsteroidal AIs and structural data on aromatase, two pharmacophore models—one ligand-based (LB) and the other structure-based (SB), respectively,—were developed. The LB pharmacophore model was generated using Molecular Operating Environment (MOE)³⁴ while the SB pharmacophore model was generated on LigandScout.^{35,36} These two models were then merged using MOE to generate a 'Merged' Model that combined the different strengths of each original model.

Over the past several years, numerous potential inhibitors of human aromatase have been tested for biological efficacy. To

create the LB pharmacophore model, a database of 56 active nonsteroidal aromatase inhibitors, presented in the literature as having been tested in human placental microsome assays, plus nine inactive nonsteroidal compounds, was compiled. Homologous half maximal inhibitory concentration (IC₅₀) values for each were also noted. This collection of 65 total nonsteroidal compounds was divided into three sets: a Training Set (Table 1) to create the model and two Test Sets—one Active (Table 2) and one Inactive (Table 3)—to validate it.

The Training Set was comprised of 20 of the most active yet structurally diverse nonsteroidal AIs. The Active Test Set differs from the Training Set in that it contains 36 AIs that are slightly less potent, yet just as structurally diverse. The Inactive Test Set contains nine potential inhibitors that have been determined to be inactive against aromatase. Molecules in the Training Set and in the two Test Sets were created using ChemSketch³⁷ and were compiled into databases using the Molecule Builder function in MOE.³⁴

The LB pharmacophore model, generated with MOE, was derived from the Training Set of 20 of the most potent yet structurally diverse nonsteroidal AIs known. This was done using the PCHD scheme in the Pharmacophore Elucidation function in MOE,³⁶ defining H-bond acceptors and donors features, as well as putative points from hydrogen bond donors and acceptors that are projected in the approximate direction of the hydrogen bond.³⁶ Different conformations of the molecules of the Training Set were taken into consideration using the Conformation Import function during the Pharmacophore Elucidation.³⁶

The resulting LB Model identified four pharmacophore features: two hydrophobic/aromatic (**Hyd|Aro**), one hydrogen-bond acceptor (**Acc**), and one hydrogen-bond acceptor projection (**Acc2**). The 3D and 2D representations of the LB Model are shown in Figures 1 and 2, respectively. In Figure 1, the solid spheres represent the four pharmacophore features identified by the LB Model. Figure 2 shows the dimensions of the model and the distances between the pharmacophoric features. Featuring coverage of 20 out of 20 of the best structurally diverse AIs, the LB Model was assigned an overlap score of 12.7354 out of a maximum of 20 and an accuracy of 100%.

Originally, attempts at generating the ligand-based (LB) pharmacophore model were made based on sets of AIs with activity levels that are lower and more comparable to biologically relevant potencies (data not shown). However, several problems with this method soon became apparent. Since, in general, less potent AIs tend to differ considerably in chemical features and structural characteristics, lower-potency ligand-based (LB) models based on these AIs could be made to encompass only a maximum of only 11 or 12 lower-potency AIs. With some lower-potency Training Sets, only a two-point LB model could be generated, if at all. On the other hand, the final LB model based on the most potent AIs allowed for more than 20 compounds to align well enough for at least four-point pharmacophore generation. It was also observed that, if fewer than 20 of the most potent compounds were used, a five-point model could be developed, as well (data not shown).

A three-dimensional (3D) pharmacophore model (i.e., structure-based pharmacophore model) of the CYP19 binding pocket was created with LigandScout^{35,36} using the X-ray structure deposited in the Protein Data Bank (PDB) [PDB code: 3EQM].³³ The model was based on interactions that define aromatase inhibition, such as hydrophobic interactions, hydrogen bonding, and electrostatic interactions.^{35,55,56} Features identified by the LigandScout software are those that take into consideration chemical functionality but not strict structural topology or definite functional groups. As a result, completely new potential pharmacophores can be identified through database screening. Moreover, to increase selectivity, the LigandScout model includes spatial information regarding areas inaccessible to any potential ligand, thus reflecting possible steric

Table 1

Training Set of 20 highly active and structurally diverse aromatase inhibitors

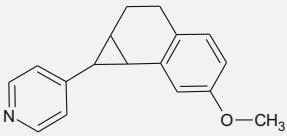
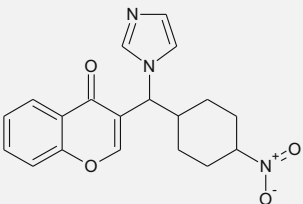
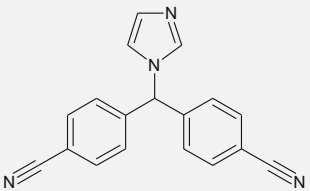
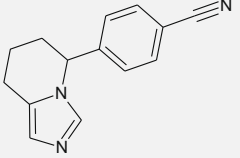
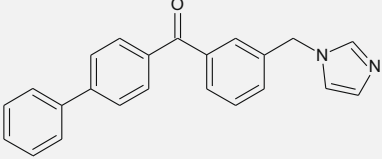
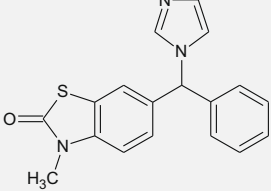
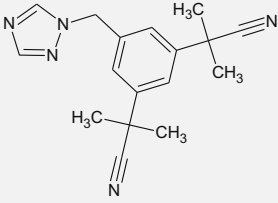
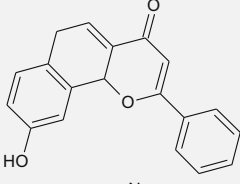
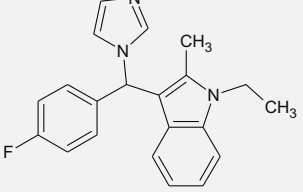
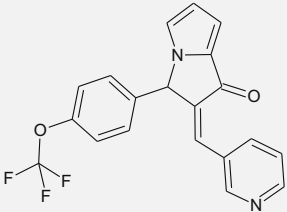
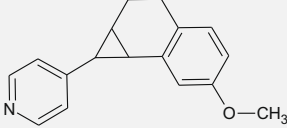
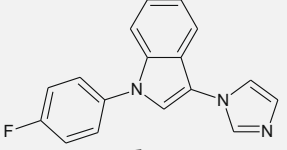
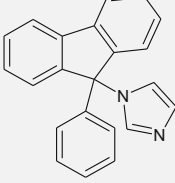
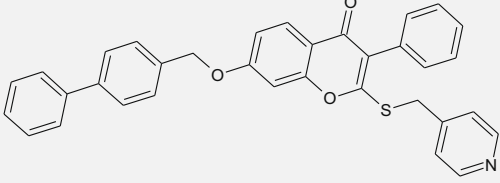
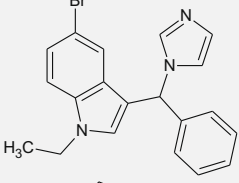
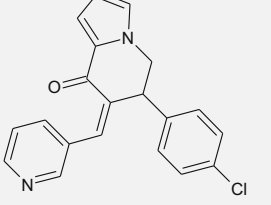
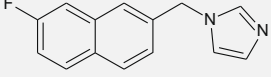
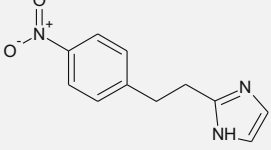
#	Compound name	Chemical structure	IC ₅₀ (nM)
1	Vorozole ³⁸		1.38
2	3-[Imidazol-1-yl-(4-nitro-phenyl)-methyl]-chromen-4-one ⁴⁰		2.3
3	CGS 18320B ⁴²		3
4	Fadrozole ⁴⁴		5
5	1d ⁴⁶		5.3
6	(R,S)-6-[1-(Imidazol-1-yl)-1phenylmethyl]-3-methyl-1,3-benzo-thiazol-2(3H)-one ⁴⁷		13
7	Anastrozole ¹⁴		15
8	9-Hydroxy-7,8-benzoflavone ⁵⁰		20
9	1-Ethyl-3-l(4-fluoro-phenyl)(1H-imidazol-1-yl)-methyl]-2-methyl-1H-indole ¹⁴		32

Table 1 (continued)

#	Compound name	Chemical structure	IC ₅₀ (nM)
10	22e ⁵³		40
11	4a ³⁹		61
12	1-(4-Fluoro-benzy1)-3-1(H-I-imidazolyl-methyl)-IH-indole ⁴¹		71.8
13	1-(9-Phenyl-9H-fluoren-9-yl)1H-imidazole ⁴³		74
14	7c ⁴⁵		85
15	5-Bromo-1-ethyl-3-[(1-H-I-imidazolyl) (phenyl)methyl]1H-indole ⁴¹		102
16	MR 20494 ⁴⁸		110
17	1-[(7-Fluoronaphth-2-yl)methyl]-1H-imidazole ⁴⁹		160
18	N-[2-(4'-Nitrophenyl) ethyl]-imidazole ⁵¹		160

(continued on next page)

Table 1 (continued)

#	Compound name	Chemical structure	IC ₅₀ (nM)
19	5-[(Imidazol-1-yl)-methyl]-5,6,7,8-tetrahydroquinoline ⁵²		160
20	4',5,7-OH-8-Prenylnaringenin ⁵⁴		200

Table 2

Active Test Set of 36 active and structurally diverse aromatase inhibitors

#	Compound name	Chemical structure	IC ₅₀ (nM)
1	Liarozole ⁵⁷		5
2	5-[(4-Chlorophen-yl)(1H-imidazol-1-yl)methyl]-1H-indole ⁵⁸		9
3	Letrozole ¹⁴		11.5
4	7,8-Benzoflavone ⁶¹		70
5	MR 20492 ⁶³		200
6	35075-39-7 ⁶⁵		200

Table 2 (continued)

#	Compound name	Chemical structure	IC ₅₀ (nM)
7	5j ⁴⁵		210
8	4f ⁵¹		310
9	Miconazole ⁶⁰		600
10	MR 16089 ⁴⁸		650
11	Apigenin ⁵⁹		900
12	TM-8 ⁵⁴		1.0
13	Chrysin ^{54,59}		1
14	Clotrimazole ⁶⁰		1.8
15	11β,13-Dihydro-10-epi-8-deoxycumam-brin B ⁶²		2

(continued on next page)

Table 2 (continued)

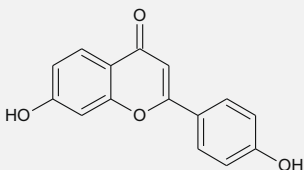
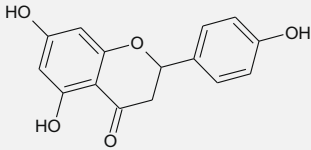
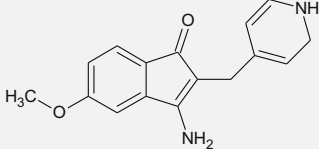
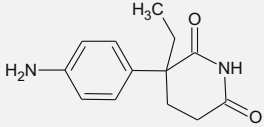
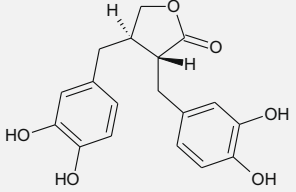
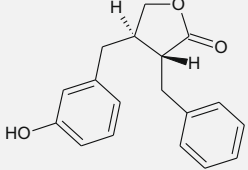
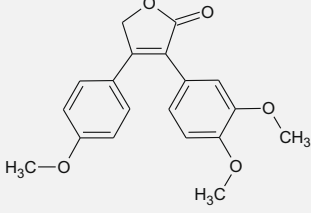
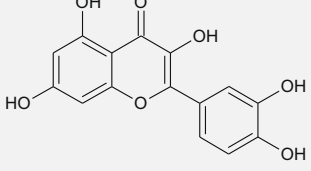
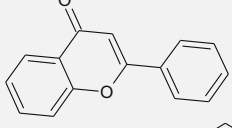
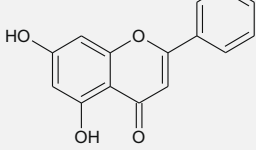
#	Compound name	Chemical structure	IC ₅₀ (nM)
16	7,4'-Dihydroxyflavone ⁶⁴		2.0
17	Naringenin ⁵⁴		2.9
18	MR 20814 ⁶⁶		3.5
19	Aminoglutethimide ⁶⁷		5.2
20	4,4'-Dihydroxy enterolactone ⁶⁸		6
21	Enterolactone ⁶⁸		6
22	TM-7 ⁵⁴		6.9
23	Quercetin ⁵⁰		12
24	Flavone ⁶¹		8.0
25	Flavanone ⁵⁹		8.7

Table 2 (continued)

#	Compound name	Chemical structure	IC ₅₀ (nM)
26	4c ⁶⁹		8.8
27	Nordihydroguaiaretic acid ⁶⁸		11
28	Dehydroleucodin ⁶²		15
29	TAN-931 ⁷⁰		17.2
30	Ludartin ⁶²		55
31	Ketoconazole ⁶⁰		60
32	10-Epi-8-deoxy-cumambrin B ⁶²		70
33	Biochanin A ⁵⁹		10.2
34	TM-9 ⁵⁴		13.7

(continued on next page)

Table 2 (continued)

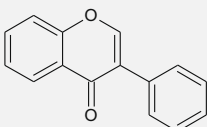
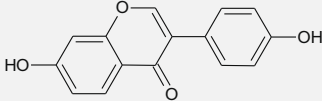
#	Compound name	Chemical structure	IC ₅₀ (nM)
35	2s ⁷⁰		17.2
36	7-Hydroxy-flavone ⁵⁹		30.5

Table 3

Inactive Test Set of nine inactive, yet structurally diverse aromatase inhibitors

#	Compound name	Chemical structure
1	3-Phenyl-7(phenyl methoxy)-2-[(4'-pyridylmethyl)thio]-4H-1-benzo-pyran-4-one ¹⁴	
2	9a ⁵⁷	
3	3',4'-Dihydroxy-flavone ⁶⁴	
4	5,6-Benzoflavone ⁵⁰	
5	3b ⁷¹	
6	12b ⁵⁷	
7	9c ⁵⁷	

Table 3 (continued)

#	Compound name	Chemical structure
8	Isoflavone ⁶⁴	
9	Daidzein ⁶⁴	

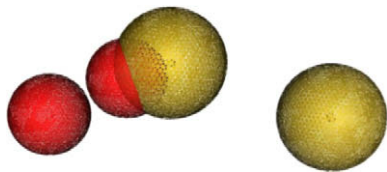


Figure 1. A 3D representation of the ligand-based (LB) pharmacophore model. Two hydrophobic groups (light yellow spheres), and two hydrogen-bond acceptors (red spheres) are shown.

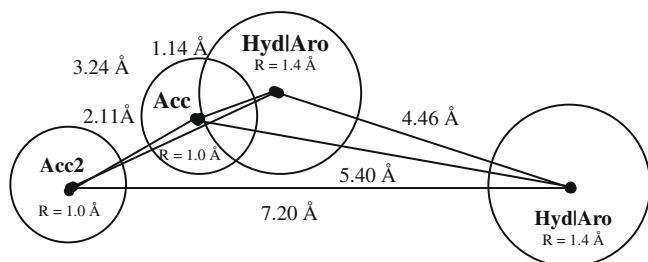


Figure 2. A 2D representation of the ligand-based (LB) pharmacophore model.

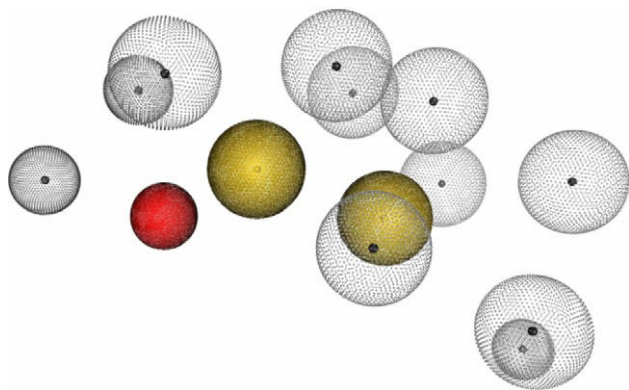


Figure 3. A 3D representation of the features in the structure-based (SB) pharmacophore model generated with LigandScout⁵⁵ from human aromatase- androstenedione complex [PDB code: 3EQM].³³ Two hydrophobic groups (light yellow spheres), one hydrogen-bond acceptor projection (red projected point), and 11 excluded volumes (gray spheres) are shown.

restrictions. In particular, excluded volume spheres placed in positions that are sterically disallowed are automatically added to the generated pharmacophore model. In this way, the aromatase-derived pharmacophore model contains the pharmacophore elements of the candidate ligands in response to the protein's active site environment.

The obtained pharmacophore hypothesis contained three chemical features: one hydrogen-bond acceptor and two hydrophobic

groups. The software automatically included 11 excluded volumes, which are spaces that inhibitor atoms are not permitted to occupy, to prevent the identification of inhibitor leads that would fit the pharmacophore elements well but that would overlap with receptor atoms. This resulted in a 14-point pharmacophore model with 11 excluded volumes, 1 hydrogen-bond acceptor projection (Acc2), and 2 hydrophobic/aromatic (HydAro) features. The 3D representation of these features is shown in Figure 3, as is the 2D representation in Figure 4. Interactions between the ligand (androstenedione), the heme moiety, and the active site residues of human aromatase are depicted in Figure 5.

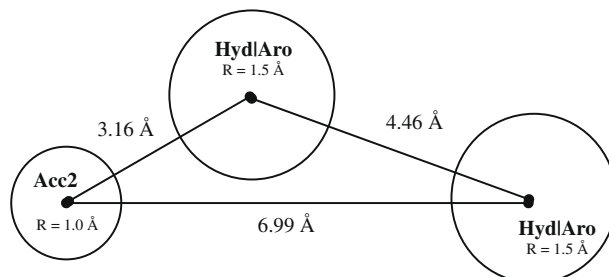


Figure 4. A 2D representation of the structure-based (SB) pharmacophore model.

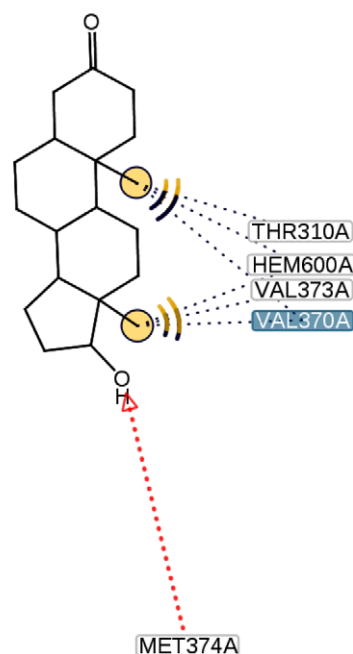


Figure 5. Interactions between the ligand (androstenedione), the heme moiety, and the active site residues of human aromatase.

The LB and SB Models were fused in MOE to generate a 'Merged' Model. Using SVL files acquired from MOE support, each original pharmacophore model (i.e., the LB and SB models) was converted into molecules (ph4_to_mol.svl)—shown in Figures 6 (A and B)—that were then superposed using the interactive superpose feature on MOE until the smallest value for the root mean squared deviation (RMSD = 0.11 Å) was achieved. Next, these molecular models were converted back into individual pharmacophore queries (mol_to_ph4.svl) and manually merged (merge_ph4.svl) to create the Merged Model. These SVL files are available upon request from MOE support.

The **Acc2** feature of the LB Model aligned with the **Acc2** feature of the SB Model, and the two pairs of **Hyd|Aro** features were also overlaid. These molecular structures and their superposition are shown in Figure 6c. The paired features were then averaged: using the Meters function on MOE,³⁶ the dimensions of the model were determined and were used to create averaged features in place of the individual, paired ones. Table 4 lists these distances and the average radii that were generated. The 3D representation of the final Merged Model is shown in Figure 7 as is its 2D representation in Figure 8.

As was mentioned when describing the LB pharmacophore model development, lower-potency LB models had been developed prior to the high-potency model. When these lower-potency models were superposed with the SB pharmacophore model, it was found that they did not align as well as the high-potency LB model. This is evidenced by the difference in typical RMSD values: 1.24 Å with the lower-potency model versus 0.11 Å with the high-potency model. Specifically, when attempting to align the SB model with the lower-potency LB model, it was found that the **Acc2** of the SB model is superposed upon the **Acc** chemical feature of the LB model, which is not chemically possible. The Merged Model derived from the high-potency LB model, however, features almost impeccable alignment with the SB model: the **Acc2** of the high-potency LB model aligns directly with that of the SB model, and the **Acc** feature remains a feature uniquely derived through ligand-based pharmacophore development (Fig. 6C). Thus, it was based on these general trends that 20 of the most potent AIs were used in the Training Set to generate the ligand-based pharmacophore model.

The quality of the three pharmacophore models was tested using an Active Test Set of 36 highly active nonsteroidal aromatase inhibitors (Table 2) as well as an Inactive Test Set of nine inactive compounds (Table 3) to see if the models could identify active AIs and differentiate between the active and inactive inhibitors. These sets of molecules were screened against the pharmacophore models using the Pharmacophore Search function in MOE. A 'hit' was defined as a compound that matched all of the characteristics of a model: the compounds were tested against the four chemical features of the LB model; the three chemical features and 11 excluded volumes of the SB model; the four total chemical features and 11 excluded volumes of the Merged Model.

All three models successfully identified 20 out of 20 of the Training Set AIs (Table 1). The test molecules were mapped onto each of the pharmacophore models (data not shown). Thirty-six

Table 4
Averaging of the SB and LB features

Feature type	Feature radius (Å)	Distance between features (Å)	New average radius (Å)
(LB) Acc2	1.000	0.130	1.065
(SB) Acc2	1.000		
(LB) Hyd Aro	1.400	0.100	1.455
(SB) Hyd Aro	1.500		
(LB) Hyd Aro	1.400	0.100	1.455
(SB) Hyd Aro	1.500		

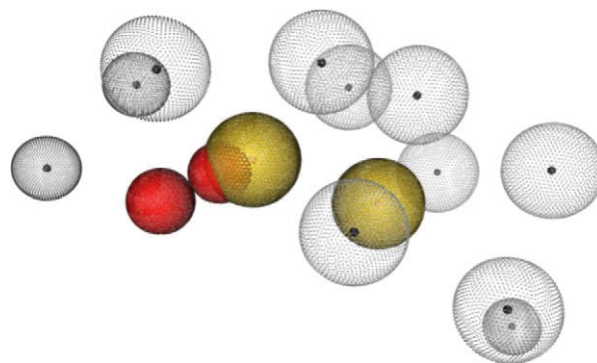


Figure 7. A 3D representation of the Merged Pharmacophore Model. Two hydrophobic groups (light yellow spheres), one hydrogen-bond acceptor projection (red projected point), and 11 excluded volumes (gray spheres) are shown.

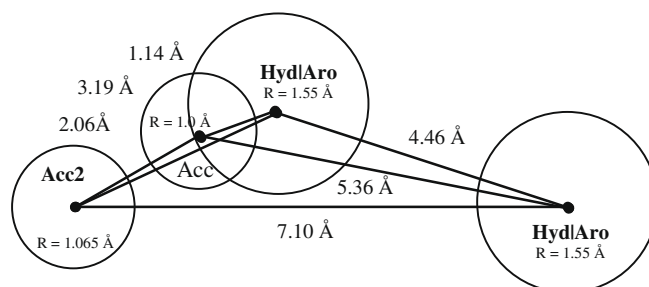


Figure 8. A 2D representation of the features in the Merged Pharmacophore Model.

molecules out of 36 were predicted as highly active, in perfect agreement with the actual biological data. However, the SB model was not able to distinguish very well between active versus inactive compounds (i.e., 8 out of 9 inactive molecules were identified by the model as being active). As shown in Table 5, the LB model is more discriminating than the SB Model, reporting only six false positives in the Inactive Test Set out of nine possible. The SB Model, therefore, performs better in identifying active inhibitors, while the LB model performs better in differentiating between active and inactive compounds. The Merged Model, however, combines

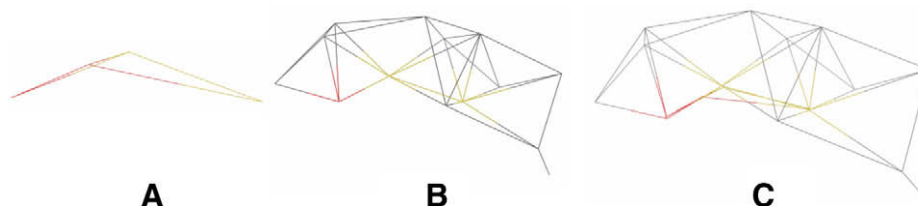


Figure 6. Molecular representations of the ligand-based pharmacophore model (LB) (A), structure-based pharmacophore model (SB) (B), and the SB and LB Models Superimposed (C).

Table 5

Validation results of the three models using the Training Set and the two Test Sets

	Training Set	Active Test Set	Inactive Test Set
LB Model	20/20	34/36	6/9
SB Model	20/20	36/36	8/9
MERGED Model	20/20	36/36	6/9

these strengths and proves superior to both of the original pharmacophore models.

Flexible docking studies were performed to see if plausible initial binding locations or orientations could be identified for some of the known aromatase inhibitors. The Molegro Virtual Docker program,⁷² which carries out docking studies based on a new hybrid search algorithm called guided differential evolution, was used for these studies. The guided differential evolution algorithm combines the *differential evolution* (DE) optimization technique with a cavity prediction algorithm that is dynamically used during the docking process.

The DE optimization technique works as follows. First, all individual ligands are initialized and evaluated according to the fitness function. The fitness of a candidate solution is the sum of the intermolecular interaction energy between the ligand and the protein and the intramolecular interaction energy of the ligand. In this study, the fitness was calculated using the docking score function described below. Next, for each individual in the population, an offspring is created by adding a weighted difference of the parent solutions, which are randomly selected from the population. Afterward, the offspring replaces the parent, if and only if it is fitter. Otherwise, the parent survives and is passed on to the next generation, which is an iteration of the algorithm. The termination condition used stops the search process when the current number of fitness—in other words, energy—evaluations performed exceeded the maximum number of evaluations allowed, which is the maximum evaluations parameter setting. This process is conducted as long as the termination condition is not fulfilled.

The docking scoring function of MolDock is based on a piecewise linear potential (PLP) introduced by Gehlhaar et al.^{73,74} and further extended in GEMDOCK by Yang and Chen.⁷⁵ In MolDock, the docking scoring function is extended with a new term, taking hydrogen bond directionality into account. The docking scoring function, E_{score} , is defined by the following energy terms:

$$E_{\text{score}} = E_{\text{inter}} + E_{\text{intra}}$$

where E_{inter} is the ligand–protein interaction energy and is defined as follows:

$$E_{\text{inter}} = \sum_{i \in \text{ligand}} \sum_{j \in \text{protein}} \left[E_{\text{PLP}}(r_{ij}) + 332.0 \frac{q_i q_j}{4r_{ij}^2} \right]$$

The summation runs over all heavy atoms in the ligand and all heavy atoms in the protein. The E_{PLP} term is the piecewise linear potential described below. The second term describes the electrostatic interactions between charged atoms. It is a Coulomb potential with a distance dependent dielectric constant given by $D(r) = 4r$. The numerical value of 332.0 fixes the units of the electrostatic energy to kcal/mol. For distances less than 2.0 Å, the electrostatic energy is cut off at the level corresponding to a distance of 2.0 Å to insure that no energy contribution can be higher than the clash penalty.

E_{PLP} is a piecewise linear potential using two different sets of parameters: one set for approximating the steric van der Waals term between atoms and the other stronger potential for hydrogen bonds. A bond is considered a hydrogen bond if one of the atoms can donate a hydrogen atom and the other atom can accept it. The PLP hydrogen bond term mentioned above only depends on the distance between atoms. E_{intra} is the internal energy of the ligand:

$$E_{\text{intra}} = \sum_{i \in \text{ligand}} \sum_{j \in \text{ligand}} E_{\text{PLP}}(r_{ij}) + \sum_{\text{flexible bonds}} A[1 - \cos(m \cdot \theta - \theta_0)] + E_{\text{clash}}$$

The double summation is between all atom pairs in the ligand, excluding atom pairs that are connected by two bonds or less. The second term is a torsional energy term, parametrized according to the hybridization types of the bonded atoms. The last term, E_{clash} , assigns a penalty of 1000 if the distance between two atoms (more than two bonds apart) is less than 2.0 Å. Thus, the E_{clash} term punishes infeasible ligand conformations.

Docking analyses in this study were used to analyze predictions for interactions between aromatase and two of the most potent inhibitors featured in this study, the results for which are explained below. The MVD algorithm for docking combines an algorithm for cavity detection in the 3D structure of the target protein with an optimization algorithm that evaluates different poses for the protein–ligand complex. Since the PDB file for the crystal structure of aromatase [PDB code: 3EQM]³³ contains the preferred substrate androstenedione bound to it at the active site, it became possible also to identify residues surrounding the binding pocket. Fifty-two such residues were identified and adjusted to take into account side-chain flexibility, and a search sphere radius of 17 Å was necessary to encompass them.

One large cavity (volume = 515.072 Å³) was detected and is shown in Figure 9. This volume agrees with the accepted literature value for the volume of the binding pocket (volume ≤ 400 Å³)³³ since the binding pocket is also included in the cavity. The shape of this cavity is also supported by literature stating that the access channel of this protein links its active site with the environment on the outer surface, allowing preferred androgenic substrates to diffuse through the channel toward the active site directly from within the membrane of the endoplasmic reticulum to which aromatase is lipid-linked.³³

The Refined Docking scheme features a three-stage process that, through optimization and individualized pose selection, focuses in on the most stable binding conformation of each separate ligand. The three stages are explained below:

Stage One: entry into the binding pocket. In setting up the docking runs, receptor flexibility was set to soften potentials during docking and minimize the receptor for each pose found. MolDock Score [GRID]⁷² was chosen as the scoring function. The origin was set to

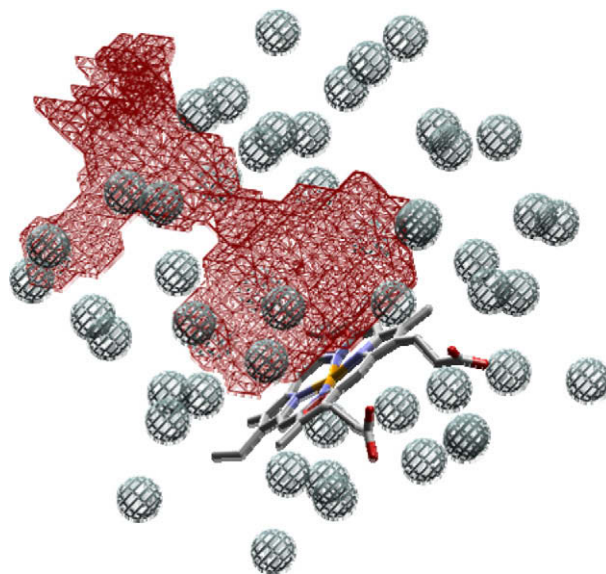


Figure 9. Access channel and 52 flexible side chains of aromatase represented in gray mesh spheres. The heme is represented in stick and color coded by atom.

Cavity 1 (volume: 515.07 Å³, surface: 884.48 Å²), and the radius of the constraint sphere was set to 10 Å. MolDock SE (simplex evolution) search algorithm was used for the search algorithm, and 50 runs with a population size of 100 were set for generation. Poses were constrained to the cavity, and pose clustering was set to return one pose per run. From the 50 poses generated per ligand docked, the pose with the lowest interaction energy was chosen in proceeding to Stage Two. In this way, the first stage of docking elucidates how the inhibitor fits into the enzyme, particularly whether or not it will find space in the binding pocket. Also, in the case of larger nonsteroidal inhibitors with long side chains, this stage of docking gives clues as to the most probable orientation of the molecule with respect to the access channel and active site area.

Stage Two: heteroatom coordination predictions. Ligands derived from the Stage One pose for each ligand were docked in Stage Two as described above, except that the origin of docking was chosen individually according to the size and shape of each ligand, and the constraint sphere radius was decreased from 10 to 7 Å. This modification was made in an effort to focus the ligand to an area in which the ligand side chains could be contained within the generated binding pocket and the entire ligand itself could be situated close enough to the heme moiety for enzyme inhibition. From among the 50 poses generated per ligand, low-energy poses that featured heteroatom coordination were selected. Heteroatom coordination was classified as any heteroatom, typically nitrogen, but also oxygen and sulfur atoms,^{22,45} at a distance of 4 Å or less from the iron atom of the aromatase heme moiety. Stage Two poses for each ligand were converted into new ligands that were then prepared for Stage Three docking.

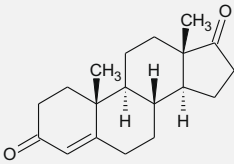
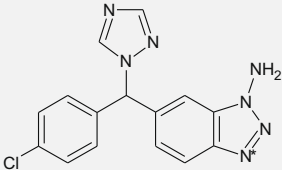
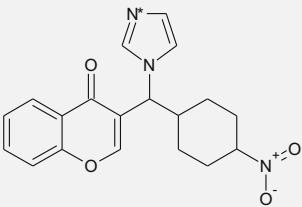
The second stage of this docking protocol therefore serves to implicate heteroatoms of nonsteroidal AIs as possible coordination partners to the iron atom of the aromatase heme moiety. Coordination as the sixth ligand of this heme group is predicted to be the main mechanism of action for nonsteroidal AIs, and this stage of the protocol generates predictions as to how this is accomplished. In addition, the second stage of docking allows for further orientation and refinement of steroidal ligand conformations in the binding pocket.

Stage Three: refinement of best predictions. Ligands derived from the Stage Two poses were docked in Stage Three as described above, except that each ligand was now used as its own reference ligand, which changes the origin of docking to the reference ligand itself. The radius of the constraint sphere was held constant at 7 Å. As such, Stage Three of Refined Docking allowed for optimization of the most important generated poses. From among all of the poses generated for each separate ligand, the lowest-energy poses were chosen and identified as representing the conformation in which the ligand is most stable in the binding pocket and the conformation in which the ligand spends most of its time. Therefore, the third and final stage of docking refines and optimizes the predictions made in the previous stage. After this stage of docking, an optimal orientation for inhibition as well as the identity of the heteroatom responsible for heme coordination can be determined with confidence.

For validation purposes the Refined Docking protocol was first applied to the androstenedione substrate, and the docked pose was found to be very close to the position of the substrate in the X-ray structure [PDB code: 3EQM]³³ (i.e., RMSD = 0.32 Å). The same protocol was applied for the two most potent nonsteroidal, vorozole³⁸ and 3-[imidazol-1-yl-(4-nitrophenyl)-methyl]-chromen-4-one.⁴⁰ The results showed that the MolDock Scores and Interaction Energy values for the two nonsteroidal AIs analyzed are comparable to the respective values for the preferred substrate androstenedione, and that a heteroatom for heme coordination is likely to be implicated in each nonsteroidal AI (see Table 6 and Fig. 10). Therefore, the high capability of prediction of the Refined Docking protocol proves that it can be used to virtually evaluate the potency of members of new classes of nonsteroidal aromatase inhibitors that will later be identified through the pharmacophore modeling and screening strategy described above.

In this study, the power of merging the two types of pharmacophore models—ligand-based (LB) and structure-based (SB)—has been demonstrated. The LB Model contributed its specificity to discriminate against inactive compounds while the SB Model contributed its ability to identify all active AIs. Thus, both models and both types of information—known active and inactive inhibitors as well as the structure of the enzyme—are essential to the development of

Table 6
Results generated from virtual refined docking

Compound name	IC ₅₀ value (nM)	Chemical structure	Interaction energy (kJ/mol)	MolDock score
Androstenedione (preferred substrate)			−133.713	−143.464
Vorozole ³⁸	1.38		−174.87	−155.107
3-[Imidazol-1-yl-(4-nitrophenyl)-methyl]-chromen-4-one ⁴⁰	2.3		−180.541	−175.752

* Heteroatom involved in heme coordination.

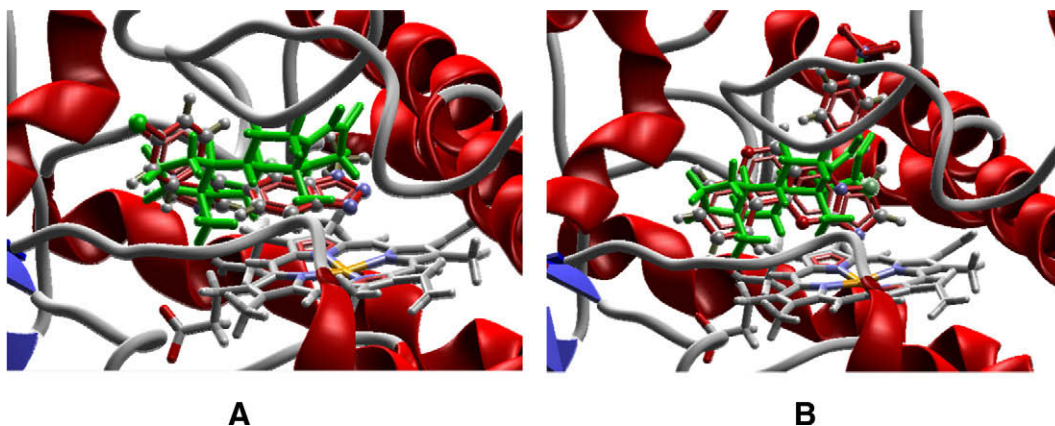


Figure 10. Best Poses for Vorozole (A) and 3-Imidazolyl (B). The heme is represented in stick and colored by atom; the androstenedione substrate is colored in green; the Als are shown in ball-and-stick.

a successful pharmacophore model and, subsequently, to the identification of more potent, highly specific, and potentially less toxic aromatase inhibitors.

The utility of our 'Merged' Pharmacophore is shown by the fact that the model elaborated within this study was able to accurately predict known inhibitors and successfully discriminate between active and inactive compounds. The mapping information based on the pharmacophore model that was developed is now being used to identify novel lead compounds with improved inhibitory activity through 3D database searches. The implementation of this in silico structure-based drug design approach to identify and develop aromatase inhibitors represents a viable strategy to discover new, more efficacious aromatase inhibitors with less adverse side effects for breast cancer treatment. If successfully carried out, this study will enable the determination of the in vivo efficacy of the newly identified inhibitor(s) and launch their further development as potential drugs for breast cancer patients.

Acknowledgments

We thank the Chemical Computing Group for their generous software support. We would also like to acknowledge the national BBSI program (<http://bbsi.eeicom.com>), a joint initiative of the NIH-NIBIB and NSF-EEC, and the BBSI @ Pitt, supported by the National Science Foundation under Grant EEC-0234002.

References and notes

- Cancer Facts & Figures 2008. American Cancer Society. 2009.
- Brueggemeier, R. W.; Hackett, J. C.; az-Cruz, E. S. *Endocr. Rev.* **2005**, *26*, 331.
- Meunier, B.; de Visser, S. P.; Shaik, S. *Chem. Rev.* **2004**, *104*, 3947.
- Simpson, E. R.; Mahendroo, M. S.; Means, G. D.; Kilgore, M. W.; Hinshelwood, M. M.; Graham-Lorence, S.; Amarneh, B.; Ito, Y.; Fisher, C. R.; Michael, M. D. *Endocr. Rev.* **1994**, *15*, 342.
- Miller, W. R. *Cancer Treat. Rev.* **1997**, *23*, 171.
- Miller, W. R.; Bartlett, J.; Brodie, A. M.; Brueggemeier, R. W.; di, S. E.; Lonning, P. E.; Llombart, A.; Maass, N.; Maudelonde, T.; Sasano, H.; Goss, P. E. *Oncologist* **2008**, *13*, 829.
- Brodie, A. M.; Njar, V. C. J. *Steroid Biochem. Mol. Biol.* **1998**, *66*, 1.
- Simpson, E. R.; Davis, S. R. *Endocrinology* **2001**, *142*, 4589.
- Brodie, A.; Sabnis, G.; Jelovac, D. J. *Steroid Biochem. Mol. Biol.* **2006**, *102*, 97.
- Castellano, S.; Stefancich, G.; Ragno, R.; Schewe, K.; Santoriello, M.; Caroli, A.; Hartmann, R. W.; Sbardella, G. *Bioorg. Med. Chem.* **2008**, *16*, 8349.
- Colozza, M.; Califano, R.; Minenza, E.; Dinh, P.; Azambuja, E. *Mini-Rev. Med. Chem.* **2008**, *8*, 564.
- Dutta, U.; Pant, K. *Med. Oncol.* **2008**, *25*, 113.
- Eisen, A.; Trudeau, M.; Shelley, W.; Messersmith, H.; Pritchard, K. I. *Cancer Treat. Rev.* **2008**, *34*, 157.
- Gobbi, S.; Cavalli, A.; Bisi, A.; Recanatini, M. *Curr. Top. Med. Chem.* **2008**, *8*, 869.
- Jackson, T.; Woo, L. W.; Trusselle, M. N.; Purohit, A.; Reed, M. J.; Potter, B. V. *ChemMedChem* **2008**, *3*, 603.
- Osborne, C.; Tripathy, D. *Annu. Rev. Med.* **2005**, *56*, 103.
- Osborne, C. K.; Schiff, R. J. *Steroid Biochem. Mol. Biol.* **2005**, *95*, 183.
- Recanatini, M.; Cavalli, A. *Bioorg. Med. Chem.* **1998**, *6*, 377.
- Recanatini, M.; Cavalli, A.; Valenti, P. *Med. Res. Rev.* **2002**, *22*, 282.
- Spinelli, G. P.; Tomao, F.; Miele, E.; Pasciuti, G.; Russillo, M.; Tomao, S. *Recenti Prog. Med.* **2008**, *99*, 34.
- Hong, Y.; Yu, B.; Sherman, M.; Yuan, Y. C.; Zhou, D.; Chen, S. *Mol. Endocrinol.* **2007**, *21*, 401.
- Neves, M. A.; Dinis, T. C.; Colombo, G.; Sae Melo, M. L. *J. Med. Chem.* **2009**, *52*, 143.
- Geisler, J.; Lonning, P. E. *Clin. Cancer Res.* **2005**, *11*, 2809.
- A Report From Breast Cancer Action. SIDE EFFECTS REVISITED: Women's Experiences With Aromatase Inhibitors. A Report From Breast Cancer Action. 2008.
- Favia, A. D.; Cavalli, A.; Masetti, M.; Carotti, A.; Recanatini, M. *Proteins* **2006**, *62*, 1074.
- Karkola, S.; Holtje, H. D.; Wahala, K. *J. Steroid Biochem. Mol. Biol.* **2007**, *105*, 63.
- Loge, C.; Le, B. M.; Marchand, P.; Robert, J. M.; Le, B. G.; Palzer, M.; Hartmann, R. W. *J. Enzyme Inhib. Med. Chem.* **2005**, *20*, 581.
- Murthy, J. N.; Nagaraju, M.; Sastry, G. M.; Rao, A. R.; Sastry, G. N. *J. Comput. Aided Mol. Des.* **2005**, *19*, 857.
- Cavalli, A.; Greco, G.; Novellino, E.; Recanatini, M. *Bioorg. Med. Chem.* **2000**, *8*, 2771.
- Janicke, F. *Anti-Cancer Drugs* **2008**, *19*, S7.
- Neves, M. A.; Dinis, T. C.; Colombo, G.; Sae Melo, M. L. *ChemMedChem* **2007**, *2*, 1750.
- Numazawa, M.; Komatsu, S.; Tominaga, T.; Yamashita, K. *Chem. Pharm. Bull. (Tokyo)* **2008**, *56*, 1304.
- Ghosh, D.; Griswold, J.; Erman, M.; Pangborn, W. *Nature* **2009**, *457*, 219.
- Chemical Computing Group. Pharmacophore Modeling. Chemical Computing Group. 2008.
- Wolber, G.; Langer, T. *J. Chem. Inf. Model.* **2005**.
- Wolber, G. *Drug Discovery Today* **2008**.
- ACD/ChemSketch: Version 8.0 for Microsoft Windows. Advanced Chemistry Development, Inc. 2004.
- Goss, P. E. *Breast Cancer Res. Treat.* **1998**, *49*, S59.
- Hartmann, R. W.; Bayer, H.; Grun, G.; Sergejew, U. B.; Mitrenga, M. *J. Med. Chem.* **1995**, *38*, 2103.
- Cavalli, A.; Bisi, A.; Bertucci, C.; Rosini, C.; Paluszczak, A.; Gobbi, S.; Giorgio, E.; Rampa, A.; Belluti, F.; Piazzi, L.; Valenti, P.; Hartmann, R. W.; Recanatini, M. *J. Med. Chem.* **2005**, *48*, 7282.
- Le, B. M.; Marchand, P.; Duflos, M.; Delevoye-Seiller, B.; Piessard-Robert, S.; Le, B. G.; Hartmann, R. W.; Palzer, M. *Arch. Pharm.* **1997**, *330*, 141.
- Bossche, H. V. *J. Steroid Biochem. Mol. Biol.* **1992**, *43*, 1003.
- Leonetti, F.; Favia, A. D.; Rao, A. R.; Aliano, R.; Paluszczak, A.; Hartmann, R. W.; Carotti, A. *J. Med. Chem.* **2004**, *47*, 6792.
- Simpson, E. R.; Dowsett, M. *Endocr. Soc.* **2002**.
- Hackett, J. C. PhD Thesis: Computational Investigations of Cytochrome P450 Aromatase Catalysis and Biological Evaluation of Isoflavone Aromatase Inhibitors. The Ohio State University. 2004.
- Gobbi, S.; Cavalli, A.; Negri, M.; Schewe, K.; Belluti, F.; Piazzi, L.; Hartmann, R. W.; Recanatini, M.; Bisi, A. *J. Med. Chem.* **2007**, *50*, 3420.
- Nativelle-Serpentini, C.; Moslemi, S.; Yous, S.; Park, C. H.; Lesieur, D.; Sourdaine, P.; Seralini, G. *J. Enzyme Inhib. Med. Chem.* **2004**, *19*, 119.
- Sonnet, P.; Dallenmagne, P.; Guillon, J.; Enguehard, C.; Stiebing, S.; Tanguy, J.; Bureau, R.; Rault, S.; Auvray, P.; Moslemi, S.; Sourdaine, P.; Seralini, G. *Bioorg. Med. Chem.* **2000**, *8*, 945.
- Hartmann, R. W.; Paluszczak, A.; Lacan, F.; Ricci, G.; Ruzziconi, R. *J. Enzyme Inhib. Med. Chem.* **2004**, *19*, 145.

50. Kellis, J. T.; Vickery, L. E. *Science* **1984**, 225, 1032.
51. Ahmed, S.; Smith, H. J.; Nicholls, P. J.; Whomsley, R.; Cariuk, P. *Drug Des Discov.* **1995**, 13, 27.
52. Jacobs, C.; Frotscher, M.; Dannhardt, G.; Hartmann, R. W. *J. Med. Chem.* **2000**, 43, 1841.
53. Recanatini, M.; Bisi, A.; Belluti, F.; Gobbi, S.; Rampa, A.; Valenti, P.; Palzer, M.; Paluszczak, A.; Hartmann, R. W. *J. Med. Chem.* **2001**, 44, 672.
54. van Meeuwen, J. A.; Nijmeijer, S.; Mutarapat, T.; Ruchirawat, S.; de Jong, P. C.; Piersma, A. H.; van den Berg, M. *Toxicol. Appl. Pharmacol.* **2008**, 228, 269.
55. Wolber, G.; Langer, T. J. *Chem. Inf. Model.* **2005**, 45, 160.
56. Wolber, G.; Seidel, T.; Bendix, F.; Langer, T. *Drug Discovery Today* **2008**, 13, 23.
57. Koymans, L. M. H.; Moereels, H.; Vanden Bossche, H. *J. Steroid Biochem. Mol. Biol.* **1995**, 53, 191.
58. Leze, M.; Le, B. M.; Pinson, P.; Paluszczak, A.; Duflos, M.; Le, B. G.; Hartmann, R. W. *Bioorg. Med. Chem. Lett.* **2006**, 16, 1134.
59. Jeong, H.; Shin, Y. G.; Kim, I.; Pezzuto, J. M. *Arch. Pharm. Res.* **1999**, 22, 309.
60. Mason, J. I.; Barbara, A. M.; Ollcott, M.; Sheets, J. L. *Biochem. Pharmacol.* **1985**, 34, 1087.
61. Brueggemeier, R. W.; Richards, J. A.; Joomprabutra, S.; Bhat, A. S.; Whetstone, J. L. *J. Steroid Biochem. Mol. Biol.* **2001**, 79, 75.
62. Blanco, J. G.; Gill, R. R.; Alvarez, C. I.; Patrito, L. C.; Genti-Raimondi, S.; Flury, A. *FEBS Lett.* **1997**, 409, 396.
63. Sonnet, P.; Guillon, J.; Enguehard, C.; Dallenmagne, P.; Bureau, R.; Rault, S. *Bioorg. Med. Chem. Lett.* **1998**, 8, 1041.
64. Ibrahim, A. R.; Abul-Hajj, Y. *J. Steroid Biochem. Mol. Biol.* **1990**, 37, 257.
65. Bayer, H.; Batzl, C.; Hartmann, R. W.; Mannschreck, A. *J. Med. Chem.* **1991**, 34, 2685.
66. Auvray, P.; Moslemi, S.; Sourdain, P.; Galopin, S.; Seralini, G.; Enguehard, C.; Dallenmagne, P.; Bureau, R.; Sonnet, P.; Rault, S. *Eur. J. Med. Chem.* **1998**, 33, 451.
67. Pouget, C.; Fagnere, C.; Basly, J.-P.; Habrioux, G.; Chulia, A. *J. Bioorg. Med. Chem. Lett.* **2002**, 12, 2859.
68. Adlercreutz, H.; Bannwart, C.; Wahala, K.; Makela, T.; Brunow, G.; Hase, T.; Arosemena, P. J.; Kellis, J. T.; Vickery, L. E. *J. Steroid Biochem. Mol. Biol.* **1993**, 44, 147.
69. Whomsley, R.; Fernandez, E.; Nicholls, P. J.; Smith, H. J.; Lombardi, P.; Pestellini, V. *J. Steroid Biochem. Mol. Biol.* **1993**, 44, 675.
70. Hida, T.; Ishii, T.; Kanamaru, T.; Muroi, M. *J. Antibiot.* **1991**, 44, 600.
71. Kim, Y.-W.; Hackett, J. C.; Brueggemeier, R. W. *J. Med. Chem.* **2004**, 47, 4032.
72. Thomsen, R.; Christensen, M. H. *J. Med. Chem.* **2006**, 49, 3315.
73. Gehlhaar, D. K.; Verkhivker, G. Proceedings of the Fourth International Conference on Evolutionary Programming. 1995, pp 615–627.
74. Gehlhaar, D. K.; Bouzida, D. Proceedings of the Seventh International Conference on Evolutionary Programming. 1998; pp 449–461.
75. Yang, J. M.; Chen, C. C. *Proteins* **2004**, 55, 288.