

A Ligand-Based Approach To Identify Quantitative Structure–Activity Relationships for the Androgen Receptor

Casey E. Bohl,^{†,‡} Cheng Chang,^{§,‡} Michael L. Mohler,[#] Jiyun Chen,[†] Duane D. Miller,[#] Peter W. Swaan,[§] and James T. Dalton^{*,†}

Division of Pharmaceutics and Pharmaceutical Chemistry, College of Pharmacy, The Ohio State University, Columbus, Ohio 43210, Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland 21201, and Department of Pharmaceutical Sciences, College of Pharmacy, University of Tennessee, Memphis, Tennessee 38163

Received February 3, 2004

We examined the three-dimensional quantitative structure–activity relationship (QSAR) of a group of endogenous and synthetic compounds for the androgen receptor (AR) using comparative molecular field analysis (CoMFA). The goal of these studies was to identify structural features necessary for high binding affinity and optimization of selective androgen receptor modulators (SARMs). A homology model of the AR was used as a scaffold to align six lead compounds that served as templates for alignment of the remaining 116 structures prior to CoMFA modeling. The conventional r^2 and cross-validated q^2 relating observed and predicted relative binding affinity (RBA) were 0.949 and 0.593, respectively. Comparison of predicted and observed RBA for a test set of 10 compounds resulted in an r^2 of 0.954, demonstrating the excellent predictive ability of the model. These integrated homology modeling and CoMFA studies identified critical amino acids for SARM interactions and provided QSAR data as the basis for mechanistic studies of AR structure, function, and design of optimized SARMs.

Introduction

Selective androgen receptor modulators (SARMs) were discovered in our laboratory and were shown to produce desirable anabolic responses while minimizing androgenic effects.^{1,2} Recent *in vivo* studies in rats demonstrated that two novel SARMs (denoted **S-1** and **S-4** in ref 5) elicit potent and tissue-selective pharmacologic effects. These nonsteroidal SARMs have many theoretical advantages over conventional steroid therapies, which are not selective between anabolic and androgenic tissue, have poor oral bioavailability, and are associated with side effects from cross-reactivity with other steroid receptors.³ Since the majority of nonsteroidal androgen receptor (AR) ligands demonstrate binding affinities 10-fold lower than the leading androgenic steroids,⁴ we sought to improve upon the favorable receptor–ligand interactions to obtain more potent compounds. We applied a traditional medicinal chemistry and pharmacologic approach to design, synthesize, and evaluate a large number of derivatives with varying AR binding affinities. This trial and error search resulted in the refinement of structure–activity relationships (SAR) for AR binding. However, this approach is both time- and cost-intensive and is not amenable to high-throughput screening. An integrated approach using molecular modeling (SYBYL 6.8) and CoMFA was employed to avoid these constraints, to create a visual database incorporating our ligands and other structurally diverse compounds, and to provide a rationale for

the synthesis of higher affinity AR ligands with improved activity.

Structure–activity relationships have been previously described for analogues of hydroxyflutamide and bicalutamide.⁵ These include an electron-withdrawing group at the 4-position of the A-ring and an amide linkage attached to the chiral center. The chiral center favors the *R* configurations in S-linked and SO₂-linked bicalutamide analogues⁶ and the *S* configurations in N- and O-linked analogues⁴ because of the lower priority of the nitrogen and oxygen with respect to the amide. Thus, all high-affinity N-, O-, S-, and SO₂-linked bicalutamide derivatives are of the same three-dimensional (3D) configuration. The presence of an NO₂ at the para position of the A-ring appears to be favored over a CN group because of stronger electron-withdrawing properties.⁷

Waller et al.,⁸ using steroidal ligands, hydroxyflutamide, and a number of pesticides, developed a CoMFA model with a cross-validated r^2 of 0.792 from a training set of 20 compounds. Alignment was performed via overlap of the A-ring of DHT to six-membered rings found in each of the other ligands with an additional energy minimization after field fit and internal coordinate adjustments to provide maximal overlap with DHT.⁸ Steric bulk was favorable at several sites of the steroidal ring system, including the B-ring at the C6 and C7 positions, C-ring at the C11 and C12 position, and D ring at the C17 position of steroids. In addition, a positive charge at the C17 position and a negative charge at the C3 position of the steroid ring system resulted in increased affinity.⁸ Steric bulk at positions more distant from the C17 position was found to be unfavorable. Our approach to design a CoMFA model differed considerably for two reasons: (1) the receptor

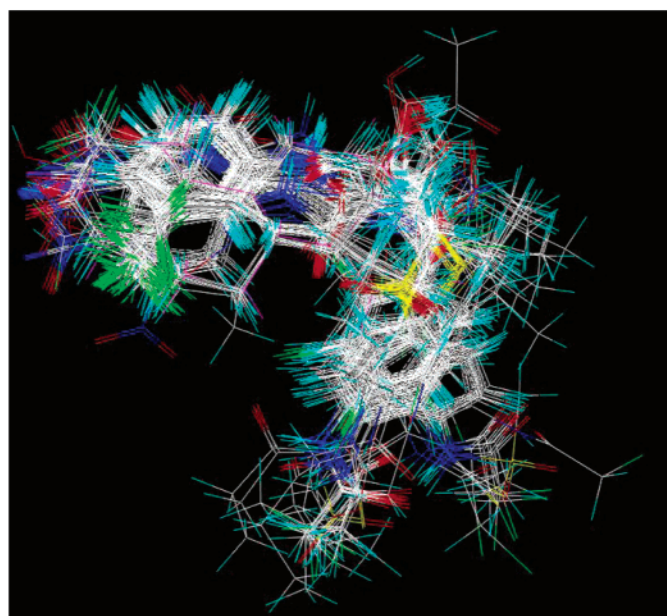
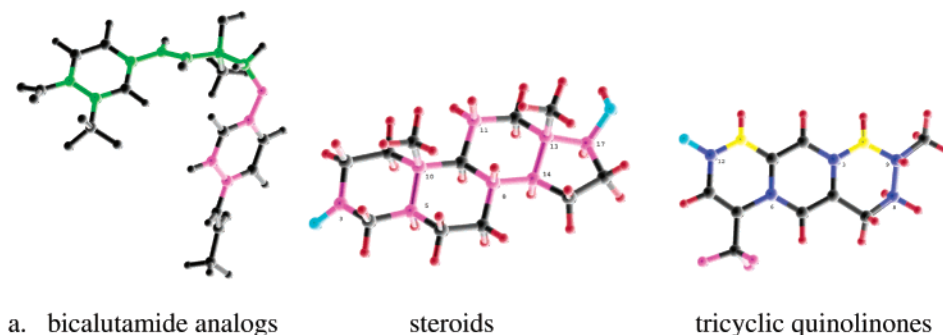
* To whom correspondence should be addressed. Office phone: 614-688-3797. Fax: 614-292-7766. E-mail: dalton.1@osu.edu.

[†] The Ohio State University.

[‡] These authors contributed equally to the work presented in this manuscript.

[§] University of Maryland.

[#] University of Tennessee.



b.

Figure 1. Alignment points for each set of compounds. (a) Conformation alignment points. Hydroxyflutamide analogues were aligned to the ether-linked SARM, **S-4** (compound **5**), by the points shown in green. Bicalutamide analogues were aligned using the pink and green points highlighted in **S-4**. Compounds **24**, **29**, **S-4**, and **60** were used as representatives of sulfonyl, thioether, ether, and amide-linked ligands, respectively, for docking. DHT was docked into the homology receptor, and other steroids were then aligned by the points shown in pink to DHT. Compound **85** was docked into the homology receptor, and other tricyclic quinolinones were aligned to this ligand by the points shown in blue. (b) Overlap of ligands generated from field fit. Training and test compounds were aligned to their analogous structure by the points shown in panel a prior to CoMFA analyses.

was used as a template to align diverse structures, and (2) the majority of the structures used in our model were bicalutamide and hydroxyflutamide analogues. These modifications allow us to combine docking information into our ligand predictions as well as provide a model that is specific for ligands resembling hydroxyflutamide and bicalutamide.

The final CoMFA model is dependent on the 3D conformation of the individual molecules in the training set, as well as their overlap. Since the AR ligands synthesized thus far are conformationally mobile, the question becomes the following: how do we assign significance to one potential conformation relative to another? Originally, 3D quantitative SAR (QSAR) techniques such as CoMFA were developed for ligand design problems in the absence of structural information about the target receptor. However, it is increasingly applied in situations in which information regarding receptor structure exists in the form of biophysical data (X-ray, NMR, etc.) or theoretical models.⁹ In these situations, a firm basis for conformational selection exists and can be expected to improve the overall effectiveness of the CoMFA-based ligand design approach. While the crystal

structure for the AR ligand binding domain (LBD) is available, most studies suggest that bicalutamide and its analogues adopt a unique conformation in solution. Poujol et al.¹⁰ suggested that the aromatic ring of bicalutamide interfered with appropriate AR LBD folding. We developed a homology model of the AR to the progesterone receptor that allows docking of these compounds and explains the mechanism of binding for chiral nonsteroidal ligands.¹¹ Importantly, this model also provides a scaffold for alignment purposes. Studies herein utilize this approach to superimpose the CoMFA contours and the receptor LBD to better understand the amino acid interactions that are important to obtain high AR binding affinity. However, it is important to note that the activity of the SARMS that we recently reported cannot be explained by crystal structure, suggesting that it may be inaccurate due to crystal packing effects. Thus, the current work is also aimed at providing a rationale for site-directed mutagenesis in order to further investigate the amino acid contour relationship and the structure of the AR LBD during interaction with SARMS.

Materials and Methods

Competitive Binding Assay. Binding data for bicalutamide and its derivatives were determined in our laboratory by a method previously described.⁴ Cytosolic AR was obtained from ventral prostates of Sprague-Dawley rats. Prostates were excised and immediately immersed in ice-cold homogenization buffer (10 mM Tris, 1.5 mM disodium EDTA, 0.25 M sucrose, 10 mM sodium molybdate, and 1 mM PMSF). Prostates were minced with scissors, homogenized, and centrifuged (model L8-M, Beckman Instruments Inc., Palo Alto, CA). The supernatant (cytosol) was collected and stored at -80°C until use. Aliquots (50 μL) of AR cytosol were incubated with a saturating concentration of [³H]-MIB (1 nM) and 1 μM of triamcinolone acetonide at 4°C for 18 h in the absence or presence of increasing concentrations of the compound of interest (10^{-1} – 10^4 nM). Nonspecific binding of [³H]-MIB was determined by adding 1000 nM MIB to the incubate. Bound and free radioligand were separated using hydroxyapatite, and the concentration of bound [³H]-MIB was determined. The IC_{50} (concentration of the test compound that inhibited the specific binding of [³H]-MIB by 50%) was determined using nonlinear regression. Relative binding affinity (RBA) was calculated as $\text{RBA} = (\text{IC}_{50} \text{ of DHT})/(\text{IC}_{50} \text{ of compound of interest})$. Binding data for the steroidal ligands and tricyclic quinolinones were obtained from publications by Waller et al.⁸ and Zhi et al.,¹² respectively. Only data with a reported DHT value were used in this study to normalize for interlaboratory variations. Biological data for the CoMFA model was input as pRBA.

Molecular Structure Building. The starting geometries of ligands were built using standard bond distances and bond angles with the sketch module of SYBYL. The correct conformation of the ring structures was used during the structure building process. The AM1 Hamiltonian within the MOPAC suite of programs was used to assign the point charges of atoms as well as to optimize the geometry by minimizing the energy. Molecular mechanics corrections were made to any CO–NH bond using the keyword MMOK. The keyword PRECISE, which increases the criterion for terminating all optimizations by a factor of 100, was also used. Of 122 ligands, 10 were randomly selected as test compounds; the remaining 112 ligands were used as the training data set.

Ligand Alignment. An important consideration in the creation of a CoMFA model is how to align the different compounds. Often, similar functional groups or common atoms are chosen as alignment points to determine how compounds will overlap. Docking ligands to a receptor has also been employed as a rationale for alignment. In the present study we used a combination of these methods by first solving the theoretical AR-bound conformations of several template AR ligands and then fitting the analogous molecules in the CoMFA training set to these template conformations using common atoms as alignment points. Compound **S-4** (compound **5**), DHT, and compound **85** were used as templates for the hydroxy-flutamide derivatives, steroids, and tricyclic quinolinones, respectively. The templates for the two-ring model were the ether-linked compound **S-4**, the sulfone-linked bicalutamide, a secondary amine-linked derivative, and a thioether-linked derivative (see Figure 1 for alignment points). These representative ligands were docked into the AR LBD homology model (detail below) to determine their relative conformations. Analogous molecules were then fitted to their respective template ligands using SYBYL/MULTIFIT.¹³ The spring constant was set to 20 for all compounds. The geometries of the resulting molecules were initially optimized by the SIMPLEX method and then by MAXMIN2 using the POWELL method. An energy gradient change convergence of 0.05 kcal/mol or a maximum of 1000 iterations, whichever occurred first, was used to terminate the minimization.

Two-ring ligands, such as bicalutamide, adopted a conformation very different from that of the MOPAC-optimized analogues. To avoid severely distorted structures, bicalutamide-like compounds were rebuilt on the basis of the conformation of the docked templates. A moderate energy minimization was performed (rms gradient of 1.0 and maximum

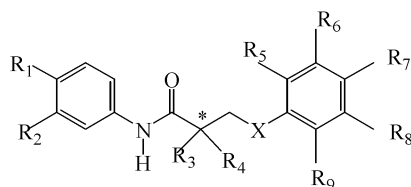
number of steps of 1000) after structure generation in order to maintain the bound conformation without the receptor environment constraints. Alignment points were selected on the basis of important functional groups and common atoms for each set of molecules as shown in Figure 1. The final overlap was generated by merging all six sets of Fieldfit aligned molecules and maintaining their relative docked position.

FlexX Docking. We developed a model of the hAR LBD based on homology to the human progesterone receptor and postulated the binding modes for testosterone and several chiral nilutamide derivatives.¹¹ In the current study, we docked a variety of molecules into the AR G1 homology model (direct binding to ligand as opposed to water-mediated binding) using a similar methodology as previously employed.¹¹ Briefly, the docking was performed with FlexX. The AR G1 model was prepared for docking by removing all hydrogens. The ligand binding site was defined as all residues within 6.5 Å of testosterone. The dihedral angle orienting the hydroxyl of T877 was rotated to $+120^{\circ}$ in the active site such that it could act as a hydrogen bond donor with bound ligands. Ligands were docked into the ligand binding site with hydrogens present, and formal charges were assigned by FlexX. The FlexX docking solutions were initially relaxed in FlexX, and scores were assigned for both relaxed and unrelaxed docking solutions. The top 30 scoring (unrelaxed) docking solutions were saved.

Ligand–AR Complex Energy Refinement. Ligands that had multiple docking solutions were evaluated to determine those that would be chosen for further energy refinement. Docking solutions were prioritized by C-Score of the relaxed conformations to eliminate all complexes except those for which all four scoring functions were predicted to be in the top 50% of the 30 docking solutions. Manual inspection of the ligand–AR interactions of these complexes allowed further elimination of complexes with fewer interactions. If multiple conformations remained, the final selection criterion was the best total score for the relaxed conformer. Energy refinement was achieved by merging the docked ligand with the homology model (without testosterone) and energy-minimizing a subset of this ligand–receptor complex. T877 was rotated to 120° as described above, protons were added to the receptor, and MMFF94s charges were applied to the receptor for energy minimization. SYBYL 6.8 was employed to minimize the subset of this complex, allowing the assignment of a hot radius (unrestrained movement) and an interesting radius (limited movement). The hot radius was 4.0 Å and the interesting radius was 8.0 Å from the ligand. The minimizations proceeded for 100 000 iterations or to a 0.005 kcal/mol termination gradient. The resulting ligand conformations served as the conformational templates (see above) for the CoMFA training sets.

CoMFA. SYBYL/CoMFA is an especially useful QSAR technique to explain the observed biological properties based on the sampling of electrostatic (Coulombic interactions) and steric (van der Waals interactions) fields surrounding a given set of compounds. Partial least-squares (PLS) was used to correlate the field descriptors with biological activities of the AR ligands. Both fields were calculated using an sp^3 carbon probe atom (+1 charge and 1.52 Å van der Waals radius) on a 2.0 Å spaced lattice, which extended beyond the dimensions of each structure by 4.0 Å in all directions. When the steric repulsion was greater than 30.0 kcal/mol, it was set to 30.0 kcal/mol, which lessens the distortion due to extreme energy terms in the final model. The indicator fields¹⁴ and hydrogen bond fields¹⁵ generated by the advanced CoMFA module were also included in the analysis. In the indicator fields, all electrostatic energies below 1.0 kcal/mol and steric energies below 10.0 kcal/mol were set to zero to eliminate noise. The hydrogen bond field descriptors were set to zero at sterically prohibited points. Sterically allowed points close to hydrogen bond acceptors were assigned a nominal steric potential, while the points close to hydrogen bond donors were assigned a nominal electrostatic potential. As a result, the steric field represented the acceptor component while the electrostatic

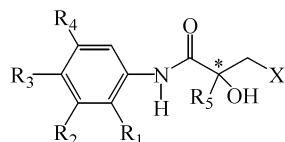
Table 1.



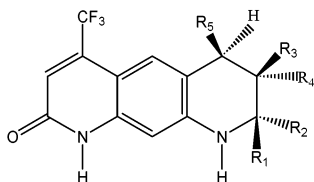
cpds	actual	predicted	residual	R1	R2	R3	R4	X	R5	R6	R7	R8	R9	isomer
1	1.35	1.36	-0.01	NO ₂	CF ₃	OH	CH ₃	O	H	H	F	H	H	S
2	2.92	2.63	0.29	NO ₂	CF ₃	OH	CH ₃	O	H	H	F	H	H	R
3	2.13	1.88	0.25	NO ₂	CF ₃	OH	CH ₃	O	H	H	COCH ₃	H	H	S
4	1.35	1.55	-0.2	NO ₂	CF ₃	OH	CH ₃	O	H	H	COCH ₂ CH ₃	H	H	S
5	1.17	1.18	-0.01	NO ₂	CF ₃	OH	CH ₃	O	H	H	NHCOCH ₃	H	H	S
6	0.96	1.43	-0.47	CN	CF ₃	OH	CH ₃	O	H	H	F	H	H	S
7	0.98	1.41	-0.43	CN	CF ₃	OH	CH ₃	O	H	H	Cl	H	H	S
8	0.75	1.12	-0.37	NO ₂	CF ₃	OH	CH ₃	O	H	H	F	F	H	S
9	1.51	0.95	0.56	NO ₂	CF ₃	OH	CH ₃	O	H	H	F	F	F	S
10	1.47	1.19	0.28	NO ₂	CF ₃	OH	CH ₃	O	H	F	H	F	F	S
11	1.24	1.18	0.06	NO ₂	CF ₃	OH	CH ₃	O	H	H	F	H	CH ₃	S
12	0.96	1.12	-0.16	NO ₂	CF ₃	OH	CH ₃	O	H	H	F	H	F	S
13	1.23	0.89	0.34	NO ₂	CF ₃	OH	CH ₃	O	H	H	H	H	F	S
14	0.88	0.95	-0.07	NO ₂	CF ₃	OH	CH ₃	O	F	F	F	F	F	S
15	1.04	1.17	-0.13	NO ₂	CF ₃	OH	CH ₃	O	H	H	Cl	F	H	S
16	1.37	1.09	0.28	NO ₂	CF ₃	OH	CH ₃	O	H	H	F	Cl	H	S
17	0.69	1.14	-0.45	NO ₂	CF ₃	OH	CH ₃	O	H	H	Cl	Cl	H	S
18	1.67	1.33	0.34	NO ₂	CF ₃	OH	CH ₃	O	H	H	NHCOCH ₂ Br	H	H	S
19	0.8	1.03	-0.23	NO ₂	CF ₃	OH	CH ₃	O	H	H	NHCOCH ₂ Cl	H	H	S
20	1.61	1.83	-0.22	Cl	CF ₃	OH	CH ₃	O	H	H	F	H	H	S
21	1.96	1.7	0.26	F	CF ₃	OH	CH ₃	O	H	H	F	H	H	S
22	1.1	1.26	-0.16	CN	Cl	OH	CH ₃	O	H	H	F	H	H	S
23	2.16	1.86	0.3	Br	CF ₃	OH	CH ₃	O	H	H	Cl	H	H	S
24	1.61	1.85	-0.24	CN	CF ₃	OH	CH ₃	SO ₂	H	H	F	H	H	R
25	3.13	2.95	0.18	CN	CF ₃	OH	CH ₃	SO ₂	H	H	F	H	H	S
26	2.53	2.4	0.13	CN	CF ₃	OH	CH ₃	S	H	H	NH ₂	H	H	R
27	3.74	3.94	-0.2	CN	CF ₃	OH	CH ₃	S	H	H	NH ₂	H	H	S
28	2.07	2.3	-0.23	CN	CF ₃	OH	CH ₃	SO ₂	H	H	NH ₂	H	H	R
29	1.26	1.5	-0.24	CN	CF ₃	OH	CH ₃	S	H	H	NHCOCH ₃	H	H	R
30	3.71	3.09	0.62	CN	CF ₃	OH	CH ₃	S	H	H	NHCOCH ₃	H	H	S
31	2.59	2.46	0.13	CN	CF ₃	OH	CH ₃	S	H	H	N(COCH ₃) ₂	H	H	R
32	1.77	1.7	0.07	CN	CF ₃	OH	CH ₃	SO ₂	H	H	NHCOCH ₃	H	H	R
33	3.39	3.51	-0.12	CN	CF ₃	OH	CH ₃	S	H	H	N(COCH ₂ CH ₃) ₂	H	H	R

Table 1. (Continued)

cpds	actual	predicted	residual	R1	R2	R3	R4	X	R5	R6	R7	R8	R9	isomer
34	2.61	2	0.61	CN	CF ₃	OH	CH ₃	S	H	H	NHCOCH ₂ Br	H	H	R
35	3.74	4.06	-0.32	CN	CF ₃	OH	CH ₃	S	H	H	H	NHCOCH ₂ Br	H	R
36	0.8	1.17	-0.37	CN	CF ₃	OH	CH ₃	S	H	H	NHCOCH ₂ Cl	H	H	R
37	3.18	2.95	0.23	CN	CF ₃	OH	CH ₃	S	H	H	NHCOCH ₂ Cl	H	H	S
38	1.41	1.67	-0.26	CN	CF ₃	OH	CH ₃	SO ₂	H	H	NHCOCH ₂ Cl	H	H	R
39	2.2	2.17	0.03	CN	CF ₃	OH	CH ₃	SO ₂	H	H	H	NHCOCH ₂ Cl	H	R
40	2.95	2.53	0.42	CN	CF ₃	OH	CH ₃	S	H	H	H	NCS	H	R
41	3.55	3.97	-0.42	CN	CF ₃	OH	CH ₃	S	H	H	H	NCS	H	S
42	3.74	3.71	0.03	CN	CF ₃	OCOCH ₃	CH ₃	S	H	H	N(COCH ₂ CH ₃) ₂	H	H	R
43	2.32	2.07	0.25	NO ₂	CF ₃	OH	CH ₃	S	H	H	NH ₂	H	H	R
44	3.29	3.54	-0.25	NO ₂	CF ₃	OH	CH ₃	S	H	H	NH ₂	H	H	S
45	1.03	1.07	-0.04	NO ₂	CF ₃	OH	CH ₃	S	H	H	NHCOCH ₃	H	H	R
46	2.66	2.7	-0.04	NO ₂	CF ₃	OH	CH ₃	S	H	H	NHCOCH ₃	H	H	S
47	1.54	1.25	0.29	NO ₂	CF ₃	OH	CH ₃	SO ₂	H	H	NHCOCH ₃	H	H	R
48	0.98	1.14	-0.16	NO ₂	CF ₃	OH	CH ₃	S	H	H	NHCOCF ₃	H	H	R
49	1.36	1.1	0.26	NO ₂	CF ₃	OH	CH ₃	SO ₂	H	H	NHCOCF ₃	H	H	R
50	1.09	0.84	0.25	NO ₂	CF ₃	OH	CH ₃	S	H	H	NHCOCH ₂ Cl	H	H	R
51	2.56	2.7	-0.14	NO ₂	CF ₃	OH	CH ₃	S	H	H	NHCOCH ₂ Cl	H	H	S
52	1.41	1.71	-0.3	NO ₂	CF ₃	OH	CH ₃	SO ₂	H	H	NHCOCH ₂ Cl	H	H	R
53	2.94	3.06	-0.12	NO ₂	CF ₃	OH	CH ₃	SO ₂	H	H	NHCOCH ₂ Cl	H	H	S
54	2.27	2.01	0.26	NO ₂	CF ₃	OH	CH ₃	S	H	H	NHSO ₂ CH ₃	H	H	R
55	1.74	1.46	0.28	NO ₂	CF ₃	OH	CH ₃	SO ₂	H	H	NHSO ₂ CH ₃	H	H	R
56	1.55	1.52	0.03	NO ₂	CF ₃	OH	CH ₃	O	H	H	Cl	H	H	S
57	2.05	1.93	0.12	NO ₂	CF ₃	OH	CH ₃	O	H	H	I	H	H	S
58	1.63	1.76	-0.13	NO ₂	CF ₃	OH	CH ₃	O	H	H	Br	H	H	S
59	2.11	1.78	0.33	NO ₂	CF ₃	OH	CH ₃	O	H	H	CH ₃	H	H	S
60	1.47	1.43	0.04	NO ₂	CF ₃	OH	CH ₃	NH	H	H	F	H	H	S
61	0.99	1.19	-0.2	CN	I	OH	CH ₃	O	H	H	F	H	H	S
62	1.85	2.22	-0.37	CN	CF ₃	OH		S	H	H	NO ₂	H	H	R
63	1.87	2.1	-0.23	NO ₂	CF ₃	OH		S	H	H	NO ₂	H	H	R
64	1.27	1.17	0.1	NO ₂	CF ₃	OH		S	H	H	NHCOCH ₂ Cl	H	H	R
65	1.27	1.52	-0.25	NO ₂	CF ₃	OH		SO ₂	H	H	NHCOCH ₂ Cl	H	H	R
66	1.38	1.38	0.0011	CN	CF ₃	OH	CH ₃	O	H	H	NHCOCH ₃	H	H	R
67	3.09	3.24	-0.15	NO ₂	CF ₃	OH	CH ₃	O	H	H	NHCO ₂ C(CH ₃) ₃	H	H	R
68	1.23	1.43	-0.2	NO ₂	CF ₃	OH	CH ₃	O	H	H	NCS	H	H	R
69	2.68	2.52	0.16	NO ₂	CF ₃	OH		NH	H	H	NHCOCH ₃	H	H	R

Table 1. (Continued)

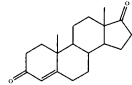
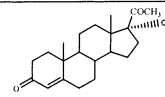
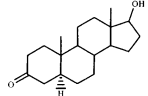
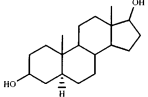
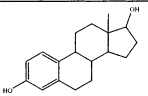
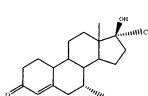
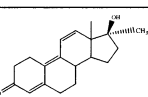
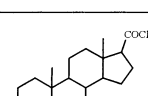
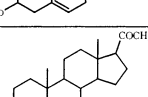
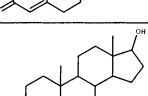
cpds	actual	predicted	residual	R1	R2	R3	R4	R5	X	Isomer
70	1.47	1.54	-0.07	H	CF ₃	CN	H	CH ₃	Br	R
71	3.64	3.55	0.09	NO ₂	H	NO ₂	H	CH ₃	Br	R
72	3.74	3.59	0.15	NO ₂	H	NO ₂	H	CH ₃	Br	S
73	2.48	2.62	-0.14	H	NO ₂	H	NO ₂	CH ₃	Br	R
74	3.74	3.57	0.17	H	H	CF ₃	H	CH ₃	H	N.A.
75	3.74	3.81	-0.07	CF ₃	H	H	H	CH ₃	H	N.A.
76	3.74	3.67	0.07	H	H	NO ₂	H	CH ₃	H	N.A.
77	3.74	3.63	0.11	NO ₂	H	H	H	CH ₃	H	N.A.
78	0.71	0.84	-0.13	H	CF ₃	NO ₂	H	CH ₃	Br	R
79	1.12	1.39	-0.27	H	CF ₃	NO ₂	H	CH ₂ CH ₃	CH ₃	N.A.
80	1.65	1.71	-0.06	H	CF ₃	NO ₂	H	CH ₃	CH ₃	R
81	1.58	1.78	-0.2	H	CF ₃	NO ₂	H	CH ₃	CH ₂ CH ₃	R
82	2.05	2.15	-0.1	H	CF ₃	NH ₂	H	CH ₃	Br	R



cpds	actual	predicted	residual	R1	R2	R3	R4	R5
83	1.11	0.94	0.17	CH ₃	CH ₃	H	H	H
84	0.54	0.82	-0.27	CH ₃	H	H	H	H
85	0.6	0.68	-0.08	CH ₂ CH ₃	H	H	H	H
86	1.74	1.45	0.29	CH(CH ₃) ₂	H	H	H	H
87	0.48	0.66	-0.19	CH ₃	H	H	H	CH ₃
88	1.06	0.76	0.3	CH ₃	H	H	H	CH ₂ CH ₃
89	0.65	1.05	-0.4	CH ₂ CH ₃	H	H	H	CH ₂ CH ₃
90	1.02	1.14	-0.12	CH ₂ CH ₃	H	H	H	CH ₃
91	1.04	1.01	0.04	CH ₃	H	H	CH ₃	H
92	0.88	1	-0.12	CH ₃	H	CH ₃	H	H
93	1.61	1.1	0.52	CH ₃	H	CH ₃	CH ₃	H
94	0.78	0.92	-0.15	CH ₂ CH ₃	H	H	CH ₃	H
95	0.74	0.76	-0.01	CH ₂ CH ₃	H	CH ₃	H	H
96	1.4	1.08	0.31	CH ₂ CH ₃	H	CH ₃	CH ₃	H
97	0.74	0.78	-0.04	CH ₃	H	H	CH ₂ CH ₃	H
98	0.88	0.9	-0.02	CH ₃	H	CH ₂ CH ₃	H	H
99	0.9	1.13	-0.23	CH ₂ CH ₃	H	H	CH ₂ CH ₃	H

Table 1. (Continued)

cpds	actual	predicted	residual	R1	R2	R3	R4	R5
100	1.16	1.01	0.15	CH ₂ CH ₃	H	CH ₂ CH ₃	H	H
101	1.32	1.53	-0.21	CH ₂ CH ₃	H	CH ₂ CH ₃	CH ₂ CH ₃	H
102	1.2	1.19	0.02	CH ₂ CH ₃	H	H	CH ₂ CH ₂ CH ₃	H

cpds	actual	predicted	residual	
Δ4-ANDROSTENEDIONE	1.7	1.52	0.18	
17α-HYDROXYPROGESTERONE	4.7	4.28	0.42	
DIHYDROTESTOSTERONE	0	-0.02	0.02	
ESTRADIOL	1.6	1.21	0.39	
ESTROGEN	1.34	1.48	-0.14	
MIBOLERONE	-0.7	-0.77	0.07	
METHYLTRIENOLONE	-0.7	-0.68	-0.02	
PREGNENOLONE	5	5.52	-0.52	
PROGESTERONE	4.7	4.32	0.38	
TESTOSTERONE	0.48	0.64	-0.16	

field indicated the donor component. Equal weights were assigned to these fields using the CoMFA standard scaling option. After the generation of field descriptors, a factor analysis was performed to help understand the clustering of AR ligands and to locate outliers. Outliers were defined as compounds that had a residual value of greater than 1. A leave-one-out (LOO) cross-validated PLS was performed to determine the cross-validated r^2 (q^2) and the optimum number of components in the model. This was used as the basis for conventional (non-cross-validated) PLS. The predicted pRBA versus experimental pRBA plot was constructed to further identify outliers. The process was repeated until no improvements in q^2 or no outliers could be identified. Corticosterone was identified as the only outlier according to this method. It was therefore excluded from the test set, and the CoMFA model was regenerated. Results from different descriptor fields were compared, and only the model with the highest q^2 was kept. A contour plot of $\text{stdev} \cdot \text{coeff}$ enclosing the highest 20%

value was created for each model. The "predict properties" command in the QSAR module was used to predict the pRBA of 10 test compounds.

Results

QSAR Validation. We used receptor binding and structural information from 122 ligands (Table 1) to identify important structure–activity relationships for AR interaction. RBA to DHT ranged from 0.02% to 20% for hydroxyflutamide analogues, from 0.02% to 16% for bicalutamide analogues, from 1.8% to 29% for tricyclic quinolinones, and from 0.002% to 501% for steroids. Molecular modeling and CoMFA were used in a ligand-based approach in an attempt to integrate unique aspects of ligand and receptor conformation and to better understand current and emerging data regarding

Table 2. Test Set

		actual	predicted	residual
a		1.55	1.34	0.21
b		1.64	1.88	-0.24
c		1.71	1.6	0.11
d		3.34	2.64	0.7
e		0.97	1.25	-0.28
f		3.55	3.73	-0.18
g		1.8	1.31	0.49
h		2	1.96	0.04
i		3.74	3.93	-0.19
j		1.08	0.62	0.46

AR structure and function. This approach resulted in a QSAR model that was highly predictive of RBA for the AR (Table 2). Conventional PLS analysis (Figure 2a,b and Table 3) showed that the actual and predicted pRBA were highly correlated ($r^2 = 0.974$) with residual values randomly scattered around zero. The q^2 was 0.593, corroborating the statistical validity of the QSAR model. Press was calculated as 0.737, which is the root mean predictive error sum of squares representing error in the cross-validated PLS correlation. The s value of 0.262 represents the root mean sum of the squares error for the conventional PLS correlation. The CoMFA model comprises 39% steric and 61% electrostatic interactions. Corticosterone was excluded from the model because of its high residual, suggesting that either the 11- or 21-hydroxyl group distinguishes it in terms of AR binding.

We used this model to predict the RBA of 10 randomly selected compounds. The RBA for these compounds ranged from 0.2% to 11%, indicating that they represented a reasonable range of RBA observed in our studies. Further, each of the structural classes of molecules was represented in this test set. A plot of actual vs predicted pRBA for these compounds demonstrated a correlation of 0.953 (Figure 2c), corroborative of the predictive capability of the QSAR model. Residual pRBA values were randomly scattered around zero.

Steric and Electrostatic Interactions with the Ligand. Coefficient contour maps were used to examine the steric and electrostatic fields associated with AR binding. We focused our analysis on compound **S-4**. Previous studies in our laboratory⁵ showed that this

ligand is a potent and tissue-selective nonsteroidal androgen. Figure 3 illustrates the contours at the 5% and 30% contour level. Polyhedra in each map surround all lattice points and indicate points in the ligand that are most strongly associated with observed differences in RBA. The contours of the steric map (van der Waals forces) are shown in yellow and green, while the contours of the electrostatic map are shown in red and blue. Greater binding is correlated with less bulk near yellow, more bulk near green, more negative charge near red, and more positive charge near blue.

Contour levels were adjusted to 5% to describe the most important regions for high binding affinity (Figure 3a) and to 30% to emphasize the contours present around the B-ring (Figure 3b). At the 5% contour level, a red (negative charge) polyhedra near the nitro group of compound **S-4** suggests that an H-bond acceptor is favored at the para position of the A-ring. Substitution of the nitro group to cyano group maintains high binding affinity, as would be expected. A green contour representing favorable hydrophobic interaction is seen at the 3-position of the A-ring. Compound **S-4** contains a trifluoromethyl group at this position; however, substitution to iodine and chlorine maintained high binding affinity. At the ketone in the bridge of compound **S-4**, a red contour is seen likely due to its conservation among all hydroxyflutamide and bicalutamide derivatives. A blue contour (favorable positive charge) at the hydroxyl group of the chiral carbon of compound **S-4** indicates its role as an H-bond donor. Last, distant to the oxygen linkage of compound **S-4** is a yellow contour representing steric hindrance at this region.

The 30% contour level exaggerates the polyhedra surrounding the B-ring of compound **S-4**. Since most molecules used in this model do not occupy this space, these contours are not present at high confidence. It is important to note that CoMFA using only bicalutamide derivatives (data not shown) could be used to display these interactions at lower contour levels. Both sides of the B-ring of compound **S-4** are surrounded by yellow contours, demonstrating the steric hindrance limiting the size of the substituent on the B-ring. The moderately sized acetamido group at the para position can be accommodated according to the model. Bulky para substituents and moderately sized meta and ortho substituents, however, would result in unfavorable steric interaction according to the model.

Ligand and AR Interaction. We used an overlap of the AR homology model and the CoMFA model at the 5% contour level to identify the most important residue interactions for high binding affinity (Figure 4a). The nitro group of compound **S-4** at the para position of the A-ring appears to interact with ARG752 and GLN711, taking advantage of this region of favorable negative charge identified in the CoMFA model. The trifluoromethyl group at the meta position of the A-ring of compound **S-4** is closely surrounded by VAL746 and MET742, suggesting that these residues are responsible for favorable hydrophobic interactions with the ligand. The ASN705 overlaps with a blue contour, suggesting that the O^{δ1} of ASN705 accepts the hydrogen from the chiral hydroxyl group of compound **S-4**. Only isomers allowing this interaction have high affinity for AR.

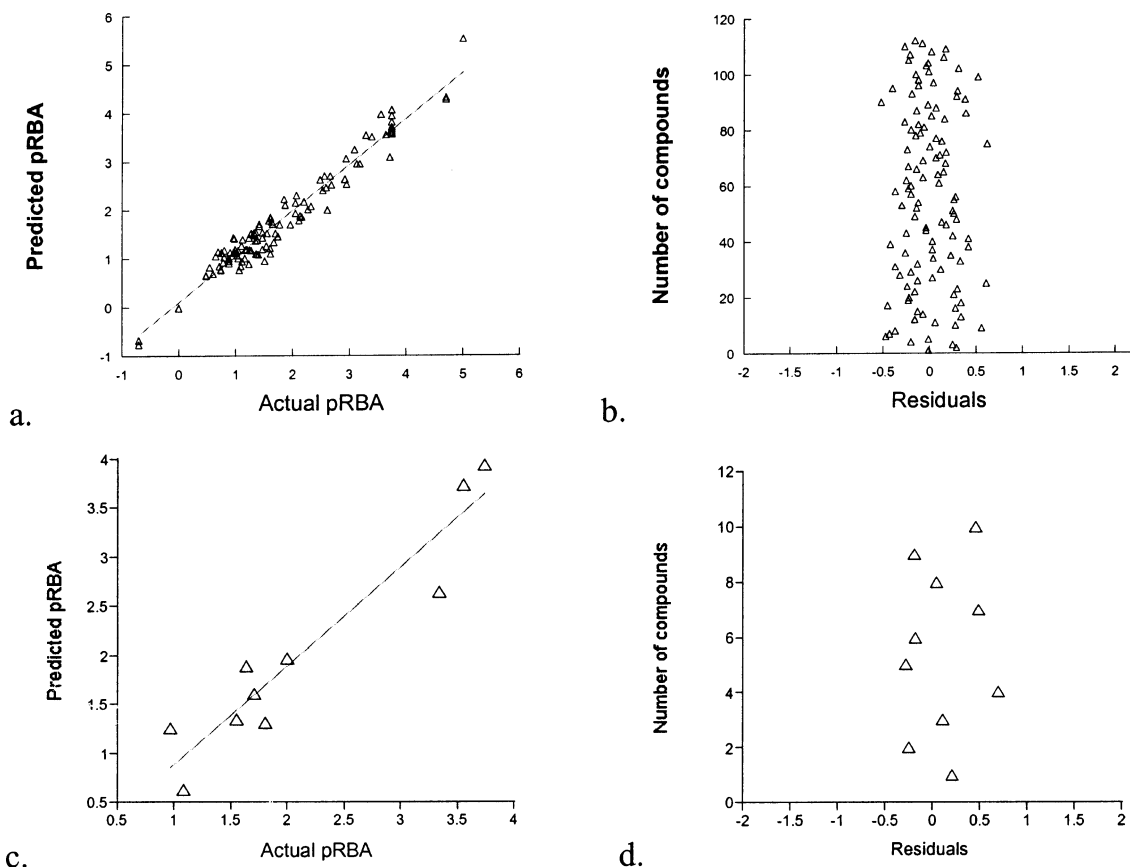


Figure 2. Plot of residuals and predictions. (a) A plot of the actual vs the predicted pRBA of the training set displays that compounds are closely scattered along the line with a slope of 0.949 and a correlation of 0.974 as calculated by KaleidaGraph 3.5. (b) Plot of the residuals following exclusion of outliers for the training set. (c) A plot of the actual vs the predicted pRBA for the test set demonstrates a correlation of 0.953. (d) Plot of the residuals from the test set.

Table 3. CoMFA Model Parameters

press	q^2	r^2	s	% steric	% electrostatic	exclusion
0.737	0.593	0.974	0.262	39	61	corticosterone

Steric hindrance seen distant to the oxygen linkage group of **S-4** overlaps with THR877.

The CoMFA model at a 30% contour level overlapped with the homology model aids in identification of amino acids bordering the B-ring (Figure 4b). According to the homology model, the B-ring lies in a subpocket bordered by MET780, CYS784, and MET787. A predominance of steric hindrance at this position is well supported by MET780 and MET787, which appear to border the acetamido group at the para position of compound **S-4**. These methionine residues also portray why moderately sized meta substituents result in poor binding affinity. Bulky residues at the para position also decrease binding affinity likely because of unfavorable steric interaction with the residues of this subpocket.

Discussion

The model described herein significantly advances our understanding of AR–ligand interactions. We incorporated a large number of chemically diverse ligands in an integrated approach using CoMFA and homology modeling. A previous CoMFA study by Waller et al.⁸ used the A-ring of steroids and hydroxyflutamide for alignment. However, our homology modeling and docking studies of the AR suggest that the A-ring of steroids

lies in a different region (Figure 5). The 3-keto group of DHT and the NO₂ group of bicalutamide-like derivatives share a similar space. At the 3-position, the trifluoromethyl group of **S-4** and the ring system of steroids overlap with the methyl group of tricyclic quinolinones in the region shown to favor hydrophobicity. The 17-OH of DHT as shown overlaps close to the linkage region of **S-4**; however, the hydroxyl group of **S-4** likely substitutes for its role as an H-bond donor. As visualized in this overlap, **S-4** occupies an additional space in the receptor not identifiable with steroids, hydroxyflutamide analogues, and tricyclic quinolinones.

Favorable negative charge on the ligand at the 4-position of the A-ring of bicalutamide and hydroxyflutamide analogues is supported by the location of ARG752 and GLN711 in the homology and crystal structures of the AR and by the activity of the receptor mutants. These residues act as H-bond donors in the crystal structure to the 3-keto group of R1881¹⁶ and in the homology model to the 4-nitro group of hydroxyflutamide.¹¹ Substitution of a cyano group at this position in bicalutamide derivatives maintains high AR binding affinity. Mutations of ARG752 (e.g., R752Q mutants observed in partial androgen insensitivity syndrome (PAIS)) result in loss of this interaction and poor binding affinity.¹⁶

The blue contour bordering the chiral hydroxyl group of bicalutamide analogues indicates the interaction with ASN705. This amino acid is also important for ligand discrimination and binding. N705S mutants observed

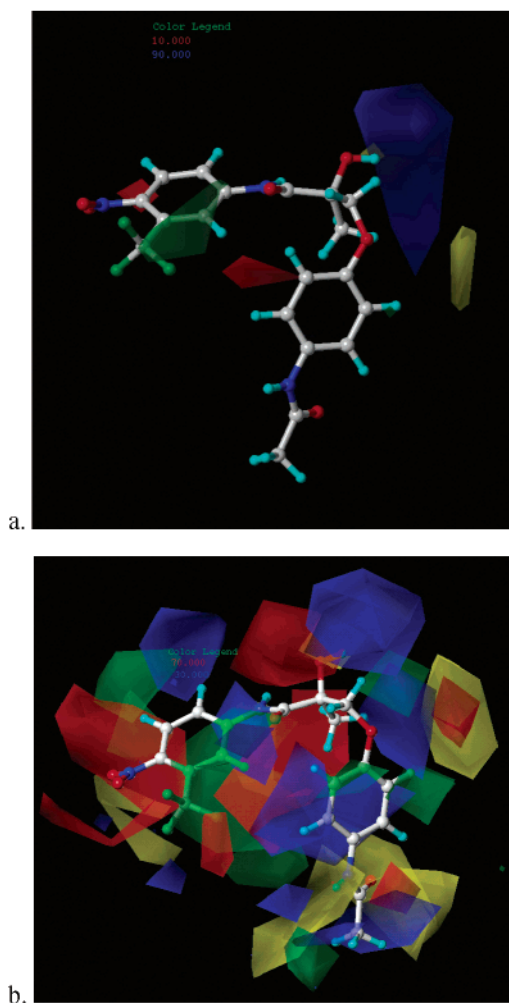


Figure 3. CoMFA model and ligand interactions. (a) The contour plot shown at 5% contour levels demonstrates the most important areas of ligand interactions. (b) The 30% contour levels allow the contours surrounding the B-ring of compound **S-4** to be visualized.

in complete androgen insensitivity syndrome (CAIS) corroborate the importance of this interaction.¹⁶ This area overlaps the 17-position of steroids.¹¹ Poujol et al.¹⁰ demonstrated near complete loss of the antagonist properties of hydroxyflutamide and bicalutamide in the N705A mutation, while RU486 and cyproterone acetate maintained the ability to repress transcriptional activation.¹⁰ This suggests that the interaction with hydroxyflutamide and bicalutamide derivatives is essential for binding. Further, the published crystal structure of the AR demonstrates that both ASN705 O^{δ1} and THR877 O^γ are hydrogen-bonded to R1881.¹⁶ The chiral hydroxyl group of bicalutamide and hydroxyflutamide does not interact with THR877 according to their docked conformation.

The yellow contour flanking the oxygen linkage group of **S-4** overlaps with THR877. Similar contours were seen in the model by Waller et al.⁸ This is an important discriminatory region in AR, disfavoring interactions with steroids (e.g., mineralocorticoids, estrogens, progestins) that contain bulkier groups at the 17-position. The T877A mutation as is observed in the LNCaP cancer cell line allows ligands with more bulky constituents at this position to bind the AR and results in less specificity in binding.¹⁷

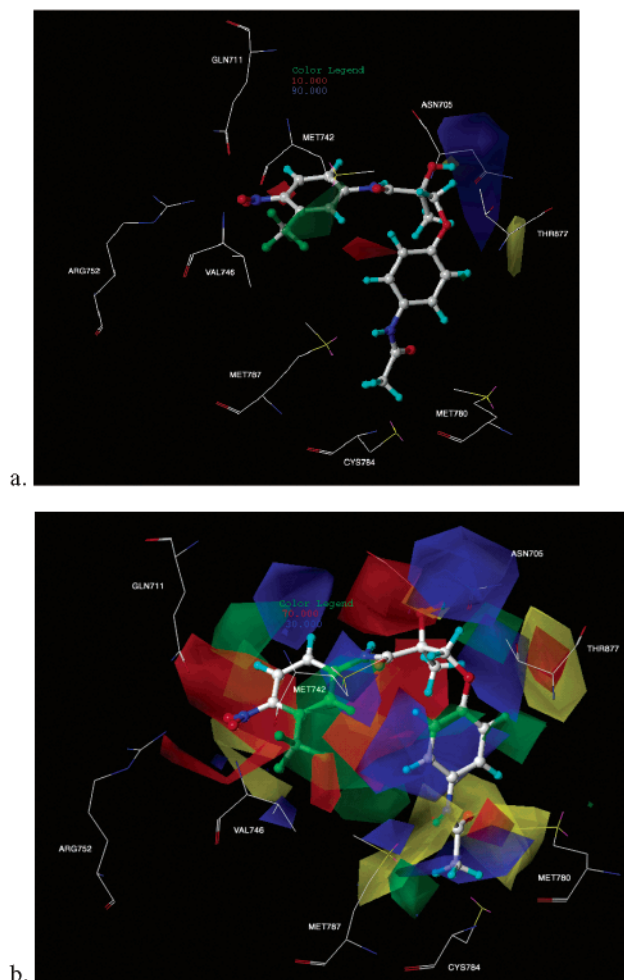


Figure 4. (a) CoMFA model displayed at 5% contour levels emphasizes critical residues for AR binding. (b) CoMFA model displayed at 30% contour levels emphasizes steric hindrance of the B-ring by MET780 and MET787.

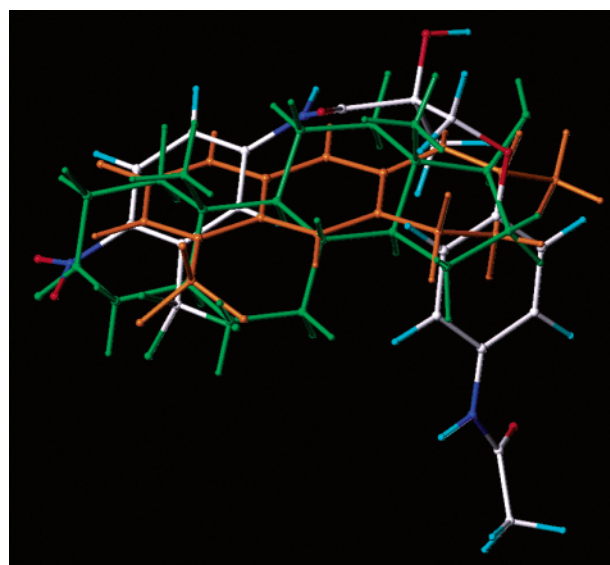


Figure 5. Overlap of **S-4** (colored by atom), DHT (green), and a tricyclic quinolinone from the docked conformation into the AR homology receptor. Compound **S-4** requires an additional space in the receptor compared with steroids and tricyclic quinolinones according to this model.

A hydrophobic favorable region seen in green at the 3-position of the A-ring in hydroxyflutamide and bi-

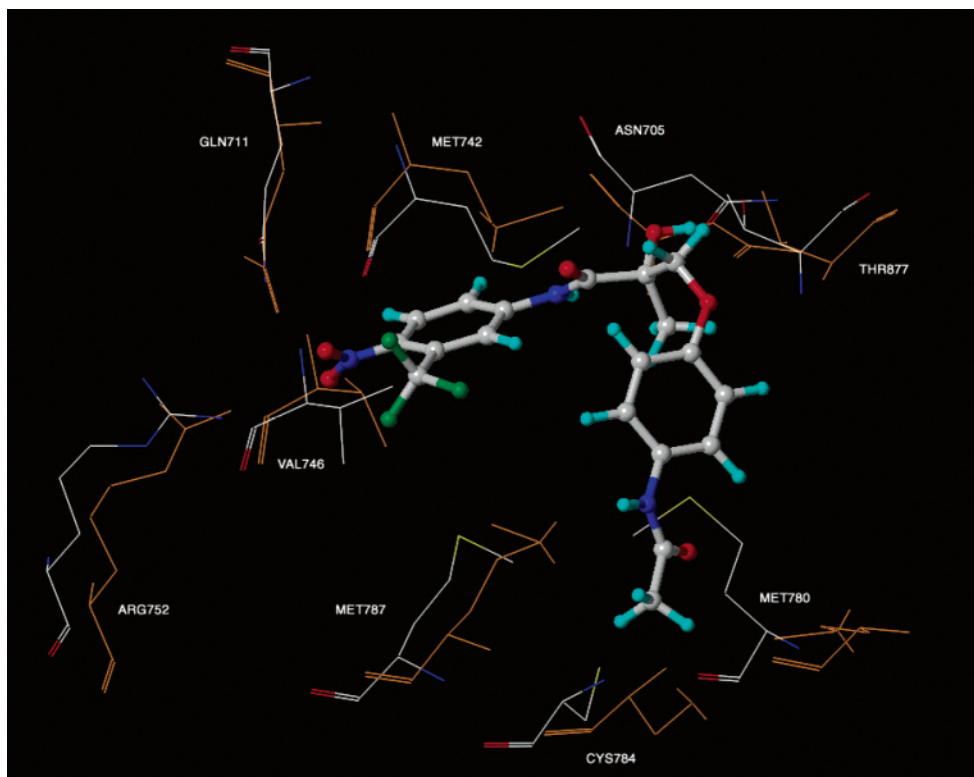


Figure 6. Overlap of AR crystal structure (colored by atom) and homology model (orange) binding pocket. The major difference between our homology model and the crystal structure is the position of MET780. Docking of bicalutamide analogues to the crystal structure will not occur because of this discrepancy.

calutamide analogues is present with high confidence as well. This position corresponds to carbons 5, 6, and 7 in steroids and to a methyl group in tricyclic quinolinones. In hydroxyflutamide and bicalutamide analogues, substitutions of chlorine and iodine for the trifluoromethyl group at this position sustained high binding affinity, while the substitution of a hydrogen resulted in poor binding affinity to the AR. VAL746 and MET742 appear to be responsible for this contour, as can be easily visualized in the overlap of the receptor and the CoMFA model. Mutations of these residues have been noted in PAIS, further supporting their importance.^{3,18}

The region bordering the B-ring of the bicalutamide analogues is an important aspect of the CoMFA model because of the variation between the homology model and the crystal structure in this region.^{11,16} Since only bicalutamide derivatives occupy the space of the bicalutamide B-ring, contours were set to 30% to exaggerate the contours in this area. Steric contours closely border the B-ring. These interactions appear to be mediated chiefly by MET787 and MET780 upon comparison to the homology model. Similar contours are seen at higher confidence levels when performing CoMFA with a data set incorporating only bicalutamide-like derivatives. Alignment to the crystal structure was not possible because of the tighter packing amino acids in this pocket in the crystal structure compared to the homology model. MET780 in the crystal structure appears to forbid the docked conformation of bicalutamide analogues (Figure 6). Functional groups such as the acetamido group can only be accommodated at the para position of the B-ring. The current study indicates

that this is due to steric hindrance from hydrophobic residues surrounding this area of the binding pocket.

It appears from the model that the highest affinity SARMS synthesized in our laboratory share all of the common electrostatic interactions with the highest binding steroidal androgens. Androgenic steroids such as DHT have more than 10 times the affinity for the AR compared to the SARMS developed in our laboratory. This is most likely from additional hydrophobic interactions with the AR binding pocket. Manipulating the hydrophobicity of certain regions of our SARMS to exploit some of these interactions might be key to increasing binding affinity. CoMFA will continue to play a role as such compounds are synthesized and tested.

Conclusion

The presented CoMFA model for the AR portrays a high predictive capacity for the binding of the AR ligand. Further, interactions identified using the CoMFA contours correspond appropriately to important amino acid contact sites identified in molecular modeling, the crystal structure, and mutational analysis. The use of bicalutamide analogues, which occupy a larger and apparently novel space in the LBD compared to other ligands (e.g., steroids, tricyclic quinolinones, and hydroxyflutamide analogues), provides insight to a previously unknown region in the binding pocket of the AR. Steric contours closely surrounding each side of the B-ring suggest that both MET787 and MET780 play an important role in preventing bulky B-ring substituents from binding the AR. This is the first study to suggest that these amino acids may be important for AR ligand binding. Ongoing studies in our laboratory are targeted

at elucidating the importance of these residues for the binding of bicalutamide-like derivatives and SARMs. We also showed that the CF₃ position of nonsteroidal AR ligands is a site that favors hydrophobic interaction. Last, other documented interactions, such as the hydrogen bond accepting at the 3-keto and hydrogen bond donating at the 17-OH group of steroidal androgens, support the CoMFA model. In summary, we used an integrated approach with CoMFA and homology modeling to investigate important interactions for AR ligand binding.

Acknowledgment. This work was partially supported by NIH Grant R01 DK59800. The authors are grateful to Dr. Craig Marhefka for providing the homology model of the AR and docking studies.

References

- (1) Dalton, J. T.; et al. Discovery of nonsteroidal androgens. *Biochem. Biophys. Res. Commun.* **1998**, *244* (1), 1–4.
- (2) Yin, D.; et al. Pharmacology, pharmacokinetics, and metabolism of acetothiolutamide, a novel nonsteroidal agonist for the androgen receptor. *J. Pharmacol. Exp. Ther.* **2003**, *304* (3), 1323–1333.
- (3) Negro-Vilar, A. Selective androgen receptor modulators (SARMs): a novel approach to androgen therapy for the new millennium. *J. Clin. Endocrinol. Metab.* **1999**, *84* (10), 3459–3462.
- (4) Yin, D.; et al. Key structural features of nonsteroidal ligands for binding and activation of the androgen receptor. *Mol. Pharmacol.* **2003**, *63* (1), 211–223.
- (5) Yin, D.; et al. Pharmacodynamics of selective androgen receptor modulators. *J. Pharmacol. Exp. Ther.* **2003**, *304* (3), 1334–1340.
- (6) Mukherjee, A.; et al. Affinity labeling of the androgen receptor with nonsteroidal chemoaffinity ligands. *Biochem. Pharmacol.* **1999**, *58* (8), 1259–1267.
- (7) He, Y.; et al. Novel nonsteroidal ligands with high binding affinity and potent functional activity for the androgen receptor. *Eur. J. Med. Chem.* **2002**, *37* (8), 619–634.
- (8) Waller, C. L.; et al. Three-dimensional quantitative structure–activity relationships for androgen receptor ligands. *Toxicol. Appl. Pharmacol.* **1996**, *137* (2), 219–227.
- (9) Gohlke, H.; Klebe, G. DrugScore meets CoMFA: adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein. *J. Med. Chem.* **2002**, *45* (19), 4153–4170.
- (10) Poujol, N.; et al. Specific recognition of androgens by their nuclear receptor. A structure–function study. *J. Biol. Chem.* **2000**, *275* (31), 24022–24031.
- (11) Marhefka, C. A.; et al. Homology modeling using multiple molecular dynamics simulations and docking studies of the human androgen receptor ligand binding domain bound to testosterone and nonsteroidal ligands. *J. Med. Chem.* **2001**, *44* (11), 1729–1740.
- (12) Zhi, L.; et al. Switching androgen receptor antagonists to agonists by modifying C-ring substituents on piperidino[3,2-*g*]quinolinone. *Bioorg. Med. Chem. Lett.* **1999**, *9* (7), 1009–1012.
- (13) Clark, M.; et al. Comparative Molecular Field Analysis (CoMFA) 2. Toward the use with 3D structural database. *Tetrahedron Comput. Methodol.* **1990**, *3*, 47–59.
- (14) Kroemer, R. T.; Hecht, P. Replacement of steric 6-12 potential-derived interaction energies by atom-based indicator variables in CoMFA leads to models of higher consistency. *J. Comput.-Aided Mol. Des.* **1995**, *9* (3), 205–212.
- (15) Bohacek, R. S.; McMartin, C. Definition and display of steric, hydrophobic, and hydrogen-bonding properties of ligand binding sites in proteins using Lee and Richards accessible surface: validation of a high-resolution graphical tool for drug design. *J. Med. Chem.* **1992**, *35* (10), 1671–1684.
- (16) Matias, P. M.; et al. Structural evidence for ligand specificity in the binding domain of the human androgen receptor. Implications for pathogenic gene mutations. *J. Biol. Chem.* **2000**, *275* (34), 26164–26171.
- (17) Veldscholte, J.; et al. The androgen receptor in LNCaP cells contains a mutation in the ligand binding domain which affects steroid binding characteristics and response to antiandrogens. *J. Steroid Biochem. Mol. Biol.* **1992**, *41* (3–8), 665–659.
- (18) Batch, J. A.; et al. Androgen receptor gene mutations identified by SSCP in fourteen subjects with androgen insensitivity syndrome. *Hum. Mol. Genet.* **1992**, *1* (7), 497–503.

JM0499007