JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

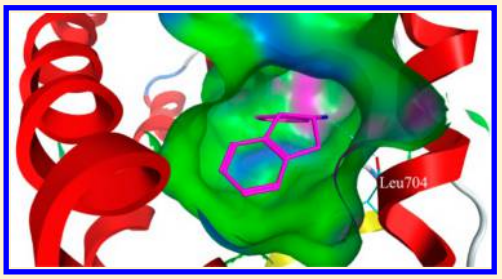
Identification of Novel Androgen Receptor Antagonists Using Structure- and Ligand-Based Methods

Huifang Li,*[,‡,†] Xin Ren,[‡,†] Eric Leblanc,[†] Kate Frewin,[†] Paul S. Rennie,[§,†] and Artem Cherkasov[§,†]

[†]Vancouver Prostate Centre, University of British Columbia, 2660 Oak Street, Vancouver, British Columbia V6H 3Z6, Canada

Supporting Information

ABSTRACT: Androgen receptor (AR) plays a critical role in the development and progression of prostate cancer (PCa). The AR hormone-binding site (HBS) is intensively studied and represents the target area for current antiandrogens including Bicalutamide and structurally related Enzalutamide. As resistance to antiandrogens invariably emerges in advanced prostate cancer, there exists a high medical need for the identification and development of novel AR antagonists of different chemotypes. Given the wealth of structural information on the AR in complex with a variety of ligands, we have applied an integrated structure- and ligand-based virtual screening methodology to identify novel AR antagonists. Virtual hits generated by a consensus voting approach were experimentally evaluated and resulted in the discovery of a number of structurally diverse submicromolar antagonists of the AR. In particular, one identified compound demonstrated anti-AR potency *in vitro* that is comparable to the clinically used Bicalutamide. These results set a ground for the development of novel classes of PCa drugs that are structurally different from current AR antagonists.



INTRODUCTION

Prostate cancer (PCa) is the second leading cause of male cancer death.[1] If the cancer is confined to the prostate, it is often curable by surgery or radiation treatment. However, in many cases, it has already spread to other tissues, particularly the bone, at diagnosis and usually requires some form of antimale sex hormone therapy to keep it in check. The antiandrogen such as Bicalutamide and Enzalutamide (initially called MDV3100) target the androgen receptor (AR), which is responsible for growth and progression of the disease. However, the emergence of recurrent, metastatic forms of castration-resistant/androgen-independent PCa continues to be a major challenge, with a median survival of only ~18 months.[2] The mechanism of resistance to current antiandrogens is unclear,[3] but for the first-generation antiandrogens like Hydroxyflutamide and Bicalutamide, it is known that certain mutations in the hormone binding site (HBS) of the AR (including T877A and W741L/C) will convert the drugs from receptor's antagonists to agonists.[4,5] The recently approved antiandrogen Enzalutamide has been developed as a second-generation AR antagonist aiming to solve the problem of resistance.[6] However, the resistance to Enzalutamide was already reported during the clinical trials.[7] Part of the problem is that all known antiandrogens, including Enzalutamide, share the same structural motif responsible for effective AR HBS binding which may attribute to occurring resistance. Therefore, more efforts are expected to discover AR antagonists possessing novel chemical scaffolds that may partially address the problem of resistance in PCa.

There is a wealth of structural information on a wild type AR complexed with diverse steroidal and nonsteroidal agonists bound to the AR HBS, including testosterone, dihydrotesterone (DHT), and R1881 among others.[8−10] The known antiandrogens such as Hydroxyflutamide and Bicalutamide (Figure S1) have only been cocrystallized with the mutated forms of the AR, interacting with the compounds in agonistic manner.[4,5] These structures provide valuable insights into essential interactions between the AR HBS and agonistic ligands. However, there is no structural data for antagonistic interactions with the AR and the mechanism of exact action of antiandrogens is still not clear. This fact imposes limitations on structure-based discovery and rational design of novel AR HBS antagonists (and, perhaps, contributed to limited structural variability among clinically used AR antagonists). On the other hand, a large number of diverse AR binders have also been reported in the literature which provides important information for ligand-based models using methods of quantitative structure−activity relationship (QSAR).

In this study, we have combined structure- and ligand-based strategies to identify novel AR antagonists of diverse chemical classes. First, the structure-based virtual screening was used to filter out compounds from public databases that are unlikely to have good binding affinity to the AR HBS. Although current docking programs can generate correct poses and capture essential interactions between ligands and the receptor, the corresponding scoring functions are not always sufficiently accurate. Thus, to complement the structure-based screening, we have developed a series of QSAR models that were subsequently used to select the final list of virtual hits. By taking

Received: October 24, 2012
Published: January 1, 2013

advantage of both methods, we were able to achieve a higher hit rate and provide wider structural coverage of a chemical space, which led us to identify a series of novel and potent AR antagonists which are structurally distinct from current antiandrogens.

MATERIALS AND METHODS

1. Structure-Based Virtual Screening. *Protein and Ligand Preparation.* The crystal structures were prepared using the Protein Preparation Wizard implemented in Maestro package from Schrödinger.[11] The solvent molecules were deleted, missing hydrogen atoms were added, and side chains were minimized using OPLS-2005 force field, and the receptor grid was defined using a 12 Å box centered on the crystallographic ligand. No constraints were applied, and all other adjustable settings were kept as default. A subset of ZINC database (ZINC 11)[12] was imported to Molecular Operating Environment (MOE) 2011.[13] All these structures were protonated/deprotonated by a washing process, added partial charges and minimized with the MMFF94x force field to a gradient of 0.0001 kcal/mol Å. Duplicate compounds in the database were removed using the db_unique.svl module from the MOE.

Molecular Docking-Based Virtual Screening. Two docking programs - Glide[14] and eHiTs[15] - were used for virtual high-throughput screening implemented on a Sun Grid Engine cluster. The Glide SP mode was initially used for filtering out compounds with low docking scores (>−8). The active site was defined from the coordination of crystallographic ligand (PDB code: 2PNU.pdb), using the default settings. To avoid any bias in docking programs, compounds with favorable Glide scores were subsequently docked by eHiTs, and compounds with eHiTs score higher than a cutoff value (−3) were removed. The root-mean-square deviation (RMSD) of the docked poses generated by Glide and eHiTs were then calculated to keep compounds with consistent docked poses. Compounds with high RMSD values (>2 Å) were removed, and Lipinski's rule[16] was applied to the remaining compounds, which were advanced to the next stage of analysis.

2. Ligand-Based Virtual Screening. *Data Set.* The success of any QSAR model depends on accurate and clean training data,[17] proper representative descriptor selection methods,[18] suitable statistical methods,[17] and, most critically, both internal and external validation of resulting models.[19]

A total of 1220 chemical structures were taken from the literature and public databases, and reported activities were converted into binary format. Part of the set consisted of 625 known AR binders reported previously,[20] including 394 chemicals measured by a Danish lab with activities in binary form, and 231 chemicals with IC_{25} values which were converted into binary format. In particular, a given chemical exhibiting IC_{25} value lower than the test concentration (10 μM) was classified as active or inactive if the IC_{25} value was above 10 μM or the cytotoxicity (IC_{50}) was over 3 μM. Another set of 595 known AR binders was assembled from public databases such as CHEMBL,[21] BindingDB,[22] DrugBank,[23] and ChemSpider.[24] Among those, chemicals with an IC_{50} value lower than 20 μM were defined as active AR inhibitors. Additionally, the external set (89 molecules) of known AR inhibitors previously used in the literature[20] was also considered to validate the developed QSAR models. The structures of the studied chemicals were optimized by the MMFF94x force field implemented in MOE program.

Descriptor Selection. INDUCTIVE descriptors,[25−27] DRAGON 5.5 descriptors,[28] and MOE descriptors were calculated for ZINC compounds to be screened. To avoid possible cross-correlation among independent variables, the pairwise correlation among all the QSAR parameters were computed, and descriptors with either low variability (value <0.1) or high correlation ($R \geq 0.95$) were removed. Besides, the recursive feature elimination (RFE) routine[29] based on the support vector machines (SVM) method was used for selecting molecular descriptors associated with AR antagonist activity. The RFE is an iterative procedure for backward feature elimination, which can be executed in the WEKA package,[30] where the descriptors were normalized by default.

Modeling Methods. Nine classification methods were used to build QSAR models for AR binders, including k-Nearest Neighbors (kNN),[31] Local Lazy method (lazy IB1),[32] Alternating Decision Tree (ADTree),[33] Artificial Neural Network (ANN),[34] K-star method,[35] Bagging method,[36] LogitBoost,[37] Decorate,[38] and Random Forest.[39] These methods were implemented within the WEKA in default setting.

Evaluation of QSAR Models. Validation is the process by which reliability and relevance of a procedure are established for a specific purpose.[40] For validation of QSAR models, three strategies were adopted[41,42] including 10-fold cross-validation, validation on external sets, and evaluation of the corresponding applicability domain (AD) criteria.

3. In Vitro and Cell-Based Assay. *Chemical Sources.* Selected chemicals were purchased from the established suppliers, including Asinex, Chembridge, and Sigma. The identity and purity of all chemicals were confirmed by mass spectroscopy (MS) and liquid chromatography−tandem mass spectrometry (LC−MS/MS).

Androgen Displacement Assay. Androgen displacement was assessed with the Polar Screen Androgen Receptor Competitor Green Assay Kit as per the instructions of the manufacturer.[43]

eGFP Cellular Transcription Assay. The AR transcriptional activity was assayed as previously described.[44] Briefly, stably transfected eGFP-expressing LNCaP human prostate cancer cells (LN-ARR2PB-eGFP) containing an androgen-responsive probasin-derived promoter (ARR2PB) were grown in phenol-red-free RPMI 1640 supplemented with 5% CSS. After 5 days, the cells were plated into a 96-well plate (35,000 cells/well) with 0.1 nM R1881 and increasing concentrations (0−100 μM) of chemicals. The cells were incubated for 3 days, and the fluorescence was then measured (excitation, 485 nm; emission, 535 nm).

Prostate-Specific Antigen (PSA) Assay. The evaluation of PSA secreted into the media was performed in parallel to the eGFP assay using the same plates (see description above). After the cells were incubated for 3 days, and following the eGFP reading, 150 μL of the media was taken from each well and added to 150 μL of PBS. PSA levels were then evaluated using Cobas e 411 analyzer instrument (Roche Diagnostics) according to the manufacturer's instructions.

Cell Proliferation Assay. LNCaP cells and PC3 cells were plated at 3,000 cells per well in RPMI 1640 containing 5% CSS in a 96-well plate and treated with 0.1 nM R1881 and chemicals (0−12 μM) for 96 h. After 4 days of treatment, cell density was measured using the 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium assay ac-

dx.doi.org/10.1021/ci300514v | J. Chem. Inf. Model. 2013, 53, 123−130

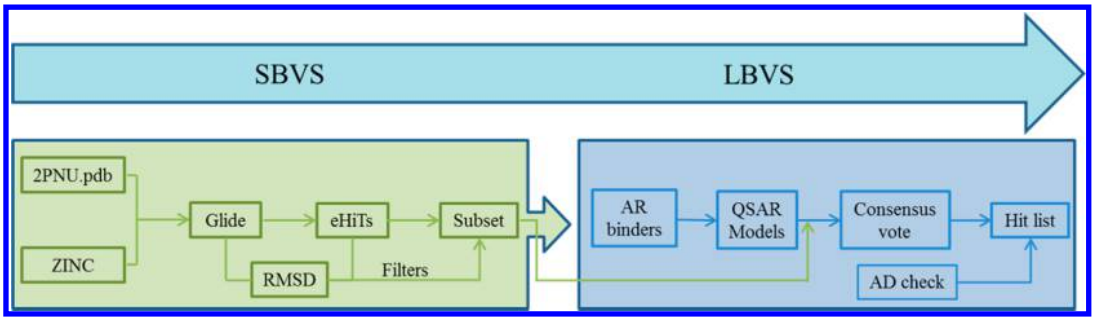


Figure 1. The pipeline of the discovery process for AR antagonists.

Table 1. RMSD Values between the Docked and Bound Conformations

Glide	2AM9	3L3X	2PNU	1Z95	eHiTs	2AM9	3L3X	2PNU	1Z95
testosterone	0.79	0.22	0.48	3.14	testosterone	0.70	1.04	3.35	1.11
DHT	0.66	0.27	0.55	1.08	DHT	0.92	0.75	0.56	1.11
EM5744	NA[b]	NA[b]	0.28	1.5	EM5744	7.56	4.63	1.00	1.85
Bic[a]	4.81	NA[b]	1.29	0.34	Bic[a]	3.08	4.15	0.65	2.21

[a]Represents Bicalutamide. [b]NA: ligands were skipped by glide sp docking program.

Table 2. Enrichment Factor (EF) Values of the Screening of DUD Using Two Crystal Structures

	PDB code							
	2AM9		3L3X		1Z95		2PNU	
	Glide[a]	eHiTs[a]	Glide[a]	eHiTs[a]	Glide[a]	eHiTs[a]	Glide[a]	eHiTs[a]
$EF_{1\%}$	26	34.7	26	32.2	11	35.9	26	35.9
$EF_{20\%}$	3.6	4.7	3.6	4.7	2.8	4.6	4.1	5.0

[a]Docking program.

cording to the manufacturer's protocol (CellTiter 961 Aqueous One Solution Reagent, Promega).

RESULTS AND DISCUSSIONS

Structure- and ligand-based virtual screening approaches have previously been applied to independently identify various AR antagonists.[20,45] It was shown that virtual screening data fusion using structure- and ligand-based methods is a more effective tool, with the potential to outperform a single method.[46] Herein, a combination of structure- and ligand-based virtual screening methodologies (Figure 1) was used in this study to achieve a higher hit rate and to provide wider structural coverage of a chemical space.

1. Structure-Based Virtual Screening. *Self- and Cross-Docking.* The AR HBS is a flexible site, and it can accommodate compounds of very different sizes and chemotypes. As there are many crystal structures of AR C-terminal ligand binding domain (LBD) available, self- and cross-docking routines were used to evaluate AR structures that are most representative for virtual screening. Based on the volume of the AR HBS cavity and types of bound ligands, four high-resolution crystals (2AM9.pdb, 3L3X.pdb, 2PNU.pdb, and 1Z95.pdb)[4,10,47,48] were selected. Those PDB entries corresponded to the AR in complex with testosterone, DHT, bulky steroidal derivative EM5744, and nonsteroidal AR antagonist Bicalutamide, respectively. The RMSD values between the docked poses and the bound conformation of the crystallographic ligands were calculated and reported in Table 1 as performed in other papers.[49] The bound conformations of smaller native ligands in 2AM9 and 3L3X structures could be well reproduced by docking, while bulkier EM5744 and Bicalutamide could not be accurately fitted into smaller 2AM9 and 3L3X. However, the self- and cross-docking of these ligands into 2PNU and 1Z95 were accurate and reproducible in both docking programs (see Table 1).

Prescreening with the Directory of Useful Decoys (DUD). The four crystal complexes were also evaluated by prescreening with the directory of useful decoys (DUD) data set.[50] The docking enrichment factor (EF) established for known AR binders from the DUD were similar and suitable in cases of 2AM9 and 3L3X structures (Table 2), while docking with 1Z95 target resulted in low enrichment. The AR HBS structure corresponding to 2PNU protein database entry yielded the highest EF indicating its better ability to differentiate AR binders. Therefore, consistent with the self- and cross-docking evaluations, the crystal structure 2PNU was selected for virtual screening.

Structure-Based Virtual Screening. Molecular docking was initially performed with the Glide SP program using a subset of 3 million compounds from ZINC database. The Glide SP docking score cutoff (>−8) was used to identify low- or no affinity ligands to be discarded. The compounds left (702,151) were further redocked into 2PNU using eHiTs, and only compounds with eHiTs score less than −3 (571,666) were kept. The docked poses were compared with those from the Glide experiment, and compounds with low RMSD values (<2 Å) were kept. These compounds were further filtered following Lipinski's rule[16] to keep compounds with potential good bioavailability. Finally, 136,595 compounds were advanced into ligand-based virtual screening.

2. Ligand-Based Virtual Screening. *The Development and Validation of QSAR Models.* QSAR models for AR

dx.doi.org/10.1021/ci300514v | J. Chem. Inf. Model. 2013, 53, 123−130

antagonists were generated using the WEKA software, which is a collection of machine learning algorithms typically used in data mining studies. In addition, WEKA includes functionalities for data preprocessing, classification, regression, clustering, association, and visualization. It is also well-suited for developing new machine learning schemes.

For all the studied compounds, their anti-AR activity values were transformed into binary form, and three sets of molecular descriptors including DRAGON, INDUCTIVE, and MOE were calculated. The RFE method[29] was applied to select the most effective descriptors, while the descriptors with low variance (<0.1) or correlated (the corresponding $R \geq 0.95$) were removed. Finally, the top 30 descriptors were selected, and from them we sampled anywhere from 5 to 30 descriptors to generate QSAR models (see Supporting Information Tables S1 and S2). Further, various machine learning methods were applied including kNN, lazy IB1, ADTree, ANN, K-star, and Bagging as well as LogitBoost, Decorate, and Random Forest to relate the selected QSAR descriptors of compounds to their binary activity parameters. The optimal number of QSAR descriptors for model generation was established in the range of 25−30 parameters (Figure 2).

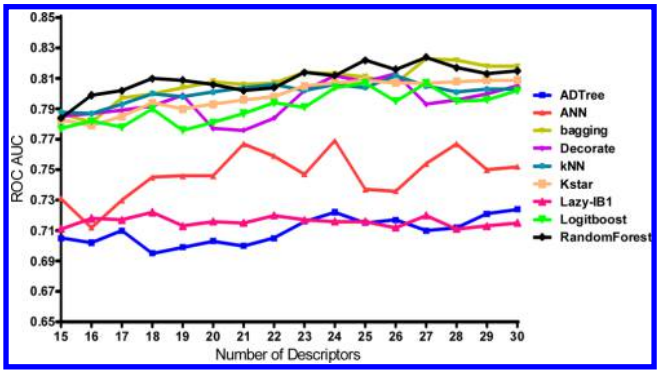


Figure 2. Correlation of ROC AUC and the number of descriptors of the training set.

To validate the performance of the developed models, an external test set was also used as an unbiased benchmark (Table 3). The developed QSAR models performed well on the external test set, with the lowest ROC AUC around 0.7, and some models like ANN could predict the external test set correctly with an AUC value close to 1.

QSAR-Based Virtual Screening. With the optimized descriptor number and highest ROC AUC values, 9 selected QSAR models were applied to screen ZINC database preprocessed with the docking as described above. In particular, 136,595 molecules emerged from Glide and eHiTs virtual screening were processed with 9 pretrained QSAR models.

Consensus Voting. A consensus voting protocol was implemented to rank the QSAR results and select hit compounds. In particular, if a given molecule was predicted as active or inactive by one model, it will receive a binary 1.0 vote. The final cumulative vote, with the maximum possible value of 9, was then used to rank the processed ZINC entries. This protocol was also evaluated by applying to the external test set of 89 known AR binders, and the corresponding consensus AUC value of 0.837 (Table 3) turned out to be higher than AUC from most of individual QSAR models, suggesting the suitability of the consensus protocol. Thus, the consensus voting was used for the screening, and on the basis of the cumulative count, the data set was shrunk to 2198 compounds with a consensus score of 8 and 9. Among these chemicals, half of these compounds (1202 entries) belong to steroidal category and were excluded. By discarding broken structures, a list of 800 compounds was retained for virtual hit selection.

Preliminary Evaluation of Selected Compounds in Ligand- and Structure-Based Manner. The range of descriptors selected in each model was compared to evaluate the applicability domain (AD) criteria for the developed models. This procedure estimated the reliability of predictions for the top-voted hits. Values of each descriptor for every top-voted chemical were checked to fit in the training descriptors value range. From this analysis we verified that all of the selected chemicals fit into suitable AD, with the percentage of reliable predictions being 100%. Following the workflow of structure- and ligand-based virtual screening, the list of virtual hits was further checked through receptor−ligand interaction. The docked poses of these selected compounds were visualized to retain compounds forming favorable interactions with the receptor and exclude those showing steric clashes with the receptor, and, finally, 37 diversified compounds were selected for further bioactivity testing (Figure S2).

3. Cell-Based and *in Vitro* Evaluation of Virtual Hit Compounds. Compounds that bind to the AR HBS can be identified by their capability to compete with a fluorescently tagged androgen ligand using an androgen displacement assay kit (Polar Screen). An initial screening at 50 μM with our selected 37 compounds identified 9 structurally diverse chemicals which demonstrated significant androgen displacement. The AR binding affinity was further evaluated by measuring the IC_{50} of these 9 chemicals. All of them are active compounds with IC_{50} less than 20 μM, and, particularly, 6 compounds demonstrated an IC_{50} under 5 μM (Table 4).

Subsequently, we evaluated the effect of these compounds on the transcriptional activity of AR using LNCaP cells, which stably transfected with an androgen-responsive probasin-derived promoter fused to an eGFP reporter (LN-ARR2PB-eGFP). Out of these 9 chemicals, 7 demonstrated significant transcriptional inhibition (IC_{50} < 5 μM) (Table 4), and one compound (VPC-12060) showed low submicromolar activity (IC_{50} = 0.378 μM) which is comparable to clinical antiandrogen Bicalutamide (IC_{50} = 0.382 μM, Figure 3).

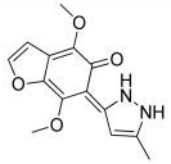
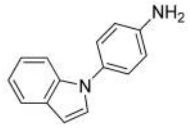
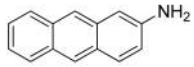
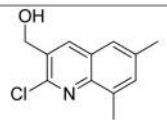
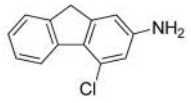
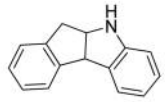
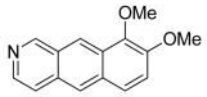
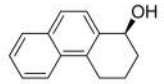
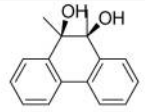
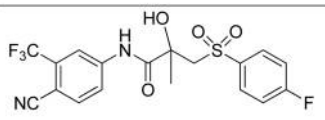
To confirm the inhibition of AR transcription by these compounds, we also evaluated the amount of PSA protein secreted in the media by eGFP LNCaP cells. PSA is a protein

Table 3. Validation Using the External Test Set

	ADTree	ANN	Bagging	Decorate	kNN	K-star	Lazy-IB1	Logit-boost	Random Forest	Consensus
AUC (ROC)	0.78	0.862	0.817	0.79	0.707	0.707	0.714	0.768	0.732	0.837
true positive fraction	0.735	1	0.915	0.795	0.807	0.911	0.764	0.765	0.881	0.875
true negative fraction	0.714	0.793	0.714	0.651	0.714	0.857	0.857	0.714	0.857	0.845
accuracy	0.776	0.809	0.809	0.738	0.708	0.719	0.685	0.764	0.742	0.82

dx.doi.org/10.1021/ci300514v | J. Chem. Inf. Model. 2013, 53, 123−130

Table 4. Structures and Activities of Identified Active Compounds

VPC-ID	Structure	Androgen Displacement IC_{50} (μM)	eGFP IC_{50} (μM)	PSA IC_{50} (μM)	Tanimoto Coefficient
12002		12.29	2.04	N/A	0.27
12007		1.84	1.06	0.703	0.18
12051		1.23	1.95	2.299	0.18
12052		3.25	1.76	1.063	0.25
12058		2.02	2.54	2.1	0.23
12060		3.43	0.38	0.173	0.18
12061		3.49	1.92	2.569	0.19
12066		18.63	9.35	N/A	0.20
12068		5.16	16.18	N/A	0.28
Bic		**1.06**	**0.38**	**0.131**	**1**

that is naturally transcribed under the control of the AR using the probasin promoter. A concentration-dependent reduction in PSA secretion was induced by these compounds, with IC_{50}s correlated with the eGFP inhibition. For VPC-12060, the PSA level (PSA IC_{50} = 0.138 μM) is comparable to that of Bicalutamide (PSA IC_{50} = 0.131 μM, Figure 3). Other active compounds also showed low PSA level, together with high AR transcriptional inhibition and strong androgen displacement potency (Table 4).

To determine whether the identified AR inhibitors affect tumor cell proliferation, we have carried out the cell proliferation assay with two prostate cancer cell lines: AR-positive LNCaP cells and AR-negative PC3 prostate cancer cells. The lead chemical VPC-12060 can antagonize the proliferative effect of the androgen R1881 without any significant effect on the growth of AR-negative PC3 cells at the same concentrations, which suggests that the inhibition of proliferation observed in AR-positive LNCaP cells is mediated through antagonism of AR.

dx.doi.org/10.1021/ci300514v | J. Chem. Inf. Model. 2013, 53, 123−130

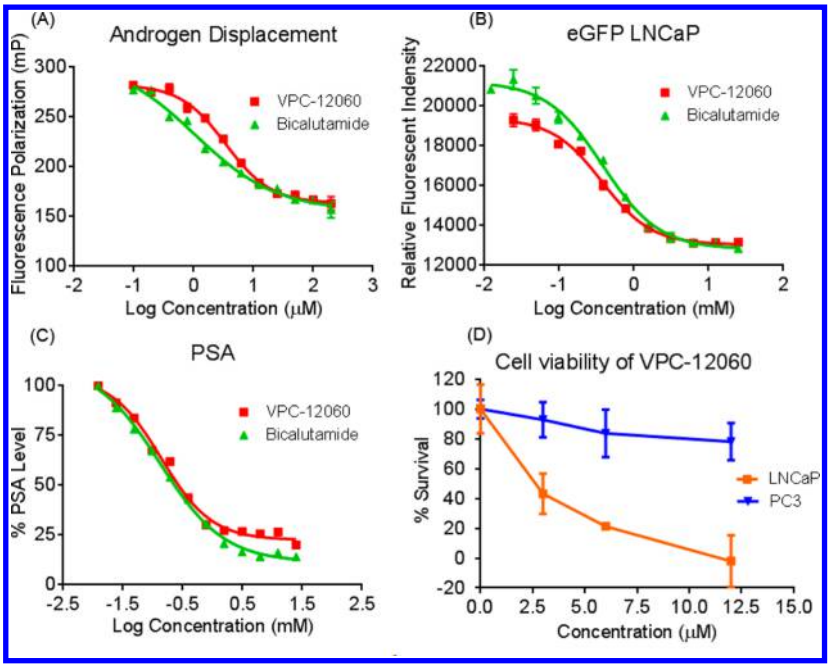


Figure 3. Dose–response curve of VPC-12060 and Bicalutamide for (A) androgen displacement activity; (B) AR transcriptional inhibition; (C) PSA level; (D) cell viability of VPC-12060.

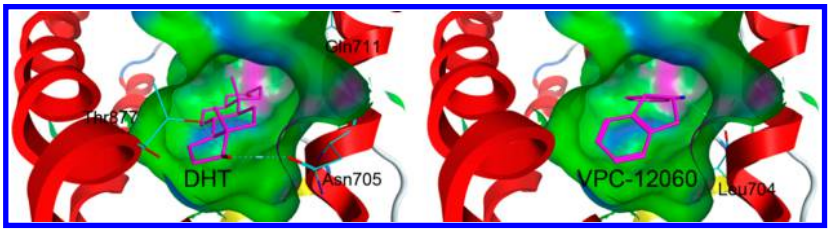


Figure 4. The molecular binding modes of DHT (left) and VPC-12060 in the AR HBS (right).

4. Molecular Docking Analysis of Hit Compounds. The identified nonsteroidal AR binders are structurally distinct from the current AR antagonists, with low structural similarity demonstrated by the Tanimoto coefficient (see Table 4), but still demonstrated sufficient binding to the AR HBS. To understand the essential features responsible for the potency of these compounds, the receptor–ligand interaction at the binding site was examined based on the docked poses from the structure-based method. The AR HBS is well-characterized as a hydrophobic cavity that forms strong hydrophobic interactions with a steroidal core of androgens. Meanwhile, there are two polar patches Arg752 and Gln711 at one end and Thr877 and Asn705 at the other end of the site. The crystal structures of mutated, agonistic AR variants (T877A and W741L/C) containing antiandrogens Flutamide and Bicalutamide, respectively, demonstrate similar binding modes of the ligands. In that prospective, the identified compounds all possess hydrophobic scaffolds with one or two polar groups attached, and the docked poses of these compounds adopt similar orientations in the AR HBS compared to crystallographic ligands (Figure 4). As observed from the figure, polar groups of VPC-12060 form hydrogen bonds with Leu704. For the majority of the identified compounds there is only one polar group capable of forming hydrogen bond interactions with the AR HBS residues, which eliminates the option of anchoring the compound to Arg752 or Gln711 at the other side of the site. The binding modes of these compounds suggest that hydrophobic interactions may be essential for the ligand coordination, while hydrogen bonding may not play a determining role.

CONCLUSION

In this study, a combination of structure- and ligand-based virtual screening strategies was applied to identify novel androgen receptor antagonists belonging to novel and diverse chemical series that are different from the current PCa drugs. The adopted combined *in silico* screening strategy revealed a sufficiently high hit rate of the discovery and identified active compounds demonstrating strong androgen displacement ability and high AR inhibitory potency. Several of the identified hits exhibited strong inhibitory effect on tumor cell proliferation and could be considered as a good starting point for further optimization and development of novel classes of AR antagonists.

ASSOCIATED CONTENT

Supporting Information

Descriptors used in this study for building QSAR models, the ROC AUC values of different methods, chemical structures of antiandrogens, and selected virtual hits. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: 604-875-4111. Fax: 604-875-5654. E-mail: hli@prostatecentre.com.

Author Contributions

‡Authors contributed equally to this work.

Author Contributions

§Dr. Cherkasov's and Dr. Rennie's laboratories made equal contributions to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by a Proof-of Principle CIHR grant and the PC-STAR Project, which is funded by Prostate Cancer Canada with the support of Safeway. The authors thank Jeffrey Leong and Dennis Ma for helping the *in vitro* work and Hans Adomat for the purity and identity measurement.

REFERENCES

(1) Siegel, R.; Ward, E.; Brawley, O.; Jemal, A. Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *Ca-Cancer J. Clin.* **2011**, *61*, 212−236.

(2) Scher, H. I.; Sawyers, C. L. Biology of progressive, castration-resistant prostate cancer: directed therapies targeting the androgen-receptor signaling axis. *J. Clin. Oncol.* **2005**, *23*, 8253−8261.

(3) Yamaoka, M.; Hara, T.; Kusaka, M. Overcoming persistent dependency on androgen signaling after progression to castration-resistant prostate cancer. *Clin. Cancer Res.* **2010**, *16*, 4319−4324.

(4) Bohl, C. E.; Gao, W.; Miller, D. D.; Bell, C. E.; Dalton, J. T. Structural basis for antagonism and resistance of bicalutamide in prostate cancer. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6201−6206.

(5) Bohl, C. E.; Miller, D. D.; Chen, J.; Bell, C. E.; Dalton, J. T. Structural basis for accommodation of nonsteroidal ligands in the androgen receptor. *J. Biol. Chem.* **2005**, *280*, 37747−37754.

(6) Tran, C.; Ouk, S.; Clegg, N. J.; Chen, Y.; Watson, P. A.; Arora, V.; Wongvipat, J.; Smith-Jones, P. M.; Yoo, D.; Kwon, A.; Wasielewska, T.; Welsbie, D.; Chen, C. D.; Higano, C. S.; Beer, T. M.; Hung, D. T.; Scher, H. I.; Jung, M. E.; Sawyers, C. L. Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science* **2009**, *324*, 787−790.

(7) Shen, H. C.; Balk, S. P. Development of androgen receptor antagonists with promising activity in castration-resistant prostate cancer. *Cancer Cell* **2009**, *15*, 461−463.

(8) Matias, P. M.; Donner, P.; Coelho, R.; Thomaz, M.; Peixoto, C.; Macedo, S.; Otto, N.; Joschko, S.; Scholz, P.; Wegg, A.; Basler, S.; Schafer, M.; Egner, U.; Carrondo, M. A. Structural evidence for ligand specificity in the binding domain of the human androgen receptor - implications for pathogenic gene mutations. *J. Biol. Chem.* **2000**, *275*, 26164−26171.

(9) Sack, J. S.; Kish, K. F.; Wang, C.; Attar, R. M.; Kiefer, S. E.; An, Y.; Wu, G. Y.; Scheffler, J. E.; Salvati, M. E.; Krystek, S. R., Jr.; Weinmann, R.; Einspahr, H. M. Crystallographic structures of the ligand-binding domains of the androgen receptor and its T877A mutant complexed with the natural agonist dihydrotestosterone. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 4904−4909.

(10) De Jesus-Tran, K. P.; Cote, P. L.; Cantin, L.; Blanchet, J.; Labrie, F.; Breton, R. Comparison of crystal structures of human androgen receptor ligand-binding domain complexed with various agonists reveals molecular determinants responsible for binding affinity. *Protein Sci.* **2006**, *15*, 987−999.

(11) Maestro, version 9.3; Schrödinger, LLC: Portland, US, 2012.

(12) Irwin, J. J.; Shoichet, B. K. ZINC - a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177−182.

(13) MOE, version 2011; Chemical Computing Group, Inc.: Montreal, Canada, 2012.

(14) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(15) Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, S. B.; Johnson, A. P. eHiTS: a new fast, exhaustive flexible ligand docking system. *J. Mol. Graphics Modell.* **2007**, *26*, 198−212.

(16) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **2001**, *46*, 3−26.

(17) Roy, P. P.; Leonard, J. T.; Roy, K. Exploring the impact of size of training sets for the development of predictive QSAR models. *Chemom. Intell. Lab. Syst.* **2008**, *90*, 31−42.

(18) Roy, P. P.; Roy, K. On some aspects of variable selection for partial least squares regression models. *QSAR Comb. Sci.* **2008**, *27*, 302−313.

(19) Rao, C. R.; Wu, Y. Linear model selection by cross-validation. *J. Stat. Plann. Inference* **2005**, *128*, 231−240.

(20) Li, J. Z.; Gramatica, P. Classification and virtual screening of androgen receptor antagonists. *J. Chem. Inf. Model.* **2010**, *50*, 861−874.

(21) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*, D1100−D1107.

(22) Gilson, M. K.; Chen, X.; Liu, M. BindingDB: An on-line molecular recognition database. *Biophys. J.* **2001**, *80*, 33a−33a.

(23) Wishart, D. S.; Knox, C.; Guo, A. C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **2008**, *36*, D901−D906.

(24) Williams, A. J.; Tkachenko, V. ChemSpider: How an online resource of chemical compounds, reaction syntheses, and property data can support green chemistry. *Abstr. Pap. Am. Chem. Soc.* **2010**, 239.

(25) Cherkasov, A.; Shi, Z.; Li, Y.; Jones, S. J. M.; Fallahi, M.; Hammond, G. L. 'Inductive' charges on atoms in proteins: Comparative docking with the extended steroid benchmark set and discovery of a novel SHBG ligand. *J. Chem. Inf. Model.* **2005**, *45*, 1842−1853.

(26) Cherkasov, A. Inductive electronegativity scale. Iterative calculation of inductive partial charges. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 2039−2047.

(27) Cherkasov, A. 'Inductive' Descriptors: 10 Successful Years in QSAR. *Curr. Comput.-Aided Drug Des.* **2005**, *1*, 21−42.

(28) Talete, s. DRAGON for Window, Version 5.5; TALETE srl: Milano, Italy, 2007.

(29) Sikonja, M. R.; Kononenko, I. An adaptation of Relief for attribute estimation in regression. In *Fourteenth International Conference on Machine Learning*, Nashville, TN, USA, July 8−12, 1997; Douglas, H. F., Ed.; Morgan Kaufmann: San Fransisco, 1997; pp 296−304.

(30) Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. H. The WEKA data mining software: an update. *SIGKDD Explorations* **2009**, *11*, 10−18.

(31) Kachigan, S. K. In *MultiVariate Statistical Analysis: A Conceptual Introduction*; Radius Press: New York, 1991.

(32) Aha, D. W.; Kibler, D. Instance-based learning algorithms. *Mach. Learn.* **1991**, *6*, 37−66.

(33) Pfahringer, B.; Holmes, G.; Kirkby, R. B. Optimizing the Induction of Alternating Decision Trees. In *Proceedings of the Fifth Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, Hong Kong, China, April 16−18; Cheung, D. W., Williams, G. J., Li, Q., Eds.; Springer: Hong Kong, China, 2001; pp 477−487.

(34) Egmont-Petersen, M.; de Ridder, D.; Handels, H. Image processing with neural networks - a review. *Pattern Recognit.* **2002**, *35*, 2279−2301.
(35) Cleary, J. G.; Trigg, L. E. An Instance-based Learner Using an Entropic Distance Measure. In *12th International Conference on Machine Learning*, Tahoe City, California, USA, July 9−12, 1995; Prieditis, A., Russell, S. J.; Eds.; Morgan Kaufmann: Tahoe City, California, USA, 1995; pp 108−114.
(36) Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123−140.
(37) Friedman, J.; Hastie, T.; Tibshirani, R. Additive logistic regression: a statistical view of boosting. *Ann. Statist.* **2000**, *28*, 337−407.
(38) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5−32.
(39) Melville, P.; Mooney, R. J. Constructing Diverse Classifier Ensembles Using Artificial Training Examples. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, Acapulco, Mexico, Aug 9−15, 2003; Morgan Kaufmann: San Francisco, 2003; pp 505−510.
(40) Roy, K. On some aspects of validation of predictive quantitative structure-activity relationship models. *Expert Opin. Drug Discovery* **2007**, *2*, 1567−1577.
(41) Konovalov, D. A.; Sim, N.; Deconinck, E.; Heyden, Y. V.; Coomans, D. Statistical confidence for variable selection in QSAR models via Monte Carlo cross-validation. *J. Chem. Inf. Model.* **2008**, *48*, 370−383.
(42) Hawkins, D. M.; Basak, S. C.; Mills, D. Assessing model fit by cross-validation. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 579−586.
(43) Invitrogen. www.invitrogen.com.
(44) Tavassoli, P.; Snoek, R.; Ray, M.; Rao, L. G.; Rennie, P. S. Rapid, non-destructive, cell-based screening assays for agents that modulate growth, death, and androgen receptor activation in prostate cancer cells. *Prostate* **2007**, *67*, 416−426.
(45) Song, C. H.; Yang, S. H.; Park, E.; Cho, S. H.; Gong, E. Y.; Khadka, D. B.; Cho, W. J.; Lee, K. Structure-based virtual screening and identification of a novel androgen receptor antagonist. *J. Biol. Chem.* **2012**, *287*, 30769−30780.
(46) Svensson, F.; Karlen, A.; Skold, C. Virtual screening data fusion using both structure- and ligand-based methods. *J. Chem. Inf. Model.* **2012**, *52*, 225−232.
(47) Cantin, L.; Faucher, F.; Couture, J. F.; de Jesus-Tran, K. P.; Legrand, P.; Ciobanu, L. C.; Frechette, Y.; Labrecque, R.; Singh, S. M.; Labrie, F.; Breton, R. Structural characterization of the human androgen receptor ligand-binding domain complexed with EM5744, a rationally designed steroidal ligand bearing a bulky chain directed toward helix 12. *J. Biol. Chem.* **2007**, *282*, 30910−30919.
(48) Zhou, X. E.; Suino-Powell, K. M.; Li, J.; He, Y. Z.; MacKeigan, J. P.; Melcher, K.; Yong, E. L.; Xu, H. E. Identification of SRC3/AIB1 as a preferred coactivator for hormone-activated androgen receptor. *J. Biol. Chem.* **2010**, *285*, 9161−9171.
(49) Feher, M.; Williams, C. I. Numerical errors and chaotic behavior in docking simulations. *J. Chem. Inf. Model.* **2012**, *52*, 724−738.
(50) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789−6801.