

Discovery of novel antagonists targeting the DNA binding domain of androgen receptor by integrated docking-based virtual screening and bioassays

Jinping Pang¹, Chao Shen¹, Wenfang Zhou¹, Yun Xia Wang¹, Luhui Shan², Xin Chai¹, Ying Shao¹, Xueping Hu¹, Feng Zhu¹, Danyan Zhu¹, Li Xiao³, Lei Xu⁴, Xiaohong Xu², Dan Li¹, and Tingjun Hou¹

¹Zhejiang University

²Zhejiang Cancer Hospital

³Huzhou University

⁴Jiangsu University of Technology

June 8, 2020

Abstract

Background and Purpose: Androgen receptor (AR), a ligand-activated transcription factor, is a master regulator in the development and progress of prostate cancer (PCa). A major challenge for the clinically used AR antagonists is the rapid emergence of resistance induced by some point mutations in the AR ligand binding domain (LBD), and therefore discovery of novel anti-AR therapeutics that can combat mutation-induced resistance is quite demanding. Therein, blocking the interaction between AR and DNA represents an innovative strategy to overcome resistance of traditional antagonists towards the AR LBD. **Experimental Approach:** In this study, an integrated docking-based virtual screening (VS) strategy based on the crystal structure of the DNA binding domain (DBD) of AR was conducted to search for novel AR antagonists, and then a series of bioassays including bio-layer interferometry (BLI) and RNA-seq were used to evaluate the biological activities of these compounds. **Key Results:** Among the tested compounds, 2-(2-butyl-1,3-dioxoisindoline-5-carbox-amido)-4,5-dimeth-oxybenzoicacid (Cpd39) was identified as a potential hit, which was competent to block the binding of the AR DBD to androgen receptor response (ARE) and showed decent potency against AR transcriptional activity. Furthermore, Cpd39 was capable of effectively inhibiting the proliferation of PCa cell lines (i.e., LNCaP, PC3, DU145, and 22RV1) and reducing the expression of not only the full-length AR but also the splice variant AR-V7. **Conclusion and Implications:** The novel AR DBD-ARE blocker Cpd39 could serve as a starting point for the development of new therapeutics for castration-resistant PCa.

Introduction

Prostate cancer (PCa) is one of the major fatal cancers for males worldwide. Androgen deprivation therapy (ADT) along with the use of potent anti-androgens has been the cornerstone of treatment for PCa (Huggins, Stevens & Hodges, 1941). However, it is inevitable that PCa would eventually develop into castration resistant PCa (CRPC) due to the restoration of androgen receptor (AR) signaling (Bohl, Gao, Miller, Bell & Dalton, 2005; Li, Chan, Brand, Hwang, Silverstein & Dehm, 2013). Considering that AR signaling is crucial for PCa at all stages, antagonists is recognized as one of the most effective ways to treat PCa (Chen et al., 2004; Tan, Li, Xu, Melcher & Yong, 2015). Like other nuclear receptors, AR consists of four functional domains, including the N-terminal transcriptional activation domain (NTD), the conserved DNA binding

domain (DBD), the hinge region, and the C-terminal ligand binding domain (LBD) (Li et al., 2019; Shafi, Yen & Weigel, 2013; Tan, Li, Xu, Melcher & Yong, 2015). To date, all of the approved AR antagonists target the ligand binding pocket (LBP) of the LBD. Among them, flutamide, bicalutamide and enzalutamide have been widely used for the treatment of androgen-dependent PCa (Chi et al., 2019; Joseph et al., 2013; Ning et al., 2013). Recently, another two AR antagonists apalutamide (ARN-509, an enzalutamide derivative) and darolutamide (ODM-201) were approved for the treatment of non-metastatic CRPC (Fizazi et al., 2019; Schuyler, 2019). However, a broad spectrum of studies demonstrated that the potency of conventional AR antagonists would be suffered from the rapid emergence of drug resistance (Watson, Arora & Sawyers, 2015). A number of mutations at the AR LBP would induce resistance to first-generation antiandrogens, such as bicalutamide and hydroxyflutamide (Bohl, Miller, Chen, Bell & Dalton, 2005; Miller, 2010; Watson, Arora & Sawyers, 2015). A phase III study showed that the F876L mutation in the LBP of AR could induce resistance to enzalutamide and apalutamide (Joseph et al., 2013). In addition, the overexpression of AR splices variants (ARVs) without a part of or even the entire LBD in PCa specimens was observed (Dehm, Schmidt, Heemers, Vessella & Tindall, 2008; Watson, Arora & Sawyers, 2015). The predominant ARVs detected in CRPC patients include AR-V3, AR-V7 and AR-V9, and their expression levels are significantly high in 22RV1 and CWR-R1 CRPC cell lines (Hornberg et al., 2011), suggesting that ARVs would most likely play a critical role in the development of CRPC. Some ARVs are also implicated in the development of drug resistance by engaging the AR chromatin-binding sites and driving the AR transcriptional program in a constitutive and ligand-independent manner (Li, Chan, Brand, Hwang, Silverstein & Dehm, 2013; Watson, Arora & Sawyers, 2015). Therefore, the discovery of novel AR antagonists is still quite urgent.

In recent years, several other targeting sites on AR, such as the action function 2 (AF2) site and the binding function 3 (BF3) site on the LBD, and the DNA binding site on the DBD, have attracted increasing attentions (Ban et al., 2014; Caboni et al., 2012; Munuganti et al., 2014). Targeting these sites may help overcome mutation-induced resistance to traditional AR antagonists targeting the LBP. The AR DBD contains two zinc finger regions, which contribute to DNA binding and homodimerization (**Fig. 1A**). The P-box at the N terminus of the α -helix in the first zinc finger can insert directly into the DNA major groove of androgen response element (ARE). The second zinc finger with the D-box interacts with another DBD monomer with a head-to-head arrangement (Khorasanizadeh & Rastinejad, 2001; Shaffer, Jivan, Dollins, Claessens & Gewirth, 2004). It is deduced that targeting the DBD can disrupt either the formation of the AR dimer or the interaction between the DBD and ARE, or both, and then suppress AR transcriptional activity. Recently, Jones *et al.* found that VPC-14337 (pyrvinium) could strongly inhibit the full-length AR and splice variants by binding to the DBD or hinge region (Dalal et al., 2014; Lim et al., 2014). However, pyrvinium illustrated cross reactivity with other nuclear receptors and could disrupt the Wnt/ β -cat signaling pathway at nanomolar-level concentrations (Thorne et al., 2010). Li *et al.* conducted structure-based virtual screening (SBVS) and discovered a series of AR antagonists in micromole range targeting a potential binding site on the DBD (Ban, Dalal, Li, LeBlanc, Rennie & Cherkasov, 2017; Dalal et al., 2014). Then, through the structural optimization of the most active hit, they identified a better candidate, 4-(4-phenylthiazol-2-yl)morpholine, with eGFP IC_{50} and PSA IC_{50} of 0.33 μ M and 0.28 μ M, respectively (**Fig. 1B**). Further structure-activity relationship (SAR) analysis led to the discovery of VPC-14449, which showed not only comparable potency to enzalutamide but also considerable activity towards enzalutamide-resistant cells and ARVs. The project resulted in the largest academic licensing deal in Canadian history, totaling \$142M. Recently, they discovered a series of AR DBD dimer blockers that possibly target the D-box of the DBD, and the best candidate (VPC-17005) showed submicromolar AR antagonistic activity (eGFP IC_{50} = 0.734 μ M and PSA IC_{50} = 0.691 μ M) (Dalal et al., 2018). Although much efforts have been dedicated to the discovery of AR antagonists towards the AR DBD, only several potential hits have been confirmed and no candidate has been pushed into the clinical trials, therefore the discovery of AR DBD antagonist with novel scaffolds is required.

In this study, to discover novel and potent antagonists targeting the AR DBD, an integrated docking-based virtual screening (VS) towards the DBD was conducted, and 61 compounds were selected for experimental bioactivity testing. Among the identified hits, 2-(2-butyl-1,3-dioxoisindoline-5-carbox-amido)-4,5-

dimethoxybenzoic acid (Cpd39) exhibited the best potency, and it could inhibit AR transcriptional activity, suppress the growth of AR-positive cell lines, and reduce the expression of AR target genes in LNCaP cells. In particular, Cpd39 could effectively downregulate the level of an ARVs specifically regulated gene AKT1 in 22RV1 cell lines. Consistent with the predicted binding mode of Cpd39, the bio-layer interferometry (BLI) analysis proved that Cpd39 could indeed block the interaction between the AR DBD and ARE. Cell viability assays on A549 and 3T3 cell lines demonstrated that Cpd39 was inherently non-toxic. In addition, it was found that Cpd39 could impede the growth of PC3 and DU145 cell lines. Further RNA-seq study indicated that it might be related with the effect on several important biological processes involved in the mTOR and PI3K-AKT signaling pathways. Taken together, the identified hit Cpd39 provides a promising new scaffold for the development of new therapeutics against CRPC.

Methods

Protein preparation and binding site detection

The crystal structure of the AR DBD dimer bound to two hexameric half-site responds elements (PDB entry: 1R4I) (Shaffer, Jivan, Dollins, Claessens & Gewirth, 2004) was used the template structure for VS. The *Protein Preparation Wizard* in Schrödinger 2017 was utilized to prepare the protein (Sastry, Adzhigirey, Day, Annabhimoju & Sherman, 2013), including assigning bond orders, adding hydrogen atoms, filling in missing side chains, and minimizing the system with the OPLS2005 force field until the root-mean-square deviation (RMSD) of heavy atoms converged to 0.30 Å. The protonation states of residues at pH = 7.0 were determined by PROPKA (Olsson, Sondergaard, Rostkowski & Jensen, 2011), and only the single protomer was remained for the final VS. If the used docking program had its own protein preparation function, it would be employed to handle the protein further.

The potential binding site in the DBD was determined by the *Site Finder* module in Molecular Operating Environment (MOE) just as Li *et al.* did in their study (Li et al., 2014). The identified binding site located at the AR DBD-ARE binding interface was embraced by the residues Ser579, Val582, Phe583, Arg586, Arg609, Lys610, Pro613 and Arg616.

Docking-based virtual screening

Three molecular docking programs, namely Glide (Friesner et al., 2004), GOLD (Jones, Willett, Glen, Leach & Taylor, 1997) and AutoDock Vina (Trott & Olson, 2010), were used in docking-based VS. All the parameters were set to the default values, unless otherwise noted as followed.

Glide . The *Receptor Grid Generation* utility of Glide was used to generate the receptor grid, which is located in a binding box with the size of 10 Å × 10 Å × 10 Å centered on the centroid of the residues that form the pocket. The standard precision (SP) scoring mode was used in molecular docking.

GOLD. The protein was prepared by the built-in protein preparation function, and the binding site was defined by all the atoms within 10 Å of the pocket. The genetic algorithm (GA) method with the “automatic” settings and the default Piecewise Linear Potential (CHEMPLP) scoring function were utilized for sampling and scoring, respectively.

Autodock Vina . The structures of the protein and ligands were converted into the *pdbqt* formats by AutoDockTools, along with the addition of hydrogen atoms, assignment of Gasteiger charges and cleanup of unwanted elements. The binding site was determined by the center of the specified residues in the pocket, and the searching space was set to 18.75 Å × 18.75 Å × 18.75 Å.

The overall VS workflow is depicted in **Fig. 1C** . The Specs library with around 210,000 small molecules was firstly pretreated by the Pan-Assay Interference Compounds (PAINS) rule (Baell & Holloway, 2010) and Rapid Elimination Of Swill (REOS) rule (Walters & Namchuk, 2003) in Canvas to remove the noisy molecules with undesirable functional groups. Then, the Lipinski’s rule-of-five (Ro5) (Lipinski, Lombardo,

Dominy & Feeney, 1997) and Opera rules (Oprea, 2000) were used to eliminate non-drug-like molecules with the number of chiral centers ≥ 3 , the violation count of the Oprea rule ≥ 3 and the violation count of Ro5 ≥ 2 . Next, each molecule was prepared by the *LigPrep* module in Schrodinger to assign its ionized states, tautomers, stereoisomers and low-energy conformation. The ionized states and tautomers at pH = 7.0 \pm 2.0 were generated by using the *Epik* method (Shelley, Cholleti, Frye, Greenwood, Timlin & Uchimaya, 2007), and the maximum number of the stereoisomers for each molecule was set to 4. The other parameters for *Ligprep* were set to the default settings.

Because the binding site on the DBD is relatively flat, it is quite possible that the binding poses of molecules cannot be reliably predicted by a single docking program. In order to improve the prediction accuracy of binding poses, three docking programs (i.e., Glide SP, GOLD and AutoDock Vina) were used to conduct three individual docking calculations. Then, the pairwise root-mean-square-deviations (RMSDs) of the binding poses predicted by the three different programs for each molecule were calculated by the *obrms* utility in OpenBabel (O’Boyle, Banck, James, Morley, Vandermeersch & Hutchison, 2011). If two of the three RMSD values for a molecule were less than 2.0 Å, this molecule would be considered in further analysis, and otherwise abandoned. The above operations might lead to the loss of some potential binders, but to some extent they could enhance the reliability of VS. Next, the remaining molecules were ranked by integrating the three docking scores, and the 171 compounds with the docking scores higher than the predefined cutoffs (-3.5, 55 and -6 for Glide SP, Gold CHEMPLP and Autodock Vina, respectively) were clustered based on the Tanimoto coefficients using the MACCS fingerprints. Finally, 61 molecules were selected through visual inspection for bioassays.

Protein expression and purification

The AR DBD (residues: 440-561) was cloned into the NdeI and XhoI sites of pET28a vector. The sequence of the AR DBD is HMCLICGDEASGCHYGALTCGSKVFFKRAAEGKQKY

LCASRNDCTIDKFRRKNCPSCRLRKCYEAGMTLGARKLKKLGNLKLQEEGEASSTTSPTEETTQKLTVSHIEGYEQ

The plasmid with the mutated residue (Tyr594Asp) in the predicted binding site was constructed with the PCR technology. The forward cloning primer was 5'-CAGAAGGAC CTGTGCGCCAGCAGAAATGATTGC, and the reverse cloning primer was 5'-GCACAGGTC CTTCTGTTTCCCTTCAGCGGCTCT (the mutated residues are underlined). The plasmids were transformed into *E. coli* BL21 (DE3).

Cultures were grown to an OD₆₀₀ of 0.8 in two liters of LB media supplemented with 50 µg/ml Kanamycin before induced with 0.1 mM IPTG for 3 h at 37 °C. The cells were collected by centrifugation at 4000 rpm for 15 min and then resuspended in 20 ml buffer A (50 mM Tris-HCl, pH=8.0, 300 mM NaCl, 5% glycerol) supplemented with 10 mM imidazole and 0.1% PMSF. Cell lysis was achieved by sonication, followed by centrifugation at 14500 × g for 30 min at 4 °C. The protein was purified with HisPur Ni-NTA Resin (Thermo Scientific; USA) and dialyzed with buffer A.

AR transcriptional activity assay

A cell line of LNCaP that stably expresses eGFP under the regulation of an androgen response element was generated to investigate the agonist/antagonist activity of the selected compounds as previously described (Tavassoli, Snoek, Ray, Rao & Rennie, 2007; Zhou et al., 2018). The LNCaP-ARR2PB-eGFP cells were grown in phenol-red-free RPMI 1640 supplemented with 5% charcoal stripped serum (CSS) for 5 days. Then cells were seeded into a 96-well plate (3.5 × 10⁴ cells per well) for 24 h, and then treated by a single concentration or increasing concentrations (0-50 µM) of compounds with 10 nM DHT for screening. The fluorescence was measured after incubating for 72 h (Synergy H1, BioTek. Excitation, 485 nm; Emission, 535 nm).

Prostate specific antigen (PSA) Assay.

The PSA secreted into the media was evaluated in parallel with the AR transcriptional activity assay using the same samples. After incubation of 3 days, 400 μ L of the media was collected and sent to Cancer Hospital of University of Chinese Academy of Sciences, Zhejiang Cancer Hospital (Hangzhou, Zhejiang). The PSA level was evaluated with the IMMULITE®2000 XPi immunoassay system (Siemens Ltd, Erlangen, Germany).

Cell proliferation assay

Cell proliferation was measured using the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2-H-tetrazolium bromide (MTT) colorimetric assay. A total of four types of PCa cell lines were involved, including LNCaP (CLS Cat# 300265/p761_LNCaP, RRID:CVCL_0395), 22RV1 (DSMZ Cat# ACC-438, RRID:CVCL_1045), PC3 (NCBI Iran Cat# C427, RRID:CVCL_0035) and DU145 (ATCC Cat# HTB-81, RRID:CVCL_0105). These cells were cultured in RPMI-1640 media (5% CSS) with 2.5×10^3 cells/well in 96-well plates (5×10^3 for LNCaP, 3×10^3 for 22RV1, and 2×10^3 for PC3 and DU145). After incubation at 37 °C for 24 h, cells were treated with serial dilutions of the tested compounds (10 nM DHT in extra for LNCaP culture) and incubated for another 3 days. Afterward, 10 μ L of 5 mg[?] mL^{-1} MTT solution was added into each well and incubated for another 4 h. Then, 100 μ L of triplex 10% SDS-5% isobutyl alcohol-0.012 mol[?] L^{-1} HCl (w/v/v) solution was added to dissolve the formazan crystals. The absorbance at 570 nm was measured with the reference wavelength at 650 nm using a spectrophotometer (Bioteck Eon, Winooski, VT).

Cell cytotoxicity assay

To rule out the inherent toxicity of the tested compounds, A549 (NCI-DTP Cat# A549, RRID:CVCL_0023) and 3T3 (CLS Cat# 400101/p677_NIH-3T3, RRID:CVCL_0594) cell lines were seeded in DMEM media at a density of 2000 cells per well and treated with the concentration gradient of the tested compounds for 24 h. Cell viability was then measured as described above.

Androgen displacement Assay

Androgen displacement was assessed with the PolarScreen™ Androgen Receptor Competitor Green Assay Kit following the instructions of the manufacturer (Invitrogen Life Technologies, Inc.).

BLI assay

BLI experiments were conducted on a FortéBio Octet Red system. The biotinylated ARE was formed by annealing the following complementary oligonucleotides in H_2O : upper strand, 5'-Biotin-TACAAATAGGTTCTTGGAGTACT TTACTAGGC ATGGACAATG, and lower strand, 5'-CATGTCCATGCC TAG TAAAGTACT CCAAGAACCTATTTGTA (AREs are underlined). All experiments were conducted with the biotinylated ARE and purified AR DBD protein in PBST with 2.5% DMSO. The ARE (50 nM) was loaded onto streptavidin sensors in 200 μ L PBS for 300 s, and then the ARE-loaded sensors were pre-equilibrated in PBS. The kinetics of ARE-protein association was monitored by soaking sensors into wells containing 1 μ M AR DBD and 50 μ M compound for 300 s, followed by dissociation in the same buffer deprived of protein for additional 180 s. All the experiments were carried out independently at least two times.

Confocal Microscopy

LNCaP cells were seeded at a concentration of 5×10^4 per well on sterile coverslips placed within 12-well plates. After incubation at 37 °C for 24 h, cells were treated with 10 μ M compounds for 12 h and then treated with 5 nM DHT for 1.5 h. After aspiration of the media, cells were fixed in 4% paraformaldehyde at 25 °C for 20 min and incubated with AR antibodies (Cell Signaling Technology Cat# 5153, RRID:AB_10691711) overnight after washed with PBS three times. An Alexa-488 conjugated goat-anti rabbit IgG (Cell Signaling Technology Cat# 4412, RRID:AB_1904025) diluted at 1:1000 was used as the secondary antibody. The

counterstain DAPI was employed to visualize cell nucleus. Images were taken at 60 magnification using the Nikon A1R confocal spinning disk microscope and followed by the analysis with the NIS-Elements Viewer (Northern Eclipse, Empix Imaging, Inc.).

Western blotting

LNCaP cells were cultured in phenol-red-free RPMI 1640 (5% CSS) for 2 days and then seeded in 6-well plates (3×10^5 cells per well). After overnight incubation, the cells were treated with 10 μ M of compounds under the presence of 10 nM DHT. After 48 h incubation, the cells were lysed with RIPA and subjected to 10% SDS-polyacrylamide for western blotting. Protein was transferred to the methanol charged PVDF membrane and probed with rabbit anti-AR (Cell Signaling Technology Cat# 5153, RRID:AB_10691711) and mouse anti-TMPRSS2 (Santa Cruz Biotechnology Cat# sc-101847, RRID:AB_2205599), along with monoclonal anti-GAPDH (Cell Signaling Technology Cat# 5174, RRID:AB_10622025) utilized to show equal loading. Proteins were visualized using anti-mouse or anti-rabbit HRP-conjugated secondary antibodies (Sangon Biotech, Shanghai, China). The protein bands were detected using the Automatic Gel Imaging Analysis System developer (Peiqing Science and Technology, Guangzhou, China).

Quantitative real-time reverse transcription polymerase chain reaction (Q-PCR)

Total RNA was isolated from cells using the EZ-10 DNAaway RNA Mini-Preps Kit (Sangon Biotech, Shanghai, China) according to the instructions of manufacturer. cDNA was generated using the Hifair® III 1st Strand cDNA Synthesis SuperMix (YEASEN, Shanghai, China). Diluted cDNA was mixed with the forward primer, reverse primer, SYBR Green PCR Master Mix (YEASEN, Shanghai, China), and RNase-free water in a 96-well plate. Analysis of mRNA expression was carried out using the Applied Biosystems QuantStudio 3. All samples were normalized to the level of GAPDH. The threshold cycles (Ct) for the control (GAPDH) and gene of interest were determined, and the relative mRNA levels were calculated by the $2^{-\Delta C_t}$ method. The details of the primer sequences used in the study are shown in the supporting information.

RNA-seq

DU145 cells were cultured in RPMI-1640 media and then exposed to DMSO and 10 μ M Cpd39, respectively, for 48 h. Total RNA was extracted with Trizol Reagent and sequenced using the Hiseq-PE150 high-throughput sequencing platform (Novogene, Beijing, China). The raw expression data were processed and normalized as followed.

First, the human reference genome GRCh37 (version 19) was downloaded from the Ensembl website (<https://www.ensembl.org>), a database project providing access to human genome annotation (Hubbard et al., 2002). Four FASTQ files containing the paired-end sequence reads were aligned to the human reference genome using HISAT2 (HISAT2, RRID:SCR_015530) (Hierarchical Indexing for Spliced Alignment of Transcripts) (version 2.1.0) (Kim, Langmead & Salzberg, 2015). The output SAM (sequencing alignment/map) files were converted to the BAM (binary alignment/map) files and sorted using SAMtools (SAMTOOLS, RRID:SCR_002105) (version 1.9) (Khalil et al., 2009). Next, StringTie (StringTie, RRID:SCR_016323) (Pertea, Pertea, Antonescu, Chang, Mendell & Salzberg, 2015) (version 2.0) was used to assemble the genes for the data set in each BAM file separately, calculate the expression levels of each gene and each isoform, and then merge all the gene structures found in any of the samples together using the full set of assemblies. The merged transcripts were passed to StringTie again so that it can re-calculate the transcript abundances. Finally, the Ballgown package (Frazee, Pertea, Jaffe, Langmead, Salzberg & Leek, 2015) grouped all the transcripts and abundances from StringTie by experimental condition, and determined the expressed genes (DEGs) between conditions using a linear model for a log transformation of the FPKM values attached to transcripts. DEGs were defined as those with the false discovery rate (FDR) adjusted p-value less than 0.05 and the fold change between conditions more than 2. To identify the biological processes and pathways that were significantly enriched by DEGs, the gene list was analyzed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (DAVID, RRID:SCR_001881) (version 6.8) (Huang da, Sherman

& Lempicki, 2009) and KOBAS 2.0 (KOBAS, RRID:SCR_006350) (Xie et al., 2011).

Data and statistical analysis

Data were analyzed in compliance with the *British Journal of Pharmacology* guidance for experimental design and analysis in preclinical pharmacological research (Curtis et al., 2018). The number of replicates was 5 or 6 per group for each data set, and the results were presented as the mean \pm SEM unless stated otherwise. Statistical analyses were performed with the GraphPad Prism software v 6.0 (San Diego, CA, USA, RRID:SCR_002798). The differences of two groups were analyzed using a two-tailed Student's t-test, and the differences between more than two groups were analyzed using one-way analysis of variance (ANOVA), followed by Dunnett's post hoc test. Some results were normalized to the control to avoid unwanted sources of variation. $P < 0.05$ was considered to be significant.

Materials

VPC-14228 (4-(4-phenylthiazol-2-yl) morpholine) and the tested compounds were purchased from Specs (Maryland, USA), and enzalutamide was purchased from MedChemExpress (New Jersey, USA).

Results and discussion

Evaluation of candidate compounds

The binding pocket formed by Ser579, Val582, Phe583, Arg586, Arg609, Lys610, Pro613 and Arg616 on the DBD-ARE binding interface was detected as the potential binding site, and then the Specs chemical library was virtually screened (**Fig. 1C**). The 61 compounds identified from the VS were tested for their antagonistic activities using a nondestructive enhanced green fluorescent protein (eGFP) assay (Trott & Olson, 2010). Vpc-14228, a reported compound targeting the AR DBD, and enzalutamide were used as positive controls (Dalal et al., 2014; Lack et al., 2011). At first, the 61 compounds were tested at the concentration of 5 $\mu\text{g}[\text{?}]\text{mL}^{-1}$, and compound Cpd31, Cpd34, Cpd55 and Cpd39, showed over 50% antagonistic activity of enzalutamide. Further testing under a concentration gradient (0.1-100 μM) illustrated that Cpd31 ($\text{IC}_{50} = 39.46 \mu\text{M}$) and Cpd39 ($\text{IC}_{50} = 10.94 \mu\text{M}$) exhibited a dose-dependent inhibition (**Fig. 1D**). Since Cpd39 showed better activity than Cpd31 (**Fig. S1**), it was selected as a potential antagonist hit for further studies. To avoid false-positive detection by the eGFP assay, a complementary PSA assay was employed (Li et al., 2014). PSA, an AR downstream target gene of AR, is a clinically important serum biomarker for PCa and has been widely used as an indicator to evaluate AR transcriptional activity. As shown in **Fig. 1D**, Cpd39 inhibited the expression of PSA in a dose-dependent manner with $\text{IC}_{50} = 20.16 \mu\text{M}$ and those for Vpc-14228 and enzalutamide were 1.26 μM and 0.329 μM , respectively.

Structure activity relationship (SAR) analysis for the analogues of Cpd39

To search for more potent analogues and explore their initial SAR of the new scaffold of Cpd39, similarity- and substructure-based analogue searching was performed using Cpd39 as the query. A total 27 analogues of Cpd39 were then identified in the ChemDiv library for activity assessment. The chemical structures of these molecules and their bioactivities are summarized in **Table S1**. In general, no significant activity improvement was observed, and only Cpd39-12 showed comparable bioactivity to Cpd39. However, valuable information can be captured from the preliminary SAR analysis, which would provide helpful guidance for further structural optimization of Cpd39. According to the predicted binding mode of Cpd39 within the binding pocket on the DBD (**Fig. 1E**), the carboxyl group on the phenyl ring could form two H-bonds with Arg586 and a H-bond with Ser579, and the amide group on the linker could form a H-bond with Arg616. These H-bond interactions are most likely the primary driving force for the binding of Cpd39. Apart from the H-bond interactions, the hydrophobic interactions between Cpd39 and some surrounding residues should

also contribute a lot. For example, the phenyl ring linked to the carboxyl group is located exactly in the center of Phe583, Tyr594 and Pro613, and thereby can form favorable van der Waals interactions with the hydrophobic side chains of these residues. In addition, the alkyl group in another phenyl ring may also play an important role in the binding by forming favorable hydrophobic interactions with Lys606 and Phe607. Another important factor may lie on the existing of Arg586 and Lys610, which can construct a natural groove on the surface to fit the binding of Cpd39. The bioactivities of the analogues are well consistent with the binding mode analysis. The ortho substituents of H-bond acceptors on Ring A, such as the carboxylic group, ester group and amide group, were favorable for ligand binding as the validation of the binding mode. For the analogues without any ortho substituent on Ring A, including Cpd39-1, Cpd39-2, Cpd39-3, Cpd39-10, Cpd39-11, Cpd39-20 and Cpd39-23, were almost inactive due to the absence of the crucial H-bond interactions with Arg586 or Ser579. The size and flexibility of the substituents on Ring C are also critical to ligand binding. Larger substituents, such as tert-pentyl (Cpd39) and tert-butyl (Cpd39-9 and Cpd39-12), led to enhanced binding affinity, while smaller substituents, such as isopropyl (Cpd39-4 and Cpd39-5), methyl (Cpd39-6) and halogen atoms (Cpd39-17 and Cpd39-25), did not show favorable effect. As for the compounds Cpd39-8, Cpd39-16 and Cpd39-24, their poor activities most likely stemmed from the rigidity of their Ring C. As shown in **Fig. 1G**, only the substituent groups with appropriate size and flexibility can stretch into the groove formed by Arg586 and Lys610. Regarding the other analogues, the reason of their low activities need to be further explored.

Cpd39 effectively inhibits AR-mediated PCa cell proliferation

To figure out the antitumor effect of Cpd39 on AR-expressing PCa cells, LNCaP, an androgen-dependent cell line, and 22RV1, an androgen-insensitive cell line, were utilized for testing (Walters & Namchuk, 2003). 22RV1 cell line is derived from the CWR22 xenograft, which expresses high level of AR-V7 and is resistant to enzalutamide. As a result, Cpd39 exerted a dose-dependent manner to inhibit the growth of LNCaP and 22RV1 cells (**Fig. 2A**). At a lower concentration of 1 μM , the potency of Cpd39 was worse than those of the controls, but at higher concentrations, Cpd39 outperformed both Vpc-14228 and enzalutamide. Cpd39 could suppress the cell viability of LNCaP to 75% at 10 μM and even less than 20% at 50 μM . In addition, Cpd39 was more effective than Vpc-14228 and enzalutamide in suppressing 22RV1 growth. Upon the treatment of 20 μM and 50 μM Cpd39 for 72 h, the growth of 22RV1 cells was significantly suppressed to 85% and 50%, respectively. In contrast, Vpc-14228 exhibited a minor effect on the growth of 22RV1.

Cpd39 inhibits the proliferation of AR-independent PCa cell with no toxicity.

To determine whether the anti-proliferative activity of Cpd39 is caused by AR down-regulated, the compound was further evaluated on two androgen-independent cell lines, PC3 and DU145. As shown in **Fig. 2D**, the cell viability curves of PC3 for Cpd39, Vpc-14228 and enzalutamide at different concentrations ranging from 0 to 50 μM have a similar shape. The IC_{50} values of enzalutamide and Vpc-14228 were higher than 100 μM for both PC3 and DU145, suggesting that the cytotoxicity of the compounds was specifically mediated through AR. However, Cpd39 could reduce cell viability to 50-60% for PC3 cell line at 50 μM and the corresponding IC_{50} value was 47.98 μM . Cpd39 exhibited an even stronger inhibitory effect on the proliferation of DU145 ($\text{IC}_{50} = 13.28 \mu\text{M}$) than on PC3 cells. To assess whether the compound is inherently toxic, 3T3 mouse fibroblasts and A549 human lung adenocarcinoma cell lines were selected for in vitro cytotoxicity assessment. To our surprise, 50 μM of Cpd39 showed a negligible effect on both 3T3 and A549 cell lines. The IC_{50} values of Cpd39 for 3T3 and A549 cell lines were larger than 80 μM and 100 μM , respectively, suggesting that the toxicity of Cpd39 to PCa cells is specific. In addition, it is obvious that the antiproliferative effect of Cpd39 on AR negative cells PC3 and DU145 was not mediated through AR, which might be linked with some unknown mechanisms (Watson, Arora & Sawyers, 2015).

To explore the inhibition mechanism of Cpd39 towards DU145 cell line, HiSeq-PE150 was used to conduct sequencing analysis. The genes expression differences were analyzed by comparing the transcriptomes before and after the treatment of 10 μM Cpd39. At first, those up-regulated or down-regulated genes were

examined to confirm whether Cpd39 could yield any potential genotoxic effect. Generally, the genes with genotoxicity including a number of p53 target genes and those involved in apoptosis, DNA repair, DNA damage response or stress response were checked. None of them showed significant response, indicating that Cpd39 is inherently nontoxic (Ellinger-Ziegelbauer et al., 2009).

Enrichment analysis of Kyoto encyclopedia of genes and genomes (KEGG) pathways was also conducted to identify the essential pathways potentially involved. The results revealed that the “metabolic pathways”, “pathways in cancer”, “proteoglycans in cancer” and “ribosome” (**Fig.2E**) were mainly affected by the treatment of Cpd39, suggesting that the changes in these pathways/processes might account for its anticancer efficacy. After an in-depth analysis, it was noticed that the mTOR and PI3K-AKT signaling pathways showed the most prominent differences, which might be the main reason why Cpd39 could induce the apoptosis of DU145 (**Fig. S2 and Fig. S3**). The genes encoding RHOA (Ras homolog family member A, fold change = 27.4, $p = 0.013$), PIK3CB (phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit beta, fold change = 470.1, $p = 0.044$) and ERBB2 (erb-b2 receptor tyrosine kinase 2, fold change = 44.8, $p = 0.012$) were synergistically trans-repressed. All of them are involved in the PI3K and AKT signaling networks, which were known to contribute significantly to the CRPC development (Li et al., 2018). RHOA, a member of the Rho family of GTP-binding proteins, is over-expressed in many forms of malignant cancers and plays an important role in focal adhesion regulation, actin stress fiber formation, and cell migration of castration-resistance transformation of PCa (Xie et al., 2011). PIK3CB is the gene that encodes the p110 β lipid kinase catalytic subunit of PI3K β . PI3K β is a critically important factor in the PI3K signaling pathway and can drive tumorigenic cell growth and migration in several tumor types. To demonstrate the efficacy of Cpd39, Q-PCR was performed to validate whether these three key genes were truly trans-repressed. The results revealed Cpd39 could significantly down-regulate the expression of these three genes and could exert a deep effect on the growth and metastasis of malignant cells, with no cytotoxicity induced (**Fig. 2F**). Taken together, we deduced that the antitumor effect of Cpd39 on DU145 probably owes to its influence on the PI3K and AKT signaling pathways.

Cpd39 is a novel AR antagonist targeting AR DBD.

To validate whether the antitumor potency of Cpd39 is attributed to the disruption of the binding of the AR DBD to ARE, BLI, a label-free technology was utilized. BLI has been widely employed in drug discovery and life science studies *in vitro* to measure biomolecular interactions. In this study, we initially attempted to confirm whether the compound could affect the binding of the AR DBD to dsDNA bearing the ARE2 sequence at a concentration of 50 μ M, with enzalutamide and Vpc-14228 as the references. In this task, the biotinylated dsDNA (ARE) was attached to the streptavidin BLI sensors, and then exposed to the solution containing both of the AR DBD and tested compound, which is capable of resulting in a large shift for the signal of wavelength compared to the solution containing only the AR DBD if the compound does interrupt the interaction between the AR DBD and ARE. As illustrated in **Fig. 3A**, the behaviors of Cpd39 and Vpc-14228 were similar and the signal decreased ~12-15% in comparison with the AR DBD alone. In contrast, the wavelength change generated by enzalutamide was little, which supported the fact that enzalutamide had no ability to impede the binding between the AR DBD and ARE. To further understand the inhibition behavior of Cpd39, this experiment was carried out by testing a set of concentrations of the compound. As shown in **Fig. 3B**, the inhibitory activity of Cpd39 steadily increased with the increase of concentrations, and Cpd39 could even impede the interaction between the AR DBD and ARE at a concentration of 500 μ M. The behavior of Cpd39 was dissimilar to that of Vpc-14228, as Vpc-14228 functioned back at 500 μ M. Furthermore, to confirm the location of the binding site, we performed the same BLI experiments with the mutated AR DBD protein. Previous studies showed that the Tyr594Asp mutation at the DBD-ARE binding interface could maintain the AR transcriptional activity, so this Tyr594Asp muted AR DBD was used in this study (Dalal et al., 2014). In our experiments, both of Cpd39 and Vpc-14228 lost their ability to block the binding of the mutated AR DBD to ARE. The association kinetics of dsDNA with the mutated AR DBD was significantly different from that of dsDNA with the wild type AR DBD (**Fig. 3C**). That is to say, similar to Vpc-14228, Cpd39 was unable to affect either the association or dissociation of dsDNA with the

mutated protein, indicating that Cpd39 did target on the site of Tyr594 at the DBD-ARE binding interface. Together, Cpd39 could inhibit the protein-DNA interactions in vitro and bind to the proposed site at the AR DBD.

In addition, we also ruled out the possibility of the binding of these compounds to the AR DBD or ARE individually (**Fig. S4**). The K_d value between the AR DBD and Cpd39 was 6.3×10^{-3} M, which was usually regarded as a weak interaction (**Fig. S4A and B**). And Cpd39 did not show a direct interaction with ARE (**Fig. S4C**).

Cpd39 does not target LBD or block nuclear localization

We then attempted to exclude the possibility that Cpd39 binds to the traditional LBP site. Accordingly, a PolarScreen™ AR competitor assay was carried out. This assay was conducted for Cpd31 and Cpd39 at a concentration of 10 μ M with DMSO and DHT as the controls. As shown in **Fig. S4D**, the behavior of Cpd39 was similar to that of the vehicle control (DMSO) but significantly different from that of the positive control (DHT), supporting that Cpd39 does not interact with the LBP. In addition, the translocation of AR into nucleus is a key step in AR signaling. It has been proven that some of the drugs targeting the AR LBD are able to exert their functions by blocking the nuclear translocation of AR, thus preventing the acceptor from the initiation of transcription (2017; Jones, Willett, Glen, Leach & Taylor, 1997). However, the compounds targeting the DBD should theoretically function in the nucleus to block the recognition of AR to ARE. Therefore, we utilized the confocal assay to evaluate whether Cpd39 could affect the DHT-induced translocation of AR from the cytoplasm to the nucleus. As shown in **Fig. 4**, the group without the treatment of DHT or enzalutamide showed low levels of AR in the nucleus compared with the group only treated with DMSO. Conversely, VPC-14228 and Cpd39 could not prevent the DHT-stimulated nuclear localization of AR with considerable fluorescence signal observed in the nucleus. The results indicated that Cpd39 could not impede the DHT-induced AR nuclear translocation, confirming the direct action of Cpd39 on the DBD.

Cpd39 down regulated not only the AR but also ARVs responsive genes in PCa Cells

To assess the ability of blocking the transcription of AR regulated genes, Cpd39 was tested on LNCaP cells. The mRNA levels of three AR target-genes (Friesner et al., 2004) KLK3 (PSA), TMPRSS2 and FKBP5 were evaluated by Q-PCR. As shown in **Fig. 2B**, KLK3, TMPRSS2 and FKBP5 are androgen-upregulated genes and their mRNA levels substantially decreased compared to that of the DHT control. Somehow, the gene of ACPG previously reported as an androgen-downregulated gene did not show expected higher transcriptional level when treated with Cpd39. To further confirm the antiandrogenic activity of Cpd39, the western blot analysis was carried out to check its effect on the expression of TMPRSS2 protein and AR. Consistent with the result of Q-PCR, the expression of TMPRSS2 protein was down-regulated at the presence of Cpd39, while AR seems unaffected.

Furthermore, to investigate whether Cpd39 has an impact on AR-V7 dependent genes, ARVs over-expressed cell line of 22RV1 was selected for testing. The cells were treated with 10 μ M and 50 μ M Cpd39 in charcoal-stripped serum (5% CSS)-containing medium which represented an androgen-independent state. AKT1, reported preferentially as a subset of genes uniquely upregulated by ARVs rather than the prototype AR (Guo et al., 2009), was analyzed by Q-PCR. Remarkably, in our test AKT1 was downregulated by Cpd39 at both the concentrations of 10 μ M and 50 μ M compared to the control group (**Fig. 2G**), though the effect is significant only under 10 μ M. And DHT has no effect. Based on the results above, we infer that Cpd39 is a novel AR antagonist capable of fighting ARVs.

Conclusion

In this study, we have conducted an integrated structure-based VS to search for ligands targeting the AR DBD, and a novel bioactive compound Cpd39 was discovered as an AR antagonist, which showed decent potency against AR transcriptional activity and the expression of PSA. The identified Cpd39 could suppress the growth of AR-positive LNCaP cell line and downregulate the expression of AR target genes PSA, TMPRSS2 and FKBP5. Molecular docking studies suggested that Cpd39 could bind to the P-box of the AR DBD. Consistent with the predictions, a series of BLI experiments showed that Cpd39 could effectively interfere the interaction between the AR DBD and ARE. Meanwhile, Cpd39 could efficiently downregulate the level of an ARVs specifically regulated gene AKT1 when compared with the DMSO treated group in 22RV1 cell line, supporting that Cpd39 may have impact on androgen-responsive genes by blocking the binding of the AR DBD to ARE. In addition, Cpd39 showed impressive growth inhibition on AR independent cell lines PC3 and DU145 with no toxicity. The mechanism was investigated by RNA-seq on DU145. The data presented in this study showed that genotoxin-responsive genes are unaffected by Cpd39, while RHOA, PIK3CB and ERBB2 involved in the PI3K and AKT pathways known to contribute to CRPC development are prominently down regulated by Cpd39. These findings implicated that Cpd39 is a promising hit for the treatment of PCa. However, Cpd39 still needs extensive structural optimization to improve activity, and our studies on Cpd39 analogues have provided some referential information.

To combat acquired resistance to available antiandrogens, CRPC patients require new treatments with non-conventional mechanisms. DNA binding and dimerization are essential for the activation of both AR and ARVs, and it is a promising strategy to target the AR DBD to overcome drug resistance with emergence of gained mutations and ARVs. To develop such AR antagonists, sustained collaborative drug discovery efforts are still needed.

Acknowledgements

This study was supported by the Key R&D Program of Zhejiang Province (2020C03010), National Natural Science Foundation of China (81773632), Zhejiang Provincial Natural Science Foundation of China (LZ19H300001), and the Key R&D Program of Zhejiang Province (2019C02024).

Author's contribution

T.H. and D.L. initiated and supervised the research. C.S. and J.P. designed the experiments. C.S., J.P., conducted virtual screening, compound validations and biological assays. W.Z., Y.W., L.S., X.C., Y.S., X.H., F.Z., L.X., D.Z., and L.X. helped performing part of biological assays. J.P., C.S., D.L. and T.H. wrote the manuscript, and the other authors contributed specific parts of the manuscript. T.H. and D.L. assume responsibility for the manuscript in its entirety. All authors have critically revised the manuscript and approved its final version.

Conflict of interest

The authors declare no conflict of interest.

Declaration of transparency and scientific rigour

This declaration acknowledges that this paper adheres to the principles for transparent reporting and scientific rigour of preclinical research as stated in the BJP guidelines for Design & Analysis, Immunoblotting

and Immunochemistry and Animal Experimentation and as recommended by funding agencies, publishers and other organisations engaged with supporting research.

References

Canvas, version 3.4, Schrödinger, LLC, New York, NY, 2017

Molecular Operating Environment (MOE), 2016.08; Chemical Computing Group Inc., 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2016.

(2017). LigPrep, version 4.0, Schrödinger, LLC, New York, NY.

Baell JB, & Holloway GA (2010). New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *J Med Chem* 53: 2719-2740.

Ban F, Dalal K, Li H, LeBlanc E, Rennie PS, & Cherkasov A (2017). Best Practices of Computer-Aided Drug Discovery: Lessons Learned from the Development of a Preclinical Candidate for Prostate Cancer with a New Mechanism of Action. *J Chem Inf Model* 57: 1018-1028.

Ban F, Leblanc E, Li H, Munuganti RS, Frewin K, Rennie PS, *et al.*(2014). Discovery of 1H-indole-2-carboxamides as novel inhibitors of the androgen receptor binding function 3 (BF3). *J Med Chem* 57:6867-6872.

Bohl CE, Gao W, Miller DD, Bell CE, & Dalton JT (2005). Structural basis for antagonism and resistance of bicalutamide in prostate cancer. *Proc Natl Acad Sci U S A* 102: 6201-6206.

Bohl CE, Miller DD, Chen J, Bell CE, & Dalton JT (2005). Structural basis for accommodation of nonsteroidal ligands in the androgen receptor. *J Biol Chem* 280: 37747-37754.

Caboni L, Kinsella GK, Blanco F, Fayne D, Jagoe WN, Carr M, *et al.* (2012). "True" Antiandrogens-Selective Non-Ligand-Binding Pocket Disruptors of Androgen Receptor-Coactivator Interactions: Novel Tools for Prostate Cancer. *J Med Chem* 55: 1635-1644.

Chen CD, Welsbie DS, Tran C, Baek SH, Chen R, Vessella R, *et al.*(2004). Molecular determinants of resistance to antiandrogen therapy. *Nat Med* 10: 33-39.

Chi KN, Agarwal N, Bjartell A, Chung BH, Pereira de Santana Gomes AJ, Given R, *et al.* (2019). Apalutamide for Metastatic, Castration-Sensitive Prostate Cancer. *N Engl J Med* 381: 13-24.

Curtis MJ, Alexander S, Cirino G, Docherty JR, George CH, Giembycz MA, *et al.* (2018). Experimental design and analysis and their reporting II: updated and simplified guidance for authors and peer reviewers. *Br J Pharmacol* 175: 987-993.

Dalal K, Ban FQ, Li HF, Morin H, Roshan-Moniri M, Tam KJ, *et al.*(2018). Selectively targeting the dimerization interface of human androgen receptor with small-molecules to treat castration-resistant prostate cancer. *Cancer Lett* 437: 35-43.

Dalal K, Che M, Que NS, Sharma A, Yang R, Lallous N, *et al.*(2017). Bypassing Drug Resistance Mechanisms of Prostate Cancer with Small Molecules that Target Androgen Receptor-Chromatin Interactions. *Mol Cancer Ther* 16: 2281-2291.

Dalal K, Roshan-Moniri M, Sharma A, Li HF, Ban FQ, Hessein M, *et al.* (2014). Selectively targeting the DNA-binding domain of the androgen receptor as a prospective therapy for prostate cancer (vol 289, pg 26417, 2014). *J Biol Chem* 289: 33877-33877.

Dehm SM, Schmidt LJ, Heemers HV, Vessella RL, & Tindall DJ (2008). Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance. *Cancer Res* 68: 5469-5477.

- Ellinger-Ziegelbauer H, Fostel JM, Aruga C, Bauer D, Boitier E, Deng S, *et al.* (2009). Characterization and interlaboratory comparison of a gene expression signature for differentiating genotoxic mechanisms. *Toxicol Sci* 110: 341-352.
- Fizazi K, Shore N, Tammela TL, Ulys A, Vjaters E, Polyakov S, *et al.* (2019). Darolutamide in Nonmetastatic, Castration-Resistant Prostate Cancer. *N Engl J Med* 380: 1235-1246.
- Frazee AC, Pertea G, Jaffe AE, Langmead B, Salzberg SL, & Leek JT (2015). Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat Biotechnol* 33: 243-246.
- Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, *et al.* (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 47: 1739-1749.
- Guo ZY, Yang X, Sun F, Jiang RC, Linn DE, Chen HG, *et al.* (2009). A Novel Androgen Receptor Splice Variant Is Up-regulated during Prostate Cancer Progression and Promotes Androgen Depletion-Resistant Growth. *Cancer Res* 69: 2305-2313.
- Hornberg E, Ylitalo EB, Crnalic S, Antti H, Stattin P, Widmark A, *et al.* (2011). Expression of androgen receptor splice variants in prostate cancer bone metastases is associated with castration-resistance and short survival. *PLoS One* 6: e19059.
- Huang da W, Sherman BT, & Lempicki RA (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37: 1-13.
- Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, *et al.* (2002). The Ensembl genome database project. *Nucleic Acids Res* 30: 38-41.
- Huggins C, Stevens RE, & Hodges CV (1941). Studies on prostate cancer II The effects of castration on advanced carcinoma of the prostate gland. *Arch Surg* 43: 209-223.
- Jones G, Willett P, Glen RC, Leach AR, & Taylor R (1997). Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 267: 727-748.
- Joseph JD, Lu N, Qian J, Sensintaffar J, Shao G, Brigham D, *et al.* (2013). A clinically relevant androgen receptor mutation confers resistance to second-generation antiandrogens enzalutamide and ARN-509. *Cancer Discov* 3: 1020-1029.
- Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, *et al.* (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A* 106: 11667-11672.
- Khorasanizadeh S, & Rastinejad F (2001). Nuclear-receptor interactions on DNA-response elements. *Trends Biochem Sci* 26: 384-390.
- Kim D, Langmead B, & Salzberg SL (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12: 357-360.
- Lack NA, Axerio-Cilies P, Tavassoli P, Han FQ, Chan KH, Feau C, *et al.* (2011). Targeting the Binding Function 3 (BF3) Site of the Human Androgen Receptor through Virtual Screening. *J Med Chem* 54:8563-8573.
- Li D, Zhou WF, Pang JP, Tang Q, Zhong BL, Shen C, *et al.* (2019). A magic drug target: Androgen receptor. *Med Res Rev* 39:1485-1514.
- Li H, Ban F, Dalal K, Leblanc E, Frewin K, Ma D, *et al.* (2014). Discovery of small-molecule inhibitors selectively targeting the DNA-binding domain of the human androgen receptor. *J Med Chem* 57: 6458-6467.
- Li Q, Huang H, He Z, Sun Y, Tang Y, Shang X, *et al.* (2018). Regulatory effects of antitumor agent matrine on FOXO and PI3K-AKT pathway in castration-resistant prostate cancer cells. *Sci China Life Sci* 61: 550-558.

- Li Y, Chan SC, Brand LJ, Hwang TH, Silverstein KA, & Dehm SM (2013). Androgen receptor splice variants mediate enzalutamide resistance in castration-resistant prostate cancer cell lines. *Cancer Res* 73:483-489.
- Lim MY, Otto-Duessel M, He ML, Su L, Nguyen D, Chin E, *et al.*(2014). Ligand-Independent and Tissue-Selective Androgen Receptor Inhibition by Pyrvinium. *ACS Chem Biol* 9: 692-702.
- Lipinski CA, Lombardo F, Dominy BW, & Feeney PJ (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Del Rev* 23:3-25.
- Miller K (2010). Words of wisdom. Re: Antitumour activity of MDV3100 in castration-resistant prostate cancer: a phase 1-2 study. Scher HI, Beer TM, Higano CS, *et al.* Prostate Cancer Foundation/Department of Defense Prostate Cancer Clinical Trials Consortium. *Eur Urol* 58:464-465.
- Munuganti RSN, Hassona MDH, Leblanc E, Frewin K, Singh K, Ma D, *et al.* (2014). Identification of a Potent Antiandrogen that Targets the BF3 Site of the Androgen Receptor and Inhibits Enzalutamide-Resistant Prostate Cancer. *Chem Biol* 21: 1476-1485.
- Ning YMM, Pierce W, Maher VE, Karuri S, Tang SH, Chiu HJ, *et al.*(2013). Enzalutamide for Treatment of Patients with Metastatic Castration-Resistant Prostate Cancer Who Have Previously Received Docetaxel: U.S. Food and Drug Administration Drug Approval Summary. *Clin Cancer Res* 19: 6067-6073.
- O’Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, & Hutchison GR (2011). Open Babel: An open chemical toolbox. *J Cheminform* 3: 33.
- Olsson MHM, Sondergaard CR, Rostkowski M, & Jensen JH (2011). PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK(a) Predictions. *J Chem Theory Comput* 7: 525-537.
- Oprea TI (2000). Property distribution of drug-related chemical databases. *J Comput Aided Mol Des* 14: 251-264.
- Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, & Salzberg SL (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 33: 290-295.
- Sastry GM, Adzhigirey M, Day T, Annabhimoju R, & Sherman W (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* 27:221-234.
- Schuyler D (2019). FDA Approves Darolutamide in Nonmetastatic Castration-Resistant Prostate Cancer. *Clin Adv Hematol Oncol* 17: 498-498.
- Shaffer PL, Jivan A, Dollins DE, Claessens F, & Gewirth DT (2004). Structural basis of androgen receptor binding to selective androgen response elements. *Proc Natl Acad Sci U S A* 101: 4758-4763.
- Shafi AA, Yen AE, & Weigel NL (2013). Androgen receptors in hormone-dependent and castration-resistant prostate cancer. *Pharmacol Ther* 140: 223-238.
- Shelley JC, Cholleti A, Frye LL, Greenwood JR, Timlin MR, & Uchimaya M (2007). Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules. *J Comput Aided Mol Des* 21: 681-691.
- Tan MH, Li J, Xu HE, Melcher K, & Yong EL (2015). Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacol Sin* 36: 3-23.
- Tavassoli P, Snoek R, Ray M, Rao LG, & Rennie PS (2007). Rapid, non-destructive, cell-based screening assays for agents that modulate growth, death, and androgen receptor activation in prostate cancer cells. *Prostate* 67: 416-426.
- Thorne CA, Hanson AJ, Schneider J, Tahinci E, Orton D, Cselenyi CS, *et al.* (2010). Small-molecule inhibition of Wnt signaling through activation of casein kinase 1 alpha. *Nat Chem Biol* 6:829-836.

Trott O, & Olson AJ (2010). Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J Comput Chem* 31:455-461.

Walters WP, & Namchuk M (2003). Designing screens: How to make your hits a hit. *Nature Reviews Drug Discovery* 2: 259-266.

Watson PA, Arora VK, & Sawyers CL (2015). Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat Rev Cancer* 15: 701-711.

Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, *et al.* (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res* 39: W316-322.

Zhou W, Duan M, Fu W, Pang J, Tang Q, Sun H, *et al.* (2018). Discovery of Novel Androgen Receptor Ligands by Structure-based Virtual Screening and Bioassays. *Genomics Proteomics Bioinformatics* 16:416-427.

Legend of Figures

Figure 1. (A) The structure of the AR DBD. (B) The structures of the two bioactive compounds targeting the AR DBD and Cpd39. (C) Workflow of the integrated docking-based VS. (D) (Top) The AR transcriptional inhibition of Cpd39 in LNCaP-ARR2PB-eGFP, and (Bottom) The PSA suppression by these compounds was evaluated by measuring the PSA secreted into the media using the same LNCaP-ARR2PB-eGFP cells. (E) The 3D presentation of the interactions between Cpd39 and the AR DBD. Cpd39 is colored green, and the surrounding residues are colored purple. (F) The electrostatic potential of the binding site. The residues with the positive electrostatic potential are colored blue, and those with the negative electrostatic potential are colored red. (G) The 2D presentation of the interactions between Cpd39 and the AR DBD.

Figure 2. (A) The antiproliferative effect of Cpd39 on LNCaP (left) and 22RV1 cells (right) cells using MTT assay. (B) The mRNA expression levels of AR target genes including KLK3 (PSA), TMPRSS2, ACP and FKBP5 analyzed by quantitative analysis after LNCaP cells treated with 10 μ M compounds and 10 nM DHT. (C) Protein expression of AR and TMPRSS2 in LNCaP analyzed by Western blot analysis. (D) The antiproliferative effects of Cpd39 on four cell lines, including PC3, DU145, 3T3 and A549 cells. (E) Distribution of the DU145 transcriptome sequences among KEGG pathways. (F) The mRNA expression levels of RHOA, PI3KCB, ERBB2, METTL3 and EIF4B in DU145 cells. (G) Expression of AKT1 analyzed by quantitative analysis after 22RV1 cells were treated with indicated compounds. Values are means \pm s.e.m.. * $P < 0.05$ versus DMSO group ($n = 5$ in each group).

Figure 3. Biotinylated ARE was loaded onto streptavidin sensors for bilayer interferometry analysis. (A) $t = 0$ s. BLI sensors preloaded with ARE (50 nM) are baselined in buffer with the indicated compound (50 μ M); at the first inflection point (*arrow*), Sensors are moved to a solution of the AR DBD (1 μ M) alone (2.5% DMSO), or the AR DBD mixed with tested compounds for 300 s, and then moved back to buffer for dissociation. The y axis represents the nanometer shift in wavelength resulting from ligand (dsDNA) binding/dissociation. (B) Same as A, but with different concentrations of tested compounds (0-500 μ M). BLI results reflected the direct binding of the compounds to the AR DBD protein. (C) same as A, but with the Y594D mutated AR DBD. Non-specific interactions between the protein or compounds and sensors were subtracted from the BLI signal.

Figure 4. Effect of compounds on AR nuclear localization. Microscopic images of LNCaP cells treated with 10 μ M compounds and 5 nM DHT or DMSO only (scale bar = 20 μ m).

Figure 1

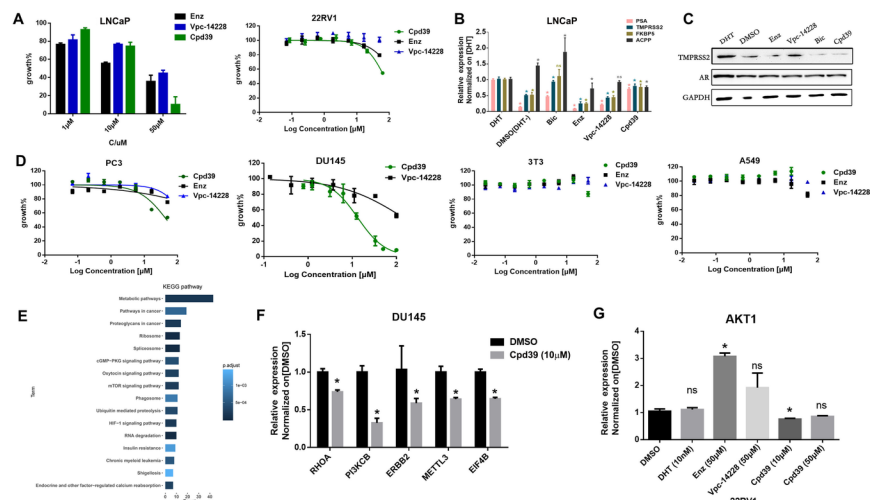


Figure 2

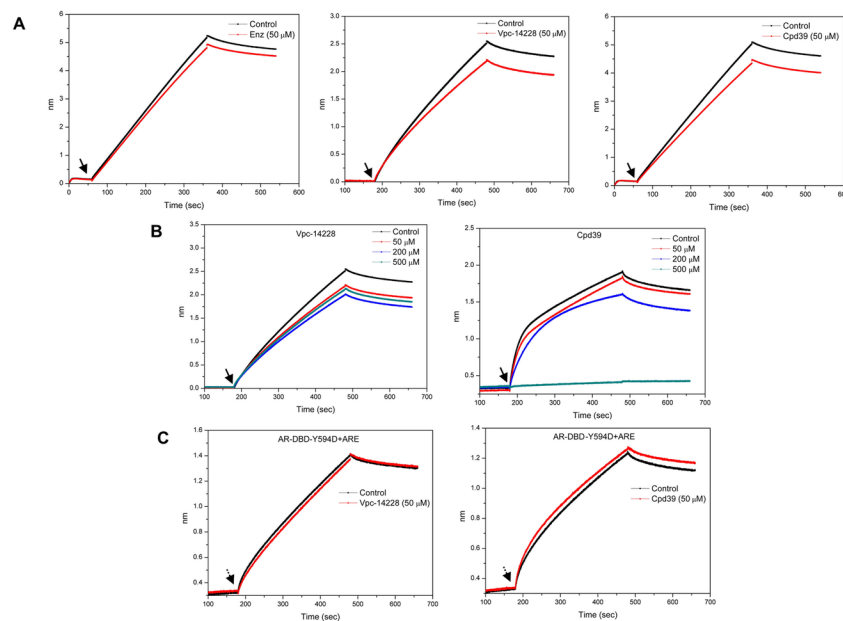


Figure 3

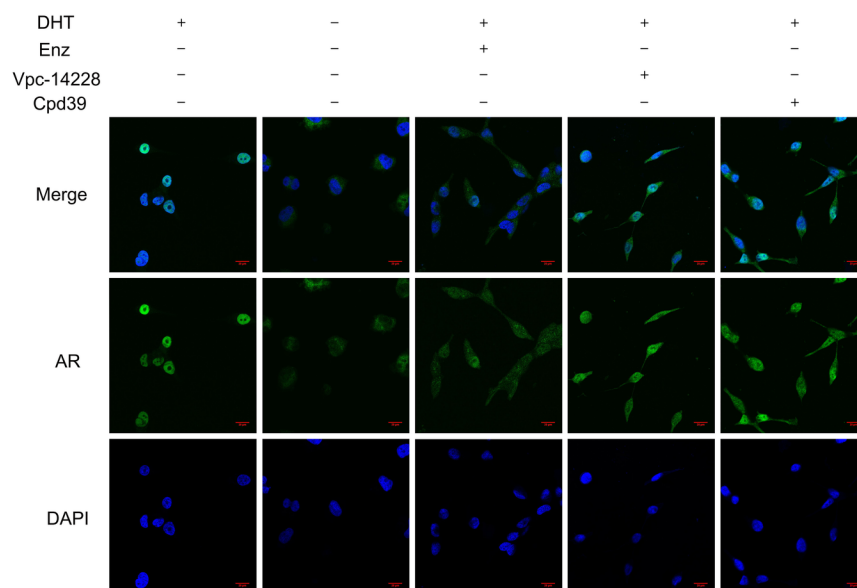


Figure 4 .

