Structures of Endocrine-Disrupting Chemicals Determine Binding to and Activation of the Estrogen Receptor α and Androgen Receptor

Haoyue Tan,† Xiaoxiang Wang,‡ Huixiao Hong, Emilio Benfenati, John P. Giesy, Giuseppina C. Gini, Rebeca Kusko, Xiaowei Zhang, Hongxia Yu, and Wei Shi*†

ABSTRACT: Endocrine-disrupting chemicals (EDCs) can interact with nuclear receptors, including estrogen receptor α (ERα) and androgen receptor (AR), to affect the normal endocrine system function, causing severe symptoms. Limited studies queried the EDC mechanisms, focusing on limited chemicals or a set of structurally similar compounds. It remained uncertain how hundreds of diverse EDCs could bind to ERα and AR and cause distinct functional consequences. Here, we employed a series of computational methodologies to investigate the structural features of EDCs that bind to and activate ERα and AR based on more than 4000 compounds. We used molecular docking and molecular dynamics simulations to elucidate the functional consequences and validated structure–function correlations experimentally using a time-resolved fluorescence resonance energy-transfer assay. We found that EDCs share three levels of key fragments. Primary (20 for ERα and 18 for AR) and secondary fragments (38 for ERα and 29 for AR) are responsible for the binding to receptors, and tertiary fragments determine the activity type (agonist, antagonist, or mixed). In summary, our study provides a general mechanism for the EDC function. Discovering the three levels of key fragments may drive fast screening and evaluation of potential EDCs from large sets of commercially used synthetic compounds.

INTRODUCTION

Endocrine-disrupting chemicals (EDCs) can cause adverse effects in humans by directly or indirectly interfering with hormone systems.1,2 EDCs’ many harmful effects spanning organ systems3,4 are ascribed to interactions with two nuclear receptor (NR) family members: estrogen receptor α (ERα) and androgen receptor (AR). The United States Environmental Protection Agency (U.S. EPA) ToxCast project showed that 12.2 and 8.4% of investigated chemicals cause harm via interactions and/or modulation of ERα or AR. EDCs trigger adverse effects via ERα/AR agonist activation, endogenous hormone antagonist repression, or generating agonist and antagonist activities simultaneously.9

Despite heroic previous attempts, it remains unknown what features or factors make chemicals active on ERα or AR as well as how these features or factors exert their functions.10 To infer important structural features driving chemical activities, researchers have used the quantitative structure–activity relationship (QSAR) method.11 Previous theoretical studies focused on specific chemical categories, and thus, the derived information is limited to said categories,12 such as hydroxylated polybrominated diphenyl ethers (HO-PBDEs)13 and bisphenol A and its analogues.14

Overcoming these deficiencies would facilitate large in silico screening of EDCs15 and would advance the understanding of EDC behaviors acting through other NRs. Here, we employed a series of computational methodologies and found that active compounds shared three levels of key characteristic fragments. Primary (20 for ERα and 18 for AR) and secondary fragments (38 for ERα and 29 for AR) are responsible for the binding to EDCs to ERα or AR, which discriminate active and inactive compounds. Tertiary fragments (66 for ERα and 56 for AR) interact with the functional lobes, directly affecting the AF-2 surface. This surface is responsible for coregulator recruitment and thus determined the activity types (agonist, antagonist, or mixed). We developed a multistep model: the NR-mediated endocrine activity model (NRMEA). NREMA can qualitatively predict compound effects on ERα or AR and provides information on characteristic fragments (https://www.veghub.eu/download/ for free download).

MATERIALS AND METHODS

EDC Dataset Collection of ERα and AR. Data from ToxCast/Tox21 and ChEMBL, including three types of in vitro
assays, were used to select active and inactive compounds following criteria 1 and 2.11,16−19

Active: Binding detected in ≥1 competitive binding assay, and activity detected in ≥1 reporter gene assay (criterion 1).

Inactive: Competitive binding assay and all reporter gene assays yielded negative results (criterion 2).

Active compounds were further classified into three different activities by criteria 3, 4, and 5.17,20

Agonist: ≥1 agonistic reporter gene assay yielded positive results, and all antagonistic reporter gene assays yielded negative results, or positive results from antagonistic assays were excluded because of cytotoxicity (criterion 3).

Antagonist: ≥1 antagonistic reporter gene assay yielded positive results and higher cytotoxicity, and all agonistic reporter gene assays yielded negative results (criterion 4).

Agonist and Antagonist (a-anta): ≥1 agonistic reporter gene assay and ≥1 antagonistic reporter gene assay yielded positive results simultaneously; moreover, positive antagonistic activity was higher than cytotoxicity (criterion 5).
Dataset S1 (ERα) and Dataset S2 (AR) contain summarized information for all compounds. More detailed definition information is in Appendix A of the Supporting Information.

Extraction of Hierarchical Characteristic Fragments of ERα and AR. To distinguish active from inactive compounds and to distinguish between the three activity types, three levels of characteristic fragments were extracted. Primary fragments were structurally required components of active compounds (small characteristic fragments, e.g., oxygen-containing aromatics and nitrogen-containing aromatics), and compounds having none were inactive. Compounds with ≥1 secondary fragment (substructures of primary fragments) were considered active. Tertiary fragments characterize different types of activity. Agonists, antagonists, and α-antas had distinct tertiary fragments.

During analysis, we randomly selected 80% of compounds with the “Partitioning Mode” of KNIME (https://www.knime.com/), an open analytics platform for innovation) as a fragment extraction training set. The remaining 20% (test set) were used for validation. More detailed information is in Appendix B of the Supporting Information.

Molecular Docking and Binding Mode Classification. We used two crystal structures [the E2-bound ERα ligand binding domain (ERα LBD) (PDB-ID: 2YJA) and the DHT-bound AR LBD (PDB-ID: 3L3X)] as active templates in molecular docking analyses. Swiss-PdbViewer reported the molecular integrity of the two proteins, and Autodock Tools 1.5.6 added hydrogens. ChemBioDraw Ultra 14.0 and Chem3D Pro 14.0 built three-dimensional (3D) ligand structures. Energy minimization was performed using SYBYL7.3 (Tripos Inc., St. Louis, MO, USA) to optimize geometries. Gasteiger–Huckel charges were assigned. We docked optimized structures of compounds into LBD binding cavities where the intrinsic small molecules are located with Autodock vina. The profile of the hydrophobic interaction and van der Waals interaction between the compound and amino acids of LBDs was obtained with the Ligplus program. We analyzed binding modes using the R ComplexHeatmap package with the Pearson correlation method. Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA) were conducted to visualize differences between ligand and receptor interactions of active and inactive compounds. The variable importance in the projection (VIP) was calculated for each amino acid to show the contribution in the ligand–receptor binding, and those amino acids with VIP ≥ 1.5 were considered to be the most relevant for the bindings. The detailed process of molecular docking and postdocking analysis are in Appendix C of the Supporting Information.

Molecular Dynamics Simulations. To explore active compound/LBD complex dynamic conformation shifts, we employed apo (unbound) conformations of ERα and AR. An experimental apo conformation of the ERα LBD (PDB-ID: 1AS2) was in the RCSB PDB database. However, the experimental apo conformation of AR LBD was not available. Thus, we used the crystallized AR LBD bound with DHT as a template (PDB-ID: 3L3X) to build the apo structure (the detailed process is described in the Supporting Information, Appendix D). Typical chemicals of each tertiary fragment (Table S13) were then tested before docking them into these two ERα and AR LBD structures to construct ligand–receptor complexes for molecular dynamics (MD) simulations. We performed MD simulations using the GROMACS 5.1.2 package, and each production simulation was run for at least 20 ns. After simulations, several complexes’ structural profiles were analyzed. Detailed information about apo-AR LBD homology modeling, the MD simulation process, and postsimulation analysis is in Appendix D of the Supporting Information.

Time-Resolved Fluorescence Resonance Energy-Transfer Assay. We performed a time-resolved fluorescence resonance energy-transfer (TR-FRET) assay to analyze the process of coregulator recruitment, and detailed information is in Appendix E of the Supporting Information.

RESULTS

Active and Inactive Compounds of ERα and AR Obtained from ToxCast/Tox21 and ChEMBL. Data from ToxCast/Tox21 and ChEMBL were categorized into three assay classes: (a) competitive binding, (b) reporter gene or transactivation, and (c) cytotoxicity assays and used to find compounds potentially affecting NR-mediated signaling (Figure 1A, Table S1). We used exclusively Homo sapiens data, and compounds were removed when cytotoxicity occurred at concentrations less than the threshold for reporter gene responses (Figure 1A, Supporting Information, Appendix A). Based on these requirements, we collated 2465 and 2845 compounds for ERα and AR, respectively (Table S2).

Active and inactive compounds were identified, and active compounds were further divided into antagonists, agonists, and α-antas as described in the Materials and Methods section. The classified compounds detailed in Table S2 cover a range of diverse structural classes, including steroids, flavonoids, phthalates, and bisphenol A (BPA) and analogues. Compounds in the same class could be active or inactive. For example, nuarimol and anilazine are both structurally similar nitrogen-containing aromatic fungicides, but the former is active for ERα and the latter is inactive (Table S3). Moreover, active compounds in the same class can exhibit various activity types (Figure 1B), such as analogues of bisphenols, flavonoids, or alklyphenols. This phenomenon also occurred in the case of AR (Figure 1B). Taken together, these data imply that structural similarity methods of previous QSAR models do not effectively predict active versus inactive compounds from a large chemical set.

We extracted essential structural fragments to distinguish active and inactive compounds, differentiate the three activity types, and further explore the mechanism.

Structural Features of Active Compounds of ERα and AR. By using 80% of the compounds studied (training set) based on the hierarchy featured fragment method, we first profiled characteristic fragments that made compounds active toward ERα and AR. For ERα active compounds, we identified 20 primary fragments divided into four types: (1) oxygen-containing aromatics, (2) aromatics, (3) oxygen-containing chains, and (4) carbon chains (Figure 1C, Table S4). Of the active ERα compounds, 99.29% possess at least one primary fragment (Table S10), while only 45.93% of compounds containing at least one of the primary fragments are active. More specific complex fragments characterizing active compounds (secondary fragments) should be further extracted based on primary fragments. We identified 38 ERα secondary fragments (Figure 1C, Table S5). Of the active compounds, 96.04% had at least one primary fragment and one secondary fragment, while 97.56% of the inactive compounds had none of the fragments or only primary fragments. The coexistence of primary and secondary fragments was necessary to produce activity. For example, TDBP, BPA, TGSH, and DiPE
Interactions between Primary/Secondary Fragments and LBDs of ERα and AR. To gain structural insights into the binding modes of active ERα and AR compounds and to identify the roles of characteristic fragments, molecular docking was used. Based on ligand–receptor complex conformations, we studied hydrogen-bond interactions and van der Waals interactions between each compound and found critical amino acids related to ligand–receptor interactions. Although compounds occupied a similar region or binding site in LBDs (Figures 2A and S1A), the interactions of different compounds with the amino acids of LBDs vary drastically.

Performing an unsupervised hierarchical cluster analysis on interactions between the ERα LBD and ligands, we found that interactions with R394 and H524 were key factors in making ligands active (Figure 2B). According to supervised OPLS score scatter analysis (Figure S2A), R394 and H524, with VIP values ≥ 1.5 (Figure S2B), were found to be significant contributors to activities (Figure 2B, red). The VIP values of the other amino acids were all <1.3. The majority of compounds with secondary fragments stabilized conformations in the LBD by forming hydrogen bonds with R394 and H524 (coefficient plot, Figure S2C). Compounds without secondary fragments do not significantly interact with these two amino acids. Thus, the ability to interact with these two amino acids may make secondary fragments to be characteristic structures of active compounds. The significance of these two amino acids has been reported previously.15 Of the 241 ligand-ERα LBD crystal structures in PDB, 180 ones have hydrogen bonds between ligands and amino acid R394, and 92 ones have hydrogen bonds between ligands and H524. These two polar amino acids are involved in a key hydrogen-bond network connecting H5 and H11, with both a part of the docking site that maintains H12 in the active position.14,25–27 Active compounds with the same secondary fragments exhibited
similar interactions with ERα's LBD (Figure S3), suggesting that secondary fragments determined the pattern of active compound binding with the ERα LBD. Various secondary fragments have their own tendency to interact with ERα LBD amino acids (Table S12). The corresponding ligand−receptor interactions of active compounds are presented in Figure 2C,D.

With the same methods, we also performed postdocking analyses for ligand−receptor compound complexes and AR LBD. N705 and T877 had the greatest VIP values (≥ 1.5, Figure S4), suggesting that they were significant determinants of activity (Figure S1B, red). However, only 14.22 and 25.61% of the compounds had hydrogen bonds with these two amino acids. A majority of active compounds that stabilized themselves in AR’s LBD do so independently of hydrogen bonds (Figure S5) but via van der Waals interactions with seven amino acids (Leu701, F876, L873, Q711, L880, M780, and Met895) (Figure S4). After determining the critical amino
acids for activity, we examined the comprehensive interactions between active compounds and amino acids in AR’s LBD. The results were consistent with ERα’s, where active compounds with the same secondary fragments have similar interactions with the LBD (Figure S5). All secondary fragments generate interactions, primarily van der Waals interactions, with two key amino acids [N705 and T877 (Table S12)]. For example, active compounds with secondary fragment 1 resulted in consistent ligand−receptor interactions (Figure S1C, left), producing entirely van der Waals binding. Active compounds with secondary fragment 2 exhibited different interactions (Figure S1C, right): they form strong hydrogen-bond interactions with N705, Q711, R752, F764, and T877 and van der Waals contacts with other amino acids. In summary, there are distinct differences between active/inactive compound interactions and LBDs. Furthermore, active compounds with the same secondary fragments result in similar ligand−receptor interactions.

Tertiary Fragments Induce Three Conformations for Coregulator Recruitment. We employed MD simulations to track the dynamic interactions between ERα and AR’s LBDs and a selected compound series covering all types of tertiary fragments (60 compounds for ERα and 52 compounds for AR, Table S13). Locations of tertiary compound fragments and entire compounds, as well as AF-2 surface dynamic changes, could be key activity type determinants. All ERα agonists stabilized in both ends of the binding pockets at location N−H5−H3−C (the N-terminal of H5 to the C-terminal of H3) and location C−H11 (the C-terminal of H11, Figure 3). The significance of these ends has been reported before. For ERα agonists, 68% of the tertiary fragments interacted directly with the abovementioned two locations (Figure 3A, Table S13), forming hydrogen bonds or van der Waals contacts (Table S14). Other tertiary agonist fragments interact with these two locations indirectly by acting as the central skeleton to swing two sides of compounds within these fragments to one of the locations. The interactions of agonists and amino acids in these two locations can influence the conformation of the AF-2 surface, driving the transduction signals from ligand binding, recruiting coregulators and regulating gene expression. Conformational superposition shows that the agonist-induced AF-2 surface analyzed in this study is similar to the surface formed by EE2 and other endogenous estrogens (Figure 3B). To characterize the dynamic AF-2 surface changes, we used the distances and center of mass angles (COMs) for H3, H5, and H12 (Figure 3C). The ERα LBD AF-2 surface with agonists was highly stabilized as the distance ranges of H3−H12 (∼5 Å) and H5−H12 (∼5 Å) were small (Figures 3D and S6). Antagonist tertiary fragments were not generally confined to two locations (Figure 3A), and 77% of these fragments were chemical skeletons (Tables S13 and S14). Correspondingly, the AF-2 surface also fluctuated slightly (Figures 3D and S7). H12 occupied the coactivator-binding groove and acted as a coactivator to interact with H3 and H5 and restrain
recruitment of both coactivators and corepressors (Figure 3B). This pattern was similar to the case of the OHT-bound LBD (PDB-ID: 3ERT). For a-antas of ERα, half of the tertiary fragments were also located in N−H5−H3−C and C−H11, while a few compounds stabilized both locations (Figure 3A, Tables S13 and S14). The majority of a-antas present an orientation close to H12. This orientation is different from that of agonists and might exert steric effects on H12, potentially perturbing the conformation of the H12/AF-2 surface.29 The AF-2 clash weakens the stability of H12,29,30 causing unpredictability in transcription, which depends on the coregulator types. Correspondingly, the AF-2 surface was also much more dynamic than those from agonists and antagonists (Figures 3D and S8).

We performed a similar analysis for AR LBD (Figure S9, Table S13). All agonist tertiary fragments stabilized at the ends of N−H5−H3−C and C−H11 (Figure S9A), and 82% of the compounds fully occupied both ends. Structural superposition revealed that agonists displayed the canonical DHT-agonist conformation, with H12 capping the LBD (Figure S9B). The majority of tertiary fragments of antagonists were chemical skeletons. The distance from H12 to H3 and H5 remained at ≈15 Å (Figures S9C and S11), implying that the AF-2 surface was very stable. A total of 83% of a-anta tertiary fragments swung to the end of the N-terminal, among which three a-antas occupied both locations. This situation is similar to ERα a-antas. As a result, H12 for a-antas was kept away from H3 (≈40 Å) and H5 (≈30 Å) and generated a wild fluctuation (Figures S9C and S12) compared with agonists and antagonists.

Moreover, all active compounds simulated in this study and compounds containing the same characteristic fragments were similar to crystal structures in the RCSB PDB database. Comparison of the simulation results to crystal structures showed that 51 ERα crystal structures and 49 AR crystal structures were consistent with our simulations, whether from the active conformation or ligand binding modes (Table S13) for all PDB-IDs of crystal structures, suggesting the reliability of our studies.

**Cofactor Recruitment by Active Compounds of ERα and AR.** We used the TR-FRET assay28 to characterize how different types of tertiary fragments affect the process of coregulator recruitment and thus determine the activity types. Fifteen and 12 typically active ERα and AR compounds were selected as model compounds (Table S16), all of which contain 8 and 7 diverse ERα and AR tertiary fragments, respectively. We investigated the binding of two well-known coregulatory proteins31,32 to ligand-ERα complexes including ERα coregulatory proteins31,32 to ligand-ERα complexes (Table S16), all of which contained 8 and 7 diverse ERα and AR tertiary fragments, respectively. We investigated the binding of two well-known coregulatory proteins31,32 to ligand-ERα complexes including pERα crystal structures and 49 AR crystal structures were consistent with our simulations, whether from the active conformation or ligand binding modes (Table S13) for all PDB-IDs of crystal structures, suggesting the reliability of our studies.

We performed a similar analysis for AR LBD (Figure S9, Table S13). All agonist tertiary fragments stabilized at the ends of N−H5−H3−C and C−H11 (Figure S9A), and 82% of the compounds fully occupied both ends. Structural superposition revealed that agonists displayed the canonical DHT-agonist conformation, with H12 capping the LBD (Figure S9B). The majority of tertiary fragments of antagonists were chemical skeletons. The distance from H12 to H3 and H5 remained at ≈15 Å (Figures S9C and S11), implying that the AF-2 surface was very stable. A total of 83% of a-anta tertiary fragments swung to the end of the N-terminal, among which three a-antas occupied both locations. This situation is similar to ERα a-antas. As a result, H12 for a-antas was kept away from H3 (≈40 Å) and H5 (≈30 Å) and generated a wild fluctuation (Figures S9C and S12) compared with agonists and antagonists.

Moreover, all active compounds simulated in this study and compounds containing the same characteristic fragments were similar to crystal structures in the RCSB PDB database. Comparison of the simulation results to crystal structures showed that 51 ERα crystal structures and 49 AR crystal structures were consistent with our simulations, whether from the active conformation or ligand binding modes (Table S13) for all PDB-IDs of crystal structures, suggesting the reliability of our studies.

We performed a similar analysis for AR LBD (Figure S9, Table S13). All agonist tertiary fragments stabilized at the ends of N−H5−H3−C and C−H11 (Figure S9A), and 82% of the compounds fully occupied both ends. Structural superposition revealed that agonists displayed the canonical DHT-agonist conformation, with H12 capping the LBD (Figure S9B). The majority of tertiary fragments of antagonists were chemical skeletons. The distance from H12 to H3 and H5 remained at ≈15 Å (Figures S9C and S11), implying that the AF-2 surface was very stable. A total of 83% of a-anta tertiary fragments swung to the end of the N-terminal, among which three a-antas occupied both locations. This situation is similar to ERα a-antas. As a result, H12 for a-antas was kept away from H3 (≈40 Å) and H5 (≈30 Å) and generated a wild fluctuation (Figures S9C and S12) compared with agonists and antagonists.

Moreover, all active compounds simulated in this study and compounds containing the same characteristic fragments were similar to crystal structures in the RCSB PDB database. Comparison of the simulation results to crystal structures showed that 51 ERα crystal structures and 49 AR crystal structures were consistent with our simulations, whether from the active conformation or ligand binding modes (Table S13) for all PDB-IDs of crystal structures, suggesting the reliability of our studies.

**Cofactor Recruitment by Active Compounds of ERα and AR.** We used the TR-FRET assay28 to characterize how different types of tertiary fragments affect the process of coregulator recruitment and thus determine the activity types. Fifteen and 12 typically active ERα and AR compounds were selected as model compounds (Table S16), all of which contain 8 and 7 diverse ERα and AR tertiary fragments, respectively. We investigated the binding of two well-known coregulatory proteins31,32 to ligand-ERα complexes (Figure 4B). As the concentrations of tested compounds are lower than 10^−6 mol L^−1, no cytotoxicity was detected in both in vitro assays. Therefore, under a limitation of 10^−6 mol L^−1, we chose three concentrations (10^−6, 10^−7, and 10^−8 mol L^−1) for TR-FRET assays. Active compounds with the same tertiary characteristic fragments exhibited similar binding profiles (Figures S13–S16). The binding affinities of all compounds at 10^−6 mol L^−1 were summarized, and all agonists (EE2, estrone, 17OHP, norethindrone, and BP-2) increased the affinity of PGC-1α and decreased the affinity of SMRT ID2. The antagonists OHT, BM, HCP, bicalutamide, and chlorhexidine decreased the affinity for PGC-1α and SMRT ID2. Five a-antas with similar tertiary fragments (TCBPA, HPTE, BPFA, BPA, and BPB) did not weaken the SMRT ID2 affinity.

We analyzed 12 AR activating compounds. In the presence of DHT and CPA (positive controls), we used two coregulatory proteins that influence AR-mediated transcription (the D11FxxLF coactivator and the SMRT ID2 corepressor) to quantify the potency of AR-coregulator binding.33,34 We observed similar results for AR-active compounds to the aforementioned ERα active compounds. Compared to the unliganded apo-AR LBD, compounds with binding affinity to coregulators demonstrated dose-dependent binding (Figures S18–S21). Active compounds with the same tertiary fragments had similar binding profiles for both concentrations. The antagonists norethindrone, 17OHP, and corticosterone increased the D11FxxLF affinity (Figure 4C) and decreased the SMRT ID2 affinity at a concentration of 10^−6 mol L^−1 (Figure 4D). The antagonists (BPFA, THPE, BPB, hexestrol, linuron, procymidine, and fennhexamid) displayed a partial opposite coregulatory affinity profile compared to agonists, decreasing the affinity of SMRT ID2 and D11FxxLF. The a-antas with similar tertiary fragments (spironolactone and estrone) caused increasing affinity for D11FxxLF and SMRT ID2. Tertiary fragments directly affected the conformation of ligand−receptor complexes, resulting in three types of coregulator recruitment, driving the activity type (agonist, antagonist, or mixed).

**DISCUSSION**

In this study, we attempted to clarify the previously unknown relationships between the structures and activities of ERα and AR-bound ligands.35 Compounds with similar chemical structures can have different abilities to activate these two receptors.36−38 The characteristic fragments presented here offer mechanism-based profiling by the use of chemical similarities. For example, steroids and 10 diverse molecules all belong to polar atom-containing carbon chains, while only steroids contain secondary fragments and thus have AR activities (Figure S22A). Steroids with long carbon-chain tertiary fragments can be a-antas, while compounds without long carbon chains were only agonists. It has been hypothesized but not clearly proven nor refuted that compounds containing oxygen or nitrogen atoms, such as DDT,39 triolocarbon,40 PBDEs,41 and their analogues are hypothetic endocrine disruptors. Here, oxygen- and nitrogen-containing aromatics were the most critical primary structural features of hundreds of potential active compounds and respectively essential for ERα and AR binding. A total of 966 (819) compounds within oxygen (nitrogen) atom-containing fragments can activate ERα (AR), while 983 (1308) compounds with oxygen (nitrogen) atoms but no secondary fragments were inactive (Figure S22B,C). Combined with mechanistic analyses, comprehensive secondary fragments forming interacting networks with LBD amino acids were the basis for the observed activities. Polar atoms (oxygen and nitrogen) played important roles in secondary fragments. The simultaneous existence of primary and secondary fragments but not atoms was decisive for determining the activities.

MD simulations revealed dynamic fragment action in receptors. Ligand fragments interacted with LBDs and caused changes in the conformation of the AF-2 surface, potentially impacting coregulator recruitment and transcriptional activity. Stabilization of H12 determined the conformation of the AF-2...
According to root-mean-square deviation (rmsd) evaluation of α-carbon atomic positions (Supporting Information, Appendix D), high H12 stabilization was observed for ERα/AR-agonist complexes (Figure S23). However, H12 fluctuated greatly in ERα/AR-antagonist complexes. Consistent with the fluctuating range of H3–H12 and H5–H12 distances, rmsd showed large deviations over the simulation and absolute values rose over time in the LBDs with α-antas.

The ERα and AR fragment hierarchy profiles were similar but not identical. ERα and AR primary fragments could be grouped into four types: (1) polar atom-containing aromatics, (2) aromatics, (3) polar atom-containing chains, and (4) carbon chains. For example, several secondary fragments can interact with both receptors: the bisphenol group (BPA and its analogues) and the long carbon chain (steroids) (Table S15). Active ERα compounds had many more secondary fragments and tertiary fragments than the active AR compounds and greater structural diversity and complexity. Comparative analyses implied that result extrapolation between NRs should be questioned. 4-Benzylphenol is an a-anta-specific ERα tertiary fragment and an antagonist-specific AR tertiary fragment. The characteristic fragment activation mechanisms of ERα and AR were similar. R394 and H524 were key amino acids for ERα activation. Tertiary ERα fragments stabilized active compounds in the two locations and resulted in different activation types. This phenomenon was also observed for AR-active compounds. N705 and T877, as critical amino acids in H5–H3–C and C–H11, respectively. The interaction between tertiary fragments for AR and the two locations ultimately dominated the types of activation.

Next, we used fragment hierarchy information to develop the NRMEA model to predict small-molecule binding to two receptors and the induced functional outcome at the cellular level (Figure S24). The activity of compounds in the test set was predicted (Datasets 1 and 2), with the sensitivity, specificity, and accuracy of this prediction. We compared NRMEA’s ERα predictions to IRFMN (estrogen receptor relative binding affinity model) and IRFMN/CERAPP (ER-mediated effect). The results of the comparison indicated that the NRMEA model had greater predictive power (Table S17). No well-developed AR prediction model is currently available. Therefore, 15 active/inactive binary classification models were developed with QSAR. These models were built using three molecular descriptor databases and five machine-learning methods (Table S17). Clearly, NRMEA had superior performance. Reference chemicals were used to validate ERα/AR in vitro assays.53,44 The activity (active/inactive, activity type) of the majority of these compounds was successfully predicted by NRMEA (Table S18). NRMEA is currently freely available in VEGA v.20 (Figure S24) and shows promise for improving virtual screening performance. The findings reported here should inspire future work to develop and characterize other NR-mediated disruptors qualitatively and quantitatively and other molecular mechanisms (e.g., receptor-DNA binding) of active compounds to probe more functions of the NR family.

### Associated Content

#### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.0c02639.

Classification methods and potency categorization, extraction of hierarchy-featured fragments of ERα and AR, molecular docking and binding mode classification, MD simulations, and TR-FRET assay (PDF)

Experimental and predicted activities of compounds for ERα (XLSX)

Experimental and predicted activities of compounds for AR (XLSX)

### Author Information

#### Corresponding Author

Wei Shi — State Key Laboratory of Pollution Control and Resources Reuse, School of the Environment, Nanjing University, Nanjing 210023, Jiangsu, China; orcid.org/0000-0001-9499-818X; Email: njushwei@nju.edu.cn

#### Authors

Haoyue Tan — State Key Laboratory of Pollution Control and Resources Reuse, School of the Environment, Nanjing University, Nanjing 210023, Jiangsu, China; orcid.org/0000-0003-1217-4588

Xiaoxiang Wang — State Environmental Protection Key Laboratory of Integrated Surface Water-Groundwater Pollution Control, School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China

Huixiao Hong — National Center for Toxicological Research US Food and Drug Administration, Jefferson 72679, Arkansas, United States; orcid.org/0000-0001-8087-3968

Emilio Benfenati — Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Sciences, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan 20156, Italy; orcid.org/0000-0002-3976-5989

John P. Giesy — Toxicology Centre and Department of Veterinary Biomedical Sciences, University of Saskatchewan, Saskatoon S7N 5B3, Canada; Department of Environmental Sciences, Baylor University, Waco 76706, Texas, United States

Giuseppina C. Gini — Department of Electronics and Information, Politecnico di Milano, Milano 20133, Italy

Rebeca Kusko — Immunocorporation, Cambridge 02142, Massachusetts, United States

Xiaowei Zhang — State Key Laboratory of Pollution Control and Resources Reuse, School of the Environment, Nanjing University, Nanjing 210023, Jiangsu, China

Hongxia Yu — State Key Laboratory of Pollution Control and Resources Reuse, School of the Environment, Nanjing University, Nanjing 210023, Jiangsu, China

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.0c02639

#### Author Contributions

* H.T. and X.W. contributed equally.

#### Notes

The authors declare no competing financial interest.

#### Acknowledgments

This work was supported by the National Key R&D Program of China (2018YFC1801604), the National Natural Science Foundation of China (21922603), the Science Fund for Excellent Young Scholars of Jiangsu Province (BK20170077), the Fundamental Research Funds for the Central Universities (021114380139), and the National Water Pollution Control
and Treatment Science and Technology Major Project (2017ZX07202-001 and 2017ZX07602-002). We appreciate Qinchang Chen, at Tongji University, for his technical guidiance on computational approaches and the TR-FRET experiment. We also thank Yi Cao at Nanjing University for his discussions, comments, and suggestions. The computational calculations were performed in the High Performance Computing Center (HPCC) of Nanjing University and the National Supercomputing Center in Shenzhen.

**REFERENCES**


