Science of the Total Environment 955 (2024) 177145

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Prediction of cytotoxicity of polycyclic aromatic hydrocarbons from first principles

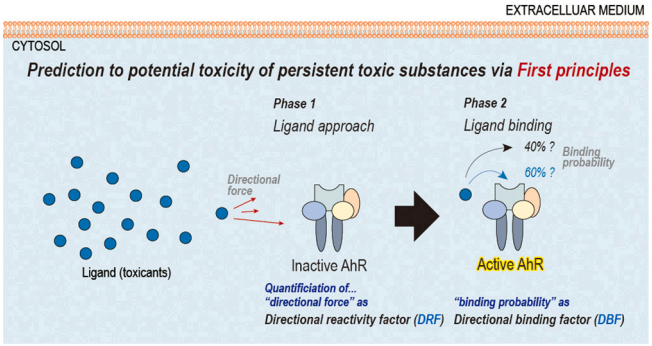
Taewoo Kim [a], Juyuan Zhen [b], Junghyun Lee [c], Shin Yeong Park [a], Changkeun Lee [a], Bong-Oh Kwon [d], Seongjin Hong [e], Hyeong-Moo Shin [f], John P. Giesy [f,g,h], Gap Soo Chang [b,*], Jong Seong Khim [a,*]

[a] School of Earth and Environmental Sciences & Research Institute of Oceanography, Seoul National University, Seoul 08826, Republic of Korea
[b] Department of Physics and Engineering Physics, University of Saskatchewan, Saskatoon, SK, S7N5E2, Canada
[c] Department of Environmental Education, Kongju National University, Gongju 32588, Republic of Korea
[d] Department of Marine Biotechnology, Kunsan National University, Kunsan 54150, Republic of Korea
[e] Department of Marine Environmental Sciences, Chungnam National University, Daejeon 34134, Republic of Korea
[f] Department of Environmental Science, Baylor University, Waco, TX 76798, United States
[g] Department of Veterinary Biomedical Sciences & Toxicology Centre, University of Saskatchewan, Saskatoon, Saskatchewan S7N5B3, Canada
[h] Department of Integrative Biology and Center for Integrative Toxicology, Michigan State University, East Lansing, MI 48824, United States

HIGHLIGHTS

- We presented simple, accurate models to predict binding of ligands to AhR.
- Development of a directional reactivity factor (DRF) as a prediction indicator.
- A directional reactive binding factor by calculating a binding rate between ligands and AhR.
- Greater accuracy than conventional toxicity prediction model (QSAR, docking model).
- The model highlighted a more profound mechanism of chemical toxicity.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Henner Hollert

Keywords:
Aryl hydrocarbon receptor-mediated activity
H4IIE-*luc* transactivation bioassay
Polycyclic aromatic hydrocarbons
Density functional theory
First principle

ABSTRACT

Ligand-specific binding interactions of xenobiotics with receptor proteins form the basis of cytotoxicity-based hazard assessment. Computational approaches enable predictive hazard assessment for a large number of chemicals in a high-throughput manner, minimizing the use of animal testing. However, *in silico* models for predicting mechanisms of toxic actions and potencies are difficult to develop because toxicity datasets or comprehensive understanding of the complicated kinetic process of ligand-receptor interactions are needed for model development. In this study, a directional reactive binding factor (DRBF) model based on first principles was used to predict cytotoxicity potencies of agonists of the aryl hydrocarbon receptor (AhR) for 16 different polycyclic aromatic hydrocarbons (PAHs). Molecular dynamics were simulated by accounting for the directional configuration factor toward receptor protein and the factor of binding to the Per-Arnt-Sim (PAS) domain. When

* Corresponding authors.
E-mail addresses: gapsoo.chang@usask.ca (G.S. Chang), jskocean@snu.ac.kr (J.S. Khim).

https://doi.org/10.1016/j.scitotenv.2024.177145
Received 11 April 2024; Received in revised form 11 October 2024; Accepted 20 October 2024
Available online 1 November 2024

comparing the experimental results of toxic potencies from *in vitro* bioassays with the predictions among two different *in silico* models, including quantitative structure-activity relationship (QSAR) and molecular docking models, the DRBF model exhibited the highest model performance ($R^2 = 0.90$ and $p < 0.01$). Our results showed that the DRBF model based on first principles and molecular and computational structural biology could serve as a novel framework to advance next generation hazard assessment for high-throughput screening of chemical substances.

1. Introduction

A large number of synthetic chemicals that are introduced to commerce or released from anthropogenic activities pose growing concern over human health and the environment, due to their potential acute and chronic toxicity. These chemicals are also known to cause or have various adverse effects, such as cytotoxicity, endocrine disruption, immune abnormalities, carcinogenesis, and lethality (AnvariFar et al., 2018; Blanchard et al., 2007). The strategy for assessing potential for adverse effects and hazard identification of chemicals has long been based on a complex set of *in vivo* animal testing. However, it is under increasing pressure to address several challenges today: i) to test a large number of existing and newly-made chemicals, many of which lack baseline toxicity data; ii) to monitor an ever-increasing list of chemical substances; iii) to minimize the use of animals for toxicity testing in accordance with the global movement for replacing animal testing; and iv) to develop high-throughput screening methods which are cost- and time-effective (Ukelis et al., 2008; Xu et al., 2021). Since conventional hazard assessment frameworks primarily relying on *in vivo* mammalian tests could not effectively meet such demands, a transition from the use of experimental animals and apical endpoints toward cell-based *in vitro* testing, based on the mode of action (MOA) and adverse outcome pathway (AOP) framework is underway (Fig. 1A). More recently, attempts of high-throughput computational models for reliable estimation of toxic potentials have voiced a further paradigm shift (Kavlock et al., 2008; Richard et al., 2008). Newer approaches focusing on both hazard

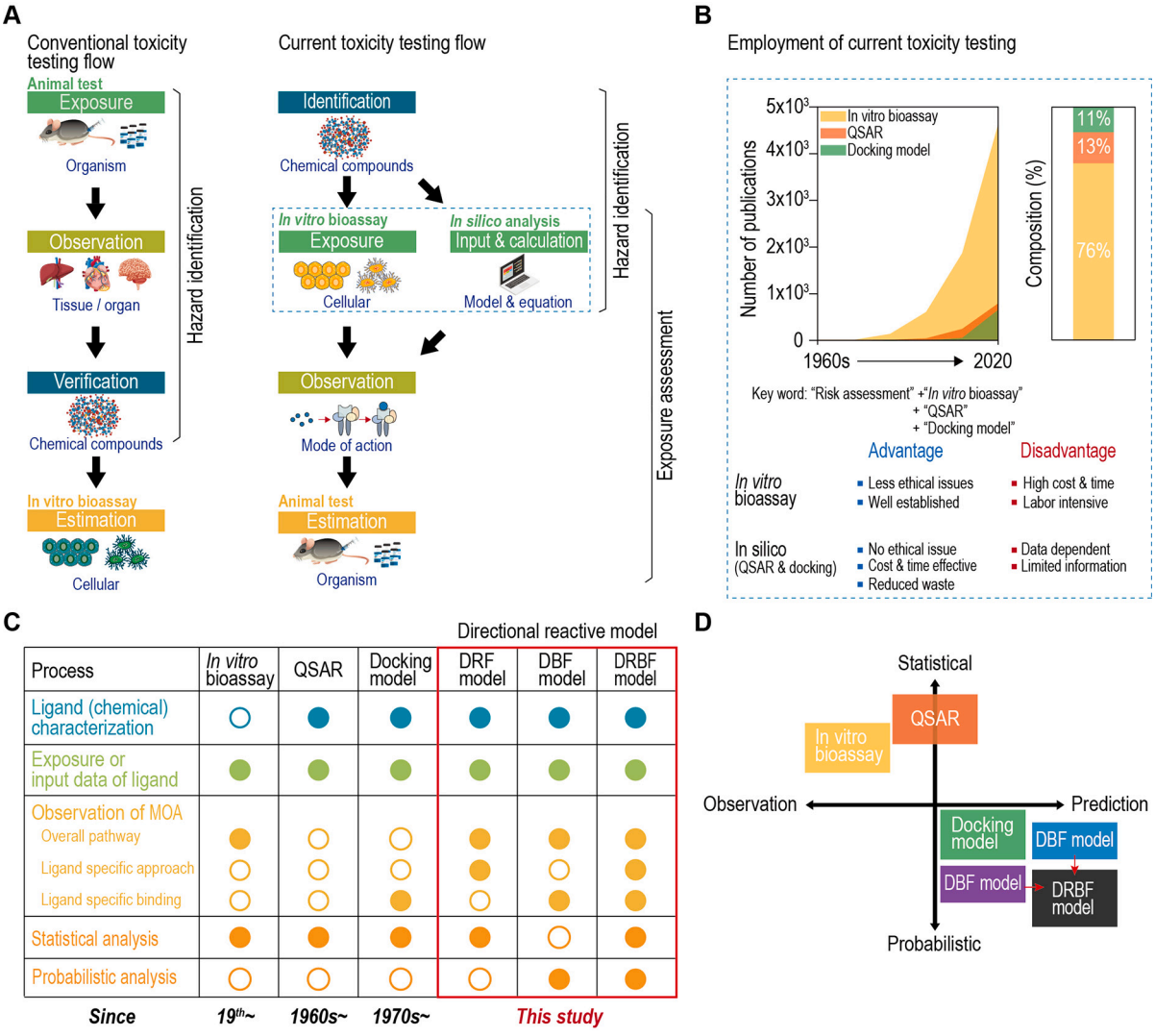


Fig. 1. Overview of methodologies used for conventional and proposed next-generation hazard assessment of organic compounds. (A) Toxicity testing flow for previous studies and the proposed toxicity testing; (B) Number of publications reporting predicted toxic potencies of persistent toxic substances (PTSs) by three types of *in silico* methods since the 1960s. (C) Characteristics of the three commonly-used toxicity modeling methods and three proposed toxicity modeling methods; (D) A quadrant plot from the perspectives of the observation–mechanistic approach and empirical–hypothetical in each method used to predict toxicity.

identification and exposure assessment have led to a better understanding of potential effects of synthetic chemicals on humans and the environment (Shin et al., 2015).

Although *in vitro* bioassays with relatively lesser ethical issues have been well established for assessing hazard and risks, bioassays utilizing recombinant cells, such as the H4llE-luciferase bioassay, H4llE-*luc*) are still relatively cost- and labor-intensive. The testing demand for a large number of potential hazardous substances inspired the emergence of computational modeling methods, such as quantitative structure-activity relationship (QSAR) and molecular docking models as alternative or complementary strategies (Chen and Ung, 2001; Politi et al., 2014; Tropsha, 2010). Strategies implementing QSAR and docking models have substantially increased since the 1990s, while in 2020 *in vitro* bioassays still account for three-quarters of toxicity assessment studies (Fig. 1B). One of the most frequently asked questions raised in QSAR research is how large a training set is needed to develop reliable predictive relationships. The answer depends on diverse factors, among which the most important determinant is the quality of the existing toxicity datasets. As regression-based models such as QSAR do not consider the detailed processes of bioactivity mechanisms but relied on the existing data, they could overestimate or underestimate the potential hazard of new target substances (Marzo et al., 2020). On the other hand, the docking model utilizes physico-chemical properties of ligands and receptors to estimate an optimized ligand binding status. When receptor-ligand binding is influenced by the whole process of molecular interaction kinetics since the introduction of ligand molecules into a recombinant cell, the sole consideration of resultant optimal binding status may not suffice to address the ligand-specific bioactivity (Yunta, 2016).

Besides, these three testing or modeling frameworks still rely upon statistical analyses of observed biological responses in the context of hazard assessment. For instance, because *in vitro* bioassays determine causes of adverse outcomes caused by individual compounds without ligand characterization and bioactivity mechanism analysis, they cannot reveal why binding events are ligand-specific (Leusch et al., 2010). The toxicity/bioactivity estimations from QSAR are made by statistical inference through descriptor-based extra- and interpolation of experimental data (Cherkasov et al., 2014; Cronin and Schultz, 2003). As a result, the lack of observation in the MOA makes an accurate estimation for a target substance difficult. The docking model has been demonstrated to be a useful tool for producing ligand-receptor complexes and estimating the binding affinity of a target substance (Friesner et al., 2006; Wei et al., 2002; Zacharias, 2003). However, because *in vitro* bioassays also involve kinetics of molecular motion of toxic substances while approaching the receptor protein, the sole consideration of the best binding affinity might be insufficient to accurately represent experimental results (Fig. 1C). With the advancement of computational capabilities, *in situ* methods for toxicity prediction are being actively developed, and models utilizing first-principles approaches have also been proposed as effective tools for toxicity assessment (Forrest et al., 2014). According to the recently proposed directional reactivity factor (DRF) model, experimental results of AhR-mediated potency for chrysene homologues from H4IIE-*luc* bioassay are significantly and strongly correlated with pre-binding factors governed by substances and receptors, such as electric dipole-driven alignment of a ligand and the location of reactive sites of the ligand with respect to the line of force action between a ligand and a receptor (Kim et al., 2020). This finding gives a clue to the evolution of hazard assessment from data-driven predictive models to first principles predictive models that account for molecular interaction kinetics resulting in the ligand-receptor binding. (Fig. 1D).

Here, we propose a novel, first principle approach referred to as a directional reactive and binding factor (DRBF) model that can predict toxic potencies of chemical substances interacting with the AhR. We applied the DRBF model to predict the toxic potencies of 16 polycyclic aromatic hydrocarbons (PAHs) without the aid of the existing experimental toxicity data (see Supporting Information (SI), Table S1). We selected these 16 PAHs to assess the DRBF model, because 1) some PAHs are known to be carcinogenic, 2) they are widely detected in the environment and urine samples of the general U.S. population, and 3) all 16 PAHs are included in the U.S. Environmental Protection Agency (EPA) priority pollutant list based on their environmental persistence, bioaccumulation, and toxicity (Andersson and Achten, 2015; Yan et al., 2004). Predicted toxic potencies of the 16 PAHs were then compared to experimental bioassay results from the H4llE-*luc* cell lines.

2. Materials and methods

2.1. Selection of study chemicals

PAHs are emitted into the atmosphere as byproducts of incomplete combustion of fossil fuels or from volcanic eruptions and forest fires (Wang et al., 2015). The selected compounds represent a range of physico-chemical properties. These are naphthalene (Nap), acenaphthene (Ace), acenaphthylene (Acl), phenanthrene (Phe), fluorene (Flu), anthracene (Ant), fluoranthene (Fla), pyrene (Pyr), benzo[*a*]anthracene (BaA), chrysene (Chr), benzo[*a*]pyrene (BaP), benzo[*b*]fluoranthene (BbF), benzo[*k*]fluoranthene (BkF), dibenzo[*a,h*]anthracene (DahA), indeno[1,2,3-*cd*]pyrene (IcdP), and benzo[*g,h,i*]perylene (BghiP). Their full names, molecular weight, and octanol-water partition coefficient (log Kow) are shown in Table S1.

2.2. Density functional theory calculations

Physico-chemical properties of the 16 PAHs were calculated based on *ab initio* density functional theory (DFT) by using the ORCA program package (Neese, 2012). Every stage of DFT computations was done by employing Becke three-parameter Lee-Yang-Parr (B3LYP) exchange functional with the polarized triple-zeta valence (def2-TZVPP) basis set. The first stage (spin-restricted) geometry optimizations were carried out for neutral structures, and then $\pm e$ charge was added for the cationic or anionic state of each molecule as the second stage (spin-unrestricted) optimizations. From the geometry-optimized structures, molecular orbitals (MO) (Fig. S1), charge populations (Mulliken and Hirshfeld), the density of states (DOS), electric dipole moments (DM), and vibrational amplitudes (VA) were computed.

2.3. Molecular dynamics for calculating the probability of binding in ligand:AhR

The computational inputs for molecular dynamics (MD) simulation were prepared using CHARMM-GUI tools (Jo et al., 2008). Specifically, ligands (PAHs) were parametrized by "Ligand Reader & Modeler" with CHARMM General Force Field (CGenFF) (Kim et al., 2017). The crystal structure of CLOCK:BMAL1 (PDB ID: 4F3L) was obtained from RCSB with H_2O molecule removed. Afterward, to facilitate the movements of small molecules, thirty molecules of a ligand were randomly placed in the surrounding of a receptor, and the volume fraction (solute) was set to 8 %. The solution system was a cubic box with a side length of about 90 nm, filled and neutralized with solvent molecules 20,000 TIP3P molecules and 0.15 % of K^+/Cl^- ions. This procedure was repeated for each of the 16 PAHs.

Simulations were conducted by using GROMACS/2019.3 and CHARMM36m force fields. Van der Waal (VDW) cut-off was set to 1.2 nm. Particle-Mesh-Ewald (PME) was employed for electrostatic interactions and the coulomb cut-off set at 1.2 nm with the Verlet method. Steep energy minimization was converged to 100 $kJ \bullet mol^{-1} \bullet nm^{-1}$ with hydrogen bonds constrained by LINCS. Temperature was coupled separately for solute and solvent using a Nose-Hoover extended ensemble, followed by 400-picoseconds (ps) NVT equilibration in which the reference temperature was 310 K. Parrinello-Rahman method and isotropic coupling type was chosen for 2-ns NPT runs, and the reference

pressure was 1.0 bar. Finally, the 20-nanoseconds (ns) production run was performed, using leap-frog algorithm (md). A 20 ns simulation time is ideal for thoroughly capturing interactions and conformational changes, supported by ligand stability, dynamic movements, and established molecular dynamics precedent (Jogalekar et al., 2010). Integration time steps were set to 0.002 ps and the energy was saved every 100 ps (total of 200 frames) for data analysis. The last frame of each simulation was used as the principal structure for further studies of ligand-receptor binding.

2.4. Aryl hydrocarbon receptor homology modeling

Three-dimensional (3D) structure of the ligand binding domain (LBD) of the aryl hydrocarbon receptor (AhR) was built from sequences of amino acids in the rat AhR (GI: 7304873 in the NCBI sequence database) by validation, using the SWISS-MODEL (Waterhouse et al., 2018), Rampage web servers, and ProSA-web (Lear and Cobb, 2016; Wiederstein and Sippl, 2007a, 2007b). A crystal structure of 4F3L (sequence identity = 27.8 %, residues 252–357) was selected as a template. The structure model, built using the Ramachandran plot from the Rampage web server, was validated by the fact that 98 % of total residues were in the favored region, with 2.4 % of the residues in the allowed region. Further, results from the ProSA analysis showed the z-score for the 3D model within the range of scores typically found for native proteins of similar size.

2.5. H4IIE-luc bioassay and calculation in median effective concentration (EC_{50})

Empirically derived acute toxic potencies of the 16 PAHs were determined based on an AhR-mediated activity using H4IIE-*luc* cells. H4IIE-*luc* cells, derived from rat hepatoma, are highly sensitive to xenobiotics and commonly used to assess PAH toxicity *via* the AhR pathway. They are widely utilized in environmental toxicology for evaluating dioxin-like activity with reliable and reproducible results. These cells express luciferase upon AhR activation, allowing precise measurement of PAH-induced cytotoxicity (Behnisch et al., 2001; Whyte et al., 2004). Considering degradability of the selected PAHs and their metabolic activity during longer durations of exposure, the 4 h exposure time was chosen for AhR-mediated potency. After exposure, the activity was expressed as relative luminescence units that were quantified using a Victor X3 multi-label plate reader (PerkinElmer, Waltham, MA). We converted responses of the H4IIE-*luc* bioassay to percentages of the maximum response of BkF (%BkF_{max}) observed for a 50 nM BkF (=100%BkF_{max}), as BkF was chosen as the reference due to its higher potency in activating the aryl hydrocarbon receptor (AhR) compared to other PAHs in our assay system. Bioassays were repeated four times in triplicate. A half-effective concentration (EC_{50}) for AhR-mediated effects of individual compounds was determined using dose-response curves fitted with the Hill equation. Compounds with ten different concentrations, using five-fold serial dilution (*viz.*, 1, 0.2, 0.04, 0.008, 0.0016, and 0.0032 μg mL^{-1}), were prepared. EC50 values were estimated from the dose-response relationship, assuming equal efficacy and parallelism with the reference compound (BkF).

2.6. In silico prediction of toxic potency

VEGA-QSAR and OpenVirtualtoxLab™ program packages were used for QSAR and molecular docking analysis, respectively (https://www.vegahub.eu). VEGA-QSAR has been performed and shown, based on a large set of toxicological data and endpoints, to be sufficient to derive predictive toxicity. EC_{50} values were obtained using SMILES notation of VEGA-QSAR and obtained toxicity data corresponding to *Daphnia magna* (Benfenati et al., 2013). Ligand:AhR binding affinities were estimated and calculated by OpenVirtualtoxLab™ (Vedani et al., 2015). Combined automated and flexible docking with multidimensional QSAR was used to simulate and quantify toxic potential and how chemicals bind to a set of currently implemented proteins that cause adverse effects.

3. Results and discussion

3.1. A kinetic factor driving ligand to receptor calculated from first principle density functional theory

DFT is a quantum mechanical modeling method used to investigate the electronic structure of atoms, molecules, and solids based on fundamental physical laws, without relying on empirical parameters. After being introduced into cells, ligands in the cytosol interact with various force fields, including electrostatic, hydrogen bonding, and van der Waals forces, which can be accurately modeled using DFT to predict their behavior and binding affinities within biological systems. Because receptor proteins are substantially larger than typical ligands in size (16 PAHs selected for this study), it is reasonable to assume that charges on receptors are fixed in space and then the ligand molecules would undergo rotation to align their dipole moment vector with the line of force action, while they approach the receptor (Fig. 2A). This directional configuration, which is due to the dipole moment vector ($\vec{D}$) of ligand, would determine the relative orientation of preferred sites in a ligand for nucleophilic or electrophilic interaction with charged amino acids in a receptor (atom-condensed Fukui function, *F*) and how strongly a ligand is attracted by charged amino acids forming the receptor protein. The strength of ligand-dependent attraction would also increase the local concentration of the ligand molecules in the vicinity of a receptor.

Based on this consideration, the DRF describing the directional configuration for each PAH was derived for a given AhR protein (Eq. (1), see Fig. 2B, S1, and Tables S2–S4) (Kim et al., 2020).

$$\mathrm{DRF} = C_{LR}Q_{R,net}\left[\sum_{n=1}^{\#A:X} F_n^{\pm}\left(\vec{\boldsymbol{D}}\bullet\hat{\boldsymbol{n}}\right)\right] = C_{LR}Q_{R,net}\left[\sum_{n=1}^{\#A:X} F_n^{\pm}\left|\vec{\boldsymbol{D}}\right|cos\phi_n\right] \quad (1)$$

where $Q_{R,net}$ denotes the net charge of AhR from charged amino acids in AhR (Fig. S2), F^+ (F^-) is the nucleophilic (electrophilic) Fukui function for each terminal carbon, bonded with hydrogen (*#A:X*) in PAHs, $\vec{\boldsymbol{D}}$ is the dipole moment vector of PAHs (Fig. S3), $\hat{\boldsymbol{n}}$ is the unit vector along the bonding direction of each terminal carbon, and C_{LR} is a ligand-receptor electric correlation coefficient of $1/4\pi\varepsilon_o\varepsilon_R r$ representing the Coulomb's constant with relative permittivity of receptor cytosol (ε_R) and the inverse proportionality to the ligand-receptor distance (*r*). The physico-chemical parameters (Eq. 1) were calculated by *ab initio* density functional theory (DFT).

The DRF equation takes into account the contribution from all atom-condensed Fukui functions along the line of action ($\hat{\boldsymbol{D}}$). This reflects that a reactive site located behind the ligand molecule with respect to the line of action would lower the reaction affinity with the receptor. Since the DRF determines the degree of the optimal reaction configuration while molecules move toward the receptor, we used the DRFs to evaluate how they are correlated with binding of the 16 PAHs to AhR protein in the ligand "approach" stage and eventually the AhR-mediated toxic potency (Fig. 2C). As a novel finding, carbon with the largest Fukui value in each PAH would be a preferred site when bound to the receptor. The carbon atom with the largest Fukui value is the most reactive site, either nucleophilic or electrophilic, and is thus more likely to interact with the receptor. This makes it the preferred binding site for PAHs, providing a theoretical basis for predicting receptor-ligand interactions. Therefore, the forward location of carbon with the largest Fukui value for the receptor would be the desirable orientation during the ligand motion toward the receptor in a cell. In addition, because DRF depends on the dipole moment of the ligand, which drives the ligand molecule to move toward the net charge of the AhR protein, DRF is expected to increase the effective concentration of the ligand molecule in the vicinity of the receptor. This means that under the same exposure conditions, a ligand

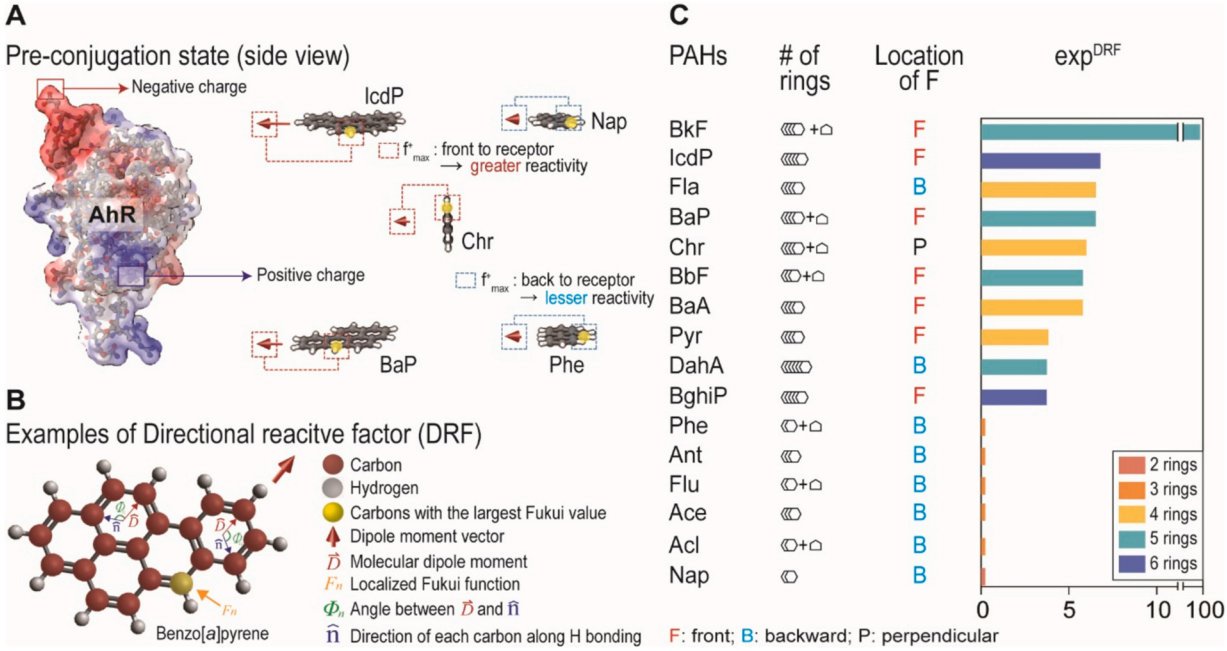


Fig. 2. DRF model for predicting a driving force of the 16 PAHs to AhR. (A) Visualization of first principles prediction model for AhR-mediated potency of benzo[*a*]pyrene before their binding to AhR (B) Various factors for calculating the directional reactive factor (DRF). Carbon atoms with the largest Fukui value in each of the 16 PAHs are highlighted in yellow, and the red arrow denotes the direction and magnitude of dipole moment; (C) Predicted potential toxicity of the 16 PAHs. Each color is the number of benzene rings for each of the 16 PAHs. F (forward), B (backward), and P (perpendicular) are indicated according to the position of the highest Fukui value in the carbon.

compound with a larger DRF would have a greater probability of binding (greater bioactivity) than one with a smaller DRF. By fitting an exponential model for dose-response, exponentiated DRF values (exp^{DRF}) were used for the hazard assessment of the 16 PAHs. Overall, the exp^{DRF} values increased with the number of rings. The PAHs with 2 to 3 cyclic rings, such as phenanthrene, Phe; antracene, Ant; fluorene, Flu; acenaphthene, Ace; acenaphthylene, Acl; and naphthalene, Nap, exhibited the least exp^{DR} values (Fig. 2C). Changes in exp^{DRF} values were generally consistent with other potential toxicities of chrysene homologues in the previous study (Kim et al., 2020). PAHs with 4–6 rings have carbons with the greatest Fukui values, located in front of the receptor and exhibited greater exp^{DRF} values. However, the exp^{DRF} value for fluoranthene (Fla) was determined to be large, although Fla. is known to show low potency. This suggests that the DRF is a necessary indicator of the binding to a receptor based on the kinetic process in the ligand approach stage but not a sufficient one requiring another ligand-specific factor in the following receptor binding stage.

3.2. PAH-AhR binding rate and determination of potential toxicity

As mentioned previously, the DRF that accounts for the physicochemical properties of ligands represents the factor driving a ligand toward the receptor protein and its optimal configuration for enhancing the binding affinity to a receptor. However, it is important to examine the state of the actual kinetic interaction between a ligand and a receptor, because the DRF model does not necessarily guarantee the formation of a ligand-AhR complex. The second stage in the molecular kinetic process for ligand-receptor binding was investigated using a MD simulation method that utilizes the final binding state of PAHs-AhR. Based on the MD results, the positions of the AhR protein were divided into a total of 18 sites where ligand molecules mainly bind to AhR (Fig. S4). Snapshots of MD binding simulation for 30 pyrene (Pyr) molecules in the vicinity of AhR are shown as an example in Fig. 3A. A Per-Arnt-Sim (PAS) domain comprises two structural repeats, which involve dimerization with Arnt (PAS A: Sites *I*, *J*, *K*, *L*, *M*, *N*, *P*, and *R*) and allow the binding of the ligands (PAS B: Sites *A*, *B*, *C*, *D*, *E*, *F*, *G*, *H*, O, and *Q*). In this study, we considered only the binding sites in PAS B domain because they are known to be responsible for enzyme regulation which could increase the potential toxicity (Freeman et al., 2019; Urban et al., 2018). Results of MD simulation suggest that the rate of PAH-AhR binding in PAS B domain is also ligand-specific (Fig. 3B) and should be taken into account in the assessment of the potential toxicity through a solvent-mediated mechanism. Another bioactive contribution parameter, termed the directional binding factor (DBF), in the simulation of molecular-scale ligand-binding kinetics processes within time scales accessible by standard MDs (~20 ns) was introduced. DBF was defined as the ratio of the number of PAH molecules binding to the sites of the PAS B domain $[n(\text{PAS B})_{binding}]$ to the total number of PAH molecules ultimately binding to the receptor among 30 simulated molecules $[n(\text{PAS A} + \text{PAS B})_{binding}]$ (Eq. (2)).

$$DBF = \frac{n(\text{PAS B})_{binding}}{n(\text{PAS B} + \text{PAS A})_{binding}} \quad (2)$$

Binding sites to AhR were different for each PAH compound (Fig. S5 and S6). Dibenzo[*a*,*h*]anthracene (DahA) exhibited the greatest binding rate, with 27 out of total 30 simulated molecules bound to either PAS A or PAS B domain (Fig. 3B). For BaP, which is known to be particularly toxic, 13 out of 30 ligands (43 %) were bound to the *A* site of PAS B domain. Among the 16 PAHs, Nap, Ace, and Flu with 2 to 3 benzene rings exhibited low binding rates (particularly in PAS B), compared to other PAHs with greater molecular weight and with 4 to 6 rings.

The DRF, representing the ligand's approach and interaction potential with the receptor, and the DBF, which quantifies the strength and stability of the ligand-receptor binding, were combined into a comprehensive toxicity indicator known as the directional reactive and binding factor (DRBF). This integrated factor captures both the reactivity of the ligand during the initial approach and its binding affinity in the final ligand-receptor complex, providing a holistic measure of the compound's overall toxicological potential. The DRBF not only accounts for the interaction dynamics between the ligand and receptor during the binding process but also describes the entire kinetic process leading up

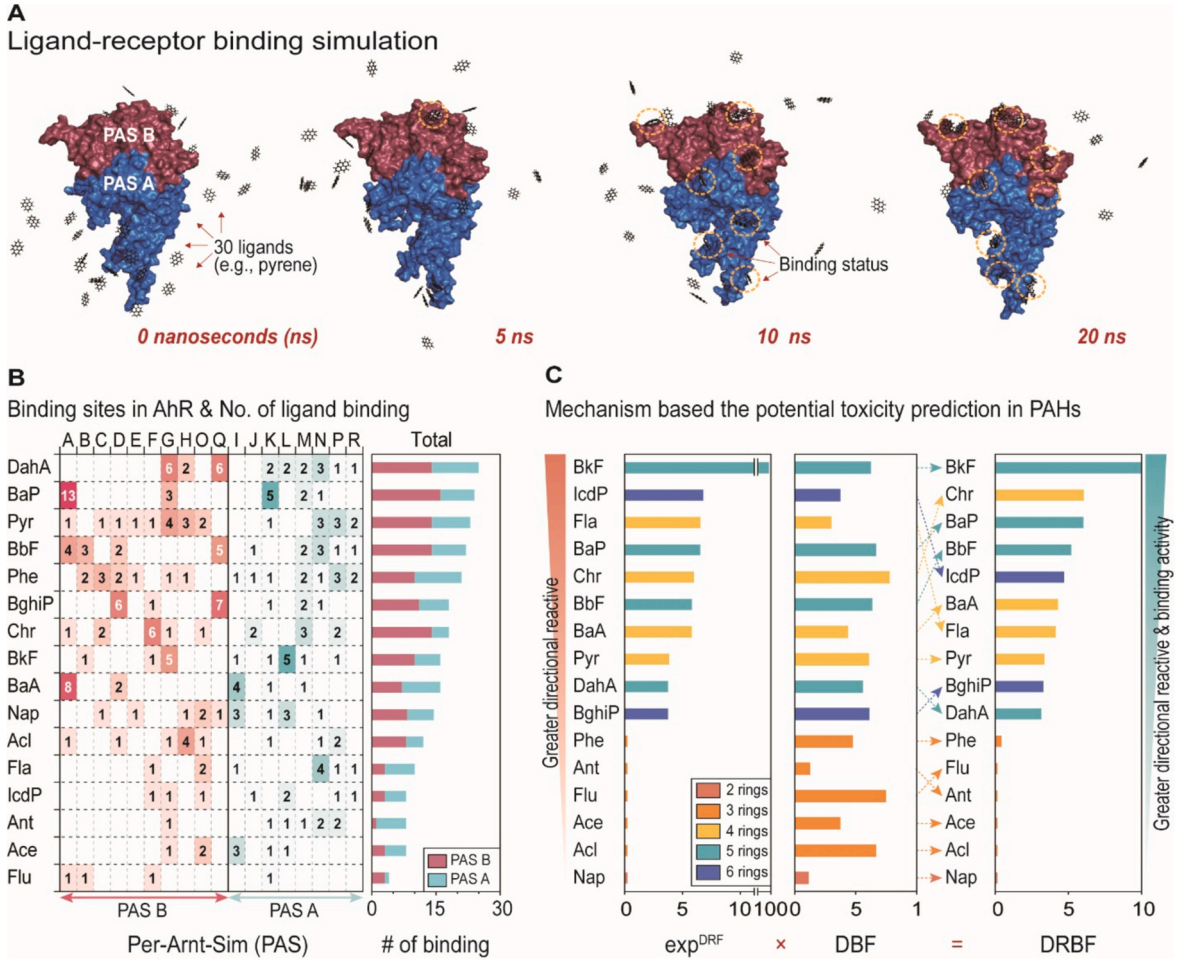


Fig. 3. Molecular dynamics simulation describing ligand-receptor binding, PAH-AhR binding rate, and predicted toxicity potencies of the 16 PAHs by DRBF. (A) Molecular dynamics of the pyrene-AhR for 20 ns. Per-Arnt-Sim (PAS) A and PAS B domains consist of AhR, and PAS B contributes to carcinogenesis after the ligands binding; (B) The number of ligand binding to the 18 binding sites (PAS B: A, B, C, D, E, F, G, H, I, J; PAS A: K, L, M, O, P, Q, and R) of each of the 16 PAHs attached to AhR through MD simulation. $P_{\mathrm{PAS\ B}}$ was calculated by dividing the number of bindings to PAS B by a total of 30 molecules: (C) Predicted toxicity of the 16 PAHs from the DRBF was calculated by multiplying $\exp^{\mathrm{DRF}}$ and DBF in the mechanism of ligand-AhR. In the ligand approach state, DRF was calculated, and in the ligand binding state, directional reactive and binding factor (DBF) was quantified.

to the formation of the ligand-receptor complex. Importantly, this indicator reflects the stages prior to the translocation of the complex into the cell nucleus, where it can regulate the expression of xenobiotic-metabolizing enzymes. By encompassing both the reactive approach phase and the binding stability, the DRBF offers a more complete picture of ligand-receptor interactions, making it a robust tool for predicting toxicity (Eq. 3).

$$DRBF = exp\left(C_{LR}Q_{R,net}\left[\sum_{n=1}^{\#A:X} F_n^{\pm}\left|\vec{D}\right|cos\phi_n\right]\right) \times \frac{n(\mathrm{PAS\ B})_{binding}}{n(\mathrm{PAS\ B} + \mathrm{PAS\ A})_{binding}} \quad (3)$$

Consideration of additional binding rate in the DRBF resulted in a change in the ranking of predicted potential toxicity of the 16 PAHs. When comparing $\exp^{\mathrm{DRF}}$ and DBF, in DRBF, indeno[*1,2,3-c,d*]pyrene (IcdP), Fla., and DahA are predicted to become less toxic than in $\exp^{\mathrm{DRF}}$, whereas greater toxicity is predicted for chrysene (Chr), BaP, benzo[*b*]fluoranthene (BbF), and benzo[*g,h,i*]perylene (BghiP) (Fig. 3C). These results indicate that it is important to consider the entire molecular kinetic process toward the ligand-receptor binding and the mechanisms of toxic expression to improve the accuracy of potential toxicity prediction. Proteins often bind with ligands in buried cavities that appear inaccessible based on the static structures. The mechanisms and pathways by which ligands reach their binding sites in such cases are often unknown (Denison and Faber, 2017; Dror et al., 2011; Zhou et al., 2010). The ligand recognition by occluded cavities can happen rapidly, indicating there is no single dominant pathway. Therefore, this approach describing ligand-receptor binding would consider all the comprehensive bindings which can make a carcinogenesis.

3.3. Experimental and predicted toxic potencies of the 16 PAHs using in vitro and in silico results

Since the DRBF-based prediction of toxic potencies includes the molecular-scale binding reaction between the 16 PAHs and AhR protein, the reliability and predictability were tested by comparison with experimental bioassay results. Empirical AhR-mediated potency for the 16 PAHs using the H4IIE-*luc* cell lines (Table S5) (Fig. 4A) was used to develop dose-response curves that were then categorized for the groups with similar molecular weight and number of benzene rings. AhR-mediated potency is proportional to the number of benzene rings. Compared to a positive control-ligand of BkF with relative potency (ReP) of 1.0, ReP values are <0.01 for the PAHs with 2–3 rings, while greater ReP values are observed in the PAHs with 4–6 rings (< 0.85). AhR-mediated potencies of PAHs varied even for PAHs of the same number of benzene rings and molecular weight (Lee et al., 2015; Pieterse et al., 2013). Construction of ligand-AhR complexes followed from the specific binding sites would be changed by the structure of ligands and their

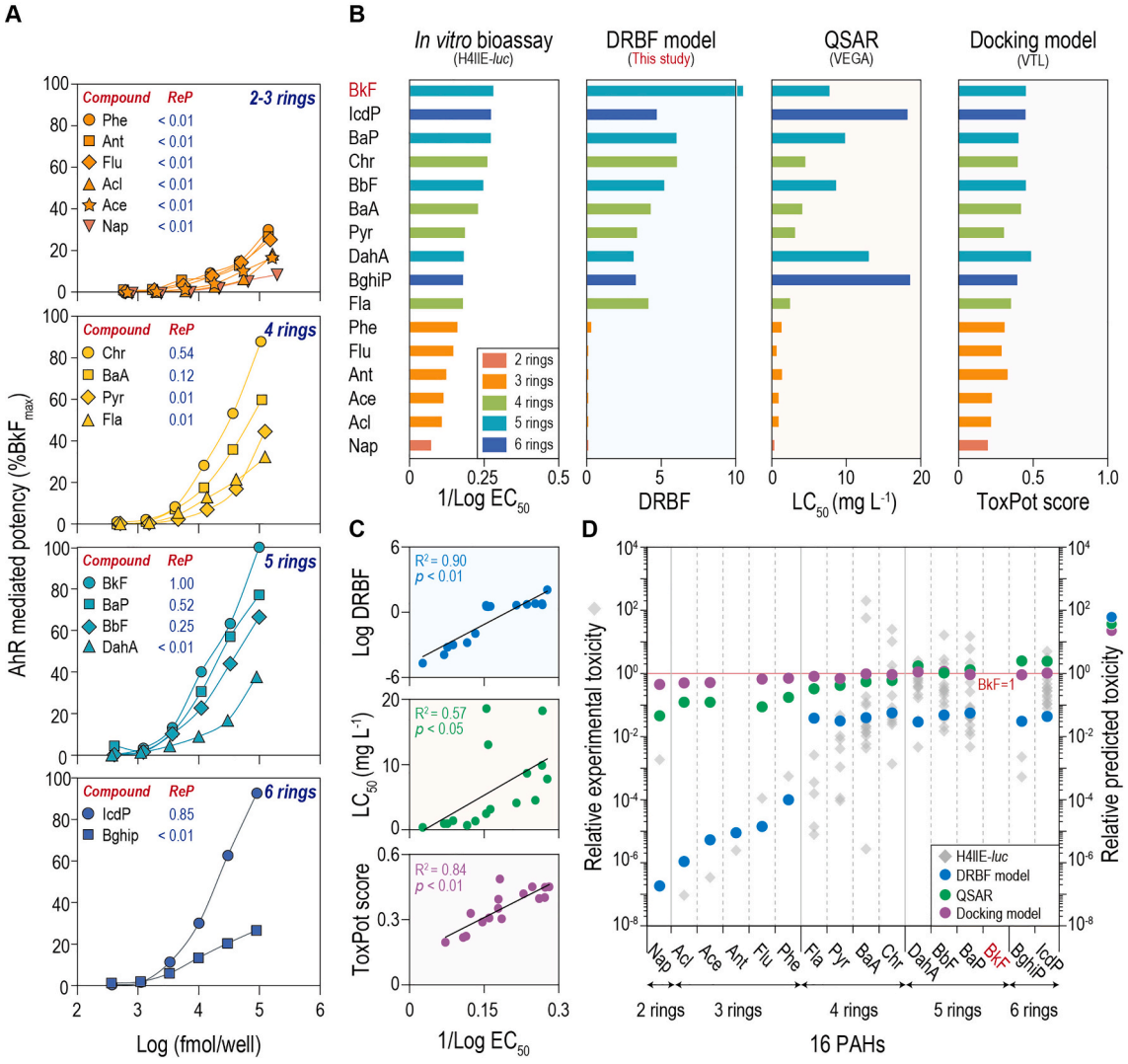


Fig. 4. Performance of the DRBF model and other *in silico* models. (A) AhR-mediated potency (%BkF$_{max}$) of the 16 PAHs determined from *in vitro* bioassay (H4IIE-*luc*); (B) Predicted potential toxicity of the 16 PAHs by the DRBF model (proposed in the current study) and two common *in silico* models (QSAR, molecular docking model). LC_{50} values were obtained from *Daphnia magna* and ToxPot scores were calculated by evaluating the binding affinity between aryl hydrocarbons (AhR) and the 16 PAHs; (C) Observed values (1/Log EC_{50}) *versus* predicted values of potential toxicity from three *in silico* methods and model performance measured by the coefficient of determination (R) with *p-value*. (D) Relative experimental and predicted toxicity of the 16 PAHs by the DRBF model and H4IIE-*luc*, QSAR, and the docking model. Each relative toxicity value was normalized against the positive control (benzo[*k*]fluoranthene; set as 1.0 of potential toxicity).

conformational change in response to the AhR protein (Barron et al., 2004; Giani Tagliabue et al., 2019).

Empirically determined toxic potency data (inverse of log(EC_{50})) are compared with DRBF values, as well as other toxicity-relevant characteristics predicted by other *in silico* models, including 1/LC_{50} from QSAR and toxic potency (ToxPot) score from the molecular docking model (Fig. 4B and Table S6). IcdP and BaP are determined to have greater DRF values than Chr, BaP, and BkF, but lesser DBF values push them up to lower rankings in the DRBF model (Fig. 4B). For PAHs consisting of 2–3 cyclic rings, such as phenanthrene (Phe), Flu, anthracene (Ant), Ace, Acl, and Nap, although additional consideration of the binding rate in the DRBF model changed their rankings of potential toxicity when comparing them from sole consideration of the DRF model, but it does not affect the overall trend. In particular, Chr, BaP, and BbF had DRBF values of 0.059, 0.049, and 0.046, respectively, which were predicted to have greater potential toxic potency than those from DRF alone. Comprehensive DRBF values, which combine physico-chemical property-based DRF and binding rates-based DBF resulted in the best performance ($R^2 = 0.90$ and $p < 0.01$) (Fig. 4C) of the model.

The overall results suggest that the DRBF model outperforms pre-existing *in silico* models of QSAR and molecular docking in predicting the toxic potencies when evaluating them with experimentally obtained values by *in vitro* bioassays. The QSAR model applied in this study predicts potential toxicity based on the number of benzene rings in PAHs. For example, the model predicts the same toxic potencies for BhiP and IcdP (LC_{50}: 0.05 mg L^{-1}) (Fig. 4B). QSAR prediction of 1/LC_{50} resulted in moderate model performance ($R^2 = 0.59$, $p < 0.05$) (Fig. 4C), partly because most QSAR approaches solely depend on the relationship between structure-dependent physico-chemical properties of substances and toxicity, while the mechanistic approach such as the DRBF model consider factors actually participating in the toxic responses. Thus, it is expected for QSAR to provide over- or under-prediction of the potential toxicity from new substances that are not previously included when training the QSAR model (Gramatica et al., 2012; Kolmar and Grulke, 2021).

A docking model predicts greater potential toxicities (ToxPot scores) for PAHs with five rings (Fig. 4B) and linear regression results indicate that the docking model accounts for 84 % of the variability between experimental and predictive values ($R^2 = 0.84$, $p < 0.05$, Fig. 4C), suggesting that the structure and physico-chemical properties of

compounds are more related to the potential toxicity than the number of benzene rings (Fig. 4B and C). Various factors affect the binding status linked to receptor activation, such as the binding sites of the ligand, the hydrophobic sites of the receptor, and the arrangement of amino acids. The long-range dipole-charge interaction sets the orientation and relative distribution of the reactive sites of ligand prior to binding to the receptor. Thus, it needs to be noted that the optimum ligand-receptor binding posture with the lowest free energy might not always be achieved in the real environment. For pre-conjugation of the ligand-receptor binding, the docking model does not account for the binding rate but the optimal binding status. Thus, the predicted potential toxicity of each ligand may not reflect the whole mechanism process of the ligand-receptor binding process (Dyla et al., 2022; Tuffery and Derreumaux, 2012).

When relative toxic potencies predicted by all three *in silico* models, after normalization to BkF, were compared with relative experimental toxicity results reported in previous studies, the DRBF model could explain more variance (measured by R^2) in experimental results than other two *in silico* models (Fig. 4D and Table 1) (Bols et al., 1999; Bosveld et al., 2002; Fent and Bätscher, 2000; Kim et al., 2019; Larsson et al., 2012; Louiz et al., 2008; Machala et al., 2001; Nisbet and Lagoy, 1992; Pieterse et al., 2013; Villeneuve et al., 2002; Vondráček et al., 2017). Based on the relative toxic potencies of the 16 PAH compounds derived by the DRBF model, the contribution of each compound to total toxicity in environmental samples, such as sediment and water, can also be calculated by a potency balance. In previous studies, using H4IIE-*luc* bioassays like this study, it was confirmed that the mixture of PAHs exhibited AhR-mediated potencies additively (Cha et al., 2019; Lee et al., 2017).

In vitro bioassays, such as those using H4IIE-*luc* cells, provide detailed and biologically relevant data by directly measuring cellular responses to chemical exposure. However, they can be resource-intensive and subject to variability in experimental conditions. In contrast, the DRBF model, a computational approach, predicts toxicity based on chemical structure and receptor interactions, enabling rapid, high-throughput screening of numerous chemicals. If the toxicity mechanisms occurring within cells are further elucidated in the future, it will be possible to predict toxicity more accurately through detailed and specific simulations (*e.g.*, concentration) using the model and offer a more comprehensive view of toxic effects.

Table 1
Comparison of toxic potencies predicted by use of the directional reactivity (DR) model with experimental potential toxicity data.

Studied compounds	*In vitro* bioassay (H4IIE-*luc*)		Mechanistic quantification model (This study)			
	Log(1/EC_{50}) (μM)	ReP[a]	DRF[b]	R_{DRF}	DRBF[c]	R_{DRBF}
Naphthalene	0.06	6.65E-14	−2.52	3.24. E-07	1.99. E-05	1.77. E-07
Acenaphthene	0.09	3.40E-07	−1.79	6.60. E-06	5.76. E-04	5.13. E-06
Acenaphthylene	0.09	9.28E-08	−2.24	1.02. E-06	1.18. E-04	1.05. E-06
Phenanthrene	0.14	5.50E-04	−1.10	1.10. E-04	1.07. E-02	9.54. E-05
Fluorene	0.12	1.11E-04	−1.64	1.23. E-05	1.53. E-03	1.36. E-05
Anthracene	0.10	2.45E-06	−1.58	1.55. E-05	9.65. E-04	8.60. E-06
Fluoranthene	0.15	2.59E-03	0.41	5.39. E-02	4.30. E+00	3.83. E-02
Pyrene	0.16	8.08E-03	0.27	3.04. E-02	3.37. E+00	3.00. E-02
Benzo[*a*] anthracene	0.20	1.18E-01	0.37	4.59. E-02	4.29. E+00	3.82. E-02
Chrysene	0.23	5.44E-01	0.40	5.18. E-02	6.61. E+00	5.88. E-02
Benzo[*a*]pyrene	0.23	5.22E-01	0.38	4.75. E-02	5.53. E+00	4.92. E-02
Benzo[*b*] fluoranthene	0.21	2.51E-01	0.37	4.59. E-02	5.20. E+00	4.64. E-02
Benzo [*k*] fluoranthene	**0.24**	**1.00E+00**	**1.12**	**1.00. E+00**	**1.12. E+02**	**1.00. E+00**
Dibenzo[*a,h*] anthracene	0.15	4.61E-03	0.27	2.97. E-02	3.14. E+00	2.80. E-02
Indeno[*1,2,3-cd*] pyrene	0.24	8.52E-01	0.41	5.38. E-02	4.70. E+00	4.19. E-02
Benzo[*g,h,i*] perlyene	0.15	2.27E-03	0.26	2.97. E-02	3.29. E+00	2.93. E-02

[a] Relative potency.
[b] Directional reactive factor.
[c] Directional reactive and binding factor.

3.4. Environmental implications

Importance of mechanistic quantification of dynamic reactions between ligands and receptors can be influenced by water molecules mediating interactions. Water molecules can have an effect on different binding modes within a receptor, ultimately dictated by the free energy of binding, but with several different contributions. In such a complex reaction, relying solely on data-driven approaches rather than integrating them with a mechanism-based approach can increase uncertainty in model predictions. Moving away from the sole reliance on traditional approaches and information sources used in hazard, exposure, and risk assessment, toward the more comprehensive use of rapidly acquired chemical information *via in vitro*, *in silico*, and targeted testing strategies will require careful consideration of the information needed and values associated with a particular decision. The data collection should be obtained after careful consideration of approaches and binding of the toxic expression mechanism.

Our study showed that a mechanistic approach based on the dynamical relationship between ligands and receptors can predict the potential toxicity of the 16 PAHs with high predictive power ($R^2 = 0.9$). This case study demonstrates two factors that help build confidence in emerging methods: i) an approach independent of inferences from statistics and ii) the ability to construct the potential toxicity upon a mechanism-based AOP. In the AOP model, molecular initiation is often mediated as a ligand-receptor protein binding and is one of the fundamental biological processes initiating toxic responses, drug efficacy, immunization, and other crucial cell functions occurring in biological organisms (Helm et al., 1991; van der Velden et al., 2020). A receptor-mediated enzyme regulation triggered by a ligand-receptor binding is ligand-specific and thus provides the base of *in vitro* bioassays in environmental toxicology and drug development for hazard assessment of target substances. If a phenomenon of xenobiotic receptor activation is therefore influenced by the entire kinetic process of ligand molecules from introduction into a cell until binding with a receptor protein, both the ligand approach stage and ligand binding stage must be considered in estimating the toxic potency of a chemical, expressed as a probability of binding.

Predictions made by the model described here are consistent with experimental results and do not require any statistical assumptions but reflect the ligand-receptor binding process quantitatively. This first-principles methodology can serve as an initial framework for analyzing the ligand-binding mechanism in the structure and functions of the AhR in a high-throughput manner to improve understanding of the variation in potential toxicities and their causes in the process of ligand-receptor response for a large number of chemicals. And, the DRBF model has the potential to predict toxicity based on the chemical structure and the receptor involved in the interaction, thus making it applicable to a broad spectrum of chemical compounds in the future. Moreover, the principles employed, particularly quantification of pre-conjugation and binding state, have an importance beyond the immediate field of toxicity screening, for example in studies of drug efficacy and monitoring of human subjects in clinical trials. Therefore, further

improvement of the DRBF model could advance molecular biology and computational structural biology and be a steppingstone for providing reliable and mechanistic interpretation screening systems for preclinical toxicology.

4. Conclusion

This study introduces and applies a new comprehensive model, DRBF, to predict the cytotoxicity of 16PAHs. The study analyzes how PAHs induce toxicity by interacting with the AhR in cells. By combining two parameters, DRF, which represents the ligand's approach potential, and DBF, which quantifies the strength and stability of ligand-receptor binding, DRBF provides a holistic description of the dynamic process of ligand-receptor interactions. This indicator captures both the ligand's reactivity during its approach and the stability of the binding, encompassing the entire kinetic process before the ligand-receptor complex translocated to the cell nucleus to regulate the expression of xenobiotic-metabolizing enzymes. The study demonstrates the utility of the DRBF in accurately predicting PAH toxicity and suggests its potential application in future toxicity assessments and environmental risk studies.

CRediT authorship contribution statement

Taewoo Kim: Writing – original draft, Investigation, Formal analysis, Data curation, Conceptualization. **Juyuan Zhen:** Investigation, Formal analysis, Data curation. **Junghyun Lee:** Investigation, Formal analysis, Conceptualization. **Shin Yeong Park:** Investigation, Formal analysis. **Changkeun Lee:** Formal analysis, Data curation. **Bong-Oh Kwon:** Project administration, Investigation. **Seongjin Hong:** Visualization, Formal analysis, Conceptualization. **Hyeong-Moo Shin:** Writing – review & editing, Formal analysis. **John P. Giesy:** Writing – review & editing, Conceptualization. **Gap Soo Chang:** Writing – review & editing, Project administration, Data curation, Conceptualization. **Jong Seong Khim:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by Korea Institute of Marine Science & Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries, Korea (20220534 and 202300239887), and National Research Foundation of Korea (NRF-2022R1A2C1092682 and RS-2023-00249256).

Appendix A. Supplementary data

Chemical compounds information of 16 PAHs used in this study (Table S1); structural and electronic configuration parameters of 16 PAHs and the potential toxicity predict-ed by directional reactivity factor (DRF) (Table S2); searching for structure homologues of aryl hydrocarbon receptors (Table S3); physico-chemical properties of aryl hydrocarbon receptors (Table S4); *in vitro* transactivation bioassay conditions for evaluating toxicities of chemical compounds. (Table S5); the estimated potential toxicity for 16 PAHs from various *in silico* models (Table S6); visualization of the HOMO and LUMO orbitals in 16 PAHs (Fig. S1); homology model and quality metrics of AhR LBD (Fig. S2); the information of physico-chemical properties of 16 PAHs (Fig. S3); AhR homology model and 16 PAHs binding sites (Fig. S4); Molecular dynamics simulations of the binding of 16 PAHs (2–6 rings) to AhR in water solution (Fig. S5 and S6) (PDF).

Data availability

Data will be made available on request.

References

Andersson, J.T., Achten, C., 2015. Time to say goodbye to the 16 EPA PAHs? Toward an up-to-date use of PACs for environmental purposes. Polycycl. Aromat. Compd. 35, 330–354. https://doi.org/10.1080/10406638.2014.991042.

AnvariFar, H., Amirkolaie, A.K., Jalali, A.M., Miandare, H., Sayed, A.H., Üçüncü, S.İ., Ouraji, H., Ceci, M., Romano, N., 2018. Environmental pollution and toxic substances: cellular apoptosis as a key parameter in a sensible model like fish. Aquat. Toxicol. 204, 144–159. https://doi.org/10.1016/j.aquatox.2018.09.010.

Barron, M.G., Heintz, R., Rice, S.D., 2004. Relative potency of PAHs and heterocycles as aryl hydrocarbon receptor agonists in fish. Mar. Environ. Res. 58, 95–100. https://doi.org/10.1016/j.marenvres.2004.03.001.

Behnisch, P.A., Hosoe, K., Sakai, S., 2001. Brominated dioxin-like compounds: in vitro assessment in H4IIE-luciferase reporter gene assay. Environ. Health Perspect. 109 (4), 369–376. https://doi.org/10.1016/S0160-4120(03)00105-3.

Benfenati, E., Manganaro, A., Gini, G.C., 2013. VEGA-QSAR: AI inside a platform for predictive toxicology. PAI@ AI* IA 1107, 21–28. http://ceur-ws.org/Vol-1107/paper8.pdf.

Blanchard, M., Teil, M.-J., Guigon, E., Larcher-Tiphagne, K., Ollivon, D., Garban, B., Chevreuil, M., 2007. Persistent toxic substance inputs to the river seine basin (France) via atmospheric deposition and urban sludge application. Sci. Total Environ. 375, 232–243. https://doi.org/10.1016/j.scitotenv.2006.12.012.

Bols, N., Schirmer, K., Joyce, E., Dixon, D., Greenberg, B., Whyte, J., 1999. Ability of polycyclic aromatic hydrocarbons to induce 7-ethoxyresorufin-o-deethylase activity in a trout liver cell line. Ecotox. Environ. Safe. 44, 118–128. https://doi.org/10.1006/eesa.1999.1808.

Bosveld, A.T., De Bie, P.A., Van den Brink, N.W., Jongepier, H., Klomp, A.V., 2002. In vitro EROD induction equivalency factors for the 10 PAHs generally monitored in risk assessment studies in the Netherlands. Chemosphere 49, 75–83. https://doi.org/10.1016/S0045-6535(02)00161-3.

Cha, J., Hong, S., Kim, J., Lee, J., Yoon, S.J., Lee, S., Moon, H.-B., Shin, K.-H., Hur, J., Giesy, J.P., Khim, J.S., 2019. Major AhR-active chemicals in sediments of Lake Sihwa, South Korea: application of effect-directed analysis combined with full-scan screening analysis. Environ. Int. 133, 105199. https://doi.org/10.1016/j.envint.2019.105199.

Chen, Y., Ung, C., 2001. Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand–protein inverse docking approach. J. Mol. Graph. Model. 20, 199–218. https://doi.org/10.1016/S1093-3263(01)00109-7.

Cherkasov, A., Muratov, E.N., Fourches, D., Varnek, A., Baskin, I.I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y.C., Todeschini, R., Consonni, V., Kuz'min, V.E., Cramer, R., Benigni, R., Yang, C., Rathman, J., Terfloth, L., Gasteiger, J., Richard, A., Tropsha, A., 2014. QSAR modeling: where have you been? Where are you going to? J. Med. Chem. 57, 4977–5010. https://doi.org/10.1021/jm4004285.

Cronin, M.T., Schultz, T.W., 2003. Pitfalls in QSAR. J. Mol. Struct. (THEOCHEM) 622, 39–51. https://doi.org/10.1016/S0166-1280(02)00616-4.

Denison, M.S., Faber, S.C., 2017. And now for something completely different: diversity in ligand-dependent activation of Ah receptor responses. Curr. Opin. Toxicol. 2, 124–131. https://doi.org/10.1016/j.cotox.2017.01.006.

Dror, R.O., Pan, A.C., Arlow, D.H., Borhani, D.W., Maragakis, P., Shan, Y., Xu, H., Shaw, D.E., 2011. Pathway and mechanism of drug binding to G-protein-coupled receptors. Proc. Natl. Acad. Sci. USA 108, 13118–13123. https://doi.org/10.1073/pnas.1104614108.

Dyla, M., González Foutel, N.S., Otzen, D.E., Kjaergaard, M., 2022. The optimal docking strength for reversibly tethered kinases. Proc. Natl. Acad. Sci. USA 119, e2203098119. https://doi.org/10.1073/pnas.2203098119.

Fent, K., Bätscher, R., 2000. Cytochrome P4501A induction potencies of polycyclic aromatic hydrocarbons in a fish hepatoma cell line: demonstration of additive interactions. Environ. Toxicol. Chem. 19, 2047–2058. https://doi.org/10.1002/etc.5620190813.

Forrest, J., Bazylewski, P., Bauer, R., Hong, S., Kim, C.Y., Giesy, J.P., Khim, J.S., Chang, G.S., 2014. A comprehensive model for chemical bioavailability and toxicity of organic chemicals based on first principles. Front. Mar. Sci. 1, 31.

Freeman, S.L., Kwon, H., Portolano, N., Parkin, G., Venkatraman Girija, U., Basran, J., Fielding, A.J., Fairall, L., Svistunenko, D.A., Moody, P.C.E., Schwabe, J.W.R., Kyriacou, C.P., Raven, E.L., 2019. Heme binding to human CLOCK affects interactions with the E-box. Proc. Natl. Acad. Sci. USA 116, 19911–19916. https://doi.org/10.1073/pnas.1905216116.

Friesner, R.A., Murphy, R.B., Repasky, M.P., Frye, L.L., Greenwood, J.R., Halgren, T.A., Sanschagrin, P.C., Mainz, D.T., 2006. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein−ligand complexes. J. Med. Chem. 49, 6177–6196. https://doi.org/10.1021/jm051256o.

Giani Tagliabue, S., Faber, S.C., Motta, S., Denison, M.S., Bonati, L., 2019. Modeling the binding of diverse ligands within the ah receptor ligand binding domain. Sci. Rep. 9, 1–14. https://doi.org/10.1038/s41598-019-47138-z.

Gramatica, P., Cassani, S., Roy, P.P., Kovarich, S., Yap, C.W., Papa, E., 2012. QSAR modeling is not "push a button and find a correlation": a case study of toxicity of (benzo-) triazoles on algae. Mol. Inform. 31, 817–835. https://doi.org/10.1002/minf.201200075.

Helm, C.A., Knoll, W., Israelachvili, J.N., 1991. Measurement of ligand-receptor interactions. Proc. Natl. Acad. Sci. USA 88, 8169–8173. https://doi.org/10.1073/pnas.88.18.8169.

Jo, S., Kim, T., Iyer, V.G., Im, W., 2008. CHARMM-GUI: a web-based graphical user interface for CHARMM. J. Comput. Chem. 29, 1859–1865. https://doi.org/10.1002/jcc.20945.

Jogalekar, A.S., Reiling, S., Vaz, R.J., 2010. Identification of optimum computational protocols for modeling the aryl hydrocarbon receptor (AHR) and its interaction with ligands. Bioorg. Med. Chem. Lett. 20 (22), 6616–6619. https://doi.org/10.1016/j.bmcl.2010.09.019.

Kavlock, R.J., Ankley, G., Blancato, J., Breen, M., Conolly, R., Dix, D., Houck, K., Hubal, E., Judson, R., Rabinowitz, J., Richard, A., Setzer, R.W., Shah, I., Villeneuve, D., Weber, E., 2008. Computational toxicology—a state of the science mini review. Toxicol. Sci. 103, 14–27. https://doi.org/10.1093/toxsci/kfm297.

Kim, S., Lee, J., Jo, S., Brooks III, C.L., Lee, H.S., Im, W., 2017. CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. J. Comput. Chem. 38, 1879–1886. https://doi.org/10.1002/jcc.24829.

Kim, J., Hong, S., Cha, J., Lee, J., Kim, T., Lee, S., Moon, H.-B., Shin, K.-H., Hur, J., Lee, J.-S., Giesy, J.P., Khim, J.S., 2019. Newly identified AhR-active compounds in the sediments of an industrial area using effect-directed analysis. Environ. Sci. Technol. 53, 10043–10052. https://doi.org/10.1021/acs.est.9b02166.

Kim, T., Zhen, J., Lee, J., Bauer, R., Lee, C., Kwon, B.-O., 2020. Influence of ligand's directional configuration, chrysenes as model compounds on the binding activity with aryl hydrocarbon receptor. Sci. Rep. 10 (1), 13821. https://doi.org/10.1038/s41598-020-70704-9.

Kolmar, S.S., Grulke, C.M., 2021. The effect of noise on the predictive limit of QSAR models. J. Chem. 13, 1–19. https://doi.org/10.1186/s13321-021-00571-7.

Larsson, M., Orbe, D., Engwall, M., 2012. Exposure time–dependent effects on the relative potencies and additivity of PAHs in the Ah receptor-based H4IIE-luc bioassay. Environ. Toxicol. Chem. 31, 1149–1157. https://doi.org/10.1002/etc.1776.

Lear, S., Cobb, S.L., 2016. Pep-Calc. com: a set of web utilities for the calculation of peptide and peptoid properties and automatic mass spectral peak assignment. J. Comput. Aid. Mol. Des. 30, 271–277. https://doi.org/10.1007/s10822-016-9902-7.

Lee, S., Shin, W.-H., Hong, S., Kang, H., Jung, D., Yim, U.H., 2015. Measured and predicted affinities of binding and relative potencies to activate the AhR of PAHs and their alkylated analogues. Chemosphere 139, 23–29. https://doi.org/10.1016/j.chemosphere.2015.05.033.

Lee, J., Hong, S., Yoon, S.J., Kwon, B.-O., Ryu, J., Giesy, J.P., 2017. Long-term changes in distributions of dioxin-like and estrogenic compounds in sediments of Lake Sihwa, Korea: revisited mass balance. Chemosphere 181, 767–777. https://doi.org/10.1016/j.chemosphere.2017.04.074.

Leusch, F.D., De Jager, C., Levi, Y., Lim, R., Puijker, L., Sacher, F., 2010. Comparison of five in vitro bioassays to measure estrogenic activity in environmental waters. Environ. Sci. Technol. 44, 3853–3860. https://doi.org/10.1021/es903899d.

Louiz, I., Kinani, S., Gouze, M.-E., Ben-Attia, M., Menif, D., Bouchonnet, S., 2008. Monitoring of dioxin-like, estrogenic and anti-androgenic activities in sediments of the Bizerta lagoon (Tunisia) by means of in vitro cell-based bioassays: contribution of low concentrations of polynuclear aromatic hydrocarbons (PAHs). Sci. Total Environ. 402, 318–329. https://doi.org/10.1016/j.scitotenv.2008.05.005.

Machala, M., Vondráček, J., Bláha, L., Ciganek, M., Neča, J., 2001. Aryl hydrocarbon receptor-mediated activity of mutagenic polycyclic aromatic hydrocarbons determined using in vitro reporter gene assay. Mutat. Res. Genet. Toxicol. Environ. Mutagen. 497, 49–62. https://doi.org/10.1016/S1383-5718(01)00240-6.

Marzo, M., Lavado, G., Como, F., Toropova, A., Toropov, A., Baderna, D., 2020. QSAR models for biocides: the example of the prediction of *Daphnia magna* acute toxicity. SAR QSAR Environ. Res. 31, 227–243. https://doi.org/10.1080/1062936X.2019.1709221.

Neese, F., 2012. The ORCA program system. Wiley Interdiscip. Rev.-Comput. Mol. Sci. 2, 73–78. https://doi.org/10.1002/wcms.81.

Nisbet, I.C., Lagoy, P.K., 1992. Toxic equivalency factors (TEFs) for polycyclic aromatic hydrocarbons (PAHs). Regul. Toxicol. Pharmacol. 16, 290–300. https://doi.org/10.1016/0273-2300(92)90009-X.

Pieterse, B., Felzel, E., Winter, R., Van Der Burg, B., Brouwer, A., 2013. PAH-CALUX, an optimized bioassay for AhR-mediated hazard identification of polycyclic aromatic hydrocarbons (PAHs) as individual compounds and in complex mixtures. Environ. Sci. Technol. 47, 11651–11659. https://doi.org/10.1021/es403810w.

Politi, R., Rusyn, I., Tropsha, A., 2014. Prediction of binding affinity and efficacy of thyroid hormone receptor ligands using QSAR and structure-based modeling methods. Toxicol. Appl. Pharmacol. 280, 177–189. https://doi.org/10.1016/j.taap.2014.07.009.

Richard, A.M., Yang, C., Judson, R.S., 2008. Toxicity data informatics: supporting a new paradigm for toxicity prediction. Toxicol. Mech. Methods 18, 103–118. https://doi.org/10.1080/15376510701857452.

Shin, H.-M., Ernstoff, A., Arnot, J.A., Wetmore, B.A., Csiszar, S.A., Fantke, P., 2015. Risk-based high-throughput chemical screening and prioritization using exposure models and in vitro bioactivity assays. Environ. Sci. Technol. 49, 6760–6771. https://doi.org/10.1021/acs.est.5b00498.

Tropsha, A., 2010. Best practices for QSAR model development, validation, and exploitation. Mol. Inf. 29, 476–488. https://doi.org/10.1002/minf.201000061.

Tuffery, P., Derreumaux, P., 2012. Flexibility and binding affinity in protein–ligand, protein–protein and multi-component protein interactions: limitations of current computational approaches. J. R. Soc. Interface 9, 20–33. https://doi.org/10.1098/rsif.2011.0584.

Ukelis, U., Kramer, P.-J., Olejniczak, K., Mueller, S.O., 2008. Replacement of in vivo acute oral toxicity studies by in vitro cytotoxicity methods: opportunities, limits and regulatory status. Regul. Toxicol. Pharmacol. 51, 108–118. https://doi.org/10.1016/j.yrtph.2008.02.002.

Urban, P., Lautier, T., Pompon, D., Truan, G., 2018. Ligand access channels in cytochrome P450 enzymes: a review. Regul. Toxicol. Pharmacol. 19, 1617. https://doi.org/10.3390/ijms19061617.

Vedani, A., Dobler, M., Hu, Z., Smieško, M., 2015. OpenVirtualToxLab—a platform for generating and exchanging in silico toxicity data. Toxicol. Lett. 232, 519–532. https://doi.org/10.1016/j.toxlet.2014.09.004.

van der Velden, W.J., Heitman, L.H., Rosenkilde, M.M., 2020. Perspective: implications of ligand–receptor binding kinetics for therapeutic targeting of G protein-coupled receptors. ACS Pharmacol. Transl. Sci. 3, 179–189. https://doi.org/10.1021/acsptsci.0c00012.

Villeneuve, D., Khim, J.S., Kannan, K., Giesy, J.P., 2002. Relative potencies of individual polycyclic aromatic hydrocarbons to induce dioxinlike and estrogenic responses in three cell lines. Environ. Toxicol. 17, 128–137. https://doi.org/10.1002/tox.10041.

Vondráček, J., Pěnčíková, K., Neča, J., Ciganek, M., Grycová, A., Dvořák, Z., Machala, M., 2017. Assessment of the aryl hydrocarbon receptor-mediated activities of polycyclic aromatic hydrocarbons in a human cell-based reporter gene assay. Environ. Pollut. 220, 307–316. https://doi.org/10.1016/j.envpol.2016.09.064.

Wang, C., Wu, S., Zhou, S., Wang, H., Li, B., Chen, H., 2015. Polycyclic aromatic hydrocarbons in soils from urban to rural areas in Nanjing: concentration, source, spatial distribution, and potential human health risk. Sci. Total Environ. 527, 375–383. https://doi.org/10.1016/j.scitotenv.2015.05.025.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46 (W1), W296–W303. https://doi.org/10.1093/nar/gky427.

Wei, B.Q., Baase, W.A., Weaver, L.H., Matthews, B.W., Shoichet, B.K., 2002. A model binding site for testing scoring functions in molecular docking. J. Mol. Biol. 322, 339–355. https://doi.org/10.1016/S0022-2836(02)00777-5.

Whyte, J.J., Jung, R.E., Schmitt, C.J., Tillitt, D.E., 2004. Ethoxyresorufin-O-deethylase (EROD) activity in fish as a biomarker of chemical exposure. Crit. Rev. Toxicol. 34 (1), 1–26. https://doi.org/10.1080/10408440091159239.

Wiederstein, M., Sippl, M.J., 2007a. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 35, 407–410. https://doi.org/10.1093/nar/gkm290.

Wiederstein, M., Sippl, M.J., 2007b. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 35, 407–410. https://doi.org/10.1093/nar/gkm290.

Xu, Y., Shrestha, N., Préat, V., Beloqui, A., 2021. An overview of in vitro, ex vivo and in vivo models for studying the transport of drugs across intestinal barriers. Adv. Drug Deliv. Rev. 175, 113795. https://doi.org/10.1016/j.addr.2021.05.005.

Yan, J., Wang, L., Fu, P.P., Yu, H., 2004. Photomutagenicity of 16 polycyclic aromatic hydrocarbons from the US EPA priority pollutant list. Mutat. Res. Genet. Toxicol. Environ. Mutagen. 557, 99–108. https://doi.org/10.1016/j.mrgentox.2003.10.004.

Yunta, M.J., 2016. Docking and ligand binding affinity: uses and pitfalls. Am. J. Model. Optim. 4, 74–114. http://pubs.sciepub.com/ajmo/4/3/2.

Zacharias, M., 2003. Protein–protein docking with a reduced protein model accounting for side-chain flexibility. Protein Sci. 12, 1271–1282. https://doi.org/10.1110/ps.0239303.

Zhou, H., Wu, H., Liao, C., Diao, X., Zhen, J., Chen, L., Xue, Q., 2010. Toxicology mechanism of the persistent organic pollutants (POPs) in fish through AhR pathway. Toxicol. Mech. Methods 20, 279–286. https://doi.org/10.3109/15376516.2010.485227.

<Submission for Science of The Total Environment>

Supplementary material for

Prediction of Cytotoxicity of Polycyclic Aromatic Hydrocarbons from First principles

Taewoo Kim, Juyuan Zhen, Junghyun Lee, Shin Yeong Park, Changkeun Lee, Bong-Oh Kwon, Seongjin Hong, Hyeong-Moo Shin, John P. Giesy, Gap Soo Chang, Jong Seong Khim

This PDF file includes:

Number of pages: 13

Number of supplementary tables: 6, Table S1 to S6

Number of supplementary figures: 6, Figure S1 to S6

[*]**Corresponding authors:**

E-mail addresses: gapsoo.chang@usask.ca (G.S. Chang); jskocean@snu.ac.kr, (J.S. Khim)

Supplementary Tables

Supplementary Figures

Supplementary Tables

Table S1. Chemical compounds information of 16 polycyclic aromatic hydrocarbons (PAHs) used in this study.

Compounds	[a]Abb.	Molecular formula	Number of benzene rings	Molecular weight	Log Kow	CAS RN	Purity (%)
Naphthalene	Na	C10H8	2	128.17	3.30	91-20-3	> 98
Acenaphthene	Ace	C12H10	3	154.21	3.92	83-32-9	> 99
Acenaphthylene	Acl	C12H8	3	152.19	3.93	208-96-8	> 98
Fluorene	Flu	C13H10	3	166.22	4.18	95270-88-5	> 98
Phenanthrene	Phe	C14H10	3	178.23	4.46	85-01-8	> 98
Anthracene	Ant	C14H10	3	178.23	4.45	120-12-7	> 98
Fluoranthene	Fla	C16H10	4	202.25	5.16	206-44-0	> 98
Pyrene	Py	C16H10	4	202.25	4.88	129-00-0	> 98
Benzo[*a*]anthracene	BaA	C18H12	4	228.3	5.76	56-55-3	> 98
Chrysene	Chr	C18H12	4	228.3	5.73	218-01-9	> 98
Benzo[*b*]fluoranthene	BbF	C20H12	5	252.3	5.78	205-99-2	> 98
Benzo[*k*]fluoranthene	BkF	C20H12	5	252.3	6.11	207-08-9	> 98
Benzo[*a*]pyrene	BaP	C20H12	5	252.3	6.13	50-32-8	> 98
Dibenzo[*a,h*]anthracene	DbahA	C22H14	5	278.3	6.50	53-70-3	> 98
Benzo[*g,h,i*]perylene	BghiP	C22H12	6	276.3	6.63	191-24-2	> 98
Indeno[*1,2,3-c,d*]pyrene	IcdP	C22H12	6	276.3	6.70	193-39-5	> 98

[a] Abb.: Abbreviations

Table S2. Structural and electronic configuration parameters of 16 PAHs and the potential toxicity predicted by directional reactivity factor (DRF).

Compounds	HOMO (eV)	LUMO (eV)	HOMO-LUMO energy gap (eV)	Dipole moment magnitude (10^2 Debye)	Fukui value F^+	Carbon atoms with the largest Fukui value	Location of carbon* to AhR	Directional reactivity factor (DRF)
Naphthalene	-6.0	-1.3	4.78	0.0	0.1	10C	B	-2.52
Acenaphthene	-5.7	-1.0	4.64	0.9	0.1	4C	B	-1.79
Acenaphthylene	-6.0	-1.0	4.96	0.5	0.1	9C	B	-2.24
Phenanthrene	-6.0	-1.3	4.71	0.0	0.1	8C	B	-1.10
Fluorene	-5.5	-1.9	3.56	0.0	0.1	7C	B	-1.64
Anthracene	-6.1	-2.1	3.91	0.3	0.1	2C	B	-1.58
Fluoranthene	-6.0	-2.0	3.98	0.3	0.1	14C	B	0.41
Pyrene	-5.6	-1.7	3.83	0.0	0.1	8C, 10C	F	0.27
Benzo[*a*]anthracene	-5.6	-1.8	3.74	0.0	0.1	9C	F	0.37
Chrysene	-5.8	-1.5	4.24	0.0	0.1	7C, 18C	P	0.40
Benzo[*a*]pyrene	-5.9	-2.0	3.98	0.4	0.1	2C	F	0.38
Benzo[*b*]fluoranthene	-5.6	-2.0	3.64	0.3	0.1	7C	F	0.37
Benzo[*k*]fluoranthene	-5.3	-2.0	3.36	0.0	0.1	7C	F	1.12
Dibenzo[*a*,*h*]anthracene	-5.4	-1.9	3.50	0.0	0.1	7C	B	0.27
Indeno[*1*,*2*,*3-cd*]pyrene	-5.6	-2.2	3.33	0.6	0.1	16C	F	0.41
Benzo[*g*,*h*,*i*]perlyene	-5.6	-1.7	3.86	0.0	0.1	10C	F	0.26

* The carbon with the largest Fukui value; forward (F), perpendicular (P), and backward (B) to the AhR.

Table S3. Searching for structure homologues of aryl hydrocarbon receptors.

Protein data Bank ID	Chain	GMQE	QMEAN	%Identity
4F3L	**B**	**0.67**	**-1.73**	**29.59**
4ZP4.1	B	0.65	-3.97	30.77
4ZP4.2	B	0.67	-2.54	30.77
4ZPH.1	B	0.67	-3.01	30.77
4ZPH.2	B	0.67	-2.54	30.77
4ZQD.2	B	0.67	-2.67	30.77
4ZQD.1	B	0.66	-3.21	30.77`
5KIZ.1	A	0.64	-4.24	31.07
6E3T.1	B	0.66	-2.62	30.77
5SY7.1	B	0.63	-2.22	25.70
4ZPR.1	B	0.59	-6.01	31.07
5SY5.1	B	0.61	-2.73	21.90
5SY5.2	B	0.62	-2.60	21.90
4H6J.1	A	0.60	-3.30	34.41
2A24.1	A	0.60	-3.99	30.77
4WN5.2	B	0.59	-2.66	24.27
4WN5.1	A	0.59	-2.40	24.27
4DJ3.1	A	0.57	-3.08	18.63

Table S4. Physico-chemical properties of aryl hydrocarbon receptors.

Single letter code	NFIFRTKHKLDFTPIGCDAKGQLILGYTEVELCTRGSGYQFIHAADILHCAESHIRMI KTGESGMTVFRLLAKHSRWRWVQSNARLIYRNGRPDYIIATQRPLTDEE
Number of residues	107
Molecular weight	12381.1 g mol^{-1}
Extinction coefficient	16500 M-1 cm^{-1}
Iso-electric point	pH 9.14
Net charge at pH 7.4	4.1
Estimated solubility	Good water solubility
[a] Equation for net charge	$\text{Net charge} = \sum_i N_i \frac{10^{pk_{a_i}}}{10^{pH} + 10^{pk_{a_i}}} - \sum_j N_j \frac{10^{pk_{a_i}}}{10^{pH} + 10^{pk_{a_j}}}$

[a] N_i is the number, and pKa_i are the pKa values, of the N-terminus and the side chains of Arginine, Lysine, and Histidine.

Table S5. In vitro transactivation bioassay conditions for evaluating toxicities of chemical compounds.

Cell line	H4IIE-*luc*
ATCC#	CRL-1548
Cell type	Recombinant (*luc*-gene)
Mode of Action	AhR-mediated potency
Endpoint	Luciferase activity
Positive control	Benzo[*a*]pyrene
Culture condition	37 ℃, 5% CO_2
Exposure time	4 h

Table S6. The estimated potential toxicity for 16 PAHs from various in silico models

Compounds	QSAR *(Daphnia magna)*		Docking model [b)]
	LC_{50} (μM)	ReP	ToxPot
Naphthalene	2.95	0.04	0.20
Acenaphthene	1.09	0.12	0.22
Acenaphthylene	1.09	0.12	0.22
Phenanthrene	0.76	0.17	0.31
Fluorene	1.52	0.08	0.29
Anthracene	0.73	0.18	0.33
Fluoranthene	0.41	0.32	0.35
Pyrene	0.32	0.40	0.30
Benzo[*a*]anthracene	0.22	0.58	0.42
Chrysene	0.24	0.55	0.40
Benzo[*a*]pyrene	0.10	1.27	0.40
Benzo[*b*]fluoranthene	0.12	1.11	0.45
Benzo[*k*]fluoranthene	0.13	1.00	0.45
Dibenzo[*a*,*h*]anthracene	0.08	1.68	0.49
Indeno[*1*,*2*,*3-cd*]pyrene	0.05	2.35	0.45
Benzo[*g*,*h*,*i*]perlyene	0.05	2.39	0.39

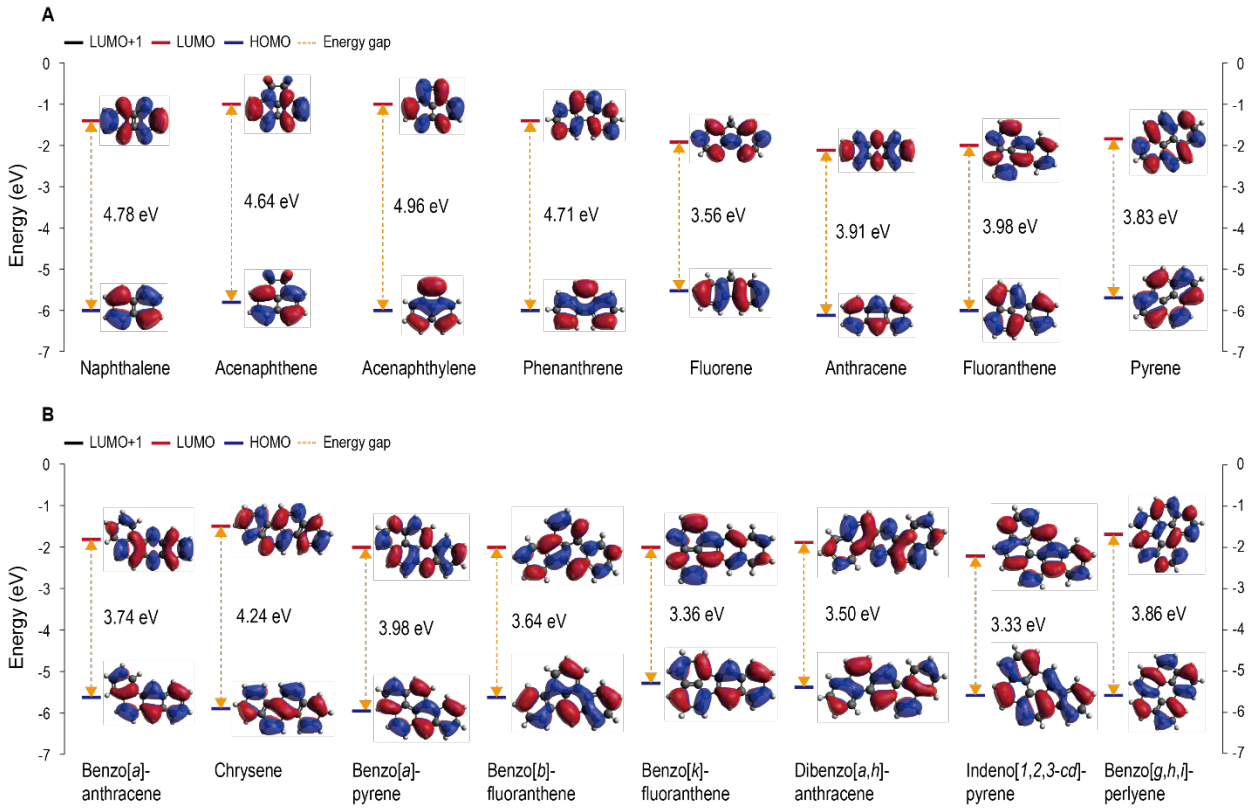


Figure S1. Visualization of the HOMO and LUMO orbitals in 16 PAHs. (A and B) Presentation of the energy levels, HOMO-LUMO gap and orbital composition distribution of the HOMO and LUMO for 16 PAHs. HOMO-LUMO gap energy ranged from 3.33 eV to 4.78 eV. The HOMO-LUMO gap is slightly difference in each PAH compound.

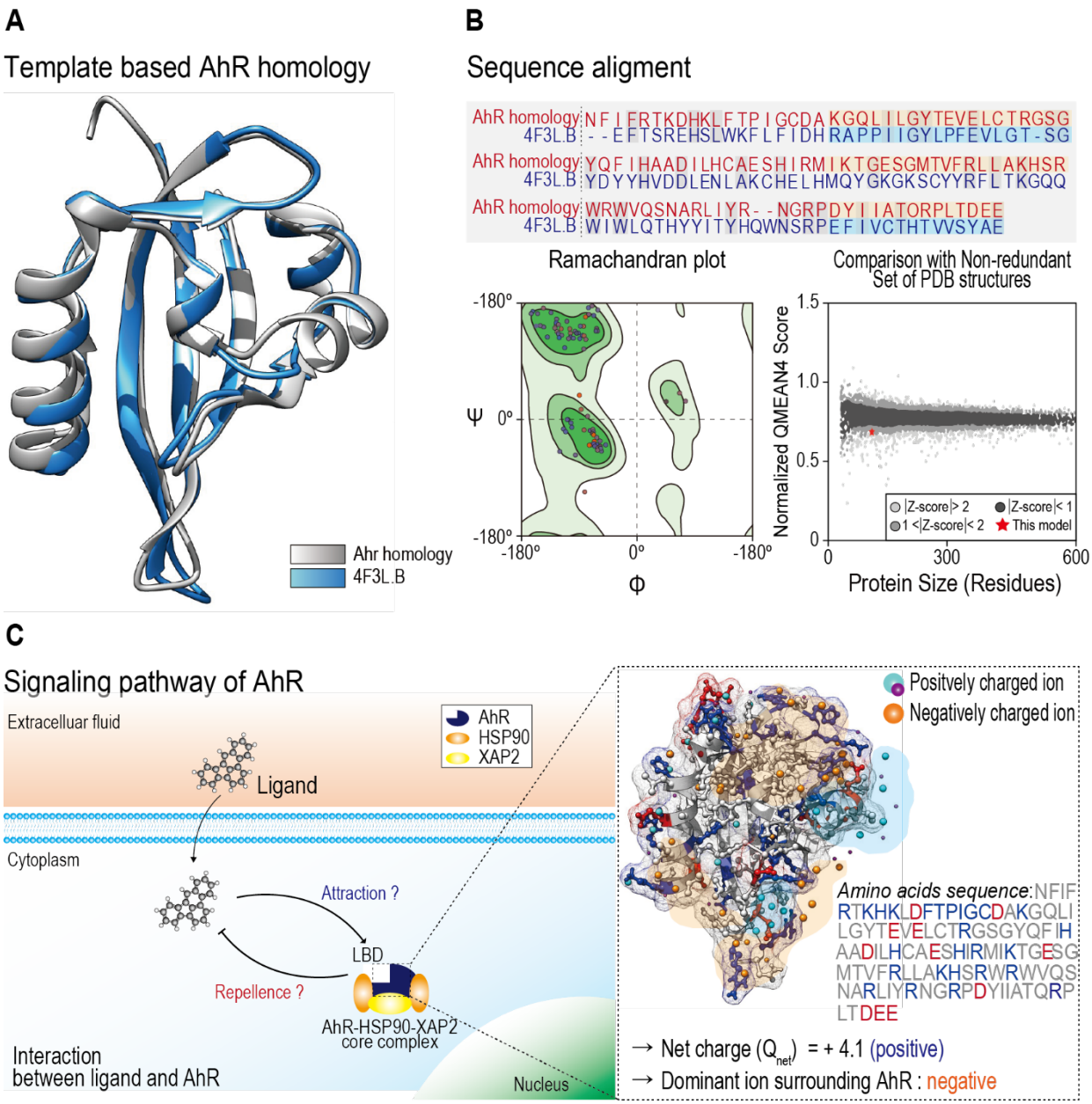


Figure S2. Homology model and quality metrics of AhR ligand binding domain (LBD). (A) Super-imposition of the template (light blue) with the AhR model (white); (B) Sequence alignment between the model and template. RMSD less than 1.0 Å and a structure overlap of 98.13. The ProSA analysis of the generated AhR structure in this study. 2 The figure represents the Prosa-web plot of template 4F3L chain A with a z-score value of − 4.07; (C) Signaling pathway of AhR and net charge of AhR homology model.

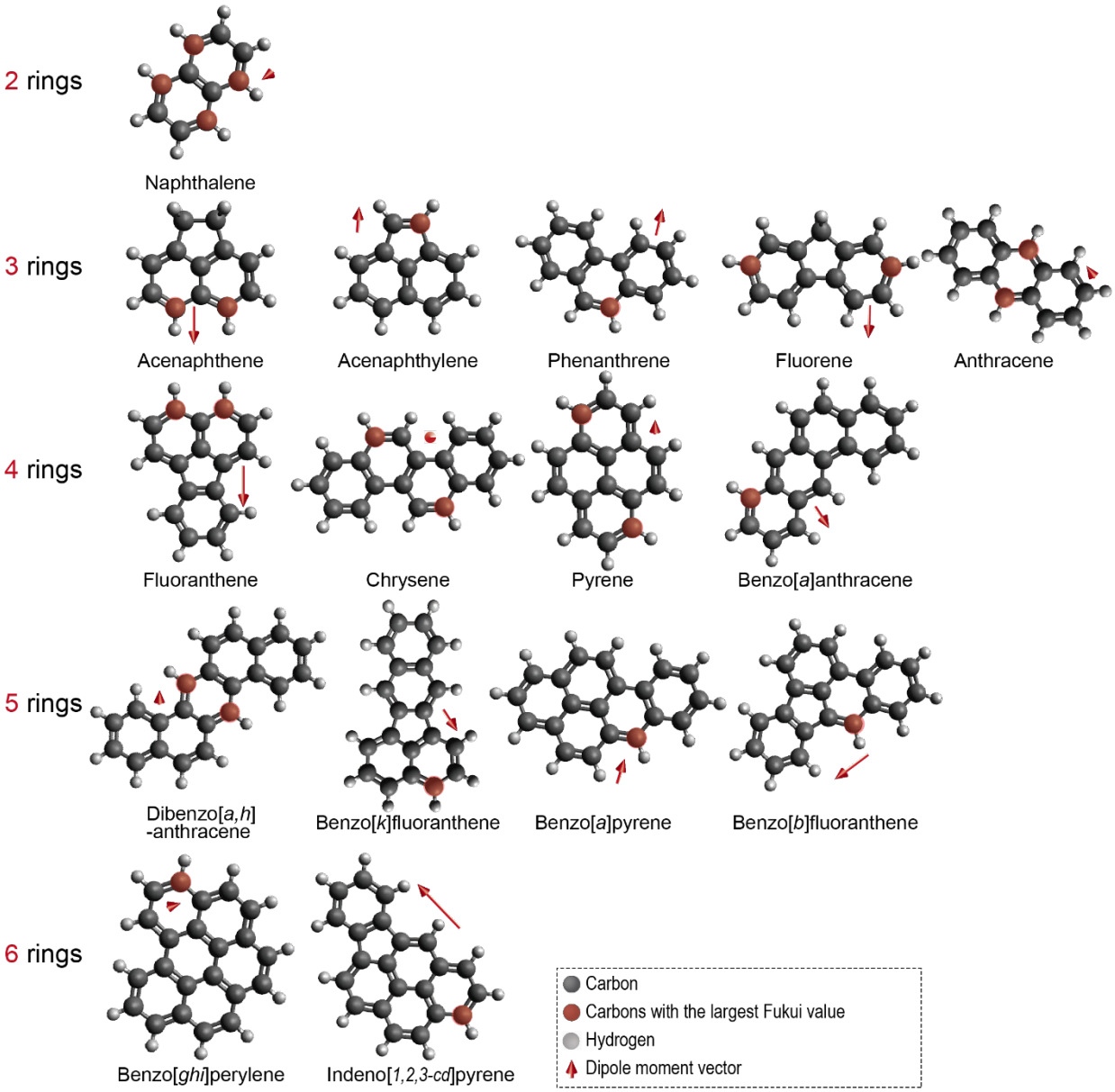


Figure S3. The information of physico-chemical properties of 16 PAHs. Carbon atoms in molecular structure are shown in dark gray, hydrogen atoms are in white, and oxygen atoms in red. A carbon with the highest Fukui value of 16 PAHs is in red, and the red arrow indicates the molecule's dipole moment. The length of the arrow corresponds to the magnitude of dipole moment vector.

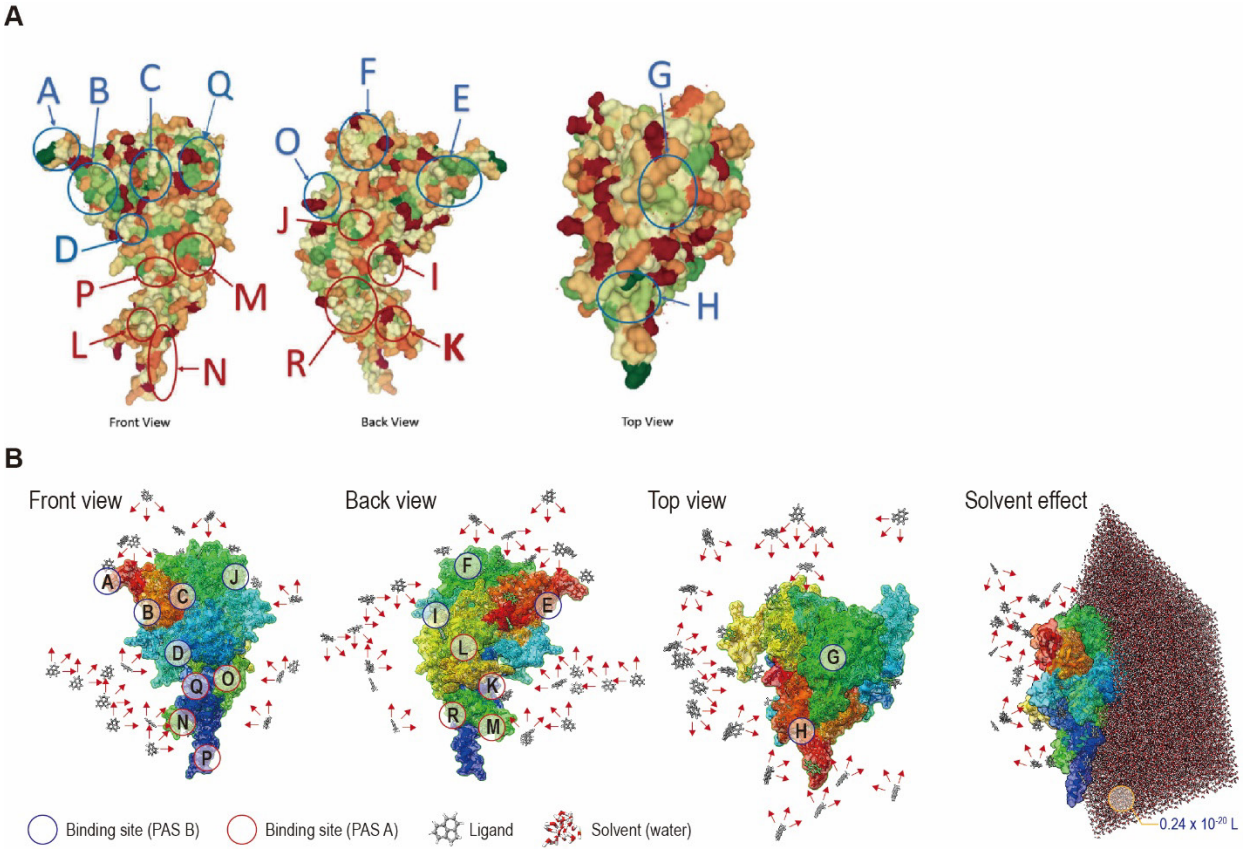


Figure S4. AhR homology model and 16 PAHs binding sites. (A) Total 18 binding sites which were bound by 16 PAHs consisted of Per-Arnt-Sim (PAS) A (red: A, B, C, D, E, F, G, H, and I) and PAS B (blue: J, K, L, M, N, O, P, Q, and R) domains. quality metrics of AhR LBD. (B) The number of total 30 ligands bound to AhR for 20 ns was measured, and the probability of binding to the PAS B domain was calculated.

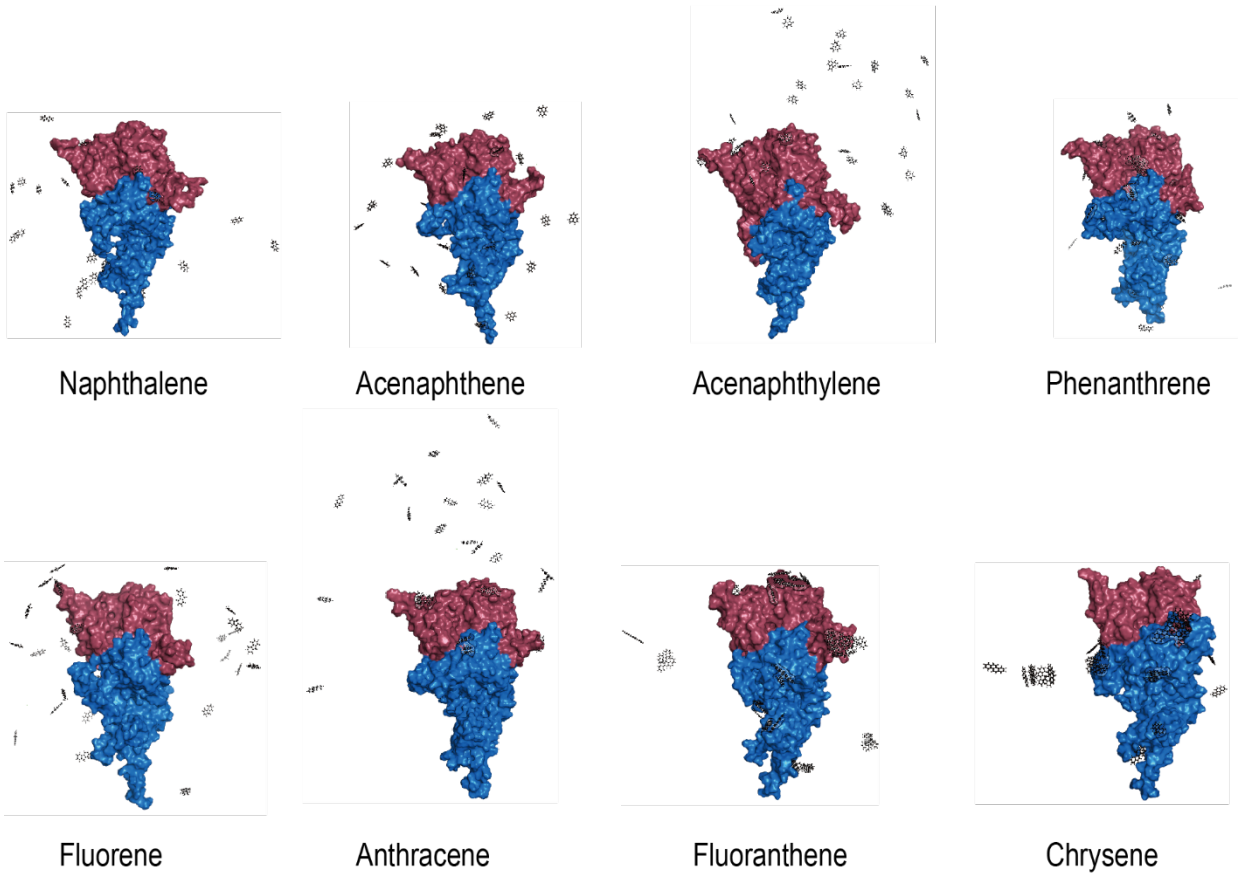


Figure S5. Molecular dynamics simulations of the binding of 16 PAHs (2–4 rings) to AhR in water solution (). AhR homology consists of total two domains including PAS A (blue) and PAS B (red). The simulation was proceeded by the input of total each 30 PAHs to AhR. The results show that PAHs do not depend on only one binding site for binding to AhR.

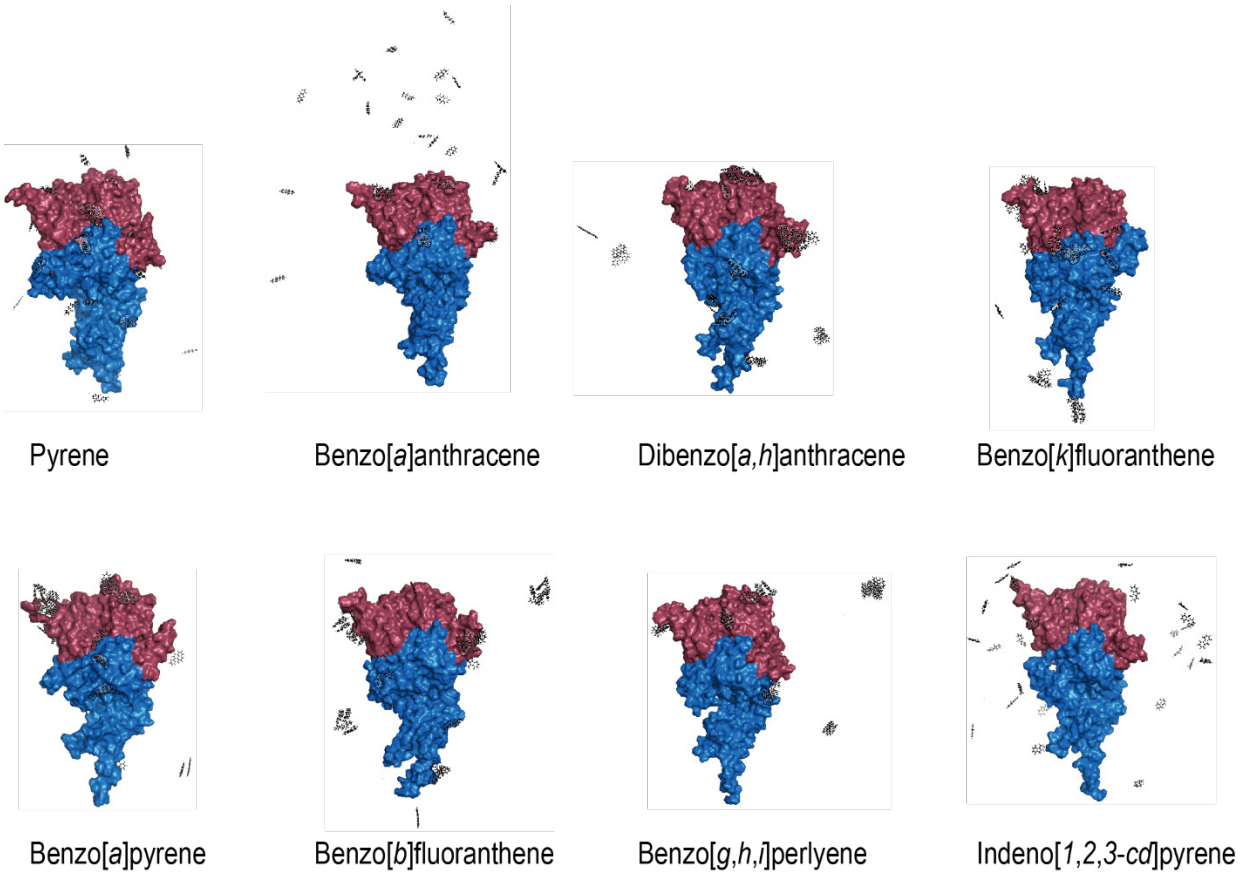


Figure S6. Molecular dynamics simulations of the binding of 16 PAHs (4–6 rings) to AhR in water solution (2). AhR homology consists of total two domains including PAS A (blue) and PAS B (red). The simulation was proceeded by the input of total each 30 PAHs to AhR. The results show that PAHs do not depend on only one binding site for binding to AhR.