Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Youni Zhao Date

Computational Analysis and Modeling of Tox21 AhR Activation High-Throughput Screening Assays

By

Youni Zhao

MPH

Environmental Health

Qiang Zhang, M.D, Ph.D.
Committee Chair

Computational Analysis and Modeling of Tox21 AhR Activation High-Throughput Screening Assays

By

Youni Zhao

Bachelor of Agriculture
Jilin University
2019

Thesis Committee Chair: Qiang Zhang, M.D, Ph.D.

An abstract of
a thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Environmental Health
2021

ABSTRACT

Computational Analysis and Modeling of Tox21 AhR Activation High-Throughput Screening Assays

By Youni Zhao

Dioxin-like compounds (DLCs) including the most potent 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) are important environmental pollutants that can cause potential harms to human health including endocrine disruption and immunosuppression. The toxicities of DLCs are mainly mediated by the aryl hydrocarbon receptor (AhR) in the cytoplasm, which after ligand binding enters the nucleus to bind to dioxin response elements (DRE) to alter the transcription of a number of genes including xenobiotic-metabolizing enzymes and many others. Induction of AhR-mediated genes has been demonstrated to be highly nonlinear in many cases, including steeply sigmoidal responses and nonmonotonic dose responses (NMDR). The present study aimed to analyze the Tox21 AhR high-throughput screening assay data to study the dose-response relationships of more than 10K chemicals. After applying a customized unsupervised and supervised machine learning approach and screening against the parallel cell viability and luciferase interference assays, 1380 active chemicals were selected and their concentration-response curves were classified into 6 shape categories: monotonic increasing, monotonic decreasing, flat, U, Bell and S shape. Among them, 2 are U shaped, 68 Bell shaped, 1092 monotonic increasing, 27 monotonic decreasing, and 1 S shaped. To investigate the nonlinearity of monotonic increasing curves, Hill function was fitted to the data and it was found that the majority of the curves have a Hill coefficient between 1-5, suggesting sigmoidal responses. To understand the potential mechanism of these sigmoidal responses, a stochastic mathematical model of positive cooperative binding between liganded AhR and DREs was constructed. The model demonstrated that with allosteric increase in binding affinity between 3, 5 and 7 DREs, the gene transcription can exhibit sigmoidal responses with Hill coefficient as high as 2.5. In summary, the present study demonstrated that AhR-mediated gene transcription induced by many chemicals can be highly nonlinear, which has significant implications in the risk assessment of these compounds.

Key words: AhR, Tox21, NMDR, Hill function

Computational Analysis and Modeling of Tox21 AhR Activation High-Throughput Screening Assays

By

Youni Zhao

Bachelor of Agriculture
Jilin University
2019

Thesis Committee Chair: Qiang Zhang, M.D, Ph.D.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Environmental Health
2021

Table of Contents

INTRODUCTION

1. Toxicity of TCDD and dioxin-like compound (DLCs)
2. Molecular initiating event (MIE) and mechanism of DLCs
3. Next-generation risk assessment (NGRA) using new approach methodology (NAM)
4. Tox21 program and EPA ToxCast program
5. Nonlinear and non-monotonic dose response (NMDR)
    5.1 Dose-response relationship
    5.2 Nonlinearity (ultrasensitivity) and NMDR of AhR-mediated responses
    5.3 Tox21 assay as a valuable resource to study AhR-mediated nonlinear dose responses

METHODS

1. Tox21 AhR assay
2. Unsupervised machine learning to cluster concentration-response curves
3. Supervised machine learning to classify concentration-response into different shapes
4. Correlation analysis to rule out cytotoxicity and luciferase interference effects
5. Curve-fitting for Hill function
6. Computational modeling of AhR-DRE activation

RESULTS

1. Statistics of Tox21 AhR response assay and viability assay data and threshold values for chemical filtering
2. Unsupervised learning for concentration-response
3. Supervised results of each screening

4. Representative concentration-response curves in each category
5. Hill function fitting
6. Mathematical modeling of positive cooperative binding between AhR and DRE.

DISCUSSION

REFERENCES

INTRODUCTION

1. Toxicity of TCDD and dioxin-like compound (DLCs)

Dioxin-like compounds (DLCs) are a group of chemicals that belongs to the polychlorinated dibenzo-p-dioxin (PCDDs) family and have the same mechanism of action as 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD), the most potent PCDD. TCDD and other DLCs widely exist as persistent environmental pollutants (POPs) [1]. TCDD and DLCs can be produced in a number of industrial activities, including steel smelting, automobile exhaust, incineration such as municipal waste burning, and paper production [1]. TCDD can also exist as a byproduct during the synthesis of organic compounds such as herbicides. The toxicity of DLCs varies greatly, and TCDD is considered to be the most toxic. Animals exposed to dioxins can exhibit developmental defects, immunotoxicity, and liver injury [2]. TCDD also has the role in cancer promotion, by enhancing the carcinogenicity of other carcinogens. High doses of TCDD can cause food intake reduction and wasting syndrome, and even death eventually. Human exposure to high doses of dioxins can result in chloracne, a serious skin condition. The effect of low-dose dioxin exposure in humans is not clear, and it is believed that toxic effects are unlikely to occur with the current level of dioxins in the general human population [3, 4].

2. Molecular initiating event (MIE) and mechanism of DLCs

TCDD and DLCs mainly bind to and activate the aryl hydrocarbon receptor (AhR) in the cell. AhR belongs to the basic helix-loop-helix/Per-Arnt-Sim (bHLH/PAS) family and functions as a transcription factor that regulates gene expression [5-7]. When not bound to a ligand, AhR is mainly associated with the heat shock protein 90 (HSP90) dimer and exists in the cytoplasm.

Hepatitis B virus X-associated protein 2 (XAP2) and prostaglandin E synthase 3 (p23) are also involved, forming a complex with AhR and HSP90 in the cytosol. After AhR binds to a ligand, such as TCDD, AhR first dissociates from the complex. The dissociation from XAP2 exposes the nuclear localization sequence (NLS) region of AhR, which allows AhR complex to bind to the nuclear import complex and be transported into the nucleus. After dissociating from HSP90 and other components of the complex, liganded AhR combines with aryl hydrocarbon receptor nuclear translocator (ARNT), another member of the bHLH family, in the nucleus to form an activated heterodimer. AhR/ARNT heterodimer can bind to the specific AhR-, dioxin- or xenobiotic-response elements (AHRE, DRE or XRE) in the DNA promoters of target genes such as CYP1A1, CYP1A2, and CYP1B1 to regulate their expression. The influence of DLCs on AhR-mediated responses is mainly in two aspects. One is enzyme induction where DLCs can induce several phase I P450 enzymes that can metabolize and detoxify exogenous compounds to protect the body from their toxicity [8]. However, in some cases, carcinogens can also be produced as reactive intermediates that can be electrophilic in nature which can attack DNA causing mutagenesis. In addition to the metabolic effects on xenobiotics, dioxins acting via AhR can affect the transcription of hundreds of other genes, including phase II enzymes, circadian clock genes, metabolic enzymes.

3. Next-generation risk assessment (NGRA) using new approach methodology (NAM)

NGRA is a risk-oriented, hypothesis-driven risk assessment framework. It integrates in vitro, in silico, and traditional methods [7]. The traditional method of assessing the hazards of chemical substances generally requires short-term or long-term animal studies. However, there are a series of related issues that need to be considered in the use of animal testing, such as the protection of animal welfare, the time-consuming nature and high-cost associated with animal study, lack of

usage of modern biology, and lack of relevance to human exposure in real-word scenarios, and so on. The EPA Toxic Substances Control Act (TSCA) list contains more than 85,000 chemical substances, which is increasing every year. The huge backlog of chemicals waiting for approval to come into commerce and the above-mentioned considerations converge to promote the industry and regulatory bodies to bring safe products to the market without conducting animal experiments. Considerable efforts are promoting changes in risk assessment methods.

Advancing the Next Generation of Risk Assessment (NexGen) is a project of the U.S. Environmental Protection Agency (EPA) in cooperation with multiple departments and research institutes. The purpose of the NexGen project is to promote NAMs of risk assessment to practical applications. NAMs mainly refers to non-animal-based experimental and computational methods that can be used to provide information in the context of risk assessment, including integrated approaches to testing and assessment (IATA), defined data interpretation methods, and performance-based test method evaluation [8]. Compared with traditional animal testing, NAMs can provide a large amount of information on a variety of toxic substances in a shorter period. Sometimes the results from animal testing cannot be easily translated to humans. But NAMs can provide more human-related data and are more cost-efficient.

4. Tox21 program and EPA ToxCast program

Tox21 is a federal collaboration project involving four government agencies, including the U.S. Environmental Protection Agency (EPA), National Toxicology Program (NTP) at the National Institute of Environmental Health Sciences (NIEHS), National Center for Advancing Translational Sciences (NCATS), and Food and Drug Administration (FDA) [9]. Each institution provides its

own unique data, expertise, and tools, etc. The Tox21 project consists of three stages, and the project is currently in the third stage. The first stage is mainly to use the high-throughput screening of robotic technology to obtain high-quality test results, which can reduce the use of experimental animals. The second stage is to further expand the scope of testing. There are currently more than 10,000 compounds in the database, covering industries, food, and cosmetics, etc. The Tox21 project is also evaluating more cell-based experiments to further define the impact of chemical exposure. The third phase of the project will further deepen the research. For example, the development of alternative test systems that can predict the toxicity and dose-effect response of substances, solve the technical limitations of in vitro test systems, and predict human diseases by evaluating the effects of substances in cells, etc [9-11].

ToxCast at USEPA is a high-throughput screening project somewhat overlapping with Tox21 but was initiated earlier. The ToxCast database contains data on approximately 1800 chemical substances screened for more than 700 biochemical and molecular endpoints [9]. The goals of this project are in line with EPA's plan to reduce mammalian research by 30% by 2025 and eliminate all mammalian research by 2035.

5. Nonlinear and non-monotonic dose response (NMDR)

5.1. Dose-response relationship

Dose-response relationship refers to the function of an organism's response to the exposure to a chemical substance for a certain period. Visualized on a x-y coordinate, usually the dose or concentration of the chemical is on the x-axis, and the biological response is on the y-axis. Dose-

response relationship is crucial for determining the harmful and safe levels of chemical substances and is a key component of the risk assessment framework adopted by the USEPA.

5.2. Nonlinearity (ultrasensitivity) and NMDR of AhR-mediated responses

For receptor-mediated responses, such as those through the AhR, a chemical may function as an agonist or antagonist. Receptor-mediated responses usually follow the Michaelis-Menten kinetics, i.e., low-dose linear and saturable at high doses. However, AhR-mediated dose responses have been shown to exhibit nonlinearity that departs from the Michaelis-Menten kinetics. It was found that different types of AhR ligands can cause different degrees of sigmoidal responses that can be empirically described by the Hill function. Sigmoidal responses represent ultrasensitivity wherein signal amplification occurs. There are several mechanisms of ultrasensitivity, such as positive cooperative binding, multistep signaling, molecular titration, zero-order ultrasensitivity, positive feedback [12]. It is unclear how AhR-mediated ultrasensitivity occurs. Studies have also shown that AhR ligands can produce nonmonotonic dose response curves (NMDRCs) [13]. The NMDRCs can be divided in general into U-shaped curves, Bell-shaped curves, and S-shaped curves, etc. For these nonlinear dose-response relationships, the concept of linear relationships is not suitable.

Based on the different response types, different regression methods can be applied for statistical analysis of the dose-response curves. Empirical models based on nonlinear regression are usually given priority. However, such approaches do not provide insight to the mechanisms underlying these nonlinear dose responses. Mechanistically modeling the AhR-mediated pathways

is critically important to reveal the potential molecular interactions that may be responsible for the ultrasensitive and NMDR responses elicited by AhR ligands.

5.3 Tox21 assay as a valuable resource to study AhR-mediated nonlinear dose responses

Tox21 provides screening data on activation of AhR signal transduction in response to more than 10,000 (10K) chemicals in agonist mode through quantitative high-throughput screening (qHTS) assay [10]. It also provides data on cell viability to help rule out any cytotoxicity-related effects. Containing 15-concentration points with triplicate assays for each chemical, the Tox21 AhR assay provides an excellent opportunity for us to study the nonlinearity in dose-response relationship of DLCs.

METHODS

1. Tox21 AhR assay

AhR can be used to detect dioxin pollutants in the environment. The chemically activated luciferase expression (CALUX) system is a kind of recombinant luciferase reporter cell bioassay that is based on AhR response. For the Tox21 AhR assay, HG2L7.5c1 cell line that was developed by Dr. Michael S. Denison, University of California at Davis, was used to assess the activation of AhR [14]. HepG2 (human hepatocellular carcinoma) cells were transfected with an Ah receptor-responsive firefly luciferase reporter gene plasmid, which contains 20 DREs and luciferase reporter gene, so an increase in luciferase activity can be used to sensitively detect the presence of AhR ligands. The Tox21 assay also includes cell viability data, which is obtained by parallel testing the cytotoxicity of compounds in the HG2L7.5c1 cell line. Tox21 data is publicly released and can be downloaded from the ***Tripod Development*** website *(https://tripod.nih.gov/tox21)*. The whole dataset contains 10496 chemicals, each chemical was tested for 15 different concentrations for at least three replicates.

2. Unsupervised machine learning to cluster concentration-response curves

All analyses were conducted in MatLab except for the dynamic modeling which was conducted in Julia. The purpose of unsupervised machine learning method is to cluster the concentration-response curves into several possible types, so that they can be visually inspected and consolidated into some main types of shapes, which are then used to better classify and filter the curves. The clustering approach is as follows. (1) The replicate data of each chemical were averaged to obtain the average concentration-response. (2) Then, inactive chemicals were removed if the averaged 15

response data points have variance < 30, or are within range of [-20, 20], or have obvious outlier response points (which differ from the two immediate neighboring points by > 40). (3) 5-concentration-points moving-average of the remaining curves was calculated. (4) A random chemical *i* was selected and a chemical *j* from the remaining chemicals having the highest curve correlation with *i* was found and formed a cluster [*i, j*]. Repeat the process until the Pearson correlation coefficient < 0.75. (5) Randomly select another chemical *i* from the remaining chemicals and repeat the above step 4 until all chemicals are grouped.

3. Supervised machine learning to classify concentration-response into different shapes

From the previous step, the clusters of concentration-response curves were visually inspected and consolidated into six categories, including monotonic increasing, monotonic decreasing, flat, U, Bell and S shape. To more accurately classify the curves into these categories, the following supervised learning algorithm was applied. (1) If the maximum response value of the curve is greater than the response values of the lowest and highest concentrations by a certain threshold (5, 10, or 15), and the maximum response value is not at the second lowest or second largest concentration, then this curve is classified into the bell shape. (2) The condition for U shape is similar but with negative threshold values. (3) If the curve satisfies the above two conditions at the same time, it belongs to S shape. We run both 5-concentration-point and 2 concentration-point moving-average curves against all these conditions to screen for as many NMDR curves as possible. In the remaining curves that were not classified into the above NMDR categories, if the absolute value of the difference between the maximum response and minimum response values is < 5, 10, or 15 for a curve, then the curve belongs to a flat shape. In the remaining curves, if the response value at the highest concentration is greater than the response value by 5, 10, or 15 at the

lowest concentration, then the curve belongs to monotonic increasing shape; if the response value at the highest concentration is lower than the response value at the highest concentration by 5, 10, or 15, then the curve belongs to monotonic decreasing shape.

4. Correlation analysis to rule out cytotoxicity and luciferase interference effects

Chemical-induced changes in cell viability will affect the interpretation of the response data, increasing the chance of observing false positive responses. For example, in a Bell or monotonic decreasing curve, the downward trend can be due to reduced cell viability, not true decrease in AhR activity. To identify those curves that have such characteristics, we first located the concentration that produced the maximal response and then compare the trend of the segment of the curve to the right of the maximal response concentration with the corresponding segment of the viability curve. If the Pearson correlation coefficient is > 0.6, then the response curve is likely to be influenced by cell viability and screened out. The interference of the luciferase assays by a chemical may also affect the interpretation of the response data. A similar screening for luciferase interference was conducted as the cytotoxicity screening above.

5. Curve-fitting for Hill function

For monotonic increasing shape, we used the Hill equation *ymin + (ymax - ymin)* x^n / (K^n + x^n)* to fit the data to assess their degree of nonlinearity by using the "fit" function in MatLab. The starting values of the parameters were taken from the averaged values of the replicate data provided by the Tox21 Tripod dataset.

6. Computational modeling of AhR-DRE activation

After DLCs activate AhR, multiple DREs can be occupied by the AhR/ARNT heterodimer in the nucleus. And the interactions among multiple DREs clustered together in a gene promoter have a potential to result in cooperative binding of AhR/ARNT dimer to the promoter, which may explain the concentration-response curves exhibiting high Hill coefficients. To understand this likely mechanism, we constructed a dynamic model simulating liganded AhR binding to 20 DREs as in the luciferase reporter cell line used in the AhR assay. The dynamic model was simulated by using Gillespie's stochastic simulation algorithm [15] in Julia language. Considering the structure of DNA and the length of the bound fragments, we explored the situation when one DRE occupancy can enhance the binding of the immediate 2 (AhR_3DRE), 4 (AhR_5DRE), and 6 (AhR_7DRE) neighboring DREs. The general form of the three scenarios in Julia are as follows:

AhR_3DRE_model:

*ki0f*foldincrease((AHR_ARNT_DRE(i-1) + AHR_ARNT_DRE(i+1)),base), AHR_ARNT + DRE(i) --> AHR_ARNT_DRE(i);*

ki0b, AHR_ARNT_DRE(i) --> AHR_ARNT + DRE(i)

AhR_5DRE_model:

*ki0f*foldincrease((AHR_ARNT_DRE(i -2) + AHR_ARNT_DRE(i-1) + AHR_ARNT_DRE(i+1) + AHR_ARNT_DRE(i+2)),base), AHR_ARNT + DRE(i) --> AHR_ARNT_DRE(i);*

ki0b, AHR_ARNT_DRE(i) --> AHR_ARNT + DRE(i)

AhR_7DRE_model:

*ki0f*foldincrease((AHR_ARNT_DRE(i-3) + AHR_ARNT_DRE(i-2) + AHR_ARNT_DRE(i-1) + AHR_ARNT_DRE(i+1) + AHR_ARNT_DRE(i+2) + AHR_ARNT_DRE(i+3)),base), AHR_ARNT + DRE(i) --> AHR_ARNT_DRE(i);*

ki0b, AHR_ARNT_DRE(i) --> AHR_ARNT + DRE(i).

The parameter *foldincrease* represents the fold increase in binding affinity between AhR and DRE. *ki0f* and *ki0b* represent the default association and dissociation rate constants of the binding, respectively. When evaluating the effects of multiple DREs, we assumed that each occupied DRE has an additive effect on gene expression. 10000 stochastic simulations were conducted, and the average response was calculated to obtain the dose-response curve. Hill coefficient *n* can be calculated from $\frac{log_{10}(81)}{log_{10}(EC_{90}/EC_{10})}$, where EC_{90} and EC_{10} represent the concentration producing 90% and 10% of the maximal response respectively.

RESULTS

1. Statistics of Tox21 AhR response assay and viability assay data and threshold values for chemical filtering

The data of the response assay (Fig. 1) and viability assay (Fig. 2) were visualized in the form of histograms and scatter plots to facilitate the selection of appropriate threshold values for filtering. For the response assay, a scatter plot (Fig. 1A) and a histogram (Fig. 1B) of the variance of all averaged-response data points, and a scatter plot (Fig. 1C) of all data values and a histogram (Fig. 1D) of the data value frequency were presented. Scatter plots of maximum values vs log variance (Fig. 1E) and minimum vs log variance (Fig. 1F) were presented also. The red lines in each figure indicate the threshold values used to filter out inactive chemicals that did not elicit a significant response in the AhR response assay. For chemicals that do not cause cytotoxicity at all, the lower and high boundaries in the viability assay were set between -20 and 20 (Fig. 2).

2. Unsupervised learning for concentration-response

After filtering the inactive chemicals, we subjected the remaining 1380 chemicals to the unsupervised learning algorithm, which generated 18 clusters of curve shapes. After visual inspection we consolidated these clusters into 6 shape categories: monotonic increasing, monotonic decreasing, flat, U, Bell and S shape.

3. Supervised results of each screening

In the supervised machine learning step, we tried different Bell_shape_magnitude, U_shape_magnitude, Flat_magnitude cut-off values (5, 10 and 15), and the corresponding results

are shown in Figs. 3-5. As we can see, the number of NMDR curves increased with lower magnitude cutoff values. We took a conservative approach by using 15 as the final cutoff value for subsequent analysis.

4. Representative concentration-response curves in each category

According to the supervised machine learning method, some representative dose-response curves in each category are presented in Fig. 6-11. Scatter plots for $\log_{10}$ concentration of inflexion point vs. magnitude of U and Bell shape curves are in Fig 12.

5. Hill function fitting

1092 monotonic increasing curves were fitted to the Hill function, and the distribution of Hill coefficients (n) vs. Ymax and n vs K (EC50) are presented in Fig. 13A and 13B. The majority of the curves have n values between 0-5 and some curves have n values between 15-20. Two representative fitted curves are shown in Fig. 13C and 13D.

6. Mathematical modeling of positive cooperative binding between AhR and DRE.

The AhR-DRE binding model was simulated by assuming multiple possibilities including the number of DREs that can allosterically enhance each other's binding, and the fold-increase in sequential binding. The Hill coefficient and EC50 of the concentration-response curves obtained under different parameter conditions are summarized in Table 1. It is interesting to note that regardless of the number of DREs involved in cooperative binding, the Hill coefficient seems to be maxed out around 2.5. Within each condition of DRE numbers (3, 5, or 7), the fold-increase (base) in sequential binding has a nonmonotonic effect on the Hill coefficient, which decreases if

the fold-increase is elevated further. The EC50 value responds to the fold-increase in binding monotonically, which decreases as the binding affinity increases. Some representative concentration response curves predicted by the model are shown in Fig. 14.

DISCUSSION

In this study, through the clustering and classification of the concentration-response curves of Tox21 AhR assays, we were able to identify over a thousand chemicals among the 10K library that are active in inducing AhR-mediated transcription as monitored by the luciferase reporter gene construct containing 20 DREs. The concentration-responses are dominated by monotonic increasing curves (1092 chemicals), which is not surprising given the Tox21 AhR assay is an activation or agonist assay, and therefore we should not expect any real responses below the baseline as would be exhibited by a monotonic decreasing response. There are nevertheless 68 monotonic decreasing curves, most of which can be attributed to cytotoxicity or due to assay variability since the cutoff value we used is 15, which can still be relatively small considering the noise range in the baseline response.

There are 68 curves identified as Bell shape and only 2 as U shape (Fig. 12). Most of the Bell shape curves can be attributed to cytotoxicity although there are chemicals that exhibited a strong trend of Bell shape without sign of compromised viability, such as Coumaphos and 3,3',5,5'-Tetramethylbenzidine (Fig. 7). The mechanism of the Bell-shaped response is unclear. There are a number of local and global mechanisms that can potentially generate NMDR responses [12, 15-18]. Recently Simon et al. proposed that squelching of coactivators by free AhR-ARNT dimer may explain a Bell-shaped response of AhR ligand [13]. As in the case of U shape curves, it is also not surprising that one 2 were identified. U-shaped curves are more likely to occur when there is an endogenous ligand that activates the AhR to some high baseline level such that when an exogenous chemical is added, some nonlinear interactions between receptor dimers and between receptor and

DREs may initially decrease the baseline gene transcription activity before the repressing force is saturated and taken over by an activating force requiring higher concentrations of the exogenous ligand [19, 20]. Given the fact that the AhR assay is an activation assay, and no activating ligand was added to the assay to begin with, U-shaped curves should not be expected. However, U-shaped response may still occur in vivo although the nature of the endogenous ligands of AhR is still elusive.

Gene induction by AhR ligand is known to exhibit highly switch like response, such as CYP450 enzyme induction by TCDD or other AhR ligand in the liver or in hepatocytes [12]. To investigate how common such sigmoidal responses are in AhR-mediated gene induction, we conducted curve-fitting for the 1092 monotonic increasing curves with the Hill function to assess their degree of steepness. We found that most of the curves exhibit Hill coefficients between 1-5, with a small fraction displaying very high values up to 20 (Fig. 13). A close inspection of those curves with high Hill coefficient values indicates that nearly all of them do not show saturable responses and the maxima responses are usually very small, thus reducing the confidence for the predicted high Hill coefficient values. These results demonstrated that AhR-mediated gene transcription can exhibit chemical-specific ultrasensitive responses. One potential function of ultrasensitive gene induction by AhR ligands is to upregulate xenobiotic-metabolizing enzymes to very high levels to clear the exogenous chemicals from the cell as much as possible.

There are many potential biochemical mechanisms generating ultrasensitivity, sigmoidal response at the cellular level, including positive cooperative binding, homodimerization, molecular titration, zero-order ultrasensitivity, and multistep signaling [21]. Given the reporter

gene construct for the Tox21 AhR assay, the 20 DREs in the promoter suggests highly that positive cooperative binding of AhR-ARNT may be the underlying mechanism. To explore this possibility, we constructed a mathematical model to simulate liganded AhR binding to multiple DREs and enhancement of the binding affinity among different numbers of neighboring DREs. The model simulates the situation when the number of allosterically interacting DREs is 3, 5, and 7, respectively. In each case, it includes 7 situations with fold-increase in binding affinity ranging between 1, 2, 5, 10, 20, 50, and 100. Surprisingly we found that in all conditions the maximum value of the Hill coefficient is around 2.57 (Table 1). Normally one would expect that the maximal Hill coefficient is equal to the number of DREs that can allosterically interact with one another. But it is not the case predicted by the model. Moreover, the fold-increase in sequential binding affinity does not lead to a monotonic positively associated change in the Hill coefficient - there seems to be an optimal value which produces the highest Hill coefficient.

The advantage of the present study lies in the classification and visualization of the dose response data in the AhR assay. And it effectively filters out insignificant data and those interfered with by cytotoxicity and luciferase inhibition. The research is innovative in that it is the first to examine the nonlinearity of the Tox21 data and establish a positive cooperative binding mathematical model to interprint some of the sigmoidal responses. There are also some limitations of the analyses and modeling. The Hill function curve-fitting can be further optimized to include concentration-responses that show signs of saturation. The cutoff values for identifying NMDR curves can be more stringent to eliminate those response curves with trivial magnitude that likely are due to assay variability. The mathematical model can be further strengthened to include coactivator binding that will allow us to predict B-shaped curves.

In summary, the present study demonstrated that AhR-mediated gene transcription induced by many chemicals can be highly nonlinear, which has significant implications in the risk assessment of these compounds.

FIGURES

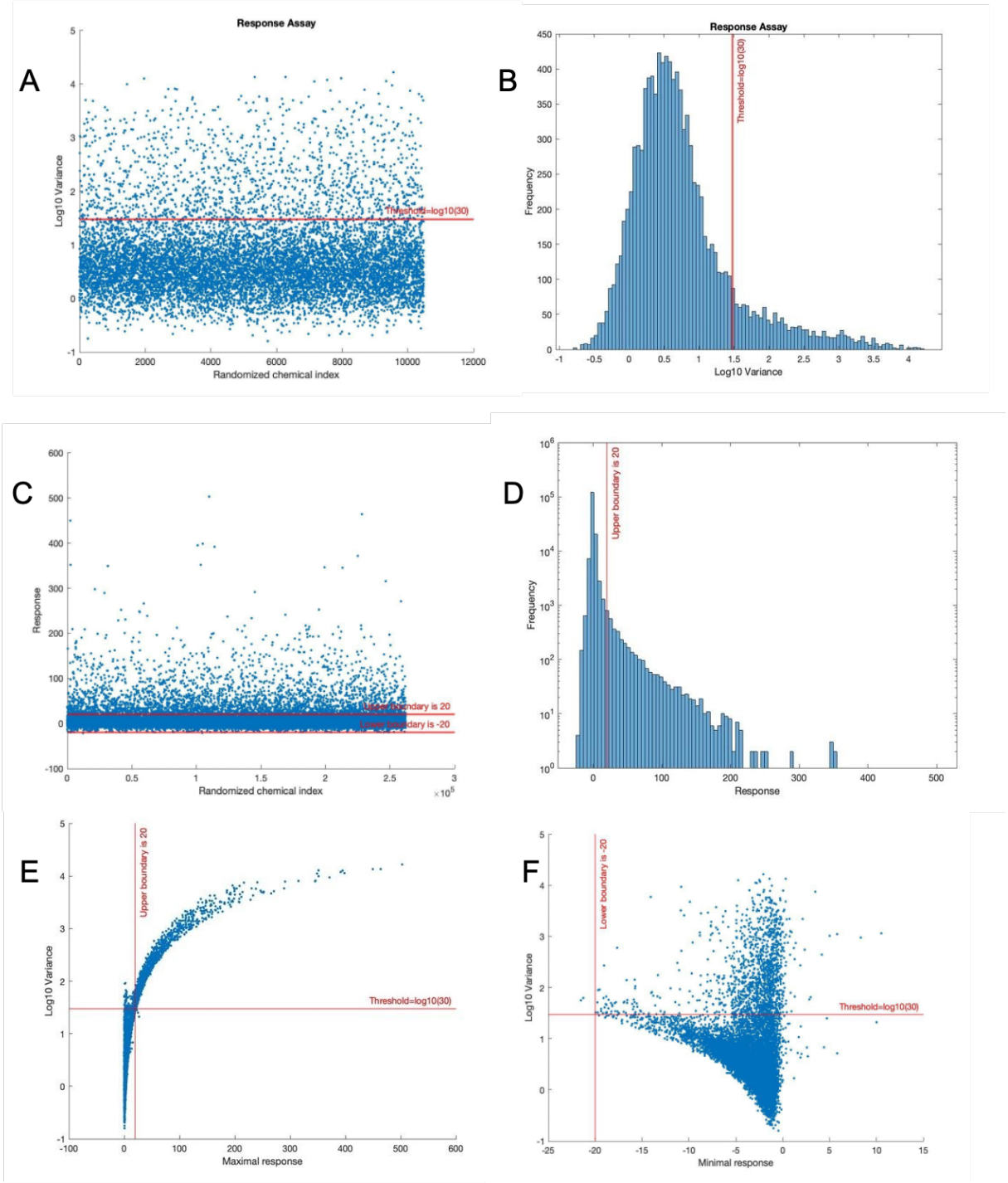

Fig 1. Visualization of the Tox21 AhR response assay data. The low variance threshold is 30, the upper boundary for response data is 20, and lower boundary for response data is -20.

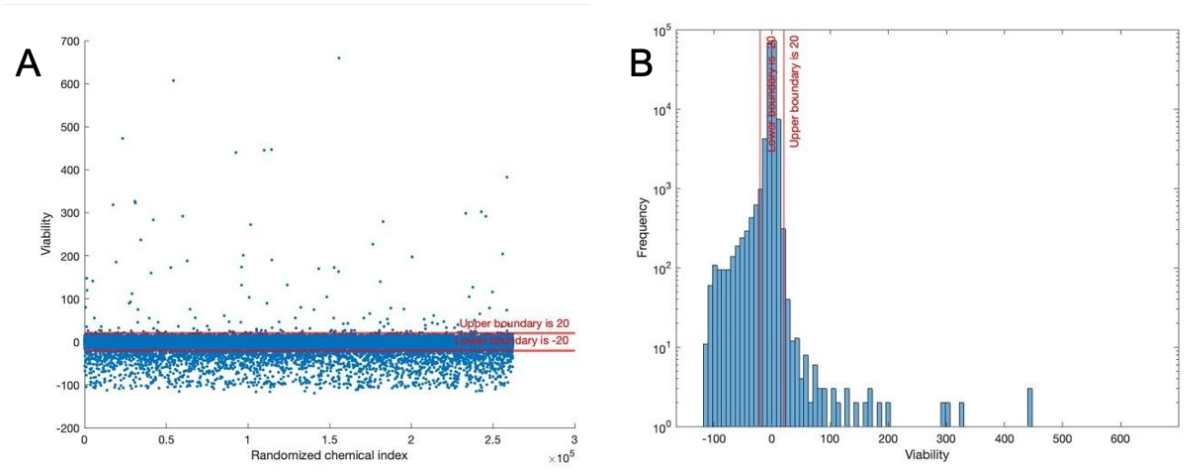

Fig 2. Visualization of the Tox21 AhR viability assay data. The upper boundary and lower boundary are [-20,20] to indicate chemicals that do not cause cytotoxicity.

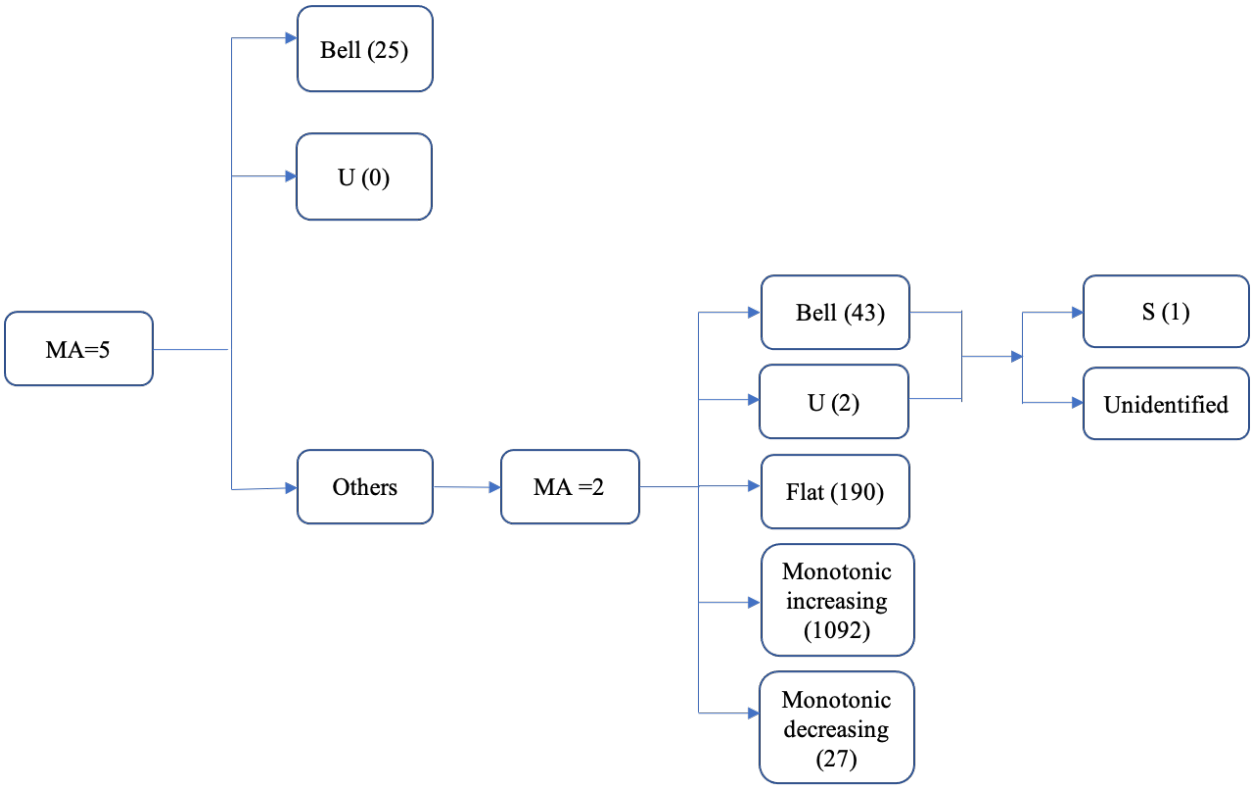

Fig 3. Classification of concentration-response curves of Tox21 AhR assay using magnitude cutoff value of 15. The number of each type of curve is shown in the boxes. The total number of U shape curves is 2 and the total number of Bell shape curves is 68.

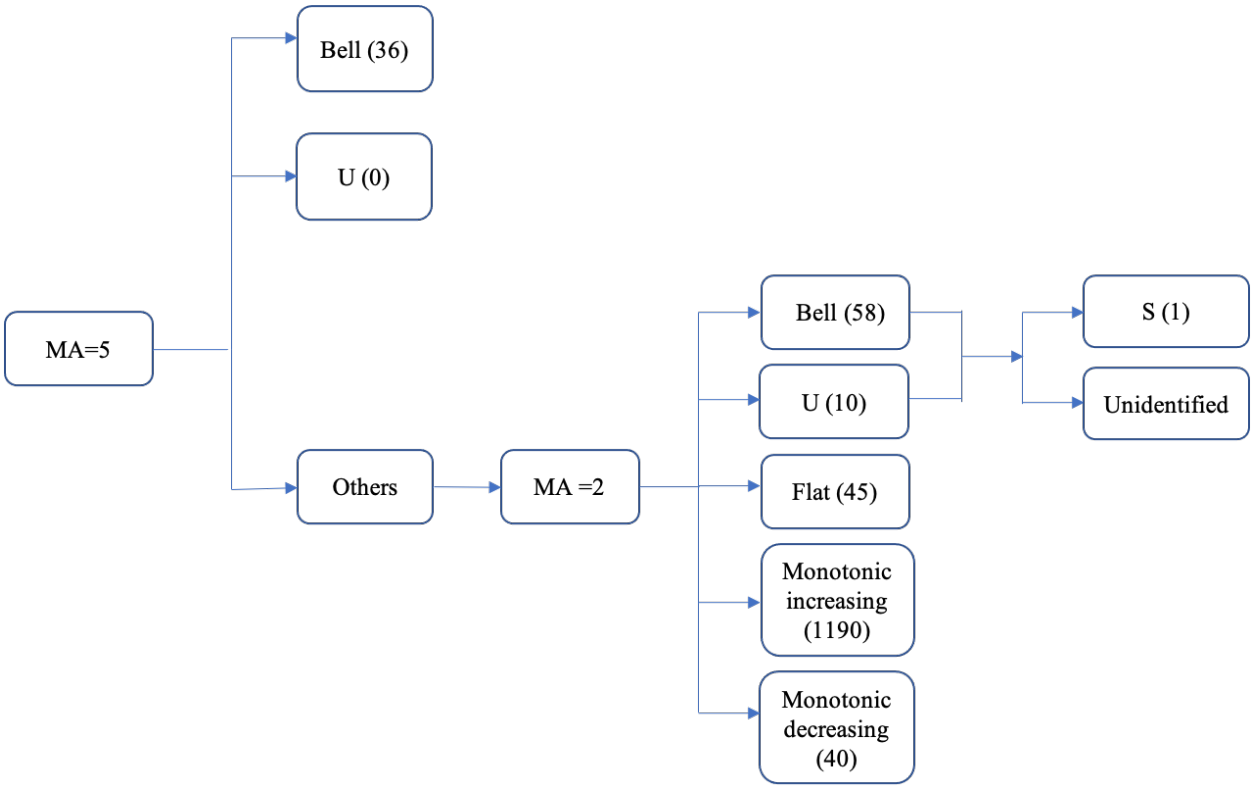

Fig 4. Classification of concentration-response curves of Tox21 AhR assay using magnitude cutoff value of 10. The number of each type of curve is shown in the boxes. The total number of U shape curves is 10 and the total number of Bell shape curves is 94.

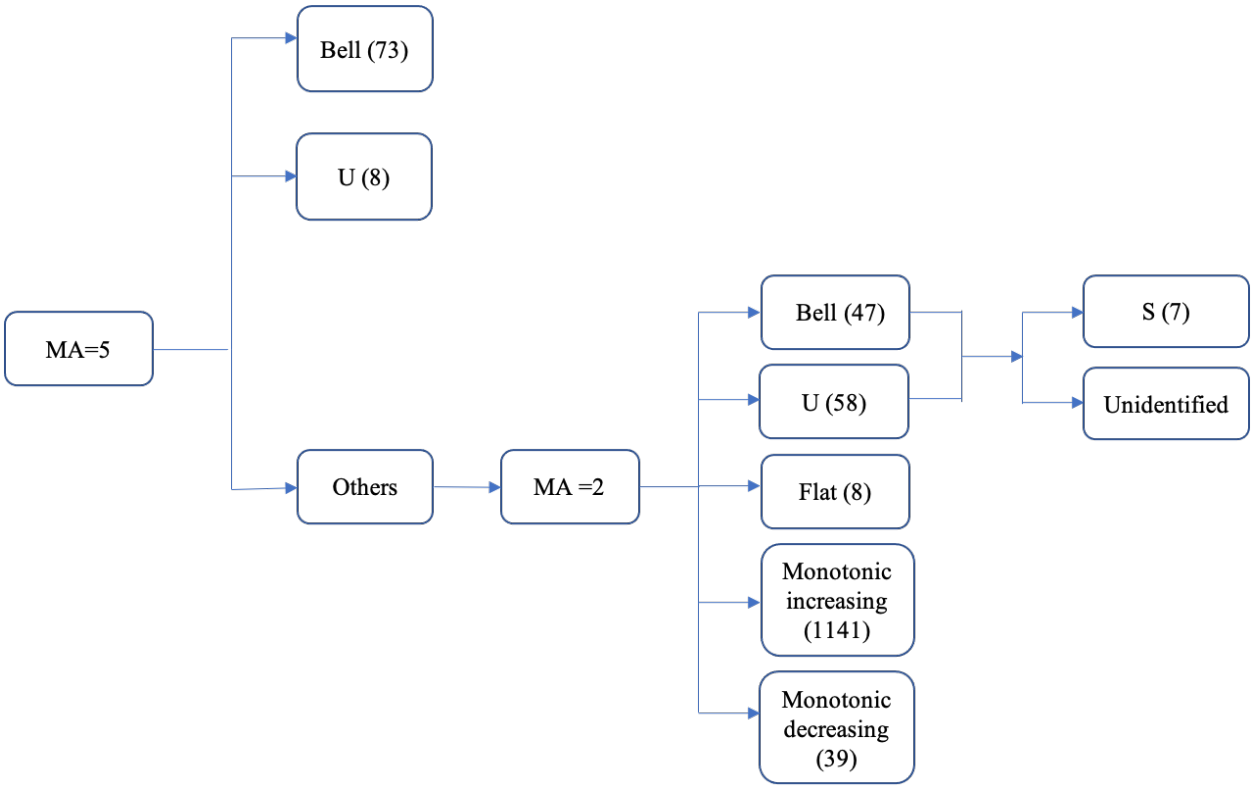

Fig 5. Classification of concentration-response curves of Tox21 AhR assay using magnitude cutoff value of 5. The number of each type of curve is shown in the boxes. The total number of U shape curves is 66 and the total number of Bell shape curves is 120.

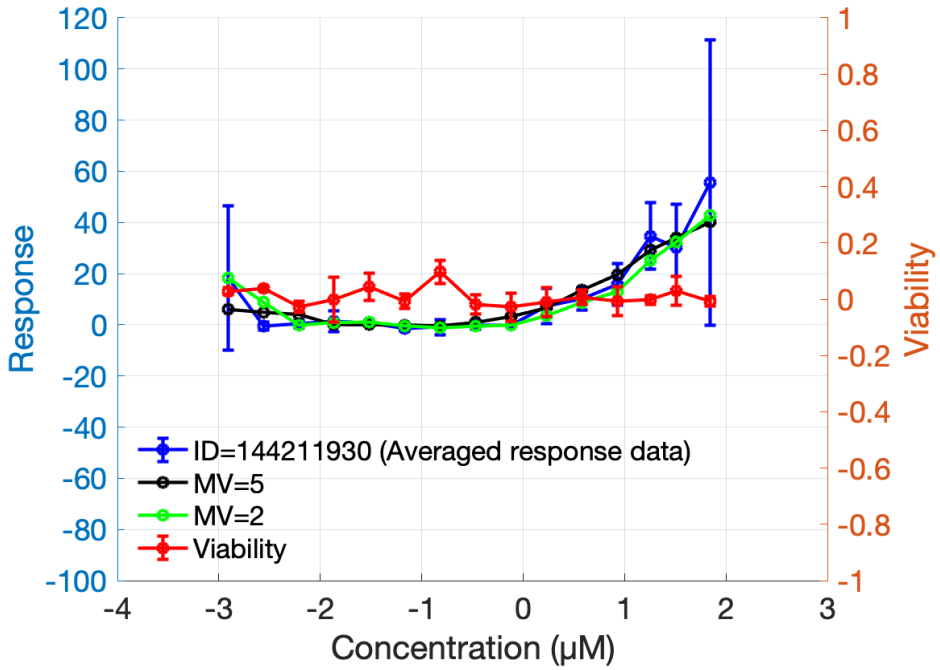

Fig 6. A typical curve of U shape, exhibited by N-Benzyl-9-(tetrahydro-2H-pyran-2-yl) adenine. The blue line represents the average of the data of all replicates, the black line represents the 5-concentration-point moving-average, the green line represents the 2-concentration-point moving average, and the red line represents the viability data.

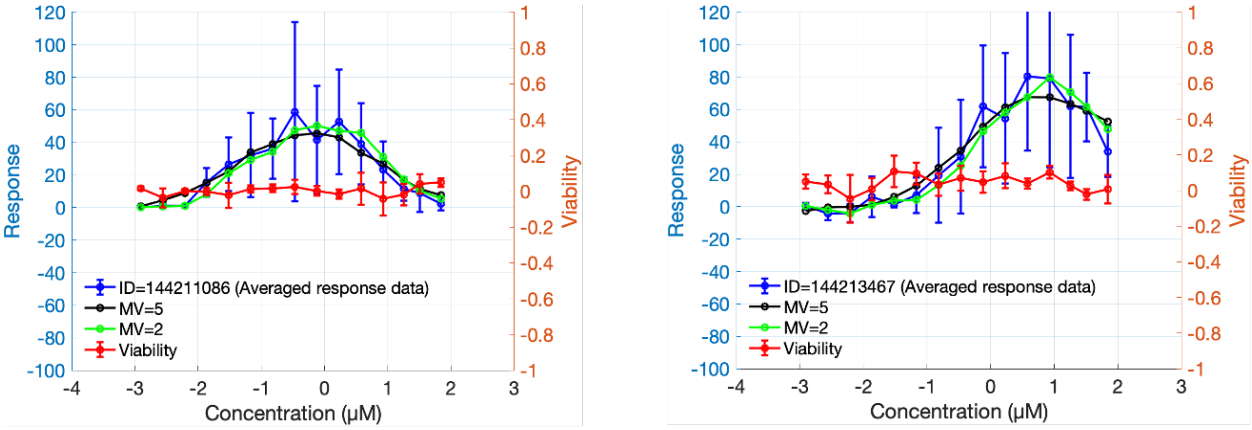

Fig 7. Two curves of Bell shape from Coumaphos and 3,3',5,5'-Tetramethylbenzidine. Color coding scheme is the same as in Fig. 6.

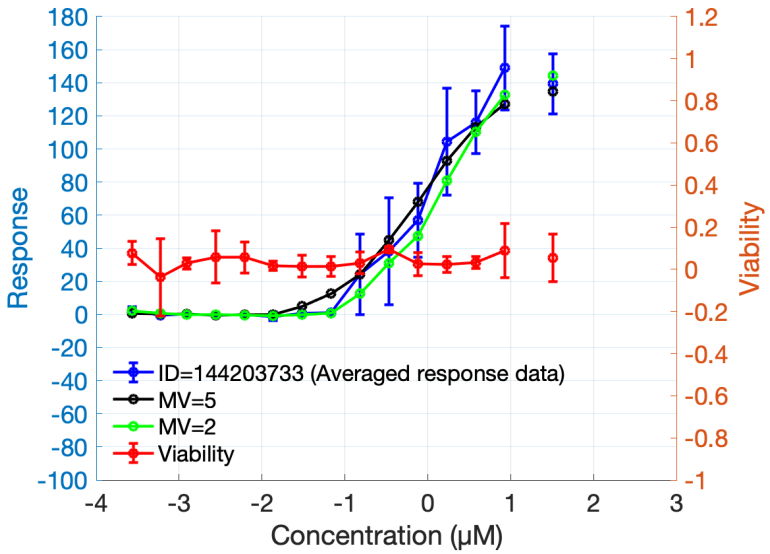

Fig 8. A typical curve of monotonic increasing shape from Leflunomide. Color coding scheme is the same as in Fig. 6.

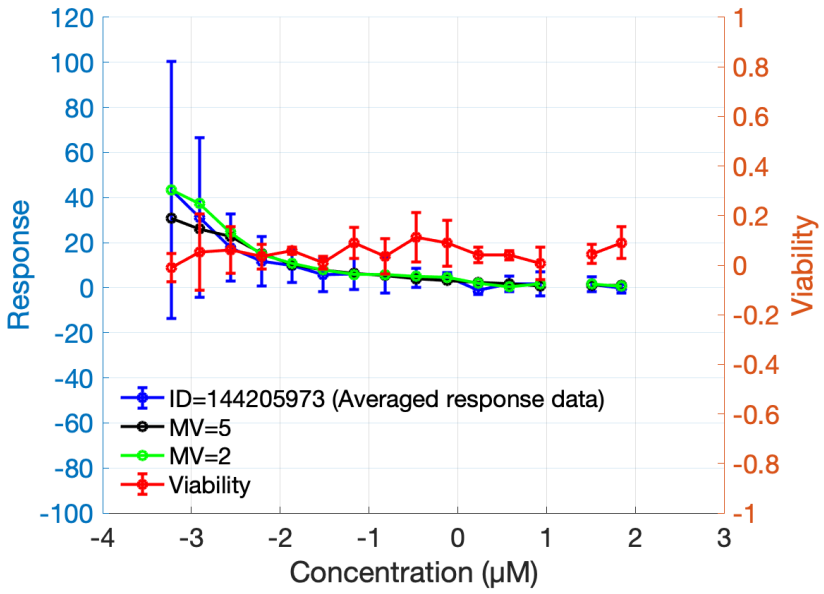

Fig 9. A typical curve of monotonic decreasing shape from Thymopentin. Color coding scheme is the same as in Fig. 6.

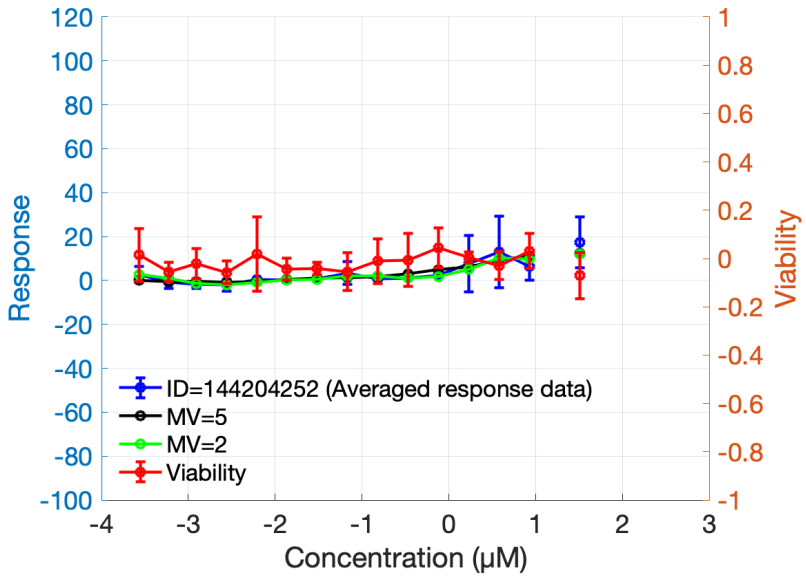

Fig 10. A typical curve of flat shape from Alfuzosin hydrochloride. Color coding scheme is the same as in Fig. 6.

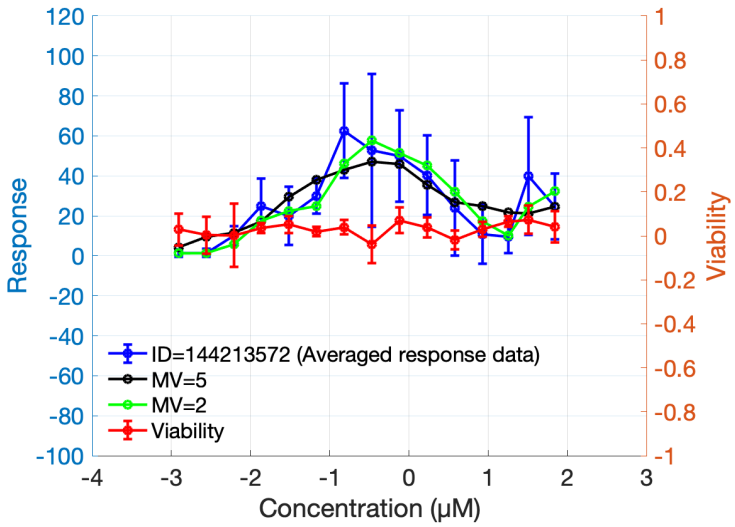

Fig 11. A typical curve of S shape from 5-Methoxypsoralen. Color coding scheme is the same as in Fig. 6.

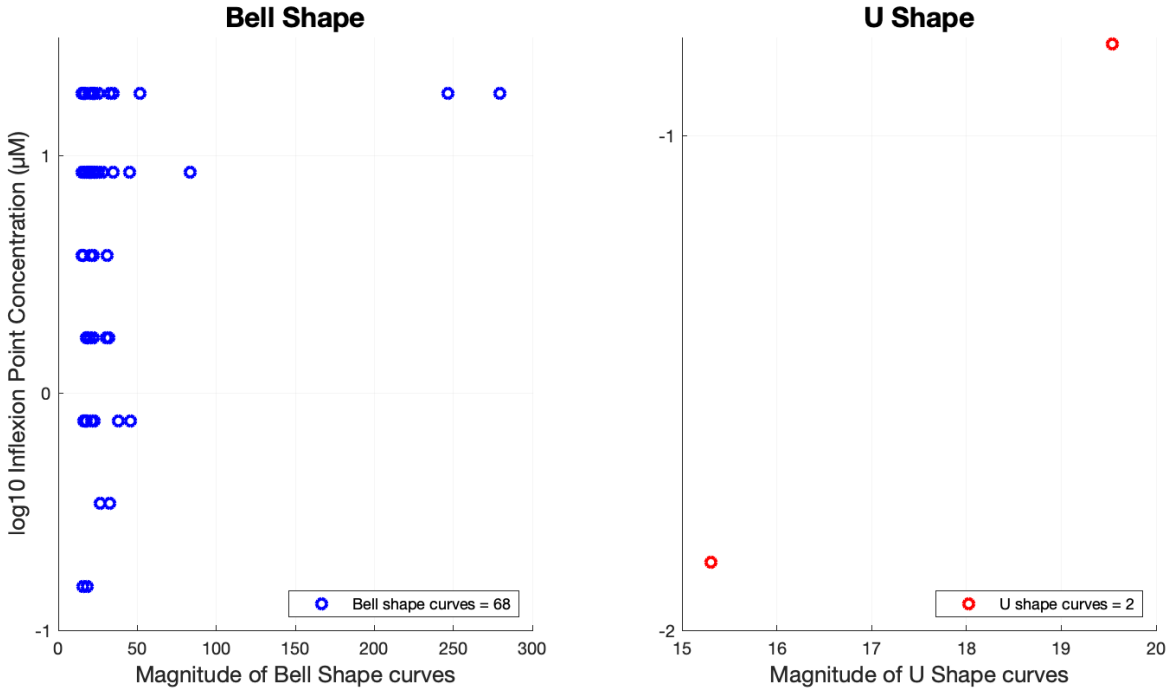

Fig 12. Scatter plots for $\log_{10}$ concentration of inflexion point vs. magnitude of U and Bell shape curves.

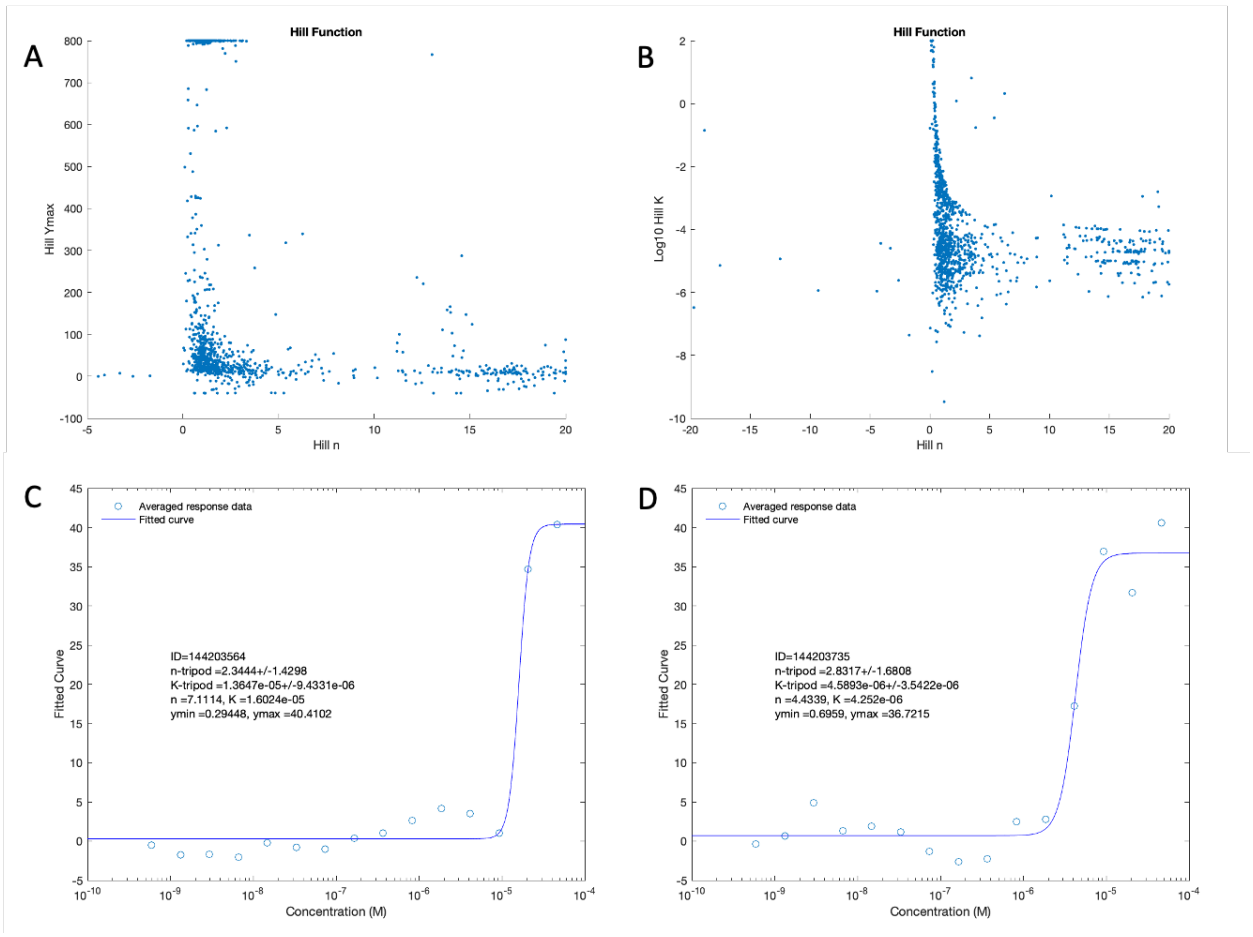

Fig 13. (A) Scatter plot of Hill coefficient (n) vs. ymax and (B) scatter plot of Hill coefficient (n) vs K (EC50) of fitted monotonic increasing curves using the Hill function. Fitted curves for Aminoquinuride dihydrochloride (C) and Lansoprazole (D). Circles represent averaged response data from multiple replicates, the lines are fitted curves.

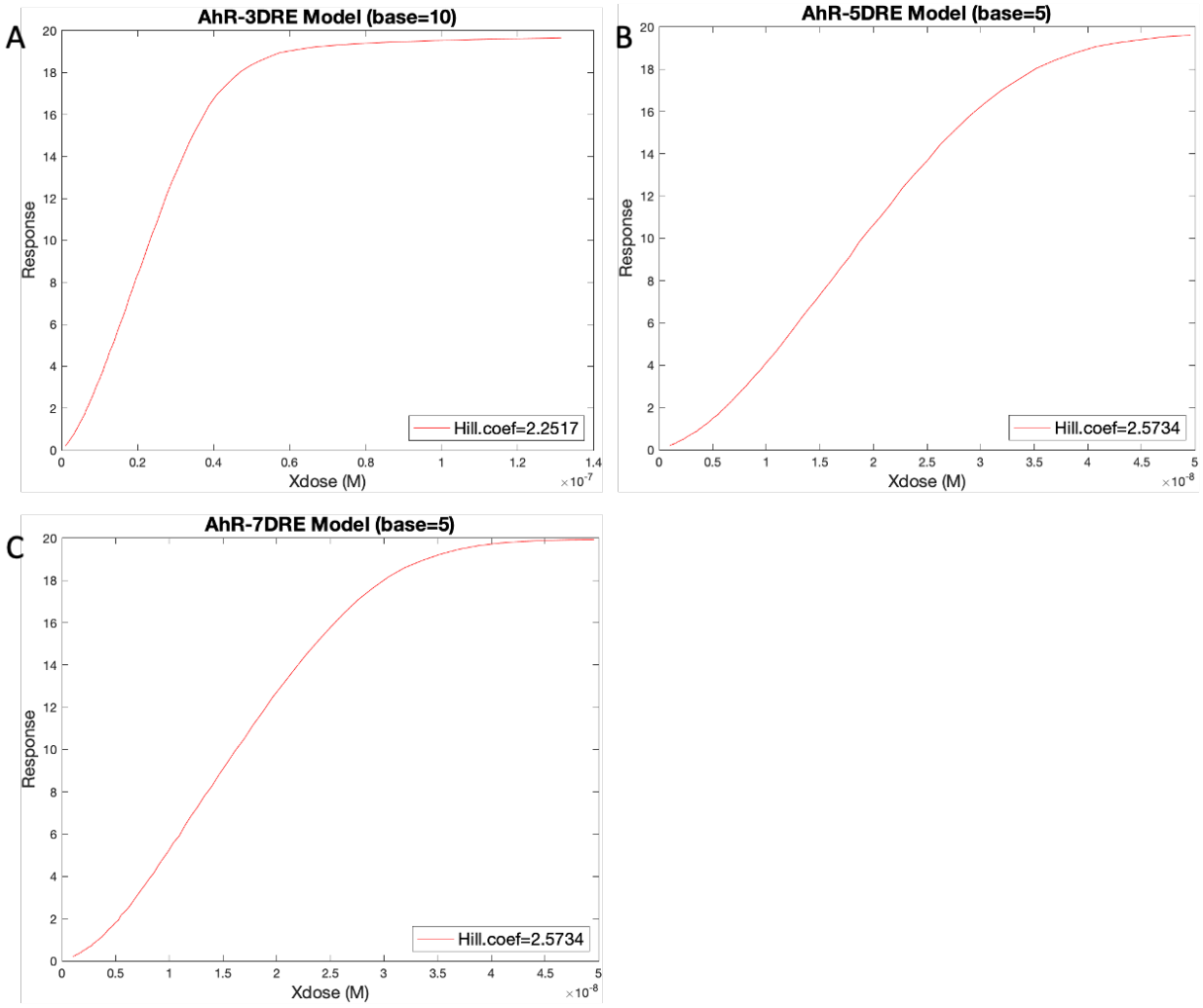

Fig 14. Simulated concentration-response curves for (A) 3 DREs with cooperative binding and fold-increase (base) for binding affinity is 10, (B) 5 DREs with cooperative binding and fold-increase (base) for binding affinity is 5, and 7 DREs with cooperative binding and fold-increase (base) for binding affinity is 5. The corresponding Hill coefficients are shown in the panels.

TABLE

Table 1. Hill Coefficient and EC50 data under various conditions of the mathematical model.

AhR-DRE Model	Base	Hill Coefficient	EC50 (M)
AhR-3DRE	1	0.96848	1.0303e-07
AhR-3DRE	2	1.2867	5.7374e-08
AhR-3DRE	5	1.9163	3.1948e-08
AhR-3DRE	10	2.2517	2.2705e-08
AhR-3DRE	20	2.3702	1.9613e-08
AhR-3DRE	50	2.3094	1.6136e-08
AhR-3DRE	100	2.2517	1.5367e-08
AhR-5DRE	1	0.96848	1.0303e-07
AhR-5DRE	2	1.766	3.6984e-08
AhR-5DRE	5	2.5734	1.8679e-08
AhR-5DRE	10	2.4343	1.6136e-08
AhR-5DRE	20	2.3094	1.5367e-08
AhR-5DRE	50	2.2517	1.4636e-08
AhR-5DRE	100	2.1968	1.4636e-08

AhR-7DRE	1	0.96848	1.0303e-07
AhR-7DRE	2	2.3094	2.6283e-08
AhR-7DRE	5	2.5734	1.6136e-08
AhR-7DRE	10	2.4343	1.5367e-08
AhR-7DRE	20	2.3094	1.4636e-08
AhR-7DRE	50	2.1968	1.4636e-08
AhR-7DRE	100	2.1445	1.4636e-08

Note: Base denotes fold-increase in affinity for sequential DRE binding.

REFERENCES

1. Rohleder, F., *Mechanism of the toxicity of 2,3,7,8-TCDD.* Derm Beruf Umwelt, 1990. **38**(3): p. 94-5.
2. Mimura, J. and Y. Fujii-Kuriyama, *Functional role of AhR in the expression of toxic effects by TCDD.* Biochim Biophys Acta, 2003. **1619**(3): p. 263-8.
3. VanEtten, S.L., et al., *Telomeres as targets for the toxicity of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) and polychlorinated biphenyls (PCBs) in rats.* Toxicol Appl Pharmacol, 2020. **408**: p. 115264.
4. Xu, S., et al., *Occurrence, Human Exposure, and Risk Assessment of Polybrominated Dibenzo-p-Dioxins and Dibenzofurans, Polychlorinated Naphthalenes, and Metals in Atmosphere Around Industrial Parks in Jiangsu, China.* Bull Environ Contam Toxicol, 2021. **106**(4): p. 683-689.
5. Dearstyne, E.A. and N.I. Kerkvliet, *Mechanism of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD)-induced decrease in anti-CD3-activated CD4(+) T cells: the roles of apoptosis, Fas, and TNF.* Toxicology, 2002. **170**(1-2): p. 139-51.
6. Larigot, L., et al., *AhR signaling pathways and regulatory functions.* Biochim Open, 2018. **7**: p. 1-9.
7. Zhu, K., et al., *Aryl hydrocarbon receptor pathway: Role, regulation and intervention in atherosclerosis therapy (Review).* Mol Med Rep, 2019. **20**(6): p. 4763-4773.
8. Bock, K.W., *From TCDD-mediated toxicity to searches of physiologic AHR functions.* Biochem Pharmacol, 2018. **155**: p. 419-424.
9. Thomas, R.S., et al., *The US Federal Tox21 Program: A strategic and operational plan for continued leadership.* ALTEX, 2018. **35**(2): p. 163-168.
10. Attene-Ramos, M.S., et al., *The Tox21 robotic platform for the assessment of environmental chemicals--from vision to reality.* Drug Discov Today, 2013. **18**(15-16): p. 716-23.
11. Hsieh, J.H., et al., *A Data Analysis Pipeline Accounting for Artifacts in Tox21 Quantitative High-Throughput Screening Assays.* J Biomol Screen, 2015. **20**(7): p. 887-97.
12. Hoffman, T.E., et al., *Ultrasensitivity dynamics of diverse aryl hydrocarbon receptor modulators in a hepatoma cell line.* Arch Toxicol, 2019. **93**(3): p. 635-647.
13. Simon, T.W., R.A. Budinsky, and J.C. Rowlands, *A model for aryl hydrocarbon receptor-activated gene expression shows potency and efficacy changes and predicts squelching due to competition for transcription co-activators.* PLoS One, 2015. **10**(6): p. e0127952.
14. Brennan, J.C., et al., *Development of Species-Specific Ah Receptor-Responsive Third Generation CALUX Cell Lines with Enhanced Responsiveness and Improved Detection Limits.* Environ Sci Technol, 2015. **49**(19): p. 11903-12.
15. Bars, R.G. and C.R. Elcombe, *Dose-dependent acinar induction of cytochromes P450 in rat liver. Evidence for a differential mechanism of induction of P450IA1 by beta-naphthoflavone and dioxin.* Biochem J, 1991. **277 (Pt 2)**: p. 577-80.
16. Tritscher, A.M., et al., *Dose-response relationships for chronic exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin in a rat tumor promotion model: quantification and*

immunolocalization of CYP1A1 and CYP1A2 in the liver. Cancer Res, 1992. **52**(12): p. 3436-42.
17. French, C.T., et al., *Induction of CYP1A1 in primary rat hepatocytes by 3,3',4,4',5-pentachlorobiphenyl: evidence for a switch circuit element.* Toxicol Sci, 2004. **78**(2): p. 276-86.
18. Bars, R.G., et al., *Induction of cytochrome P-450 in cultured rat hepatocytes. The heterogeneous localization of specific isoenzymes using immunocytochemistry.* Biochem J, 1989. **262**(1): p. 151-8.
19. Kohn, M.C. and C.J. Portier, *Effects of the mechanism of receptor-mediated gene expression on the shape of the dose-response curve.* Risk Anal, 1993. **13**(5): p. 565-72.
20. Li, L., et al., *Non-monotonic dose-response relationship in steroid hormone receptor-mediated gene expression.* J Mol Endocrinol, 2007. **38**(5): p. 569-85.
21. Zhang, Q., S. Bhattacharya, and M.E. Andersen, *Ultrasensitive response motifs: basic amplifiers in molecular signalling networks.* Open Biol, 2013. **3**(4): p. 130031.