Using ToxCast/Tox21 Assays and QSAR Modeling to Predict Androgen Receptor Pathway Activity

N Kleinstreuer[1], Q Zang[1], D Filer[2], M Martin[2], K Houck[2], D Allen[1], W Casey[3], R Judson[2]

[1]ILS/NICEATM, RTP, NC, USA; [2]EPA/ORD/NCCT, RTP, NC, USA; [3]NIH/NIEHS/DNTP/NICEATM, RTP, NC, USA

Abstract

The Tox21 and ToxCast programs include *in vitro* assays conducted in a high-throughput screening (HTS) format. Many are relevant to the androgen receptor (AR) pathway and can identify substances with potential *androgenicant*-androgenic activity *in vivo*. Here we used nine of these assays to build a mathematical model to distinguish true AR pathway activity from technology-specific assay interference. The assay battery probed perturbations of the AR pathway at multiple points (receptor binding, cofactor recruitment, gene transcription and protein production) in multiple cell types. We compiled a list of putative AR reference chemicals from the ICCVAM and OECD reference chemical lists. Chemicals included agonists, antagonists, selective androgen receptor modulators (SARMs), and inactive chemicals. The model showed 96% (22/23) concordance with reference data, including successfully identifying multiple SARMs with both agonist and antagonist activity. The model identified as agonists or antagonists all chemicals in the ToxCast library known to specifically target AR, as well as chemicals such as prochloraz with known anti-androgenic activity *in vivo*. However, fluoranthene, a putative SARM, was active in the cofactor recruitment assays but none of the other AR pathway assays, and was therefore misspecified by the model as acting via an assay-specific interference pathway. We discuss patterns of assay activity and pathway predictions across 1846 ToxCast chemicals and identify those predicted to be active against the AR pathway. The results from the AR pathway model were used to train and build a cross-validated quantitative structure–activity relationship (QSAR) model for AR binding and used to make predictions for 30,000 chemicals. Where available, we compared *in vitro* and *in silico* predictions to toxicity data from the literature to identify potential trends relating to use case and exposure scenarios. *(Data in poster abstract have been updated to reflect the most recent analyses.)*

Introduction

- U.S. (7 U.S.C. 136, 110 Stat 1613) and international regulations require the testing of certain chemicals for the detection of potential endocrine activity (estrogen, androgen, steroidogenesis, and thyroid pathways).
- As many as 30,000 chemicals may lack sufficient testing data for this purpose, with several hundred new chemicals being added each year (EPA 2011).
- The U.S. Environmental Protection Agency (EPA) ToxCast chemical research program (Dix et al. 2007; Judson et al. 2010) and the Tox21 U.S. federal partnership (Tice et al. 2013) include multiple endocrine-related high-throughput screening (HTS) assays.
- Following an approach used to model the estrogen receptor pathway (EPA and NICEATM 2014), we have constructed a mathematical model to predict chemically induced androgen receptor (AR) activity based on nine HTS assays that map to the AR pathway.

High-Throughput Screening Data

- Data on 1846 chemicals were generated during ToxCast Phases I and II using nine AR pathway assays (**Table 1**):
  - Two cell-free biochemical radioligand AR binding assays (Novascreen: Knudsen et al. 2011; Sipes et al. 2013)
  - Two cofactor recruitment assays that measure protein:protein interaction between AR and SRC1 (Odyssey Thera: Filer et al., manuscript in preparation)
  - One transactivation assay measuring reporter gene levels (Attagene: Martin et al. 2010; Franzosa et al. manuscript in preparation)
  - Two transactivation assays measuring reporter protein level readouts (Tox21: Huang et al. manuscript in preparation)
  - Two transactivation antagonist assays (Tox21: Huang et al. manuscript in preparation)
- **Figure 1** shows how the assays map to a model of the AR pathway.
- The chemicals were tested in concentration–response format in all assays except for the cell-free binding assays. These were initially tested at a single concentration (25 µM), and if significant activity was seen, the chemical was then tested in concentration–response mode.
- All concentration–response assay data were analyzed using the ToxCast data analysis pipeline, which automates the processes of baseline correction, normalization, curve-fitting, hit-calling, and AC_{20} (half-maximal activity) determination. The pipeline also detects a variety of potential confounders, which are annotated as "caution flags". The pipeline and all raw and processed data and annotations are publicly available (http://actor.epa.gov/).

Table 1. Assays Used in the AR Pathway Model

ID	Assay Name	Source	Gene	Species	Type
A1	NVS human AR	Novascreen	AR	*Homo sapiens*	Receptor binding
A2	NVS chimpanzee AR	Novascreen	AR	*P. troglodytes*	Receptor binding
A3	OT_AR_ARSRC1_0480	Odyssey Thera	AR;SRC	*Homo sapiens*	Cofactor recruitment
A4	OT_AR_ARSRC1_0960	Odyssey Thera	AR;SRC	*Homo sapiens*	Cofactor recruitment
A5	ATG_AR_TRANS	Attagene	AR	*Homo sapiens*	RNA reporter gene
A6	Tox21_AR_BLA_Agonist_ratio	NCGC	AR	*Homo sapiens*	β-lactamase reporter gene
A7	Tox21_AR_LUC_MDAKB2_Agonist	NCGC	AR	*Homo sapiens*	Luciferase reporter gene
A8	Tox21_AR_BLA_Antagonist_ratio	NCGC	AR	*Homo sapiens*	β-lactamase reporter gene
A9	Tox21_AR_LUC_MDAKB2_Antagonist	NCGC	AR	*Homo sapiens*	Luciferase reporter gene

Acknowledgements

The Intramural Research Program of the National Institute of Environmental Health Sciences (NIEHS) supported this poster. Technical support was provided by ILS under NIEHS contract HHSN273201400003C.

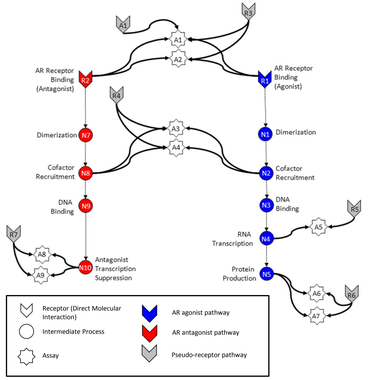
The views expressed above do not necessarily represent the official positions of any Federal agency. Since the poster was written as part of the official duties of the authors, it can be freely copied.

A summary of NICEATM activities at the 2015 SOT Annual Meeting is available on the National Toxicology Program website at http://ntp.niehs.nih.gov/go/742110.

AR Pathway Network

- **Figure 1** depicts a model of the network used to evaluate the integrated *in vitro* assay responses. The model is based on the molecular events that typically occur in a receptor-mediated response.
- The process starts with the interaction of a chemical with a nuclear AR (receptor node R1).
  - For example, an AR agonist will cause the receptors to dimerize (node N1), translocate to the nucleus and recruit cofactors to form the complete active transcription factor complex (node N2).
  - This transcription factor complex then binds to the chromatin DNA (node N3) and initiates transcription of mRNA (node N4) and subsequent translation to protein (node N5).
- Each of these processes except dimerization and DNA binding was assessed by one or more of the nine *in vitro* assays listed in **Table 1** (represented in **Figure 1** as white stars).
- **Figure 1** shows the two modes of the AR pathway: agonist (blue icons beginning with R1) and antagonist (red icons beginning with R2). The model assumes that a chemical that interacts with the AR will bind in either or both of the agonist or antagonist conformations, triggering activity in the appropriate pathway.
- Each of the *in vitro* assays is subject to processes that can lead to nonspecific activity independent of the activity of the AR pathway node that it is supposed to measure. These may be due to biological interference, artifacts, or other sources of experimental noise. These assay interference pathways are shown in **Figure 1** as alternate "pseudo-receptors" (gray arrow nodes, A1 for example).
- Examples of how a specific chemical may interact with the AR pathway are shown in **Figure 2**, in which pink highlighting represents the expected activity from a true agonist (**Figure 2a**), antagonist (**Figure 2b**), or a chemical causing assay interference (**Figure 2c**).

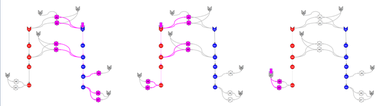
Figure 1. AR Pathway Model[a]



Abbreviation: AR = androgen receptor.

[a] Colored arrow nodes (R1/R2) represent "receptors" with which a chemical can directly interact. Colored circles represent intermediate biological processes that are not directly observable. White stars represent the assays that measure activity at the biological nodes. Arrows represent transfer of information. Grey arrow nodes represent biological interference pathways (R3–R7) or technology-specific interference (one example shown, A1).

Figure 2. Examples of Chemical Interactions with AR Pathway Model[a]



(a) R1/Agonist (b) R2/Antagonist (c) R7/Interference

Abbreviation: AR = androgen receptor.

[a] Arrows and assays highlighted pink represent the expected AR pathway model activity from a chemical that is (a) a true AR agonist, (b) a true AR antagonist, or (c) acting through an interference pathway.

Mathematical Model

- We developed a simple linear additive model to predict the relative androgenic or anti-androgenic activity of a test chemical, using data from the assays that map to the AR pathway in **Figure 1**.
- The model assumes that the value (the efficacy, *A*) returned by an assay at a given concentration is a linear sum of the contributions from the receptors that it measures:

$$A_i = \sum_j F_{ij} R_j$$

- The model then seeks a set of R_j values that minimize the difference between the predicted assay values (A_i^{pred}) and the measured ones (A_i^{meas}) for each chemical–concentration pair. A constrained least-squares minimization approach is used, where the function being minimized is:

$$\varepsilon^2 = \sum_i \left(A_i^{pred} - A_i^{meas}\right)^2 + penalty(R)$$

- The term *penalty(R)* penalizes solutions that predict that many receptors are being simultaneously activated by the chemical:

$$penalty(R) = \alpha \frac{(sR^2)}{(tRT + sR_0^2)}$$

In this equation, *SR* is the sum of R values at that concentration, SR_0 is a threshold value, and α is a small number between 0 and 1. This penalty term helps stabilize the solutions and is based on the assumption that it is unlikely that most chemicals will strongly and specifically interact with many dissimilar molecular targets.

- The model produces a response value (between 0 and 1) for each receptor at each concentration. These results are summarized as the integral across the concentration range, expressed as area under the curve (AUC):

$$AUC_j = \frac{1}{N_{conc}} \sum_{i=1}^{N_{conc}} sign(slope) \times R_j(conc_i)$$

- Because the biological response of greatest environmental concern is AR pathway antagonism, the AUC scores are normalized to yield a value of 1 for the antagonist positive control (OECD 2010).

Cytotoxicity Filter

- We developed a scheme to filter out nonselective assay hits attributed to cytotoxicity using the distance between the logACA50(assay) and the median logACA50(cytotox), with respect to the global cytotoxicity MAD (the median of the median absolute deviation [MAD] of the logACA50(cytotox) distributions across all chemicals).
- For chemicals with two or more positive responses in cytotoxicity assays, a "Z-score" was calculated for each AR pathway assay hit as:

$$Z(chemical, assay) = \frac{\log AC50(chemical, assay) - median\left[\log AC50(chemical, cytotox)\right]}{global\,cytotoxicity\,MAD}$$

- A large Z-value will occur at concentrations significantly below those causing cytotoxicity. Thus, a hit associated with this Z-value is unlikely to be caused by either cell-stress or cytotoxicity-related processes and is more likely to be associated with a target-selective mechanism.
- However, in instances where a loss of signal is observed, it is still difficult to distinguish antagonism from cytotoxicity. This is of particular concern because many environmental chemicals exhibit antagonist activity. Thus, efforts to improve cytotoxicity filtering are continuing.

Evaluation of Model Performance

- A set of 23 reference chemicals was used to evaluate model performance (**Table 2**). These were chosen based on consistent *in vitro* results in reports from ICCVAM (ICCVAM 2003) and OECD (OECD 2010) and their inclusion in the ToxCast chemical library.
- **Figure 3** summarizes the performance of the model in predicting reference chemical activity.

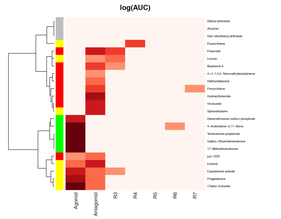
Table 2. Reference Chemicals

Chemical Name	CASRN	Activity
Dexamethasone	2392-39-4	Agonist
4-Androstenedione	63-05-8	Agonist
5a-Dihydrotestosterone	521-18-6	Agonist
Methyl testosterone	58-18-4	Agonist
Testosterone propionate	57-85-2	Agonist
Flutamide	13311-84-7	Antagonist
4-tert-Octylphenol	140-66-9	Antagonist
Procymidone	32809-16-8	Antagonist
Bisphenol A	80-05-7	Antagonist
Vinclozolin	50471-44-8	Antagonist
p,p'-DDE	72-55-9	Antagonist
Hydroxyflutamide	52806-53-8	Antagonist
Diethylstilbestrol	56-53-1	Antagonist
Di-n-butyl-phthalate	84-74-2	Inactive
Diethylhexyl phthalate	117-81-7	Inactive
Atrazine	1912-24-9	Inactive
Cyproterone acetate	427-51-0	SARM
17-β-Estradiol	50-28-2	SARM
Estrone	53-16-7	SARM
Linuron	330-55-2	SARM
Spironolactone	52-01-7	SARM
Progesterone	57-83-0	SARM
Fluoranthene	206-44-0	SARM

Abbreviations: CASRN = Chemical Abstracts Service Registry Number; SARM = selective androgen receptor modulator, which has both agonist and antagonist activity.

Evaluation of Model Performance (cont'd)

Figure 3. AR Pathway Receptor AUC Values for Reference Chemicals[a]



Abbreviations: AR = androgen receptor; AUC = area under the curve.

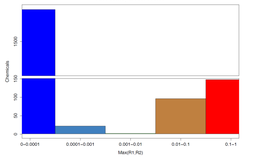
[a] Labels along the bottom of the figure refer to receptor pathways in **Figure 1**, with the agonist and antagonist pathways being R1 and R2 respectively. Reference chemicals are color-coded on the sidebar as agonist (green), antagonist (red), selective androgen receptor modulator (both agonist and antagonist activity, yellow) or inactive (grey). Red colored areas represent AUC scores of at least 0.05; darker color indicates larger AUC values.

- Overall, the model showed 96% (22/23) concordance in identifying agonist or antagonist or AR activity across the reference set, using a threshold of 0.05 as a positive AUC score.
  - The three inactive reference chemicals (atrazine, dibutyl phthalate, and diethylhexyl phthalate) were identified by the model as being inactive.
  - All five agonist reference chemicals produced a high agonist (R1) AUC score, and were correctly predicted to act via true AR agonism. One agonist chemical, androstenedione, showed potential assay interference via the R6 receptor pathway but the model score was much lower than for agonist activity.
  - Of the eight antagonist reference chemicals, all were identified as true antagonists with high antagonist (R2) AUC scores.
  - Three antagonist reference chemicals–flutamide, bisphenol A, and procymidone–were also predicted to potentially act via assay interference pathways R3 and R7, but the scores were all lower than for antagonist activity.
- The model correctly identified multiple selective androgen receptor modulators (SARMs) that have both agonist and antagonist activity.
  - Four SARMs were correctly predicted to have both agonist and antagonist activity by the model, while two SARMs (spironolactone and linuron) were only identified as antagonists.
  - Fluoranthene, also a SARM, was active in the cofactor recruitment assays but none of the other AR pathway assays, and was therefore incorrectly characterized by the model as acting via an assay-specific interference pathway (R4).
- Examples of assay concentration–response plots and model AUC predictions are shown in **Figure 4** for testosterone propionate (reference agonist), vinclozolin (reference antagonist), cyproterone acetate (reference SARM), and prochloraz (pesticide with known AR antagonist behavior *in vivo*) (Wilson et al. 2008).

Activity in the AR Pathway Model across the ToxCast Library

- **Figure 5** shows the distribution of AR model pathway scores across the ToxCast chemical library.
  - The figure includes the maximum agonist or antagonist AUC score for each chemical.
  - Of the 1846 chemicals tested, 1587 were completely inactive in the model, with both R1 and R2 scores below 0.0001, while 148 chemicals were predicted to be either androgen agonists or antagonists (R1 or R2 > 0.1). The remaining 120 chemicals had model scores in the intermediate region.
- **Figure 6** is a calibration curve to aid interpretation of the AUC distributions, showing that an AUC of 0.1 is equivalent to half-maximal activity against the AR pathway at ~100 µM.

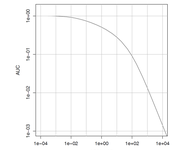
Figure 5. AR Pathway Model Scores for 1855 ToxCast Chemicals[a]



Abbreviation: AR = androgen receptor.

[a] The histogram shows AR pathway model scores, using the maximum R1 (agonist) or R2 (antagonist) value and without applying the cytotoxicity filter, across the 1855 chemicals in the ToxCast library.

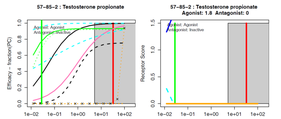
Figure 6. Calibration Curve for AR Pathway Model Scores



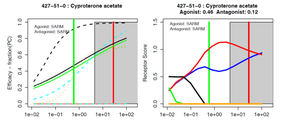
Abbreviations: AC50 = half-maximal activity concentration; AR = androgen receptor; AUC = area under the curve.

Figure 4. Examples of Chemical Activity in Assays and Receptor AUC Values from the AR Pathway Model

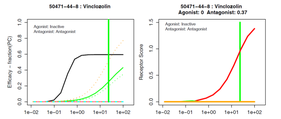
Testosterone Propionate (Reference AR Agonist)



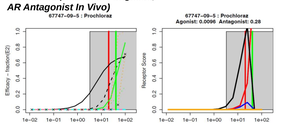
Cyproterone Acetate (Reference SARM)



Vinclozolin (Reference AR Antagonist)



Prochloraz (Conazole Fungicide, AR Antagonist In Vivo)



Abbreviations: AR = androgen receptor; SARM = selective androgen receptor modulator.

Left plots show assay responses and right plots show model predictions, corresponding to legend. The vertical green line represents the median AC50 (half maximal activity concentration) across the AR pathway assays. The gray box represents the region of potential cytotoxicity, and the vertical red line represents the median cytotoxicity AC50, where applicable.

QSAR Predictions of AR Pathway Activity

- Quantitative structure–activity relationship (QSAR) models are needed to make predictions for chemicals that do not have *in vitro* assay data available.
- We used the QikProp library of molecular descriptors and three machine learning approaches to build cross-validated QSAR models to predict AR pathway activity.
  - Support vector machine (SVM)
  - Linear discriminant analysis
  - Classification and regression tree
- The best performing model was SVM. The five-fold cross-validated statistics are shown in **Table 3**.

Table 3. Cross-Validated QSAR SVM Model Performance

Model Run	Sensitivity	Specificity	BA
1	0.71	0.76	0.74
2	0.78	0.74	0.76
3	0.71	0.80	0.75
4	0.69	0.80	0.74
5	0.69	0.77	0.73
Average	**0.72**	**0.77**	**0.74**

Abbreviations: BA = balanced accuracy; QSAR = quantitative structure–activity relationship; SVM = support vector machine.

- This model was retrained on the results from the entire ToxCast library and used to make predictions for 30,000 chemicals in the broader chemical universe. Based on this preliminary model, 20.60% (6475/31428) of these chemicals were predicted to have antagonist activity against the AR pathway. However, this initial result is very likely an overestimate that may be confounded by very weak activity or cytotoxicity; further refinement of the model should improve specificity.

Conclusions

- The AR pathway model performed well at predicting AR pathway activities of the reference chemical set, including identifying SARMs with both agonist and antagonist activities. Further, all 15 chemicals in the ToxCast library known to interact specifically with the AR were identified by the model as either agonists or antagonists with R1 or R2 > 0.1, and environmental chemicals with *in vivo* evidence of AR pathway perturbation, such as prochloraz, were also identified by the model.
- The majority of the 1846 ToxCast chemicals tested in the AR assays were not predicted by the model to have any androgenic or anti-androgenic activity. Certain environmental chemicals such as antimicrobials (e.g., triclosan and triclocarban) and plasticizers (e.g. bisphenol A and bisphenol AF) were predicted to be AR antagonists, but these results were confounded by cytotoxicity, and more targeted testing within the relevant concentration ranges may be required.
- The AR pathway model provides a biologically based mathematical approach to distinguish assay interference from true agonist or antagonist activity and to prioritize large numbers of environmental chemicals for their potential androgenic or anti-androgenic activity.
- QSAR models trained on the AR pathway model will be further developed and consensus models built to make predictions for 30,000 chemicals in the environment.

References

Dix DJ et al. 2007. Toxicol Sci. 2007 Jan;95(1):5-12.

EPA. 2011. The Incorporation of *in Silico* Models and *in Vitro* High Throughput Assays in the Endocrine Disruptor Screening Program (EDSP) for Prioritization and Screening. Summary Overview. A Part of the EDSP Comprehensive Management Plan [Internet]. Washington, DC:U.S. Environmental Protection Agency. Available: http://epa.gov/endo/pubs/edspp21_work_plan_summary%20_overview_final.pdf

EPA, NICEATM. 2014. Integrated Bioactivity and Exposure Ranking: A Computational Approach for the Prioritization and Screening of Chemicals in the Endocrine Disruptor Screening Program. Presentation to the December 2014 meeting of the Scientific Advisory Panel for the EPA Federal Insecticide, Fungicide and Rodenticide Act [Internet]. Washington, DC:U.S. Environmental Protection Agency. Available at http://www.regulations.gov/contentStreamer?objectId=09000064819119e&disposition=attachment&contentType=pdf.

Food Quality Protection Act of 1996. 7 U.S.C. 136. Public Law 104-170.

ICCVAM. 2003. ICCVAM Evaluation of In Vitro Test Methods for Detecting Potential Endocrine Disruptors: Estrogen Receptor and Androgen Receptor Binding and Transcriptional Activation Assays (NIH Publication No. 03-4503). Research Triangle Park, NC:National Institute of Environmental Health Sciences.

Judson RS et al. 2010. Environ Health Perspect 118(4): 485–492.

Knudsen TB et al. 2011. Toxicology 282(1–2):1–15.

Martin MT et al. 2010. Chem Res Toxicol 23(3):578–90.

OECD. 2010. Draft Report of Pre-validation and Inter-laboratory Validation for Androgen Receptor (AR) Mediated Stably Transfected Transcriptional Activation (AR-STTA) Assay to Detect Androgenic and Anti-androgenic Activities. Paris: OECD Publishing. Available: http://www.oecd.org/chemicalsafety/testing/46755800.pdf.

Safe Drinking Water Act Amendments of 1996. 110 Stat 1613. Public Law 104–182.

Sipes NS et al. 2013. Chem Res Toxicol 26(6):878–895.

Tice RR et al. 2013. Environ Health Perspect 121: 756–765.

Wilson VS et al. 2008. Int J Androl 31(2): 178–187