



OPEN ACCESS

*CORRESPONDENCE

Huixiao Hong,

✉ huixiao.hong@fda.hhs.gov

Barry Hardy,

✉ barry.hardy@edelweissconnect.com

RECEIVED 07 June 2024

ACCEPTED 27 August 2024

PUBLISHED 19 September 2024

CITATION

Dong F, Hardy B, Liu J, Mohoric T,
Guo W, Exner T, Tong W, Dohler J,
Bachler D and Hong H (2024)Development of a comprehensive open
access “molecules with androgenic
activity resource (MAAR)” to facilitate
risk assessment of chemicals.*Exp. Biol. Med.* 249:10279.

doi: 10.3389/ebm.2024.10279

COPYRIGHT

© 2024 Dong, Hardy, Liu, Mohoric, Guo,
Exner, Tong, Dohler, Bachler and Hong.

This is an open-access article

distributed under the terms of the

[Creative Commons Attribution License](#)
(CC BY). The use, distribution or

reproduction in other forums is

permitted, provided the original

author(s) and the copyright owner(s) are

credited and that the original

publication in this journal is cited, in

accordance with accepted academic

practice. No use, distribution or

reproduction is permitted which does

not comply with these terms.

Development of a comprehensive open access “molecules with androgenic activity resource (MAAR)” to facilitate risk assessment of chemicals

Fan Dong¹, Barry Hardy^{2*}, Jie Liu¹, Tomaz Mohoric²,
Wenjing Guo¹, Thomas Exner², Weida Tong¹, Joh Dohler²,
Daniel Bachler² and Huixiao Hong^{1*}¹National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR,
United States, ²Edelweiss Connect Inc., Durham, NC, United States

Abstract

The increasing prevalence of endocrine-disrupting chemicals (EDCs) and their potential adverse effects on human health underscore the necessity for robust tools to assess and manage associated risks. The androgen receptor (AR) is a critical component of the endocrine system, playing a pivotal role in mediating the biological effects of androgens, which are male sex hormones. Exposure to androgen-disrupting chemicals during critical periods of development, such as fetal development or puberty, may result in adverse effects on reproductive health, including altered sexual differentiation, impaired fertility, and an increased risk of reproductive disorders. Therefore, androgenic activity data is critical for chemical risk assessment. A large amount of androgenic data has been generated using various experimental protocols. Moreover, the data are reported in different formats and in diverse sources. To facilitate utilization of androgenic activity data in chemical risk assessment, the Molecules with Androgenic Activity Resource (MAAR) was developed. MAAR is the first open-access platform designed to streamline and enhance the risk assessment of chemicals with androgenic activity. MAAR's development involved the integration of diverse data sources, including data from public databases and mining literature, to establish a reliable and versatile repository. The platform employs a user-friendly interface, enabling efficient navigation and extraction of pertinent information. MAAR is poised to advance chemical risk assessment by offering unprecedented access to information crucial for evaluating the androgenic potential of a wide array of chemicals.

The open-access nature of MAAR promotes transparency and collaboration, fostering a collective effort to address the challenges posed by androgenic EDCs.

KEYWORDS

androgen receptor, risk assessment, chemicals, database, open access

Impact statement

The prevalence of endocrine-disrupting chemicals (EDCs) and their potential health impacts necessitate robust tools for risk assessment. The androgen receptor is crucial in mediating the effects of male sex hormones, with disruption during critical developmental periods leading to reproductive health issues. To address this, the Molecules with Androgenic Activity Resource (MAAR) was developed. MAAR integrates diverse data sources to create an open-access platform facilitating chemical risk assessment. By offering easy navigation and extraction of androgenic activity data, MAAR enhances transparency and collaboration in addressing the challenges posed by androgenic EDCs.

Introduction

Endocrine-active chemicals are exogenous compounds that affect the endocrine system of humans and other vertebrates. Endocrine activity of chemicals has the potential to cause numerous adverse outcomes, including disrupting physiological function of endogenous hormones and altering homeostasis [1, 2]. Evidence that certain man-made chemicals can disrupt the endocrine system by mimicking endogenous hormones sparked intense international scientific discussion and debate starting some 24 years ago [3]. These discussions culminated in issuance of legislation that reauthorized the Safe Drinking Water Act¹ and authorization of the 1996 Food Quality Protection Act mandating that the US Environmental Protection Agency (EPA) develop a program for screening and testing chemicals with endocrine disrupting potential². In 2015, the US Food and Drug Administration (FDA) published guidance to provide recommendations to sponsors of investigational new drug applications, new drug applications, and biologics license applications regulated by the FDA's Center for Drug Evaluation and Research (CDER) regarding nonclinical studies intended to identify the potential for a drug to cause endocrine-related toxicity³. FDA's National Center for Toxicological Research (NCTR) developed a program to meet the need for information

systems focused on aggregating knowledge of chemicals with experimental results relevant to endocrine activity. These efforts resulted in the development of the endocrine disruptors knowledge base (EDKB) [4] and estrogenic activity database (EADB) [5], which have been used to help identify endocrine active chemicals, develop predictive toxicology models, and prioritize chemicals for laborious and expensive testing [6–22]. However, as of today, androgenic activity data have not been comprehensively curated into a database.

Androgen receptors (ARs) are ligand-dependent transcription factors that belong to the nuclear receptor superfamily [23]. ARs are widely expressed in various tissues within the body [24]. They are the targets for drugs to treat hormone-related diseases including cancers of prostate, breast, ovary, pancreas, etc. [25]. On the other hand, chemicals can interfere with the endocrine system by interacting with ARs, which result in adverse effects [26]. Therefore, estimation of the androgenic activity of drugs and other chemicals is critical for the evaluation of drug safety and assessment of chemical risk.

Over the past decades, large numbers of chemicals have been assayed for androgenic activity by government agencies, industry, and academic research groups, with the results of these studies reported in the public domain. However, the data are distributed across different and diverse sources, obtained in multiple diverse assays, and stored in different formats, limiting the use of the data in research and regulation. Therefore, a comprehensive and reliable resource to provide open access to the data and enable modeling and prediction of androgenic activity for untested chemicals could facilitate advancement in developing strategies to mitigate the AR-driven toxicity and risk. To enable and optimize the use of the data generated by these studies, we have developed and are maintaining a comprehensive open access resource called Molecules with Androgenic Activity Resource (MAAR) to provide the scientific and regulatory communities with an up-to-date androgenic activity database for evaluating potential endocrine activity of chemicals.

Materials and methods

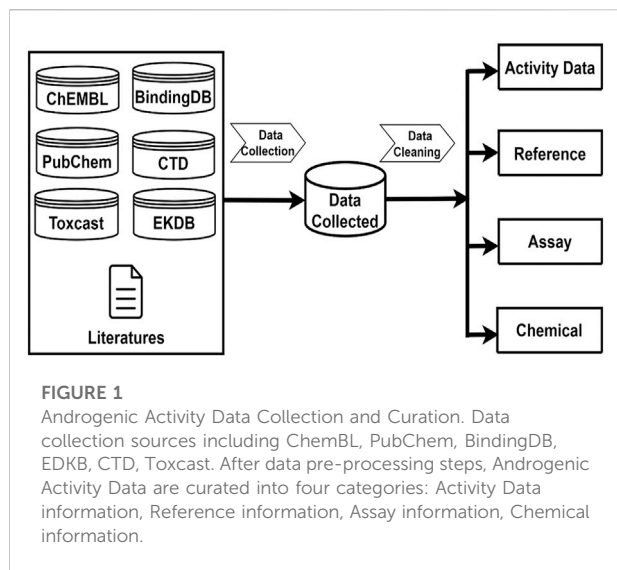
Data collection

Androgenic activity data were collected from multiple sources which encompass published literature and public

1 <http://www.epa.gov/safewater/sdwa/index.html>

2 <http://www.epa.gov/scipoly/oscpendo/>

3 <https://www.fda.gov/media/86996/download>



databases including PubChem [27], ChEMBL [28], BindingDB [29], EDKB [4], ToxCast [30], and the Comparative Toxicogenomics Database (CTD) [31]. Java programs were developed to automatically retrieve androgenic activity data points and associated data such as chemical structures, assays, species, and references from these public databases. In addition, androgenic activity data in the literature were manually searched and extracted. The collected androgenic activity data include both quantitative measurements for active compounds and qualitative descriptions for inactive chemicals. Figure 1 gives the sources from which the androgenic activity and related data were collected and the four types of data that were included in this database.

Data curation

After data were collected from individual sources, data were pre-processed and integrated before they were implemented into the database. Given potential duplications in the data collected from different sources, an automated pre-process program was devised to check and remove duplicated data records by comparing chemicals, activity data, assays, and references. This program identified and removed duplicates by comparing CID, ChEMBL ID, PubMed ID, endpoint values, and assay descriptions across data sources. Geometric and optical isomers are considered duplicates only if they have the same CID or ChEMBL ID. This program also ensured data uniformity by transforming all collected activity data into standardized units. For different activity values of a compound from different sources where inconsistencies were found, a manual review was conducted to determine the most reliable value by examining the assay details. Following a cleaning

TABLE 1 Reference data table.

Data field	Description
Reference ID	Internal ID for reference
PMID	PubMed ID for reference
Journal_Name	Reference journal name
Year	Publication year
Volume	Volume number
Issue	Issue number
First_Page	First page number
Author	Author names
Title	Publication title

TABLE 2 Assay data table.

Data field	Description
Assay ID	Internal ID for assay
Description	Description of assay
Assay_Name	Assay name
Assay_Group	Assay group, e.g., HTS, Reporter gene
Assay_Format	Assay format, e.g., cell-based, protein-based
Assay_Type	Assay type, e.g., Agonist, Antagonist
Species	Species assay based on, e.g., <i>Homo sapiens</i>
AID	Bioassay AID in PubChem
ChEMBL Assay ID	Assay ID in ChEMBL.

procedure that removed duplicates to keep unique androgenic activity data, the pre-processed data were combined to make the final data that were included in the database. This program was developed in Java. It processes text file containing all activity information, specifies columns used for comparison, and identifies both duplicate and unique activity records to ensure that non-redundant data is included in the final dataset.

Data model

The data implemented in the database were organized into four categories: androgen activity data, references, assays, and chemical information. Properties for each of the four categories are summarized in Tables 1–4. The four tables are interconnected through Chemical ID, Assay ID and Reference ID as depicted in the database schema in Figure 2.

TABLE 3 Androgen activity data table.

Data field	Description
Activity ID	Internal ID for activity data
Chemical ID	Internal ID for chemicals
Assay ID	Internal ID for assays
Reference ID	Internal ID for references
Endpoint	Activity endpoint, e.g., IC50, AC50, LogRBA
Relation	Relation to describe activity value, e.g., >, =, <
Value	Activity value from endpoint measurement
Units	Activity data unit, e.g., nM, %
Download	Database where data downloaded, e.g., PubChem, ChEMBL
Curation/Data source	Date source, e.g., literatures, US Patent, Tox21, Abbott Labs

TABLE 4 Chemical data table.

Data field	Description
Chemical ID	Internal chemical ID of chemical
IUPAC_NAME	IUPAC name of chemical
Name	Chemical name used in the system
Synonyms	Chemical synonyms (a string separated by " ")
SMILES	SMILES string of chemical
InChIKey	A fixed-length format directly derived from InChI
InChI	International Chemical Identifier
Molecular_weight	Molecular weight
Molecular_formula	Molecular formula
CAS	Chemical CAS registry number
CID	Compound ID in PubChem
CHEMBL_ID	Compound ID in ChEMBL
BindingDB_ID	Compound ID in BindingDB

Database design

The curated tables were put into a cloud-based database based on EdelweissData that was developed by Edelweiss Connect, GmbH to tackle data management issues in life sciences. Some of the advantages offered by the EdelweissData solution are:

- Each published dataset is assigned a unique URL and is easily accessible through a web browser.
- Published datasets are automatically versioned.

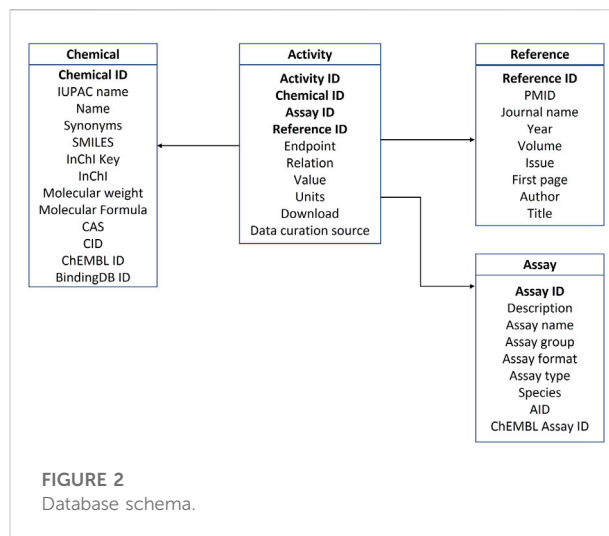


FIGURE 2
Database schema.

- Published datasets are static, i.e., they cannot be changed unless a newer version is published.
- Published datasets are immediately available through a web service and can be consumed by numerous data analysis tools (Python, R, Excel, KNIME, etc.) via REST API (see also section Data model).
- Flexibility - there is no predefined schema for published datasets. Instead, the schema is inferred from the data during publishing. This allows for a quick and easy consumption of datasets with various structures.

Database implementation - EdelweissData

The MAAR Database is built as a simple web application with a back-end supported by EdelweissData and a front-end that lets the user easily explore the database. The most common use cases (such as search by compound or chemical similarity) are well covered by the web application as is, while more advanced queries or analyses could be made through the API (Application Programming Interface) enabled by EdelweissData.

Results

Data collected and curated

In total, 125,519 androgenic activity data points for 13,648 chemicals were collected and curated from multiple sources and included in the MAAR database. These data were obtained from 923 assays. Table 5 lists the statistics of the data collected.

The androgenic activity data are presented in two types. The first type is quantitative value and 48,273 data points are in this type. A quantitative activity value indicates the androgenic

TABLE 5 Statistics of the data collected in the androgenic activity database.

Chemicals		13,648	
Activity data	Quantitative data	48,273	
	Qualitative data	Active	723
		Inactive	71,630
		Not determined	4,893
Total		125,519	
Assays	Binding	379	
	Reporter gene	358	
	Cell proliferation	86	
	<i>In vivo</i>	60	
	HTS	24	
	Other	16	
	Total		923
Species		6	

activity is numerically determined. Another type of data is qualitative androgenic activity that is described using qualitative terms: active or positive indicates a chemical was tested using an assay and activity was observed but could not be numerically determined; inactive or negative means a chemical did not show androgenic activity in an assay; inconclusion or not determined or unspecified implies activity of a chemical in an assay was not able to be determined. There are 77,246 data that are qualitative. The data were generated using 923 assays, including 379 binding assays, 358 reporter gene assays, 86 cell proliferation assays, 60 *in vivo* assays, 24 high-throughput screening assays (HTS), and 16 other assays that could not be clearly put into any type of assays. Species used in the activity testing are also included in the database, including *Bos taurus*, Chimpanzee, Chinese hamster, *Homo sapiens*, *Mus musculus*, and *Rattus norvegicus*. Information on species for some assays could not be determined in the sources, and thus they are missed for the data generated using such assays.

A chemical could be tested in many laboratories using multiple assays. All androgenic activity data for the same chemicals were collected and presented in this database. The distribution of androgenic activity data for the same chemicals is given in Figure 3.

The same chemical is often tested by a variety of assays and has multiple data records. Of the 13,648 chemicals in the database, 2,504 have only one androgenic activity data and the remaining 11,144 have more than one data. Many chemicals have more than 10 androgenic activity data reported and are included in this database. For example, 3,481 chemicals have 11 data records, and 54 chemicals even have more than 40 data records. The androgenic activity data

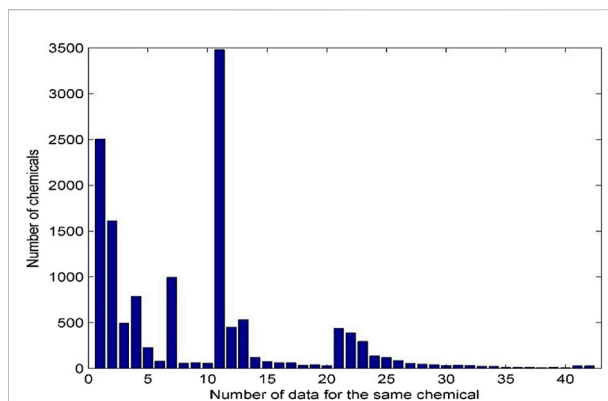


FIGURE 3

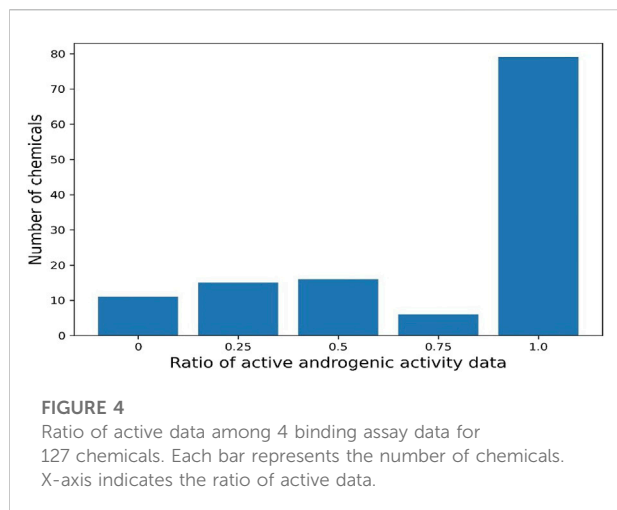
Distribution of androgenic activity data for the same chemical. Each bar represents the number of chemicals. X-axis indicates the number of data records for the same chemicals. The chemicals with 41–50 data records were grouped into the bar with x-axis value 41 and the chemicals with more than 50 data were grouped in the last bar with x-axis value 42.

obtained from the same type of assays in different laboratories could be inconsistent. For example, 127 chemicals each have four androgenic activity data generated using binding assays. As shown in Figure 4, 79 chemicals are active for all four data (100% active) and 11 chemicals consistently show inactive (0% active), while the other (37 chemicals) have inconsistent androgenic activity data: one active and three inactive for 15 chemicals, three active and one inactive for six chemicals, and two active and two inactive for 16 chemicals. This database presents all androgenic activity data reported in different sources. Assessing data quality and selecting data for specific applications such as QSAR (quantitative structure-activity relationship) modelling are critically important. Users should make decisions on how to use the data tailored to their applications.

Web resource

The MAAR database is made available through a web portal as an open science resource based on open data provided according to a Creative Commons license. We have established the resource as part of the OpenTox open knowledge infrastructure located at⁴. The main initial functionality supported allows the user to search for compounds or chemically similar compounds in the database (Figure 5). The portal also supports the location of community-generated notebooks providing additional analysis of the data, starting with illustrative examples we have provided (see sections Method and Application Programming Interface).

⁴ <https://opentox.net/MAAR/>



Application programming interface (API)

The MAAR database comes with a versatile API that simplifies the consumption of the data into other applications. Common data analysis tools that support Representational State Transfer (REST) APIs can obtain data in the database through a simple web request. In this way data can be easily transferred into a Python or R script/notebook, KNIME, Microsoft Excel, etc. To make it even easier for users, an example of an API call in Python and curl⁵ is provided in the web application and could be copy-pasted to the user's script/notebook. API documentation is available from the EdelweissData main website⁶.

For the purpose of demonstrating programmatic data retrieval from the database, we show an example of how a particular dataset could be accessed with a web request. Each assay dataset in the database has its own unique ID and when the URL pointing to that dataset is called the database returns the dataset in the JSON format. For a dataset inside the database the URL for dataset with ID "21b033c5-d048-41f5-b8a1-d5d8492f7048" would be the following:⁷. And the response from the database is shown in Figure 6.

Notebooks

The REST API service mentioned above is very well suited for different interactive notebooks that are nowadays a common tool for data analysis and visualization. There are many different notebooks available today that differ in the language,

interactivity, etc. To build an interactive notebook for visualization of the MAAR data we decided to use Observable HQ notebooks⁸ as they offer in our opinion a very good user experience even for technically less skilled users. The programming language in Observable notebooks is JavaScript, which is typically not the language of first choice for data analysis, however, it is very well suited for interactive visualizations that work as a web page.

The Observable notebook⁹ for the MAAR database is available through a URL and can be easily shared with anyone. The notebook addresses a simple use case where a user wants to search the database for a particular compound. The notebook returns a list of chemically most similar compounds (based on the Tanimoto chemical similarity – see Figure 7, left) together with their activities in the assays. In the next step, users can narrow down the set of activities by filtering the assays based on format, group, type, species, or endpoint. Finally, the subset of compounds (on x axis) and their activities (colors) in various assays (y axis) is displayed as a heatmap (Figure 7, right).

Discussion

Androgens are hormones that play a key role in the development and maintenance of male characteristics. Understanding the androgenic activity of chemicals is important for assessing chemical risk through endocrine disruption. Therefore, androgenic activity data are important for comprehensive chemical risk assessments, providing insight into the potential endocrine-disrupting effects of substances and helping to establish guidelines and regulations to protect human health and the environment.

Vast amounts of androgenic activity data have been generated and reported in the public domain for many chemicals. However, accessing and using androgenic activity data in the public domain may pose several challenges. First, androgenic activity data are contained in different and diverse sources in the public domain. The lack of comprehensive datasets can hinder applications in chemical risk assessment. Second, the importance of data quality and reliability in scientific research cannot be overstated [32, 33]. Sound scientific conclusions rest on the foundation of accurate and trustworthy data. The reliability and accuracy of available androgenic activity data vary. Incomplete or poorly curated datasets can compromise the validity of research findings. Third, androgenic activity data are sourced from various studies, experiments, or databases, leading to heterogeneity in data formats and measurement techniques. Lack of

⁵ <https://en.wikipedia.org/wiki/CURL>

⁶ <https://edelweissdata.com/docs/about>

⁷ <https://api.aadb.cloud.edelweissconnect.com/datasets/21b033c5-d048-41f5-b8a1-d5d8492f7048/versions/latest>

⁸ <https://observablehq.com/>

⁹ <https://observablehq.com/@saferworldbydesign/aadb-notebook>

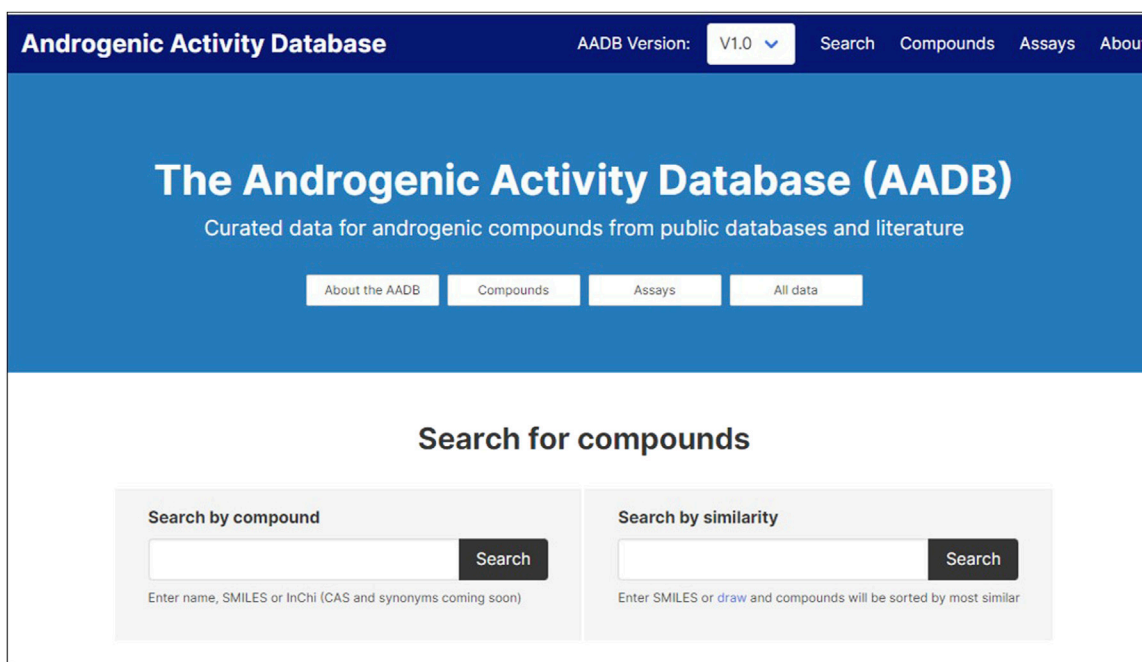


FIGURE 5
User interface of the web resource.

```
▼ id:  
  id: "21b033c5-d048-41f5-b8a1-d5d8492f7048"  
  version: 1  
  name: "FDA-AADB-V1.0_Compounds"  
  ▶ schema: {}  
  rowcount: 1  
  created: "2021-09-02T08:48:46.8446470+00:00"  
  description: ""  
  ▼ metadata:  
    Data Collection: "FDA Androgenic Activity Database (AADB)"  
    Dataset Type: "Compounds"  
    Original Filename: "FDA-AADB-V1.0_Compounds.csv"  
    Version: "V1.0"  
    isPublic: true
```

FIGURE 6
JSON response from the database.

standardized protocols for androgenic activity assessment can make it challenging to compare data from different sources. Finally, without sufficient metadata or contextual information, it may be challenging to interpret and utilize androgenic activity data accurately. Inter-laboratory and species-specific variations in androgenic responses can complicate the interpretation of androgenic activity data. Therefore, to facilitate utilization of

the available androgenic data in chemical risk assessment, we aim to develop an open resource of androgenic activity data of molecules so that the huge amount of androgenic activity data generated in the scientific community could be used to accelerate and improve chemical risk assessment.

In this article we report the development of an open science data resource for androgenic activity data. We followed

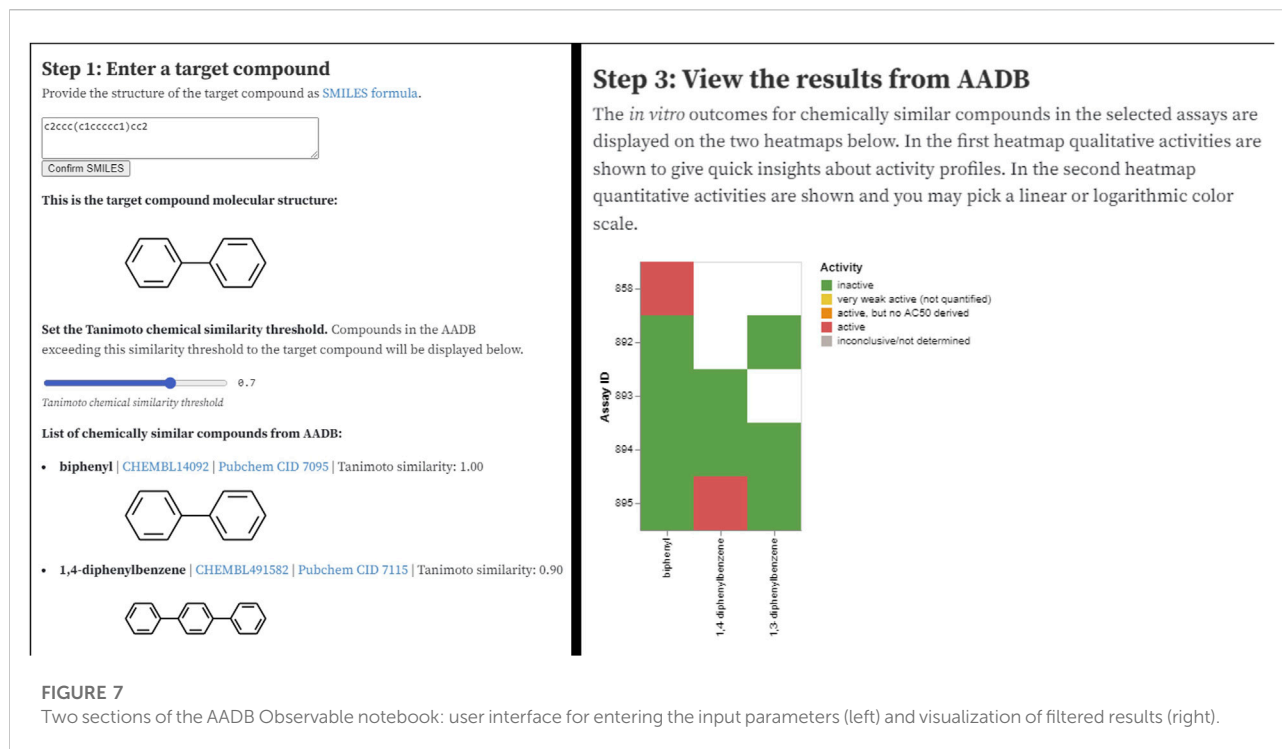
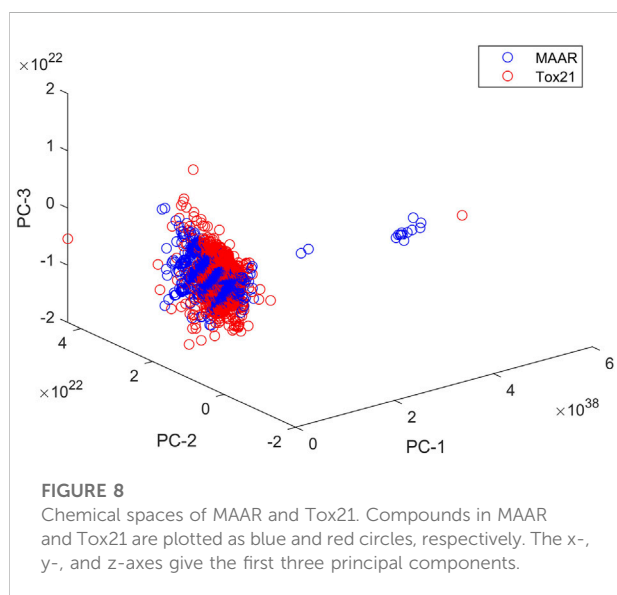


FIGURE 7

Two sections of the AADB Observable notebook: user interface for entering the input parameters (left) and visualization of filtered results (right).



principles established in previous projects including OpenTox¹⁰, OpenRiskNet¹¹ and EU-ToxRisk¹² [34–41]. The work includes the careful collection and curation of data entering the database

¹⁰ <https://opentox.net/>

¹¹ <https://openrisknet.org/>

¹² www.eu-toxrisk.eu

and a data model which includes harmonized data to structure the data in a database. Resource functionalities aligned to the FAIR (findability, accessibility, interoperability, and reusability) principles in the preparation and sharing of open science data and supporting further initiatives and use of the project knowledge. We also paid attention to data integrity principles in the construction of the database and the provision of data through harmonized application programming interfaces, supporting the building of web applications making reliable use of the data. This approach should support analysis and modelling goals of the community in making use of the open knowledge resource created by this work.

The MAAR database is an extensive compilation of chemical compounds, systematically curated and annotated for their androgenic properties, providing researchers, regulators, and industry stakeholders with a comprehensive resource for in-depth investigations. To evaluate the structural coverage of chemicals in the MAAR, we computed chemical spaces for both the MAAR and Tox21 [42] datasets using Mold2 descriptors [43, 44]. Following the methodology outlined in our previous studies [19, 45], we performed principal component analysis to represent the chemical space for each dataset. Figure 8 illustrates the first three principal components of the compounds in MAAR and Tox21, demonstrating that the structural coverage of MAAR closely resembles that of Tox21. This comparison confirms that MAAR includes structurally diverse set of compounds, making it suitable for a wide range of applications. Development of the

MAAR database represents a significant stride towards a more comprehensive and accessible approach to assessing the androgenic activity of chemicals. By providing a centralized platform for data integration and analysis, the MAAR database is poised to enhance our understanding of androgenic endocrine disruption and contribute to the development of effective risk management strategies in the face of evolving chemical landscapes.

Conclusions

We have reported here on a useful curated database for androgenic activity provided as an open science resource to the community, and available to enable searches for relevant information on the presence or absence of evidence on androgenic activity of compounds. We have also provided a model and resource with interfaces supporting additional community members to build additional analysis and modelling applications that work with the database. We hope the resource will prove useful and encourage additional development of the resource including addition of new data and its analysis.

Author contributions

Conceptualization, HH and BH; methodology, FD, TM, JL, TE, WG, TE, BH, and HH; software, TM, TE, JD, and BH; data curation, FD, JL, WG, and HH; writing–original draft preparation, FD, JL, TM, BH, and HH; writing–review and editing, WT, BH, and HH; supervision, BH and HH. All authors have read and agreed to the published version of the manuscript.

References

- Adebayo OA, Adesanoye OA, Abolaji OA, Kehinde AO, Adaramoye OA. First-line antituberculosis drugs disrupt endocrine balance and induce ovarian and uterine oxidative stress in rats. *J Basic Clin Physiol Pharmacol* (2018) **29**(2): 131–40. doi:10.1515/jbcp-2017-0087
- Danzo BJ. Environmental xenobiotics may disrupt normal endocrine function by interfering with the binding of physiological ligands to steroid receptors and binding proteins. *Environ Health Perspect* (1997) **105**(3):294–301. doi:10.1289/ehp.97105294
- Kavlock RJ, Daston GP, DeRosa C, Fenner-Crisp P, Gray LE, Kaattari S, et al. Research needs for the risk assessment of health and environmental effects of endocrine disruptors: a report of the U.S. EPA-sponsored workshop. *Environ Health Perspect* (1996) **104**(Suppl. 4):715–40. doi:10.2307/3432708
- Ding D, Xu L, Fang H, Hong H, Perkins R, Harris S, et al. The EDKB: an established knowledge base for endocrine disrupting chemicals. *BMC Bioinformatics* (2010) **11**(Suppl. 6):S5. doi:10.1186/1471-2105-11-s6-s5
- Shen J, Xu L, Fang H, Richard AM, Bray JD, Judson RS, et al. EADB: an estrogenic activity database for assessing potential endocrine activity. *Toxicol Sci* (2013) **135**(2):277–91. doi:10.1093/toxsci/kft164
- Hong H, Tong W, Fang H, Shi LM, Xie Q, Wu J, et al. Prediction of Estrogen Receptor Binding for 58,000 chemicals Using an Integrated system of a tree-based model with structural alerts. *Environ Health Perspect* (2002) **110**(1):29–36. doi:10.1289/ehp.0211029

Author disclaimer

This article reflects the views of the authors and does not necessarily reflect those of the U.S. Food and Drug Administration. Any mention of commercial products is for clarification only and is not intended as approval, endorsement, or recommendation.

Data availability

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://github.com/FANMISUA/AADB.git>.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

Authors BH, TM, TE, JD, and DB were employed by Edelweiss Connect Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

- Tong W, Perkins R, Fa Perkinsng H, Hong H, Xie Q, Branham SW, et al. Development of Quantitative Structure-Activity Relationships (QSARs) and their use for priority setting in the testing strategy of endocrine disruptors. *Regul Res Perspect* (2002) **1**(3):1–16.
- Hong H, Fang H, Xie Q, Perkins R, Sheehan DM, Tong W. Comparative molecular field analysis (CoMFA) model using a large diverse set of natural, synthetic and environmental chemicals for binding to the androgen receptor. *SAR QSAR Environ Res* (2003) **14**(5-6):373–88. doi:10.1080/10629360310001623962
- Shi LM, Tong W, Fang H, Xie Q, Hong H, Perkins R, et al. An integrated 4-Phase approach for setting endocrine disruption screening priorities - phase I and II predictions of estrogen receptor binding affinity. *SAR QSAR Environ Res* (2002) **13**(1):69–88. doi:10.1080/10629360290002235
- Sakkiah S, Guo W, Pan B, Kusko R, Tong W, Hong H. Computational prediction models for assessing endocrine disrupting potential of chemicals. *J Environ Sci Health C* (2018) **36**(4):192–218. doi:10.1080/10590501.2018.1537132
- Ng HW, Zhang W, Shu M, Luo H, Ge W, Perkins R, et al. Competitive molecular docking approach for predicting estrogen receptor subtype α agonists and antagonists. *BMC Bioinformatics* (2014) **15**(Suppl. 11):S4. doi:10.1186/1471-2105-15-s11-s4
- Ng HW, Doughty SW, Luo H, Ye H, Ge W, Tong W, et al. Development and validation of decision forest model for estrogen receptor binding prediction of chemicals using large data sets. *Chem Res Toxicol* (2015) **28**(12):2343–51. doi:10.1021/acs.chemrestox.5b00358

13. Ng HW, Shu M, Luo H, Ye H, Ge W, Perkins R, et al. Estrogenic activity data extraction and *in silico* prediction show the endocrine disruption potential of bisphenol A replacement compounds. *Chem Res Toxicol* (2015) **28**(9):1784–95. doi:10.1021/acs.chemrestox.5b00243
14. Hong H, Rua D, Sakkiah S, Selvaraj C, Ge W, Tong W. Consensus modeling for prediction of estrogenic activity of ingredients commonly used in sunscreen products. *Int J Environ Res Public Health* (2016) **13**(10):958. doi:10.3390/ijerph13100958
15. Hong H, Harvey BG, Palmese GR, Stanzione JF, Ng HW, Sakkiah S, et al. Experimental data extraction and *in silico* prediction of the estrogenic activity of renewable replacements for bisphenol A. *Int J Environ Res Public Health* (2016) **13**(7):705. doi:10.3390/ijerph13070705
16. Ye H, Luo H, Ng HW, Meehan J, Ge W, Tong W, et al. Applying network analysis and Nebula (neighbor-edges based and unbiased leverage algorithm) to ToxCast data. *Environ Int* (2016) **89–90**:81–92. doi:10.1016/j.envint.2016.01.010
17. Sakkiah S, Kusko R, Tong W, Hong H. Applications of molecular dynamics simulations in computational toxicology. In: Hong H, editor. *Advances in computational toxicology: methodologies and applications in regulatory science*. Cham: Springer International Publishing (2019). p. 181–212.
18. Sakkiah S, Selvaraj C, Guo W, Liu J, Ge W, Patterson TA, et al. Elucidation of agonist and antagonist dynamic binding patterns in ER- α by integration of molecular docking, molecular dynamics simulations and quantum mechanical calculations. *Int J Mol Sci* (2021) **22**(17):9371. doi:10.3390/ijms22179371
19. Tan H, Wang X, Hong H, Benfenati E, Giesy JP, Gini GC, et al. Structures of endocrine-disrupting chemicals determine binding to and activation of the estrogen receptor α and androgen receptor. *Environ Sci Technol* (2020) **54**(18):11424–33. doi:10.1021/acs.est.0c02639
20. Banerjee A, De P, Kumar V, Kar S, Roy K. Quick and efficient quantitative predictions of androgen receptor binding affinity for screening Endocrine Disruptor Chemicals using 2D-QSAR and Chemical Read-Across. *Chemosphere* (2022) **309**(Pt 1):136579. doi:10.1016/j.chemosphere.2022.136579
21. Wilkes JG, Stoyanova-Slavova IB, Buzatu DA. Alignment-independent technique for 3D QSAR analysis. *J Comput Aided Mol Des* (2016) **30**(4):331–45. doi:10.1007/s10822-016-9909-0
22. Zhang L, Sedykh A, Tripathi A, Zhu H, Afantitis A, Mouchlis VD, et al. Identification of putative estrogen receptor-mediated endocrine disrupting chemicals using QSAR- and structure-based virtual screening approaches. *Toxicol Appl Pharmacol* (2013) **272**(1):67–76. doi:10.1016/j.taap.2013.04.032
23. Lu NZ, Wardell SE, Burnstein KL, Defranco D, Fuller PJ, Giguere V, et al. International Union of Pharmacology. LXV. The pharmacology and classification of the nuclear receptor superfamily: glucocorticoid, mineralocorticoid, progesterone, and androgen receptors. *Pharmacol Rev* (2006) **58**(4):782–97. doi:10.1124/pr.58.4.9
24. Sakkiah S, Ng HW, Tong W, Hong H. Structures of androgen receptor bound with ligands: advancing understanding of biological functions and drug discovery. *Expert Opin Ther Targets* (2016) **20**(10):1267–82. doi:10.1080/14728222.2016.1192131
25. Mooradian AD, Morley JE, Korenman SG. Biological actions of androgens. *Endocr Rev* (1987) **8**(1):1–28. doi:10.1210/edrv-8-1-1
26. Matsumoto T, Sakari M, Okada M, Yokoyama A, Takahashi S, Kouzmenko A, et al. The androgen receptor in health and disease. *Annu Rev Physiol* (2013) **75**:201–24. doi:10.1146/annurev-physiol-030212-183656
27. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res* (2019) **47**(D1):D1102–D1109. doi:10.1093/nar/gky1033
28. Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, et al. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res* (2019) **47**(D1):D930–D940. doi:10.1093/nar/gky1075
29. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res* (2007) **35**:D198–201. doi:10.1093/nar/gkl999
30. Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ. The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol Sci* (2007) **95**(1):5–12. doi:10.1093/toxsci/kfl103
31. Davis AP, Grondin CJ, Johnson RJ, Sciacy D, Wiegiers J, Wiegiers TC, et al. Comparative toxicogenomics database (CTD): update 2021. *Nucleic Acids Res* (2021) **49**(D1):D1138–43. doi:10.1093/nar/gkaa891
32. Hong H, Xu L, Liu J, Jones WD, Su Z, Ning B, et al. Technical reproducibility of genotyping SNP arrays used in genome-wide association studies. *PLoS One* (2012) **7**(9):e44483. doi:10.1371/journal.pone.0044483
33. Pan B, Kusko R, Xiao W, Zheng Y, Liu Z, Xiao C, et al. Correction to: similarities and differences between variants called with human reference genome HG19 or HG38. *BMC Bioinformatics* (2019) **20**(1):252. doi:10.1186/s12859-019-2776-7
34. Hardy B, Douglas N, Helma C, Rautenberg M, Jeliakzova N, Jeliakzov V, et al. Collaborative development of predictive toxicology applications. *J Cheminformatics* (2010) **2**(7):7. doi:10.1186/1758-2946-2-7
35. Hardy B, Apic G, Carthew P, Clark D, Cook D, Dix I, et al. A toxicology ontology roadmap. *ALTEX* (2012) **29**:129–37. doi:10.14573/altex.2012.2.129
36. Hardy B, Apic G, Carthew P, Clark D, Cook D, Dix I, et al. Toxicology ontology perspectives. *ALTEX* (2012) **29**:139–56. doi:10.14573/altex.2012.2.139
37. Kohonen P, Benfenati E, Bower D, Ceder R, Crump M, Cross K, et al. The ToxBank data warehouse: supporting the replacement of *in vivo* repeated dose systemic toxicity testing. *Mol Inform* (2013) **32**(Issue 1):47–63. doi:10.1002/minf.201200114
38. Fourches D, Muratov E, Tropsha A. Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J Chem Inf Model* (2010) **50**:1189–204. doi:10.1021/ci100176x
39. Fourches D, Muratov E, Tropsha A. Curation of chemogenomics data. *Nat Chem Biol* (2015) **11**:535. doi:10.1038/nchembio.1881
40. Fourches D, Muratov E, Tropsha A. Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation. *J Chem Inf Model*. (2016) **56**(7):1243–1252. doi:10.1021/acs.jcim.6b00129
41. Tropsha A. Best practices for QSAR model development, validation, and exploitation. *Mol Inform* (2010) **29**(6-7):476–88. doi:10.1002/minf.201000061
42. Idakwo G, Thangapandian S, Luttrell J, Li Y, Wang N, Zhou Z, et al. Structure-activity relationship-based chemical classification of highly imbalanced Tox21 datasets. *J Cheminform* (2020) **12**(1):66. doi:10.1186/s13321-020-00468-x
43. Hong H, Xie Q, Ge W, Qian F, Fang H, Shi L, et al. Mold(2), molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics. *J Chem Inf Model* (2008) **48**(7):1337–44. doi:10.1021/ci800038f
44. Hong H, Liu J, Ge W, Sakkiah S, Guo W, Yavas G, et al. Mold2 descriptors facilitate development of machine learning and deep learning models for predicting toxicity of chemicals. In: Hong H, editor. *Machine learning and deep learning in computational toxicology*. Cham: Springer International Publishing (2023). p. 297–321.
45. Liu J, Xu L, Guo W, Li Z, Khan MKH, Ge W, et al. Developing a SARS-CoV-2 main protease binding prediction random forest model for drug repurposing for COVID-19 treatment. *Exp Biol Med (Maywood)* (2023) **248**(21):1927–36. doi:10.1177/15353702231209413