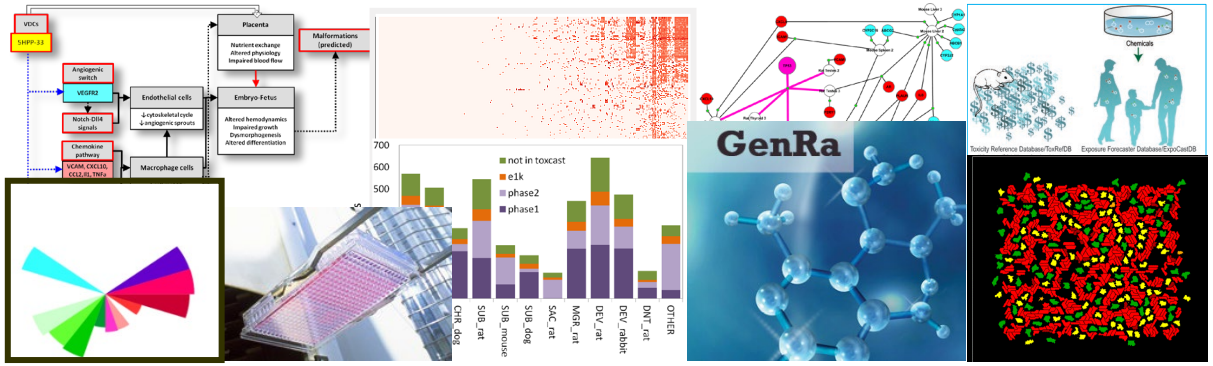
Navigating the Minefield of Computational Toxicology – Charting Progress from BluePrint to Implementation



5 June 2023

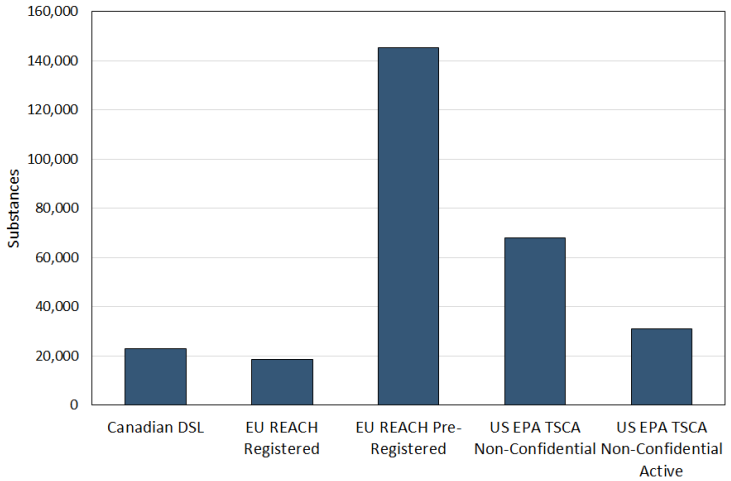
Grace Patlewicz
Center for Computational Toxicology and Exposure
Office of Research and Development

The views expressed in this presentation are those of the presenter and do not necessarily reflect the views or policies of the U.S. EPA

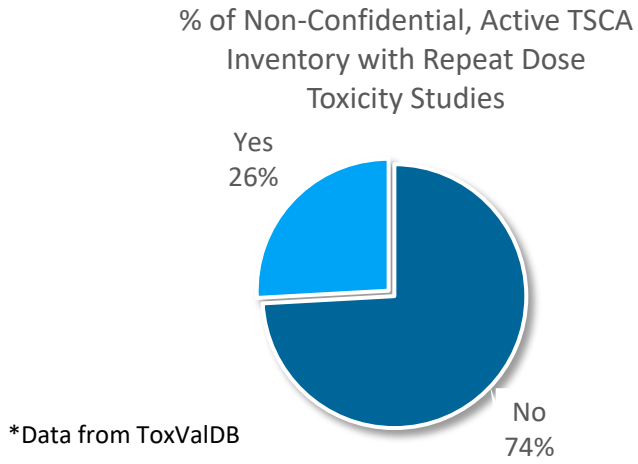
Why New Approach Methodologies (NAMs)?

United States Environmental Protection Agency

Number of Substances

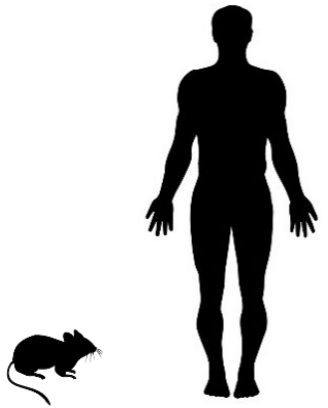


Amount of Data

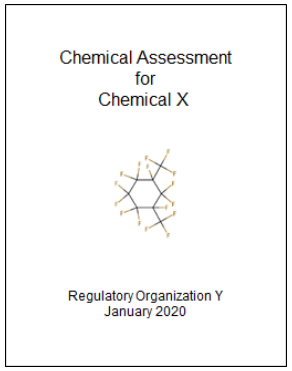


*Data from ToxValDB

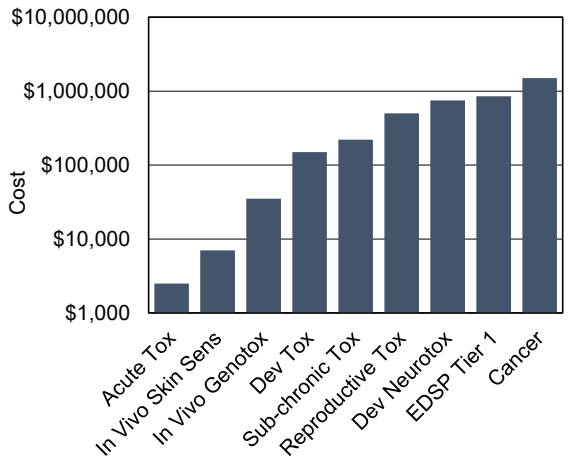
Reliability/Relevance



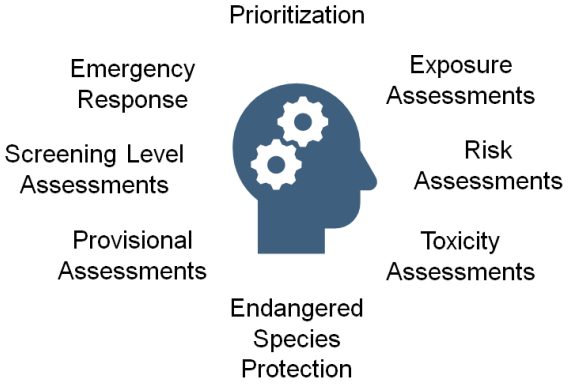
Time



Economics



Broad Range of Decision Contexts



Selected Milestones

2004: DSSTox

2007: NRC Toxicity in the 21st Century (2007)

EPA's NCCT established, EPA ToxCast programme launched

Tox21 established (2008)

2018: Facilitating selection of candidates for prioritisation within TSCA
2019: CompTox BluePrint
2021: PFAS National Testing Strategy

2003

2004-2008

2012-2017

2018-2021

2023

Framework for a Computational Toxicology Research Programme within EPA
Outlines a research plan for 10-15 year horizon

Translation to Regulatory Application
Endocrine Disruption Screening Programme
EDSP

Scope of today's presentation

TSCA New Chemicals Collaborative Research Programme NCCRP

EPA's CompTox Research BluePrint

TOXICOLOGICAL SCIENCES, 169(2), 2019, 317–332

doi: 10.1093/toxsci/kfz058
Advance Access Publication Date: March 5, 2019
Forum

FORUM

The Next Generation Blueprint of Computational Toxicology at the U.S. Environmental Protection Agency

Russell S. Thomas,*,1 Tina Bahadori,† Timothy J. Buckley,‡ John Cowden,* Chad Deisenroth,* Kathie L. Dionisio,‡ Jeffrey B. Frithsen,§ Christopher M. Grulke,* Maureen R. Gwinn,* Joshua A. Harrill,* Mark Higuchi,¶ Keith A. Houck,* Michael F. Hughes,¶ E. Sidney Hunter, III,¶ Kristin K. Isaacs,‡ Richard S. Judson,* Thomas B. Knudsen,* Jason C. Lambert,‖ Monica Linnenbrink,* Todd M. Martin,‖ Seth R. Newton,‡ Stephanie Padilla,¶ Grace Patlewicz,* Katie Paul-Friedman,* Katherine A. Phillips,‡ Ann M. Richard,* Reeder Sams,* Timothy J. Shafer,¶ R. Woodrow Setzer,* Imran Shah,* Jane E. Simmons,¶ Steven O. Simmons,* Amar Singh,* Jon R. Sobus,‡ Mark Strynar,‡ Adam Swank,‡ Rogelio Tornero-Valez,‡ Elin M. Ulrich,‡ Daniel L. Villeneuve,‖ John F. Wambaugh,* Barbara A. Wetmore,‡ and Antony J. Williams*

*National Center for Computational Toxicology, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, †National Center for Environmental Assessment, U.S. Environmental Protection Agency, Washington, D.C. 20004, ‡National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, §Chemical Safety for Sustainability National Research Program, U.S. Environmental Protection Agency, Washington, D.C. 20004, ¶National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, ‖National Center for Environmental Assessment, U.S. Environmental Protection Agency, Cincinnati, OH 45220, ‖National Risk Management Research Laboratory, U.S. Environmental Protection Agency, Cincinnati, OH 45220, and ‖National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Duluth, MN 55804

1To whom correspondence should be addressed at National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, 109 T.W. Alexander Drive, Room D110-D, Mail Code: D143-02, Research Triangle Park, NC 27711. Fax: (919) 541-1194. E-mail: thomas.russell@epa.gov.

Disclaimer: The U.S. Environmental Protection Agency has provided administrative review and has approved this article for publication. The views expressed in this article are those of the authors and do not necessarily reflect the views of the U.S. Environmental Protection Agency.

ABSTRACT

The U.S. Environmental Protection Agency (EPA) is faced with the challenge of efficiently and credibly evaluating chemical safety often with limited or no available toxicity data. The expanding number of chemicals found in commerce and the environment, coupled with time and resource requirements for traditional toxicity testing and exposure characterization,

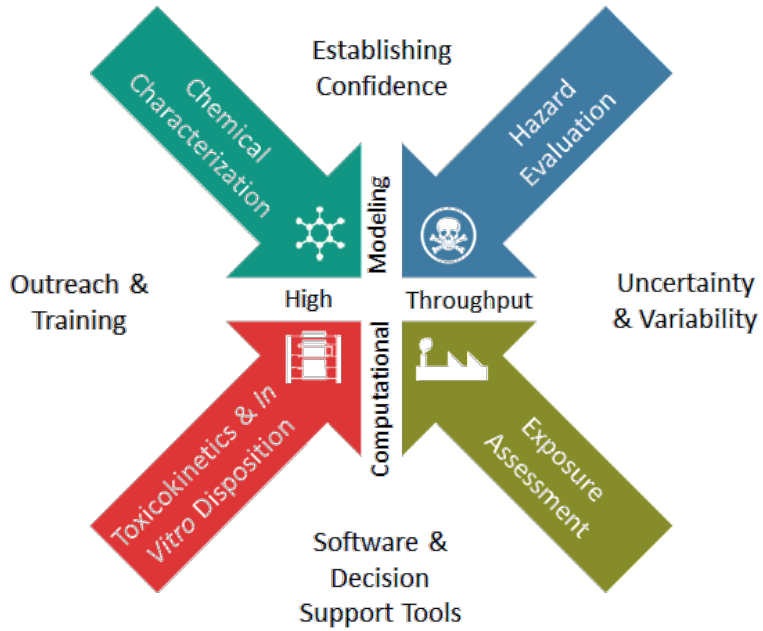
Published by Oxford University Press on behalf of the Society of Toxicology 2019.

- DSSTox
- Chemical library
- Read across
- SAR/QSAR modeling
- Chemotypes
- TTC
- Literature Curation (ChemProp)

- Communities of Practice
- NAM Training courses/ videos

- HTTK assays (metabolism, bioavailability, binding)
- Partition coefficients
- HTTK R package
- Multi-route models
- Literature Curation (CvT)
- In vitro disposition

- Case Studies
- Reference Materials
- Reporting Templates



- CompTox Chemicals Dashboard
- RapidTox
- Factotum
- ECOTOX
- SeqAPASS
- GenRA
- TEST

- In Vitro Assays (HTTr, HTPP, ToxCast)
- Tiered testing
- Organotypic models
- Addressing limitations (metabolism, chemical space)
- Statistical and Biologically-based Modeling
- AOPs
- Literature Curation (ToxVal, ToxRefDB)
  - SEEM
  - ToxBoot
  - HTTK
  - ENTACT

- ExpoCast
- NTA/SSA
- Literature and External Source Curation (CPDat, CPCat, ChemExpoDB)
- Product emissivity

For Toxicology, NAM Development and Application are Being Integrated in a Tiered Framework

OXFORD SOT Society of Toxicology
www.toxsci.oxfordjournals.org

TOXICOLOGICAL SCIENCES, 169(2), 2019, 317–332
doi: 10.1093/toxsci/kfz058
Advance Access Publication Date: March 5, 2019
Forum

FORUM

The Next Generation Blueprint of Computational Toxicology at the U.S. Environmental Protection Agency

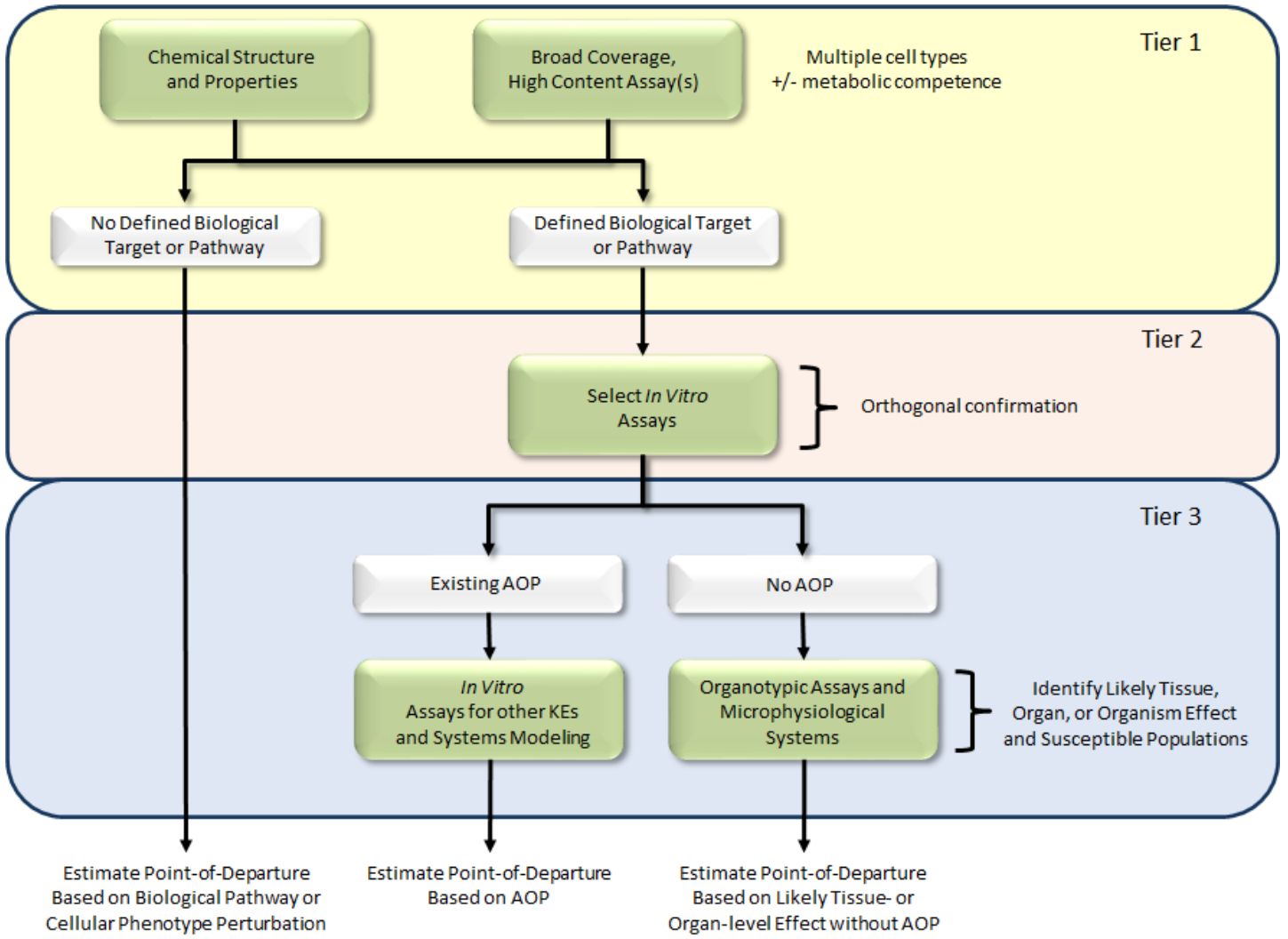
Russell S. Thomas,*,[1] Tina Bahadori,† Timothy J. Buckley,‡ John Cowden,* Chad Deisenroth,* Kathie L. Dionisio,‡ Jeffrey B. Frithsen,§ Christopher M. Grulke,* Maureen R. Gwinn,* Joshua A. Harrill,* Mark Higuchi,¶ Keith A. Houck,* Michael F. Hughes,¶ E. Sidney Hunter, III,¶ Kristin K. Isaacs,‡ Richard S. Judson,* Thomas B. Knudsen,* Jason C. Lambert,‖ Monica Linnenbrink,* Todd M. Martin,‖ Seth R. Newton,‡ Stephanie Padilla,¶ Grace Patlewicz,* Katie Paul-Friedman,* Katherine A. Phillips,‡ Ann M. Richard,* Reeder Sams,* Timothy J. Shafer,¶ R. Woodrow Setzer,* Imran Shah,* Jane E. Simmons,¶ Steven O. Simmons,* Amar Singh,* Jon R. Sobus,‡ Mark Strynar,‡ Adam Swank,‡ Rogelio Tornero-Valez,‡ Elin M. Ulrich,‡ Daniel L. Villeneuve,‖ John F. Wambaugh,* Barbara A. Wetmore,‡ and Antony J. Williams*

*National Center for Computational Toxicology, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, †National Center for Environmental Assessment, U.S. Environmental Protection Agency, Washington, D.C. 20004, ‡National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, §Chemical Safety for Sustainability National Research Program, U.S. Environmental Protection Agency, Washington, D.C. 20004, ¶National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, ‖National Center for Environmental Assessment, U.S. Environmental Protection Agency, Cincinnati, OH 45220, ‖National Risk Management Research Laboratory, U.S. Environmental Protection Agency, Cincinnati, OH 45220, and ‖National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Duluth, MN 55804

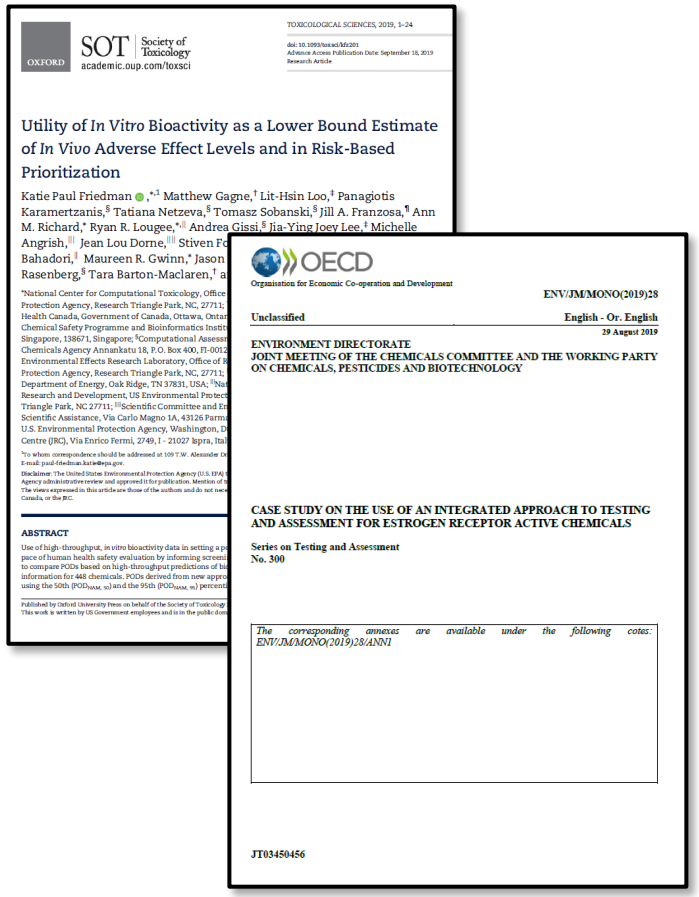
[1]To whom correspondence should be addressed at National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, 109 T.W. Alexander Drive, Room D110-D, Mail Code: D143-02, Research Triangle Park, NC 27711. Fax: (919) 541-1194. E-mail: thomas.russell@epa.gov.

ABSTRACT

The U.S. Environmental Protection Agency (EPA) is faced with the challenge of efficiently and credibly evaluating chemical safety often with limited or no available toxicity data. The expanding number of chemicals found in commerce and the environment, coupled with time and resource requirements for traditional toxicity testing and exposure characterization,



Case Studies to Build Confidence and Help Translate to Regulatory Application



SOT | Society of Toxicology
academic.oup.com/toxsci

TOXICOLOGICAL SCIENCES, 2019, 1–24

Utility of *In Vitro* Bioactivity as a Lower Bound Estimate of *In Vivo* Adverse Effect Levels and in Risk-Based Prioritization

OECD
Organisation for Economic Co-operation and Development

ENV/JM/MONO(2019)28

Unclassified — English - Or. English

29 August 2019

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

CASE STUDY ON THE USE OF AN INTEGRATED APPROACH TO TESTING AND ASSESSMENT FOR ESTROGEN RECEPTOR ACTIVE CHEMICALS

Series on Testing and Assessment
No. 300

The corresponding annexes are available under the following cotes: ENV/JM/MONO(2019)28/ANN1

JT03450456

Completed case studies

Ongoing and New Case Studies

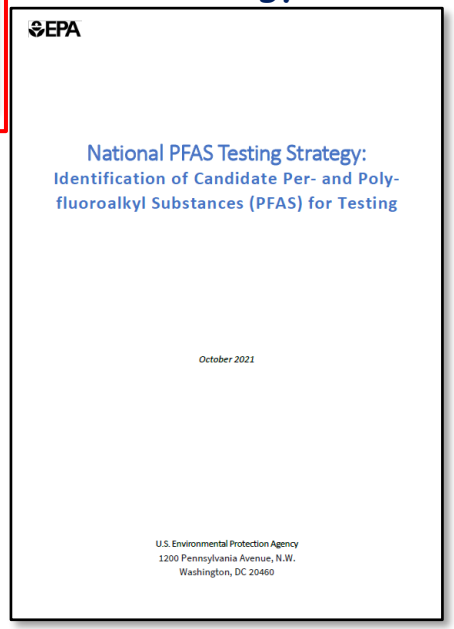
- Use NAMs on selected pesticides with established MOAs
- Develop and apply NAMs for evaluating developmental neurotoxicity
- Integrating NAMs to screen candidates for prioritization under TSCA
- Application of *in vitro* bioactivity and HTTK for screening-level assessments in biosolids
- Prospective case study on application of *in vitro* assays for hazard characterization
- Using NAMs to inform chemical categorisation
- Computational approaches for rapid exposure estimates
- Using *in vitro* bioactivity to inform quantitative ecological hazard assessments
- Evaluating predictivity of HTTK methods

Supporting the Regulatory Partners within EPA Using Computational Toxicology and Exposure in Many Different Areas

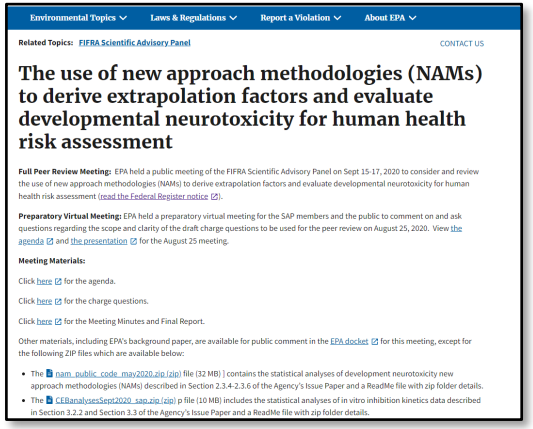
United States Environmental Protection Agency

TSCA New Chemicals Collaborative Research Programme

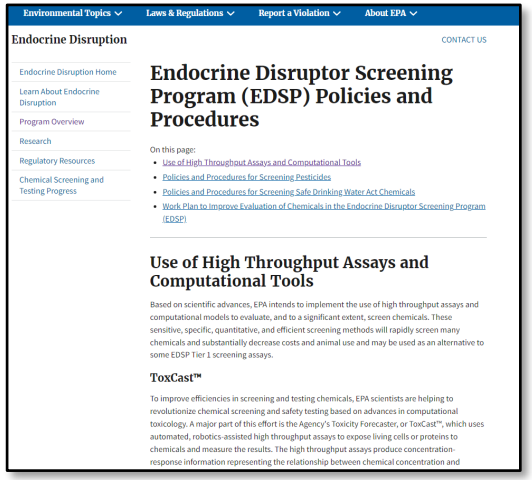
National PFAS Testing Strategy



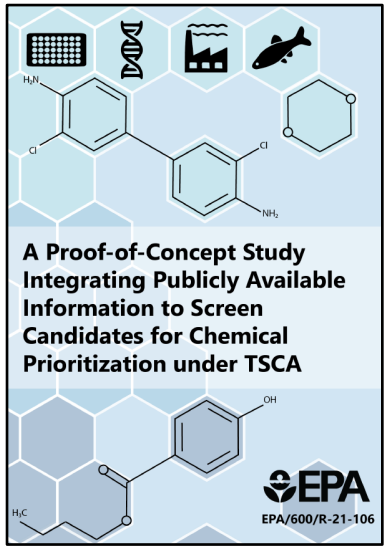
Using New Approaches to Evaluate Developmental Neurotoxicity for FQPA



Endocrine Disruptor Screening Programme



Prioritising Existing Chemicals Under TSCA



Background on TSCA and New Chemical Evaluations

- The TSCA New Chemicals program serves a "gatekeeper" role to manage potential risk to human health and environment from chemicals new to the marketplace; EPA receives ~ 500 new chemical submissions annually.
- TSCA section 5 requires that any person planning to manufacture or import a non-exempt new chemical substance (i.e., a chemical not on the TSCA Inventory) notify EPA before beginning that activity. This notice is known as a premanufacture notice (PMN).

Background on TSCA and New Chemical Evaluations

- EPA is generally required to review these PMNs within 90 days, which consists of assessing the potential risks to human health and the environment of the chemical under the conditions of use, and to make an affirmative determination.
- Where the chemical substance presents or may present an unreasonable risk, EPA must take action to prevent those risks before the chemical can enter commerce.

EPA United States Environmental Protection Agency

Challenges and Opportunities in New Chemical Evaluations

- New chemical submissions typically lack chemical-specific data on human and environmental hazards, exposure, physical chemical properties and environmental fate/transport.
- EPA must make an affirmative determination for all new chemical submissions within the 90-day time period.
- EPA must evaluate new chemical risks under intended, known, and reasonably foreseen conditions of use.

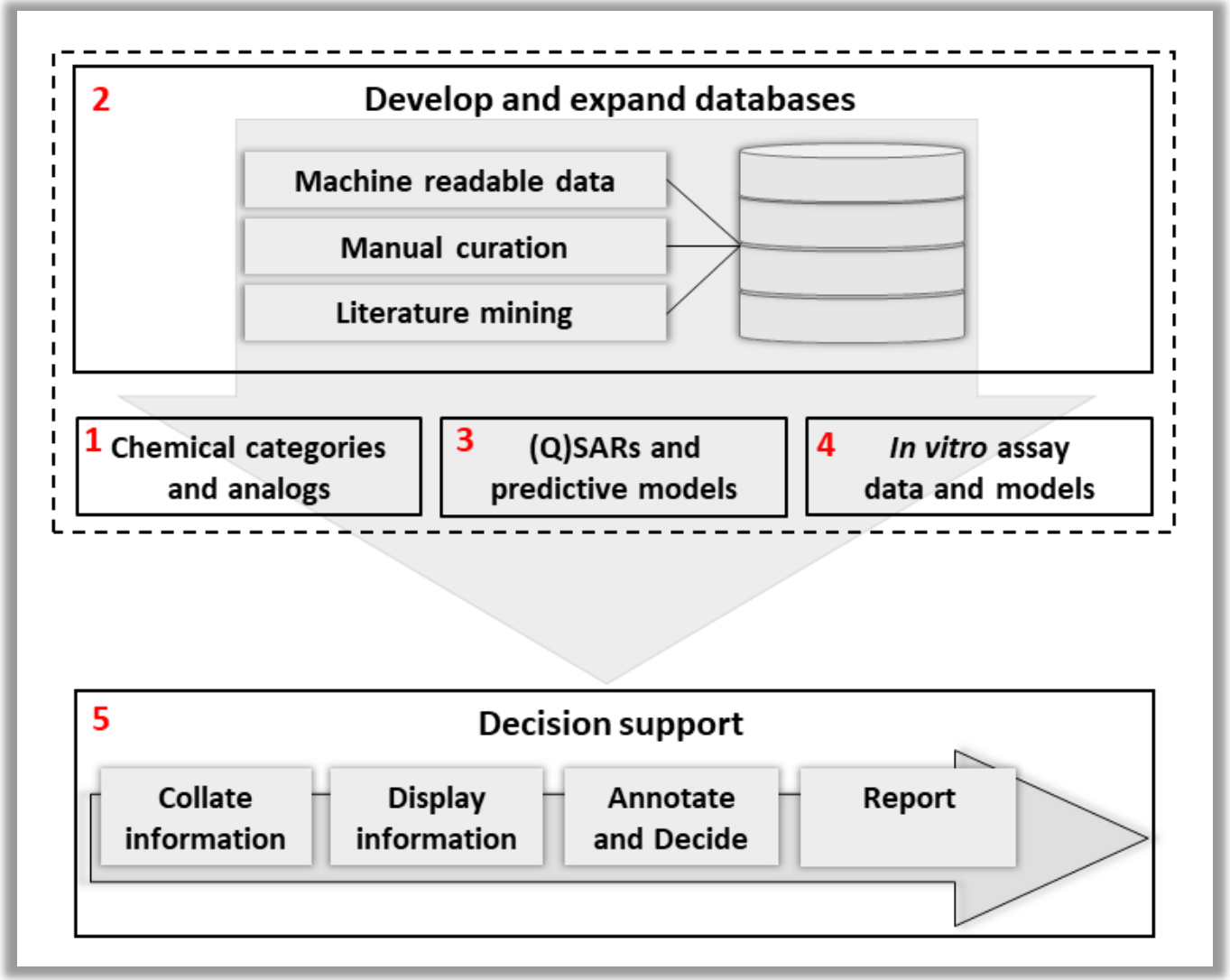
Challenges and Opportunities in New Chemical Evaluations

- EPA's chemical data management infrastructure is outdated. Chemical safety data submissions are scattered across multiple databases, file management systems, and paper files making searches and integration of chemical information inefficient and time consuming.
- EPA is required to reduce and replace vertebrate animal testing.

New approach methods (NAMs), along with data curation and decision support tools, may address additional hazard data gaps, identify potential conditions of use, and furnish more information for making the required determination

Focus Areas in the TSCA New Chemicals Collaborative Research Programme

1) Update and refine chemical categories
2) Develop and expand databases containing TSCA chemical information
3) Develop and refine (Q)SAR and predictive models for physicochemical properties, environmental fate/transport, hazard, exposure, and toxicokinetics
4) Explore ways to integrate and apply NAMs in New Chemical Assessments
5) Develop a TSCA new chemicals decision support tool to modernize the process

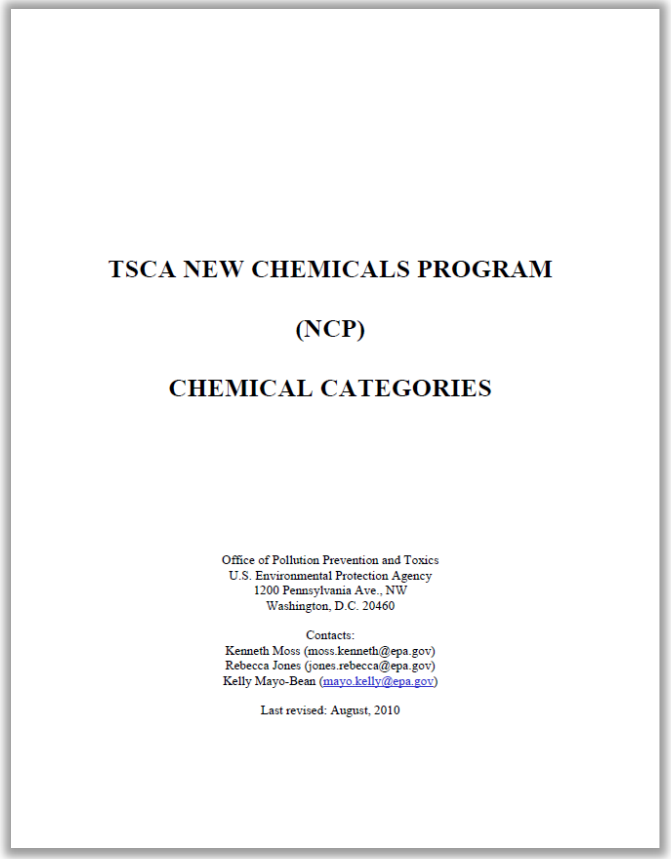


https://epa.figshare.com/articles/code/PubMed_Abstract_Sifter/10324379

Update and refine chemical categories

	Research Area	Challenge	Approach	Expected Outcome(s)
1	Update and Refine Chemical Categories	Currently 56 TSCA categories, last updated 2010	Systematically define chemical categories and analogues for read-across using structural (and other) boundaries; physicochemical properties; structural alerts for hazard, fate, exposure, and/or functional uses; existing hazard data; and/or, *in vitro* mechanistic and toxicokinetic data from NAMs	This will increase the efficiency of new chemical reviews and promote the use of the best available data to protect human health and the environment.

Update and refine chemical categories



TSCA NEW CHEMICALS PROGRAM

(NCP)

CHEMICAL CATEGORIES

Office of Pollution Prevention and Toxics
U.S. Environmental Protection Agency
1200 Pennsylvania Ave., NW
Washington, D.C. 20460

Contacts:
Kenneth Moss (moss.kenneth@epa.gov)
Rebecca Jones (jones.rebecca@epa.gov)
Kelly Mayo-Bean (mayo.kelly@epa.gov)

Last revised: August, 2010

- 56 existing NCCs are characterised largely by structural features and in some cases by physicochemical properties.
- The chemical categories are used to identify potential hazard concerns and testing strategies for new chemical submissions.

- The key goals of collaborative research in this area are to implement the chemical categories in a transparent and reproducible manner that would permit updates with new information, such as additional structure descriptors, physicochemical data, or NAM data.
- Further, planned research will investigate to what extent new categories are needed to capture substances in the TSCA active inventory that could not be readily assigned to one of the 56 existing NCCs.

https://www.epa.gov/reviewing-new-chemicals-under-toxic-substances-control-act-tsca/chemical-categories-used-review-new

Update and refine chemical categories

National PFAS Testing Strategy: Identification of Candidate Per- and Poly-fluoroalkyl Substances (PFAS) for Testing (October 2021)

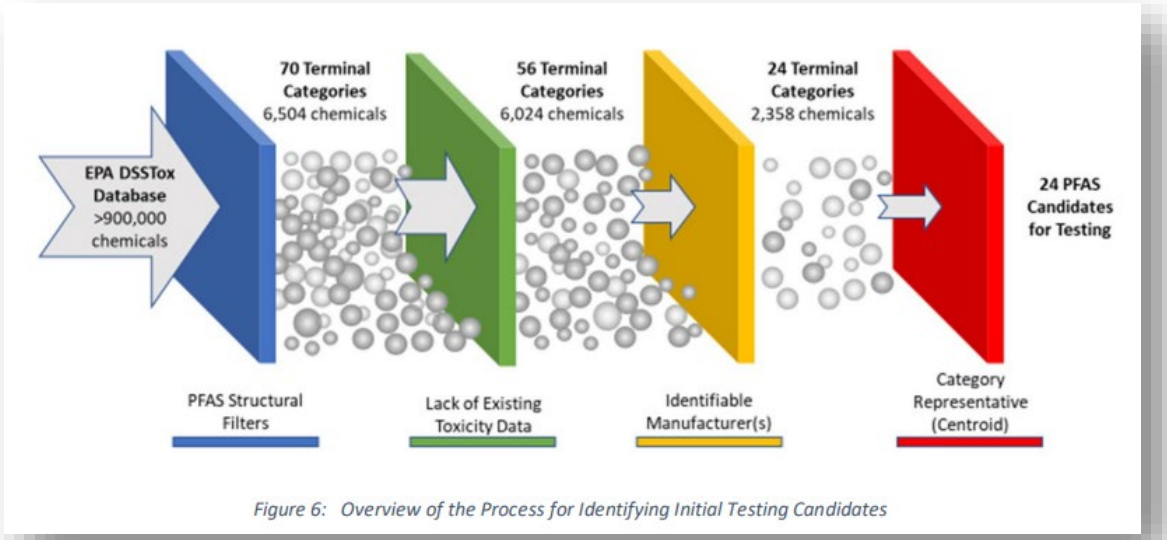
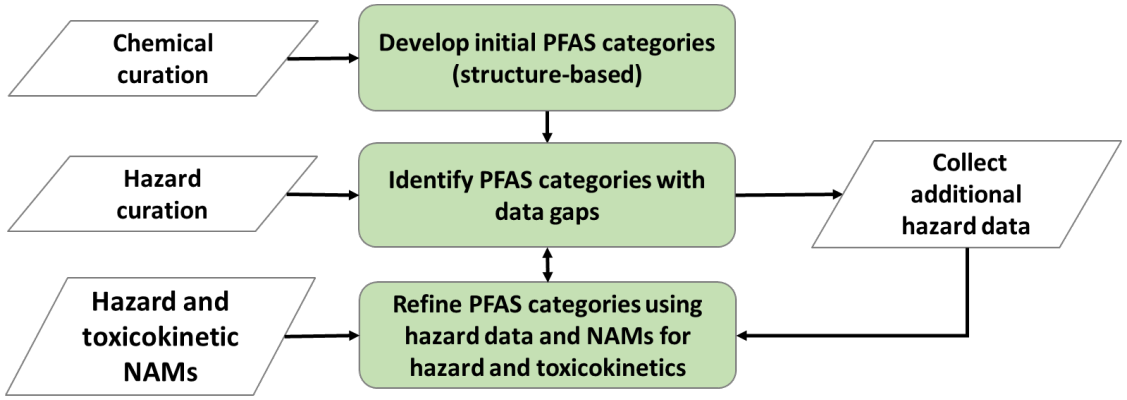


Figure 6: Overview of the Process for Identifying Initial Testing Candidates

https://www.epa.gov/system/files/documents/2021-10/pfas-natl-test-strategy.pdf



Chemical categories may be developed by a combination of one or more of the following:

- structural descriptors,
- physicochemical properties,
- predicted metabolism,
- *in vitro* mechanistic and toxicokinetic, and/or
- *in vivo* toxicity data (human or ecological health).

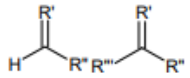
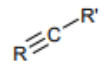
Category research will help address additional questions:

- **To what extent the TSCA chemicals fall within the applicability domain for existing (Q)SAR models or structural alert scheme.**
- **What is a proof-of-principle scheme using chemical categories, read-across, and (Q)SARs to inform *in silico* evaluation of the TSCA active inventory?**

Existing new chemical categories (NCCs) will be turned into a machine- readable format, facilitating profiling and comparison

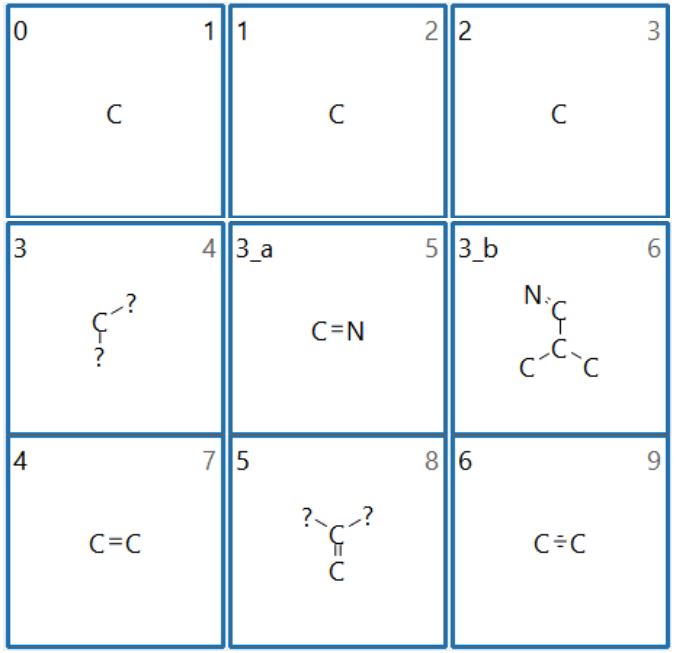
- Structure information built into the current NCCs will be turned into a machine-readable format to enable substructure searching and mapping to other types of structural descriptors, such as ToxPrints (Yang et al., 2015).
- The TSCA non-confidential active chemical inventory will be profiled using the newly codified NCCs to assign them into their respective categories.
- Here, we show an example of translating AIM fragments to machine read-able format.

AIM fragments
[first 5 Analog Identification Methodology (AIM) fragments]

Fragment	Number	Modified Name	General Comments	Structure
-CH3 [aliphatic carbon]	0	Methyl; Primary carbon		
-CH2- [aliphatic carbon]	1	Secondary carbon		
-CH [aliphatic carbon]	2	Tertiary carbon		
C [aliphatic carbon - No H, not tert]	3	Quaternary carbon		
=CH2 [olefinic carbon]	4	Alkene		
=CH- or =C< [olefinc carbon]	5	Vinyl	R's are undefined	
#C [acetylenic carbon]	6	Alkyne		

Multiple steps including initial conversion to SMARTs

Chemical Subgraphs and Reactions Mark-up Language (CSRML) [used for ToxPrints]



This research will enable computational approaches to chemical grouping based on one or more types of structural descriptor(s) as well as other pertinent information.

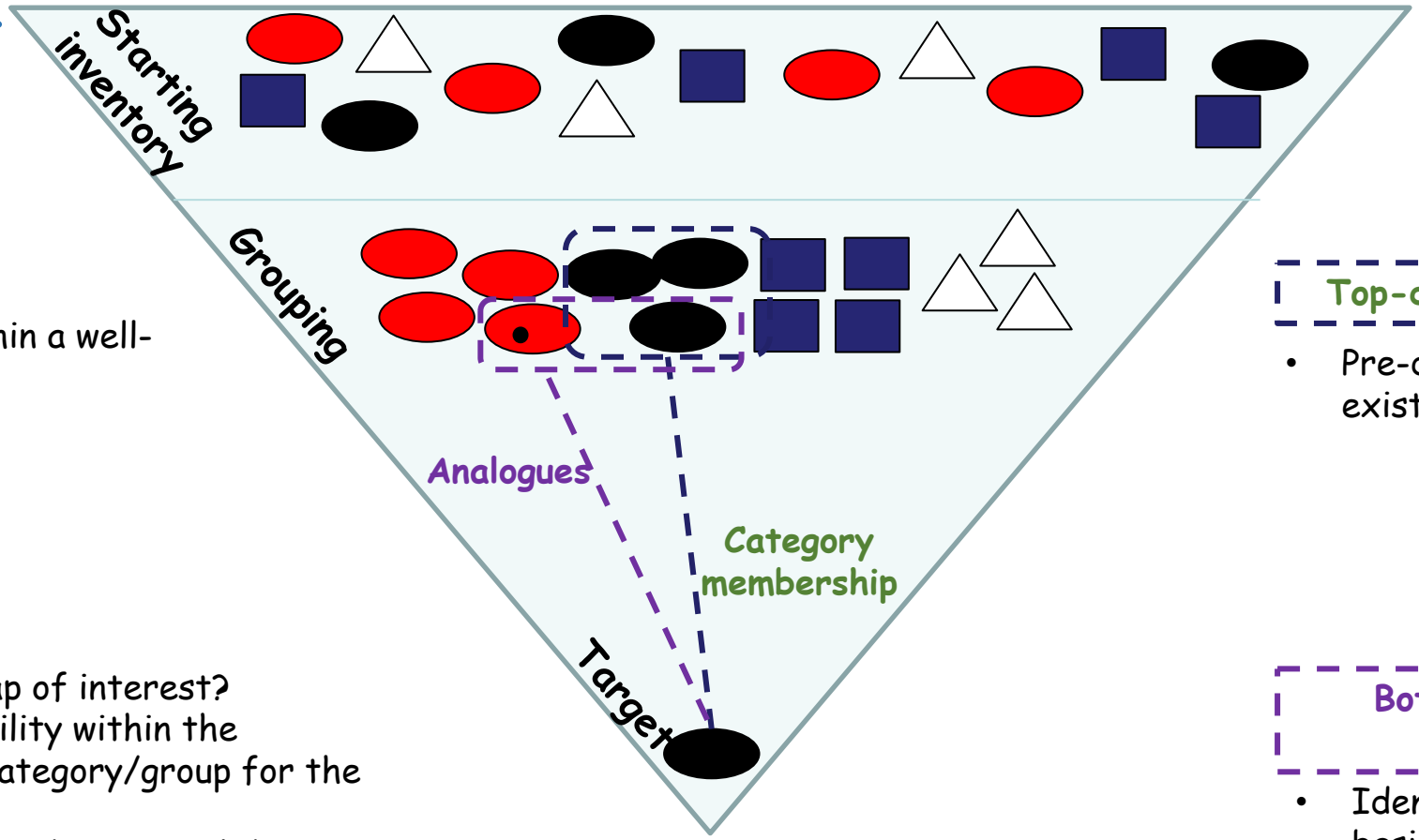
Categories and read-across are two complementary strategies for approaching data-gap filling for data poor chemicals

United States Environmental Protection Agency

Problem formulation context

- Does the target fall within a well-characterised group?
- Is there a specific data gap of interest?
- Is there significant variability within the category/grouping, or no category/group for the target?
- Are there close analogues with empirical data?

Approach



Top-down category approach

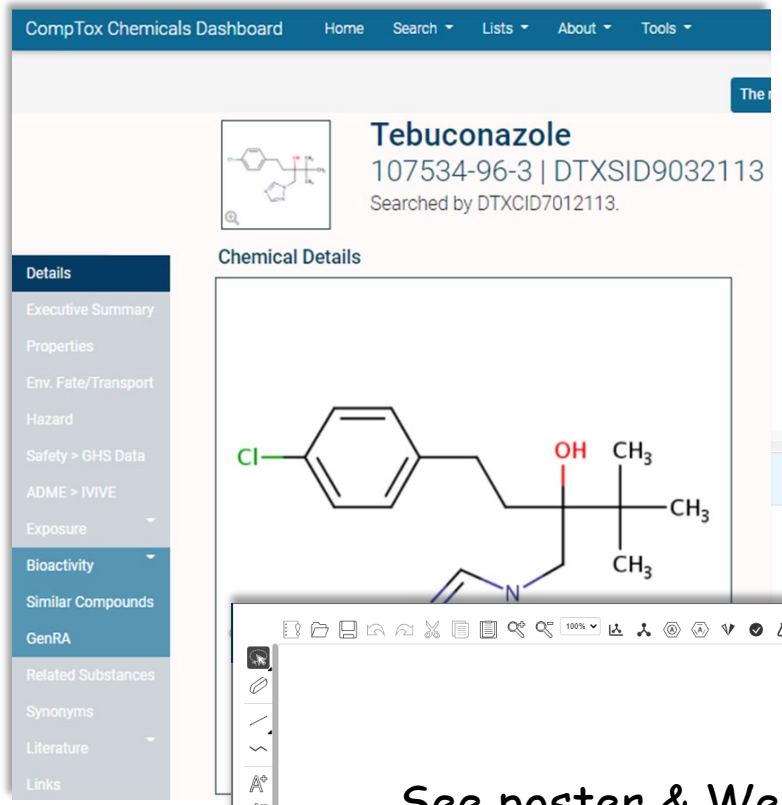
- Pre-define groupings based on existing chemical universe

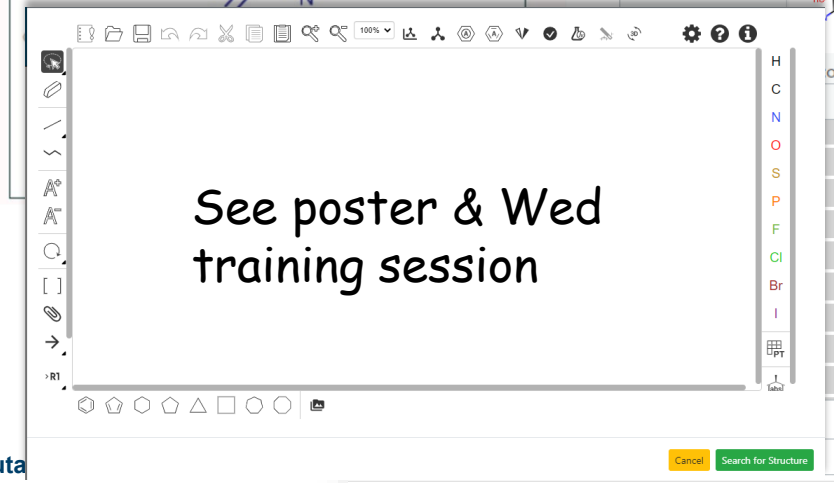
Bottom-up read-across approach

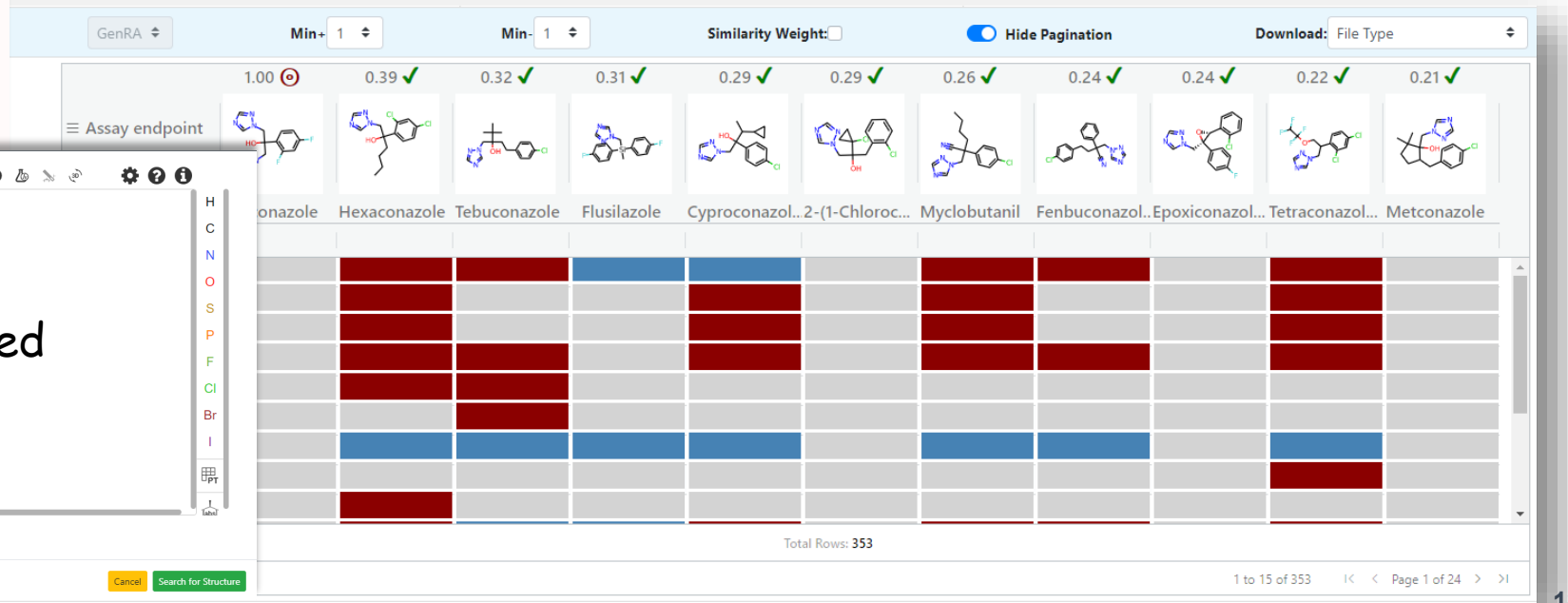
- Identify analogues on the basis of similarity in different contexts

Currently available public tools for systematic read-across may be informative

- GenRA v3.2
- Quantitative evaluation of similarity and confidence in predictions
- Web application and standalone python package (genra-py)
- Interactive workflow to:
  - search for target or draw it;
  - define fingerprints for similarity and number of analogs;
  - Hybrid descriptors now available;
  - Examine what data exist for source analogs;
  - Inspect the consistency, concordance, and range of effects for analogs
  - Understand confidence in the prediction(s)

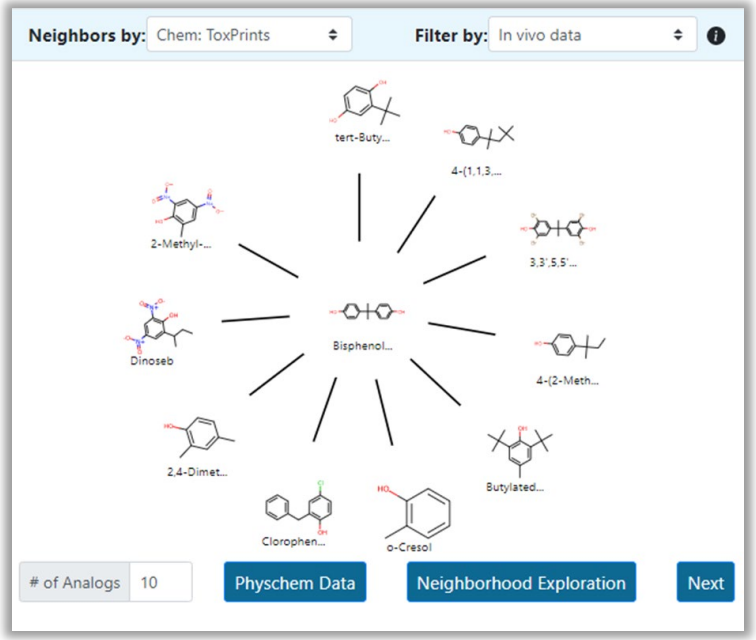






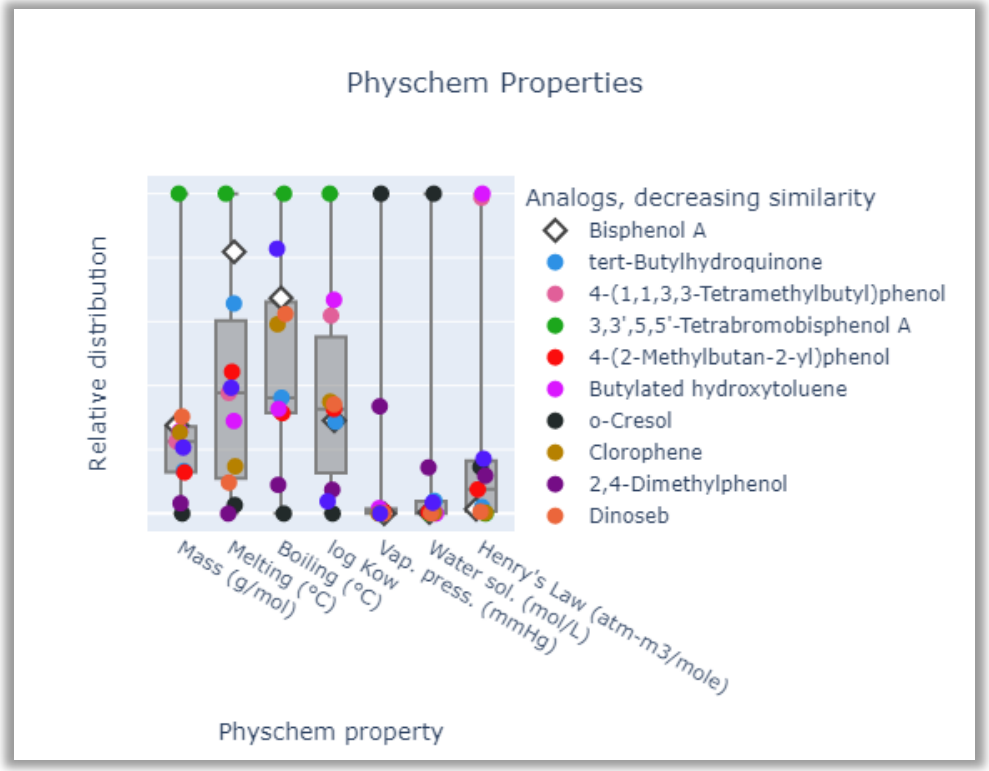
GenRA capabilities will continue to be expanded in Research Area 1

Different fingerprint types and fingerprint hybrids can be used to define neighbourhoods of chemicals associated with available hazard data



Research will examine the impact of hybrid features on GenRA performance

Some analogues defined by one or more fingerprint methods may have more similar physicochemical profiles



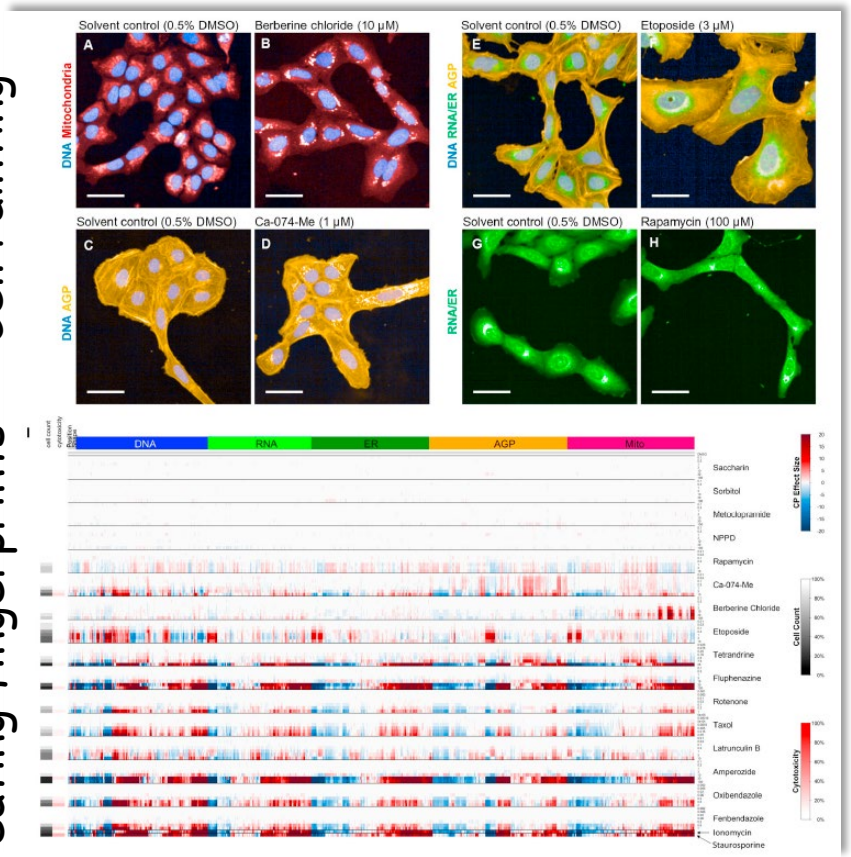
Research will extend similarity contexts to additional types of bioactivity, toxicokinetic, and metabolism data to inform analogue identification and evaluation

Extending similarity context to other types of data: bioactivity and toxicokinetics

Example of NAM-based fingerprints for hazard profile

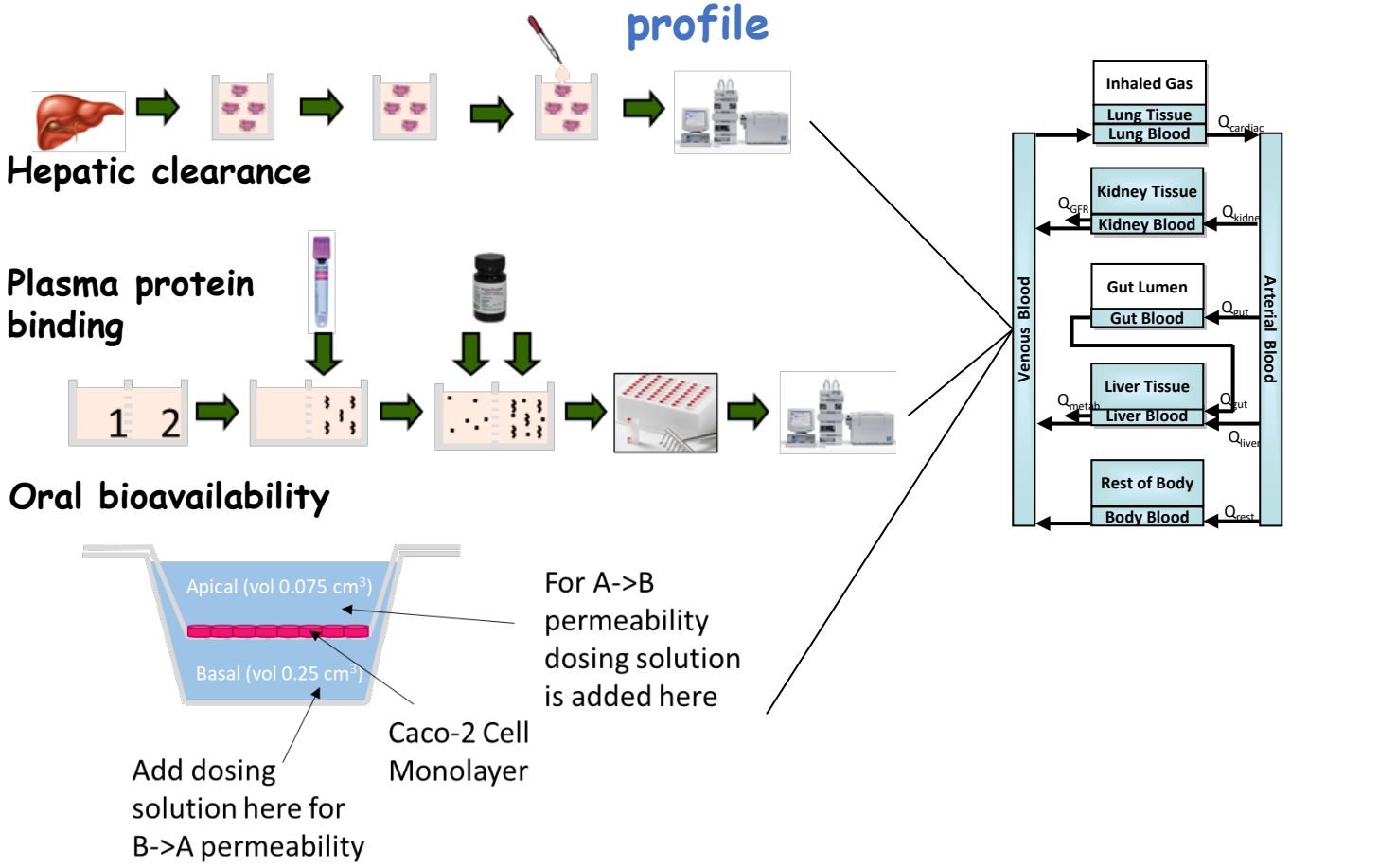
Cell Painting

Creating fingerprints



Nyffeler et al. 2020
10.1016/j.taap.2019.114876

NAM-based fingerprints for toxicokinetic profile



HTTK, led by Drs. John Wambaugh and Barbara Wetmore

Extending similarity context to other types of data: metabolism and metabolites

How to leverage information about the metabolic pathways of substances in order to increase confidence in toxicological assessment via read-across

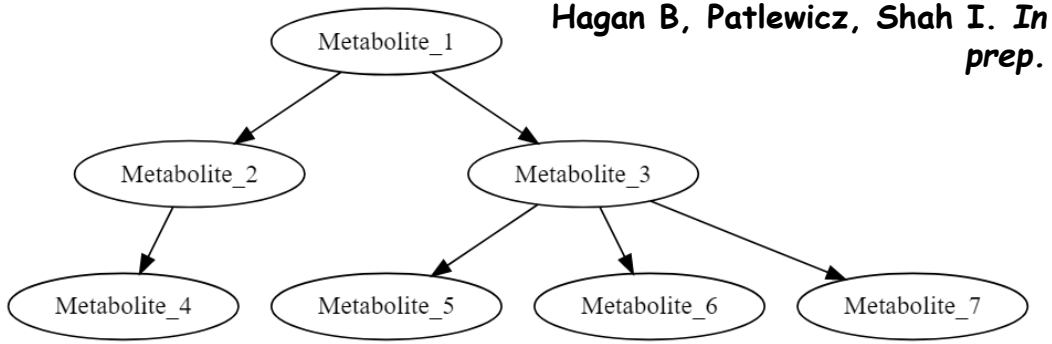
- Metabolic similarity is an important consideration in evaluating analogue suitability in read-across, but current practice often relies on expert judgement and/or empirical (*in vivo*) metabolism data.

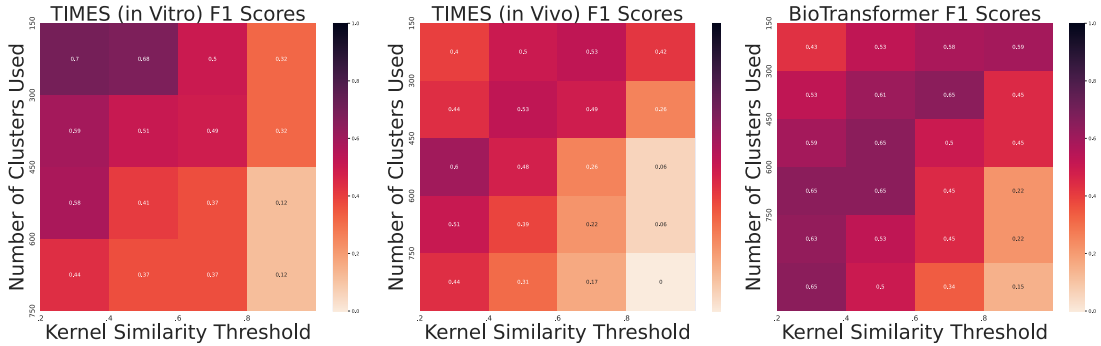
- As empirical metabolism data is limited, we will make use of predicted metabolism data from different tools such as BioTransformer and TIMES. How well these tools perform relative to reported empirical data will be evaluated.

- One approach being explored to codify 'metabolic information' is to construct metabolic graphs from the predictions generated from different tools and evaluate their similarity with different metrics such as kernel approaches.

Metabolic Graph Construction

Hagan B, Patlewicz, Shah I. *In prep.*





TIMES *(in vitro)* **TIMES *(in vivo)*** **BioTransformer**

Comparing the correspondence between structural analogues pairs vs using kernel approach

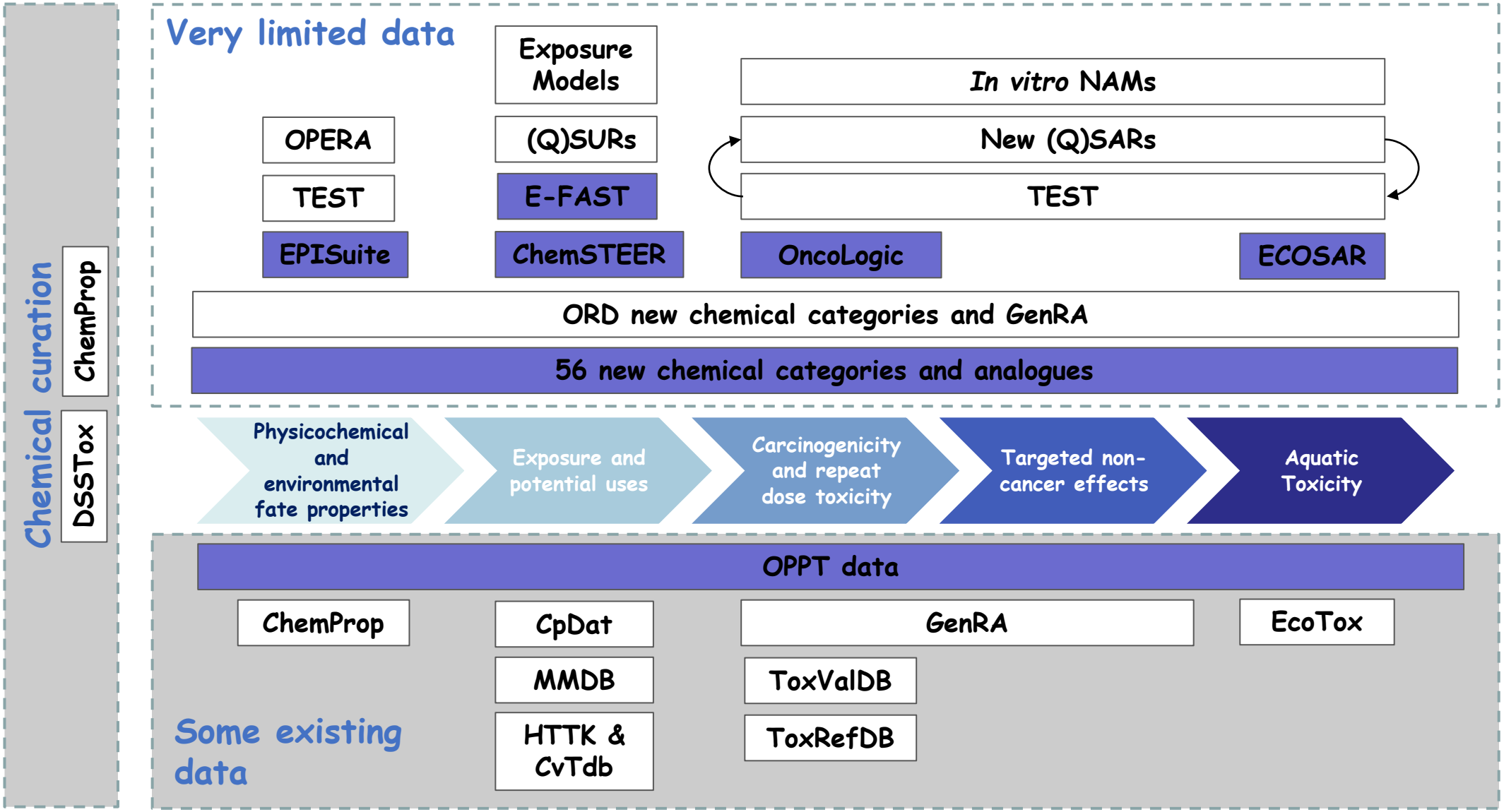
Summary of Research Area 1

- Produce a computational approach to chemical grouping into categories based on structure and other descriptors, including structural descriptors, physicochemical and environmental fate properties, predicted metabolism and/or metabolites, NAM-based hazard and toxicokinetic information, and/or *in vivo* hazard data (human and/or ecological health)
- Continue enhancing GenRA capabilities to include:
  - evaluating the impact of hybrid features on GenRA performance;
  - extending similarity contexts to additional types of bioactivity data;
  - evaluating the contribution of metabolism data to inform analogue identification and evaluation; and,
  - additional case studies to build confidence in the use of GenRA versus other read-across approaches
- Characterise the chemical structure space encompassed within the TSCA non-confidential active chemical inventory and evaluate to what extent the chemicals on this inventory fall within the applicability domain for (Q)SAR models or other structural alert schemes (either existing or in development)

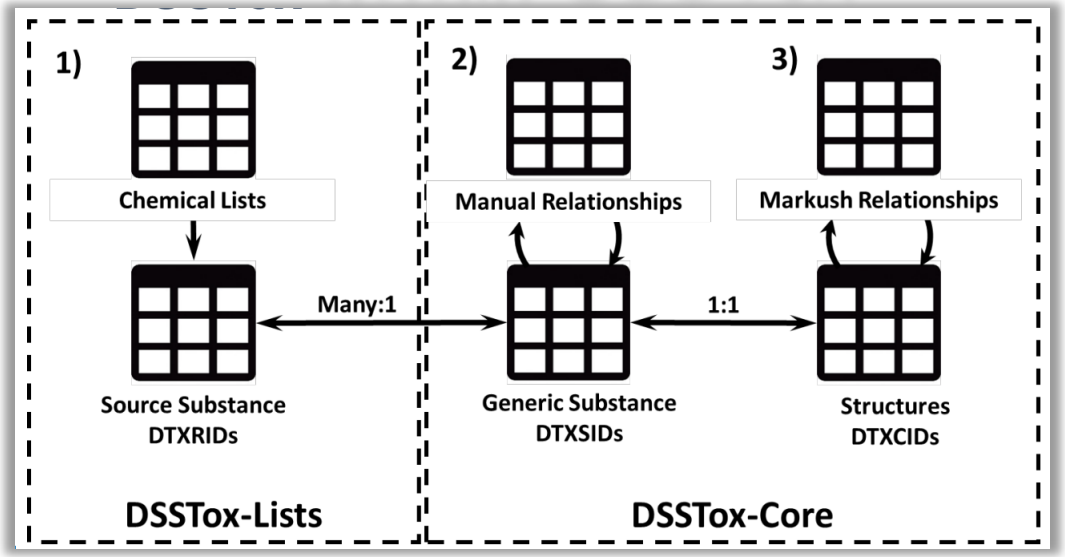
Research Area 2

Research Area		Challenge	Approach	Expected Outcome(s)
2	Develop and Expand Databases Containing TSCA Chemical Information	Existing TSCA information is not computationally accessible or easily searchable	Continue extraction and curation of physical-chemical property, environmental fate, hazard, and exposure information (non-CBI) in ORD databases Map information in ORD databases to standardised reporting templates and store in an International Uniform Chemical Information Database (IUCLID)	Publicly available sources can expand the amount of information available, enhancing chemical reviews and enabling efficient sharing of chemical information across EPA.

Different data scenarios



Expanded chemical curation of chemical identity makes more chemistry accessible for applications within DSSTox



Credit: Williams AJ, Richard AM, Grulke C

- >1.2 million unique substances
- Informs accurate structure-data linkages for screening projects, read-across, non-targeted analysis, structure-based modeling

As the TSCA active inventory of chemicals grows each year, expansion of the DSSTox database to include these chemicals as well as existing and emerging chemicals of interest for modeling applications is essential.

- Curation enables programmatic access to any data that can be linked to a DSSTox identifier.
- New chemical submissions under TSCA may be for defined or complex mixtures, and chemistry curation can provide solutions for better linking appropriate data to these mixtures to facilitate read-across or other downstream predictions (e.g., PFAS, substances of unknown or variable composition, complex reaction products and biological materials).

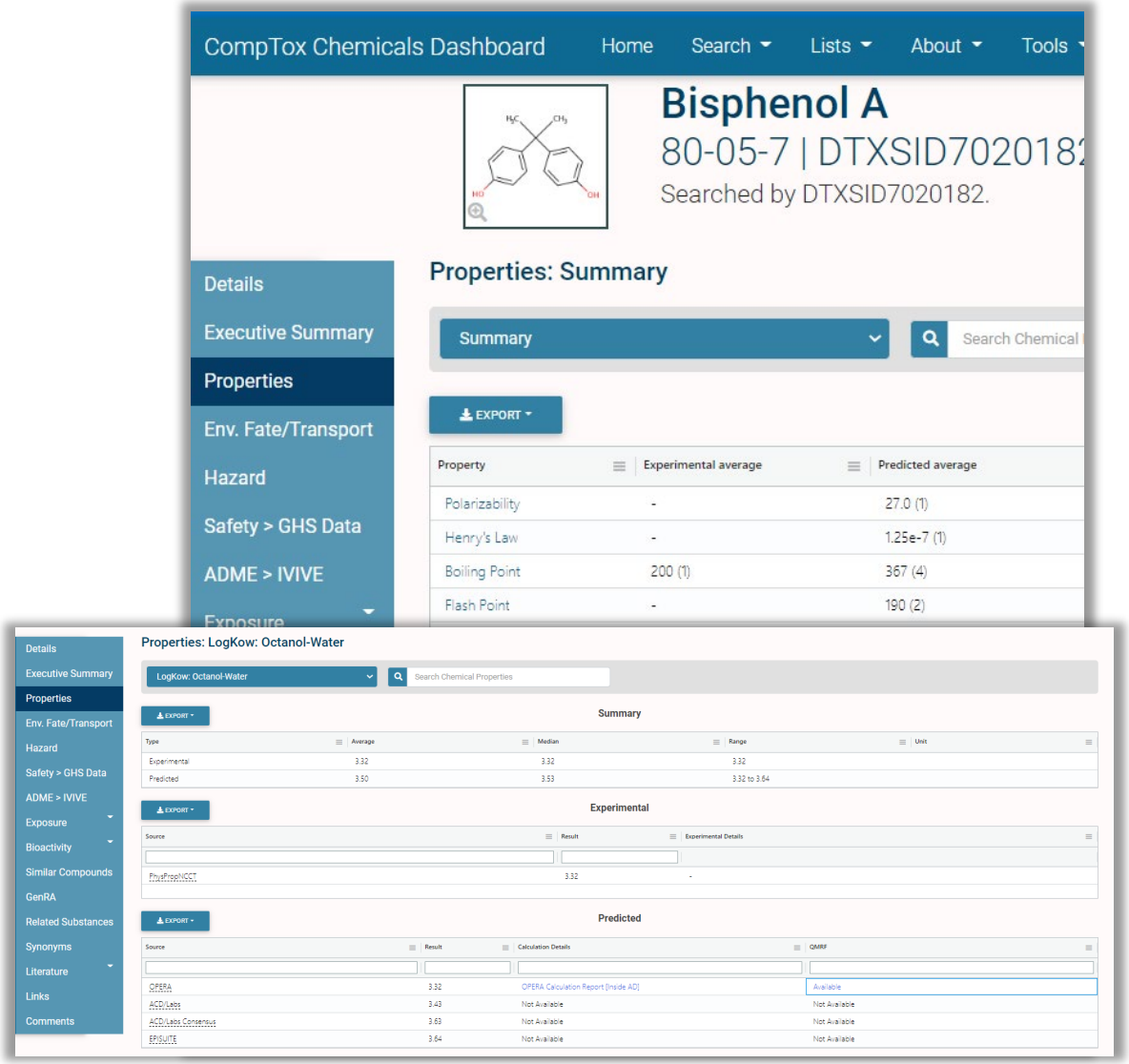
As more chemicals are added to the TSCA active nonconfidential inventory, or chemistries with limited available information are identified, more structure, physicochemical, and environmental fate property data curation is needed to support decision making and data interoperability.

DSSTox = Distributed Structure-Searchable Toxicity (database)

Expanded physicochemical and environmental fate properties and predictions inform (Q)SARs: ChemProp

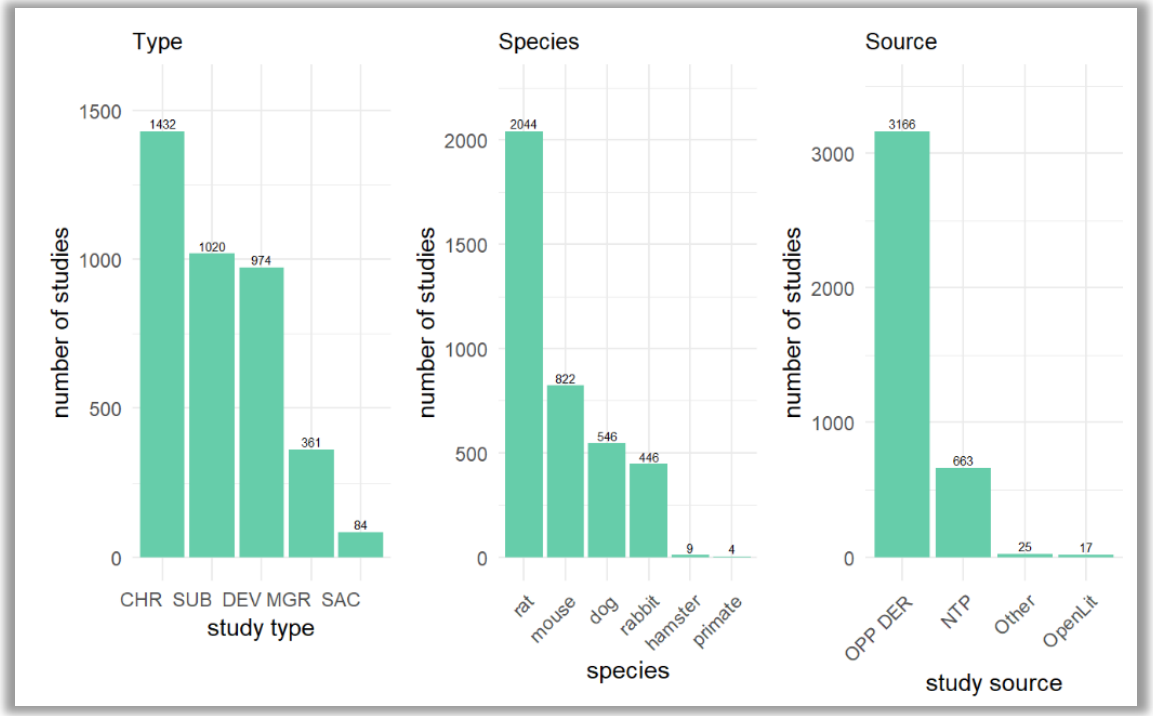
- ChemProp stores both experimental and predicted physicochemical and environmental fate property data for access by applications such as the CompTox Chemicals Dashboard and (Q)SAR development.
- Experimental data have been harvested and curated from online sources such as the PHYSPROP database, ECOTOX, online public sources (eChemportal, PubChem, LookChem, OChem), and peer-reviewed literature.
- Predicted data are generated using OPERA, TEST, EPISuite, ECOSAR, and ACD/Labs.

Making forward predictions of these properties for new chemical submissions may improve with more curated data from existing TSCA-relevant chemicals.



Expanding databases with hazard data will improve *in silico* approaches, including (Q)SAR and read-across, for new chemical evaluation: ToxRefDB

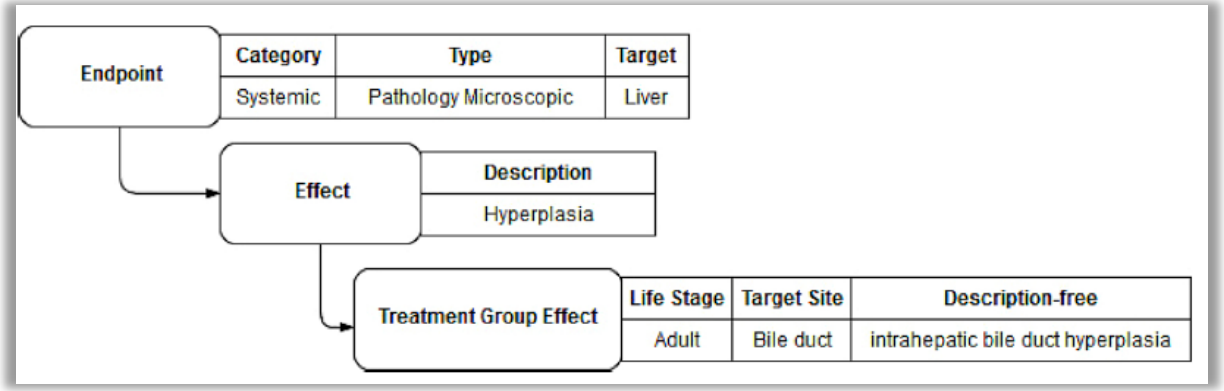
ToxRefDB v2.1 contains summary information for 1143 chemicals and 5986 studies, with quantitative dose-response data extracted for 3871 studies.



The study designs with highest frequency in the database include chronic (CHR), sub-chronic (SUB), developmental (DEV), subacute (SAC), multigeneration reproductive (MGR.

Credit Madison Feshuk, Dr. Katie Paul Friedman, Dr. Sean Watford

- Includes guideline or guideline-like studies, with guideline profiles developed for OCSPP series 870 Health Effects and some NTP study types to allow inference of negative effects
- Study design and meta-data, dose-response data, and detailed effect terminology linked to study type/guideline
- Standardised effect terminology developed for ToxRefDB mapped to terms from the United Medical Language System (UMLS)
- An application was developed to manage two curator and manager reviews of manually curated data from source

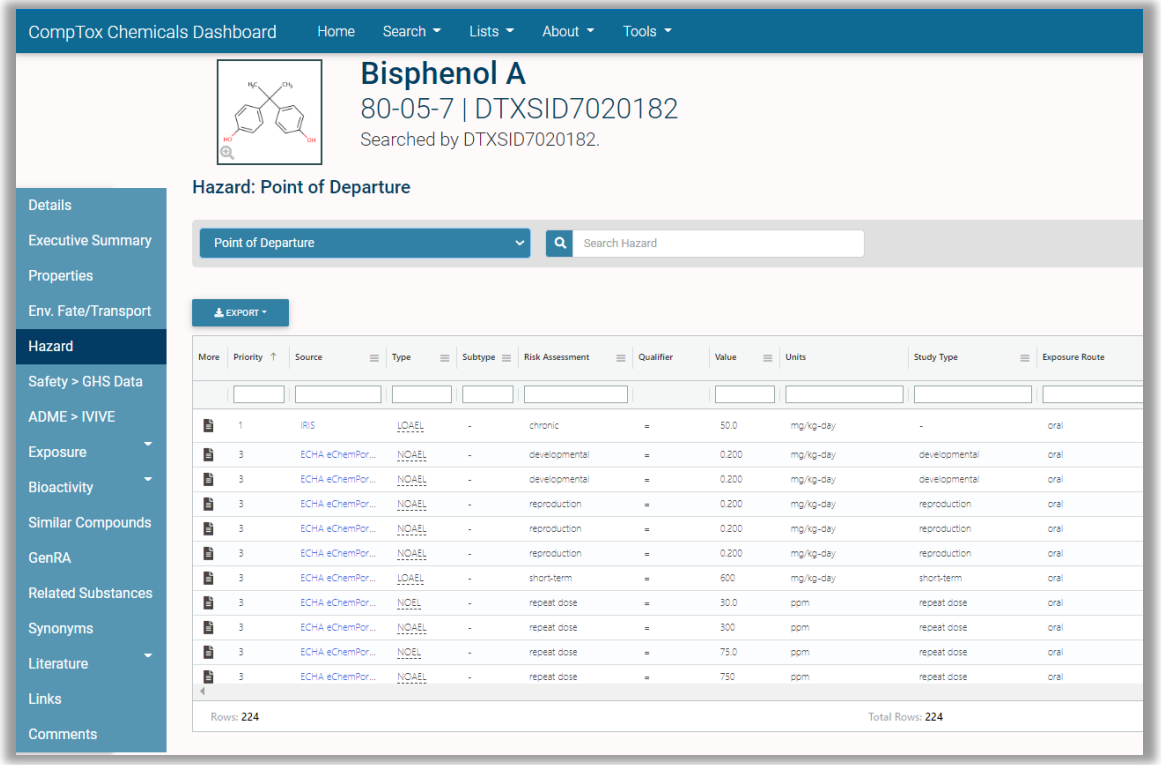


ToxRefDB v2.0: Watford S et al. 2019 DOI [10.1016/j.reprotox.2019.07.012](https://doi.org/10.1016/j.reprotox.2019.07.012)

https://doi.org/10.23645/epacomptox.6062545.v4

Expanding databases with hazard data will improve *in silico* approaches, including (Q)SAR and read-across, for new chemical evaluation: ToxValDB

- ToxValDB is a collection of quantitative information on chemicals and *in vivo* toxicology summary values
  - Experimental *in vivo* toxicology records
    - PODs (LOAEL, NOAEL, BMD), effects, species, exposure routes, study types
    - Human health and ecological health
    - Example sources: HPVIS, ToxRefDB, ECOTOX, HAWC, EFSA, ECHA, COSMOS, HESS
  - Risk assessments
    - RfD, RfC, cancer slope factors, caner unit risk
    - Example sources; IRIS, PPRTV, ATSDR, Cal OEHHA
  - Air, water and soil quality values, worker exposure limits
    - Example sources: RSL, OSHA, NIOSH
- Data is computationally extracted from source documents / databases and mapped to common terms
- Currently, 47 sources and > 50,000 chemicals with at least one value (ToxValDB v9.2).
- ToxValDB v9.4 about to be released



https://doi.org/10.23645/epacomptox.20394501.v3

Credit Dr. Richard Judson

Summary of Research Area 2

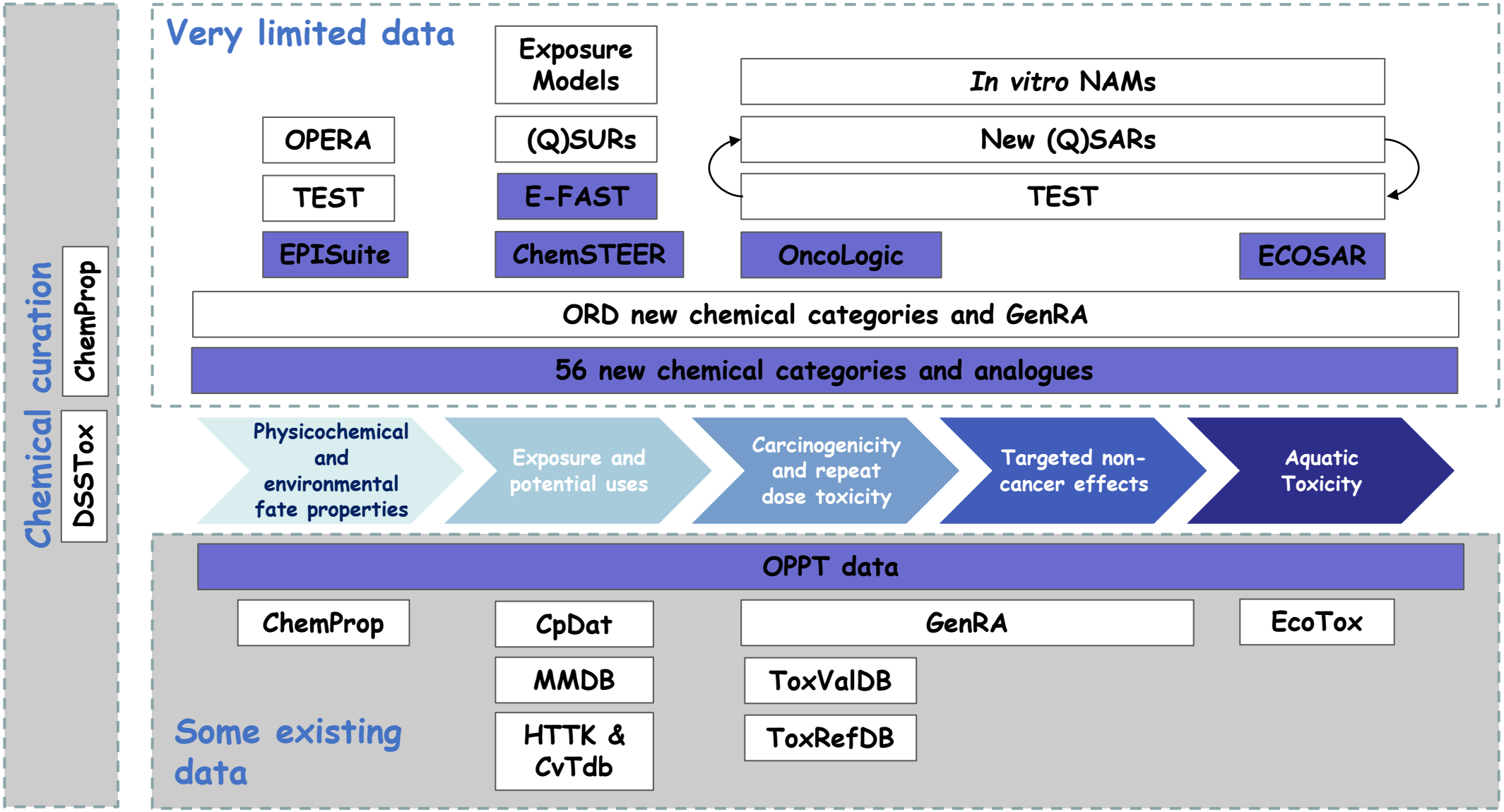
Chemical data for TSCA-relevant chemicals will be curated across the myriad ongoing curation activities in ORD

Data type curated	Relevant Database(s) and links to full data
Chemistry and properties	DSSTox, ChemProp: https://comptox.epa.gov/dashboard/
In vivo hazard in human health relevant models	ToxRefDB: https://doi.org/10.23645/epacomptox.6062545.v4 ToxValDB: https://doi.org/10.23645/epacomptox.20394501.v3
In vivo hazard in ecologically relevant species	ECOTOX Knowledgebase: www.epa.gov/ecotox
Monitoring, release, and product information for exposure	MMDB: https://clowder.edap-cluster.com/datasets/606cc2bd9932c7c0b50a73af CpDat: https://doi.org/10.23645/epacomptox.5352997
Toxicokinetic data	HTTK: https://cran.r-project.org/web/packages/httk/index.html CvTdb: https://github.com/USEPA/CompTox-PK-CvTdb

Research Area 3

	Research Area	Challenge	Approach	Expected Outcome(s)
3	Develop and Refine QSAR and Predictive Models for Physical-Chemical Properties, Environmental Fate/Transport, Hazard, Exposure, and Toxicokinetics	Currently used models are not always publicly accessible, easy to update with additional chemicals, or the best performing for all chemistries	Develop and update QSAR and predictive models using existing data and curated data from Research Area #2 Evaluate models to determine the best suite for use by OPPT for regulatory purposes	Updated models that reflect the best available science, increase transparency, and a process for updating these models as science allows.

Different data scenarios



Existing use of QSAR and structure alerts can be enhanced by ongoing work in ORD to publish QSARs for real-time prediction

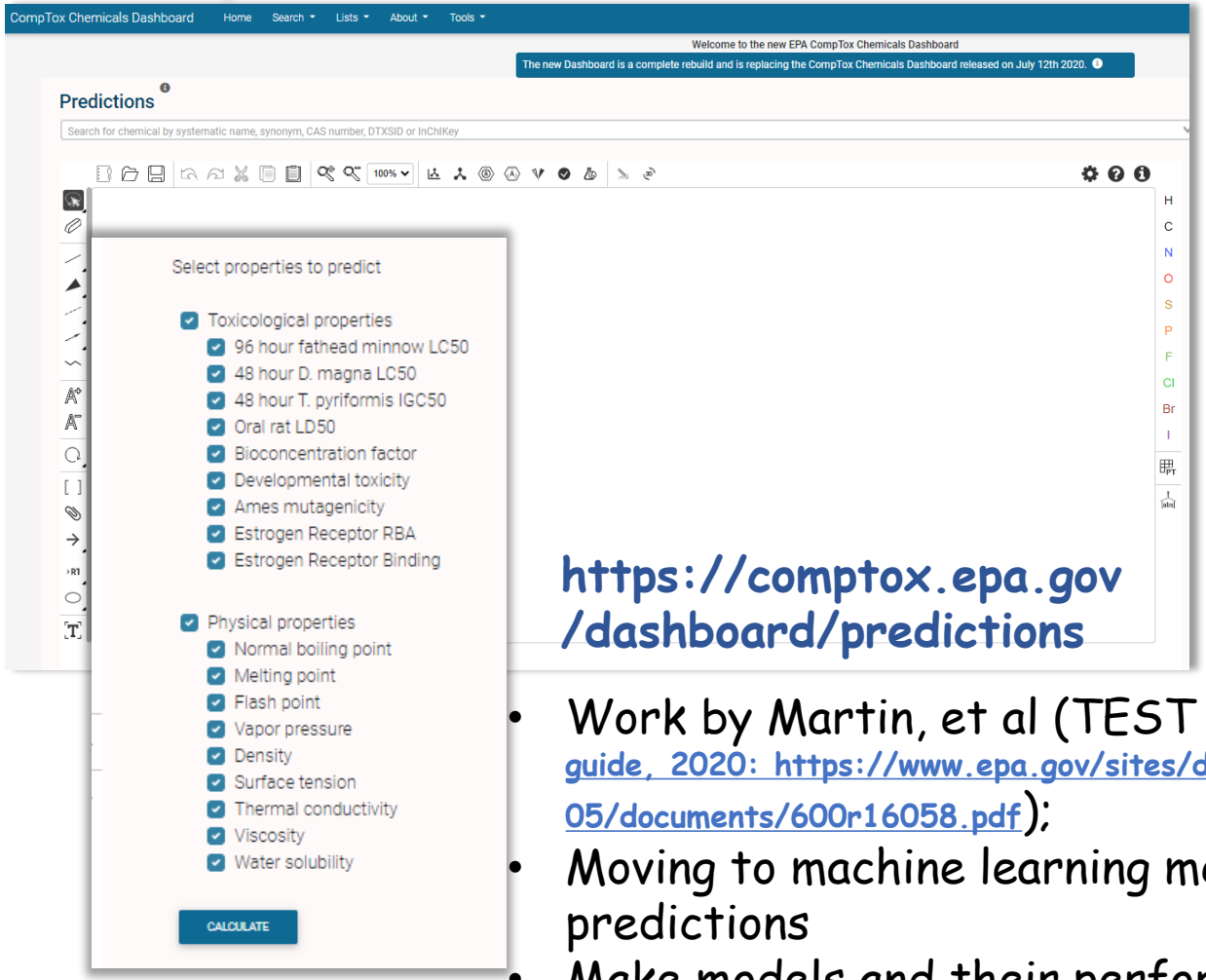
Existing QSAR and structure profiling strategies, such as:

- EPI Suite
- ECOSAR
- OncoLogic
- OECD QSAR Toolbox structure-based profilers

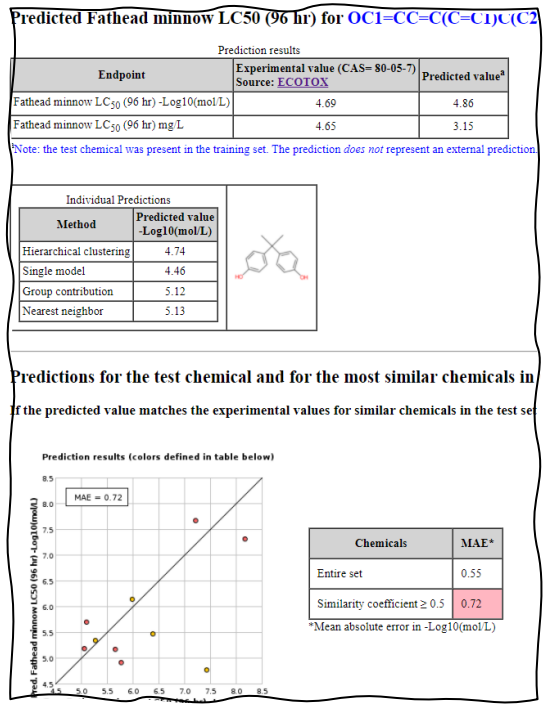
Existing OECD guidance, such as:

Guidance document on the validation of quantitative structure-activity relationships models (2007)

See presentation by Drs Martin & Charest



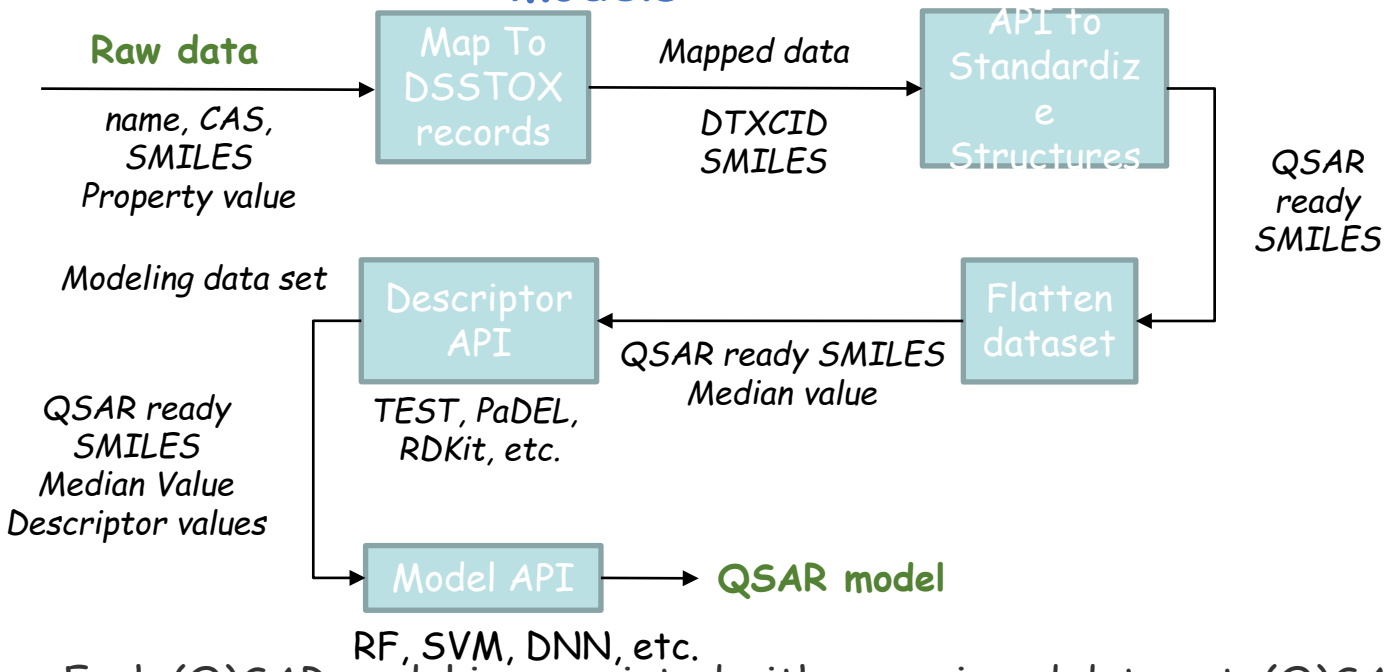
https://comptox.epa.gov/dashboard/predictions



- Work by Martin, et al (TEST models and software, User guide, 2020: https://www.epa.gov/sites/default/files/2016-05/documents/600r16058.pdf);
- Moving to machine learning methods and their consensus predictions
- Make models and their performance reports publicly available, including evaluation of applicability domain

WebTEST2.0 as a model registration and model development platform

Standardised workflow for developing models



Capability to build machine-learning and consensus models within the WebTEST2.0 platform

- Python-based machine learning methods including:
  - RF - Random Forest
  - SVM – Support Vector Machine
  - DNN – Deep Neural Network
  - XGBoost – eXtreme Gradient Boosting
  - kNN- k nearest neighbors
- Consensus of machine-learning methods
  - Consensus – average of above methods

- Each (Q)SAR model is associated with a versioned data set, (Q)SAR methodology, and molecular descriptor set (all stored in a database) so that the predictions are reproducible
- **Easily implementable as web services for both model building and real-time model prediction that will provide deployable (Q)SAR models with appropriate documentation.**

Credit Dr. Todd Martin

WebTEST2.0 as a model registration and model development platform

- Models for physicochemical properties (*e.g.*, octanol water partition coefficient (logKow), vapor pressure, and Henry's law constant) are being developed using the WebTEST2.0 workflow.
- Revised toxicity models will be developed by expanding the toxicity datasets for WebTEST1.0 (*e.g.*, acute aquatic toxicity).
- In addition, models will be developed for additional toxicity endpoints (*e.g.*, carcinogenicity, repeat dose toxicity, skin sensitisation) to support TSCA new chemical evaluations.

WebTEST2.0 as a model registration and model development platform

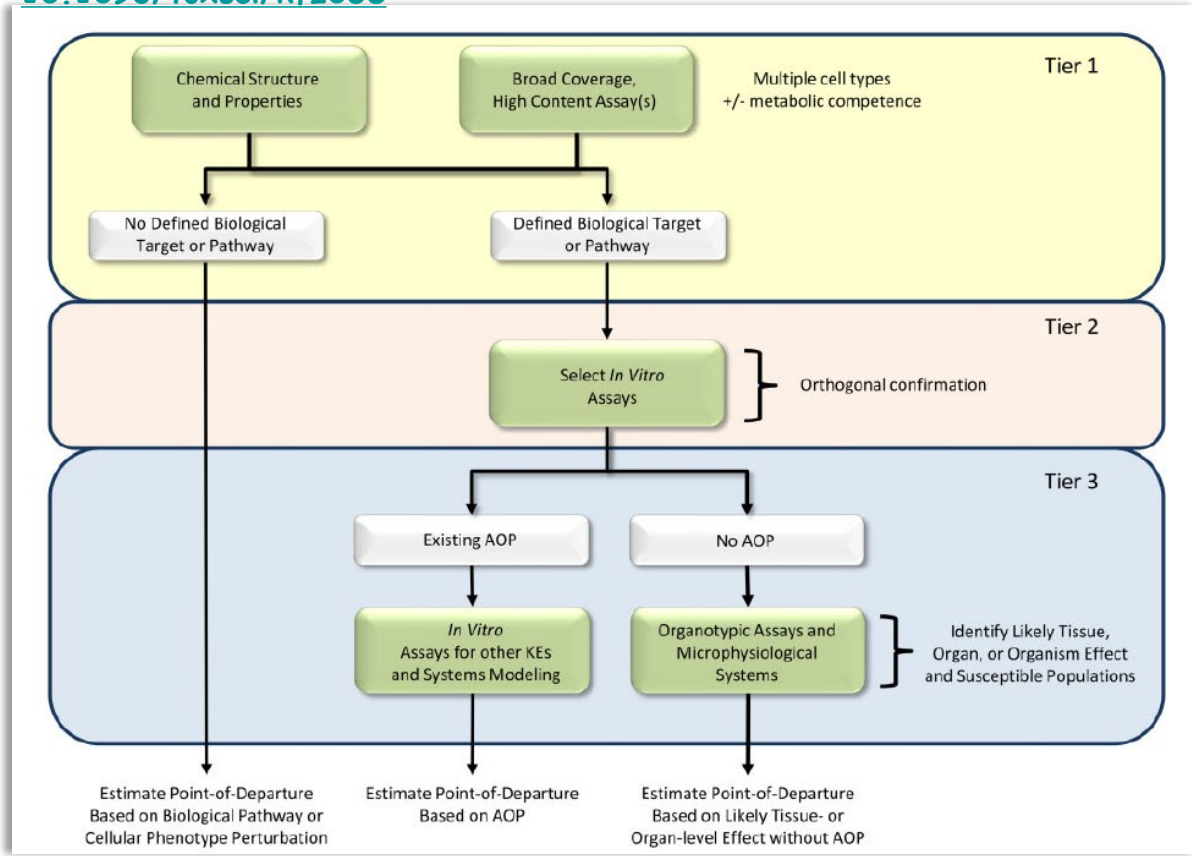
- Add externally developed models via web services
  - Models developed outside of the WebTEST platform will be implemented via Docker containers or via API calls to external webservices.
  - OPERA, EPI Suite, and WebTEST1.0 models will be incorporated into WebTEST2.0 via webservices.
  - Additionally, bioactivity-based models for estrogen receptor (Judson et al., 2017; Judson et al., 2015), androgen receptor (Judson et al., 2020; Kleinstreuer et al., 2017), steroidogenesis (Haggard et al., 2018; Haggard et al., 2019), and potentially other bioactivities based on *in vitro* NAM data, will be included in the WebTEST2.0 model registration platform.
  - Registration of all models, regardless of their development within or outside of the WebTEST platform, will include meta-data on the input features used in the modeling, the model output, and version information about that model; this constitutes an important goal for WebTEST2.0 and for rapid integration of information from disparate sources for next generation risk assessment.

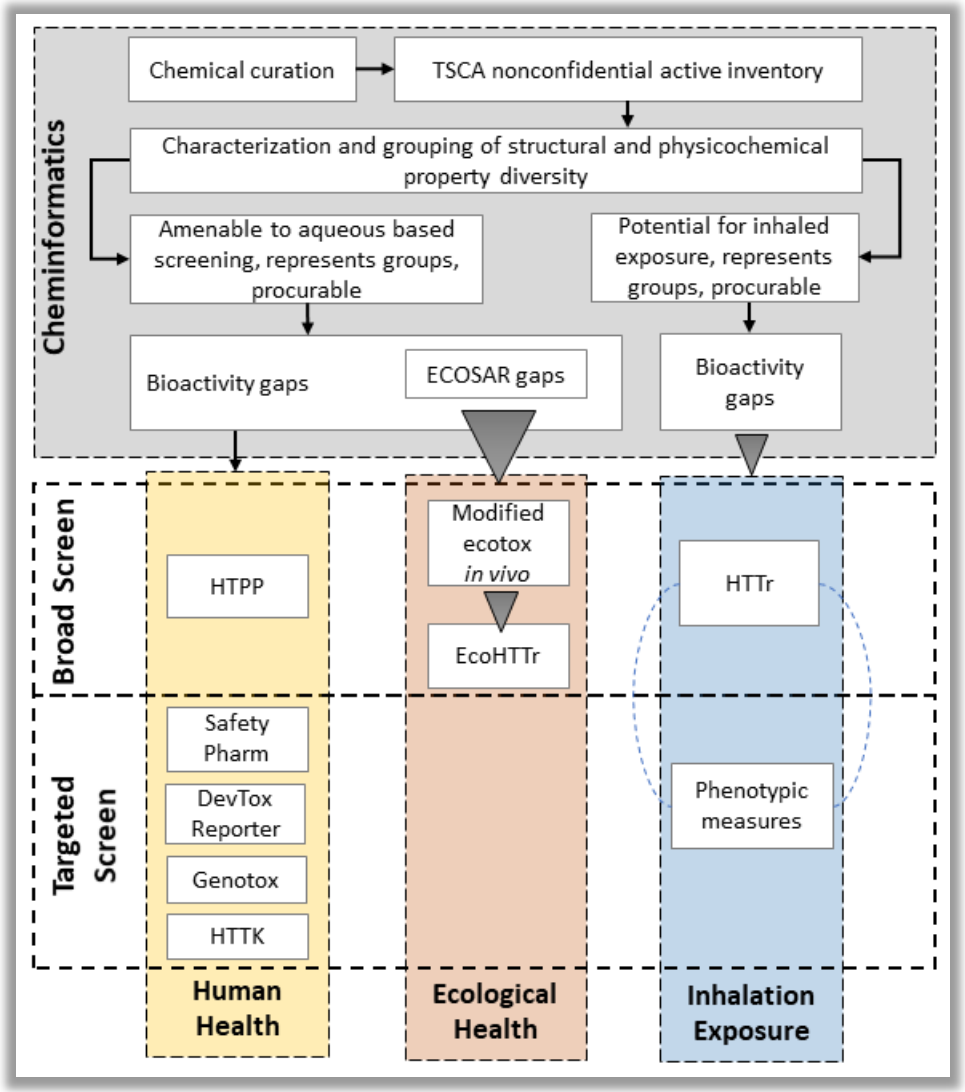
Research Area 4

	Research Area	Challenge	Approach	Expected Outcome(s)
4	Explore Ways to Integrate and Apply NAMs in New Chemical Assessments	Reduction in the use of vertebrate animals in accordance with TSCA Section 4(h) Many PMN submissions are data poor Amended TSCA requires affirmative determination regarding unreasonable risk	Develop and evaluate a suite of *in vitro* NAMs for informing new chemical evaluations Use mechanistic and toxicokinetic *in vitro* NAMs to inform and refine chemical categories in Research Area #1	A suite of NAMs that could be used by external stakeholders for testing and data submissions under TSCA as well as informing and expanding new chemical categories

Research Area 4 will combine broad and targeted screens to inform estimates of a bioactivity-based point-of-departure and address specific biology

Thomas et al. 2019
10.1093/toxsci/kfz058

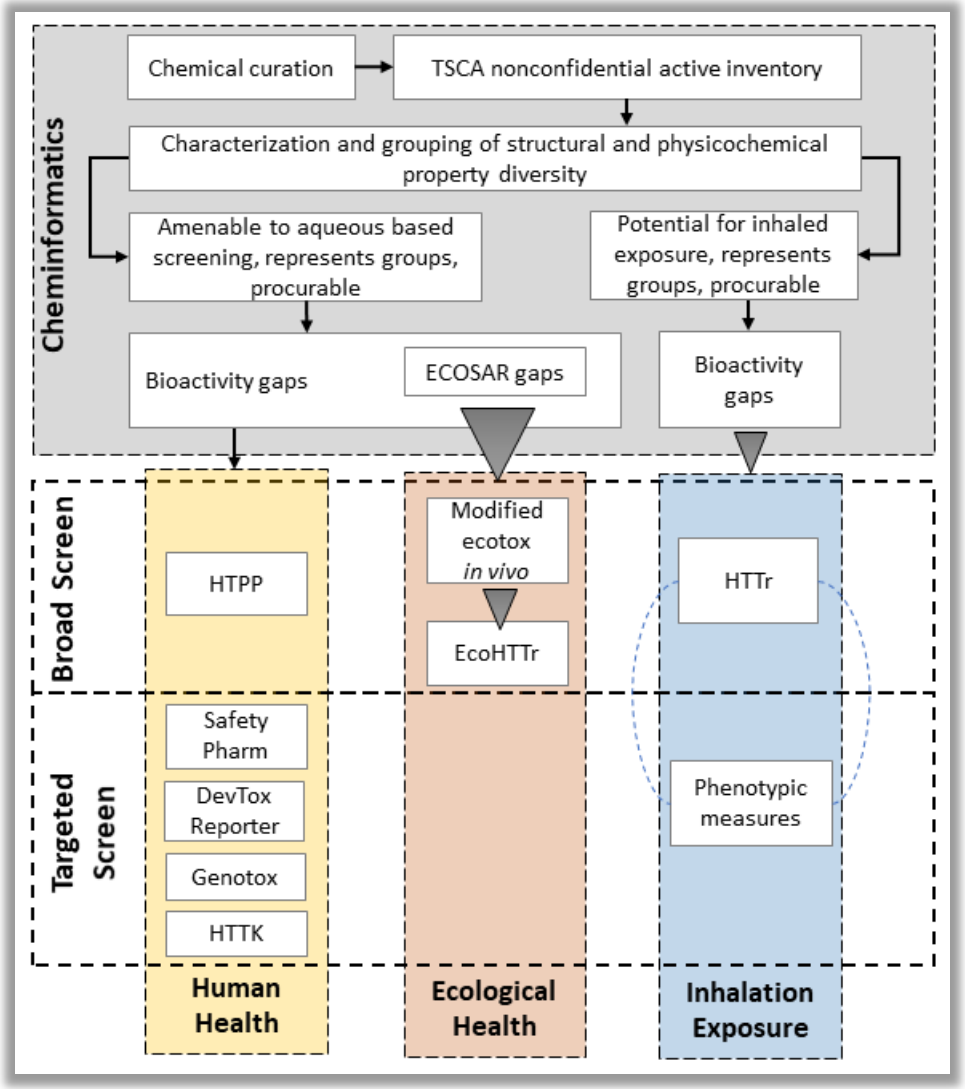




In Research Area 4, ORD will collect *in vitro* NAM data to demonstrate how NAMs for bioactivity and toxicokinetics can be used in a NAM-informed assessment of data-poor chemicals.

Cheminformatics and reference chemical knowledge will drive the selection of 200-300 chemicals for an initial case study

- In a first step, ORD will focus on development of a dataset for 200-300 chemicals, including some reference chemicals as well as TSCA-relevant chemicals from the nonconfidential inventory, to increase scientific confidence in application of this suite of bioactivity NAMs for informing chemical safety.
- These data will be needed to evaluate performance of these NAMs for further application and may also inform evolving frameworks for using multiple data streams to inform bioactivity-based dose-response assessment and hazard identification. .



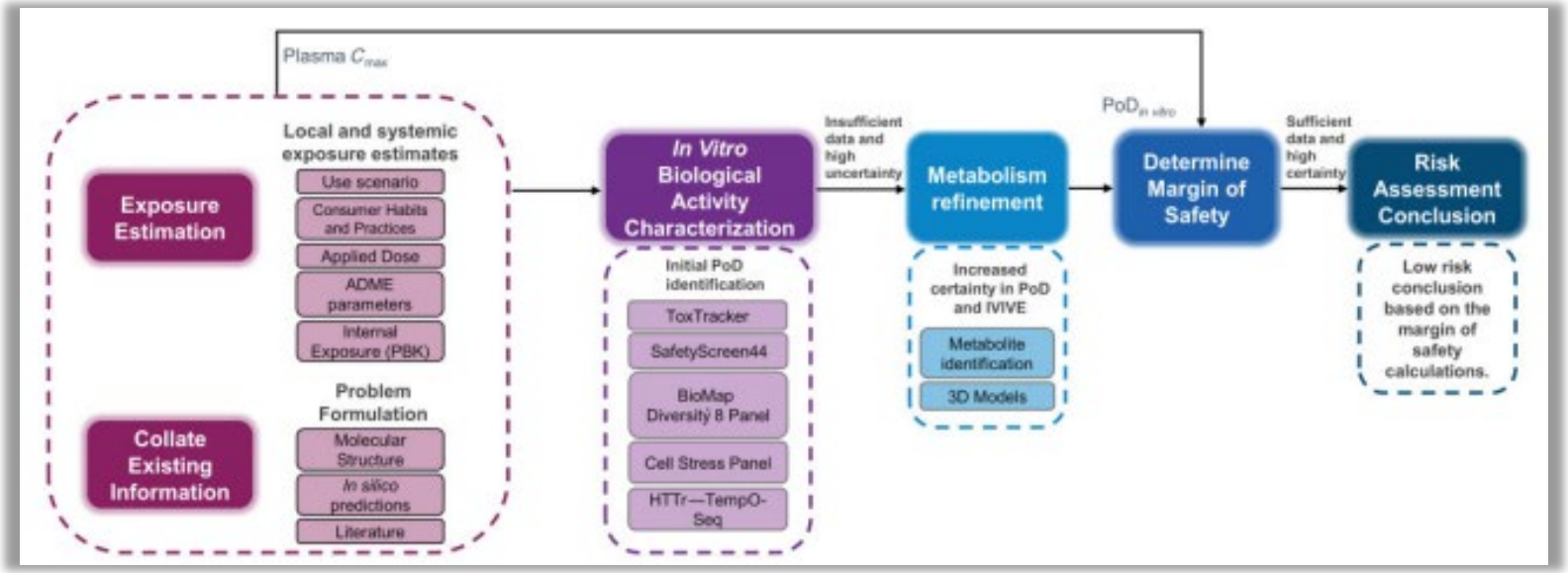
See poster by Patlewicz et al

Research Area 5

	Research Area	Challenge	Approach	Expected Outcome(s)
5	Develop a TSCA New Chemicals Decision Suppor t Tool to Modernize the Process	Searching, collating, and integrating data for new chemical assessments is inefficient and costly	Build proof of concept software workflow that integrates all data streams in a new chemical risk decision context	A decision support tool that will efficiently integrate all the data streams (*e.g.*, chemistry, fate, exposures, hazards) into a final risk assessment and transparently document the decisions and assumptions made. This will facilitate the new chemicals program tracking decisions over time and evaluating consistency within and across chemistries.

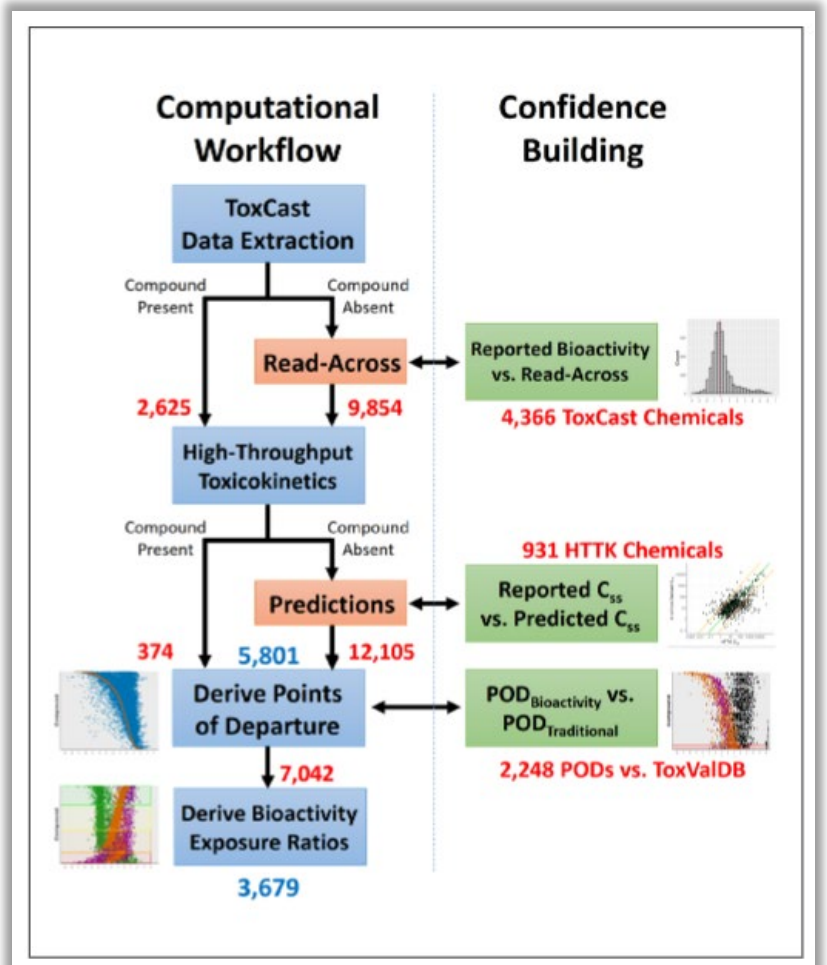
Examples of next generation risk assessment workflows that incorporate NAMs

Baltazar et al. 2020, "A Next Generation Risk Assessment Case Study for Coumarin in Cosmetic Products." 10.1093/toxsci/kfaa048

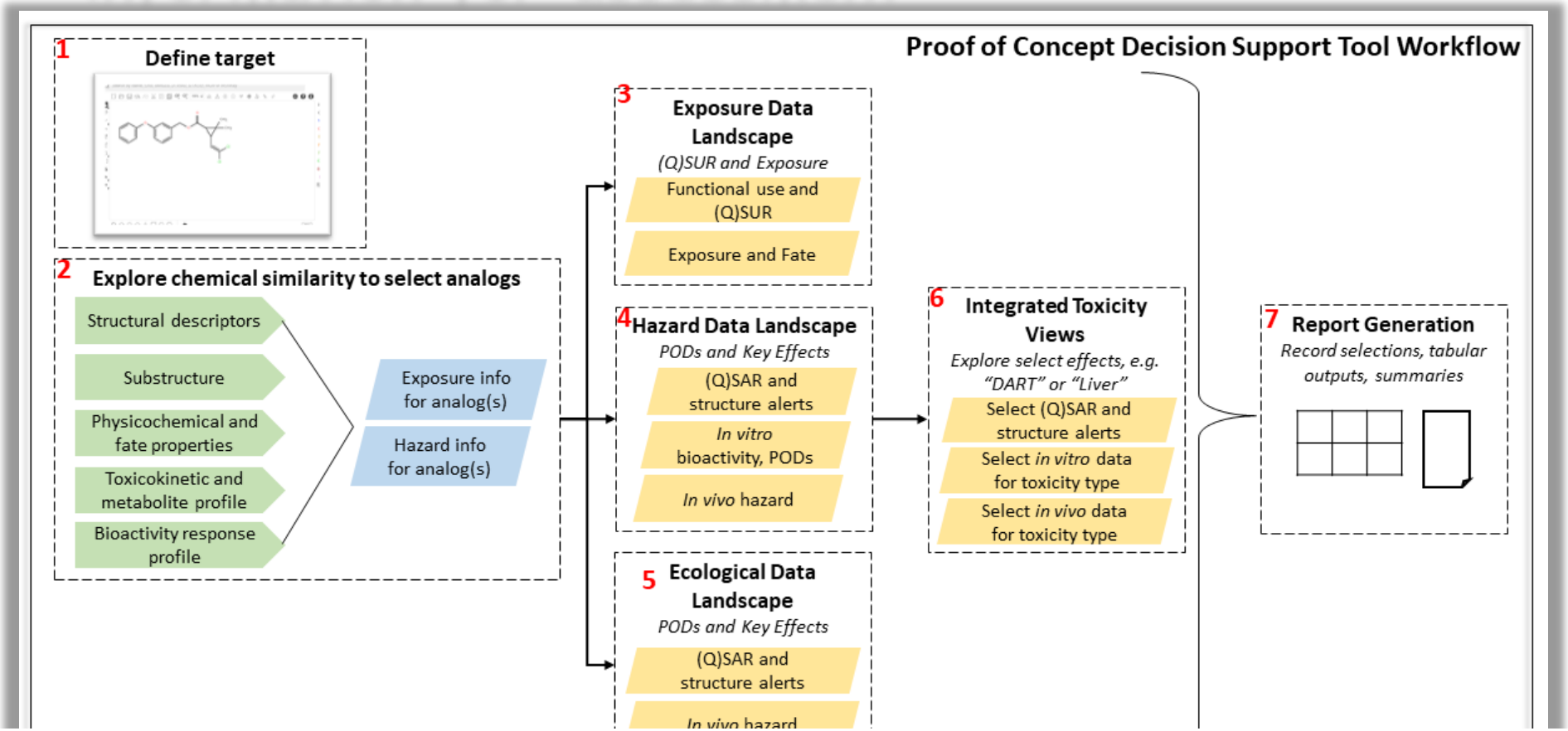


- Customised, modular decision support tool will be needed
- Blend existing data and models with NAMs
- Software that allows user to interact with data, make selections, and record these selections

Beal et al. 2020, "Implementing in vitro bioactivity data to modernize priority setting of chemical inventories." 10.14573/altex.2106171

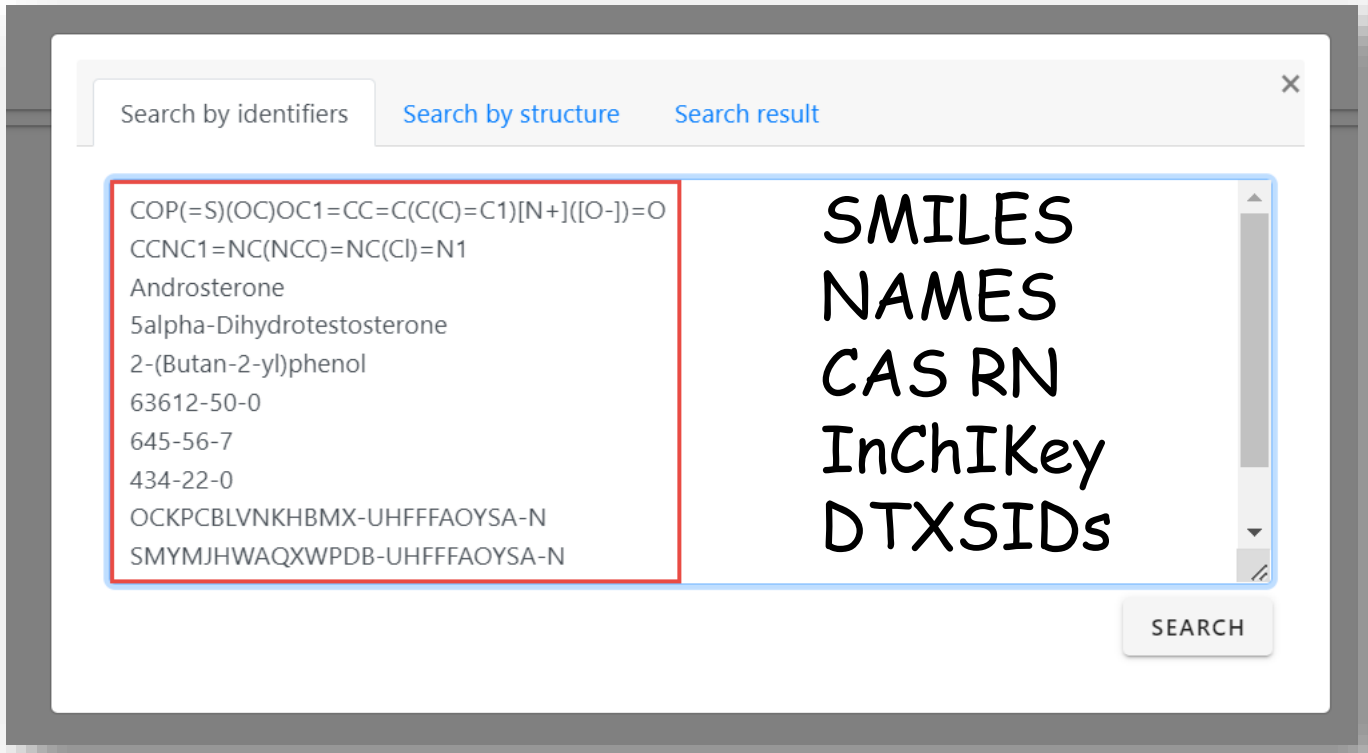


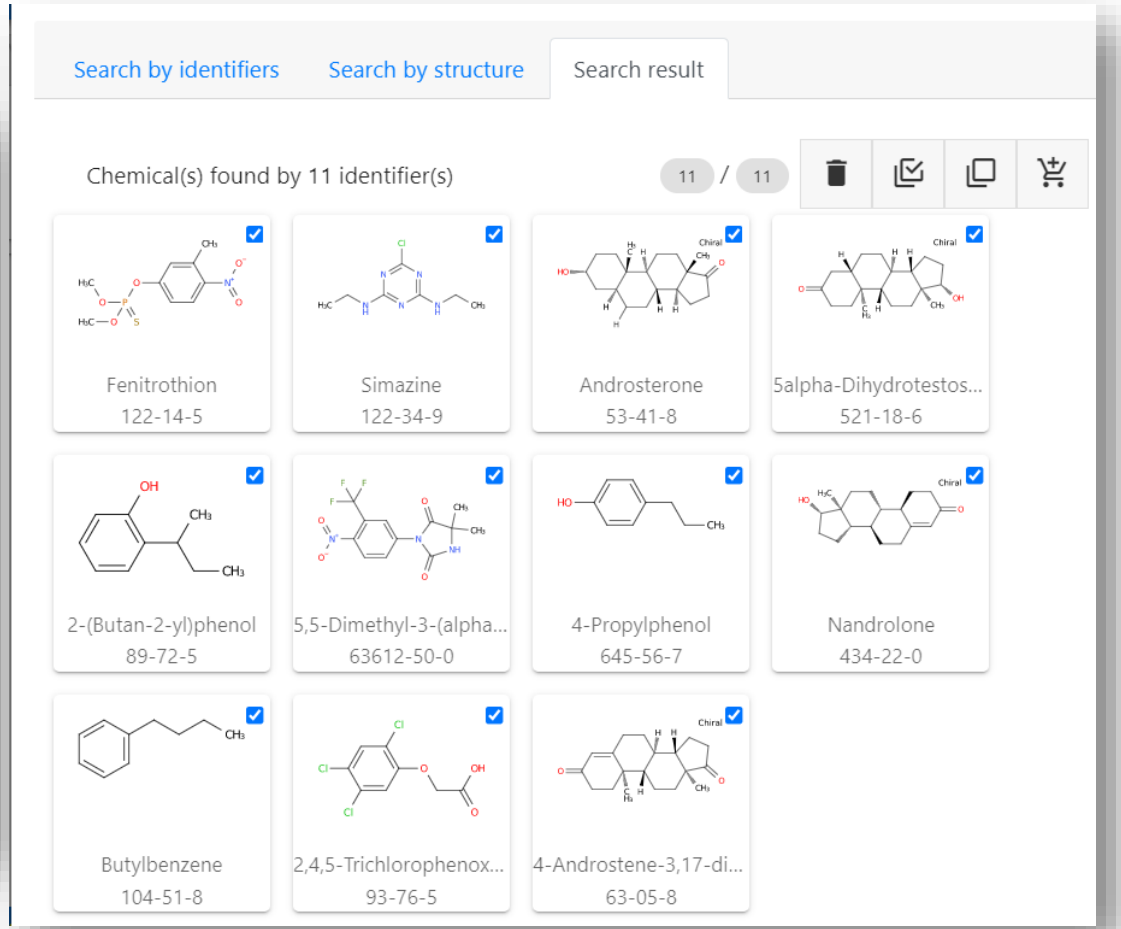
Components of a proof-of-concept decision support tool may include modules to display, select, and download key information for assessment



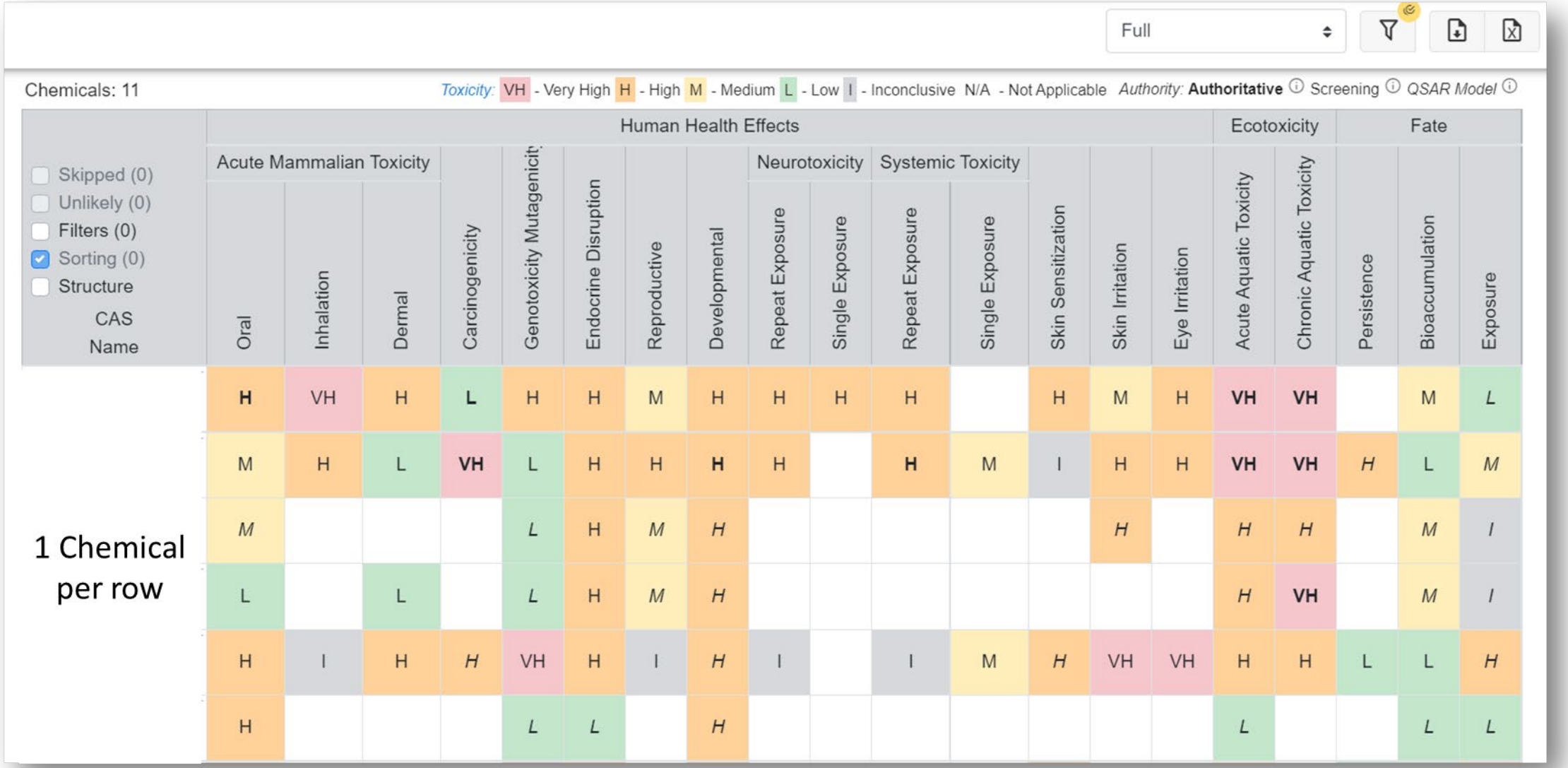
Overall, this collaboration to support digitisation, integration, and ultimate conversion to IUCLID-compatible formats will support collation of these data for utilisation by applications, such as a decision support tool.

Cheminformatics PoC Modules could be adapted to address NCCRP tools

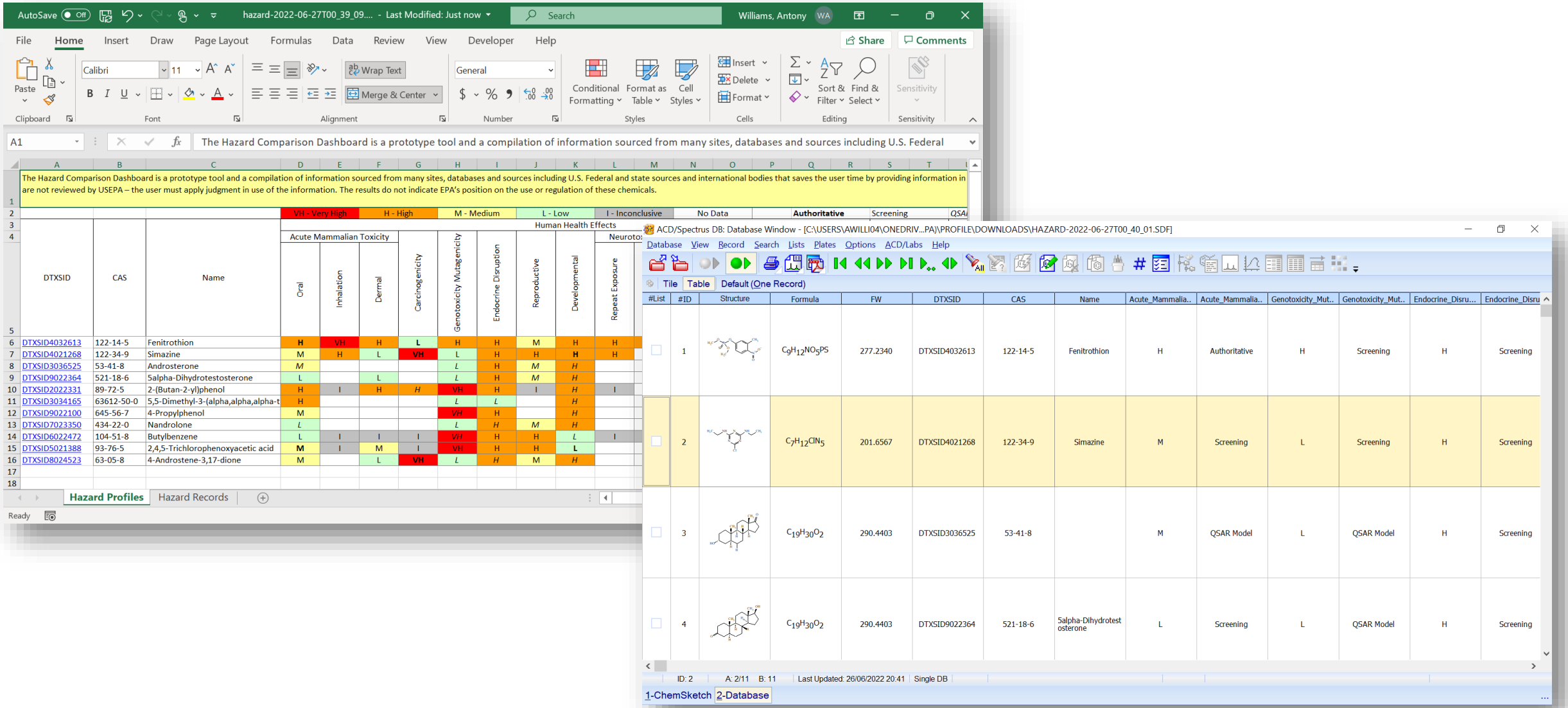




From visualising the hazard profiles…



Easy export of all data to Excel or SDF



Work of the NCCRP has begun in earnest

- Research during the 2023-2026 timeline will provide a foundation for continued improvement to meet the needs of new chemicals evaluation.
- Complement the EPA NAMs WorkPlan:
  - modernise available approaches, including decision support tools, for new chemicals evaluation
  - impact the engineering of the databases, models, and tools that ORD is building for multiple stakeholders to execute the vision of the CompTox BluePrint (Thomas et al., 2019) and the EPA NAMs WorkPlan (USEPA, 2021b)
- Achieve common goals:
  - greater acceptance and scientific confidence in NAMs applied within the NCCRP;
  - greater understanding of the future needs of NAM development; and
  - decision support tools that provide consistent, but iteratively improving, access to and integration of myriad data sources with chemical information, including data derived from NAMs
- Build external partnerships:
  - OPPT, ORD, and NIH (DTT/NIEHS, NICEATM, and NCATS)
  - Other regulatory partners, such as ECHA

EPA Core Planning Team

OCSPP	ORD	OGC
Madison Le	Annette Guiseppi-Elie	Sonja Rodman
Shari Barash	Russell (Rusty) Thomas	Don Sadowsky
Jeff Dawson	Katie Paul Friedman	Scott Albright
Anna Lowit	Alison Harrill	Victoria Clarke
Louis Scarano	Samantha Jones	Chris Kaczmorek
Tala Henry	Beth Owens	Devi Chandrasekaran
Keith Salazar		
Meghan Tierney		
Stan Barone		
Kellie Fay		
Martin Phillips		
Keith Avery		

Project Team..

Katie Paul Friedman
Kristin Isaacs
Caroline Ring
Katherine Phillips
John Wambaugh
Antony Williams
Grace Patlewicz
Todd Martin
Nate Charest
Dale Hoff
Brett Blackwell
Richard Kolanczyk
Risa Sayre
Norm Adkins
Madison Feshuk

Shaun McCullough
Mark Higuchi
Dan Villeneuve
Kevin Flynn
Chad Deisenroth
Joshua Harrill
Richard Judson
Sean Watford
Jill Franzosa
Tim Buckley
Sid Hunter
Peter Egeghy
Michael DeVito
Jonathan Wall

And more…