United States
Consumer Product Safety Commission

February, 2025

# CPSC Staff Statement on: Guidance Document for Conducting Qualitative Class-Based Hazard Assessment of Organohalogen Flame Retardants

The U.S. Consumer Product Safety Commission (CPSC or Commission) contracted with ICF (Contract BPA No. 61320622A0005, Order No. 61320622F2011) to identify approaches to class-based hazard assessment, to complete a guide that can be used to conduct class-based hazard assessments, and to complete a case study to apply the approaches documented in the guide.

This statement was prepared by the CPSC staff. ICF produced for CPSC staff the accompanying report: "Guidance Document for Conducting Qualitative Class-Based Hazard Assessment of Organohalogen Flame Retardants" (the Guide). The statement and report have not been reviewed or approved by, and may not represent the views of, the Commission.

The National Academies of Science, Engineering and Medicine (NASEM) assigned the known OFR chemicals and analogs to 14 subclasses of organohalogen flame retardants (OFRs).[1] CPSC staff subsequently updated the list of OFRs and subclass assignments.[2] ICF reviewed the identified analogs for all 14 OFR subclasses, identified metabolites of interest for all 14 subclasses, and developed this read across process guide.[3]

The purpose of the Guide is to describe a process to conduct qualitative class-based read across for OFRs, focusing on evaluation of the grouping of chemicals as a subclass. This will help provide an updated list of the chemicals in a subclass, along with identified analogs and metabolites, and a statement of confidence for each OFR belonging to the subclass. The Guide also addresses the NASEM subclass assignments by reconsidering their classification using an end-goal paradigm of grouping the chemicals by their health effects or other properties suggestive of similar bioactivity instead of primarily focusing on their structure.

The Guide is divided into three parts: identification and description of selected read across tools; subclass evaluation steps to develop and implement objective criteria for evaluation of the read across predictive results; and guidance to evaluate the quality of those read across predictions against existing empirical data. The subclass evaluation steps are organized into five profiles: 1) physicochemical; 2) mechanistic; 3) toxicokinetic; 4) metabolic; and 5) adverse effects. Under the same contract order, ICF applied the Guide to the polyhalogenated organophosphate (PHOP) subclass in the associated report titled, "Class-Based Qualitative Hazard Assessment of Polyhalogenated Organophosphate (PHOP) Flame Retardants."[4]

---

[1] National Academies of Sciences, Engineering, and Medicine 2019. A Class Approach to Hazard Assessment of Organohalogen Flame Retardants. Washington, DC: The National Academies Press. https://doi.org/10.17226/25412. Available at: http://nap.edu/25412.

[2] Available at https://doi.org/10.1038/s41597-022-01351-0.

[3] Read-across is a method to predict a property/endpoint for a substance of interest (target substance) from known information on the same property/endpoint from a 'similar' substance (source analogue) usually based on structural similarity (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6820193/).

[4] Available at https://www.cpsc.gov/Business--Manufacturing/Organohalogen-Flame-Retardant-Chemicals-Assessment.

# Guidance Document for Conducting Qualitative Class-Based Hazard Assessment of Organohalogen Flame Retardants

Call Order No. 61320622F2011
CPSC BPA No. 61320622A0005
April 29, 2024

Submitted To:
U.S. Consumer Product Safety Commission
5 Research Place
Rockville, MD 20850

Submitted By:
ICF
1902 Reston Metro Plaza
Reston, VA 20190 USA

# Table of Contents

# Tables



# Figures

# Acronyms and Abbreviations

| Acronym / Abbreviation | Term/Definition |
|---|---|
| ADME | absorption, distribution, metabolism, and excretion |
| AI | artificial intelligence |
| AOP | adverse outcome pathway |
| API | application programming interface |
| ATSDR | Agency for Toxic Substances and Disease Registry |
| CAS RN | Chemical Abstracts Service Registry Number |
| cHTS | curated high throughput screening |
| CPSC | Consumer Product Safety Commission |
| CTD | Comparative Toxicogenomics Database |
| CTS | Chemical Transformation Simulator |
| CYP | cytochrome P450 |
| DNEL | derived no-effect level |
| DTT | Division of Translational Toxicology, of the National Institute of Environmental Health Sciences (NIEHS) (formerly DNTP – Division of the National Toxicology Program) |
| ECHA | European Chemicals Agency |
| EFSA | European Food Safety Authority |
| EPA | U.S. Environmental Protection Agency |
| httk | high throughput toxicokinetics |
| IARC | International Agency for Research on Cancer |
| ICE | Integrated Chemical Environment |
| IPCS | International Programme on Chemical Safety |
| IUCLID | International Uniform Chemical Information Database |
| JECFA | Joint FAO/WHO Expert Committee on Food Additives |
| JMPR | Joint Meeting on Pesticide Residues |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| LASSO | least absolute shrinkage and selection operator |
| MACCS | Molecular ACCess System |
| MIE | molecular initiating event |
| NAM | new approach methodologies |

| Acronym / Abbreviation | Term/Definition |
|---|---|
| NASEM | National Academies of Science, Engineering, and Medicine |
| NIH | National Institutes of Health |
| NIOSH | National Institute for Occupational Safety and Health |
| NTP | National Toxicology Program |
| OECD | Organisation for Economic Co-operation and Development |
| OEL | occupational exposure levels |
| OFR | organohalogen flame retardant |
| OPERA | Open (Quantitative) Structure-activity/property Relationship Application |
| PBK | physiologically based kinetic |
| PBTK | physiologically based toxicokinetic |
| PCA | principal component analysis |
| PECO | Population, Exposure, Comparator, Outcome |
| PHA | polyhalogenated alicycles |
| PHACbx | polyhalogenated aliphatic carboxylates |
| PHACh | polyhalogenated aliphatic chains |
| PHBA | polyhalogenated benzene alicycle |
| PHBzAF | polyhalogenated benzene aliphatics and functionalized |
| PHB | polyhalogenated benzene |
| PHBAF | polyhalogenated bisphenol aliphatics and functionalized |
| PHC | polyhalogenated carbocycle |
| PHDE | polyhalogenated diphenyl ether |
| PHOP | polyhalogenated organophosphate |
| PHPhAE | polyhalogenated phenol aliphatic ether |
| PHPhD | polyhalogenated phenol derivative |
| PHPBI | polyhalogenated phthalates/benzoates/imide |
| PHT | polyhalogenated triazine |
| QC | quality control |
| QSAR | quantitative structure-activity relationship |
| QSPR | quantitative structure–property relationship |
| SMARTS | SMiles ARbitrary Target Specification |
| SMILES | Simplified Molecular Input Line Entry System |
| t-SNE | t-Distributed Stochastic Neighbor Embedding |

# 1. Introduction

In 2015, several organizations petitioned the Chemical Product Safety Commission (CPSC) to regulate the use of organohalogen flame retardants (OFRs) in certain consumer products due to their potential toxicity (CPSC, 2017). In response to this petition, the National Academies of Sciences, Engineering, and Medicine (NASEM) published a plan to assess OFRs as a class for potential chronic health hazards (NASEM, 2019). Rather than evaluating the OFRs as a single class, the NASEM plan focuses on evaluating the toxicity of OFRs as 14 different subclasses based on chemical structure, physicochemical properties, and predicted biological activity. As part of their 2019 report, NASEM identified 161 OFR chemicals and 1,073 analogs. The NASEM's publication set the foundation for class-based read across methods for OFRs, and this guide builds on the concepts that NASEM introduced. The purpose of this guide is to describe a process developed for conducting qualitative class-based read across for OFRs, focusing on evaluation of the grouping of chemicals as a subclass.

In 2022, CPSC published an article titled "Development of a Flame Retardant and an Organohalogen Flame Retardant Chemical Inventory", which included a downloadable inventory of chemicals used as flame retardants. This inventory was developed by compiling diverse sets of publicly available data sources from governmental organizations and open literature (Bevington et al., 2022). The goal of CPSC's flame retardant inventory was to describe the heterogeneity of compounds that can be used as flame retardants and changes in formulation chemistry over time. The inventory was developed as a resource for scientists interested in better understanding properties of flame retardant and OFR subclasses. Moreover, the 2022 CPSC inventory also expanded NASEM's 2019 OFR inventory; a summary of these changes is shown in Table 1. The present document was developed to guide the assignment of the chemicals in CPSC's OFR inventory to one (or more) of the 14 subclasses identified by NASEM and provide a preliminary qualitative assessment. The methods and decision logic described in the guide are sufficiently general for use with all 14 subclasses of OFRs, but illustrative examples based on the polyhalogenated organophosphates (PHOPs) subclass are included from an accompanying case study (Bevington et al., 2022).

**Table 1. Number of Chemicals in the 14 OFR Subclasses**

| OFR Subclass | Subclass Abbreviation | NASEM Original Number of Chemicals[1] | CPSC Expanded Number Chemicals[1] |
|---|---|---|---|
| Polyhalogenated alicycles | PHA | 17 | 22 (22)[2] |
| Polyhalogenated aliphatic carboxylates | PHACbx | 4 | 3 (3) |
| Polyhalogenated aliphatic chains | PHACh | 12 | 47 (20) |
| Polyhalogenated benzene alicycles | PHBA | 4 | 4 (4) |
| Polyhalogenated benzene aliphatics and functionalized | PHBzAF | 19 | 20 (20) |

| OFR Subclass | Subclass Abbreviation | NASEM Original Number of Chemicals[1] | CPSC Expanded Number Chemicals[1] |
|---|---|---|---|
| Polyhalogenated benzenes | PHB | 19 | 50 (38) |
| Polyhalogenated bisphenol aliphatics and functionalized | PHBAF | 11 | 14 (14) |
| Polyhalogenated carbocycles | PHC | 15 | 21 (21) |
| Polyhalogenated diphenyl ethers | PHDE | 12 | 223 (58) |
| Polyhalogenated organophosphates | PHOP | 22 | 42 (42) |
| Polyhalogenated phenol aliphatic ethers | PHPhAE | 9 | 11 (11) |
| Polyhalogenated phenol derivatives | PHPhD | 7 | 8 (8) |
| Polyhalogenated phthalates/benzoates/imides | PHPBI | 11 | 19 (19) |
| Polyhalogenated triazines | PHT | 6 | 6 (6) |
| Total | - | 161 | 488 (278) |

[1]Two OFR chemicals were categorized by using two chemotypes and included in two subclasses (polyhalogenated benzene aliphatics and functionalized and polyhalogenated carbocycles).
[2]The number in parentheses is the number of chemicals in the subclass after removing mixtures and uncommonly reported congeners.

This guide also addresses the subclass assignments suggested by NASEM by reconsidering their classification using an end-goal paradigm of grouping the chemicals by their health effects or other properties suggestive of similar bioactivity instead of primarily focusing on their structure. The same process used to evaluate the grouping of chemicals is also used to gather the information required to make a class-based read across assessment. Many OFRs lack comprehensive hazard data, and class-based read across can help bridge these gaps by extrapolating information from chemically similar substances within a subclass or their analogs. A class-based approach assesses multiple chemicals at one time, using the data available for data-rich members of the subclass or structural analogs and extrapolating to data-poor members in the subclass.

Familiarity with read across and the computational tools described herein is required to use this guide. Specific instructions on how to use any of the tools are not provided and can be found at the tool source URL. The source URLs for all tools and databases described in this guide are listed in Appendix C. This guide does not cover determining whether there are sufficient data and modeling approaches to conduct an assessment for a group of chemicals; this is already addressed for each of the 14 subclasses of OFRs through their scoping documents published by CPSC (CPSC, n.d.).

## 1.1. Guide Organization

The remainder of this guide is split into three parts that should be executed consecutively. Getting Started (Section 2) provides steps to consider before getting started on the assessment. It is crucial to follow these steps prior to conducting the assessment, in order to ensure accuracy and quality control while managing changes and additions to data for many chemicals at once, and for evaluating whether the subclass is sufficiently structurally similar for a class-based assessment.

The subclass evaluation steps (Section 3) are organized into five profiles:

1. A physicochemical profile to determine physiochemical similarities or differences.
2. A mechanistic effect profile based on existing in vitro data and/or in vivo data to build an understanding of common bioactivity and to define subclass in the absence of animal data, if necessary.
3. A toxicokinetic profile using modeling tools that can provide insight into the toxicological relevance of chemicals.
4. A metabolite profile to determine metabolic similarities and support adverse effect conclusions of data poor OFRs based on similar metabolite profiles shared with data rich OFRs.
5. An adverse effect profile based on in vivo data, epidemiological data, and New Alternative Methods (NAMs) to determine the adverse effects of the subclass.

The five profiles provide an organized grouping of information to evaluate chemical classification and hazard outcomes. The order of the profiles is not necessarily a recommended order of operations. Profiles should be built in an order appropriate to the chemicals and available information and tools. Each profile section will provide examples of tool or database choices that could be used to build the profile. This guide is not meant to be a review of all available tools; select examples were chosen to be presented based on CPSC's priorities and the OFRs' chemistries. Suggested criteria for tool selection, as well as strengths and weaknesses of available tools. Are provided, when appropriate.

Guidance on collating the information from the initial structural evaluation and five profiles is then provided at the end of the guide (Section 4). This guide and the accompanying case study provide a process for determining probable subclass hazard(s) by using the available data. Additional steps for extracting and evaluating effect data in the mechanistic and adverse effect profiles are required for hazard conclusions, as discussed in several of the later sections. A suggested framework for evaluating confidence of subclass inclusion for each chemical is also outlined, which can be used to inform confidence in the overall conclusions of the probable

subclass hazard(s). The described process, with some modifications, may also be useful for classes of chemicals other than OFRs.

# 2. Getting Started

## 2.1. Quality Control

A standardized method of chemical identification and tracking must be employed to ensure quality control of the data and consistency across steps when working with many chemicals and many pieces of information connected to each chemical. Chemical lists for class-based read across should be standardized by applying quantitative structure-activity relationship (QSAR) ready Simplified Molecular Input Line Entry System (SMILES). However, there is no universal standard for writing a SMILES string, and two different software programs or code strings may generate distinct SMILES representations for the same chemical. Therefore, consistent use of the same code for developing SMILES must be applied throughout the entire evaluation and assessment. After standardization, chemical lists should be deduplicated by SMILES string to ensure uniqueness, and all subsequent information should be aligned with the SMILES. Additionally, any newly introduced chemicals from diverse sources should always be checked to ensure compliance with the chosen SMILES standard.

A workflow using RDKit and implemented in Python using the in-house library CompDesc (https://test.pypi.org/project/CompDesc/) was developed for applying this guide. This package mimics the QSAR-ready preparation workflow developed by Mansouri et al. (2024) and Mansouri et al. (2016). This cleaning process removes chemical mixtures, counter ions, eliminates hydrogen atoms, and does not discriminate by stereochemistry. The resulting structures can be referred to as QSAR-ready structures, given their compatibility with QSAR and are referred to as “cleaned SMILES” or “cleaned structures” herein. A detailed description of how to clean and manage chemical lists is provided in Appendix A.

## 2.2. Initial Chemical Classification

Read across predominantly relies on the concept of structure-activity relationships, wherein similar chemicals exhibit similar bioactivity. For a class-based assessment, the initial classification of the target chemicals involves an expert-driven process primarily based on their chemical structures. One outcome of the class-based read across process is an evaluation of whether chemicals can be grouped together for hazard assessment, but this process is dependent on the initial grouping being representative of the target subclass.

Before conducting any analysis, the first step is to validate the initial grouping by determining whether the initial structural classification is suitable for assessment. This is done by evaluating both inter- and intra-subclass similarity; effective assessment is feasible when different subclasses exhibit significant differences from each other, and chemicals within each subclass share adequate similarity. Employing computational clustering allows for quantifying intra- and inter-subclass similarity, emphasizing the need for structural differences for subclass recognition

and high similarity for effective inter-subclass comparisons. Visual inspection of the chemical structure aids in intuitive understanding and should be used for this step.

The CPSC Inventory of NASEM's divided the OFRs into 14 subclasses (NASEM, 2019) (Table 1) which provide initial structural groupings for OFRs. NASEM delineated different structural components (such as triazines and solely aliphatic compounds) into different subclasses but did not further evaluate these subgroups by bioactivity or related classification. One way to assess the initial subclass grouping is hierarchal clustering of the subclasses using topological fingerprints and a Tanimoto distance, then visualizing this through a dendrogram (Figure 1). For any computational clustering based on chemical similarity, careful attention needs to be given to the methods used to represent chemicals (e.g., the type of fingerprint), as well as the metric used to compare two chemicals (e.g., Tanimoto distance) (see Appendix B). The representation and comparison methods should be appropriate for delineating differences between chemicals. For example, if the goal is to differentiate between small aliphatic compounds with carbonyls or alcohols, then fingerprints that differentiate between those groups are most appropriate. In contrast, a fingerprint that uses bond distance may be more appropriate if comparing large ring structures to small-functionalized chains.

Figure 1 is an example of two dendrograms created to evaluate the NASEM OFR subclass categories in the accompanying case study. Each dendrogram uses a pair-wise Tanimoto score to compute similarity but the dendrograms differ from each other by the fingerprint applied to describe the OFRs: PubChem fingerprints and RDKit's implementation of the Daylight-like fingerprint (Daylight Chemical Information Systems Inc., 2011b) based on hashing molecular subgraphs (Daylight Chemical Information Systems Inc., 2011a, 2011b; Landrum, n.d.). The RDKit topological fingerprints recognize structural subgraphs and provided up to 1,400 bits per chemical for the OFRs whereas the PubChem fingerprints do not use structure subgraphs but instead map atoms and structural substructures and had up to 120 bits per chemical for the OFRs. PubChem had less specificity than RDKit, but overall, they both show that for most subclasses, most of the OFRs in the subclass are grouped close to each other, with some individual chemical exceptions. This confirms that a class-based read across could be constructed for most OFR subclasses as they are currently defined by NASEM. Additionally, both types of fingerprints in Figure 1 provide a reasonable comparison of the OFR subclasses without significant differences in the overall grouping of subclasses.

On both dendrograms, OFRs in the polyhalogenated benzene alicycles (PHBA) subclass are spread across approximately 180º. This is an example of an initial classification of chemicals that may not be suitable for read across. Consideration should be given to evaluating those chemicals with the subclasses each chemical is most similar to in the dendrograms.

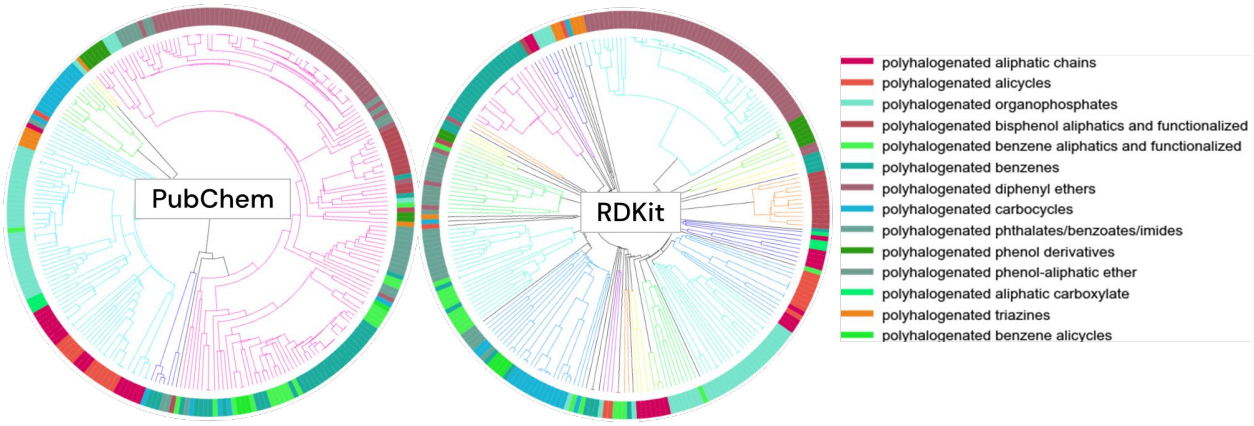


**Figure 1. Dendrograms comparing structural grouping using hierarchical clustering based on topological Fingerprint and Tanimoto score of CPSC OFR subclasses.**

Each subclass is defined by a unique color on the circumference, and each chemical is a slice matched to a unique inner branch line. Each computationed hierarchical cluster is a different colored set of branching within the circle. The branch colors inside the dendrogram are automatically generated by the hclust at a certain level of the tree and can be optimized, if desired.

## 2.3. Analog Search

An analog approach is commonly employed in read across to augment chemical subclasses for addressing data gaps. Analogs are generally selected based on their structural similarity to the target compounds because structurally similar compounds are expected to elicit reasonably similar toxic effects (OECD, n.d.).The selection of analogs is an important step in the read across process, and the quality of the analogs can significantly impact the quality of any computational predictions or subsequent conclusions, especially if analogs are the main anchor chemicals for making conclusions. Traditional, single compound read across typically involves considering several factors beyond structural similarity, including (but not limited to) physicochemical properties and kinetic properties, when available, to determine the best analog(s) to draw conclusions about the target chemical.

Finding analogs for class-based read across should follow a rational similarity approach, computed through a similarity score between two chemicals, following a similar approach to that classically used in read across to capture the overall chemical similarity. This method is more appropriate than a substructure search, for example, as developed in virtual high-throughput screening, where the search is based solely on the presence of substructures and does not provide information on the overall similarity between chemicals.

For structural computational analog searches, the following must be defined before the search:

- Chemical Database: Identifying the source from which analogs will be extracted.
- Chemical Representation: Determining the representation method, such as molecular descriptors, SMILES vectorization, or fingerprints.

- Similarity Metric: Choosing metric for similarity calculation, such as Tanimoto, Euclidean, or Manhattan distance.
- Similarity Criterion: Establishing a cutoff on the similarity metric to determine acceptable analog matches.

In a class-based read across approach for OFRs, it is likely that multiple chemicals within a subclass will exhibit varying degrees of structural dissimilarity. The criteria for selecting analogs within a subclass need to be less stringent compared to traditional read across methods for one chemical to accommodate the inherent structural diversity of the target subclass. Capturing maximum chemical diversity within the subclass minimizes the likelihood that analogs will only be captured for limited number of similar chemicals defining the subclass, therefore poorly representing the overall subclass. Consideration for the database, representation, similarity metric, and similarity criterion choices are addressed in the following subsections. Additional detail on similarity metrics and criterion is provided in Appendix B.

### 2.3.1. Chemical Databases for Analog Identification

The 2019 NASEM report identified >1,000 analogs for the 14 subclasses of OFRs using chemotypes available in the U.S. Environmental Protection Agency's (EPA's) Chemistry Dashboard. Initial data availability of these analogs was screened as part of CPSC's scope documents and associated evidence maps. This was completed prior to the expansion of the OFR inventory list by CPSC in 2022. Searching and screening new analogs related to newly identified OFRs in the expanded list was part of the purpose and scope of this Guide and the accompanying PHOP case study.

Choosing the chemical database for analog identification is an important decision and entirely dependent on the goals of the assessment. Table 2 summarizes 6 commonly used and publicly available chemical databases that have chemical similarity searching capabilities. This is not an exhaustive list; there are many other databases with tools for identifying analogs, both public and proprietary. It may be desirable to evaluate more than just these 6 databases for appropriateness to the target subclass prior to choosing one for identifying analogs. Table 2 highlights the 6 databases and their characteristics relative to the priorities and goals expressed by CPSC staff for choosing OFR analogs.

Criteria for a database choice should be prioritized, then used to evaluate databases for use. The highest priority criteria for class-based read across are the ability to batch search and the quantity of similar chemicals in the database. Batch searching is necessary since analogs for many chemicals will need to be identified. As the number of seed chemicals (i.e., number of chemicals used to identify analogs) increase, completing individual searches for each chemical becomes unreasonable. Also due to the nature of searching for many analogs for many chemicals at once, it is crucial that the database has a large set of chemicals of similar size and structure so that it can identify analogs for all seed chemicals. Other considerations include the type(s) of fingerprints available, the ability to search for analogs by characteristics other than structure

(e.g., metabolism, mechanistic pathways, environmental fate), if the application programming interface (API) can be queried or the database is easily accessible through user-friendly interface, if the fingerprints and/or chemical database is public or proprietary, and if toxicity data is already available in the database (therefore avoiding additional data searching steps). Multiple databases could also be considered for analog searches to compound benefits, but the advantage of using multiple databases should be weighed with the search and collation time required.

Some of the OFRs are rare structures and it is unlikely that all will be included in all the databases listed in Table 2. The database choice should include most of the OFR compounds from all subclasses to validate the database as a source for analogs. It is preferable, but not required, that the database is publicly available to establish transparency in any CPSC assessments wherever possible. However, proprietary options may be a better fit for some of the OFRs (or other target chemicals) and were considered in the initial evaluation of databases in the accompanying case study. Other considerations include time to compute and compatibility with the fingerprint choice. The fingerprints choice for the OFR analog search should be nonspecific enough to identify as many reasonable analogs as possible, but specific enough to sufficiently cluster each OFR subclass by structure. Nonproprietary fingerprints also are of higher desirability.

**Table 2. Publicly Available Databases With Tools for Identifying Analogs[1]**

| Database | Chemical information | Batch search capability | Human relevant toxicity data available | Similarity search method |
|---|---|---|---|---|
| PubChem | • >310 million chemicals and substances<br>• No cleaning of submissions; significant chemical duplication<br>• Outputs all available chemical identifiers<br>• Biased toward small molecular weight molecules | Yes, through API query | Yes | PubChem Fingerprint (publicly available) and a Tanimoto score |
| OECD QSAR Toolbox | • Includes >100,000 individual chemicals total, and >45,000 chemicals with human health hazard endpoint values<br>• Analogs can be searched using non-structural characteristics including metabolism, mechanism, and toxicity | Yes, user can install the underlying database | Yes | Based on substructure search |
| EPA CompTox Chemical Dashboard | • >1,200,000 chemicals and substances<br>• Includes data from multiple data sources<br>• Use of the GenRA tool interface can search analogs for a single chemical | Yes, through a python package (genra-py) (Shah et al., 2021) | Yes | Choice of different fingerprints from Morgan, Torsion, ToxPrint and a mix of three previous one with a Tanimoto score |
| NTP ICE Chemical Quest | • The underlying database include the EPA DSSTox database<br>• Can calculate structural similarity using fingerprints generated using Saagar fingerprints | No | Yes and evaluated for quality | Saagar proprietary fingerprint with a Tanimoto score, (Sedykh et al., 2021) |
| Chemspider | • >129 million chemical structures | Yes, through commercial API | No | Substructure matching and similarity based on a LASSO score |
| ZINC | • >230 million chemicals<br>Chemicals are primarily drugs and drug derivatives aimed at protein docking | No | No | DICE Fingerprint with a Tanimoto score, but also allows for a substructure search |

[1]Based on information acquired in April 2024.

Depending on the context and the familiarity with the target chemicals, a curated or restricted database might be preferred over a comprehensive one. For example, DSSTox (within the EPA CompTox Dashboard) could be favored when investigating specific toxicological endpoints of chemicals similar to current and past pesticides. Such chemicals are well-studied and heavily-regulated, and it is likely that many such chemicals have data in DSSTox (Grulke et al., 2019).

Conversely, for read across analysis involving less prevalent or less studied chemicals, an analog search in a larger database could be preferred, in order to capture the maximum number of analog possibilities and increase chances of finding relevant data. This latter situation is more appropriate for most of the OFR subclasses. The 2019 NASEM report and CPSC's scoping documents found that there are some data -rich OFR compounds in 11 of the 14 OFR subclass, but many OFRs in each subclass are data-poor or have no data available at all. This makes it important to cast a wide net for potential analogs with the end goal of finding additional data-rich chemicals with sufficient similarity to the seed OFR, meaning the chosen database should have an extensive library of chemicals to identify as potential analogs.

Database searching requires input criteria to search for similar chemicals to the target chemicals in the subclass of interest, as described in the following subsection. Once the analog search is complete, the database is likely to have many redundancies. A chemical may be represented multiple times—perhaps in a mixture, stereoisomers, or charged compounds with different counter ions. Database outputs need to be standardized to clean SMILES and deduplicated using the methods described in Section 2.1.

### 2.3.2. Similarity Search

The Tanimoto coefficient is a ratio ranging from 0 (no similarity) to 1 (identical) that is traditionally used in read across to determine similarity between two chemicals. Employing a Tanimoto score for the similarity metric requires defining chemicals by molecular representation, meaning that the chemical representation choice influences the similarity metric. Coupling molecular fingerprints with a Tanimoto metric is preferred for structural analog searches because fingerprints can be quickly computed from SMILES resulting in the rapid calculation of a Tanimoto score between two fingerprints. The quick computation grants the ability to screen a database containing several million chemicals within a reasonable time on standard computers. The fingerprint and Tanimoto coefficient combination is also advantageous because there are many types of fingerprints available and fingerprint choice can be fit to the target subclass. Furthermore, Tanimoto distance of ≥0.85 has been shown to increase the probability of shared biological activity between chemicals (Patterson et al., 1996).

For these reasons, analog searches for the OFRs should use the Tanimoto coefficient metric with a fingerprint choice that does a good job of distinguishing between subclasses. A similarity criterion should be tested at 0.80 to 0.90 (0.85 ± 0.05) to determine the coefficient that balances a high degree of similarity between parent and analog compounds while allowing sufficient room for diversity, but also not capturing overly diverse compounds. Since the OFRs are all halogenated and large (>200 g/mol), the degree of halogenation and molecular weight should be used to evaluate similarity metrics and establish the best value for the final analog search list.

Descriptions of other chemical representations and chemical similarity metrics are provided in Appendix B for reference.

## 2.4. Data Searching and Visualization Methods

The end goal of class-based assessments is to make toxicity conclusions using available data and/or to inform future toxicity testing if necessary. Finding the available data plays a critical role in a successful assessment. Data-rich chemicals are required as anchors for read across, and having more chemicals with data increases the assessment confidence for data poor or data absent chemicals. The following section describes two methods (using toxicological databases and a literature review, respectively) for determining data availability for subclasses of chemicals, their analogs, and their metabolites. Additional detail on issues specific to metabolites is present in Section 3.5.

Evidence maps visually report data availability by visually presenting the availability of toxicity data by chemical and by assay type. The numbers presented in evidence maps developed for the OFRs should not be interpreted as absolute measures of data availability (i.e., a review/assessment may incorporate results from a single study or many, and it is not always clear if databases are reporting duplicate data), but instead reflect relative amounts of data (i.e., Chemical A has more data than Chemical B). Prior to screening literature or collating database data, criteria for relevance need to be developed, which should include the metadata categories used to tag each study. Tags are then collated into evidence maps for use in an assessment, summarizing the number of references or database hits addressing the metadata combinations. For OFRs and their analogs and metabolites, metadata categories are chemical identity and assay types tested (e.g., acute toxicity, carcinogenicity).

### 2.4.1. Toxicological Database Downloads and Evidence Maps

This section describes the general process for identifying, searching, and processing data from toxicological databases, which is outlined in Figure 2. For full details on the process of downloading data from toxicological databases and assembling evidence maps, refer to the literature survey guide produced under contract number CPSC-D-17-0001, Task Order 61320621F1001 (RSC, 2022; see "Literature Survey Guide" at https://www.cpsc.gov/Business--Manufacturing/Organohalogen-Flame-Retardant-Chemicals-Assessment).

Data from toxicological databases can be used to characterize the overall availability of hazard data for chemicals. Downloading existing data from toxicological databases requires less time and effort relative to conducting a literature review, and database evidence maps should be prioritized for scoping data availability prior to conducting a literature review. Additional level of effort is saved if pre-existing methods (logic and code) exist to assemble data from multiple databases, as is the case with the existing R code from the literature survey used in the application of this guide for the accompanying case study on PHOPs. This advantage holds even when some edits to logic and code are needed to apply to new data, or to the same databases after they have been updated.

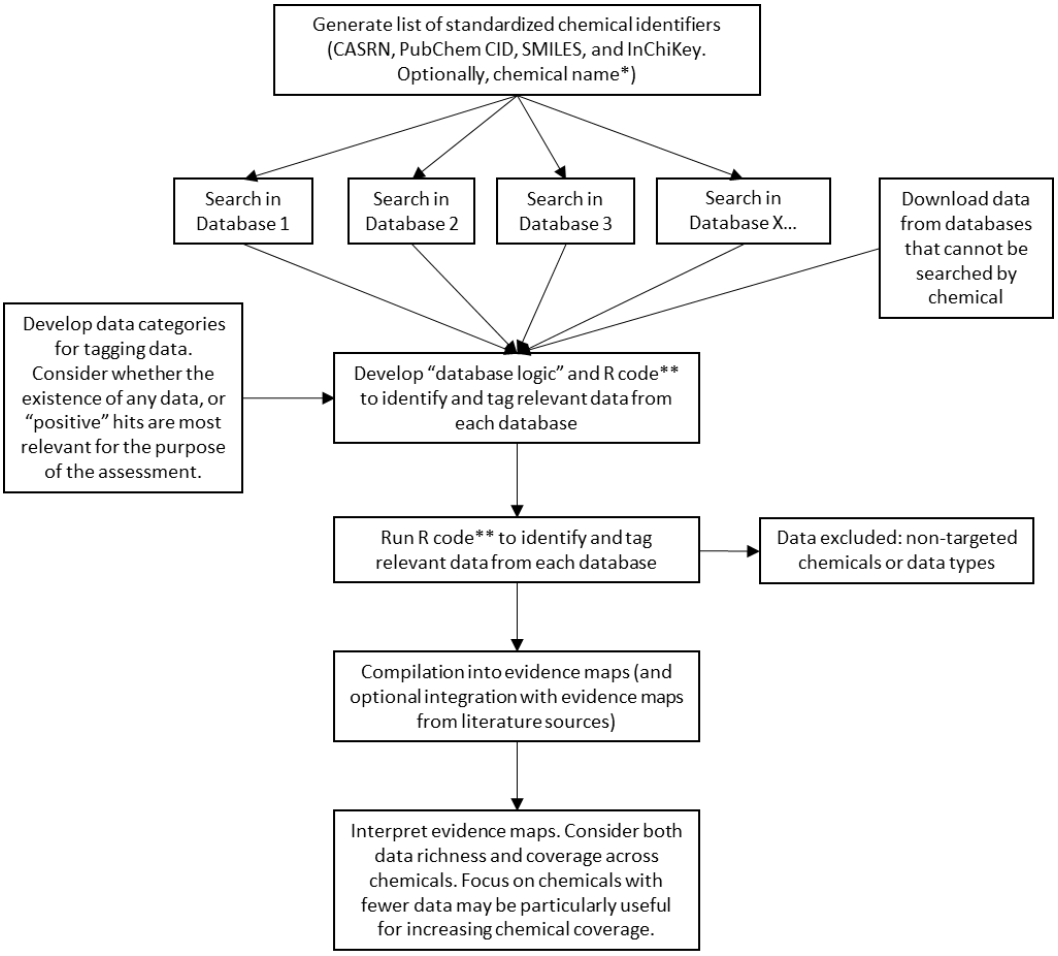


**Figure 2. Flow diagram showing the process for database searching and compiling data availability into evidence maps.**

*Chemical name is not usually used for searching databases, but is useful for human readability of output.
**Previously developed R code is to be used for application of this guide in the accompanying case study

Briefly, data from several primarily publicly available toxicological databases (Table 3) can be downloaded and subsequently summarized into evidence maps to show whether a chemical has data in a database (and what kind of data). Most toxicological databases can be searched by specific chemical identifiers and associated data can be downloaded. Other databases are provided as a single spreadsheet download that includes all chemicals in the database, whether they are chemicals of interest for a given assessment (e.g., the National Institute for Occupational Safety and Health [NIOSH] Pocket Guide). Data "hits" are counted for each database to be summarized in the evidence map. For the OFR assessment, hits are defined as reported data for a chemical, regardless of whether the data produced informative results (such as a positive or negative finding) or null results. For some assessment purposes, this higher level of data availability may be sufficient. For other purposes, it may be useful to invest additional time to distinguish between informative and null results or to only count informative results. The application of this case study using the aforementioned pre-existing methods does not include this additional step.

Many of these toxicological databases curate or otherwise collect data from multiple sources. Therefore, deduplicating results from different sources is needed. An existing R code described in the literature survey guide was developed by the University of Cincinnati to assemble data hits from across these databases into evidence maps by assay type and chemical. These evidence maps provide an overarching view of data availability by chemical and assay type (e.g., acute, mechanistic, epidemiology). This data overview can aid in identifying toxicological endpoints that have been investigated for multiple members of a chemical subclass. Shared endpoints are key for a class-based hazard assessment, and such an overview combined with targeted information from the literature can be used to identify such shared endpoints.

**Table 3. Publicly Available Toxicity Databases[1]**

| Database | Organization | Data Categories(s) |
|---|---|---|
| Data from the CompTox Chemicals Dashboard, summarized and categorized by Vegosen and Martin (2020). | United States Environmental Protection Agency (U.S. EPA) | • In vivo adverse effects and categorical severity (and quantitative values for acute toxicity)<br>• QSAR predictions of adverse effects and categorical severity<br>• Categorizations and severity by authoritative bodies |
| Toxicity Values database (ToxVal) | United States Environmental Protection Agency (U.S. EPA) | • Quantitative values<br>• In vivo and in vitro adverse effects<br>• Limited human data<br>• Limited toxicokinetic data |
| Quantitative Structure-Activity Relationship (QSAR) Toolbox Database | Organisation for Economic Co-operation and Development (OECD) | • In vitro and in vivo adverse effects<br>• Limited toxicokinetic data |
| Danish (Q)SAR database / (Q)SAR Models | Technical University of Denmark | • QSAR predictions of adverse effects, as well as some mechanistic and toxicokinetic predictions based on 3 commercially available models: CASE Ultra, Leadscope, and SciQSAR. It also includes a “battery prediction” that integrates results from the 3 models into an overall “call”<br>• Limited human and animal experimental data |
| NIOSH Pocket Guide | National Institute for Occupational Safety and Health (NIOSH) | • Human adverse effects and toxicokinetic data<br>• Quantitative values (occupational exposure levels, OELs) |
| PubChem Bioassay | National Institutes of Health (NIH) | • In vitro mechanistic data |

| Database | Organization | Data Categories(s) |
|---|---|---|
| Comparative Toxicogenomics Database (CTD) | MDI Biological Laboratory and NC State University | • In vitro and in vivo mechanistic data<br>• Chemical-gene interactions also grouped into gene ontology terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways |
| European Chemicals Agency (ECHA) Dossier Study Results obtained via International Uniform Chemical Information Database (IUCLID) | IUCLID, ECHA | • Primarily in vivo animal data<br>• Limited human and in vitro data and QSAR predictions<br>• Limited toxicokinetic data<br>• Limited derived no-effect levels (DNELs) (quantitative) |
| Integrated Chemical Environment (ICE), Direct Download | National Toxicology Program (NTP) | • In vitro mechanistic data (Tox21, ToxCast)<br>• Animal and in vitro adverse effects including acute oral toxicity, cancer, developmental and reproductive toxicity, eye irritation skin sensitization, and absorption, distribution, metabolism, and excretion (ADME)<br>• In vivo and in vitro mechanistic and endpoint-specific data<br>• In vitro toxicokinetic data (experimental and predicted data) |

[1]The NIOSH Pocket Guide can be obtained in machine-readable format by request from NIOSH staff.

Table 3 is not an exhaustive list of useful databases but reflects those used for OFRs (and for consistency, for their analogs and metabolites). In light of the expected availability of data and the ongoing development of these and other databases, this list should only serve as a starting point for consideration when applying this process beyond the OFR work. More broadly, if animal data are prevalent across members of a potential subclass, there may be no need to include QSAR-based predictions at all.

From the existing evidence maps on OFRs, some general trends in the utility of different kinds of data and different databases emerged (Bradley et al., 2023). In general, traditional animal toxicology data and in vitro data were complementary. While new approach methodologies (NAM) data tended to exist for a greater number of chemicals, such data does not always clearly map to health endpoint categories. Traditional data can generally be mapped to health endpoint categories in evidence maps, but traditional data was available for fewer chemicals and there were fewer hits per category. A few improvements in the way NAM data are reported in databases would increase their utility, such as a clear connection to a health endpoint (including

an adverse outcome pathway [AOP] if possible) and reporting of both positive and null results with clear indication of the direction of adversity. There are databases that provide some of this information in some instances, but this is the minority occurrence. Determining usefulness of different sources of in vitro data is recommended in hazard assessment for OFRs.

There are a number of key challenges/limitations to assembling and interpreting evidence maps based on these (Table 3) or other databases. Many of these relate to database accessibility and integration, including (but not limited to): the availability of batch search, the file format of output data, the complexity and completeness of output data, distinguishing between positive/negative and null results, determining the direction of adversity, and the variety of chemical identifiers used by different databases. While many of these challenges have already been addressed in the existing R code for evidence map assembly used in the application of this guide, continuing updates to the source databases require verifying that these or other new issues have not arisen due to database updates.

The evidence maps already exist for the OFR subclasses and these should be used for the application of this guide (see the following section). For analogs and metabolites identified for OFR subclasses, it is recommended that the existing R code be used to assemble evidence maps for the same databases as for OFRs except for Danish QSAR Database/Models and the ECHA Dossiers obtained via IUCLID. Based on the OFR evidence map results, these databases are unlikely to add any additional useful information and require substantially greater level of effort due to the lack of full batch search capabilities.

### 2.4.2. Literature Reviews and Evidence Maps

For full detail on the methods of conducting literature searches and constructing evidence maps use for OFRs, refer to the literature survey guide produced under contract number CPSC-D-17-0001, Task Order 61320621F1001 (RSC, 2022). Additionally, many other sources are available that describe literature review methods, including rigorous procedures for systematic review (European Food Safety Authority (EFSA), 2010; Higgins et al., 2023; Johnson et al., 2014; National Toxicology Program (NTP), 2019; Silbergeld & Scherer, 2013; Woodruff & Sutton, 2014).This section of the guide provides only a general overview of the literature review and evidence map creation processes used for previous CPSC projects and in application of this guide for OFRs.

Figure 3 outlines the process of conducting a literature review and constructing evidence maps. The literature review process includes literature searching, screening the identified literature for relevance, and categorizing relevance literature by metadata. Extracting data from literature (with varying degrees of detail) can also be part of the literature review process. Predefined data categories should be used to tag each study for assembly into evidence maps.

Data obtained from a literature review are a complement to data from toxicological databases (Section 2.4), providing additional data not included in toxicological databases. Data from

literature are organized into evidence maps using the same chemical and data type categories as used for the database evidence maps. Additionally, literature reviews can 1) identify key assessments or other reviews that may have considered some members of a potential chemical sub-class simultaneously, and 2) identify key studies that could be used for future steps in risk assessment (e.g., dose-response assessment).

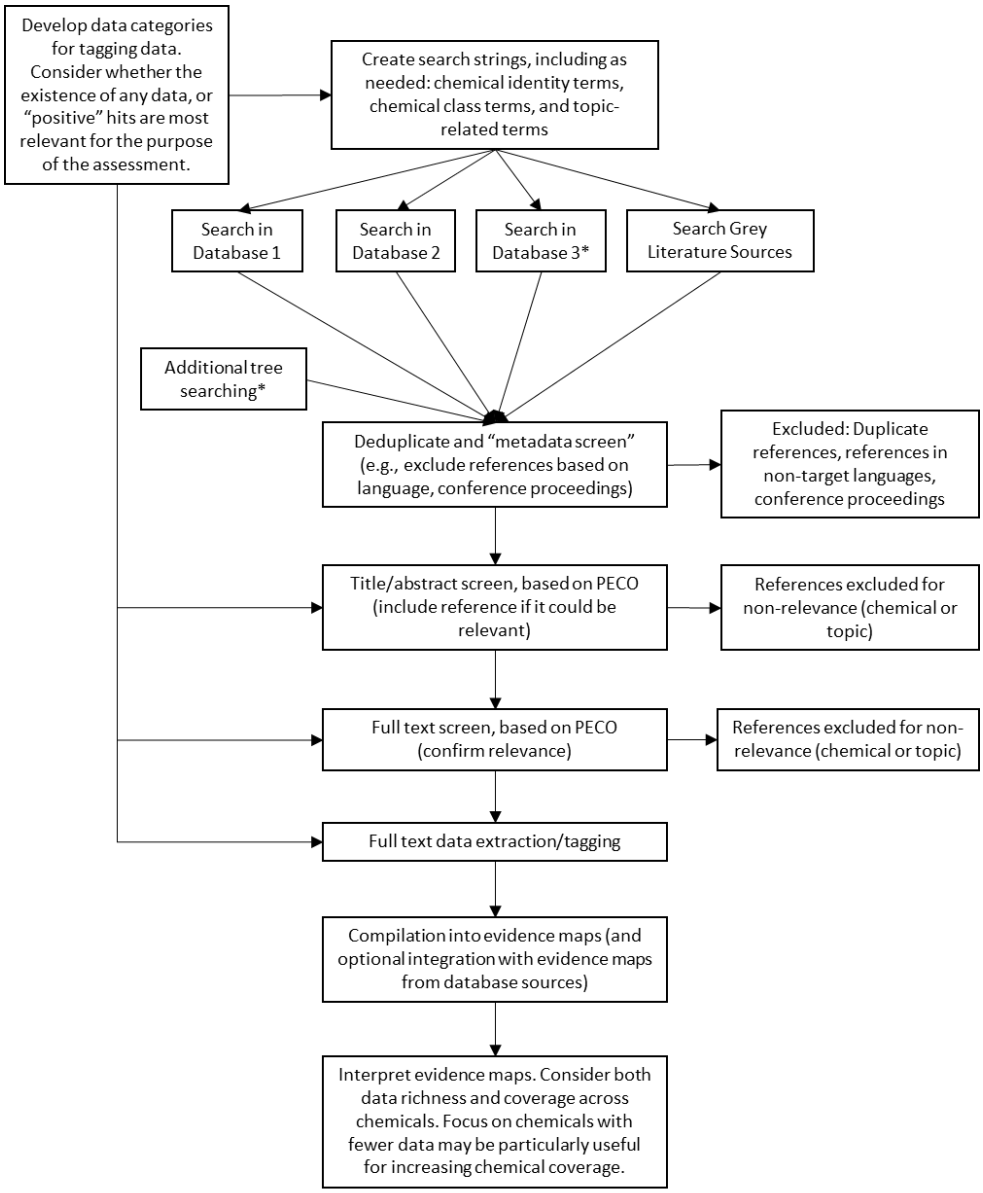


**Figure 3. Process diagram describing the steps for a literature review.**

PECO = Population, Exposure, Comparator, Outcome; PECO defines relevance criteria.
*Indicates the number of literature databases searched may vary by assessment.

There are several key challenges to conducting a literature search and identifying relevant toxicological data for a class-based hazard assessment. The overall level of effort is high, though there may be opportunities to use artificial intelligence (AI) to reduce the level of effort.

Toxicological databases (Section 2.4.2) could also be used to inform whether a literature search is needed or likely to produce results. If toxicological database evidence maps provide a lot of information on many chemicals and endpoints, a literature search may be unnecessary. Many analog and metabolite chemicals may lack sufficient data of any kind to justify the level of effort for a literature search. Test searches in one or more literature databases may also be used to inform whether a literature search is likely to provide additional useful information. Additionally, chemical identifiers for the same chemical may be different in different published studies, constituting a particular challenge when the target is a subclass (many chemicals) versus just a single chemical. For example, several PHOP OFRs appear in the literature under multiple abbreviations (and even multiple names), and multiple PHOPs may share the same abbreviation across studies. An effort by Bergman et al. (2012) to collect commonly used abbreviations and to propose a system of standardized abbreviations for many organic flame retardants (including some halogenated organophosphates) can help reconcile this potential confusion for flame retardants for recent and future literature. Quality control and deduplication is required when using abbreviations for searching, screening, and combining data.

# 3. Profiles for Evaluating Classification and Applicability of Class-based Assessment

## 3.1. Metabolite Profile

A class-based assessment can benefit from determining the metabolites of the target chemicals, then identifying the health outcomes associated with the metabolites, if possible. Health outcomes from exposure to chemicals are often tied to the metabolites, particularly when the half-life of metabolites is greater than that of the parents themselves. Additionally, identifying the metabolites for each chemical in a subclass can lead to the identification of shared metabolites. This can strengthen the class-based approach to hazard assessment by linking chemicals by similar metabolism or by linking parents to the same toxicological effects as a result of their shared metabolite(s). Metabolic pathways may not always be the same across metabolites, but similar/parallel metabolic pathways may also be considered.

There are two avenues for determining metabolite identities: 1) literature and/or database searching for empirically identified metabolites, or 2) predicting metabolites using metabolite prediction tools. The following subsections describe methods for determining the empirical and predicted metabolites, and guidance on how to establish the toxicologically relevant metabolites.

### 3.1.1. Finding Empirical Metabolites

In most cases, empirical metabolites will have a more limited role than predicted metabolites in the class-based hazard assessment, because empirical metabolites are likely reported for only a subset of the subclass members, if at all. At a minimum, empirical metabolites provide a useful check and validation of the results of the search for predicted metabolites. Information on empirical metabolites can be used to confirm the formation of specific metabolites and/or the involvement of metabolic pathways implicated in the metabolite predictions. Data on empirical metabolites may also be useful for identifying shared metabolites, if, for example, the prediction program does not include some necessary aspects of metabolism for the parent chemicals.

Empirical metabolites are likely best identified via a literature search (see Section 2.4). For chemical identity search terms, names and other identifiers of the parent chemicals should be used. In many cases, this may be all that is available. If some metabolites are already known, names or identifiers for these may be specified along with the corresponding parent chemical names to identify any less well-known metabolites. Metabolism-related search terms should be used to target results so that they are focused on metabolism and metabolites. A non-exhaustive list might include: metabolite, metabolites, metabolism, transformation, phase I, and phase II. If a specific type of metabolism is expected, terms related to relevant enzymes, metabolic processes, or conjugates may also be considered. If it is known or expected that some metabolites might be used as biomarkers to measure exposure, terms such as biomarker and biomonitoring may also be appropriate. Given the additional challenges related to chemical

identify of metabolites (see next paragraph), tree searching may be particularly helpful to find relevant metabolites.

Extraction of data from the identified literature requires a significant level of manual effort and organization. This is exacerbated by the fact that reporting on metabolite identities is likely varied and sparse. Metabolites may be reported by their chemical name, structural image, placeholder name (e.g., “Metabolite A”), or even just a chromatogram peak. This poses a challenge for extraction of data, and for correct identification. At a minimum, a structure is required to create a SMILES string, and ideally determine the chemical CASRN. However, many metabolites will not be associated with a CASRN or any other standardized identifier because they are often rare or unknown chemicals with representative structures derived based on the structure of the parents. This can make identification and/or deduplicating challenging and will add to the overall level of effort.

Some best practices should be employed to ensure extraction accuracy and efficacy of comparing the results to the predicted metabolite list. First, the identity of the corresponding parent chemical should also be extracted alongside the metabolite. Any potential chemical identifiers reported for the metabolites should be extracted, including but not limited to: chemical names, abbreviations, CAS numbers, SMILES strings, or other text-based chemical identifiers. If included, images of chemical structure should also be captured, as they can be manually converted into SMILES strings. In addition, the biology of a metabolic pathway often can be understood more easily based on chemical structures than based on other identifiers. Having a variety of chemical identifiers maximizes the chances of accurately identifying each unique metabolite.

It may also be worth noting the metabolic processes that occur (or are hypothesized), even when a specific metabolite structure is not explicitly identified. This can be used to help identify potential shared metabolism pathways across members of a subclass and facilitate comparison to predicted metabolites when prediction tools also provide information on metabolic process (e.g., Chemical Transformation Simulator [CTS]).

Depending on the assessment, it is possible to reduce the level of effort by not extracting duplicate parent-metabolite pairs repeatedly. Well-known metabolites may be identified by multiple studies and extracting only unique combinations of species and chemical (or other aspects as appropriate) may be sufficient for the purposes of metabolite and pathway identification.

Conducting quality control (QC) on extracted data is necessary given that it is a manual process. One way of approaching this is a primary extractor and a QC extractor. The QC extractor may check each extracted field, or some percentage of extracted fields. QC may also entail a higher-level check to make sure that metabolites extracted make sense chemically and metabolically.

Assigning all the extracted empirical metabolites a standard chemical identifier is also necessary to identify unique parent-metabolite pairs. Given that the initial list will be a compilation of chemical names, structures, and other identifiers, it is easiest to translate these into SMILES. because SMILES can be created more easily than other identifiers when only a chemical structure image is available. Once SMILES are created for all extracted metabolites, each should be standardized into QSAR-ready SMILES and deduplicated using the same process as all other chemicals lists in the assessment. See Appendix A for more details.

### 3.1.2. Predicting Metabolites

Numerous tools are available for predicting metabolites from parent chemicals. Choosing the tool (or tools) to predict metabolites depends on the quantity and qualities of the parent chemicals, as well as the reproducibility requirements of the assessment. The assessment-specific criteria should first be defined, then an analysis of the available tools should be conducted prior to any predictions. Additionally, metabolite predictions should be validated based on plausibility of metabolites and consistency of results with empirical data.

#### 3.1.2.1. Metabolite Predictor Choice

Relevant criteria that could be used to determine tool appropriateness depend on the goals of the assessment. For the purposes of class-based read across of OFRs, criteria should include whether the tools are publicly available, if they are regularly maintained, if they have publications evaluating their performance with measured data, and if they can be run in batch mode. Publicly available tools are ideal for reproducibility of published assessments and may be required for regulatory applications; tools that allow for bulk uploads and predictions are preferable for all class-based assessments. Table 4 is a non-exhaustive list of metabolite prediction tools, their descriptions, their availability for public use, and information on their ability to run in batch mode.

If there is a specific metabolite model tailored to the input chemicals, then that tool is preferred. For example, some metabolite prediction models are developed specifically for a certain type of drug (Litsa et al., 2020a). An additional consideration when choosing tool(s) is the methodologies the tool uses to make predictions. Predictive tools predominantly use two distinct approaches (Nguyen et al., 2018):

- **Rules-Based Approach:** This method involves predicting metabolites from parent chemicals by leveraging known chemical reactions and applying established reactions to anticipate the formation of metabolites. This approach is also called dictionary approach, where reactions/rules are derived from expert judgement or/and using statistical analysis from empirical experiment.
- **Machine Learning-Based Approach:** Distinct from relying on predefined reactions, this method seeks to construct a comprehensive model for metabolite generation based on existing data. The model considers the entirety of the metabolic landscape, extracting patterns from diverse datasets to augment its predictive capabilities.

Newer metabolite predicting tools often combine both approaches and other emerging method (Smith et al., 2022). A strategy for increasing confidence in the predicted metabolites is to use two tools with different methodologies and/or training sets and then to compare and combine the resulting list of metabolites. Using two tools can reduce tool bias and provide confidence in the plausibility of metabolites if they are predicted by both tools. Additionally, using two tools can prevent under-predictions (i.e., too few plausible results). Although it is also possible to over-predict (i.e., have too many results), this is less of a concern for the purposes of creating the metabolite profile for class-based assessment because it is more important to cover the entire chemical space. Plus, the quantity of chemicals will be naturally reduced by data availability and assessment of toxicological importance (Section 3.2.1).

**Table 4. Metabolite Prediction Tools**

| Tool | Qualities | Publicly Available/ Batch Mode[1] |
|---|---|---|
| BioTransformer 3.0 (Wishart et al., 2022) | BioTransformer focuses exclusively on predicting the metabolism of xenobiotics combining rule-based with machine learning approaches. | Yes/Yes |
| ADMET Predictor® metabolism module (SimulationsPlus) | Contains a collection of QSAR models for cytochrome P450s (CYP) metabolite prediction, CYP inhibition models, and more. | No/NA |
| GLORYx (de Bruyn Kops et al., 2021) | Combines a rules-based approach with a custom-built reaction library with FAME2 (a machine learning approach to predict sites of CYP reactions). Its library is weighted with pharmaceuticals. | Yes/Yes |
| MultiCASE-Meta-Ultra (Chakravarti & Saiakhov, 2022) | Prediction is derived from QSAR models to predict the Sites of Metabolism and a reaction of xenobiotic metabolism Database. | No/NA |
| Xenosite (Dang et al., 2020) | A web-based tool that predicts the metabolism of xenobiotics by CYP enzymes using a deep learning approach. This approach mostly works on machine learning modeling. | Yes/No |
| SymCyp | Rule-based approach from expert knowledge that includes 175 biotransformation reactions. It is available as a web-based tool. | Yes/Yes |
| MetaTrans (Litsa et al., 2020b) | Machine learning approach for metabolite structure prediction. | Yes/Yes |
| OECD QSAR Toolbox (OECD, 2024) | OECD QSAR includes 11 metabolic simulators and 5 metabolic databases. It covers both rule-based approaches as well as machine learning approaches. | Yes/Yes |

[1]NA = not applicable because tool is proprietary, not available without subscription, high costs, and/or does not allow for adequate evaluation of the methodology used and underlying assumptions.

Given the diverse range of chemical structures within the list of OFR chemicals, metabolite predictors applied to the OFRs should not be specific to a particular type of chemical. Additionally, few modeling tools have incorporated environmental chemicals into their

frameworks and these tools should be preferred. Benchmark analysis could also be developed using a subset of known chemicals to identify the software that best matches the chemicals of interest. However, this study could be challenging due to the need for defining quality criteria and determining the subset of chemicals to consider. Ultimately, developing a consensus approach utilizing available software, ideally combining machine learning with reaction-based approaches, could provide a compromise for robust modeling of metabolite prediction for the chemicals of interest.

For the OFRs, the selection of GLORYx and the OECD QSAR Toolbox is recommended to ensure a predictor encompassing a wide diversity of reactions (GLORYx) and one that incorporated environmental chemicals during its development (OECD QSAR Toolbox). Other factors supporting the selection of these two software tools were their availability, lack of licensing requirements, and widespread usage for read across.

Once the tool(s) are chosen, metabolite predictions should be made and resulting metabolites should be transformed into QSAR ready SMILES, deduplicated, and managed according to Appendix A, and data availability should be assessed using one or both methods described in Section 2.4.

#### 3.1.2.2. Limitation and Challenges of Predicted Metabolites

Predictive tools come with limitations. A recent benchmarking study of metabolite prediction tools for agrochemicals showed that most of the tested tools (including GLORYx, BioTransformer 3.0, and MetaTrans) could properly predict two-thirds of the experimentally observed first-generation metabolites (Scholz et al., 2023). However, a significant number of metabolites were predicted but not found experimentally. On average, the accuracy of these prediction tools was around 18%.

Many freely available metabolite prediction tools lack the capability to provide information on the kinetics of chemical transformation, resulting in uncertainty regarding the concentration of metabolites produced. Consequently, a significant portion of predicted metabolites may exist in low enough quantities that they are toxicologically irrelevant. As discussed in Section 3.2, information on elimination half-life and peak plasma concentration can be used to focus the list of metabolites to those more likely to be of toxicological significance. In addition, advanced commercial software addresses this limitation by offering predictions of metabolite quantities based on enzyme reactions and physiologically based toxicokinetic (PBTK) modeling. Another constraint is that metabolite predictions often center around a single reaction originating from the parent compound, neglecting potential chain reactions among metabolites. This limitation can be mitigated in machine learning-based approaches wherein predictions rely on input parent chemicals and output metabolite chemicals derived from a training set typically sourced from the literature without chemical reaction requirements. However, the effectiveness of machine learning-based approaches is contingent upon the comprehensiveness of the training set, which may be limited.

## 3.2. Physiologically Based Toxicokinetic Profile

A PBTK profile offers insights into the dynamics of chemical concentration and metabolism within diverse biological compartments. Major compartments play pivotal roles in understanding chemical toxicity. For example, plasma is crucial for transport of the chemical within the body, and the kidneys are vital for removal of the chemical from the body. However, it is important to note that a comprehensive PBTK profile is available for only a limited number of chemicals. For most substances, these profiles are derived from computational models that simulate biological compartments, predicting chemical concentrations over time. PBTK models leverage physicochemical properties as inputs, which can be either empirical or predicted. Table 5 lists five PBTK modeling tools that could be considered for class-based read across, but there are many other tools avaialble (Madden et al., 2019).

**Table 5. Physiologically Based Toxicokinetic Modeling Tools[1]**

| Tool | Qualities | Publicly Available/ Batch Mode[1] |
|---|---|---|
| EPA's high-throughput toxicokinetics package (httk) | • R package that estimates concentration in central compartment and key organs based on experimental or predicted hepatic clearance and plasma fraction unbound.<br>• Trained on more environmental chemicals (versus drugs) than other tools.<br>• Integrated (to some degree) with ICE, allowing for use without knowledge of R code. | Yes/Yes |
| GastroPlus | • Set of 10 modules that can customize predictions by the classic exposure routes including oral, dermal, inhalation, ocular and intramuscular administration for multiple species.<br>• Provides the physiochemical and ADME parameters as PBPK model input. | No/Uknown |
| Simcyp Simulator | • Incorporates databases of genetic, physiological, and epidemiological information to enable simulation of different populations (including modules for pediatrics and non-human species).<br>• Incorporates an automated sensitivity analysis tool that can be used to assess the influence of changing specific parameters. | No/Unknown |
| PK-Sim and MoBi | • PK-Sim<br>• Physiologically based kinetic (PBK) modelling tool with integrated database of anatomical and physiological parameters for humans, mouse, rat, dog and monkey.<br>• Mobi<br>• Software for multiscale physiological modelling and simulation.<br>• A range of biological models can be imported (e.g., PBK model imported from PK-Sim) or developed de novo. Software is compatible with MATLAB and R. | Yes/Yes |

[1]Adapted from Madden et al. (2019)

Choosing the tool to use should be individualized for each OFR subclass. One of the challenges that many OFR subclasses will encounter is that lipophilic chemicals are particularly difficult to model since they bioaccumulate (e.g., PHB, PHPBI). In these scenarios, models with "lipophilicity" or "adipose" compartments included are preferred. Also, if literature exists reporting successful applications (e.g., good prediction for plasma or tissue concentration) for a modeling platform for chemicals with similar physical chemical properties to the OFR subclass(es) of interest, then that would provide strong supportive evidence for the application of the model in class-based read across.

### 3.2.1. Toxicological Importance

When evaluating large amounts of PBTK data for a chemical list, whether parent or metabolite, it is important to provide criteria to prioritize chemicals that are more likely to be of toxicological relevance (i.e., likely to cause a toxic effect). PBTK modeling can provide estimated elimination half-life values and peak plasma concentrations for each substance. An understanding of the toxicokinetics of representative substances and expert derived cut-offs of half-life and plasma concentration (or other organ level concentrations) can be used to narrow down the list of substances to those that are likely to be toxicologically relevant. These cut-offs can be based on values from the target chemical subclass or can be defined by expert analysis of similar compounds or endpoints. If possible, these predictions should be compared with any available empirical data to determine the relative accuracy.

For the OFRs, PBPK data was reviewed for low toxicity and high toxicity compounds (acetaminophen and organophosphate pesticides) to determine a plasma concentration of toxicological relevance for the two opposing hazard categories. These concentrations were then used to develop a reasonable range of toxicologically relevant plasma concentration for comparison with the values calculated for the OFRs and their metabolites. While there are no direct correlations between any of the OFRs and the selected chemical types for creating these bounds, the toxicology of these groups are well understood and helpful in creating the proposed half-life and plasma concentration boundaries. For acetaminophen (low toxicity), a daily oral dosage of 4,000 mg for the average human (57 mg/kg/day) results in a plasma concentration of 97 µM (14.6 µg/mL). Safe concentrations in blood of acetaminophen are 20 mg/L (Agrawal & Khazaeni, 2023). In contrast, organophosphate pesticides are considered highly acutely toxic (Eddleston et al., 2008). A specific example, chlorpyrifos, is considered fatal at a plasma concentration of 4.73 µM, but non-fatal (yet still hazardous) at 1.1 µM. Taking this into account, OFRs and their metabolites can be estimated as toxicologically relevant if their calculated plasma concentrations are greater than 1 µM and their half-lives are greater than 1 hour. Anything circulating less than 1 hour is unlikely to be toxicologically significant. The exception to these bounds was made for compounds with half-lives greater than 20 hours to reflect the potential for accumulation with continuous exposure and is toxicologically relevant regardless of

plasma concentration given the long period of time when the chemical would be circulating in the body.

## 3.3. Physicochemical Properties

Many physicochemical properties influence bioavailability, clearance, and chemical hazard (Committee on the Design and Evaluation of Safer Chemical Substitutions, 2014). In class-based read across, a certain degree of similarity in these properties is expected among chemicals belonging to any given OFR subclass. Two types of physicochemical properties can be found in chemical databases or in the literature: empirical and predicted. Empirical data are typically limited to a few chemicals in a subclass, especially if most of the chemicals are data poor. For this reason, predicted data are predominantly used for large sets of chemicals, such as those used in class-based read across approaches.

### 3.3.1. Empirical Physicochemical Properties

Chemical databases, such as the EPA Chemical Dashboard, PubChem, or ChemSpider, include measured physicochemical properties that can be easily retrieved. However, these measured properties are typically limited to a subset of the chemicals in a subclass and are unlikely to cover the entire list of chemicals of interest comprehensively. In addition, depending on the database cycle, some databases do not integrate the latest data available in the literature. A deep dive into literature search is usually needed to complete the available empirical physicochemical properties. Empirical values should always be preferred to the predicted values for analysis if it is practical.

### 3.3.2. Predicted Physicochemical Properties

QSAR modeling can be used to enhance the coverage of physicochemical properties in a group of chemicals and make intra- and inter-class comparisons. Various suites of QSAR models, as well as individual models, have been developed to estimate the physicochemical properties of chemicals. Table 6 summarizes some of the more widely used tools for predicting physicochemical properties.

**Table 6. Physicochemical Prediction Tools**

| Tool | Description | Training Set Database | Public Availability/ Batch Mode |
|---|---|---|---|
| Open (Quantitative) Structure-activity/property Relationship Application (OPERA) | A suite of QSAR/ Quantitative structure–property relationship (QSPR) models that predict a wide variety of physicochemical properties, environmental fate parameters, and toxicity endpoints.<br>The models also include accuracy assessments, confidence ranges, and experimental values where available. | PHYSPROP | Yes/Yes |

| Tool | Description | Training Set Database | Public Availability/ Batch Mode |
|---|---|---|---|
| EPI Suite™ | A suite of estimation programs for physicochemical properties, environmental fate, and ecotoxicity. | PHYSPROP | Yes/Yes |
| ACD/Labs PhysChem | Commercial product requiring licensing for predicting physicochemical properties. | Experimental data curated by ACD/Labs of an unknown number and classification of compounds | No/Unknown |
| Cactvs Toolkit by PubChem | A machine learning based QSAR approach to predict the physicochemical properties and toxicities of compounds. | Unknown source, number and classification, of compounds. | Yes/Yes |
| SwissADME | Computes physicochemical descriptors and predicts ADME parameters, pharmacokinetic properties, drug-like nature, and medicinal chemistry friendliness of one or multiple small molecules to support drug discovery. | Unknown source, number and classification, of compounds. | Yes/Yes |

While QSAR models offer helpful and powerful predictions of physicochemical properties, they come with a set of limitations. Not all models perform equally well, and this variability is especially tied to the applicability domain of each model. The applicability domain defines the chemical space within which the model can provide accurate predictions. Typically, this corresponds to the region of the chemical space well-covered by the training set of chemicals used to develop the model. Although there is no universally accepted standard for estimating a model's applicability domain, best practices dictate that each model should specify the domain of applicability.

A model's chemical space should align with the target chemicals to the best degree possible. However, in some cases there may be no confident predictions for the chemicals of interest. For a class-based read across the physicochemical profile is unlikely to be a primary contributor to making conclusions without a high-confidence QSAR model and multiple empirical values to represent the subclass chemical space.

### 3.3.3. Building the Physicochemical Profile

The best practice involves combining predictions with experimental values and blending predictions from different models by selecting the one with the most suitable applicability domain for the chemical of interest. However, implementing this recommendation can be challenging, particularly in the case of class-based read across, where the majority of data points need to be predicted. In such instances, the most comprehensive and complete model is likely to

be used, combined with empirical data. For example, software such as OPERA already returns the empirical value if it is included in its training set.

Comparing predictive tools is challenging because the best prediction tool may differ across chemicals within a subclass. A general recommendation is to prioritize software and models with the largest training sets or use a model that has been specifically tailored for chemicals similar to the targets. However, there are no models developed for OFRs. Choosing a computational model is best determined on a subclass basis using an expert opinion on the chemistry of the subclass

## 3.4. Mechanistic Effect

Anchoring chemicals into specific subclasses can be assisted through understanding of the mechanistic effects of key chemicals within the subclass of interest. Here, mechanistic effect refers to any cellular level alterations that may occur due to interaction with a chemical of interest, including changes in gene expression, protein signaling, cell/tissue viability, hormone synthesis, etc. Through analysis of in vitro and in vivo data for key chemicals, it is possible to build a signaling profile of a subclass of chemicals. The mechanistic effect profile can provide key signaling pathways that are altered by exposure to the chemical and help to identify the overall effects (increase or decrease in signaling) of those signaling changes. This analysis for data-rich chemicals can help provide a target profile for data-poor chemicals in the chemical subclass.

Multiple databases exist that provide information on perturbations to mechanistic pathways that occur following exposure to a chemical of interest (Table 7). By building a mechanistic profile, adverse effects and signaling pathways can be connected to chemical subclasses. While several of these databases are anchored in the same data, the details provided by each are different and can help build out a profile. It should be noted, chemicals that are poorly soluble in water may not have a large amount of data or the data may be incomplete in these databases, as the approaches used are mainly dependent on aqueous media exposure. Highly common effects, such as cytotoxicity and liver enzyme effects that may affect multiple chemical subclasses can be removed for analyses with data further filtered to identify key pathway types. For the PHOPs case study, the databases used were ICE curated high throughput screening (cHTS), PubChem, and the Comparative Toxicogenomics Database.

**Table 7. Databases With Mechanistic Data**

| Database | Qualities |
|---|---|
| Toxcast/Tox21 | In vitro assays covering a range of mechanisms. |
| NTP ICE cHTS | Includes some ToxCast and Tox21 data after additional curation.<br>Groups assays that reflect a common endpoint, but the direction of adversity is not clear for each assay.<br>Active, inactive, inconclusive, and unspecified results are included. |

| | |
|---|---|
| PubChem | In vitro assays covering a range of mechanisms.<br>Active, inactive, inconclusive, and unspecified results are included. |
| Comparative Toxicogenomics Database (CTD) | Chemical-gene interactions, individually by gene and grouped by gene ontology or pathway (KEGG).<br>Only includes active hits and not inactive hits. |
| AOP-DB | Identifies AOPs for which a chemical is noted as a stressor associated with the molecular initiating event (MIE).<br>Only includes active hits and not inactive hits.<br>Genes, pathways, and diseases associated with each AOP are also listed (though these may or may not be directly associated with the searched chemical).<br>Data is updated automatically, but only a few AOPs have been through peer review. |

Visualizing patterns can often be the best way to create a mechanistic profile for a chemical subclass with many chemicals. As seen in Figure 4, the top two chemicals are data rich while others are data poor. Ideally, identifying overlapping signaling pathways among the data rich chemicals and/or chemicals with some data allows for determining where mechanistic effects are common across the subclass, thus building a mechanistic profile.

As briefly noted above, certain limitations and challenges exist when using these databases, both from the perspective of the chemical/substance of interest and for what each database can provide. While certain databases often contain the exact same data (e.g., CompTox Dashboard/ToxCast, ICE cHTS), their curation approaches for displaying or cataloguing the data may be different. How the mechanistic targets are defined also may be different based on the use case of the database, such as by target species or host species. Having an innate understanding of the intent of each database and the applicability domain of test types is key to understanding how to use the outputs.

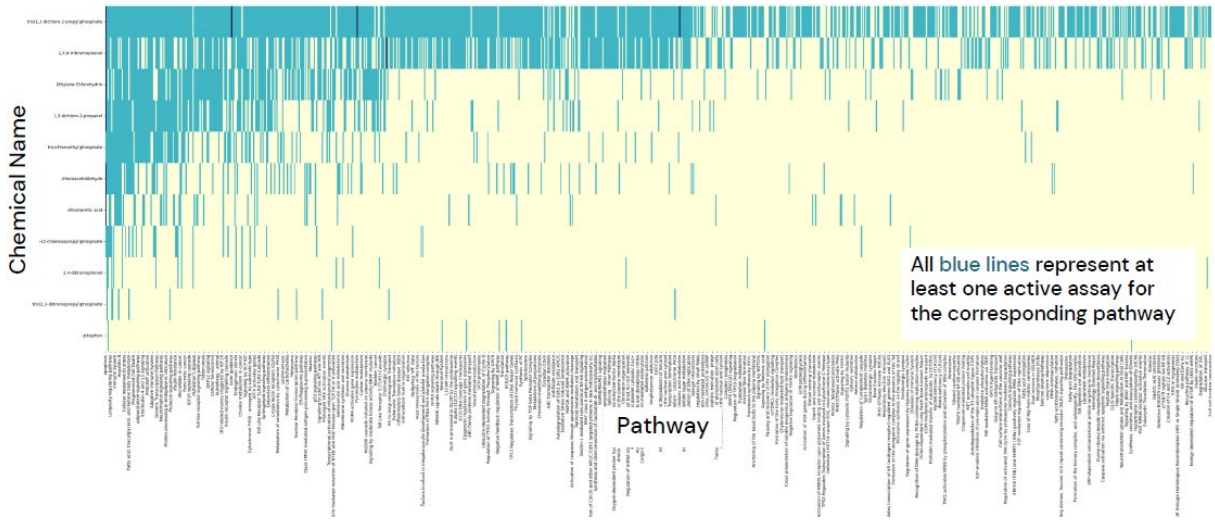


**Figure 4. Example heatmap from the PHOP case study showing if a pathway in CTD (x-axis) had at least one active hit for a chemical (y-axis).**

Additional issues come from an overall lack of data availability for assays. Chemicals may be inactive or not even tested, limiting the potential for building a class-specific profile, especially if the only available data are from assays that are not specific to any defined mechanistic profile, such as cytotoxicity or liver enzyme activity. In these cases, building a profile will be extremely difficult. Even with one data rich chemical in a subclass, it can be hard to build a profile, as anchoring an entire subclass by one single chemical may not be feasible and is restrictive. Where there are adequate data, the time required to build a profile may be prohibitive, as one needs to investigate each assay type and determine through a weight-of-evidence evaluation whether that assay contributes meaningfully to the overall profile. Such an evaluation is further complicated if a physiological outcome from perturbation of a specific pathway is unclear.

## 3.5. Adverse Effect

An adverse effect is a biochemical change, functional impairment, or pathologic lesion that impairs performance or reduces the ability of an organism to respond to additional challenge. Examples of potential adverse effects from exposure to a chemical or substance include liver damage, development of cancer, decreased fertility, or birth defects. Adverse effect data are high priority for determining commonalities between chemicals that likely belong within a subclass and for making assessment conclusions about the subclass.

A wealth of tools and resources are available to gather relevant data. For OFRs, much of the following information on adverse effects has already been identified, through literature searches and database searches as outlined in Section 2, CPSC's scoping documents, and results of a systematic review of OFRs currently in progress by DTT (NIEHS, 2022). Data for analogs and metabolites for OFRs can be obtained using the methods in Section 2, as described in the accompanying case study.

Additional publicly available literature reviews that are unlikely to appear in literature search databases such as PubMed may be identified through agencies such as Division of Translational Toxicology (DTT), U.S. Environmental Protection Agency (EPA), European Food Safety Authority (EFSA), European Chemicals Agency (ECHA), Health Canada, Agency for Toxic Substances and Disease Registry (ATSDR), International Agency for Research on Cancer (IARC), International Programme on Chemical Safety (IPCS), Joint FAO/WHO Expert Committee on Food Additives (JECFA), and Joint Meeting on Pesticide Residues (JMPR), among others. Reviews from many of these agencies have been identified for chemicals belonging to several of the OFR subclasses and should be considered in an assessment, when feasible.

A key step for a class-based hazard analysis is to identify endpoints or targets that are shared among multiple chemicals within a subclass. It may seem intuitive to focus primarily on the more data-rich members of the subclass to identify such shared effects but focusing on shared effects for all chemicals with data builds confidence in the hazard assessment. Identifying the

probable shared adverse effects can be accomplished through direct visualization with a heatmap. Figure 5 provides two heatmap examples used in the accompanying case study to identify shared data availability among PHOP chemicals.

Heatmaps can show multiple levels of data, both by the number of blocks that are filled in for a particular category and by the color to show the number of hits within a specific block. For example, Figure 5 shows the different PHOPs chemicals across the top horizontal axis, different health effects (vertical access) and the number of data hits (color of the block). Heatmaps can provide a simple way to see which chemicals have the most data and what type of data is shared among chemicals in the subclass and their analogs.

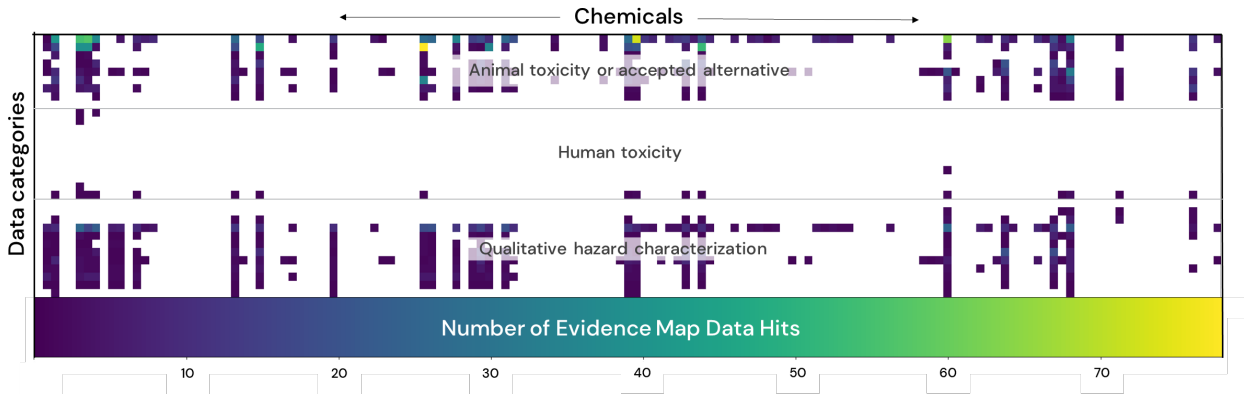


**Figure 5. Example heatmap showing the number of data hits in evidence maps for chemicals.**

Data availability from the DTT literature should be examined for shared health endpoints among OFRs and their NASEM analogs using the data published in a public Tableau heatmap in 2023 (DTT, 2023). (Additional analogs identified using the methods described in this guide and the accompanying case study were not included in the DTT systematic review.) Co-exposures, mixtures, and nonmammalian systems should be filtered out before analysis. At the time this guide was written, the DTT published Tableau heatmap included data availability only for animal and epidemiology studies (mechanistic studies were not divided into assay types or outcomes). However, the DTT literature review was still in progress and future application of this guide should use all details available.

It is important to identify potential health effects reported for metabolites and to compare them to those reported for OFRs, particularly if the PBTK profile suggests the metabolites are more toxicologically relevant than parent chemicals. The same heatmap style comparison of evidence maps can accomplish this. Given that the adverse effect profile is built on data availability, it is important to note that data counts are subject to several biases, reflecting in particular that a data hit for the OFRs means that an endpoint was evaluated, not necessarily that there was an effect. This means that the endpoints with hits are closely tied to the study types conducted. For example, the availability of multiple guideline repeat dose animal studies means that there will be high data counts for systems and targets that are required for the guideline. Thus, there may be a large number of hits for mortality or whole body effects compared to other outcomes even if

there are no mortalities and no significant changes in body weights (or other whole body outcomes). This bias should be considered when analyzing the data availability and comparing Tableau results to the information reported in the scoping documents for OFRs. If multiple studies reported mortality due to the chemical or subclass of interest, it is likely that this would have been noted during the scoping process, meaning the scoping documents provide key clues when assessing data availability.

This guide and the accompanying case study only cover the data availability. This means that application of the guide results in only probable effects, not confirmed reported effects. In order to build confidence in the adverse effects profile of OFRs and the resulting read across assessment, it is necessary to extract and analyze the data from the primary studies. Data extraction can be a laborious effort depending on the number of studies and quality of the studies. If data extraction is pursued as a follow-on step to this guide, then there are several methods for reducing the level of effort. For example, prioritizing studies based on presumed quality (i.e., guideline studies over ECHA dossiers) or using machine learning tools to assist in data extraction are two ways to reduce the level of effort. Further discussion on data extraction is outside the scope of this guide.

# 4. Assessment and Conclusions

## 4.1. Identifying Anchor Chemical(s) and Defining the Chemical Subclass

The first step in the overall assessment is to identify the anchor chemical(s) representing the subclass. These must have some data at a minimum, and ideally be data rich. Chemicals within the subclass, analogs, or a combination of both can be considered anchor chemicals. The main requirement is the abundance of hazard data, but the anchor chemicals should also represent the overall subclass structurally and in each of the five profiles described in Section 3. Ideally, anchor chemicals should:

- Collectively cover the chemical space of a subclass based on structural similarity and physicochemical properties.
- Have toxicity data available for a relatively wide range of endpoints, and a range of mechanistic data.
- Have shared and/or structurally similar metabolites with a significant number of other chemicals in the subclass.

After completing all five profiles, the choice of anchor chemical(s) should be clear. If there is significant ambiguity, then the subclass itself is likely not well defined for class-based assessment and should be evaluated for splitting into multiple subclasses.

Once one or more anchor chemicals are identified, confidence of subclass inclusion can be quantified for each proposed OFR. Table 8 provides criteria for making quantitative assessments of subclass inclusion according to the information gleamed in each profile and additional structural similarity evaluations using the analog lists and structural comparisons from Section 2. Confidence in the inclusion of each OFR in its respective subclass is assessed using a rating scale, rating the chemical as 0 (low/no confidence) to 2 (high confidence) for each classifying factor described in Table 8. The criteria in Table 8 may be customized for each subclass based on the subclass specific results.

**Table 8. Confidence Rating Scale Criteria for Subclass Inclusion**[1]

| Classifier | Low/No Confidence (0) | Medium Confidence (1) | High Confidence (2) |
|---|---|---|---|
| **Metabolite Profile** | No shared metabolites with other OFRs. | No shared metabolites with other OFRs but have metabolites that are structurally similar to the metabolites of the anchor chemical(s) or other OFRs in the subclass. | Shared metabolites with other OFRs. |
| **PBTK Profile** | Predicted TK values of OFR or metabolites are outliers. | Predicted and/or experimental TK values of OFR or metabolites are within | Predicted and/or experimental values of OFR or metabolites |

| Classifier | Low/No Confidence (0) | Medium Confidence (1) | High Confidence (2) |
|---|---|---|---|
| | | reasonable range of main distribution. | are within close range of main distribution. |
| **Physicochemical Profile** | Outliers to distributions. | Qualitatively different from subclass distributions according to expert opinion. | Within the subclass distribution. |
| **Mechanistic Profile** | Not enough data hits available to evaluate confidence; or data availability hits do not meet criteria for medium confidence. | Has some data and data availability suggests effects are shared with 1 or more anchor chemicals. | Data rich, with 3 or more shared data hit effect categories and/or target systems. |
| **Adverse Effect Profile** | No or little data availability limiting confidence evaluation; or data availability hits do not meet criteria for medium confidence. | Data hits of OFR includes at least 2 endpoints shared by anchor chemical(s) in a literature review; or data hits of OFR includes at least 3 of the assay types shared by anchor chemical(s) in evidence maps. | Known shared health effects with anchor chemical(s) through a scoping document or other source. Anchor OFRs are rated 2 by default. |
| **Metabolite Adverse Effect** | No or little data availability, limiting confidence evaluation; or Data availability hits do not meet criteria for medium confidence. | Data hits of at least 1 metabolite includes at least 2 endpoints shared by anchor chemical(s) in a literature review; or data hits of at least one metabolite includes at least 3 of the assay types shared by anchor chemical(s) in evidence maps. | Data hits of at least 3 metabolites included at least 2 endpoints shared by anchor chemical(s) in a literature review; or data hits of at least three metabolite includes at least 3 of the assay types shared by anchor chemical(s) in evidence maps. |
| **Structural Similarity** | Clear or potential outlier on structural analysis visualization | Within the main subclass grouping on structural analysis visualization | Within the main subclass grouping on structural analysis visualization and considered an analog of anchor chemicals (or anchor chemical is an analog of the OFR). |

[1]Statistical ranges or distribution analysis should be created for each individual group and profile based on the distribution type.

The sum of the confidence ratings can then be used to determine the overall confidence for including each OFR in the subclass. The combination of the confidence criteria above for all five profiles provides a quantitative way to describe the confidence in grouping chemicals together even when very little or no data is available for many chemicals in the group. The anchor chemicals need to be in the highest confidence category, or they are not proper anchor chemicals. This means that there could be data rich chemicals that are not anchor chemicals after assessment. Using the total ranking to assign categories of confidence (e.g., high, medium, low) will likely require ranges customized to each subclass. In the PHOPs case study, there were four categories: very high, high, medium, and low.

Anything considered low confidence should be excluded from the subclass and reevaluated with another subclass or as a new subclass altogether. This is because the predictive tools that can be used for the metabolite, PBTK, and physicochemical properties profiles and the structural analysis provide four mechanisms for comparing chemicals with no data to the anchor chemicals, and a resulting low confidence score provides computational evidence that the chemicals are outliers from the target subclass. A medium confidence rating has some ambiguity, and the additional information that empirical data would provide may promote (or demote) the confidence ratings for those chemicals.

## 4.2. Conclusions

The OFRs rated as medium to high confidence for subclass inclusion are considered as a target subclass for determining health outcomes. For the qualitative hazard assessment, evidence should be summarized based on the adverse effects profile and supported by additional profiles. The outcome of applying this guide will be an updated list of the chemicals in a subclass, along with identified analogs and metabolites, and a statement of confidence for each OFR belonging to the subclass. In addition, draft hazard effects, such as likely target systems, will be identified for the subclass. Where possible, analogs should support these conclusions. Recommendations for next steps to increase confidence in the assessment (such as key studies to address data gaps) should be discussed.

# References


Agrawal, S., & Khazaeni, B. (2023). Acetaminophen toxicity. In *StatPearls [Internet]*. StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK441917/

Bergman, Å., Rydén, A., Law, R. J., de Boer, J., Covaci, A., Alaee, M., Birnbaum, L., Petreas, M., Rose, M., Sakai, S., Van den Eede, N., & van der Veen, I. (2012). A novel abbreviation standard for organobromine, organochlorine and organophosphorus flame retardants and some characteristics of the chemicals. *Environment International*, *49*, 57-82. https://doi.org/10.1016/j.envint.2012.08.003

Bevington, C., Williams, A. J., Guider, C., Baker, N. C., Meyer, B., Babich, M. A., Robinson, S., Jones, A., & Phillips, K. A. (2022). Development of a flame retardant and an organohalogen flame retardant chemical inventory. *Scientific Data*, *9*(1), 295. https://doi.org/10.1038/s41597-022-01351-0

Bradley, M. A., Christianson, A., & Haber, L. T. (2023). *Efficient assembly of high-throughput human health toxicity data across databases* (Abstract #5104). (The Toxicologist: Late-Breaking Supplement, a Supplement to Toxicological Sciences, 192, Issue S2). https://www.toxicology.org/pubs/docs/Tox/2023-Late-Breaking-Abstract-Supplement.pdf

Chakravarti, S. K., & Saiakhov, R. D. (2022). MultiCASE platform for in silico toxicology. *Methods Molecular Biology*, *2425*, 497-518. https://doi.org/10.1007/978-1-0716-1960-5_19

Committee on the Design and Evaluation of Safer Chemical Substitutions: A Framework to Inform Government and Industry Decision, Board on Chemical Sciences and Technology, Board on Environmental Studies and Toxicology, Division on Earth and Life Studies, & National Research Council. (2014). 5, Physiochemical properties and environmental fate. In *A Framework to Guide Selection of Chemical Alternatives*. National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK253956/

Consumer Product Safety Commission (CPSC). (2017). *Staff briefing package in response to petition HP15-1, requesting rulemaking on certain products containing organohalogen flame retardants*. https://www.cpsc.gov/content/Commission-Briefing-Package-Petition-HP-15-1-Requesting-Rulemaking-on-Certain-Products-Containing-Organohalogen-Flame-Retardants

Consumer Product Safety Commission (CPSC). (n.d.). *Organohalogen flame retardant chemicals assessment*. https://www.cpsc.gov/Business--Manufacturing/Organohalogen-Flame-Retardant-Chemicals-Assessment

Dang, N. L., Matlock, M. K., Hughes, T. B., & Swamidass, S. J. (2020). The metabolic rainbow: Deep learning phase I metabolism in five colors. *Journal of Chemical Information and Modeling*, *60*(3), 1146-1164. https://doi.org/10.1021/acs.jcim.9b00836

Daylight Chemical Information Systems Inc. (2011a). *Daylight Theory Manual*. https://www.daylight.com/dayhtml/doc/theory/index.html

Daylight Chemical Information Systems Inc. (2011b). *Fingerprints - Screening and similarity* (Daylight Theory Manual, Issue. https://www.daylight.com/dayhtml/doc/theory/theory.finger.html

de Bruyn Kops, C., Šícho, M., Mazzolari, A., & Kirchmair, J. (2021). GLORYx: Prediction of the metabolites resulting from phase 1 and phase 2 biotransformations of xenobiotics. *Chemical Research in Toxicology*, *34*(2), 286-299. https://doi.org/10.1021/acs.chemrestox.0c00224

Division of Translational Toxicology (DTT) Visualizations. (2023). *OFR Evidence Maps (SOT 2023)*. https://public.tableau.com/app/profile/division.of.translational.toxicology.dtt./viz/OFREvidenceMapsSOT2023/ReadMe?publish=yes

Durant, J. L., Leland, B. A., Henry, D. R., & Nourse, J. G. (2002). Reoptimization of MDL keys for use in drug discovery. *Journal of Chemical Information and Computer Sciences*, *42*(6), 1273-1280. https://doi.org/10.1021/ci010132r

Eddleston, M., Eyer, P., Worek, F., Sheriff, M. H., & Buckley, N. A. (2008). Predicting outcome using butyrylcholinesterase activity in organophosphorus pesticide self-poisoning. *QJM*, *101*(6), 467-474. https://doi.org/10.1093/qjmed/hcn026

European Food Safety Authority (EFSA). (2010). Application of systematic review methodology to food and feed safety assessments to support decision making. *EFSA Journal*, *8*(6), 1637. https://doi.org/10.2903/j.efsa.2010.1637

Fourches, D., Muratov, E., & Tropsha, A. (2010). Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research. *Journal of Chemical Information and Modeling*, *50*(7), 1189-1204. https://doi.org/10.1021/ci100176x

Fourches, D., Muratov, E., & Tropsha, A. (2016). Trust, but verify II: A practical guide to chemogenomics data curation. *Journal of Chemical Information and Modeling*, *56*(7), 1243-1252. https://doi.org/10.1021/acs.jcim.6b00129

Fournier, R. (2003). Reviews in Computational Chemistry. Volume 18. *Journal of the American Chemical Society*, *125*(22), 6836-6836. https://doi.org/10.1021/ja025334y

Grulke, C. M., Williams, A. J., Thillanadarajah, I., & Richard, A. M. (2019). EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research. *Computational Toxicology*, *12*, 100096. https://doi.org/10.1016/j.comtox.2019.100096

Higgins, J., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V. (2023). *Cochrane handbook for systematic reviews of interventions*. Cochrane. www.training.cochrane.org/handbook

Jaeger, S., Fulle, S., & Turk, S. (2018). Mol2vec: Unsupervised machine learning approach with chemical intuition. *Journal of Chemical Information and Modeling*, *58*(1), 27-35. https://doi.org/10.1021/acs.jcim.7b00616

Johnson, P. I., Sutton, P., Atchley, D. S., Koustas, E., Lam, J., Sen, S., Robinson, K. A., Axelrad, D. A., & Woodruff, T. J. (2014). The Navigation Guide—Evidence-based medicine meets environmental health: Systematic review of human evidence for PFOA effects on fetal growth. *Environmental Health Perspectives*, *122*(10), 1028-1039. https://doi.org/10.1289/ehp.1307893

Landrum, G. (n.d.). *Getting started with the RDKit in python - RDKit fingerprint*. https://www.rdkit.org/docs/GettingStartedInPython.html#rdkit-fingerprints

Li, T., Tong, W., Roberts, R., Liu, Z., & Thakkar, S. (2021). DeepCarc: Deep learning-powered carcinogenicity prediction using model-level representation [Original Research]. *Frontiers in Artificial Intelligence*, *4*. https://doi.org/10.3389/frai.2021.757780

Litsa, E. E., Das, P., & Kavraki, L. E. (2020a). Prediction of drug metabolites using neural machine translation. *Chemical Science (Royal Society of Chemistry: 2010)*, *11*(47), 12777-12788. https://doi.org/10.1039/d0sc02639e

Litsa, E. E., Das, P., & Kavraki, L. E. (2020b). Prediction of drug metabolites using neural machine translation. *Chemical Science*, *11*(47), 12777-12788. https://doi.org/10.1039/D0SC02639E

Madden, J. C., Pawar, G., Cronin, M. T. D., Webb, S., Tan, Y.-M., & Paini, A. (2019). In silico resources to assist in the development and evaluation of physiologically-based kinetic models. *Computational Toxicology*, *11*, 33-49. https://doi.org/10.1016/j.comtox.2019.03.001

Mansouri, K., Grulke, C. M., Richard, A. M., Judson, R. S., & Williams, A. J. (2016). An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modelling. *SAR QSAR Environ Res*, *27*(11), 939-965. https://doi.org/10.1080/1062936x.2016.1253611

Mansouri, K., Moreira-Filho, J. T., Lowe, C. N., Charest, N., Martin, T., Tkachenko, V., Judson, R., Conway, M., Kleinstreuer, N. C., & Williams, A. J. (2024). Free and open-source QSAR-ready workflow for automated standardization of chemical structures in support of QSAR modeling. *Journal of Cheminformatics*, *16*(1), 19. https://doi.org/10.1186/s13321-024-00814-3

National Academies of Sciences Engineering and Medicine. (2019). *A class approach to hazard assessment of organohalogen flame retardants*. The National Academies Press. https://doi.org/10.17226/25412

National Institute of Environmental Health Sciences (NIEHS). (2022). *Consumer products and theraputics program*. https://www.niehs.nih.gov/research/atniehs/dtt/strategic-plan/exposure/consumer

National Toxicology Program (NTP). (2019). *Handbook for conducting a literature-based health assessment using OHAT approach for systematic review and evidence integration*. https://ntp.niehs.nih.gov/sites/default/files/ntp/ohat/pubs/handbookmarch2019_508.pdf

Nguyen, D. H., Nguyen, C. H., & Mamitsuka, H. (2018). Recent advances and prospects of computational methods for metabolite identification: A review with emphasis on machine learning approaches. *Briefings in Bioinformatics*, *20*(6), 2028-2043. https://doi.org/10.1093/bib/bby066

Organisation for Economic Co-operation and Development (OECD). (2024). *QSAR toolbox: Metabolisms (observed and simulated)*. https://qsartoolbox.org/resources/metabolisms/

Organisation for Economic Co-operation and Development (OECD). (n.d.). *Introduction to (quantitative) structure activity relationships*. https://www.oecd.org/chemicalsafety/risk-assessment/introductiontoquantitativestructureactivityrelationships.htm

Patterson, D. E., Cramer, R. D., Ferguson, A. M., Clark, R. D., & Weinberger, L. E. (1996). Neighborhood behavior: A useful concept for validation of "Molecular Diversity" descriptors. *Journal of Medicinal Chemistry*, *39*(16), 3049-3059. https://doi.org/10.1021/jm960290n

Scholz, V.-A., Stork, C., Frericks, M., & Kirchmair, J. (2023). Computational prediction of the metabolites of agrochemicals formed in rats. *Science of the Total Environment*, *895*, 165039. https://doi.org/https://doi.org/10.1016/j.scitotenv.2023.165039

Sedykh, A. Y., Shah, R. R., Kleinstreuer, N. C., Auerbach, S. S., & Gombar, V. K. (2021). Saagar-A new, extensible set of molecular substructures for QSAR/QSPR and read-across predictions. *Chemical Research in Toxicology*, *34*(2), 634-640. https://doi.org/10.1021/acs.chemrestox.0c00464

Shah, I., Tate, T., & Patlewicz, G. (2021). Generalized read-across prediction using genra-py. *Bioinformatics*, *37*(19), 3380-3381. https://doi.org/10.1093/bioinformatics/btab210

Silbergeld, E., & Scherer, R. W. (2013). Evidence-based toxicology: Strait is the gate, but the road is worth taking. *ALTEX - Alternatives to animal experimentation*, *30*(1), 67-73. https://doi.org/10.14573/altex.2013.1.067

Smith, A. M. E., Lanevskij, K., Sazonovas, A., & Harris, J. (2022). Impact of established and emerging software tools on the metabolite identification landscape. *Front Toxicol*, *4*, 932445. https://doi.org/10.3389/ftox.2022.932445

Todeschini, R., Consonni, V., Xiang, H., Holliday, J., Buscema, M., & Willett, P. (2012). Similarity coefficients for binary chemoinformatics data: Overview and extended comparison using simulated and real data sets. *Journal of Chemical Information and Modeling*, *52*(11), 2884-2901. https://doi.org/10.1021/ci300261r

University of Cincinnati Risk Science Center (RSC). (2022). *Literature survey guide draft 6*. https://www.cpsc.gov/s3fs-public/Literature-survey-guide-draft-6.pdf?VersionId=_j214PR_dTRvc0xJuXMt4EZYc3LyIxop

Wishart, D. S., Tian, S., Allen, D., Oler, E., Peters, H., Lui, Vicki W., Gautam, V., Djoumbou-Feunang, Y., Greiner, R., & Metz, Thomas O. (2022). BioTransformer 3.0—a web server for accurately predicting metabolic transformation products. *Nucleic Acids Research*, *50*(W1), W115-W123. https://doi.org/10.1093/nar/gkac313

Woodruff, T. J., & Sutton, P. (2014). The Navigation Guide systematic review methodology: A rigorous and transparent method for translating environmental health science into better health outcomes. *Environmental Health Perspectives*, *122*(10), 1007-1014. https://doi.org/10.1289/ehp.1307175

Yang, C., Tarkhov, A., Marusczyk, J., Bienfait, B., Gasteiger, J., Kleinoeder, T., Magdziarz, T., Sacher, O., Schwab, C. H., Schwoebel, J., Terfloth, L., Arvidson, K., Richard, A., Worth, A., & Rathman, J. (2015). New publicly available chemical query language, CSRML, to support chemotype representations for application to data mining and modeling. *Journal of Chemical Information and Modeling*, *55*(3), 510-528. https://doi.org/10.1021/ci500667v

# Appendix A. Chemical List Management

## A.1. Chemical Structure Format

A standardized method of chemical identification and tracking must be employed to ensure quality control of the data and consistency across steps when working with many chemicals and many pieces of information connected to each chemical. The SMILES format is commonly employed to represent chemicals as strings and make them machine readable. However, there is no universal standard for writing a SMILES string, and two different software programs or code libraries may generate distinct SMILES representations for the same chemical. Throughout any assessment, it is crucial to consistently adhere to a specific SMILES format and use the same software to process other chemical identifiers into their SMILES. Using the same SMILES format for chemical list management will avoid unintended duplication of chemicals or incorrect analysis due to information being tagged to varying SMILES strings for the same chemical. Any new chemicals (analogs, metabolites, or additional target chemicals) that are added to the assessment during the process should always be converted to the chosen SMILES standard as the first step after chemical retrieval.

After identifying SMILES strings for each chemical, certain preparation steps are necessary to standardize the structure and remove duplicates. Deduplication is an important step in chemical list management and using a homogenously applied SMILES format is the most effective way to accomplish this. Cheminformatics standards are available on how to prepare a chemical's structure (Fourches et al., 2010, 2016; Mansouri et al., 2016).

Classic structure preparation includes:

- standardizing SMILES in the chosen SMILES format,
- removal of salts and metal connections,
- elimination of all hydrogen atoms,
- neutralization of the chemical,
- removing stereochemistry,
- treatment of tautomers (discreet or duplicates),
- stripping of chemical solvents from the chemical identifier,
- exclusion of mixtures, and
- deduplication of chemicals.

Additional chemical list preparation steps could include more specific procedures depending on the chemical list and the intent for the list. For example, it may involve a step to remove any chemicals that contain inorganic atoms if inorganic complexes are irrelevant to the assessment.

For application of this guide on the CPSC expanded OFR list, a QSAR ready SMILES workflow was developed using RDKit and implemented in Python using the in house library CompDesc

(https://test.pypi.org/project/CompDesc/). This package mimics the QSAR ready preparation workflow developed by Mansouri et al. (2024) and Mansouri et al. (2016). More specifically, this package includes sequential steps: normalization of the SMILES, which removes chemical stereochemistry and standardizes the SMILES; a metal/ion removal step to strip any isolated metals and counter ions; a step to remove the charge from the molecule; a step to remove hydrogen; and finally, a step to eliminate mixtures and inorganic chemicals (i.e., no SMILES will generate for mixtures or inorganic chemicals).

## A.2. Version Control

Scripts used for developing read across should be included in a version control system, employing distributed version control systems such as Git or Mercurial to ensure reproducibility and the effective tracking of changes. Additionally, the origin of input files should be documented, including version information if available, and the date of assessment. It is essential that input files remain unmodified.

# Appendix B. Chemical Similarity and Clustering

## B.1. Chemical Representations

Chemicals can be computationally represented through various approaches, including classic fingerprinting, molecular descriptors, and SMILES vectorization. Each of these representations has its advantages and disadvantages depending on the usage end goal. The approach to chemical representation should be based on the representation's ability to cluster chemicals within a subclass effectively when compared to chemicals not in the subclass, ensuring the accurate capture of subclass characteristics while minimizing the overlap with other subclasses,

Molecular descriptors enable the assignment of quantitative values to depict diverse properties, such as physicochemical properties or binding affinity. This method provides a comprehensive way to capture a range of characteristics for each chemical. Examples of molecular descriptors include physical chemical properties such as LogP (characterizing a molecule's lipophilicity) and pKa (acid dissociation constant).

Vectorization is a representation that employs SMILES strings and machine learning techniques to represent chemicals (Jaeger et al., 2018). This approach proves particularly useful for QSAR modeling, as demonstrated in Li et al. (2021) but provides descriptors that cannot be understood by a human, meaning data interpretation requires an additional step converting the vectorization to a chemical identifier.

Fingerprints involve a binary vector where each property is indicated by a bit (0 or 1). Each bit can represent, for example, the presence of a chemical substructure such as a carbonyl. This method is favored for large database similarity searches due to its rapid computation and the ability to encapsulate numerous properties in a single string. However, a limitation of fingerprints is their inability to represent chemical quantitative properties. The literature offers a variety of fingerprint types, each with its specific characteristics. Table B.1 is a non-exhaustive list of some common fingerprints to consider using as chemical representations.

**Table B.1. Examples Common Chemical Fingerprints for Read-Across Methods**

| Fingerprint | Description | Qualities |
|---|---|---|
| Topological RDKit (Landrum, n.d.) | Based on a hash definition of a molecule. The algorithm finds all subgraphs between paths from 1 to 7 chemical bonds. Subgraphs are next converted into bits. | • Very quick to compute.<br>• Only characterizes the molecular topology (e.g., the path in the molecular structure) and not the chemical functional groups. |
| Molecular ACCess System (MACCS) (Durant et al., 2002) | SMARTS-based (SMiles ARbitrary Target Specification language based) implementation of the 166 public MACCS keys translated into bits. | • A simple fingerprint based on substructure search that can be easily interpretable for comparing compounds.<br>• Limited to 166 substructures and thus cannot completely characterize a |

| Fingerprint | Description | Qualities |
|---|---|---|
| | | molecule with substructures outside of the predefined 166. |
| PubChem | Developed specifically to screen the PubChem database. It includes 880 bits covering element and ring counts, chemical topology (including bond types), and substructure search. | • Provides chemical characterization that includes multiple chemical representations.<br>• Can be computed directly with the PubChem API, therefore making applying chemical representation assignment and database searching for chemical similarity seamless. |
| Saagar (Sedykh et al., 2021) | Specifically developed for read across purposes and to integrate with toxicity databases. It includes 834 bits mostly based on substructures. | • Optimized for read across analysis and showed good performance in a read across benchmark study.<br>• Proprietary fingerprints.<br>• Are available on ICE web tool but not available in batch mode. |
| ToxPrint (Yang et al., 2015) | Based on chemotype representation of a molecule (e.g., SMARTS patterns) or reaction transformations (e.g., Simplified Molecular Input Reaction Kode System, reaction SMILES). ToxPrint was designed to cover the environmental, regulatory, and commercial-use chemical space, and to represent patterns relevant to toxicity. It includes 729 bits. | • Capability of considering reaction transformation.<br>• Developed specifically for environmental chemicals and toxicity evaluations.<br>• Available through the EPA chemicals dashboard in batch mode. |

The selection of a fingerprint should enable the accurate definition of each chemical with its unique characteristics, therefore facilitating the differentiation of chemicals from one another. However, fingerprints with too much granularity for defining chemicals may be restrictive when searching for analogs.

## B.2. Similarity Metric

Similarities between chemicals are calculated using a chosen distance metric or coefficient. The choice of metric often depends on the chemical representations employed. For example, Euclidean or Manhattan metrics can be used for molecular descriptors, while Tanimoto/Jaccard or Dice distances can be used for fingerprint representations.

Specifically for fingerprints, comparisons have been made by comparing similarity metrics with different benchmark datasets, and it was concluded that the Tanimoto/Jaccard coefficient performed well in comparison to the other metrics (Todeschini et al., 2012). There is no direct advantage to test different metrics in a class-based read across for analog search of a specific chemical set. The application of this guide to class-based read across of OFRs will use the Tanimoto similarity coefficient.

## B.3. Chemical Clustering

Chemical clustering usually occurs after pairwise similarity scores are computed among all chemicals in the dataset. It is used to organize the set of chemicals and group together those that are most similar. To develop chemical clustering, users need to choose the clustering algorithm as well as the type of representations they want to use to visualize clusters. There are two main types of clustering algorithms: hierarchical and non-hierarchical.

Briefly, hierarchical clustering approaches typically use an agglomerative or a divisive method to group chemicals based on an iterative combination of clusters. Hierarchical clustering has the advantage of creating a hierarchy of clusters, allowing users to explore relationships at various levels of granularity.

A non-hierarchical clustering algorithm employs a different approach in which chemicals are randomly assigned to clusters. The algorithm iteratively seeks the optimal arrangement of clusters by aiming to minimize the intra-cluster distance and maximize the inter-cluster distance. Importantly, in this approach, the user defines the number of clusters desired before initiating the clustering process.

Various algorithms and approaches are available (Fournier, 2003). There is no approach that clearly outperforms others, and no clear guidelines are available to choose the clustering approach. This choice is mostly dependent on preference or expertise. Visualization of clustering can be accomplished with multiple methods, such as:

- The principal component analysis (PCA) is a dimensionality reduction technique that can help visualize the clustering of chemicals in a lower-dimensional space. Each dimension is defined with the most significant features contributing to the variance in the data.
- The t-Distributed Stochastic Neighbor Embedding (t-SNE) is another dimensionality reduction technique.
- The Hierarchical Clustering Dendrogram is a tree-like structure (dendrogram) that represents the relationships between different chemicals. It is good for understanding the hierarchy and grouping of chemicals.
- The Heatmap can be used to directly visualize the similarity or dissimilarity matrix. Usually rows and columns represent chemicals, and the color intensity indicates the degree of similarity.

Again, there are no best choices for visualization and the choice mostly depends on expert judgement. However, this decision is usually influenced by the size of the datasets. For example, visualizing a dendrogram with more than a few hundred chemicals can be challenging, while using PCA with only a few chemicals may provide limited information.

# Appendix C. Database and Tool Sources

**Table C.1. Publicly Available Toxicity Database Sources[1]**

| Database | Source URL |
|---|---|
| Data from the CompTox Chemicals Dashboard, summarized and categorized by Vegosen and Martin (2020). | Published supplementary material available at: https://doi.org/10.1007/s10098-019-01795-w |
| Toxicity Values database (ToxVal) | Provided by U.S. EPA for early phases of this work, now available at: https://comptox.epa.gov/dashboard/ |
| Quantitative Structure-Activity Relationship (QSAR) Toolbox Database | https://qsartoolbox.org/ |
| Danish (Q)SAR database / (Q)SAR Models | https://qsarmodels.food.dtu.dk |
| NIOSH Pocket Guide | Provided in machine-readable format by NIOSH for this work |
| PubChem Bioassay | https://pubchemdocs.ncbi.nlm.nih.gov/pug-rest-tutorial |
| Comparative Toxicogenomics Database (CTD) | http://ctdbase.org/ |
| European Chemicals Agency (ECHA) Dossier Study Results obtained via International Uniform Chemical Information Database (IUCLID) | https://iuclid6.echa.europa.eu/reach-study-results |
| Integrated Chemical Environment (ICE), Direct Download | https://ice.ntp.niehs.nih.gov/ |

[1]Although the NIOSH Pocket Guide is publicly available, a machine-readable version is not publicly available at time of publication.

**Table C.2. Physicochemical Predictive Tools**

| Tool | Source |
|---|---|
| OPERA | github.com/NIEHS/OPERA |
| EPISUITE | www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface |
| TEST | www.epa.gov/chemical-research/toxicity-estimation-software-tool-test |
| ACD/Labs | www.acdlabs.com/products/percepta-platform/ |
| Cactvs (PubChem) | xemistry.com/faq.htm |
| SwissADME | http://www.swissadme.ch/ |

**Table C.3. Metabolite Prediction Tools**

| Tool | Source |
|---|---|
| Biotransformer Tool | https://biotransformer.ca/ |

| | |
|---|---|
| ADMET predictor metabolism module | https://www.simulations-plus.com/software/admetpredictor/metabolism/ |
| EPA Chemical Transformation Simulator (CTS) | https://qed.epa.gov/cts/ |
| GLORY | https://nerdd.univie.ac.at/glory/ |
| MultiCASE-Meta-Ultra | http://www.multicase.com/meta-ultra |
| Xenosite | https://swami.wustl.edu/xenosite/about |
| SymCyp | https://smartcyp.sund.ku.dk/mol_to_som |
| OECD QSAR Toolbox in vivo Rat metabolism, or<br>Rat liver S9 metabolism simulator | https://qsartoolbox.org/ |
| Gastroplus | https://www.simulations-plus.com/software/gastroplus/ |

**Table C.4. Chemical Databases with Tools for Identifying Analogs**

| Databases | Source |
|---|---|
| PubChem | https://pubchem.ncbi.nlm.nih.gov/ |
| DSSTox | (see EPA CompTox Dashboard)<br>https://comptox.epa.gov/ |
| Chemspider | https://www.chemspider.com/ |
| Zinc | https://zinc12.docking.org/ |
| OECD QSAR Toolbox | https://qsartoolbox.org/<br>In particular: https://qsartoolbox.org/features/grouping/ |
| EPA CompTox Dashboard, GenRA | https://comptox.epa.gov/genra/<br>https://github.com/i-shah/genra-py |
| NTP ICE Chemical Quest | https://ice.ntp.niehs.nih.gov/Tools |

**Table C.5. Tools for Physiologically Based Toxicokinetic Predictions**

| Tool | Source |
|---|---|
| EPA's high-throughput toxicokinetics package (httk) | https://cran.r-project.org/web/packages/httk/index.html |
| GastroPlus | https://www.simulations-plus.com/software/overview/ |
| Simcyp Simulator | https://www.certara.com/software/ |
| PK-Sim and MoBi | http://www.systems-biology.com/products/PK-Sim.html |