

Development of Chemical Categories for Per- and Polyfluoroalkyl substances (PFAS) to Faciliate the Selection of Potential Candidates for Tiered Toxicological Testing and Human Health Assessment

PFAS categories for tiered toxicity assessment

Grace Patlewicz^{a,*}, Richard Judson^a, John Cowden^a, Antony Williams^a, Kelly Carstens^a, Sigmund Degitz^a, Stephanie Padilla^a, Katie Paul Friedman^a, John Wambaugh^a, Barbara Wetmore^a, Stan Barone^b, Jeff Dawson^b, Anna Lowitt^b, Tala Henry^b, Russell S Thomas^a

^aCenter for Computational Toxicology and Exposure (CCTE), US Environmental Protection Agency, Durham, USA,

^bOffice of Chemical Safety and Pollution Prevention (OSCPP), US Environmental Protection Agency, DC, USA,

Abstract

Per- and Polyfluoroalkyl substances (PFAS) are a class of man-made chemicals that are in widespread use and many present concerns for persistence, bioaccumulation and toxicity. Whilst a handful of PFAS have been characterized for their hazard profiles, the vast majority of PFAS have not been extensively studied. Herein, a generalizable chemical category approach was developed and applied to non-polymer PFAS that could be readily characterized by a distinct chemical structure. The PFAS definition as described in the TSCA Inactive Significant New Use Rule (SNUR) was applied to the Distributed Structure-Searchable Toxicity (DSSTox) database to retrieve an initial list of 10,576 candidate PFAS. Plausible degradation products from the 617 PFAS on the non-confidential TSCA Inventory were simulated using the Catalogic expert system, and the unique PFAS degradants (3126) were added to the list resulting in a set of 13702 candidate PFAS. Each PFAS was then assigned into a primary category using Organisation for Economic Co-operation and Development (OECD) structure-based classifications. The primary categories were subdivided into secondary categories based on a chain length threshold (≥ 7 vs < 7). Secondary categories were subcategorized using chemical fingerprints to achieve a balance between total number of structural categories vs. level of structural similarity within a category based on the Jaccard index. A set of 85 structural categories were derived from which a subset of representative candidates could be proposed for potential tiered testing, taking into account considerations such as the sparsity of relevant toxicity data within each category as a whole, presence on environmental monitoring lists, and the ability to identify plausible manufacturers/importers. Refinements to the approach considering ways in which the categories could be updated by new approach method (NAM) mechanistic data and physicochemical property information are also described. This categorization approach shows promise as a means to identify candidates for testing with related applications in PFAS QSAR development, use of read-across as well as targeted evaluation of PFAS.

Keywords: Per- and Polyfluoroalkyl substances (PFAS), Chemical categories, read-across, New Approach Methods (NAMs), tiered testing, Toxic Substances Control Act (TSCA)

*Corresponding author
Email address: patlewicz.grace@epa.gov (Grace Patlewicz)

FUNDING

The work presented in this manuscript was solely supported by appropriated funds of the US Environmental Protection Agency (US EPA).

NOTES

The views expressed in this manuscript are those of the authors and do not necessarily reflect the views or policies of the US Environmental Protection Agency (US EPA). Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

1. Introduction

1.1. Background

Per- and Polyfluoroalkyl substances (PFAS) are a large class of man-made chemicals that have been manufactured and used in a variety of industries since the 1940s^{1,2,3}. PFAS have been or are currently being synthesized for a myriad of different uses, including adhesives, stain resistant coatings for clothes or furniture, fire retardants, and many more uses. In addition to consumer and industrial applications, PFAS are being released into the environment during manufacturing and use [<https://factor.niehs.nih.gov/2019/3/feature/2-feature-pfas>]. PFAS and products containing them are regularly disposed of in landfills or incinerated which can also lead to further release into soil, groundwater, and air^{4,5}. They are also found in biosolids from wastewater treatment facilities which have been spread onto agricultural fields⁶; [<https://www.ewg.org/news-insights/news/2022/04/ewg-forever-chemicals-may-taint-nearly-20-million-cropland-acres>].

Characterizing the scope and scale of the 'PFAS class' has been challenging in the absence of a harmonized PFAS definition. Some have cited thousands of PFAS being in the environment (estimates range from 4700 to greater than 9000) (See <https://www.scientificamerican.com/article/forever-chemicals-are-widespread-in-u-s-drinking-water/>), but there is likely to be an increasing number identified given that analytical methods are continually being evolved to detect them. An Organisation for Economic Co-operation and Development (OECD) working group defined PFAS as 'fluorinated substances that contain at least one fully fluorinated methyl or methylene carbon atom (without any H/Cl/Br/I atom attached to it); that is, any chemical with at least a perfluorinated methyl group (-CF₃) or a perfluorinated methylene group (-CF₂-)^{7,8}. This broad OECD definition would make estimates of a few thousand PFAS too low; however, the OECD working group also acknowledges that a chemistry definition of PFAS does not equate to how PFAS should be necessarily assessed in terms of their hazard profile or to what extent subcategorizations of PFAS are appropriate depending on

different legislative frameworks. Indeed, if the OECD definition were applied to a large inventory such as the US EPA's Distributed Structure-Searchable (DSSTox) Database project⁹ estimates of the number of PFAS would be in the range of 30,000 members. For contrast, PubChem's Classification Browser (<https://pubchem.ncbi.nlm.nih.gov/classification/#hid=120>) has tagged over 6.4 million substances as meeting the OECD PFAS definition. Any substance containing a CF3 would be classified as a "PFAS" even though it might fall within the remit of other regulatory frameworks. The US EPA's Office of Pollution Prevention and Toxics (OPPT) recently proposed a structural definition for defining PFAS for the purposes of a rule that requires EPA review and approval prior to manufacturing and/or processing a chemical substance meeting the definition that is not currently "active in U.S. commerce" as provided by TSCA. This proposal is known as the TSCA Inactive Significant New Use Rule (SNUR)¹⁰. This definition is narrower in scope than the OECD chemistry definition yet still identifies several thousand PFAS candidates. Under the proposed TSCA Inactive SNUR, a PFAS is defined as 'including at least one of three substructures: 1) R-(CF2)-CF(R')R'', where both the CF2 and CF moieties are saturated carbons; 2) R-CF2OCF2-R', where R and R' can either be F, O, or saturated carbons; and 3) CF3C(CF3)R'R'', where R' and R'' can either be F or saturated carbons. Of the many thousands of PFAS, few have been studied extensively in terms of their toxicity profile. Beyond PFAS such as perfluoroocanoic acid (PFOA) and perfluorooctane sulfonic acid (PFOS), the vast majority of PFAS lack data to facilitate a robust characterization of their potential toxicity¹¹. In an effort to address these data gaps, Congress directed EPA (15 USC 8962) to develop a process for prioritizing which PFAS or 'class' of PFAS should be subject to additional research efforts based on potential for human exposure to, potential toxicity of, and other available information. This is described in more detail in EPA's National Testing Strategy [<https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/national-pfas-testing-strategy>] that was published in October 2021.

The notion of a 'class' underpins grouping approaches which includes the concept of developing categories to perform associated read-across. Rather than assessing each PFAS individually, closely related PFAS could be, in principle, grouped together into categories. Thus, in a category approach, not every PFAS needs to be tested for every single endpoint. Instead, the overall data for that category could potentially prove applicable to support a hazard assessment for other members of the category.

Grouping approaches have been in use in regulatory programmes for many years dating back to 1998 when guidance was developed by the EPA in support of the US High Production Volume (HPV) Challenge Program (https://hero.epa.gov/hero/index.cfm/reference/details/reference_id/1060798). The concepts of grouping, categories and read-across are extensively described in OECD's grouping guidance document, last revised in 2017¹² and presently undergoing revision. Moreover, the state of art in

read-across has been described extensively in the literature; from workflows which outline the steps undertaken to develop category and analogue approaches through to the evaluation, justification and documentation of the read-across prediction made^{13,14,15,16}. More recently the notion of enhancing structure-based groupings with new approach methods (NAMs) has also been an evolving topic, for example, to what extent can structural categories be further justified by NAM data by providing a mechanistic underpinning^{14,15,17,16}.

This study describes the approach taken to further refine a relevant PFAS landscape of interest to EPA from which an initial set of structural categories were derived. The work here is a continuation of the initial categorization efforts described in the EPA National Testing Strategy. For the set of categories developed, the data gaps were assessed to help identify which categories were particularly data poor (e.g., lacking relevant repeat dose toxicity data) and/or associated with known exposures and therefore would benefit from data collection or new test data generation (using both NAMs or traditional approaches) to better characterize the category as a whole. The aims of this manuscript are as follows:

1. Summarize the process of constructing a PFAS landscape;
2. Profile the PFAS landscape to assign substances into broad structural categories in combination with chain length;
3. Evaluate the degree of structural similarity within each category and determine which categories needed to be further subset to maximize their structural similarity whilst maintaining a pragmatic total number of categories;
4. Facilitate the identification of potential candidate PFAS for tiered testing by capturing additional considerations such as availability of a known manufacturer/importer; Agency and/or State priorities, environmental monitoring information and structural diversity within the category;
5. Evaluate the categories based on their predicted physical state and physicochemical properties;
6. Consider the utility of the structural categories developed in performing read-across, as well as refinements such as incorporating mechanistic and toxicokinetic data derived from NAMs. Separately, EPA has been leading a research programme to test a targeted set of ~150 PFAS through an array of different NAM approaches as part of a category approach^{18,19,20,21,17,22,23,24}. The mechanistic insights derived from this parallel effort offer potential opportunities to refine the structurally-based categories developed.

2. Methods

2.1. Defining the PFAS landscape of interest

To define the PFAS landscape for the purpose of this study, the starting point was to search the DSSTox database^{9,25} using a series of structure-based queries that capture the PFAS definition described in the TSCA Inactive Significant New Use Rule (SNUR)¹⁰. DSSTox forms the basis of the EPA CompTox Chemicals Dashboard (referred to herein as the Dashboard)^{9,25} and comprises 1,200,059 substances (at the time of writing, March 2023). (<https://comptox.epa.gov/dashboard/>). As a result of the search, 10,576 substances were identified as forming the initial PFAS landscape for this study. This set was cross referenced with the TSCA inventory (see Section 2.11) to identify matches. For each of the TSCA PFAS, degradation products were simulated using the biodegradation model, Catalogic 301C v12.17 within the commercial software tool, OASIS Catalogic v5.15.2.14 (University As Zlatarov, Laboratory of Mathematical Chemistry, Bourgas, Bulgaria; <http://oasis-lmc.org/>). The set of PFAS degradation products (3126) for the parent TSCA substances were added to the initial landscape such that the final PFAS landscape used in this study comprised 13,702 substances. Note only degradation products meeting the SNUR definition were considered. Chemicals were represented by unique DSSTox Substance Identifiers (DTXSID)⁹, Simplified Molecular-Input-Line-Entry System (SMILES) (<https://www.daylight.com/dayhtml/doc/theory/theory.smiles.html>), chemical names and CAS Registry Numbers (CASRN). International Chemical Identifier keys (InChIKeys), (hashed InChI)²⁶ were used as identifiers for the degradation products. Chemical substances in the DSSTox database have been curated and standardized to ensure correctness in chemical structure as well as their associations to chemical names and other identifiers such as CASRN. Examples of this curation include checking for errors and mismatches in chemical structure formats and mapping to identifiers, as well as structure validation issues like hyper-valency, tautomerism etc⁹.

2.2. Biodegradation potential

Biodegradation predictions were made for PFAS in the landscape that were on the TSCA inventory using the Catalogic 301C v12.17 model within the commercial software tool, OASIS Catalogic v5.15.2.14. The intent was to enrich the landscape by substances more likely to be found in the environment as they originated from substances in commerce. The biodegradation Catalogic 301C model simulates aerobic biodegradation under Ministry of International Trade and Industry, Japan (MITI) I (OECD 301C) test conditions. The modelled endpoint is the percentage of theoretical biological oxygen demand (BOD) on day 28. The underlying training set for the model comprises BOD data for 2620 substances - 745 of these were collected from the MITI I database and 806 were provided by National Institute of Technology and Evaluation (NITE), Japan. A further 1069 substances that were

proprietary were provided by NITE, Japan. In addition to BOD data, a second database underpinning the model comprised pathways for 783 organic substances, documented pathways for 587 chemicals were collected from the primary and secondary literature whereas pathways for 196 proprietary substances were provided by NITE, Japan. In brief, the Catalogic model comprises a metabolic simulator and an endpoint model. The microbial metabolism is simulated by a rule-based approach based on a set of hierarchically ordered transformations and a system of rules controlling the application of these transformations. Recursive application of the transformations allows for the simulation of metabolism and generation of biodegradation pathways. Calculation of the modelled endpoint is based on the simulated metabolic tree and the material balance of transformations used to build the tree. Predictions were made for all PFAS in the landscape that were on the non-confidential TSCA inventory (see Section 2.11 for more details). Prediction results containing the list of simulated metabolites (as SMILES) along with their DTXSID identifiers were exported as a text file. Prediction results were then processed in the following manner:

1. DTXSID identifiers were extracted for each parent substance and mapped to each metabolite. This ensured for a given parent, all metabolites could be readily associated with its corresponding parent substance.
2. A new identifier was then created for the metabolites based on the parent DTXSID identifier. That is to say, the first metabolite simulated for parent DTXSID9065256 would be tagged as DTXSID9065256_m_1 and so on. This would provide an approximate means of tracking where in the simulated pathway a metabolite appeared and which parent it was associated with.
3. SMILES that were generated by OASIS software are non-standard in their format and not readable by other cheminformatics software. This was a major limitation as without standardization, no further computation whether that be assigning metabolites to OECD categories or calculating any properties was possible. Using a series of manual ad hoc rules, the SMILES generated by the OASIS Catalogic software were converted to a standardized form. This involved transforming stereochemistry annotations and salt form (cation and anion) representations.
4. InChIKeys were then generated for all standardized SMILES, parents and simulated metabolites. Use of InChIKeys provided an unambiguous means of structurally representing the substance (rather than using SMILES that are potentially non-unique) and enabled subsequent associations to be derived between substances. The processed results were saved for subsequent analysis.

Many degradation products were found to be common across parent substances. Grouping by InChIKeys created a set of unique degradation products. These were filtered to remove non PFAS degradation products leaving a set of 3126 degradation products that were added to the starting

landscape of 10,576 substances.

To explore the coverage and relevance of the MITI training set within the Catalogic 301C model relative to the PFAS on the TSCA inventory substances, a comparison was performed to evaluate the number of PFAS substances within the training set as well as to assess the overlap in structural space as characterized by Morgan chemical fingerprints²⁷ (see Section 2.4 for details on chemical fingerprint generation). In the latter case, this structural space was projected onto a 2-dimensional (2D) scatterplot (see Figure A1) using a t-distributed stochastic neighbor embedding (t-SNE) to facilitate visualization²⁸.

2.3. Profiling PFAS into categories

2.3.1. Primary structural categories

This study aimed to develop a hierarchy of PFAS categories starting with a handful of large, diverse categories that could be subcategorized into more structurally similar categories based on other considerations (e.g., chemical fingerprints, chain length). To that end, primary categories were derived by profiling the PFAS landscape of 13,702 substances through the database framework developed by Su and Rajan²⁹ called "PFAS-Map" (https://github.com/MatInfoUB/PFAS_Map_MaDE_UB). As described in Su and Rajan²⁹, PFAS could be classified into one of at least nine broad primary categories:

- PFAS derivatives
- PFAAs
- PFAA precursors
- Non-PFAA perfluoroalkyls
- FASA-based PFAA precursors
- Fluorotelomer-based PFAA precursors
- Silicon PFAS
- Side-chain fluorinated aromatics PFAS
- Other aliphatic PFAS

PFAAs makes reference to perfluoroalkyl acids whereas FASA denotes perfluoroalkane sulfonamides. The PFAS landscape was also processed through the OPEn structure-activity/property Relationship App (OPERA) v2.8 tool³⁰ (<https://github.com/kmansouri/OPERA>) to derive QSAR READY SMILES and selected physicochemical property predictions (as discussed in Section Section 2.13). These are standardized SMILES where salts and stereochemistry are removed. QSAR READY SMILES could be generated for 13,449 substances of the PFAS landscape. Substances without QSAR READY SMILES or which could not be computationally resolved were assigned as

"Unclassified". In practice, when substances are batch processed by PFAS-Map, a first class and second-class classification are designated. For this study, the first-class designation was used as the initial primary category assignment which was then converted using a simple mapping dictionary to ensure that substances such as "Fluorotelomer PFAA precursors, cyclic" were aggregated to "Fluorotelomer PFAA precursors". This was performed to limit the number of primary structural categories which themselves were largely consistent with the hierarchy described by OECD in their Terminology 2021 guidance⁷.

2.3.2. Secondary structural categories

It is hypothesized that the length of carbon chain influences differences in toxicity and the length of time the chemical spends in the body and environment. This supposition has been underpinned by experiences with PFAAs^{31,32}. Due to the potential importance of chain length in the toxicity, persistence and bioaccumulation of PFAS, secondary structural categories were defined using a carbon chain length threshold.

Chain length determination

The maximum number of contiguous CF2 groups in a chain was determined for all 13,702 substances. This was achieved by iterating through a range of CF2 chain lengths and determining the longest chain length where there was a match with the corresponding substance. For instance, Perfluorosebacamide [DTXSID40380015] contains 8 contiguous CF2 units; hence, its chain length was denoted as 8. PFOA had a maximum chain length of 7 whereas PFOS had a maximum chain length of 8. For PFOA, although there are 8 carbons in its backbone, the 8th is part of the carboxyl group whereas in PFOS, there are 8 CF2 groups plus the sulfonate group.

For the current analysis, the chain length threshold was set at 7 (≥ 7 vs < 7). This threshold served as a pragmatic representation of what constitutes a "long chain" PFAS. The chain length threshold is broadly consistent with the EPA's 2009 PFAS action plan (https://www.epa.gov/sites/default/files/2016-01/documents/pfcs_action_plan1230_09.pdf). A PFAS with a maximum CF2 number greater than or equal to 7 was denoted "gte7". Using this threshold, both PFOS and PFOA would be assigned to the "gte7" secondary category. A PFAS with a maximum number of CF2 groups less than 7 was denoted "lt7". Defining chain lengths for PFAS with non-contiguous chains or branching is less clear but has been evaluated in more detail by Richard et al^{33,34} through the development of new PFAS specific chemical fingerprints, so-named PFAS ToxPrints, as an extension of the logic used to develop the original ToxPrints that had been defined for a broader chemistry³⁵.

A secondary category was thus denoted by its PFAS-Map assignment, akin to the primary OECD structural classification and a carbon chain length threshold e.g., Perfluorooctanesulfonamidoacetic

acid [DTXSID40440941] would thus be described as belonging to the "FASA based PFAA precursors, gte7" secondary category.

2.3.3. Terminal structural categories

The underlying motivation for the study was to identify categories that would balance maximizing structural similarity that could permit read-across within those categories versus pragmatism in terms of total number of categories. Too many categories with very few substances renders the approach less generalizable, too few categories could result in extrapolating between substances that were not sufficiently similar. To that end, an objective threshold was needed to determine how granular categories needed to be to manage this trade-off and ensure that the categorization was actionable. An objective threshold was developed, described in Section 2.6, that compared structural similarity within a category relative to the structural similarity between different categories.

2.4. Chemical fingerprints

Morgan chemical fingerprints²⁷ were calculated for all substances within each secondary category using the open-source Python library RDkit (Landrum, rdkit.org) with a radius of 3 and a bit-length of 1024. Fingerprints capture presence (denoted by 1) and absence of features (denoted as 0) and hence, by their nature they represent a binary dataset. These fingerprint (FP) files were stored for subsequent processing. Morgan fingerprints also known as Extended Connectivity Fingerprints (ECFPs) are widely used in machine learning applications for cheminformatics³⁶ especially when ranking diverse structures by similarity. These circular fingerprints map the molecular environment of every atom.

2.5. Chemical similarity

Pairwise distance matrices were calculated for each secondary category. These were generated by using the chemical fingerprint files as inputs and computing the Jaccard distance for each pair of substances. The Jaccard distance captures the proportion of FP bits between 2 substances that differ (https://en.wikipedia.org/wiki/Jaccard_index). The Jaccard distance ranges from 0 to 1 where 0 would indicate zero distance (or high similarity) and 1 would indicate high distance (or low similarity).

Distance matrices were computed for all secondary categories and stored for subsequent processing. These are referred to as 'within category' distance matrices in Section 2.6.

2.6. Objective distance threshold

The rationale underpinning the objective distance threshold was based on the expectation that the variance in the distribution of the pairwise distances for each secondary category representing the 'within category' similarity would be lower than distributions of the pairwise distances between different secondary categories ('between category'). The 'within category' distances had already been

computed as described in Section 2.5.

'Between category' combinations aimed to identify categories that did not share the same primary category root. A list of all possible binary combinations of secondary categories was created using the names of the secondary categories, "Other aliphatics, lt7" and "PFAAs, gte7" is an example of such a binary combination. These were then filtered to remove secondary categories that shared the same primary category root (i.e., a combination such as "PFAAs, lt7" and "PFAAs, gte7" would be excluded from consideration as a 'between category'). Chemical fingerprint datasets for each binary combination were created by joining the secondary category chemical fingerprint datasets together. Pairwise distance matrices were then derived for the combined category set. These were filtered to retain only the pairwise distances between starting secondary categories.

The empirical cumulative distribution functions (ECDFs) of the pairwise distances were calculated for each secondary category (see Figure A2 for a plot of the ECDFs). ECDFs were also derived for the 'between category' combinations (see Figure A3 for a plot of the first 10 ECDFs). The ECDFs permitted a visual inspection of the range of the pairwise distances across all secondary categories as well as across all 'between category' combinations. Based on visual inspection of the ECDFs, the median value for each distribution was selected as the summary metric.

Probability density functions of median values from all within and between secondary categories were plotted to explore their overlap. The 5th percentile of the 'between categories' distribution was defined as the threshold to determine whether a secondary category merited further subcategorization. A secondary category was only subcategorized if the median of its 'within category' pairwise distance distribution exceeded this threshold.

2.7. Terminal categories

Secondary categories that exceeded the threshold were subcategorized using agglomerative hierarchical clustering. The condensed form of the pairwise distance matrix computed for each secondary category that exceeded the threshold was used as an input into a hierarchical clustering using Ward's method³⁷. Ward's method is a criterion that minimizes the total within-cluster variance. For each secondary category, the dendrogram was plotted and the number of first-generation clusters was set as the maximum cluster number. Clusters were labelled as 1,2,3 etc. Each of the clustering results were combined into one table which was then merged with the starting table of primary and secondary categories.

The next generation of categories, quaternary categories, would then be processed in the same manner to determine whether any exceeded the objective threshold and needed to be subcategorized further as already described. In practice, a maximum of two generations of subcategorizations were

performed, with the expectation that this would balance the structural similarity within the category relative to total number of terminal categories.

Secondary categories or tertiary categories which did not exceed the threshold were ultimately denoted as the terminal category. Thus, a terminal category could be tagged as "PFAA, lt7", effectively a secondary category, or could be tagged as "Other aliphatics, lt7, 1", a tertiary category or following two iterations of subcategorization would be tagged as "Other aliphatics, gte7, 1, 2". Note: in the data files and figures, terminal categories without one or two iterations of subcategorization are denoted as "PFAA, lt7, nan, nan" or "Other aliphatics, lt7, 1, nan" where "nan" represents a null value.

The conceptual workflow for creating the PFAS structural categories is summarized in Figure 1.

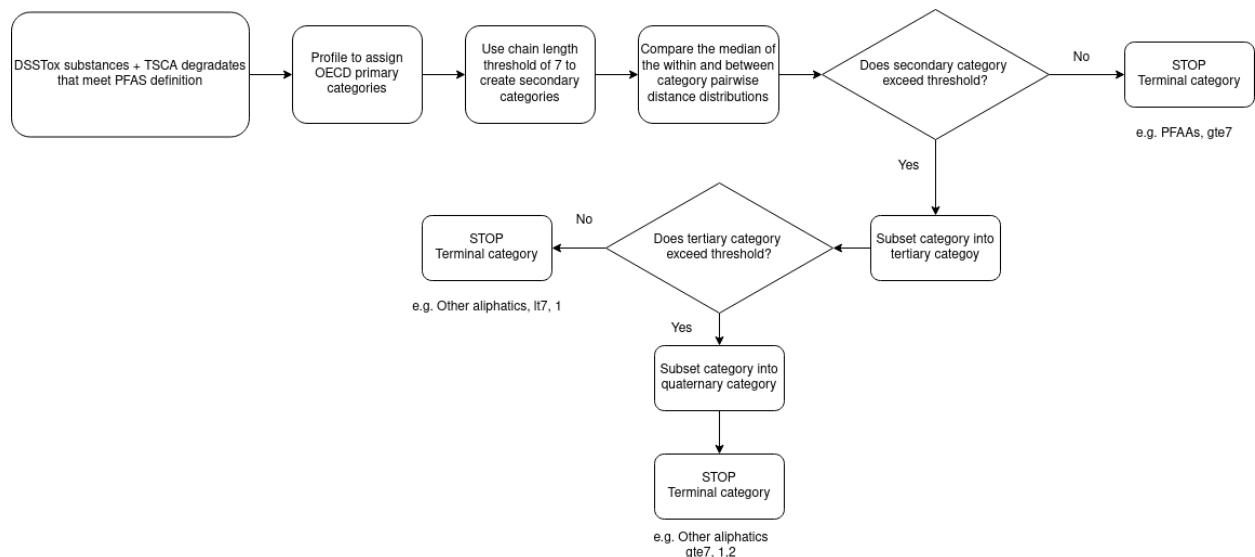


Figure 1: Conceptual workflow for generating PFAS structural categories

2.8. Centroid calculation

For each terminal category, a single substance was identified that was nominally representative of the category. This substance was the computed centroid calculated from the Jaccard pairwise distance matrices (see Section 2.5). The sum of the pairwise distances across all substances for a given structural category was computed and the substance with the minimum value was denoted as the centroid (i.e., this substance would have the lowest distance from all other category members). Technically, this calculation gives rise to the medoid of a cluster. However, for the purposes of this analysis, the term centroid is used for convenience to denote it as the 'middle' substance within the category. Distances of all category members relative to the centroid substance were also computed.

2.9. Selecting additional candidates beyond centroids

Since many of the terminal categories were likely to be large in size, a single substance might be insufficient to both characterize the category and its potential hazard profile. To address this limitation, the MaxMinPicker approach, as implemented within the RDKit Python library, was applied to identify additional substances which would in turn capture the breadth and diversity of each category³⁸. The MaxMinPicker approach proceeds as follows:

1. Molecular descriptors are generated for all substances. In this case, the Morgan fingerprints calculated for all substances within a terminal category represented the candidate pool whereas the pre-computed centroid equated to the initial seed.
2. From the substances in the terminal category, the substance that had the maximum value for its minimum distance to the picked set (initially this would be just the centroid) would then be identified. This substance would be the most distant one to those already picked so it would be transferred to the 'picked set' (now centroid + 1).
3. An iteration back to step 2 would then be performed until the desired number of substances were picked.

The MaxMin approach is a well-established algorithm for dissimilarity-based compound selection that has been applied in drug discovery for many years. The reader is referred to Snarey et al³⁹ for a comparison of the different algorithms.

The MaxMinPicker was applied to all terminal categories containing more than 5 members to select the next 3 most diverse substances within a category (centroid + up to 3 additional substances). The intention of picking additional diverse substances was to help bound the domain of the structural category. The selection of 3 most diverse substances was chosen out of convenience to propose an actionable number of additional substances. A systematic evaluation of the relationship between the number of diverse substances that could be picked relative to the structural diversity within each terminal category was also undertaken. This was approached as follows, first the ranked order by diversity of all members within a terminal category was computed. Then the pairwise distance matrices derived in Section 2.5 were filtered by the diverse substances, starting from the centroid, centroid plus first diverse chemical through to the complete set of category members. At each step the mean minimum distance was recorded. This enabled the construction of a matrix to capture the mean of the minimum pairwise distances relative to the number of diverse chemicals selected. The normalized cumulative sum of all the mean minimum distances was then computed. This provided a means of evaluating the proportion of structural diversity that was captured as a function of the number of MaxMin substances selected.

Two aspects could be assessed objectively from this calculation, namely:

1. how much structural diversity was being captured by the 3 diverse picks originally identified; and
2. how many diverse substances should ideally be selected (if practical resources were not a limiting factor) that would capture a specified level of the structural diversity. For example, how many substances would need to be selected if capturing a specific percentage of the structural diversity within a terminal category was desired, 80% is presented for illustrative purposes.

2.10. Facilitating the identification for potential candidates for tiered testing

To facilitate the identification of potential candidate PFAS for tiered testing, availability of a known manufacturer/importer, Agency and/or State priorities, environmental monitoring information were evaluated as additional considerations. These are described in turn.

2.11. Qualitative list designations

Several indicators were added to the landscape to identify substances based on their TSCA inventory status, Chemical Data Reporting under TSCA, or State/Region priorities.

The non confidential (Non-CBI) TSCA Inventory active and inactive lists were downloaded from the Dashboard and combined into one large set. Substances within this inventory are annotated by Chemical Abstract Service (CAS) Registry Number, Chemical Abstracts (CA) Index Name, and DSSTox substance identifier (DTXSID). These were matched by the DTXSID identifiers already captured in the PFAS landscape. Substances were tagged as 'inactive', 'active' or 'unclassified'. Note the predicted degradation products of substances tagged as either "inactive" or "active" had been used to augment the PFAS landscape as already described in Section 2.1.

The CDR data from 2020 was downloaded from the public EPA web address (<https://www.epa.gov/chemical-data-reporting/access-cdr-data>). The data is structured in several files depending on the granularity of information captured. The CDR data comprises information for a set of 8660 substances. DTXSID identifiers were available for 8017 of these substances when using the Batch search functionality within the Dashboard. A tag was created for CDR2020 status if a PFAS was found to be associated with a CDR record. National Aggregated Production Volume (National Agg PV) data was also extracted to highlight how this was distributed across primary categories. Since some of the production volume (PV) data was numeric and some represented in numeric ranges, the PV data was summarized into one of 10 different ranges (<25,000 lbs, <1,000,000 lbs, 25-<100,000 lbs, 100,000 <500,000 lbs, 500,000 -<1,000,000 lbs, 50,000,000 -<100,000,000 lbs, <1,000,000 lbs, 100,000,000 -< 1,000,000,000 lbs, 20,000,000 -< 100,000,000 lbs, 1,000,000 -< 20,000,000 lbs, 1,000,000 -<10,000,000 lbs).

Various Regions or States have identified PFAS of interest based on validated analytical methods or for environmental monitoring purposes. The data sources captured as part of the EPA's PFAS Analytic Tools website (<https://echo.epa.gov/trends/pfas-tools#data>) were used to construct lists of such PFAS. The specific data sources were Discharge Monitoring Data, Drinking Water (State) Data, Drinking Water (UCMR) Data, Environmental Media Data, Production Data, Toxics Release Inventory (TRI) Data - Waste Managed, TRI Data - On-Site, TRI Data - Off-Site and Production Data (all accessed 7th April 2023). Discharge Monitoring data is collected by virtue of the National Pollutant Elimination System permit. Drinking Water Data comprises UCMR, Unregulated Contaminant Monitoring Rule data and State level monitoring data. Environmental Media data comprises ambient sampling data reported by federal, state, tribal and local governments, academic and non-governmental organizations, and individuals that are submitted to the Water Quality Portal (WQP). Production data entails information reported under the Chemical Data Reporting (CDR) Rule under TSCA. TRI tracks the management of certain toxic chemicals that may pose a threat to human health or the environment by more than 21,000 facilities throughout the US and its territories. The National Defense Authorization Act of Fiscal year 2020 (NDAA) added certain PFAS to the TRI list and provided a framework for the ongoing listing of additional PFAS.

Identifiers were extracted from these source files and searched against the Dashboard to map to DTXSID records. The set of identifiers (Names and CASRN) within the entire PFAS landscape were also queried against PubMed, the National Library of Medicine's citation index for biomedical literature to determine whether a substance might have been studied in the literature. This was queried using the Landscape feature within the Abstract Sifter v7.5⁴⁰. Each of these respective priorities were then matched to the PFAS landscape to provide another attribute for consideration when selecting potential candidates for testing.

2.12. Constrained PFAS Landscape

One of the main limitations of the identification of centroids and additional diverse substances was that they might yet not yield feasible candidates due to the lack of assignable manufacturer/importer. This was articulated as potential challenge in the Testing Strategy. To address this practical constraint, the same process of computing centroids, identifying additional diverse substances and evaluating their structural diversity coverage was also performed using the terminal categories as a basis as described in Section 2.7 but constraining the landscape to only those substances on the TSCA inventory and specifically those substances that were actives on the TSCA inventory. This would provide some more flexibility in terms of selecting substances for testing that were already in commerce and/or could be more readily procured.

2.13. Evaluation of predicted physicochemical properties

The route of exposure and the types of toxicological tests that may be needed to understand the health risks associated with a substance is partly dependent on the physical state and physicochemical properties. Physicochemical properties were predicted using the open-source OPERA v2.8 tool³⁰ for all substances with QSAR READY SMILES. Predictions were possible for 13,449 substances in the PFAS landscape (as discussed in Section 2.3.1). Properties predicted were melting point, boiling point, Henry's Law constant (HLC), water solubility and vapour pressure. For the physical state, a melting point less than 25 deg C would be indicative of a liquid whereas a value greater would be a solid and a boiling point less than 25 deg C would be a gas. These are the guiding principles underpinning the EPA's Sustainable Futures Framework guidance (see https://www.epa.gov/sites/default/files/2015-05/documents/05-iad_discretes_june2013.pdf). A water solubility threshold of 100 mg/L was used to denote whether a substance was soluble/insoluble whereas a vapour pressure threshold of 75 mmHg or a HLC of 0.1 atm m³ mol⁻¹ determined volatility. Based on these properties, each substance was assigned into 1 of 4 'physchem testing tracks' (from A-D). Track A aimed to capture substances that were insoluble solids, B to identify both soluble solids and soluble non-volatile liquids, whereas C tagged soluble volatile liquids/insoluble liquids and soluble gases. Track D assigned substances as insoluble gases or highly volatile gases. Substances that could not be assigned into one of these 4 tracks was tagged with a 'no testing track' flag. For each of the terminal structural categories, Morgan finger-print representations were projected into two dimensions using a t-distributed stochastic neighbor embedding (t-SNE) to facilitate visualization²⁸. The projections were plotted as 2D kernel density distributions overlaid with physchem testing track information to help explore the extent to which members were assigned to the same track and therefore had a consistent physchem profile across a given terminal category.

2.14. Evaluation of variance of in vivo toxicity within terminal categories

Ultimately, read-across of data within categories could be performed such that the hazard profile of the category is adequate without needing to test a significant number of category members. To evaluate the feasibility of performing read-across within the terminal categories derived, an exploration of the distribution of in vivo points of departure (PODs) within and across terminal categories was performed for 2 routes of exposure; oral and inhalation.

2.14.1. Variance of in vivo PODs across and within terminal categories

From ToxValDB version 9.4, the Toxicity Values Database (Judson et al., in prep), all studies where 'oral' or 'inhalation' was the route of exposure were extracted. Only records where a point of departure was reported as a NOEL, NOAEL, NOAEC, LOAEL, LOEL, LOAEC and where the dose units

were expressed as mg/kg-bw/day or mg/m³ were retrieved. Study types were also restricted to the following: 'chronic', 'developmental', 'reproduction', 'reproduction developmental', 'subchronic', 'neuro-toxicity', and 'short-term' as captured in the 'study type' field within the database. The toxicity data extracted for each route of exposure was then merged with the PFAS substances from the landscape. The minimum study-level point of departure (POD) was taken. Aggregating all minimum values to a single value per substance depended on the availability of NOAEL(C)-type or LOAEL(C)-type data and was performed as follows. Minimum NOEL(C)/NOAEL(C) values were preferentially used for a substance but if these were not available, LOEL(C)/LOAEL(C) values were taken and adjusted by a factor of 10. The 10th percentile of all values was then reported for a substance irrespective of its study type or duration. Whilst the approach of aggregating the available information across study types and durations is a simplified assumption, the summary value provides an estimate of the POD for each substance and the expected level of variation across and within categories. Box and whisker plots were created to reflect the distribution of the PODs across the terminal categories for each route of exposure. Strip plots were overlaid to show the variation of chain length across a given terminal category for the oral route of exposure. Inspection of the plots was intended to facilitate a coarse grain evaluation of the extent of the variation in PODs as a function of chain length and category. PODs were plotted on a log scale.

2.15. Potential structure category refinements by considering mechanistic and toxicokinetic data derived from NAMs: Qualitative NAM flags

A summary of the NAM testing being undertaken for ~150 PFAS was described in Patlewicz et al.¹⁷. See Houck et al.¹⁹ for results from various nuclear receptor and oxidative stress targeted assays, Houck et al.²⁰ for 12 human primary cell-based assay models of pathophysiology including immunosuppression, Carstens et al.¹⁸ for the developmental neurotoxicity assays, and, for toxicokinetic information, Smeltz et al.²² and Kreutz et al.²¹. Manuscripts describing the remaining data streams (zebrafish developmental toxicity and thyroid pathway assays) are in preparation.

In addition to the NAM testing, a quality control (QC) evaluation of the High Throughput Screening (HTS) stocks was undertaken to confirm PFAS analyte presence and stability²³. This evaluation was warranted given recent reports of certain PFAS degrading in the aprotic solvent dimethyl sulfoxide (DMSO), readily used as the solvent of choice in HTS^{41,42}. Two hundred and five PFAS selected based on criteria described in Patlewicz et al.¹⁷ were evaluated using low resolution tandem mass spectrometric detection strategies to confirm presence of intended analyte, evaluate analyte stability and presence of isomers, and verify stock concentrations for a subset for which commercially available verified standards were available. Ultimately 57 PFAS failed QC evaluation, with three exhibiting degradation in DMSO and the

remainder not detected as present, likely due to volatilization. The pass/fail scores and informational flags as described in Smeltz et al. [smeltz_targeted_2023], and can be downloaded from https://epa.figshare.com/articles/dataset/Chemistry_Dashboard_Data_Analytical_QC_for_PFAS/22118099.

For each of the NAM data streams, substances were tagged with a qualitative flag to indicate the class of mechanistic information that could be derived from the associated assay outcome (e.g., estrogen receptor activity from a nuclear receptor assay) and an expert-derived qualitative level of confidence associated with the outcome (high, medium or low). Only NAM results from substances that passed QC were carried forward. These flags were considered as an additional line of evidence to determine whether a terminal category might merit being split based on its mechanistic or toxicokinetic information or to inform what types of higher order testing might be most impactful for a given substance drawn from said terminal category. Each set of flags are described in turn. Confidence scores across the NAM flags were standardized as appropriate to facilitate visualizations across data streams. Each flag could take on one of three values, low, medium and high concern, color coded as blue, yellow and red. The flag categories are summarized below in Table 1.

Table 1: Summary of NAM Flag Rationales

Technology	Low Concern (Blue)	Medium Concern (Yellow)	High Concern (Red)
Nuclear Receptors	No nuclear receptor activity	Activity against at least one of the receptors ER, PPARA, PPARG, PPARD, NFE2L2, PXR, RARG, RXRB at the level of two samples one assay or one sample in 2 orthogonal assays.	Activity in the yellow medium concern that is confirmed using the Eurofins assays
DNT	No activity or activity was only observed at the highest concentration related to cytotoxicity	Low number of hits but which demonstrated selective bioactivity	Moderate to high bioactivity (as measured by hitcall) and demonstrated selective bioactivity (activity below cytotoxicity AC50 as measured by AUC) and median AC50 < 10 μ M

Technology	Low Concern (Blue)	Medium Concern (Yellow)	High Concern (Red)
Zebrafish	Development was normal in all larvae	Test results were equivocal or if only 10-33% of the larvae were affected	Positive activity (i.e., elicited death, non-hatching, or malformations in at least 50% of the animals)
Thyroid	No activity greater than 50% of the model inhibitors/binders	Activity greater than 50% of the model inhibitors/binder, but the concentration necessary to result in this activity was 2 orders of magnitude higher than the model inhibitors/binders	EC50s that were within 2 orders of magnitude of the model inhibitors/binders
Immune	Selectivity scores less than $0.25 \log_{10} \mu\text{M}$	Selectivity scores of greater than $0.25 \log_{10} \mu\text{M}$	
TK	TK_PlasBInd_High: Plasma protein binding higher than 50% of non-PFAS chemicals (TK_PlasBind) ($f_{up} < 0.11$) (this corresponds to 25th percentile of PFAS ($f_{up}<0.10$)	TK_PlasBInd_Higher: Plasma protein binding higher than 50% of PFAS chemicals ($f_{up} < 0.0109$)	TK_PlasBInd_Highest: Plasma protein binding higher than 75% of PFAS ($f_{up} < 0.0039$)
Intrinsic Clear- ance (TK_Metab)	TK_Metab_Moderate: Clint in upper 75th percentile of exp PFAS data (Clint>5.97 ul/min/million heps). Max Clint = 49.86 TK_Struc_Endo	TK_Metab_Slow: Clint < 5.97 ul/min/million cells (lower 75th percentile) Non-fluorinated structure is similar to endogenous chemicals More likely to be a transporter substrate	TK_Metab_Stable: Stable in in vitro hepatocyte incubation (Clint = 0 or Clint.pvalue > 0.05)

2.15.1. In Vitro Assay Data Processing

All in vitro data (excluding the toxicokinetic data) was processed in a standard fashion, described here. Chemicals were run in each assay in concentration-response, with a range from approximately 0.01-100 mM; in some cases, single-concentration screening preceded multi-concentration screening to facilitate prioritization of chemicals for screening. Data were processed through the ToxCast Pipeline (tcpl)⁴³ (version 2.1.0) to generate concentration-response curves and estimate potency and hitcall values. Each concentration-response data set was fit to a constant, a Hill curve, and a gain-loss model (combining a rising and a falling Hill curve). The curve with the lowest Akaike Information Criteria (AIC) was selected as the winner. If the winning curve was either a Hill or gain-loss and the top of the curve exceeded a specified noise threshold, the hitcall for the curve was set to 1 (active), and otherwise set to 0 (inactive)⁴³. For most technologies, the potency is expressed as the AC50 (50% maximal activity value). The BioSeek potency metric is a LEC (lowest effect concentration), which is the first concentration at which the response exceeds the activity threshold. The reason for this is that data are generated at only 4 concentrations. The Zebrafish developmental data was treated somewhat differently, the open source R library tcplFit2 was used to curve-fit the data⁴⁴. This includes constant (response = 0), Hill, gain-loss, first- and second-order polynomials, a power function and 2, 3, 4 and 5 parameter exponential models, and continuous hitcall parameter that ranges from 0-1. For the purposes of the current analysis, a hitcall > 0.9 was considered active and hitcall ≤ 0.9 was inactive. With this application of tcplFit2, the potency is expressed as a benchmark dose (BMD) based on a benchmark response of 1.349 standard deviations from the median response of the two lowest concentrations in the index. Unless otherwise noted, hitcalls were not modified based on cytotoxicity considerations.

2.15.2. Nuclear Receptor Activity

A total of 142 PFAS were evaluated for nuclear receptor activity using an in vitro screening platform that consisted of two multiplexed transactivation assays in human hepatoma cell line HepG2, together encompassing 81 diverse transcription factor activities via the Attagene platform¹⁹. Concentration-response modeling was conducted as described above, with the design accommodating testing concentrations up to 300 μM, with assays run in both TRANS and CIS modes in duplicate. Houck et al.¹⁹ reported activity for the following nuclear receptors among the tested PFAS: Estrogen receptor (ESR1), NRF2 (NFE2L2, a sensor for oxidative stress), PPAR-alpha, -gamma and -delta (PPARA, PPARG, PPARD), PXR (NR1I2), RXR-beta (RXRB) and RAR-gamma (RARG). In addition to the Attagene data for the ESR1 target, Houck et al.¹⁹ also provided information in the form of the human T-47D estrogen-sensitive cell proliferation assay that was conducted using real-time impedance measurement as the readout from ACEA Therapeutics, Inc. (also up to 300 μM). Confirmatory assays were

run on subset of the hits in the initial screen for ESR1, NFE2L2, PPARA, PPARAG, RXRB and RARG. This process used orthogonal assays from Eurofins Inc. (<https://www.eurofins.com>). More details of this data are given elsewhere (Judson et al. in prep). These assays used several different targeted technologies, including cell-free assays that examine coactivator recruitment to indicate receptor activation using fluorescence resonance energy transfer; cell-based protein-protein interaction assays with receptor complex activation indicated by radioligand detection; and a cell-based nuclear factor erythroid 2-related factor (Nrf2) nuclear translocation assay using chemiluminescence as a marker of increased oxidative stress response (Table 2).

Table 2: Additional confirmatory assays from Eurofins

Eurofins Assay Catalog		
Target	Number	Item Name
ESR1	86-0003P-2452AG	ERalpha Human Estrogen NHR Cell Based Agonist Protein-Protein Interaction Assay
ESR1	311410-0	ERalpha Human Estrogen NHR Functional Agonist Coactivator Assay
NFE2L2 (Keap1-Nrf2)	86-0015P-2473AG	Keap1-Nrf2 Human Transcription Factor Cell Based Agonist Nuclear Translocation Assay
PPARA	86-0003P-2459AG	PPARalpha Human NHR Cell Based Agonist Protein-Protein Interaction Assay
PPARA	338210-0	PPARalpha Human NHR Functional Agonist Coactivator Assay
PPARA	2811	PPARalpha Human NHR Functional Agonist Coactivator Assay
PPARG	86-0003P-2460AG	PPARgamma Human NHR Cell Based Agonist Protein-Protein Interaction Assay
PPARG	338250-0	PPARgamma Human NHR Functional Agonist Coactivator Assay
PPARG	2771	PPARgamma Human NHR Functional Agonist Coactivator Assay
RARG	338800-0	RARgamma Human Retinoic Acid NHR Functional Agonist Coactivator Assay
RXRB	338940-0	RXRbeta Human Retinoid X NHR Functional Agonist Coactivator Assay

PFAS selected for confirmation were required to meet minimal criteria for activity in the initial screen. The substance was required to satisfy one of two types of rules: "single assay, multiple hits", meaning that the same endpoint was assessed with two independent samples and both samples of the same chemical were active in the endpoint; or "two assays same target" where a single chemical sample was active in both endpoints. The first rule could be satisfied by any one of the Attagene (CIS or TRANS) or ACEA endpoints so long as the two independent samples were positive in the same assay. The second rule could be satisfied if both the Attagene CIS and TRANS endpoints were active for a chemical-target pair for a single sample, or if one of the Attagene ESR1 endpoints and the ACEA ESR1 endpoint were together active for a single replicate. Additionally, a substance had to have an average potency <100 μ M, and with a Z-score of the potency relative to cytotoxic burst >1⁴⁵.

For each assay, two blinded positive reference chemicals were assessed, as well as two putative negatives. The positive reference chemicals were selected based on public activity data in RefChemDB⁴⁶ and availability of samples in the current ToxCast chemical inventory. This last constraint meant that some of the positive controls were not the most potent available. The putative negative chemicals were PFAS that had passed analytical QC, had shown activity in some other nuclear receptor assays, but were negative in the assays for which they were used as negative controls. Only assays with activity in the positive reference chemicals were included. The counts of chemicals tested and confirmed are given in Table 3.

Table 3: Counts of active and inactive chemicals per target in the confirmatory assays. A chemical was classified as active if it was active in at least one of the confirmatory assays for the target.

Target	Active	Inactive
ESR1	11	9
NFE2L2	1	2
PPARA	6	2
PPARG	1	1
RARG	0	2
RXRB	0	3

For this analysis, all of the original nuclear receptor actives were included as medium flags ("yellow"), but those that confirmed using a Eurofins assay were elevated to high concern ("red"). An additional three substances were active in both Attagene ESR1 endpoints and the ACEA endpoint, and these were flagged as having higher confidence ("red"). PFAS that were inactive in Attagene and ACEA were given low ("blue") flags.

2.15.3. Developmental Neurotoxicity (DNT) Activity

A set of 160 PFAS were screened in a DNT assay battery, including four assays from two assay technologies: the microelectrode array (MEA) technology and the high-content imaging (HCI) technology. The assays modelled four distinct neurodevelopment processes: proliferation, apoptosis, neurite outgrowth, and neural network formation¹⁸. The MEA neuronal network formation assay includes 17 endpoints measuring decreased neuronal network development and function and 2 endpoints measuring cytotoxicity. The HCI assays different cell-based models of neural cells to evaluate proliferation, apoptosis, and neurite outgrowth in 8 different endpoints¹⁸.

DNT bioactivity flags were determined using several metrics, including hitcall determination, potency as the AC50, and selective bioactivity (activity occurring below the cytotoxicity threshold as defined by concurrently run cytotoxicity assays). Based on these three metrics, a chemical was binned into one of four flag categories: 0 - inactive or equivocal, 1 - non-selective bioactivity, 2 - low selective DNT activity and 3 - clear selective DNT bioactivity. Clear DNT bioactivity was assigned a level 3 (high confidence) if there was moderate to high bioactivity (as measured by hitcall) and demonstrated selective bioactivity (activity below cytotoxicity AC50 as measured by AUC) and median AC50 < 10 µM. Low DNT activity was assigned a 2 (moderate confidence) if there were a low number of hits but which demonstrated selective bioactivity. Non-selective DNT bioactivity assigned a 1 (low confidence) if activity was only observed at the highest concentration related to cytotoxicity. Inactive or equivocal assigned a 0 were for zero hits or borderline curves determined to be inactive by expert review. For the current analysis, chemicals with inactive or equivocal results were set to low concern ("blue"), chemicals with non-selective DNT bioactivity were set to medium concern ("yellow"), and those with clear DNT bioactivity were flagged as of high concern ("red").

2.15.4. Zebrafish Activity

Developmental toxicity was assessed using a zebrafish embryo assay in which zebrafish embryos were exposed to each chemical until the larval stage⁴⁷; Britton et al., in prep. The larvae were then assessed for death, non-hatching and malformations (swim bladder non-inflation, craniofacial abnormalities, edema, spinal curvature, blood pooling, abnormal position in the water column, or abnormal pigmentation). The developmental toxicity of many of the chemicals were assessed twice: once in a single, high-concentration, preliminary, range-finding assay, and then again in a concentration-response assay. The flag levels were then assigned to each chemical according to the incidence (%) of adversely affected larvae for each chemical. Chemicals that were positive (i.e., elicited death, non-hatching, or malformations in at least 50% of the animals) were assigned a high concern flag ("red"). If the development was normal in all larvae, the low concern flag was assigned ("blue"). Chemicals that affected some embryos, but less than 50%, were assigned the medium concern flag ("yellow").

2.15.5. Thyroid Target Activity

The potential impacts of PFAS compounds on the thyroid axis were assessed using medium-throughput assays which rely on human recombinant enzymes. The assays test seven Molecular Initiating Events (MIEs) in the thyroid Adverse Outcome Pathways (AOPs) network. This suite of assays covers critical pathways within the thyroid axis including deiodinase enzymes (Human Deiodinase 1,2 and 3, Human Iodotyrosine deiodinase^{48,49}), human thyroid peroxidase⁵⁰, and thyroid hormone plasma binding proteins (transthyretin, and thyroxine binding globulin⁵¹). The MIEs link to 16 known or putative pathways in the AOP Wiki (<https://aopwiki.org/aops>, Society for the Advancement of AOPs, 2020).

The flags for the thyroid chemicals were set as follows, based on potency (AC50 values) and efficacy. Low concern ("blue") chemicals had no activity greater than 50% of the model inhibitors/binders. Medium concern chemicals ("yellow") had efficacy greater than 50% of the model inhibitors/binder, but the corresponding AC50 was at least 2 orders of magnitude higher than the model inhibitors/binders. High concern chemicals (red) had AC50s that were within 2 orders of magnitude of the model inhibitors/binders.

2.15.6. Immunosuppression-Relevant Bioactivity

The BioMAP panel, comprising 12 different assay systems, has been used previously, largely in preliminary toxicity profiling of pharmaceutical and consumer chemicals^{52,53,54}. These 12 assay systems include models of autoimmune disease, chronic (vascular) inflammation, allergy, monocyte activation, lung inflammation and fibrosis, cardiovascular inflammation, and dermatitis, wound healing^{55,20}.

Within the BioMAP panel, specific readouts for 3 different model assay systems were identified as immunosuppression relevant for a semi-quantitative flag of potential immunosuppression. Details of these models were published previously^{55,20}, including an analysis of the results produced by four immunosuppressive drugs²⁰. The three model systems included in the immunosuppression flag were:

1. the SAg system (T cell activation "super-antigen" model; intended to model an autoimmune or chronic inflammation states relevant to T-cell dependent conditions; uses co-cultured primary human peripheral blood mononuclear cells [PBMC] and human umbilical vein endothelial cells [HUVEC] stimulated with superantigens, i.e., T cell receptor [TCR] antigens);
2. the BT system (T cell dependent B cell activation; intended to model autoimmune, allergy, or asthma, or oncology disease states where B-cell activation and antibody production are relevant; uses co-cultured PBMC and CD19+-B cells stimulated with TCR antigens and anti-IgM); and,
3. the Mphg System (macrophage activation response; intended to model chronic inflammation and

macrophage activation relevant to conditions involving cardiovascular inflammation, restenosis, and arthritis; uses co-cultured HUVEC cells and macrophages stimulated using toll-like receptor 2 [TLR2] ligands derived from yeast).

Endpoints from these systems considered immunosuppression relevant were effects on PBMC viability in the SAg and BT systems; decreased B cell proliferation in BT system; decreased T cell proliferation in the SAg system; decreased soluble IgG production in the BT system; and decreased IL-10 production in the Mphg systems.

The flags used here reflect selective immunosuppressive activity in vitro at the endpoints specified as "immunosuppression-relevant," where selectivity is defined as immunosuppression-relevant bioactivity occurring at concentrations lower than those that elicit overt cytotoxicity. Cytotoxicity is determined using sulforhodamine B (SRB), a total protein marker, to monitor for significant loss of cellular protein (compared against historical values) in the BioMAP systems. The LEC among the immunosuppression-relevant endpoints was subtracted from the minimum LEC among the SRB cell viability endpoints. If there were no chemical effects on the SRB cell viability endpoints, the immunosuppression-relevant LEC was subtracted from 3 (equivalent to 1000 μM on a log₁₀- μM scale). Selectivity scores of less than 0.25 log₁₀- μM were coded as low concern ("blue") (i.e., no selective immunosuppression-relevant bioactivity observed. Chemicals with selectivity scores of greater than 0.25 were assigned the medium concern flag ("yellow"). No chemicals were assigned the high concern flag for this technology.

2.15.7. In Vitro Toxicokinetics (TK)

Dosimetric conversion of NAM data to an administered equivalent dose (AED) requires consideration of chemical TK in a process known as in vitro-in vivo extrapolation (IVIVE). As described previously in Wetmore et al.⁵⁶, a simplified HT-IVIVE TK approach utilizes in vitro experimental measures of hepatic clearance and plasma protein binding to estimate internal blood concentrations such as the steady-state concentration (C_{ss}). Experimental hepatic clearance assays were conducted as previously described⁵⁶ for all PFAS passing stock QC evaluation. Individual PFAS were incubated with pooled adult human mixed sex hepatocyte suspensions and monitored for substrate depletion (i.e., loss of parent PFAS) over a 240 minute time course^{21,57}. Assay reference compounds were included to ensure hepatocytes were performing as expected; and negative controls were included to monitor and correct for abiotic loss where needed. Plasma protein binding was measured using ultracentrifugation, deriving fraction unbound in plasma (f_{up}) after quantitating levels in the aqueous supernatant resulting after ultracentrifugation and the amount present in the time-matched whole plasma^{22,21}. In addition to being used in HT-IVIVE to derive administered equivalent dosages^{58,59}, these TK values

can be evaluated to inform PFAS half-life estimations, persistence and bioaccumulation in general.

TK flags that captured plasma protein binding, hepatic metabolic stability and PFAS metabolite formation were developed based on experimental findings and downstream analyses. These were binary in structure: denoting presence and absence of a flag, respectively. The three flags for 'PlasBind' were underpinned by in vitro fraction unbound in plasma (fup) data (Smeltz et al., Kreutz et al.^{22,21} – plasma protein binding that was higher than 50% of non-PFAS chemicals (and 25% of tested PFAS) ($fup < 0.11$) would be denoted 'High'; a 'Higher' flag would correspond to $fup < 0.0109$ which is higher than 50% of tested PFAS; and 'Highest' flag would correspond to $fup < 0.0039$, higher than 75% of tested PFAS. These 3 flags were collapsed into a single TK Plasma Binding flag for consistency with the other NAM flags as detailed in Table 1.

'Highest' flag was mapped to the 'high confidence' flag, whereas 'Higher' was mapped to 'moderate confidence' and 'High' to 'low confidence'. The TK_Metab designations, of which there were 3, corresponded to measures of hepatic in vitro clearance (Clint) with 'Stable' denoting a Clint = 0, 'Slow' denoting a Clint less than 5.97 uL/min/million hepatocytes (observed for 75% of PFAS); and 'Rapid' denoting a Clint greater than 5.97 uL/min/million hep^{21,57}). Again these were mapped to 'High', 'Moderate' and 'Low' as shown Table 1 for consistency with the other flags under a summary flag named 'TK_Metab'. Finally, TK_Struc_Endo corresponds to a structural flag to identify substances, due to their similarity to endogenous transporter ligands, are more likely to act as a transporter substrate⁶⁰. Substances presenting a structural flag were mapped to the "Moderate" as shown in Table 1.

Quality control (qc) flags that leverage targeted analytical chemistry evaluations that occurred during method development and TK data generation are also included to provide quality metrics for the PFAS stocks used during screening. In addition to the stock QC flag described earlier, a qc_httk flag is also included that notes significant abiotic loss (i.e., >50% loss of parent analyte within 60 minutes of assay initiation in negative controls) within the hepatic metabolic clearance assay. This observation may have implications for interpretation of the NAM bioactivity data.

3. Data analysis software and code

Data processing was conducted using the Anaconda distribution of Python 3.9 and associated libraries. Jupyter Notebooks, scripts and datasets will be made available on github at XXXX and on Figshare at XXXX.

4. Results and discussion

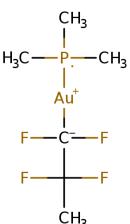
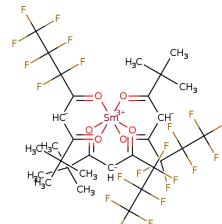
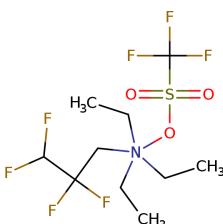
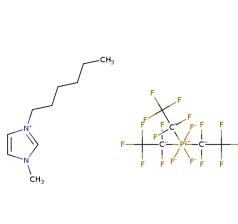
4.1. Primary and secondary structural categories

The PFAS landscape following application of the TSCA SNUR rule to DSSTox resulted in a dataset comprising 10,576 substances plus 3,126 degradation products for a total of 13,702.

Minimal structural overlap was found between the accessible training set substances from the Catalogic model and the TSCA inventory substances. Figure A1 depicts a t-SNE plot for the MITI training set substances relative to the TSCA substances using Morgan chemical fingerprints as inputs. In view of this, the degradation products simulated (comprising 22% of the landscape) should be interpreted with caution until additional experimental data are collected and new models developed.

Chain lengths could not be computed for four substances: DTXSID901222929, DTXSID101138156, DTXSID301146476, DTXSID301336809 as shown in Table 4. Based on inspection and given the presence of metal ions or mixtures, these four substances were dropped from further consideration. Thus, the final PFAS landscape used for the remainder of the analysis comprised 13,698 substances.

Table 4: Substances dropped due to lack of chain length derivation.

DTXSID	DTXSID901222929	DTXSID101138156	DTXSID301146476	DTXSID301336809
CASRN	64443-70-5	17631-69-5	885275-45-6	713512-19-7
Structure				

There were also 26 substances that were tagged as "Not PFAS" based on the OECD structure definitions used within PFAS-Map. Upon manual inspection of the substances, DTXSID9096236 [CASRN 422-69-5], DTXSID8063181 [CASRN 55364-35-7] and DTXSID5059872 [CASRN 354-83-6] were reassigned to the "Silicon PFAS" primary category. Substances DTXSID1011943 [CASRN 1780277-75-9], DTXSID3012966 [CASRN 28781-82-0], DTXSID50448610 [CASRN 79035-75-9], DTXSID8012466 [CASRN 590424-08-1] and DTXSID60667618 [CASRN 817562-60-0] were reassigned to the "Side-chain aromatics" primary category. Substance WXGNWUVNYMJENI-UHFFFAOYSA-N was reassigned to the "Other aliphatics" primary category. The remainder were

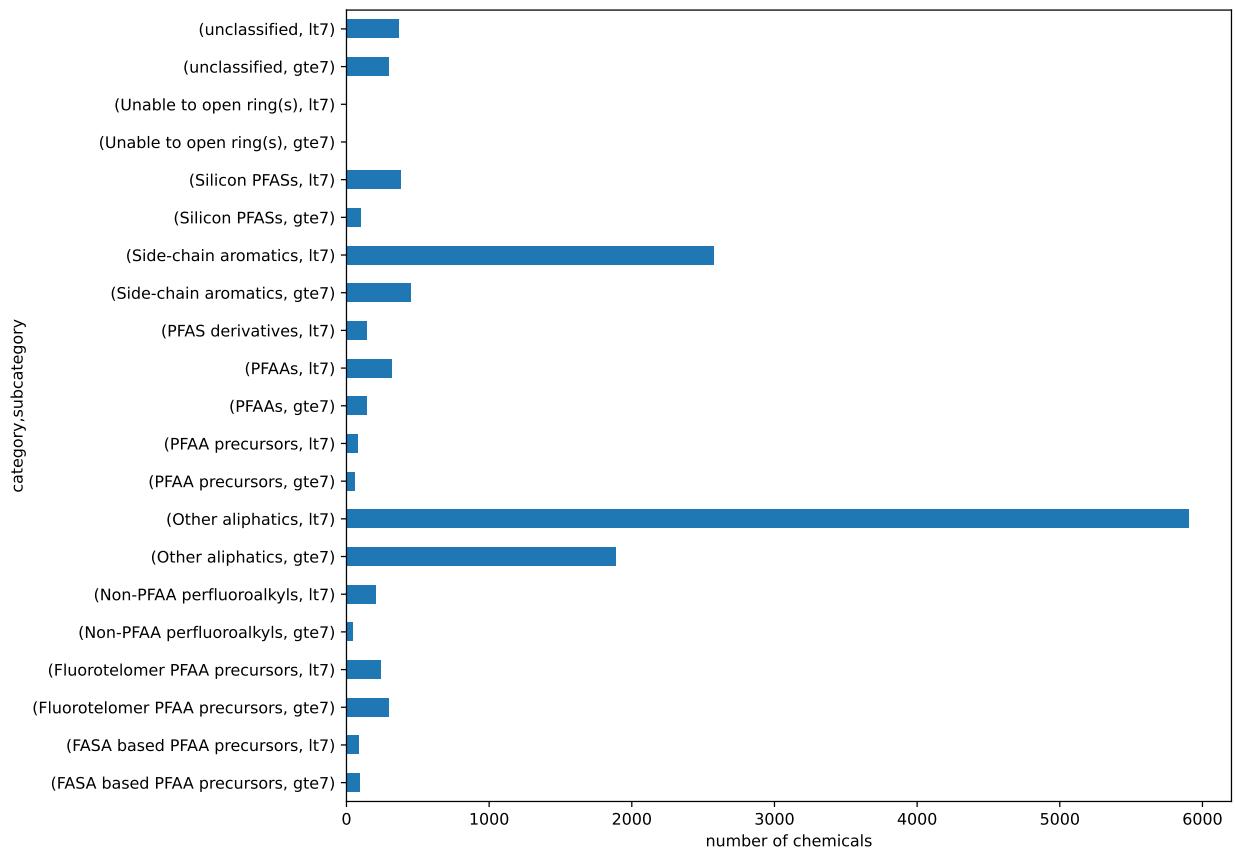
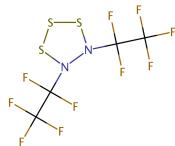
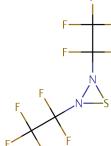
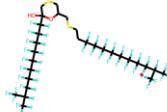
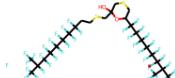
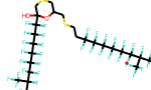


Figure 2: Bar chart showing the number of substances within each secondary category, ordered by primary category root. Methods to define primary and secondary categories are outlined in Sections 2.31 and 2.32. Lt7, chain length less than 7; gte7, chain length greater than or equal to 7.

reassigned to the "PFAS derivatives" primary category. Figure 2 is a bar chart showing the number of PFAS within each secondary category.

From Figure 2, there appear to be no substances assigned as "Unable to open ring(s)". In fact there were 5 substances that were tagged as "Unable to open ring(s)", two of which included heterocyclic moieties comprising a hydrazine linkage and a disulphide bond and the remaining three were thiane/oxane like (see Table 5).

Table 5: 'Unable to open ring(s)' substances.

DTXSID0053124	DTXSID2051982	FAQDJDLEAALLIF- UHFFFAOYSA-N
 FZRBYUVNWOOAJS- UHFFFAOYSA-N	 MBHWADCTCXPAHT- UHFFFAOYSA-N	
		

Across the more than 13,000 PFAS substances evaluated, forty-three percent of the substances fell into the "Other aliphatics, lt7" category. In addition, 667 substances fell into the "Unclassified, lt7" or "Unclassified, gte7" secondary categories. This represents a potential limitation of using broad definitions represented by the OECD primary categories themselves. A chemotype ToxPrint enrichment was explored following the approach outlined in Wang et al.⁶¹ but using the PFAS specific ToxPrints developed in Richard et al³⁴ (see Section Appendix A.1 for methodological details). This was an effort to identify whether there were specific structural features that might be helpful in splitting apart those primary categories with the largest memberships namely (i.e., "Other aliphatics" and "Unclassified"). The most enriched features for the "Unclassified" category included fluorotelomer chains and phosphorus or sulfonic acid functional groups whereas heteroatoms, nitriles, amines and epoxides featured as functional groups for the "Other aliphatics". However, these enriched features were not determined to be sufficiently distinctive to justify creation of additional primary categories.

Structural similarity was evaluated within and between secondary categories to determine which

secondary categories required further subcategorization (as discussed in Section 2.6 of the Methods). Figure 3 shows the two distributions of the median pairwise distance distributions in the between and within secondary category combinations. The objective distance threshold derived by taking the 5th percentile of the median pairwise distances from the between categories resulted in a value of 0.75.

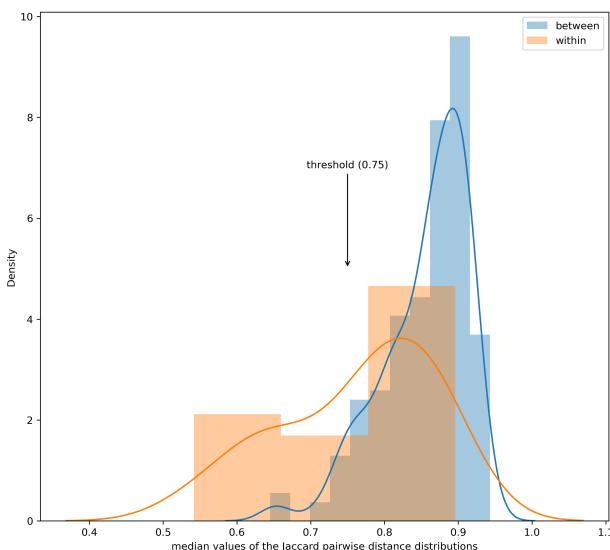


Figure 3: Probability density functions of the median Jaccard pairwise distance distributions for within (orange) and between (blue) secondary categories. Orange and blue graphed lines represent the fits to the probability density distributions.

Based on the threshold, 13 secondary categories (Table 6) were found to exceed the value that would render them subject to further subcategorization. The thirteen secondary categories included the "Other aliphatics", "Side-chain aromatics", "Silicon PFAS", "Unclassified", "PFAA precursors" and "Non-PFAA perfluoroalkyls". The "Other aliphatics" categories are of little surprise given their membership sizes were the largest out of all the secondary combinations; hence, these categories were expected to be the most diverse in terms of their structural makeup. Since the "Unable to open ring(s)" category only comprised 5 substances in total (as shown in Table 5), this category was excluded from any subcategorization.

Table 6: List of secondary categories exceeding the threshold and their corresponding median pairwise distances (rounded to 2 decimal places)

Primary-Secondary Categories	Median pairwise distance
Fluorotelomer PFAA precursors, lt7	0.79
Non-PFAA perfluoroalkyls, lt7	0.85
Other aliphatics, gte7	0.77

Primary-Secondary Categories	Median pairwise distance
Other aliphatics, <1t7	0.88
PFAA precursors, 1t7	0.83
PFAAs, 1t7	0.8
PFAS derivatives, 1t7	0.82
Side-chain aromatics, gte7	0.79
Side-chain aromatics, 1t7	0.88
Silicon PFAS, 1t7	0.83
Unable to open ring(s), 1t7	0.9
Unclassified, gte7	0.76
Unclassified, 1t7	0.87

Figure 4 shows the membership following the first generation of clusters being created for the 12 secondary categories that exceeded this objective threshold.

Following creation of the next generation categories, there were 18 tertiary categories that met or exceeded the threshold and were subcategorized further. The root primary categories were predominantly from the "Other aliphatics", "Side-chain aromatics", "Fluorotelomer PFAA precursor" and "Non-PFAA perfluoroalkyls" categories. Figure 5 reflects the quaternary categories for the 18 that were subset further.

Terminal categories were defined as either secondary or tertiary categories that did not exceed the threshold, as well as all quaternary categories. A total of 85 terminal categories were ultimately derived. This represented a trade-off in terms of the final number of terminal categories that was a practical number to characterize the landscape of PFAS balanced with maximizing structural similarity within the categories themselves. The full list of 13,768 substances together with their terminal category assignments are provided in the supplementary information. Structural similarity within categories did increase following subcategorization, Figure A4 shows the ECDFs of several terminal categories which are left shifted relative to the original ECDFs for the secondary categories (Figure A2), i.e. the pairwise distance range decreases.

4.2. Selection of representative substances

Whilst centroids were selected as the most representative substance from each terminal category, there was a recognition that a single chemical was unlikely to capture the breadth of diversity within a category. Additional substances to capture the breadth and structural diversity relied on the

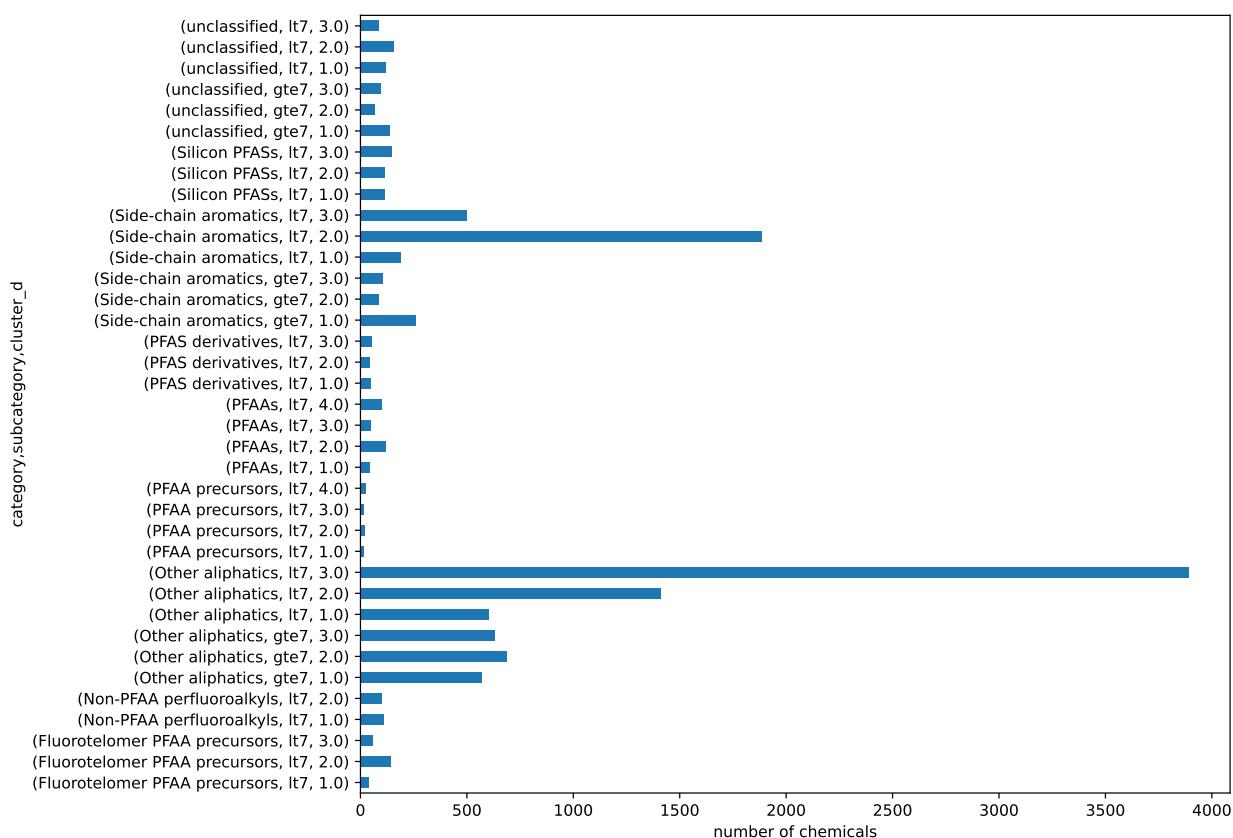


Figure 4: Bar chart showing the number of substances within each tertiary category, ordered by primary and secondary category roots. Methods to define tertiary categories are outlined in Section 2.7. Lt7, chain length less than 7; gte7, chain length greater than or equal to 7.

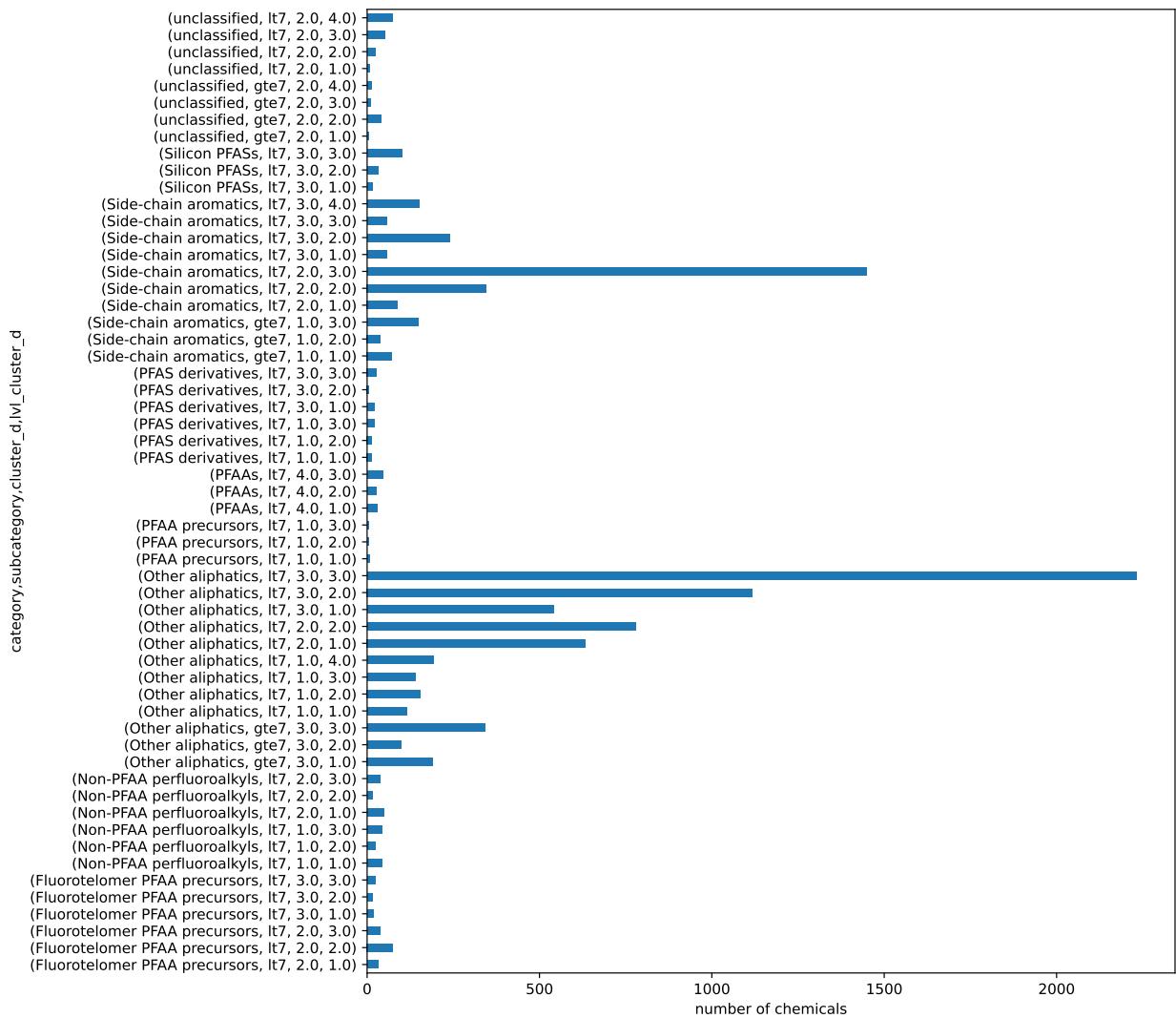


Figure 5: Bar chart showing the number of substances within each quaternary category, ordered by primary, secondary, and tertiary category roots. Methods to define quaternary categories are outlined in Section 2.7. Lt7, chain length less than 7; gte7, chain length greater than or equal to 7.

MaxMinPicker method. This method was used to select up to 3 further substances in addition to the centroid. A total of 320 substances were selected using this approach for 80 of the terminal categories. Terminal categories with 5 or fewer members did not result in any additional substances being selected (beyond the centroid) by the approach. Table 7 lists the 5 terminal categories which had insufficient membership to apply the MaxMinPicker approach.

Table 7: Terminal categories for which the MaxMin approach was not undertaken.

Terminal category	Membership
PFAA precursors, lt7, 1, 2	4
PFAA precursors, lt7, 1, 3	5
Unable to open ring(s), lt7	4
Unable to open ring(s), gte7	1
Unclassified, gte7, 2, 1	5

To evaluate the proportion of structural diversity captured by the selected representative substances, the normalized cumulative minimum distance was calculated as a function of the number of substances selected using the MaxMinPicker method as discussed in Section 2.9 . There were 11 terminal categories, out of the 80 terminal categories for which diverse substances were selected, where picking 3 substances captured 50% of the structural diversity (shown in Table 8).

Table 8: Terminal categories for which 3 representative substance selections capture more than 50% of the structural diversity.

Terminal category	Number of chemicals for 80% structural diversity	Cumulative % of Structural Diversity	Terminal Category size
Unclassified, gte7, 2.0, 3.0	1	100	9
Unclassified, lt7, 2.0, 1.0	3	92.63	7
PFAA precursors, lt7, 1.0, 1.0	3	85.29	7
PFAS derivatives, lt7, 3.0, 2.0	3	84.32	6

Terminal category	Number of chemicals for 80% structural diversity	Cumulative % of Structural Diversity	Terminal Category size
Unclassified, gte7, 2.0, 4.0	3	80.87	12
Non-PFAA perfluoroalkyls, lt7, 2.0, 2.0	4	70.63	15
PFAA precursors, gte7, nan, nan	5	68.59	58
Fluorotelomer PFAA precursors, lt7, 3.0, 2.0	5	64.18	15
Fluorotelomer PFAA precursors, lt7, 3.0, 1.0	6	58.96	18
PFAA precursors, lt7, 3.0, nan	6	54.08	18
Non-PFAA perfluoroalkyls, gte7, nan, nan	5	52.83	48

Notes: Column 1 represents the number of substances that would be required to capture 80% of the structural diversity in the category, Cumulative % of Structural Diversity represents the normalized cumulative minimum distance for up to 3 selected diverse substances. Note the 80% used as a threshold is purely for demonstration purposes only.

For the largest terminal category, "Other aliphatics, lt7, 3,3", the existing up to 3 diverse substances only captures 0.46% of the structural diversity. In order to capture 80% of the structural diversity for this terminal category, 935 substances would need to be selected for additional testing and assessment. The number of substances to select from each terminal category to capture 80% of the structural diversity varied from 1 (as shown above in Table 8) to 935 with the median number being 20.

Figure 6 shows the curves of the number of diverse selections as a function of the percentage

normalized cumulative minimum distances for 10 terminal categories. These vary in steepness showing how quickly or not the structural diversity coverage converges with number of diverse selections depending on the terminal category of interest.

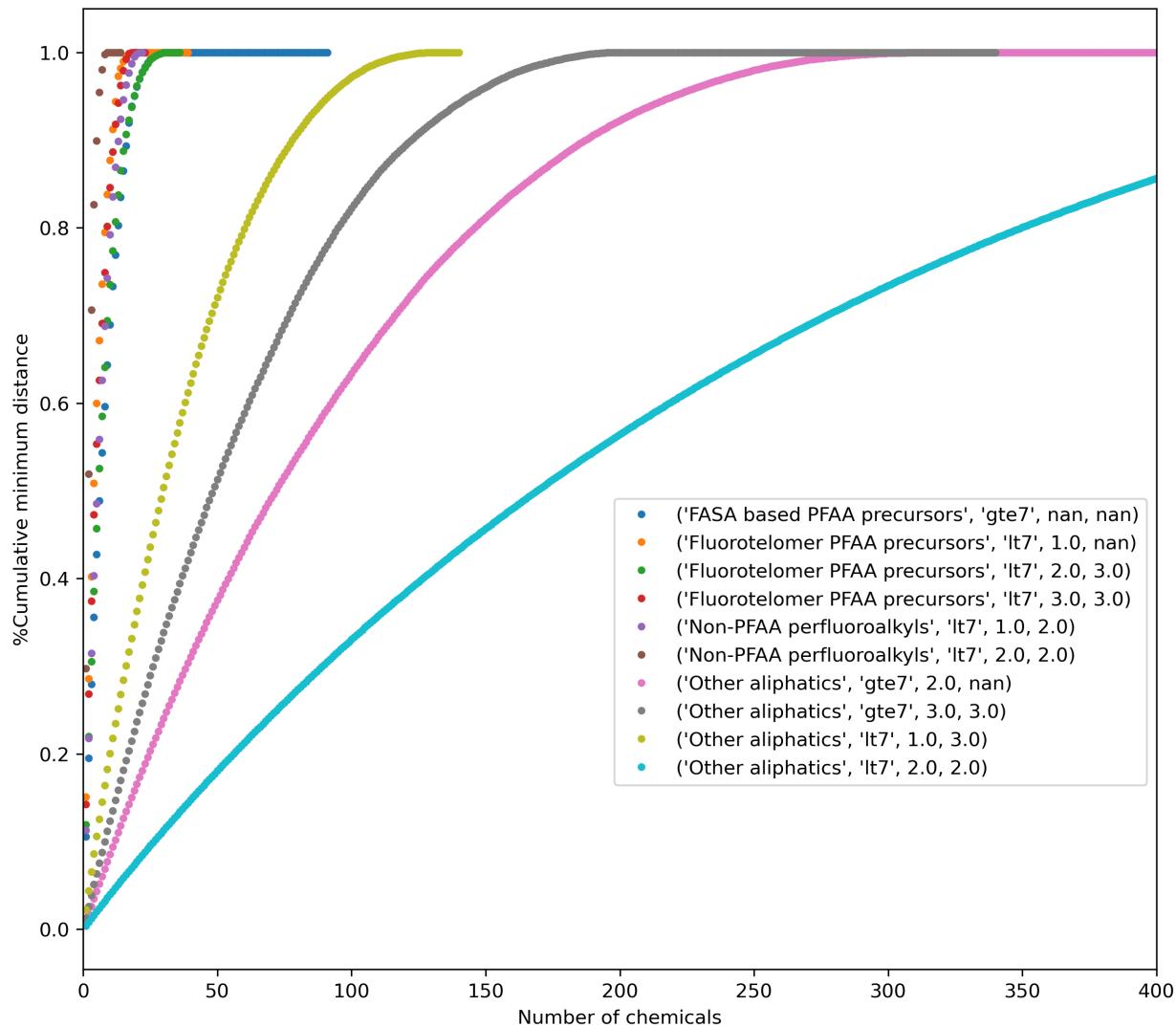


Figure 6: For a selection of terminal categories, the extent to which 100% structural diversity is captured relative to number of diverse chemicals selected varies.

Figure 7 attempts to summarize the tradeoff of the number of diverse chemicals (centroids and MaxMin) as a function of % structural diversity captured on a per terminal category basis.

The diverse selections identified earlier for the terminal categories reflects a pragmatism in terms of identifying a potential candidate list of substances. As discussed in later Section 4.6, the structural diversity captured forms one of the considerations in test candidate selection relative to those terminal categories that are data poor or contain substances that are on the TSCA inventory.

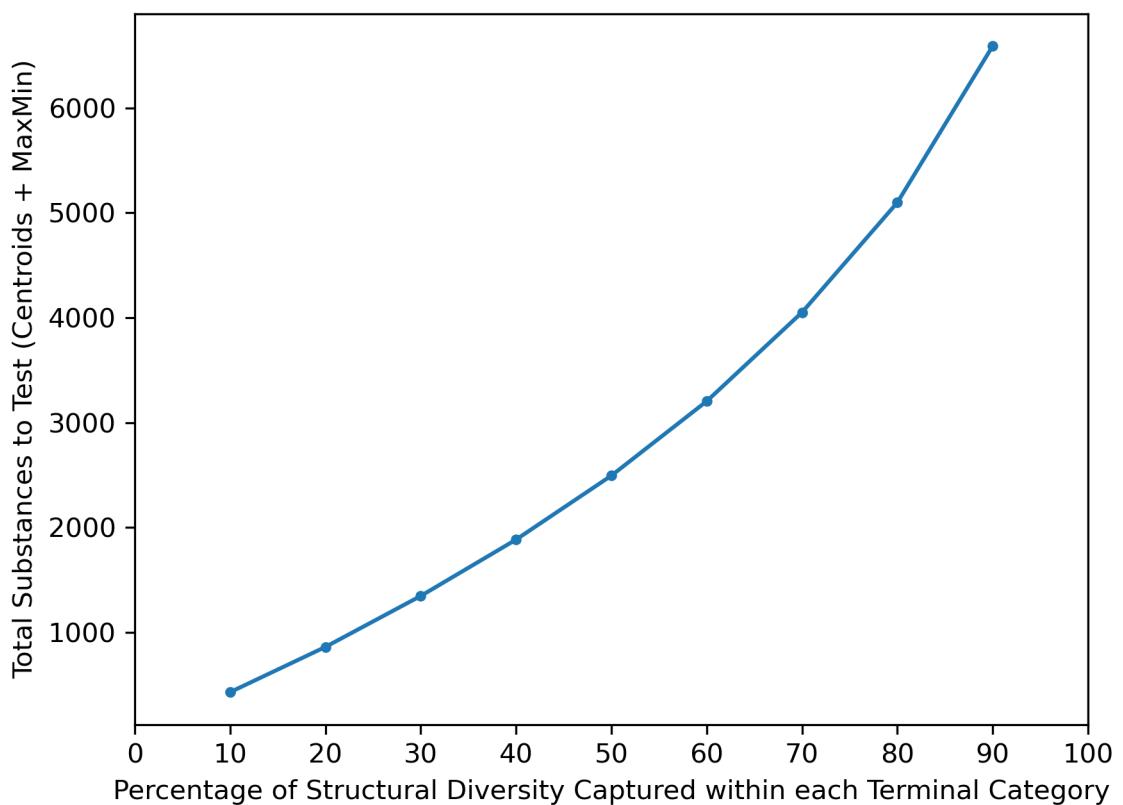


Figure 7: Lineplot showing the number of diverse substances that would need to be selected to achieve a specific % structural diversity coverage across each terminal category.

4.3. Evaluation of physchem consistency within terminal categories

For the 13,449 substances in the PFAS landscape for which OPERA predictions could be generated, 66 substances (0.5%) could not be assigned into any specific physchem testing track, 79 substances (0.6%) were assigned into testing track D, the remainder were relatively evenly distributed amongst testing tracks A-C (range: 3649-5277, 27-39%). The proportions were similar if only substances on the TSCA active inventory were considered namely testing track A (24%), testing track B (28%), testing track C (41%), testing track D (6%) and no testing track (2%). Across the terminal categories, there was a general trend of number of testing tracks increasing with size in category membership (see Figure A5). Figure 8 shows an example of one of the most diverse and largest terminal categories "Other aliphatics, lt7, 3, 3" which comprises 2232 members and spans all 4 testing tracks. Although substances predominantly lie within the testing track B, there is no discernible separation between the tracks across the structural category as characterized by Morgan fingerprints. In contrast all 92 substances belonging to terminal category "FASA based PFAA precursors, gte7, nan, nan" fell into testing track A (figure not shown) whereas the 44 substances in "PFAAs, lt7, 1, nan" fell into testing tracks A and B, with the majority of substances in testing track B. Those substances in testing track A all have estimated water solubility values over a magnitude lower (data not shown). There was a positive association between how structurally diverse a terminal category was and the consistency in physchem profile observed (as reflected by the physchem testing tracks). However the Morgan fingerprints could not resolve the differences. PFAS ToxPrint fingerprints were also explored to evaluate the extent to which they had greater resolution in differentiate features that would account for the differences in physchem profile (results not presented). For the selection of potential candidates for tiered testing, the physchem profile remains an important consideration in concert with the structural diversity described in Section 4.2.

4.4. Variation of POD values across and within terminal categories

Ultimately, the terminal categories are intended to facilitate a read-across. To explore the feasibility of this further, the 10th percentile values of minimum oral and inhalation study level PODs were derived using available data for 71 substances and 15 substances respectively. The distributions were plotted in a series of box plots. In vivo toxicity data were available for at least one chemical in 28 of the 85 terminal categories across the 2 routes of exposure. The available data allowed preliminary trends for terminal categories to be observed where the primary root was FASA based PFAA precursors, PFAAs, Fluorotelomer PFAA precursors, and some of the Other aliphatics categories. In Figure 9, boxplots for each of the routes of exposure are shown side by side though there were only 8 terminal categories ("Fluorotelomer PFAA precursors, lt7, 2, 2", "Fluorotelomer PFAA precursors, lt7, 3, 3", "Non-PFAA perfluoroalkyls, lt7, 2,3", "Other aliphatics, lt7, 1, 3", "Other aliphatics, lt7, 2,

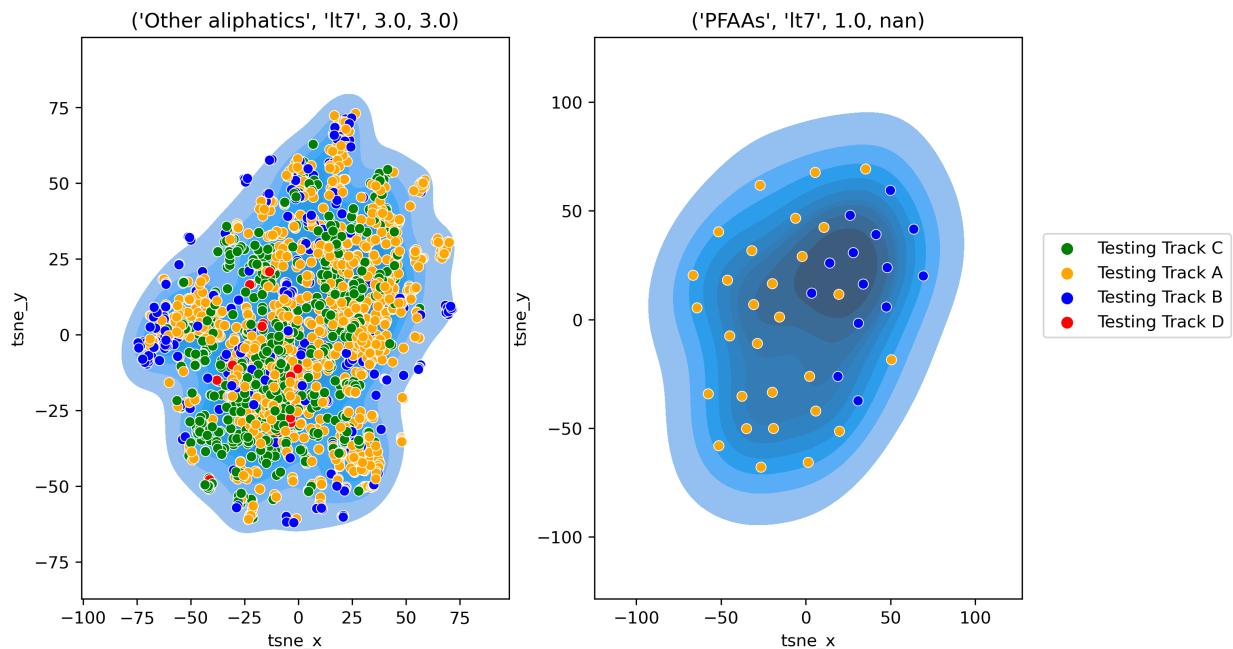


Figure 8: t-SNE projections for terminal category a) "Other aliphatics, lt7, 3, 3" and b) t-SNE projection for terminal category "PFAAs", "lt7", 1.0, nan" using Morgan chemical fingerprints with testing tracks A-D overlaid.

2", "Other aliphatics, lt7, 3, 3", "PFAAs, gte7, nan, nan", "PFAS derivatives, lt7, 2, nan") for which inhalation data was available.

Figure 10 shows boxplot and strip plots for the oral studies only. It appears that substances at each end of the spectrum of chain length within a category tended to exhibit lower toxicity, i.e., their aggregate POD is higher. The spread of POD values within a category with greater diversity in chain length tend to span > 2 orders of magnitude. Moreover, of the 71 substances with studies, 29 were associated with NOEL/NOAEL values only or a combination of NOEL/NOAEL/LOEL/LOAEL values.

Although the available toxicity data are limited, there does appear to be some separation in the potency distributions between terminal categories based on a common primary root. Inspection of Figure 10 does show a shift in potency values between the 2 FASA based PFAA precursors categories with those in the gte7 category being more potent. Similarly, there is a general shift between the PFAA categories with a left shift for those substances in the gte7 category vs the majority of the PFAA lt7 categories. However, the relatively large spread for some of the terminal categories suggests that additional refinement beyond structural similarity and chain length will likely be needed for some terminal categories prior to broader application in a read-across context.

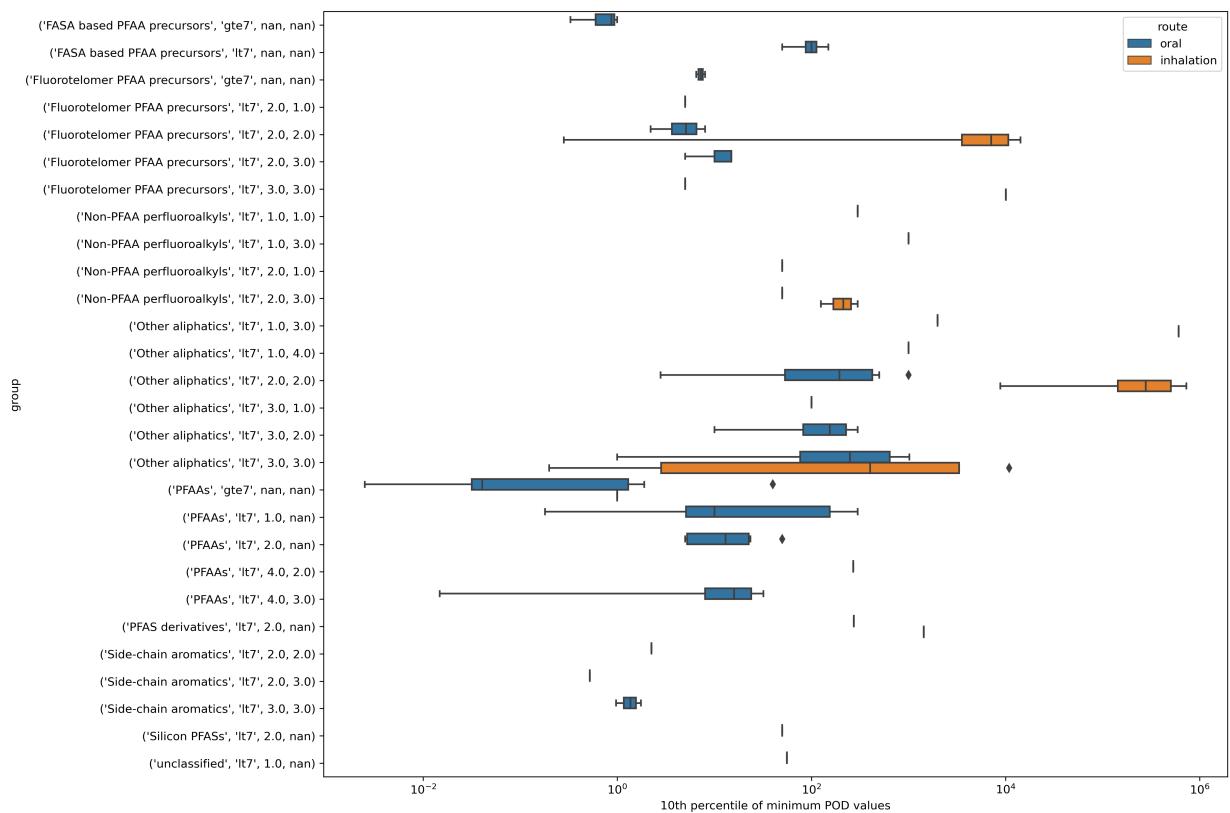


Figure 9: Boxplots of the variation of 10th percentiles of minimum point of departure values from oral and inhalation studies (units mg/kg/day). The box in the boxplot reflects the quartiles of the dataset, whilst the whiskers extend to $+ 1.5 \times \text{IQR}$. Outliers are shown as points if they exceed $1.5 \times \text{IQR}$. The inhalation boxplot is shown below the oral boxplot for a given terminal category.

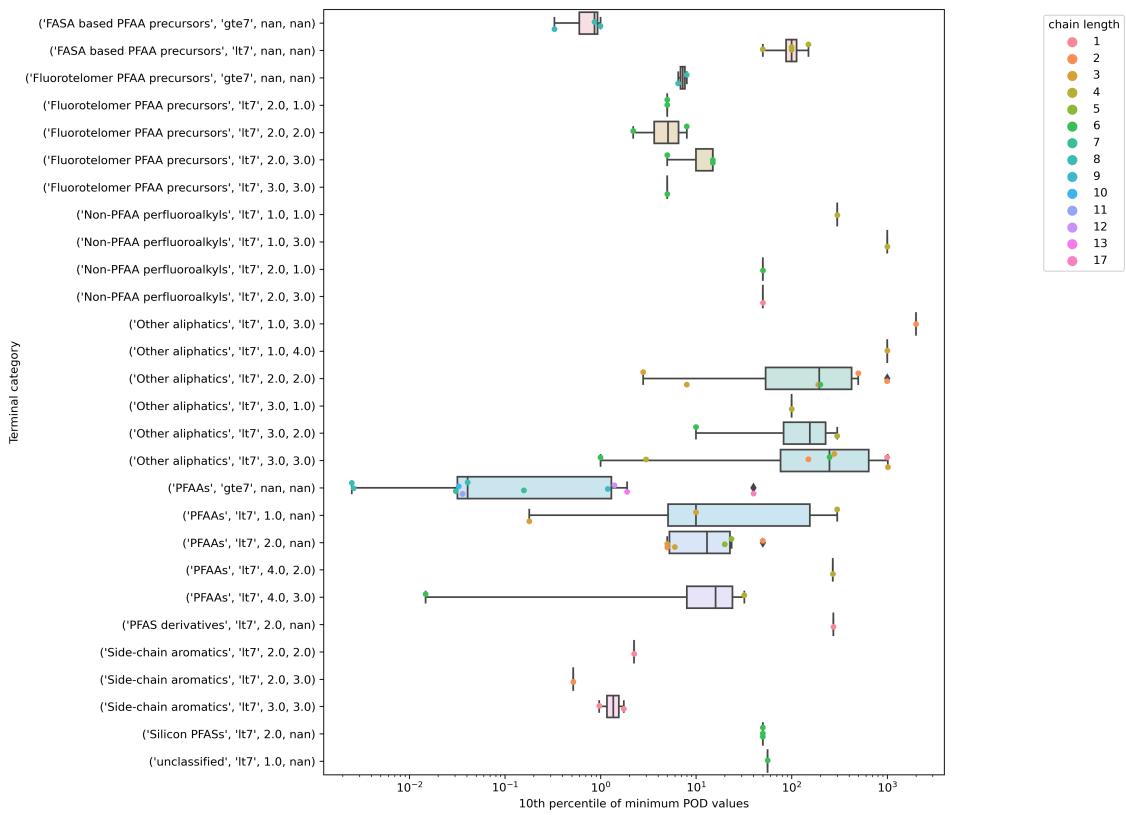


Figure 10: Boxplots showing the spread of the 10th percentile of the minimum oral POD values across and within terminal categories bounded by the carbon chain number. The box in the boxplot reflects the quartiles of the dataset, whilst the whiskers extend to + 1.5 * inter-quartile range (IQR). Outliers are shown as points if they exceed 1.5 * IQR.

4.5. Qualitative NAM flags

There were six data streams with qualitative flags assigned for the ~150 PFAS tested as part of the research project described in Patlewicz et al¹⁷ namely: 1) nuclear receptor assays (NR); 2) developmental toxicity (zebrafish testing); 3) DNT (developmental neurotoxicity); 4) thyroid toxicity; 5) immunosuppression (BioMAP assays); and 6) toxicokinetics (TK). No data were represented as null values (white colored), data available but no flag identified as denoted a 0 (blue colored), 1 denoted a medium confidence flag (yellow colored) and 2 was associated with a high confidence flag (colored in red) consistent with the descriptions described in Table 1. Figure 11 profiles all the NAM flags across the different technologies together with a stock QC flag [(author?)²³] (Pass (red)) and a qc_httk flag (Pass (red)).

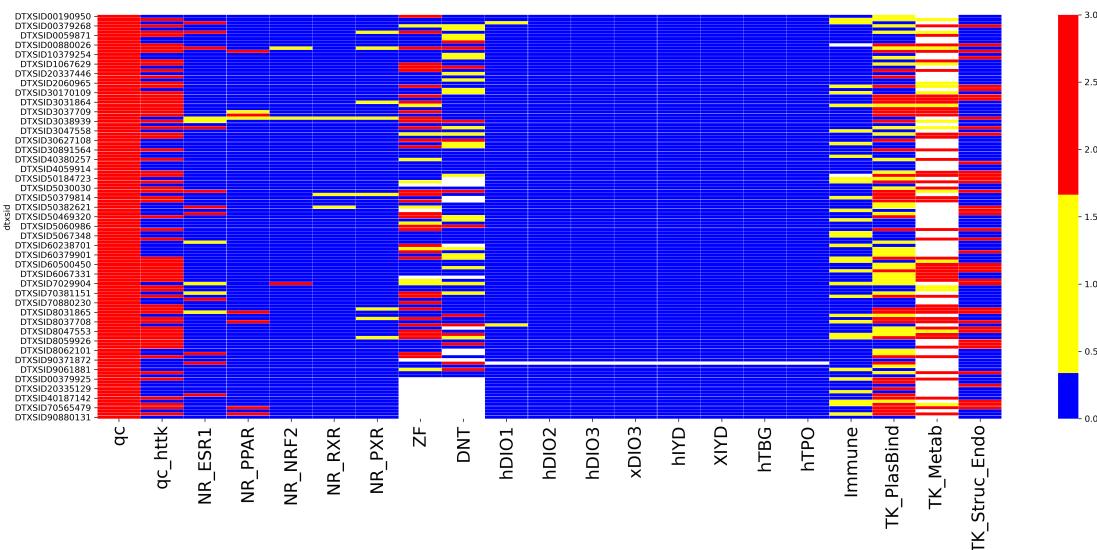


Figure 11: Heatmap of NAMs flags for the ~150 PFAS substances (~120 of which passed analytical QC) tested as part of the research programme described in Patlewicz et al¹⁷

From Figure 11, the first two columns represent the quality control (QC) information. The next 5 columns represent the NR data. The next 2 columns represent the developmental toxicity (ZF) assay and the DNT assay. The next 8 columns represent the thyroid assay outcomes followed by the integrated immunotoxicity flag from the BioMap assays. The last 3 columns represent the TK flags.

Of the 127 substances with associated NAM and TK flags, 124 matched with a substance in the PFAS landscape, permitting a closer examination of how consistent and concordant the NAM flags were within and across 26 of the 85 PFAS terminal categories (Figure A6). For 9 of the terminal categories that had at least 5 or more substances tested across the NAM assays, an enrichment analysis (using a Fischer exact test) was performed to identify whether any category was enriched for any flags. Terminal category "PFAAs, gte7" was enriched for the nuclear receptor PXR assay

whereas "Other aliphatics, gte7, 2, nan" and "Fluorotelomer PFAs, gte7" were enriched for immunosuppression. As an illustrative example, Figure 12 shows 2 contrasting terminal categories and their corresponding NAM profiles and the degree to which these were consistent across the category members.

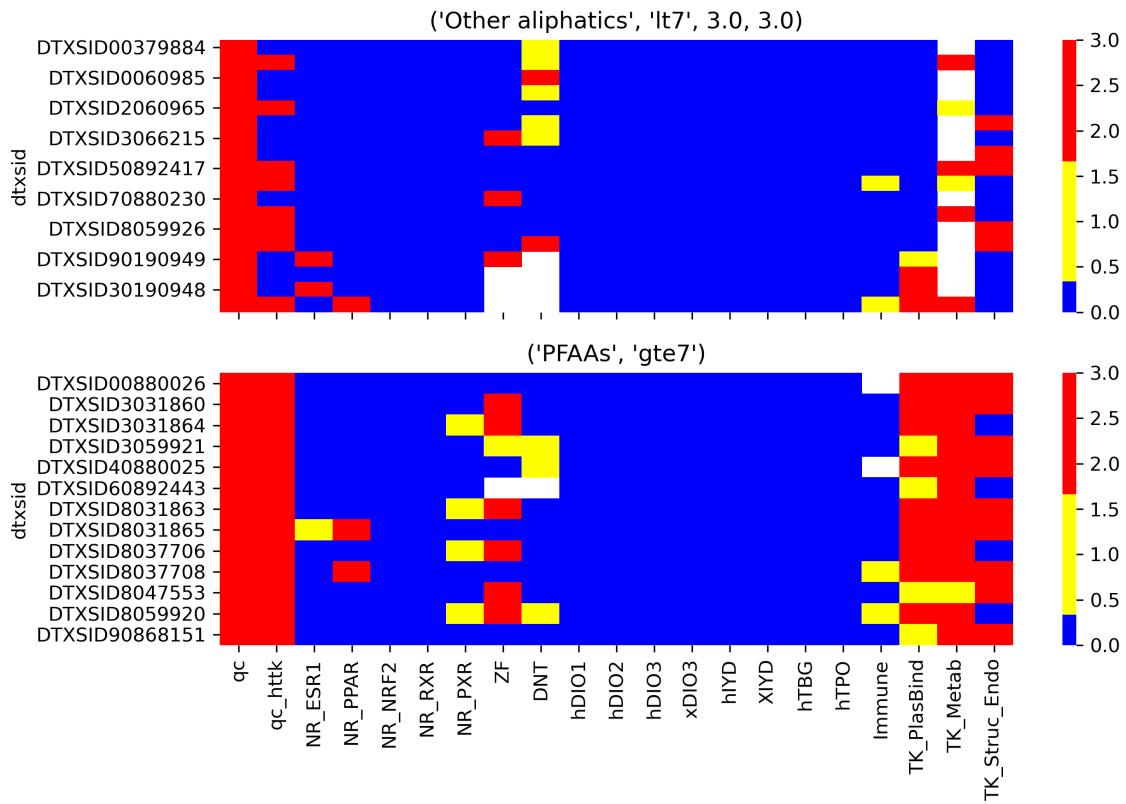


Figure 12: Heatmap for 2 of the terminal structural categories to illustrate the extent of their concordance across NAM profiles

These heatmaps might also play a role in probing to what extent the NAM profiles are able to rationalize the differences in POD values observed for specific terminal categories. As an example, the terminal category "PFAs, gte7" was associated with a large variation in 10th percentile POD values as shown in Figure 10. Figure 13 shows both the NAM profiles, chain length and POD information side by side for this category for convenience. Category members with chain lengths of 7 tended to give rise to responses for nuclear receptors ESR and PPAR, whereas PXR was activated for members with a chain length of 8. Immune responses were only observed for chain lengths of 7. None of the members gave rise to any activity in the suite of thyroid assays. TK flags showed that plasma binding was highest for substances with a chain length of 7 but intrinsic clearance was largely stable across the category. The most potent substances by in vivo toxicity were substances with a chain length of

8.

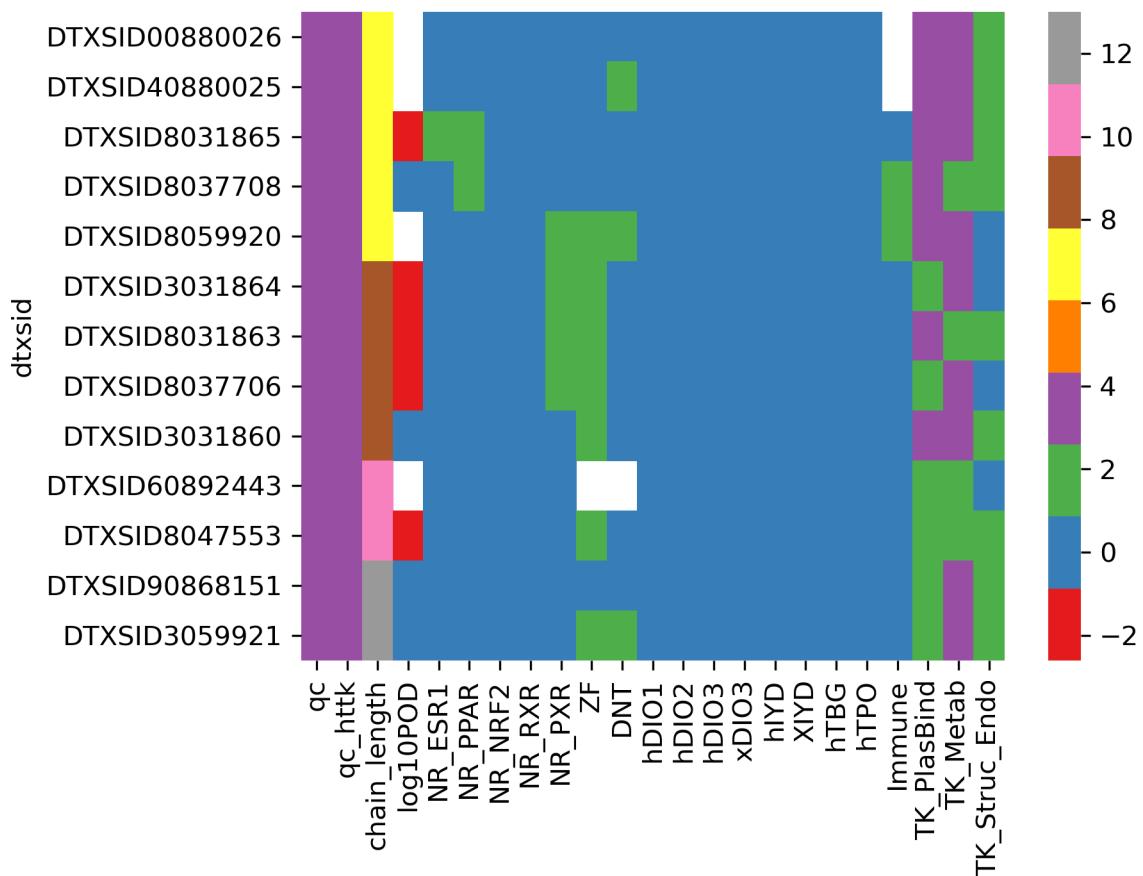


Figure 13: Heatmap of showing the log of the 10th percentile of the minimum oral POD values for terminal category "PFAAs, gte7" and its associated NAM data.

Although the NAM data is limited, the insights gleaned are useful to probe to what extent a terminal category might need to be subdivided further or whether for a candidate substance for testing, the closest neighbor by NAM profile might provide a targeted roadmap to identify the most strategic tiered testing to undertake.

4.6. Potential regulatory application to support the National Testing Strategy (NTS)

There are several considerations that come into play when identifying potential candidates for tiered testing in concert with the landscape defined. To make the NTS actionable, one consideration was to limit the landscape to one that was constrained by the TSCA active inventory to increase the feasibility of being able to identify a manufacturer/importer of the substance. A second enables the tradeoff between the number of diverse substances to select vs capturing the structural diversity to be more practically addressed. Herein the scope of terminal categories represented by the TSCA and

TSCA active inventory and the impact this has in terms of capturing structural diversity was evaluated. Finally a proposal was outlined that considers how the terminal categories could be triaged to initially focus on terminal categories which were either data poor or contained members that represented large exposure sources.

4.6.1. Constraining the landscape to the TSCA active inventory

4.6.1.1. TSCA inventory.

Of the substances in the PFAS Landscape, only 617 substances were identified to be on the TSCA inventory, of which 293 were 'active' and the remaining 324 'inactive'. Active and inactive refers to the EPA's designation of whether a substance is active in US commerce based on the rule requiring industry to report chemicals manufactured or imported or processed in the US over a 10 year period ending 21st June 2016. Figure 14 shows a bar chart of the membership of the terminal categories and how that differs when considering TSCA inventory status (overall or by active TSCA only).

The largest membership when constrained by presence on the TSCA inventory still reflects the "Other aliphatics, lt7" and "Other aliphatics, gte7" categories. Across the terminal categories, 75% of the categories (64 out of the 85 categories) contain members on the TSCA inventory. If only categories containing substances that are on the TSCA active inventory are considered, then the number of terminal categories decreases to 53, i.e., 62% coverage. The majority of the categories where there are no examples on the TSCA inventory are relatively small in size with fewer than 40 members. There were a couple of "Side-chain aromatics" categories which were larger with members between 57-346 as well as the two of the "Unclassified" categories with 52-97 members.

4.6.1.2. Selection of representative substances in the constrained TSCA active inventory.

Centroids were computed for the 53 terminal categories containing substances that were on the active TSCA inventory. For 14 of these terminal categories, membership exceeded 5, which permitted the MaxMinPicker approach to be applied to identify further analogues. An additional 56 analogues were selected from this constrained landscape. Figure 15 shows the overlap in substances (centroids and diverse) across the unconstrained and the TSCA active constrained landscapes. The minimal overlap between the sets highlights the limitations of using a constrained inventory, i.e., one which does not represent the breadth of the PFAS chemistry. However, the substances on the TSCA active inventory represent those substances that are currently in commerce in the US and potentially represent the largest exposure source.

An evaluation of the structural diversity captured using the centroids and additional MaxMin substances relative to the number of substances that would need to be selected to attain 80% structural

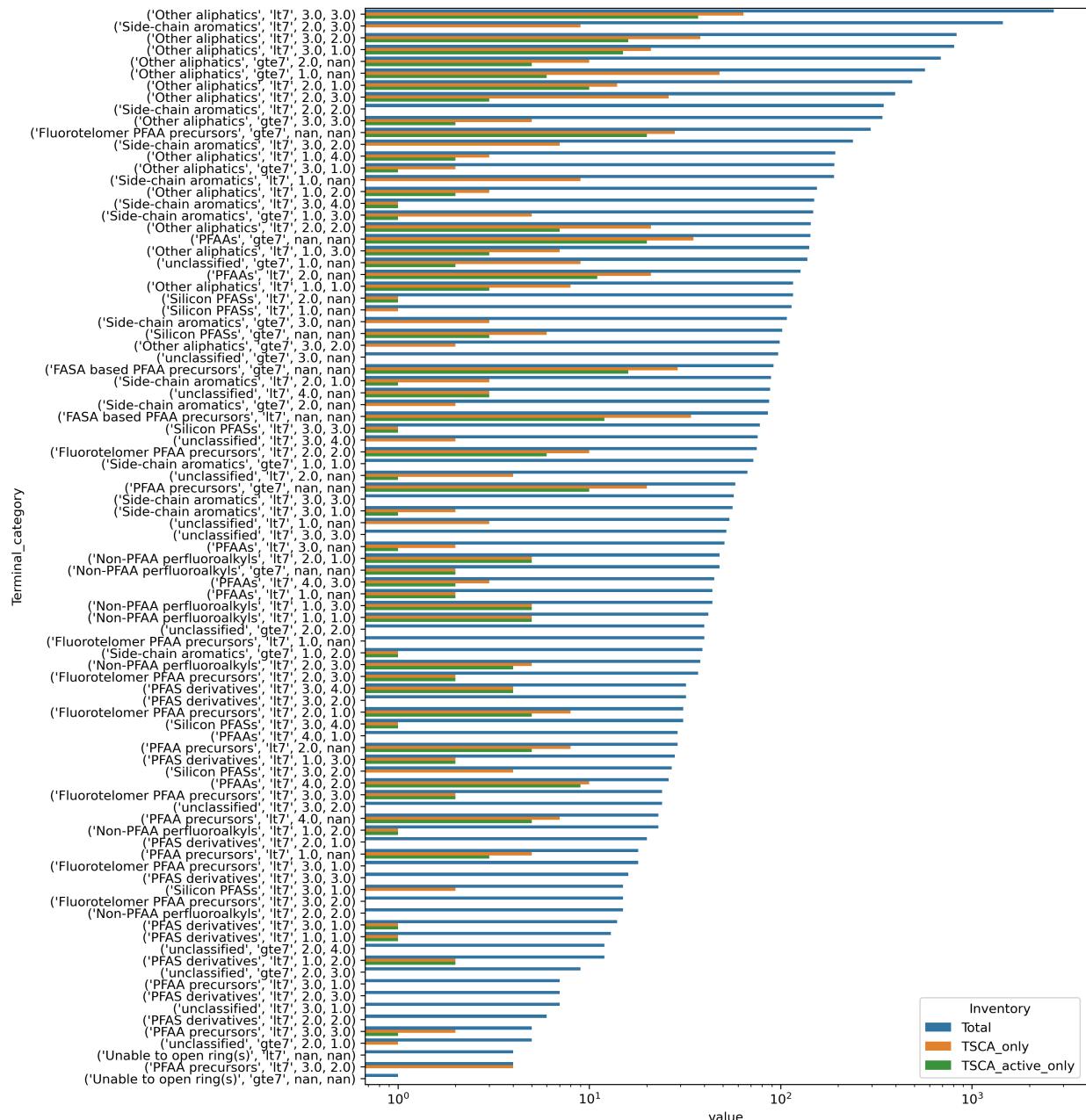


Figure 14: Bar chart showing membership of terminal categories and how that differs when constrained by TSCA inventory or TSCA active inventory.

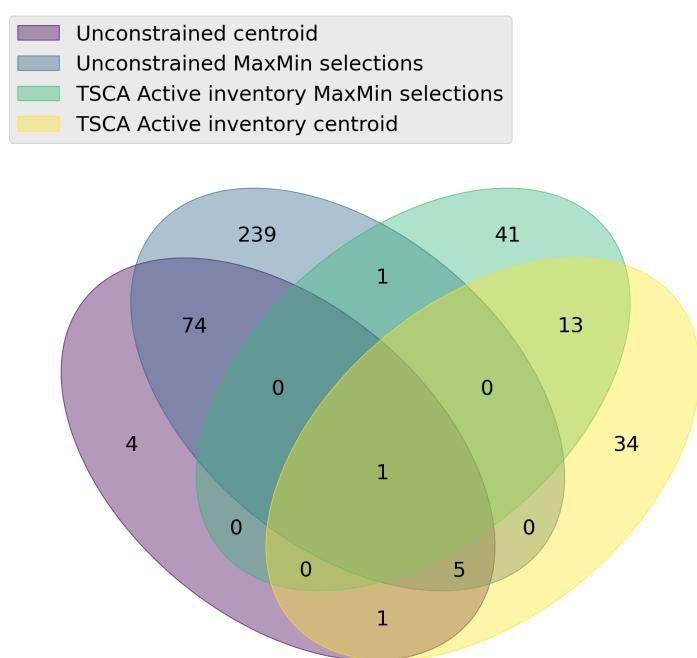


Figure 15: Venn diagram showing the overlap in substances based on whether they were identified as additional diverse picks or centroids in the PFAS landscape and that constrained by the TSCA active inventory.

diversity coverage was also undertaken in the same manner as had been performed for the full landscape. For the 11 of the 14 categories where the MaxMin approach had been applied, the diverse picks originally selected captured more than 50% of the structural diversity as shown in Table 9. This is not so surprising given the TSCA active set substantially limited the terminal category size and in turn their diversity.

Table 9: Constrained terminal categories where the MaxMin approach had been applied. Terminal categories for which 3 representative substance selections capture more than 50% of the structural diversity

Terminal category	Number of chemicals for 80% structural diversity	Cumulative % of Structural Diversity	Terminal Category size
FASA based PFAs precursors, gte7, nan, nan	6	57.63	16
FASA based PFAs precursors, lt7, nan, nan	4	69.04	12
Fluorotelomer PFAs precursors, gte7, nan, nan	3	87.81	20
Fluorotelomer PFAs precursors, lt7, 2.0, 2.0	3	86.86	6
Other aliphatics, gte7, 1.0, nan	2	100	6
Other aliphatics, lt7, 2.0, 1.0	6	55.54	13
Other aliphatics, lt7, 3.0, 1.0	4	77.95	10
PFAs precursors, gte7, nan, nan	2	100	10
PFAs, gte7, nan, nan	3	86.12	20
PFAs, lt7, 2.0, nan	5	66.64	11
PFAs, lt7, 4.0, 2.0	4	75.7	9

Notes: Number of chemicals for 80% structural diversity represents the number of diverse selections to capture 80% of the structural diversity, Cumulative % of Structural Diversity reflects the structural diversity captured by the diverse selections already made and Terminal Category size reflects the size of the terminal category if constrained by the availability of TSCA active substances.

For the largest terminal category, "Other aliphatics, It7, 3.0, 3.0", 3 diverse substance selections only captured 19.85% of the structural diversity. In order to capture 80% of the structural diversity for this terminal category, 18 substances would need to be selected for additional tiered testing. The number of substances to select from each terminal category to capture 80% of the structural diversity varied from 2 to 18 with the median number being 4. Across the entire TSCA active space, considering the 14 categories where a MaxMin approach could be applied - 74 substances would need to be selected capture an 80% structural diversity. That it is to say, in order to capture up to a 80% coverage across all the TSCA active categories, at least 115 substances (centroids + MaxMin) would be ideally selected for testing. Figure 16 summarizes the structural diversity attained across all TSCA Active terminal categories.

4.6.2. Proof of Concept Workflow: Identifying potential testing candidates

The availability of toxicity data across different study types and the presence of substances within different monitoring lists was arrayed across the terminal categories. All oral and inhalation studies from ToxValDB 9.4 were first retrieved as described in Section 2.14.1. There were 80 substances with data for one or more of the study types which were then matched on the basis of DTXSID with substances in the PFAS landscape. The resulting table was then transformed to produce a table where columns represented different study types, rows were substances and cells were labelled 1 if data for a specific study type existed for a specific substance and 0 if no data existed. The qualitative lists described in Section 2.1 were compiled together and transformed into a table where rows represented substances, columns represented the different list sources and cells were populated with a 1 or 0 to denote presence or absence on a specific list. There were 348 unique substances found across the different environmental monitoring and discharge lists which were then matched with substances in the PFAS landscape. CDR status tags and Pubmed count tags were then added to the PFAS landscape. Columns representing the toxicity study types and various lists were grouped by terminal category to produce a new table which reflected presence or absence of information (denoted by 1 or 0) on a per terminal category. Study quality was not considered - only the availability of publicly available toxicity data. The set of terminal categories were filtered to retain only those terminal categories which contained members on the TSCA active inventory (53 terminal categories). Figure 17 provides a perspective of this information, namely the toxicity data sparsity across the categories that fall within the scope of the TSCA active inventory as well as different environmental monitoring efforts

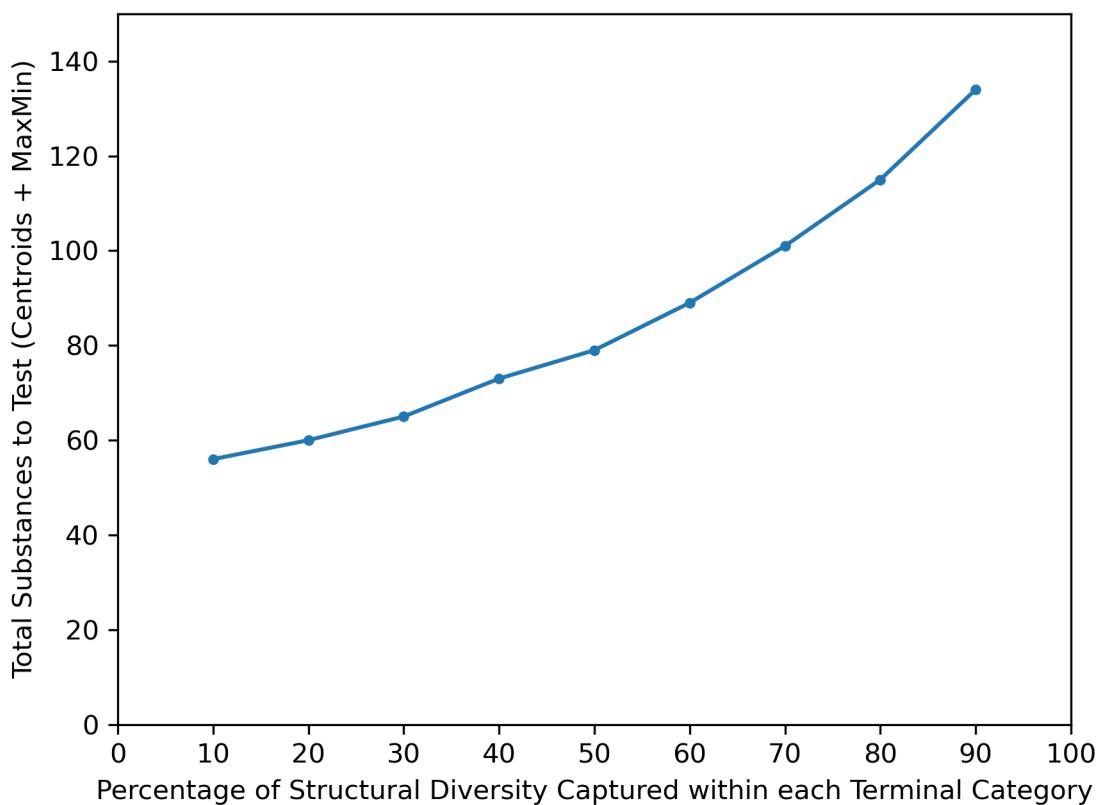


Figure 16: Lineplot of the TSCA active constrained terminal categories as a function of number of diverse substances selected and the % structural diversity captured.

or discussed in the literature. The PFAAs categories and their subcategorizations show up with data entries which is largely unsurprisingly, given the extent to which PFOA and PFOS have been studied.

Notes: #Pubmed is a tag to denote presence or absence of articles indexed in Pubmed. PROD-Data = Production data, DISCHARGE = Discharge Monitoring data, DRINKING_WATER = Drinking Water (State) Data, DRINKING_WATER-UCMR = Drinking Water data comprising Unregulated Contaminant Monitoring Rule data and State level monitoring data, TRI_Waste = Toxics Release Inventory (TRI) Data Waste Managed, TRI_On-Site = On Site TRI Data, TRI_Off-Site = Off Site TRI Data, Analytical_Mthds = PFAS with Validated Analytical Methods 533 and 537

Each of the earlier sections in of themselves highlight different lines of evidence that can inform the identification of potential test candidates. Here, an attempt was made to demonstrate how these steps can be integrated together to triage terminal categories and their potential candidates for subsequent tiered testing efforts (Figure 18). Step 1 is to consider a given terminal category and determine whether it meets the condition of being a 'data poor category'. Data-poor in this context was to consider whether this was a category that did not contain any members for which repeated dose toxicity data existed (by the oral or inhalation route and with a reported NOAEL, LOAEL, LOEL, NOEL, NEL or LEL value). There were 56 terminal categories out of the 85 total number of categories that met this condition.

The next step was to focus on terminal categories that overlapped with those which contained substances that were on the TSCA inventory.

There were 64 terminal categories that contained substances that were on the TSCA inventory of which 53 terminal categories contained substances that were on the TSCA active inventory. Of the TSCA categories, 37 also satisfied the condition of being a 'data poor' category, in contrast 27 of the TSCA active categories were 'data poor'. The following step was to consider terminal categories that contained substances that were on different environmental monitoring lists. There were 42 terminal categories that contained substances that were on one or more monitoring lists or 68 if Pubmed article availability was taken into account. Of these 42 terminal categories, 18 were also overlapping with data poor TSCA categories or 16 of the data poor TSCA active categories. The 16 terminal categories included "Other aliphatics, gte7, 1.0, nan", "Other aliphatics, lt7, 1.0, 1.0", "PFAA precursors, gte7, nan, nan", "PFAA precursors, lt7, 2.0, nan", "PFAA precursors, lt7, 4.0, nan", "PFAS derivatives, lt7, 1.0, 1.0", "PFAS derivatives, lt7, 1.0, 2.0", "PFAS derivatives, lt7, 3.0, 1.0", "Side-chain aromatics, gte7, 1.0, 2.0" and "Unclassified, lt7, 3.0, nan".

For a category that satisfied all these conditions, the next step would be identify the representative substances characterizing the category (namely the centroid and MaxMin substances and check

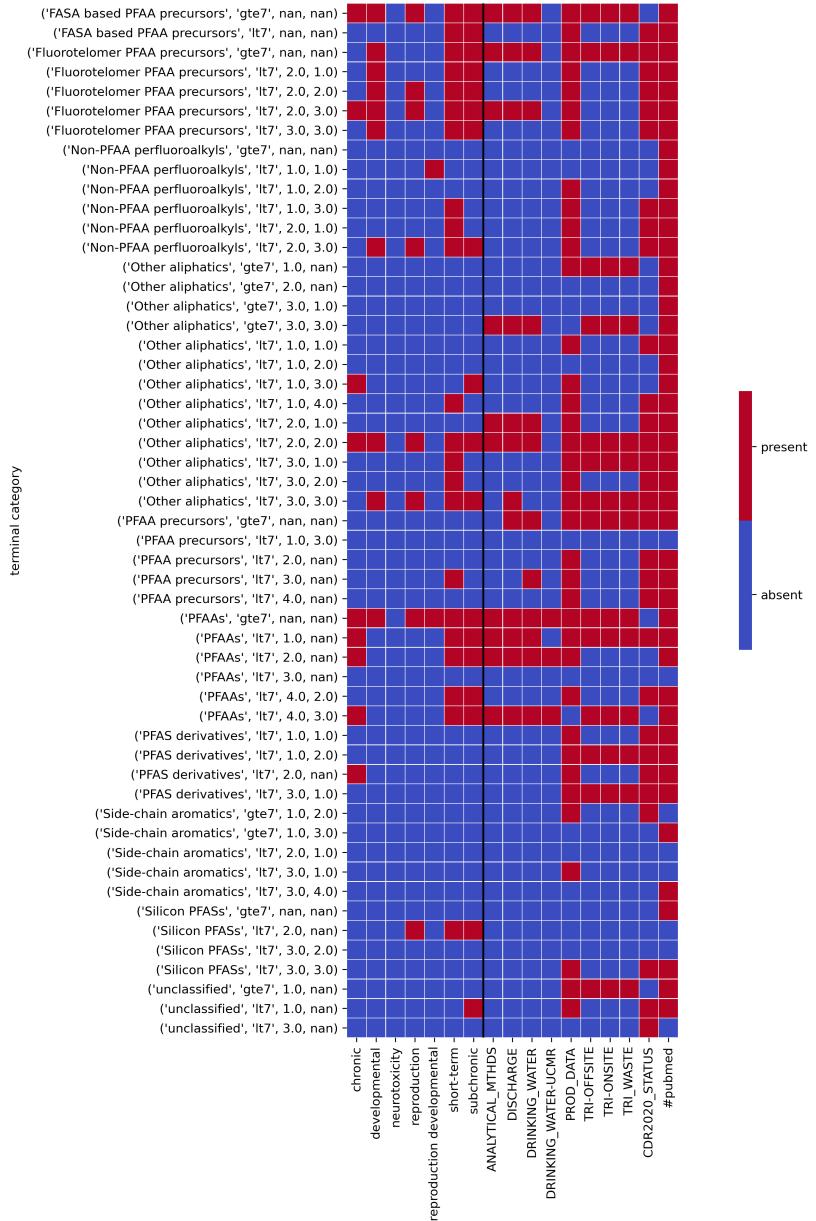


Figure 17: Heatmap of toxicity data availability and different environmental monitoring efforts nationwide

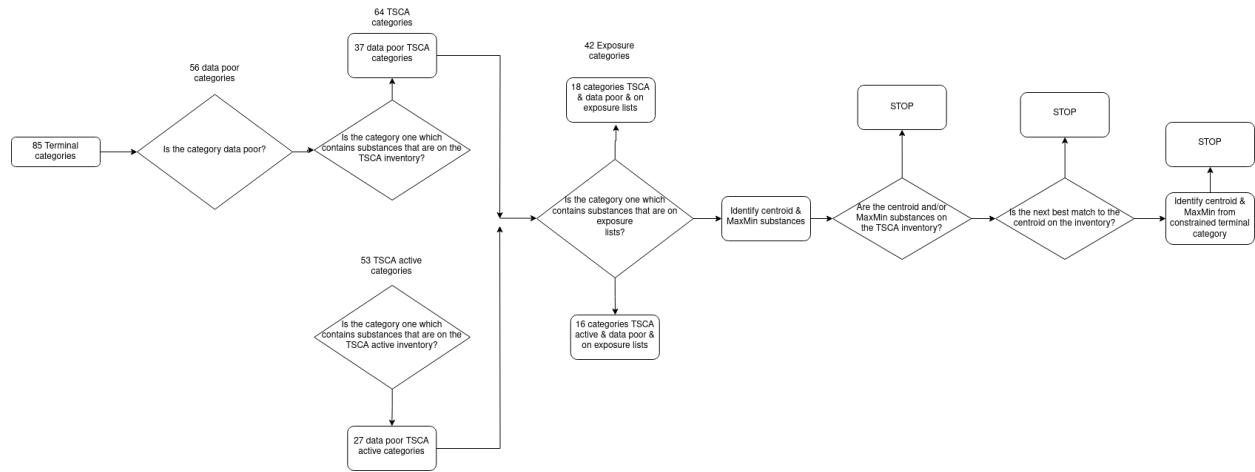


Figure 18: Workflow to highlight the main steps involved in prioritizing potential test candidate(s) selection for a given terminal category. If the Pubmed article availability was taken into account, 30 categories would overlap with the TSCA data poor categories or 23 of the TSCA active data poor categories.

whether any were on the TSCA inventory). If none of these were on the inventory, then the next step would be to check whether the next closest match to the centroid was on the inventory. If not, the next steps would be to identify the centroid and MaxMin substances from either the TSCA constrained inventory or the TSCA active constrained inventory for that terminal category. Figure 18 summarizes these steps in a conceptual workflow.

For illustrative purposes, terminal category "PFAA precursors, lt7, 2.0, nan" was identified that met the conditions of being a data poor category, containing members on the TSCA active inventory and containing members on various environmental monitoring lists. This terminal category comprises 22 members. The centroid, DTXSID60447694 was not on the TSCA inventory. Alternatively, the TSCA active centroid DTXSID0059877 [CASRN 355-38-4] could have been selected. Figure 19 shows a t-SNE project with the centroid, MaxMin and TSCA centroid substances shown for illustrative purposes to highlight their relative positions in the structural space captured within the terminal category.

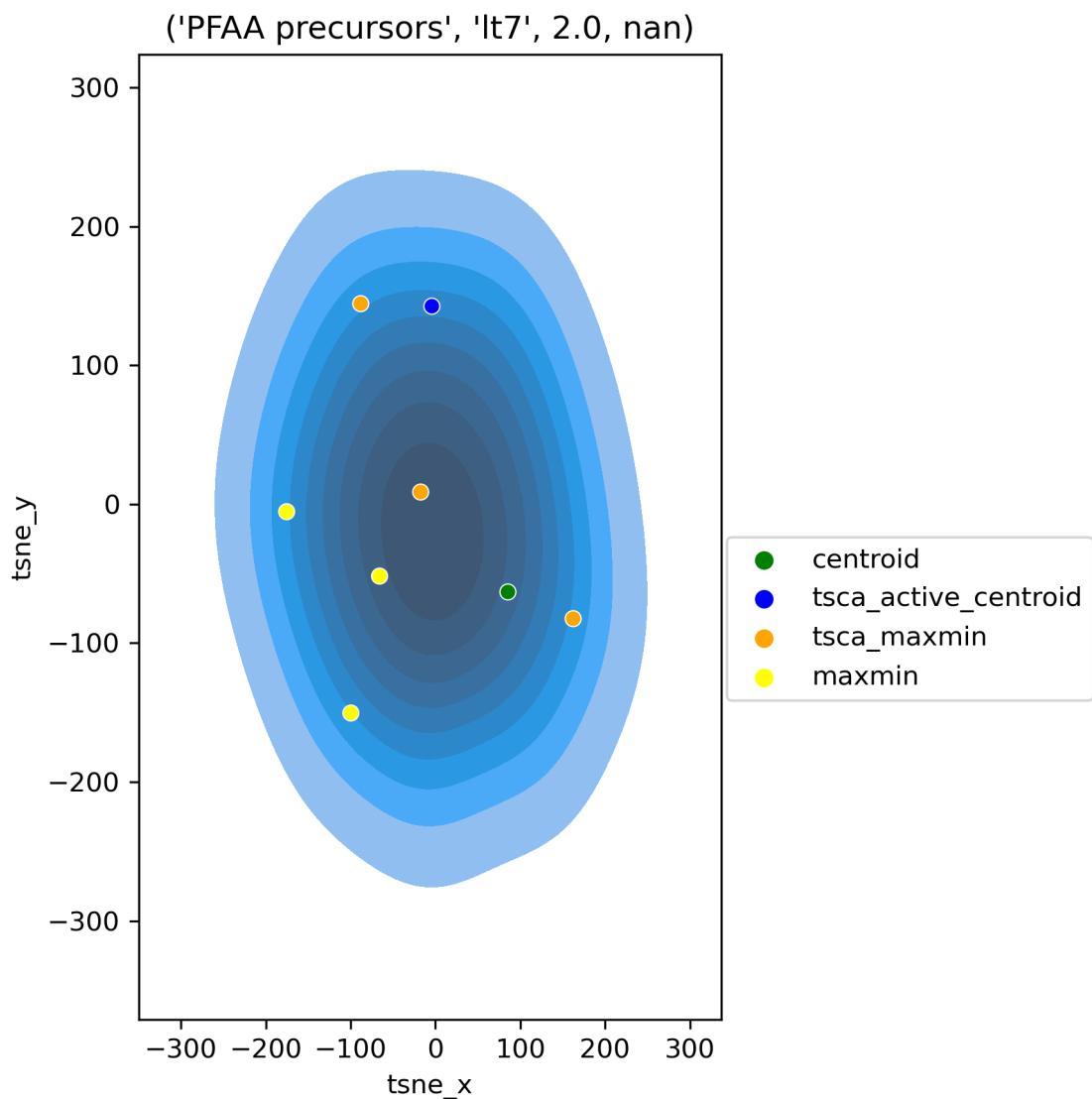


Figure 19: t-SNE projection of terminal category "PFAA precursors, It7, 2, nan" with its (TSCA active) centroid and MaxMin substances shown.

Whilst only 39% of the structural diversity is captured by the centroid and additional MaxMin substances, this increases to 66% if the constrained inventory is considered.

5. Conclusions

EPA was directed by Congress to develop a process for prioritizing which PFAS or classes of PFAS should be subject to additional research efforts based on potential for human exposure to, toxicity of, and other available information. Herein, we describe an approach to create a relevant PFAS Land-

scape using the TSCA SNUR rule definition to continue the efforts initiated in the National Testing Strategy. A landscape of 13,702 substances was created which comprised 10,576 substances in conjunction with simulated degradation products of TSCA relevant substances using the Catalogic expert system. Adding simulated degradates was intended to enrich the landscape by substances that might be expected to be found in the environment from existing substances in commerce. The simulated degradation products were derived from an expert system which includes training set substances that are PFAS though a full characterization of the model relative to the PFAS landscape was not feasible as some of the training set was proprietary in nature. For the portion of training set substances that could be evaluated - there was a minimal overlap in datasets as shown in Figure A1. The robustness of the simulated degradation products is a limitation in the approach and requires additional work but a pragmatic one given the absence of data to refine and improve the model further.

Using a scheme aligned to the OECD categories, substances were first assigned into one of 9 broad categories. These were refined by a surrogate for chain length to subset each of the substances into secondary categories such that a substance would be tagged by a primary category e.g. PFAAs and then a tag to denote greater/less than a carbon chain length of 7. The threshold of 7 was a pragmatic one to help identify long chains though the limitation was this was best suited for straight chain linear PFAS. For each of the secondary categories, a centroid was identified nominally expected to be representative of the overall category. Pairwise distances relative to the centroid were computed as a means of being able to rank order the similarity of category members as potential candidates for additional testing. Since in many cases, secondary categories were large and structurally diverse, an approach was developed to derive an objective threshold to determine if and when subcategorization using hierarchical clustering was needed to maximize structural similarity within a category. A threshold based on the 5th percentile of 'between category' distances was devised and applied as a stopping threshold. Ultimately a set of 85 terminal categories were proposed which could be prioritized based on those categories containing TSCA substances as well as considerations such as the sparsity of relevant toxicity and environmental data.

These 85 proposed terminal categories are limited in that they are structural in nature and anchored by the expectation that structural similar substances are likely to exhibit similar properties. Per current read-across technical guidance, structural similarity is only one component of a read-across assessment and many other considerations come into play. Given the sparsity of toxicity data for the PFAS landscape, this effort was very much a pragmatic starting point to prioritize PFAS into categories to that could help characterize the structural diversity relative to the toxicity space. Qualitative flags from NAM testing provided an additional component to help explore consistency of mechanistic data within a terminal category or help inform relevant higher tier data though it is

recognised that only 124 substances had been tested across the NAM assays and which had passed QC. The NAM flags have limited utility in highlighting potential tiered testing for a candidate substance, the NAM profile of the neighboring substance could provide some indications of which tiered testing might be more informative depending on category or what to expect as far as TK considerations. The NAM data also could be evaluated in concert with existing in vivo potency data where available to help rationalize the variability of the data. Publicly available in vivo data across the terminal categories was used to evaluate whether read-across could be potentially viable based on the variation of the in vivo data itself. A 10th percentile of the minimum POD values for a given substance was calculated to derive a single value per substance. Not all terminal categories were associated with toxicity data but for those categories, the following insights were noted; substances with lower carbon chain length within a category tended to exhibit lower toxicities (i.e., higher PODs), but the spread of PODs within a category could be large particularly for diverse categories based on carbon chain length, spanning 2 orders of magnitude or more. In addition to the shift in potency between terminal categories containing longer vs short chain lengths, there was also a shift between terminal categories with different functional groups e.g. Non-PFAA perfluoroalkyls tended to be less potent vs. PFAAs.

Evaluations of physical chemical similarity were also performed to explore the extent to which profiles were consistent across terminal categories.

Beyond a centroid as the nominal representative candidate substance characterizing the category, MaxMin approaches were applied to identify additional substances that could help bound the domain of the terminal category as a whole or when constrained by membership of TSCA substances or TSCA active substances only. A measure of %structural diversity coverage relative to the number of MaxMin substances selected was determined for both the unconstrained and TSCA active space - if 80% of the structural diversity needed to be captured from the full landscape across all 85 terminal categories, then over 5000 substances would need to be selected for further evaluation which would be practically challenging, whereas across the TSCA active landscape, a more manageable 115 substances would address a 80% structural diversity threshold. This highlights the initial pragmatic choice of setting a fixed number of MaxMin substances to be drawn from each terminal category would benefit from refinement to select different number of substances depending on the size of terminal category.

Refinements are expected to adjust the terminal categories based on the data generated through additional NAM testing or through the conventional data accessed. This workflow of devising structural categories shows promise as a practical and pragmatic means to help inform subsequent prioritization of candidates for additional tiered testing.

References

- [1] Z. Wang, J. C. DeWitt, C. P. Higgins, I. T. Cousins, *A Never-Ending Story of Per- and Polyfluoroalkyl Substances (PFASs)?*, Environmental Science & Technology 51 (5) (2017) 2508–2518, number: 5 Publisher: American Chemical Society. doi: [10.1021/acs.est.6b04806](https://doi.org/10.1021/acs.est.6b04806).
URL <https://doi.org/10.1021/acs.est.6b04806>
- [2] J. Glüge, M. Scheringer, I. T. Cousins, J. C. DeWitt, G. Goldenman, D. Herzke, R. Lohmann, C. A. Ng, X. Trier, Z. Wang, *An overview of the uses of per- and polyfluoroalkyl substances (PFAS)*, Environmental Science: Processes & Impacts 22 (12) (2020) 2345–2373, publisher: The Royal Society of Chemistry. doi: [10.1039/D0EM00291G](https://doi.org/10.1039/D0EM00291G).
URL <https://pubs.rsc.org/en/content/articlelanding/2020/em/d0em00291g>
- [3] L. G. T. Gaines, *Historical and current usage of per- and polyfluoroalkyl substances (PFAS): A literature review*, American Journal of Industrial Medicine 66 (5) (2023) 353–378, _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/ajim.23362>. doi: [10.1002/ajim.23362](https://doi.org/10.1002/ajim.23362).
URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/ajim.23362>
- [4] Y. Chen, H. Zhang, Y. Liu, J. A. Bowden, T. M. Tolaymat, T. G. Townsend, H. M. Solo-Gabriele, *Evaluation of per- and polyfluoroalkyl substances (PFAS) in leachate, gas condensate, stormwater and groundwater at landfills*, Chemosphere 318 (2023) 137903. doi: [10.1016/j.chemosphere.2023.137903](https://doi.org/10.1016/j.chemosphere.2023.137903).
URL <https://www.sciencedirect.com/science/article/pii/S0045653523001704>
- [5] J. Li, B. Xi, G. Zhu, Y. Yuan, W. Liu, Y. Gong, W. Tan, *A critical review of the occurrence, fate and treatment of per- and polyfluoroalkyl substances (PFASs) in landfills*, Environmental Research 218 (2023) 114980. doi: [10.1016/j.envres.2022.114980](https://doi.org/10.1016/j.envres.2022.114980).
URL <https://www.sciencedirect.com/science/article/pii/S0013935122023076>
- [6] N. Bolan, B. Sarkar, M. Vithanage, G. Singh, D. C. W. Tsang, R. Mukhopadhyay, K. Ramadass, A. Vinu, Y. Sun, S. Ramanayaka, S. A. Hoang, Y. Yan, Y. Li, J. Rinklebe, H. Li, M. B. Kirkham, *Distribution, behaviour, bioavailability and remediation of poly- and per-fluoroalkyl substances (PFAS) in solid biowastes and biowaste-treated soil*, Environment International 155 (2021) 106600. doi: [10.1016/j.envint.2021.106600](https://doi.org/10.1016/j.envint.2021.106600).
URL <https://www.sciencedirect.com/science/article/pii/S0160412021002257>
- [7] OECD, *Reconciling Terminology of the Universe of Per- and Polyfluoroalkyl Substances: Recommendations and Practical Guidance*, Series on Risk Management No. 61., Tech. rep. (2021).
- [8] Z. Wang, A. M. Buser, I. T. Cousins, S. Demattio, W. Drost, O. Johansson, K. Ohno, G. Patlewicz, A. M. Richard, G. W. Walker, G. S. White, E. Leinala, *A New OECD Definition for Per- and Polyfluoroalkyl Substances*, Environmental Science & Technology 55 (23) (2021) 15575–15578, number: 23 Publisher: American Chemical Society. doi: [10.1021/acs.est.1c06896](https://doi.org/10.1021/acs.est.1c06896).
URL <https://doi.org/10.1021/acs.est.1c06896>
- [9] C. M. Grulke, A. J. Williams, I. Thillanadarajah, A. M. Richard, *EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research*, Computational Toxicology (Amsterdam, Netherlands) 12 (Nov. 2019). doi: [10.1016/j.comtox.2019.100096](https://doi.org/10.1016/j.comtox.2019.100096).
- [10] U. EPA, *Per- and Poly-Fluoroalkyl Chemical Substances Designated as Inactive on the TSCA Inventory; Significant New Use Rule* (2023).
URL <https://www.federalregister.gov/documents/2023/01/26/2023-01156/per--and-poly-fluoroalkyl-chemical-substances-designated-as-inactive-on-the-tasca-inventory>
- [11] L. M. Carlson, M. Angrish, A. V. Shirke, E. G. Radke, B. Schulz, A. Kraft, R. Judson, G. Patlewicz, R. Blain, C. Lin, N. Vetter, C. Lemeris, P. Hartman, H. Hubbard, X. Arzuaga, A. Davis, L. V. Dishaw, I. L. Druwe, H. Hollinger, R. Jones, J. P. Kaiser, L. Lizarraga, P. D. Noyes, M. Taylor, A. J. Shapiro, A. J. Williams, K. A. Thayer, *Systematic Evidence Map for Over One*

- Hundred and Fifty Per- and Polyfluoroalkyl Substances (PFAS), Environmental Health Perspectives 130 (5) (2022) 056001, publisher: Environmental Health Perspectives. doi:10.1289/EHP10343.
URL <https://ehp.niehs.nih.gov/doi/10.1289/EHP10343>
- [12] OECD, *Guidance on Grouping of Chemicals, Second Edition OECD Series on Testing and Assessment, No 194*, OECD Publishing (2017). doi:<https://doi.org/10.1787/9789264274679-en>.
URL <https://www.oecd.org/publications/guidance-on-grouping-of-chemicals-second-edition-9789264274679-en.htm>
- [13] M. T. D. Cronin, *Chapter 1: An Introduction to Chemical Grouping, Categories and Read-Across to Predict Toxicity*, in: *Chemical Toxicity Prediction*, 2013, pp. 1-29. doi:10.1039/9781849734400-00001.
URL <https://pubs.rsc.org/en/content/chapter/bk9781849733847-00001/978-1-84973-384-7>
- [14] S. E. Escher, H. Kamp, S. H. Bennekou, A. Bitsch, C. Fisher, R. Graepel, J. G. Hengstler, M. Herzler, D. Knight, M. Leist, U. Norinder, G. Ouédraogo, M. Pastor, S. Stuard, A. White, B. Zdrazil, B. van de Water, D. Kroese, *Towards grouping concepts based on new approach methodologies in chemical hazard assessment: the read-across approach of the EU-ToxRisk project*, Archives of Toxicology 93 (12) (2019) 3643-3667, number: 12. doi:10.1007/s00204-019-02591-7.
URL <http://link.springer.com/10.1007/s00204-019-02591-7>
- [15] G. Patlewicz, M. T. Cronin, G. Helman, J. C. Lambert, L. E. Lizarraga, I. Shah, *Navigating through the minefield of read-across frameworks: A commentary perspective*, Computational Toxicology 6 (2018) 39-54. doi:10.1016/j.comtox.2018.04.002.
URL <https://linkinghub.elsevier.com/retrieve/pii/S2468111318300331>
- [16] G. Patlewicz, I. Shah, *Towards systematic read-across using Generalised Read-Across (GenRA)*, Computational Toxicology 25 (2023) 100258. doi:10.1016/j.comtox.2022.100258.
URL <https://www.sciencedirect.com/science/article/pii/S2468111322000469>
- [17] G. Patlewicz, A. M. Richard, A. J. Williams, R. S. Judson, R. S. Thomas, *Towards reproducible structure-based chemical categories for PFAS to inform and evaluate toxicity and toxicokinetic testing*, Computational Toxicology 24 (2022) 100250. doi:10.1016/j.comtox.2022.100250.
URL <https://www.sciencedirect.com/science/article/pii/S246811132200038X>
- [18] K. E. Carstens, T. Freudenrich, K. Wallace, S. Choo, A. Carpenter, M. Smeltz, M. S. Clifton, W. M. Henderson, A. M. Richard, G. Patlewicz, B. A. Wetmore, K. Paul Friedman, T. Shafer, *Evaluation of Per- and Polyfluoroalkyl Substances (PFAS) In Vitro Toxicity Testing for Developmental Neurotoxicity*, Chemical Research in Toxicology 36 (3) (2023) 402-419, publisher: American Chemical Society. doi:10.1021/acs.chemrestox.2c00344.
URL <https://doi.org/10.1021/acs.chemrestox.2c00344>
- [19] K. A. Houck, G. Patlewicz, A. M. Richard, A. J. Williams, M. A. Shobair, M. Smeltz, M. S. Clifton, B. Wetmore, A. Medvedev, S. Makarov, Bioactivity profiling of per- and polyfluoroalkyl substances (PFAS) identifies potential toxicity pathways related to molecular structure, Toxicology 457 (2021) 152789. doi:10.1016/j.tox.2021.152789.
- [20] K. A. Houck, K. P. Friedman, M. Feshuk, G. Patlewicz, M. Smeltz, M. S. Clifton, B. A. Wetmore, S. Velichko, A. Berenyi, E. L. Berg, Evaluation of 147 perfluoroalkyl substances for immunotoxic and other (patho)physiological activities through phenotypic screening of human primary cells, ALTEX 40 (2) (2023) 248-270. doi:10.14573/altex.2203041.
- [21] A. Kreutz, M. S. Clifton, W. M. Henderson, M. G. Smeltz, M. Phillips, J. F. Wambaugh, B. A. Wetmore, *Category-Based Toxicokinetic Evaluations of Data-Poor Per- and Polyfluoroalkyl Substances (PFAS) using Gas Chromatography Coupled with Mass Spectrometry*, Toxics 11 (5) (2023) 463. doi:10.3390/toxics11050463.
URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10223284/>
- [22] M. Smeltz, J. F. Wambaugh, B. A. Wetmore, *Plasma Protein Binding Evaluations of Per- and Polyfluoroalkyl Substances for Category-Based Toxicokinetic Assessment*, Chemical Research in Toxicology (May 2023). doi:10.1021/acs.chemrestox.3c00003.

- [23] M. G. Smeltz, M. S. Clifton, W. M. Henderson, L. McMillan, B. A. Wetmore, Targeted Per- and Polyfluoroalkyl substances (PFAS) assessments for high throughput screening: Analytical and testing considerations to inform a PFAS stock quality evaluation framework, *Toxicology and Applied Pharmacology* 459 (2023) 116355. doi:10.1016/j.taap.2022.116355.
- [24] T. E. Stoker, J. Wang, A. S. Murr, J. R. Bailey, A. R. Buckalew, High-Throughput Screening of ToxCast PFAS Chemical Library for Potential Inhibitors of the Human Sodium Iodide Symporter, *Chemical Research in Toxicology* 36 (3) (2023) 380–389. doi:10.1021/acs.chemrestox.2c00339.
- [25] A. J. Williams, C. M. Grulke, J. Edwards, A. D. McEachran, K. Mansouri, N. C. Baker, G. Patlewicz, I. Shah, J. F. Wambaugh, R. S. Judson, A. M. Richard, The CompTox Chemistry Dashboard: a community data resource for environmental chemistry, *Journal of Cheminformatics* 9 (1) (2017) 61. doi:10.1186/s13321-017-0247-6.
- [26] S. R. Heller, A. McNaught, I. Pletnev, S. Stein, D. Tchekhovskoi, *InChI, the IUPAC International Chemical Identifier*, *Journal of Cheminformatics* 7 (1) (2015) 23. doi:10.1186/s13321-015-0068-4.
URL <https://doi.org/10.1186/s13321-015-0068-4>
- [27] D. Rogers, M. Hahn, *Extended-Connectivity Fingerprints*, *Journal of Chemical Information and Modeling* 50 (5) (2010) 742–754, publisher: American Chemical Society. doi:10.1021/ci100050t.
URL <https://doi.org/10.1021/ci100050t>
- [28] L. van er Maaten, G. Hinton, Visualizing Data using t-SNE., *Journal of Machine Learning Research* 8 (2018) 2579–2605.
- [29] A. Su, K. Rajan, A database framework for rapid screening of structure-function relationships in PFAS chemistry, *Scientific Data* 8 (1) (2021) 14, number: 1. doi:10.1038/s41597-021-00798-x.
- [30] K. Mansouri, C. M. Grulke, R. S. Judson, A. J. Williams, OPERA models for predicting physicochemical properties and environmental fate endpoints, *Journal of Cheminformatics* 10 (1) (2018) 10. doi:10.1186/s13321-018-0263-1.
- [31] W. S. Chambers, J. G. Hopkins, S. M. Richards, *A Review of Per- and Polyfluorinated Alkyl Substance Impairment of Reproduction*, *Frontiers in Toxicology* 3 (2021).
URL <https://www.frontiersin.org/articles/10.3389/ftox.2021.732436>
- [32] K. Sznajder-Katarzyńska, M. Surma, I. Cieślik, *A Review of Perfluoroalkyl Acids (PFAAs) in terms of Sources, Applications, Human Exposure, Dietary Intake, Toxicity, Legal Regulation, and Methods of Determination*, *Journal of Chemistry* 2019 (2019) e2717528, publisher: Hindawi. doi:10.1155/2019/2717528.
URL <https://www.hindawi.com/journals/jchem/2019/2717528/>
- [33] A. M. Richard, H. Hidle, G. Patlewicz, A. J. Williams, *Identification of Branched and Linear Forms of PFOA and Potential Precursors: A User-Friendly SMILES Structure-based Approach*, *Frontiers in Environmental Science* 10 (2022).
URL <https://www.frontiersin.org/article/10.3389/fenvs.2022.865488>
- [34] A. M. Richard, R. Lougee, M. Adams, H. Hidle, C. Yang, J. Rathman, T. Magdziarz, B. Bienfait, A. J. Williams, G. Patlewicz, A New CSRML Structure-Based Fingerprint Method for Profiling and Categorizing Per- and Polyfluoroalkyl Substances (PFAS), *Chemical Research in Toxicology* 36 (3) (2023) 508–534. doi:10.1021/acs.chemrestox.2c00403.
- [35] C. Yang, A. Tarkhov, J. Marusczyk, B. Bienfait, J. Gasteiger, T. Kleinoeder, T. Magdziarz, O. Sacher, C. H. Schwab, J. Schwoebel, L. Terfloth, K. Arvidson, A. Richard, A. Worth, J. Rathman, 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) (2015) 510–528. doi:10.1021/ci500667v.
- [36] N. M. O'Boyle, R. A. Sayle, *Comparing structural fingerprints using a literature-based similarity benchmark*, *Journal of Cheminformatics* 8 (1) (2016) 36. doi:10.1186/s13321-016-0148-0.
URL <https://doi.org/10.1186/s13321-016-0148-0>
- [37] J. H. Ward, *Hierarchical Grouping to Optimize an Objective Function*, *Journal of the American Statistical Association* 58 (301) (1963) 236–244, publisher: Taylor & Francis _eprint: <https://www.tandfonline.com/doi/pdf/10.1080/01621459.1963.10500845>. doi:10.1080/01621459.1963.10500845.

- URL <https://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10500845>
- [38] M. Ashton, J. Barnard, F. Casset, M. Charlton, G. Downs, D. Gorse, J. Holliday, R. Lahana, P. Willett, *Identification of Diverse Database Subsets using Property-Based and Fragment-Based Molecular Descriptions*, Quantitative Structure-Activity Relationships 21 (6) (2002) 598-604, _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/qsar.200290002>. doi:[10.1002/qsar.200290002](https://doi.org/10.1002/qsar.200290002).
- URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/qsar.200290002>
- [39] M. Snarey, N. K. Terrett, P. Willett, D. J. Wilton, Comparison of algorithms for dissimilarity-based compound selection, Journal of Molecular Graphics & Modelling 15 (6) (1997) 372-385. doi:[10.1016/s1093-3263\(98\)00008-4](https://doi.org/10.1016/s1093-3263(98)00008-4).
- [40] N. Baker, T. Knudsen, A. Williams, *Abstract Sifter: a comprehensive front-end system to PubMed*, F1000Research 6 (2017) Chem Inf Sci-2164. doi:[10.12688/f1000research.12865.1](https://doi.org/10.12688/f1000research.12865.1).
- URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5801564/>
- [41] H. K. Liberatore, S. R. Jackson, M. J. Strynar, J. P. McCord, *Solvent Suitability for HFPO-DA ("GenX" Parent Acid) in Toxicological Studies*, Environmental Science & Technology Letters 7 (7) (2020) 477-481, publisher: American Chemical Society. doi:[10.1021/acs.estlett.0c00323](https://doi.org/10.1021/acs.estlett.0c00323).
- URL <https://doi.org/10.1021/acs.estlett.0c00323>
- [42] C. Zhang, A. C. McElroy, H. K. Liberatore, N. L. M. Alexander, D. R. U. Knappe, *Stability of Per- and Polyfluoroalkyl Substances in Solvents Relevant to Environmental and Toxicological Analysis*, Environmental Science & Technology 56 (10) (2022) 6103-6112, publisher: American Chemical Society. doi:[10.1021/acs.est.1c03979](https://doi.org/10.1021/acs.est.1c03979).
- URL <https://doi.org/10.1021/acs.est.1c03979>
- [43] D. L. Filer, P. Kothiyal, R. W. Setzer, R. S. Judson, M. T. Martin, tcpl: the ToxCast pipeline for high-throughput screening data, Bioinformatics (Oxford, England) 33 (4) (2017) 618-620. doi:[10.1093/bioinformatics/btw680](https://doi.org/10.1093/bioinformatics/btw680).
- [44] T. Sheffield, J. Brown, S. Davidson, K. P. Friedman, R. Judson, *tcplfit2: an R-language general purpose concentration-response modeling package*, Bioinformatics 38 (4) (2022) 1157-1158. doi:[10.1093/bioinformatics/btab779](https://doi.org/10.1093/bioinformatics/btab779).
- URL <https://doi.org/10.1093/bioinformatics/btab779>
- [45] R. Judson, K. Houck, M. Martin, A. M. Richard, T. B. Knudsen, I. Shah, S. Little, J. Wambaugh, R. Woodrow Setzer, P. Kothiyal, J. Phuong, D. Filer, D. Smith, D. Reif, D. Rotroff, N. Kleinstreuer, N. Sipes, M. Xia, R. Huang, K. Crofton, R. S. Thomas, *Editor's Highlight: Analysis of the Effects of Cell Stress and Cytotoxicity on In Vitro Assay Activity Across a Diverse Chemical and Assay Space*, Toxicological Sciences 152 (2) (2016) 323-339. doi:[10.1093/toxsci/kfw092](https://doi.org/10.1093/toxsci/kfw092).
- URL <https://doi.org/10.1093/toxsci/kfw092>
- [46] R. S. Judson, R. S. Thomas, N. Baker, A. Simha, X. M. Howey, C. Marable, N. C. Kleinstreuer, K. A. Houck, *Workflow for Defining Reference Chemicals for Assessing Performance of In Vitro Assays*, ALTEX 36 (2) (2019) 261-276. doi:[10.14573/altex.1809281](https://doi.org/10.14573/altex.1809281).
- URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6784312/>
- [47] S. Padilla, D. Corum, B. Padnos, D. L. Hunter, A. Beam, K. A. Houck, N. Sipes, N. Kleinstreuer, T. Knudsen, D. J. Dix, D. M. Reif, *Zebrafish developmental screening of the ToxCast™ Phase I chemical library*, Reproductive Toxicology (Elmsford, N.Y.) 33 (2) (2012) 174-187, number: 2. doi:[10.1016/j.reprotox.2011.10.018](https://doi.org/10.1016/j.reprotox.2011.10.018).
- [48] J. H. Olker, J. J. Korte, J. S. Denny, P. C. Hartig, M. C. Cardon, C. N. Knutsen, P. M. Kent, J. P. Christensen, S. J. Degitz, M. W. Hornung, *Screening the ToxCast Phase 1, Phase 2, and e1k Chemical Libraries for Inhibitors of Iodothyronine Deiodinases*, Toxicological Sciences: An Official Journal of the Society of Toxicology 168 (2) (2019) 430-442. doi:[10.1093/toxsci/kfy302](https://doi.org/10.1093/toxsci/kfy302).
- [49] J. H. Olker, J. J. Korte, J. S. Denny, J. T. Haselman, P. C. Hartig, M. C. Cardon, M. W. Hornung, S. J. Degitz, *In vitro screening for chemical inhibition of the iodide recycling enzyme, iodotyrosine deiodinase*, Toxicology in vitro: an international journal published in association with BIBRA 71 (2021) 105073. doi:[10.1016/j.tiv.2020.105073](https://doi.org/10.1016/j.tiv.2020.105073).

- [50] K. Paul Friedman, E. D. Watt, M. W. Hornung, J. M. Hedge, R. S. Judson, K. M. Crofton, K. A. Houck, S. O. Simmons, Tiered High-Throughput Screening Approach to Identify Thyroperoxidase Inhibitors Within the ToxCast Phase I and II Chemical Libraries, *Toxicological Sciences: An Official Journal of the Society of Toxicology* 151 (1) (2016) 160-180. [doi:10.1093/toxsci/kfw034](https://doi.org/10.1093/toxsci/kfw034).
- [51] M. Montaño, E. Cocco, C. Guignard, G. Marsh, L. Hoffmann, A. Bergman, A. C. Gutleb, A. J. Murk, New approaches to assess the transthyretin binding capacity of bioactivated thyroid hormone disruptors, *Toxicological Sciences: An Official Journal of the Society of Toxicology* 130 (1) (2012) 94-105. [doi:10.1093/toxsci/kfs228](https://doi.org/10.1093/toxsci/kfs228).
- [52] A. Hammitzsch, C. Tallant, O. Fedorov, A. O'Mahony, P. E. Brennan, D. A. Hay, F. O. Martinez, M. H. Al-Mossawi, J. de Wit, M. Vecellio, C. Wells, P. Wordsworth, S. Müller, S. Knapp, P. Bowness, *CBP30, a selective CBP/p300 bromodomain inhibitor, suppresses human Th17 responses*, *Proceedings of the National Academy of Sciences* 112 (34) (2015) 10768-10773, publisher: *Proceedings of the National Academy of Sciences*. [doi:10.1073/pnas.1501956112](https://doi.org/10.1073/pnas.1501956112)
URL <https://www.pnas.org/doi/10.1073/pnas.1501956112>
- [53] A. O'Mahony, M. R. John, H. Cho, M. Hashizume, E. H. Choy, Discriminating phenotypic signatures identified for tocilizumab, adalimumab, and tofacitinib monotherapy and their combinations with methotrexate, *Journal of Translational Medicine* 16 (1) (2018) 156. [doi:10.1186/s12967-018-1532-5](https://doi.org/10.1186/s12967-018-1532-5).
- [54] L. Simms, E. Mason, E. L. Berg, F. Yu, K. Rudd, L. Czekala, E. Treilles Sticken, O. Brinster, R. Wieczorek, M. Stevenson, T. Walele, Use of a rapid human primary cell-based disease screening model, to compare next generation products to combustible cigarettes, *Current Research in Toxicology* 2 (2021) 309-321. [doi:10.1016/j.crtox.2021.08.003](https://doi.org/10.1016/j.crtox.2021.08.003).
URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8408431/>
- [55] N. C. Kleinstreuer, J. Yang, E. L. Berg, T. B. Knudsen, A. M. Richard, M. T. Martin, D. M. Reif, R. S. Judson, M. Polokoff, D. J. Dix, R. J. Kavlock, K. A. Houck, Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms, *Nature Biotechnology* 32 (6) (2014) 583-591. [doi:10.1038/nbt.2914](https://doi.org/10.1038/nbt.2914).
- [56] B. A. Wetmore, J. F. Wambaugh, S. S. Ferguson, M. A. Sochaski, D. M. Rotroff, K. Freeman, H. J. Clewell, III, D. J. Dix, M. E. Andersen, K. A. Houck, B. Allen, R. S. Judson, R. Singh, R. J. Kavlock, A. M. Richard, R. S. Thomas, *Integration of Dosimetry, Exposure, and High-Throughput Screening Data in Chemical Toxicity Assessment*, *Toxicological Sciences* 125 (1) (2012) 157-174. [doi:10.1093/toxsci/kfr254](https://doi.org/10.1093/toxsci/kfr254)
URL <https://doi.org/10.1093/toxsci/kfr254>
- [57] D. Crizer, J. Zhou, K. Lavrich, B. Wetmore, S. Ferguson, M. DeVito, B. Merrick, In vitro hepatic clearance of per-and polyfluoroalkyl substances (pfas), *Toxicologist* (2023) unpublished.
- [58] B. A. Wetmore, J. F. Wambaugh, B. Allen, S. S. Ferguson, M. A. Sochaski, R. W. Setzer, K. A. Houck, C. L. Strope, K. Cantwell, R. S. Judson, E. LeCluyse, H. J. Clewell, R. S. Thomas, M. E. Andersen, *Incorporating High-Throughput Exposure Predictions With Dosimetry-Adjusted In Vitro Bioactivity to Inform Chemical Toxicity Testing*, *Toxicological Sciences* 148 (1) (2015) 121-136. [doi:10.1093/toxsci/kfv171](https://doi.org/10.1093/toxsci/kfv171)
URL <https://doi.org/10.1093/toxsci/kfv171>
- [59] K. Paul Friedman, M. Gagne, L.-H. Loo, P. Karamertzanis, T. Netzeva, T. Sobanski, J. A. Franzosa, A. M. Richard, R. R. Lougee, A. Gissi, J.-Y. J. Lee, M. Angrish, J. L. Dorne, S. Foster, K. Raffaele, T. Bahadori, M. R. Gwinn, J. Lambert, M. Whelan, M. Rasenberg, T. Barton-Maclaren, R. S. Thomas, Utility of In Vitro Bioactivity as a Lower Bound Estimate of In Vivo Adverse Effect Levels and in Risk-Based Prioritization, *Toxicological Sciences: An Official Journal of the Society of Toxicology* 173 (1) (2020) 202-225. [doi:10.1093/toxsci/kfz201](https://doi.org/10.1093/toxsci/kfz201).
- [60] D. E. Dawson, C. Lau, P. Pradeep, R. R. Sayre, R. S. Judson, R. Tornero-Velez, J. F. Wambaugh, *A Machine Learning Model to Estimate Toxicokinetic Half-Lives of Per- and Polyfluoro-Alkyl Substances (PFAS) in Multiple Species*, *Toxics* 11 (2) (2023) 98, number: 2 Publisher: Multidisciplinary Digital Publishing Institute. [doi:10.3390/toxics11020098](https://doi.org/10.3390/toxics11020098).
URL <https://www.mdpi.com/2305-6304/11/2/98>

- [61] J. Wang, D. R. Hallinger, A. S. Murr, A. R. Buckalew, R. R. Lougee, A. M. Richard, S. C. Laws, T. E. Stoker, High-throughput screening and chemotype-enrichment analysis of ToxCast phase II chemicals evaluated for human sodium-iodide symporter (NIS) inhibition, *Environment International* 126 (2019) 377-386. [doi:10.1016/j.envint.2019.02.024](https://doi.org/10.1016/j.envint.2019.02.024).

Appendix A. Supplementary information

Appendix A.1. Evaluating the feasibility of subdividing the primary categories

Corina Symphony on the command line (licensed from Molecular Networks GmbH and Altamira LLC) was used to compute the 129 PFAS ToxPrints³⁴. The Fisher's exact test was used to compute an odds ratio and associated p value for each PFAS ToxPrint relative to the OECD primary category designation. This was comparable with the methodology discussed in Wang et al.⁶¹. A PFAS ToxPrint was considered enriched if it had an odds ratio greater than or equal to 3, a one-sided Fishers exact p-value less than 0.05 (probability value of the odds ratio being greater than 1) and the number of True Positives (TP) was determined to be greater than or equal to 3. For the "Unclassified" primary category, the top 2 enriched ToxPrints were PFAS chain features: FT_n3_N, FT_n1_OP both of which represent fluorotelomer chains with either 3 or 1 CH₂ units and a nitrogen or organophosphorus terminus. On the otherhand, the "Other aliphatics" had heteroatoms, nitrile and primary amines as enriched functional groups. The intention was to explore whether certain types of features were specifically enriched in these broad primary categories to consider splitting them apart to reduce the starting membership. The full set of enrichments for all primary categories are provided as a separate data file.

Supplementary Figures

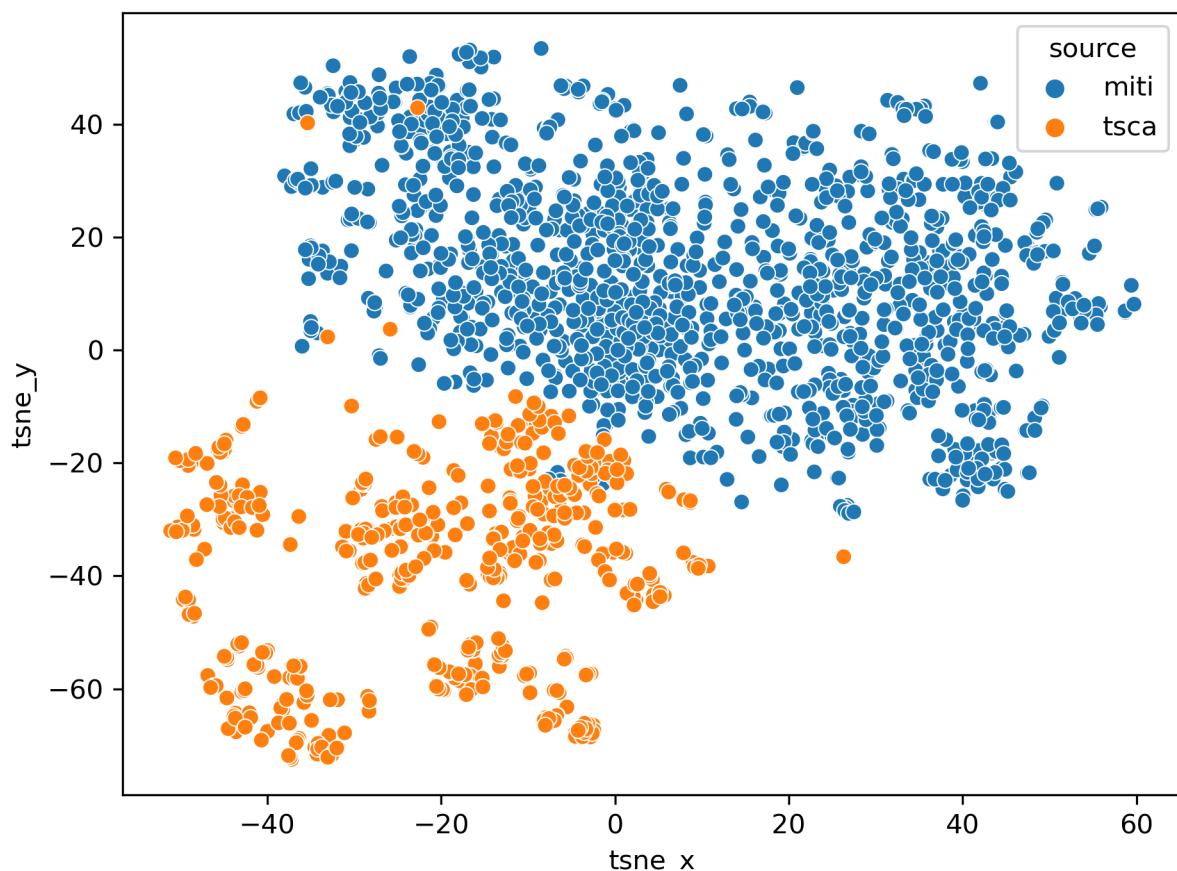


Figure A1: Overlap of MITI training data substances with TSCA substances using Morgan chemical fingers and represented in a t-SNE plot

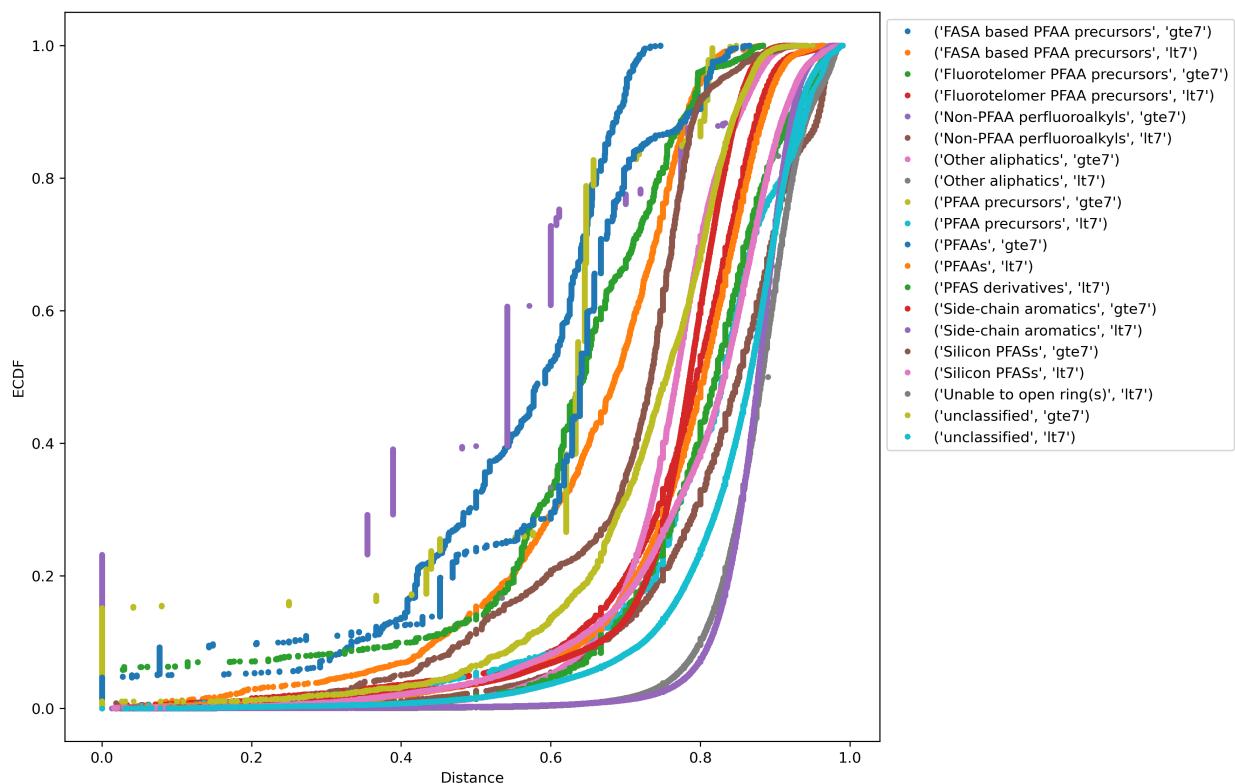


Figure A2: ECDFs of the within categories based on the chain length threshold of 7

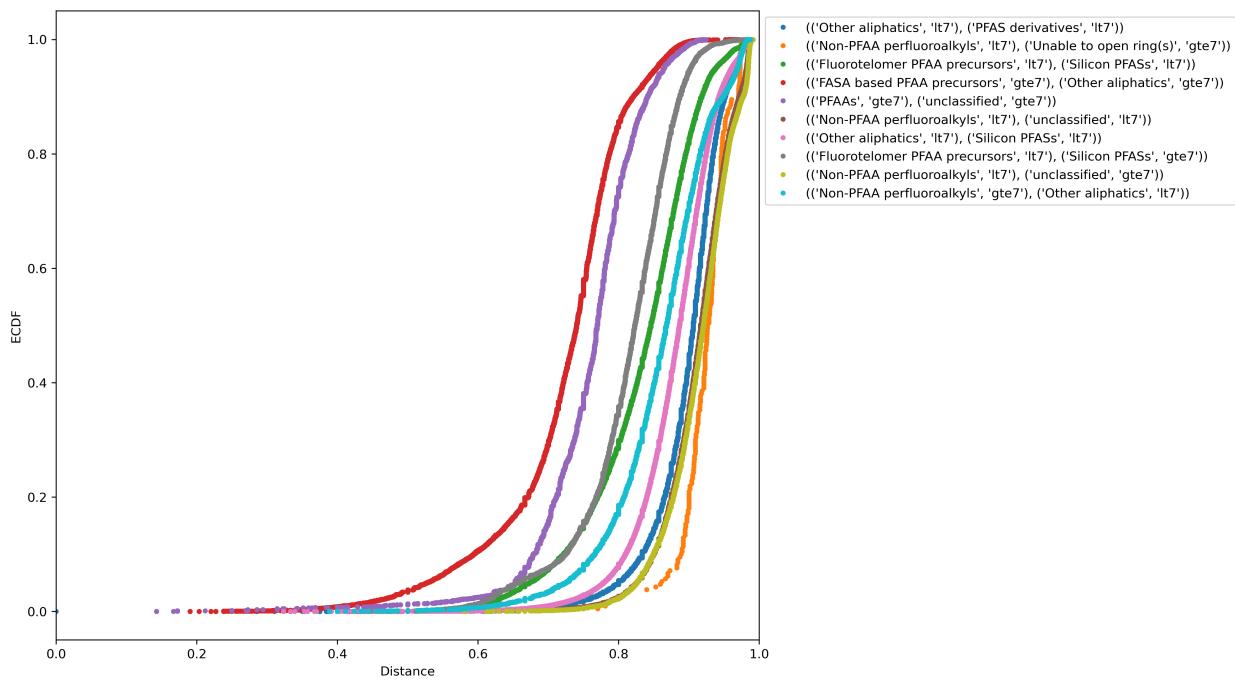


Figure A3: EDCFs for selected between category combinations for carbon chain length categories

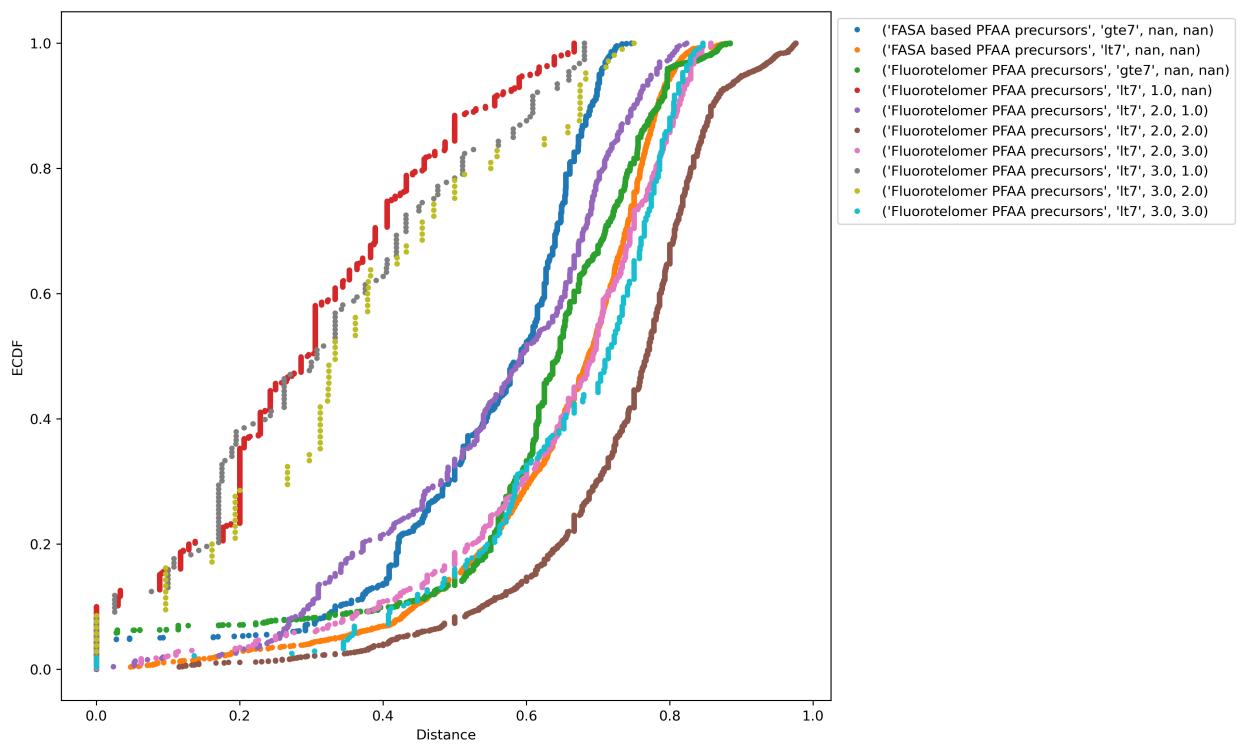


Figure A4: EDCFs for selected terminal categories to demonstrate left shift in pairwise distance

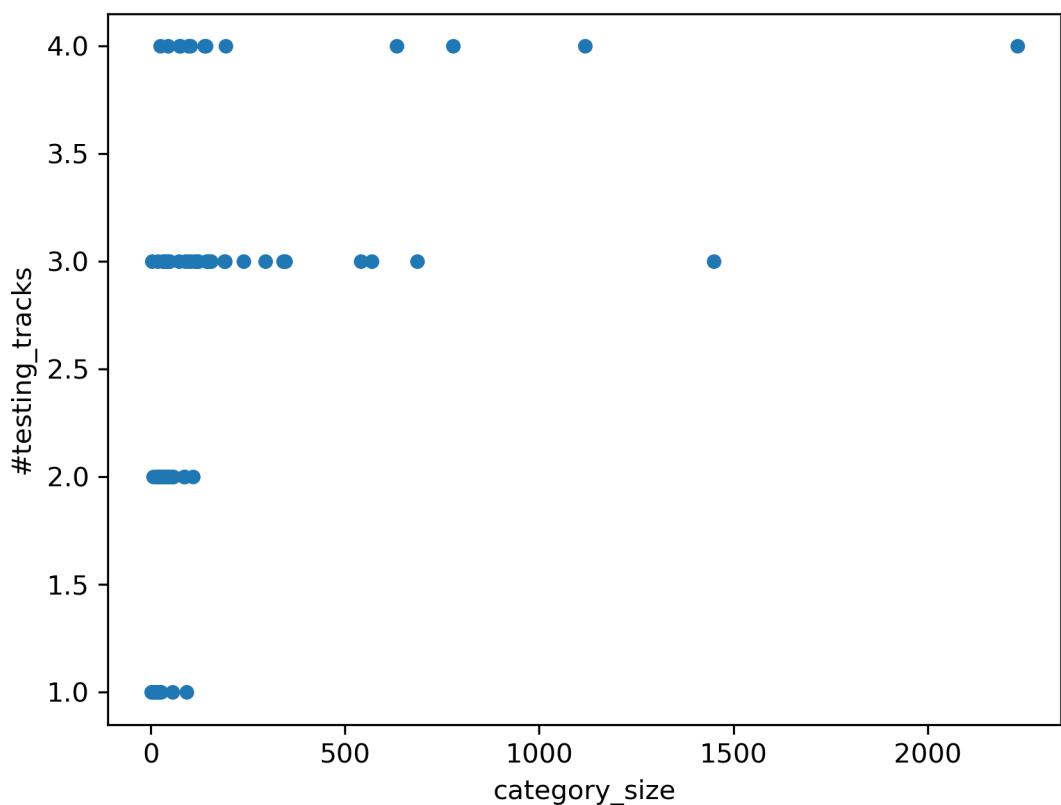


Figure A5: Correlation between terminal categories with large membership size and the number of testing tracks represented amongst their memberships

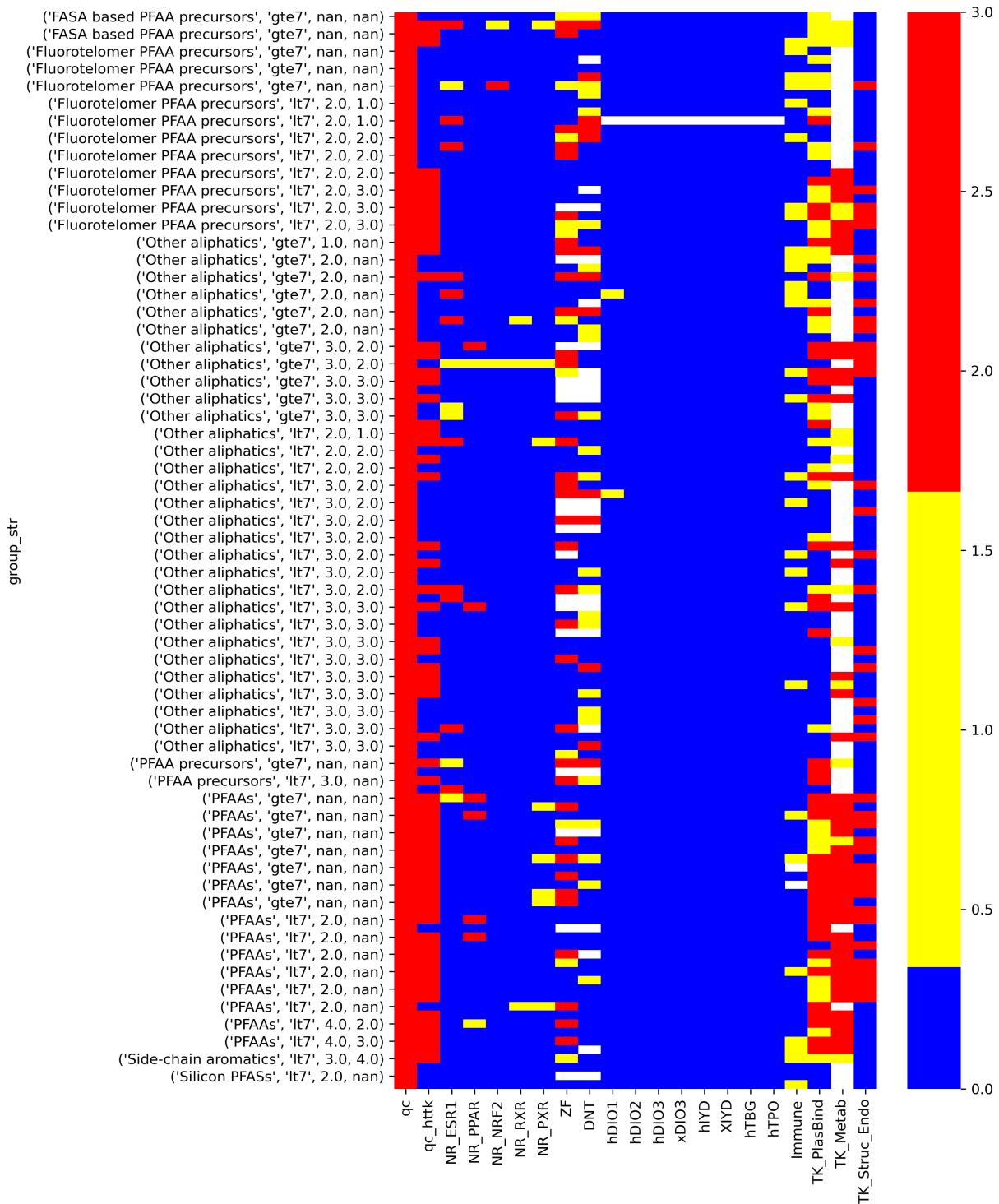


Figure A6: Heatmap of NAM flags for substances tested that overlap with the PFAS inventory