Chemical Research in Toxicology

pubs.acs.org/crt

Article

Quantifying Analogue Suitability for SAR-Based Read-Across Toxicological Assessment

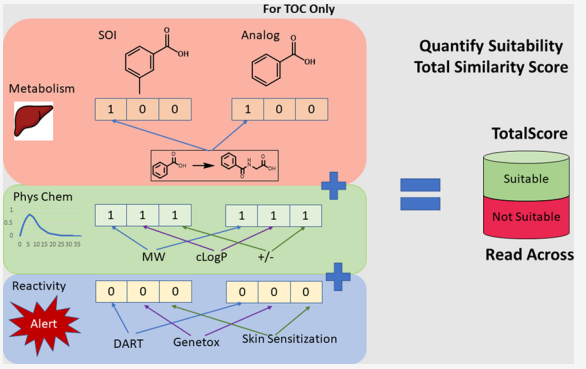
Cathy Lester,* ElLantae Byrd, Mahmoud Shobair, and Gang Yan

Cite This: *Chem. Res. Toxicol.* 2023, 36, 230−242

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Structure activity relationship (SAR)-based read-across often is an integral part of toxicological safety assessment, and justification of the prediction presents the most challenging aspect of the approach. It has been established that structural consideration alone is inadequate for selecting analogues and justifying their use, and biological relevance must be incorporated. Here we introduce an approach for considering biological and toxicological related features quantitatively to compute a similarity score that is concordant with suitability for a read-across prediction for systemic toxicity. Fingerprint keys for comparing metabolism, reactivity, and physical chemical properties are presented and used to compare these attributes for 14 case study chemicals each with a list of potential analogues. Within each case study, the sum of these nonstructural similarity scores is consistent with suitability for read-across established using an approach based on expert judgment. Machine learning is applied to determine the contributions from each of the similarity attributes revealing their importance for each structure class. This approach is used to quantify and communicate the differences between a target and a potential analogue as well as rank analogue quality when more than one is relevant. A numerical score with easily interpreted fingerprints increases transparency and consistency among experts, facilitates implementation by others, and ultimately increases chances for regulatory acceptance.



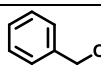
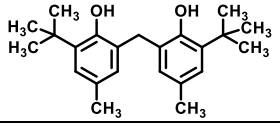
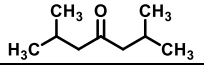
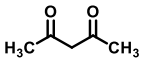
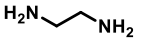
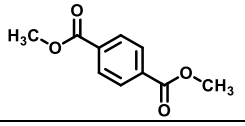
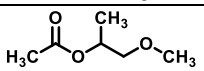
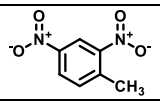
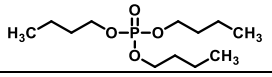
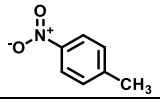
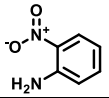
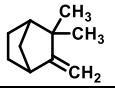
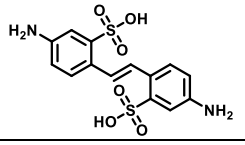
1. INTRODUCTION

Structure activity relationship (SAR)-based read-across is one of the most widely used animal alternative methods for safety assessment. The technique is used to fill toxicological data gaps and involves the use of toxicological data for a data-rich source compound or analogue to establish the safety of a data-poor target compound or structure of interest (SOI). Read-across is simple in concept but can be difficult in practice, and the most challenging aspect of the approach involves identifying analogues and justifying their use in predicting toxicity. The P&G framework for analogue selection relies on the use of expert judgment to evaluate similarity in structure, metabolism, reactivity, and physical chemical properties as it relates to systemic toxicity to assign one of the following analogue ratings: "suitable", "suitable with interpretation", "suitable with precondition", and "not suitable".[1,2] "Suitable" analogues are very close to the analogue structurally, undergo the same metabolic pathways, and display the same structural alerts with nearly identical physical chemical properties. "Suitable with interpretation" analogues are similar in all attributes but with small differences in one attribute with little impact on relative toxicity such as differences in physical chemical properties. If an analogue is a metabolite of the SOI or if the SOI and analogue potentially may converge metabolically, the analogue is rated "suitable with precondition". Finally, "not suitable" analogues possess little similarity in one or more attributes. The approach was tested using 14 case study chemicals and was shown to correlate well with in vivo toxicological data for systemic toxicological effects including the following end points: repeat-dose, developmental and reproductive (DART), skin sensitization, genetic toxicity, and carcinogenicity (genotoxic mode of action).[3]

The framework has been applied for over 10 years to establish the safety assessment of more than 1500 compounds. Recently, by analyzing approximately 4000 SOI/analogue pairs of structures, it was shown that there is little correlation between suitability for read-across and quantitative structural similarity scores computed using Tanimoto comparisons of molecular fingerprints. Instead, all SOI/analogue pairs of structures analyzed where the analogue was deemed suitable for read-across were described as matched molecular pairs (MMPs) of structures possessing a common core or scaffold with small changes between attached R groups. These small changes fell into one of five categories of changes for both

Received: September 30, 2022
Published: January 26, 2023

Table 1. Case Study Chemicals

	SOI Structure	CAS RN	Chemical Name
1		100-51-6	Benzyl alcohol
2		119-47-1	2,2'-Methylenebis[6-tert-butyl-p-cresol]
3		108-83-8	Diisobutyl ketone
4		123-54-6	Acetylacetone
5		107-15-3	Ethylenediamine
6		120-61-6	Dimethyl terephthalate
7		108-65-6	1-Methoxy-2-propyl acetate
8		121-14-2	2,4-Dinitrotoluene
9		126-73-8	Tributyl phosphate
10		99-99-0	4-Nitrotoluene
11		88-74-4	2-Nitroaniline
12		112-53-8	Lauryl alcohol
13		79-92-5	Camphene
14		81-11-8	4,4'-Diamino-2,2'-stilbenedisulfonic acid

aliphatic and aromatic compounds. A workflow describing these structural changes recently has been published to facilitate implementation of the approach by others.[4]

While current practices of read-across often require calculating a quantitative structural similarity score as a first step in analogue selection, it is well-known that structural considerations alone are insufficient for justifying analogue suitability.[2,5] However, the similarity score is appealing because it provides a quantitative comparison of two structures even if comparing molecular fingerprints is not always sensitive to differences in toxicity. For this reason, quantitation of all similarity attributes, chemical as well as biological, is needed and has been proposed to facilitate regulatory acceptance.[6] Frameworks and tools enabling a read-across toxicological assessment have been described and combined into a harmonized framework that suggests integrating quantitative measures of structural similarity with other similarity considerations such as physical chemical properties, structural alerts, and features from a metabolic profiler.[7]

There are examples of recent efforts to quantify non-structural similarities in read-across. Gadaleta et al. have combined structural, biological, and metabolic similarity in the analogue selection process using MACCS (Molecular Access System) fingerprints for structural similarity, biotransformations predicted using SyGMa (Systematic Generation of possible Metabolites) software for metabolic similarity, and the outcome of high-throughput screening assays from PubChem for biological similarity.[8,9] The authors demonstrate improved predictivity using all three similarity measures relative to structural similarity alone. Others have proposed

https://doi.org/10.1021/acs.chemrestox.2c00311
Chem. Res. Toxicol. 2023, 36, 230−242

similar methods for assessing metabolic similarity. Boyce et al. computed metabolic similarity using pathway transformations predicted by Meteor for 3,4-toluenediamine and potential analogues for defining a fingerprint with comparisons performed by calculating a Jaccard distance.[10] Yordanova et al. considered metabolic similarity by comparing side-by-side arrays of metabolites for a target/analogue pair of structures for identification of common metabolic pathways, common metabolites, as well as consideration of the formation of common reactive metabolites.[11] Yang et al. defined "analogue quality" as the geometric mean of similarity in structure and molecular/physical chemical properties for a target/analogue pair and show that this metric may be used to derive a range of NOAEL (No Observed Adverse Effect Level) values for the target based on those values determined for analogues deemed to be high quality.[12] Finally, "biological similarity" may be calculated by comparing the output from various HTS (high-throughput screening) assays that may include profiles of gene expression, assays of enzymatic activity, or measurements of cytotoxicity.[13,14]

The work presented in this Article quantifies similarity in metabolism, reactivity, and physical chemical properties between an SOI and a potential analogue to justify use of the analogue in a read-across assessment for the prediction of systemic toxicological effects which are not end point specific but may include references to structural alerts associated with known toxicity. Fingerprints representing metabolism, reactivity, and physical chemical properties are presented and defined for the 14 case study chemicals previously used to validate the P&G framework for analogue selection and listed in Table 1.[3] To characterize similarity in metabolism, the product of the similarity scores characterizing the metabolic pathway of the compound and the presence of reactive species resulting from metabolic biotransformation or limited detoxification is used to heavily weight any differences in metabolism between the two compounds. A fingerprint listing structural alerts for each compound is used to compute similarity in reactivity. Finally, a fingerprint for comparing the physical chemical properties for two chemicals is determined based on four properties including lipophilicity (logP), molecular weight (MW), charge, and volatility. Once the fingerprints are defined, they are compared for each SOI/analogue pair of compounds using the Tanimoto algorithm to compute a similarity score (S) as shown below:

$$S = \frac{c}{a + b - c} \tag{1}$$

where a is the number of bits in SOI fingerprint A, b is the number of bits in analogue fingerprint B, and c is the number of bits present in both A and B.[15]

In the study presented here, fingerprints representing biologically relevant similarity attributes are presented, and their use for justifying a read-across prediction is investigated for 14 different case study chemicals and their series of potential analogues used previously in validation of the expert-judgment-based P&G framework.[3] The correlation between structural similarity and suitability for read-across is explored using common molecular fingerprints to identify any molecular features that may be sensitive to analogue ratings. The importance of each nonstructural similarity attribute also is considered to identify which attribute dominates the suitability decision for each class of compounds. Finally, two different weighting scenarios are evaluated for optimum concordance between a weighted sum of nonstructural similarity attributes and analogue suitability.

2. MATERIALS AND METHODS

2.1. Metabolism Similarity. Metabolism is critical for assessing the suitability of an analogue for read-across. Xenobiotic metabolism can produce both nontoxic metabolites which tend to be polar and readily excreted and also reactive metabolites which can interact with macromolecules potentially resulting in toxicity. We capture both the formation of nontoxic metabolites and the formation or presence of species that may bind cellular macromolecules species when we consider similarity in metabolism as the product of a metabolism similarity score and reactive metabolism similarity score. Although the rate and extent of metabolism may differ between an SOI and potential analogues, "suitable" analogues share the same metabolic pathways and form the same reactive metabolites as the SOI. Therefore, similarities of the major metabolic pathways and reactive metabolites are compared for all SOI/analogue pairs.

The metabolic pathways of the SOI and analogues were summarized from literature data if identified or were predicted based on the presence of the functional group in a similar scaffold or substructure frequently supported by data for such compounds with similar structural features. For each SOI and analogue, a metabolism fingerprint was created that consists of key entries corresponding to biotransformations known or predicted to occur for the compound. A starting list of biotransformations used as metabolism fingerprint keys was obtained from the SyGMa (Systematic Generation of potential Metabolites) prediction tool.[9] The output of the SyGMa tool was not used, but rather only the list of biotransformations which was expanded to include those missing from the list. Transformation names were edited in some cases to remove structurally specific references. For example, in some instances it was not necessary to specify the position of hydroxylation of an alkyl chain (e.g., terminal, penultimate, or at another position along the chain) if this transformation represents a minor metabolic event along the pathway.

Because toxicity is frequently associated with the formation of reactive metabolites, it is important to highlight biotransformation keys involved in any activating processes. For each chemical considered, a reactive metabolism fingerprint also was generated to heavily weight and compare reactive metabolite formation. A current list of the biotransformations included in our fingerprints is shown in Table S1A grouped according to phase 1 and phase 2 transformations. Table S1B displays a collection of the biotransformations from the first list along with the name of the reactive species which possesses the potential to react with cellular macromolecules. In some cases, the biotransformation listed is considered detoxifying such as glutathionation of a compound along with the species that may react with cellular macromolecules when glutathione has been depleted. While toxicity associated with this functional group also is highlighted in the corresonding alert similarity, it is important to compare both the potential for transformations that may be considered detoxifying as well as transformations involving reaction with cellular macromolecules when comparing similarity in metabolism. A Tanimoto comparison of the metabolism fingerprints and the reactive metabolism fingerprints was performed for all SOI/analogue pairs to calculate a metabolism similarity score (MSS) and a reactive metabolism similarity score (RMSS). For both calculations, the counts of the biotransformation keys were included so that if a biotransformation was shown or predicted to occur more than once, the counts for the keys were included by modifying the biotransformation name accordingly, e.g., aromatic_hydroxylation, aromatic_hydroxylation-, and aromatic_hydroxylation-. A total metabolism similarity score (TMSS) was then computed as the product of the two metabolism similarity scores as shown below:

$$\text{TMSS} = \text{MSS} \times \text{RMSS} \tag{2}$$

Fingerprints and similarity scores were generated using Pipeline Pilot software (Biovia version 2021). Metabolism fingerprint data for

https://doi.org/10.1021/acs.chemrestox.2c00311
Chem. Res. Toxicol. 2023, 36, 230−242

all 14 case study chemicals are included in the Supporting Information.

2.2. Alert Similarity. Analogous to metabolism similarity, a fingerprint defining alerts for all compounds was generated, and a Tanimoto comparison was performed to compute an alert similarity score for all SOI/analogue pairs. Keys for the alert similarity fingerprint consist of alerts generated from DEREK (Lhasa Nexus 6.1.0) and an internal tool for assessing the potential of a chemical for developmental and reproductive toxicity (P&G DART Decision Tree v. 1.6).

DEREK (Deductive Estimation of Risk from Existing Knowledge) is a software package produced by Lhasa Nexus that predicts toxicological alerts for a chemical using expert knowledge, rules, and references. Structural fragments of a chemical are compared to chemical structures with known toxicity. The knowledge-base covers a range of toxicological end points including carcinogenicity, chromosome damage, mutagenicity, and skin sensitization.[16] Extensive coverage of bacterial mutagenicity and skin sensitization end points permits the software to provide reasoning describing a known absence of toxicity for a compound. For bacterial mutagenicity and skin sensitization, the software provides negative predictions of "inactive" and "non-sensitizer," respectively. In addition, DEREK has several prediction levels, and all chemicals considered in this study were evaluated at the "Plausible" prediction level.

The P&G Automated DART DT (Developmental and Reproductive Toxicity Decision Tree) is a screening tool used to predict DART toxicity by comparing the structure of an input chemical to a virtual library of structures possessing structural features associated with compounds with a known precedent for DART toxicity.[17] Tool output consists of the following five predictions: Maps Positive, Maps Scaffold, No Maps, DART Negative, and Not Covered.

An SD file containing the SOI and analogues was generated for each case study, and the output from DEREK and the DART DT were used as fingerprints to analyze the reactivity of all SOI/analogue pairs. If a DEREK alert was fired for a chemical (SOI or analogue), the reactive fingerprint consisted of the name of the toxicological end point alert concatenated to the alert value. If an alert was not fired for a chemical at the "Plausible" level, the reactive fingerprint consisted of the name of the toxicological end point concatenated to the following values: "null" when no alert was fired, "inactive" (negative for bacterial mutagenicity), or "non-sensitizer" (negative for skin sensitization). The five different predictions from the DART DT provided an additional fingerprint key. If a chemical is classified as Maps Positive, Maps Scaffold, or DART Negative, the category and/or the name of category also was included in the fingerprint. The reactive fingerprints are listed in Table S2.

2.3. Physical Chemical Similarity. Similarity in physical chemical properties is important in demonstrating that an analogue display similar pharmacokinetic profiles to the SOI including similarity in absorption, distribution, metabolism, and excretion (ADME). Absorption is important in describing how the chemical gets into the body, and distribution describes where it will go once inside the body; metabolism describes how the compound breaks down, and excretion describes how the compound leaves the body. Predictions of these properties is critical to the pharmaceutical industry in which poor pharmacokinetic properties often limit the development of new drugs. To obtain an early read on pharmacokinetic properties of potential drugs, the pharmaceutical industry employs many in vitro and in vivo methods. A set of interpretable rules of thumb or guides on structure−property relationships based on the results of ADMET (absorption, distribution, metabolism, excretion, and toxicity) assays were summarized using principal component analysis. In the exercise, it was shown that 81% of the variation in the assays can be described by the top 4 principal components including predicted values for lipophilicity (cLogP), molecular weight (MW), and positive or negative charge.[18] These properties are incorporated into a fingerprint for calculating similarity in physical chemical properties between potential analogues with an SOI.

Lipophilicity is the most important property for determining ADMET behavior, and here the absolute value of the difference in LogP (predicted using ACD Percepta software version 2020) for an analogue and SOI (shown below) is considered for fingerprint entries; these differences are binned according to the ranges listed in Table S3. Bins were chosen based on the absolute value for LogP differences calculated for more than 3000 SOI/analogue pairs for which the analogue was considered "suitable" or "suitable with interpretation" following the P&G framework.[4]

$$\Delta\text{LogP} = |\text{LogP(analog)} - \text{LogP(SOI)}| \quad (3)$$

Keys corresponding to MW bins were established in a manner similar to the bins used by Gleeson.[18] Charge was determined as the charge of the largest fraction of species calculated using the Henderson−Hasselbalch equation.[19] A final property used to determine analogue suitability is volatility. A volatile analogue would not be a suitable read-across analogue for establishing the safety of a nonvolatile SOI. Using the definition of a low-vapor-pressure volatile organic compound (VOC) specified by the California Air Resource Board, a compound is considered a VOC [true (1) or false (0)] if it contains at least one carbon atom and is consistent with one of the following:

(1) vapor pressure <0.1 mmHg

(2) boiling point <216 °C

(3) contains ≤12 carbon atoms

2.4. Total Similarity. A total similarity score (TotalScore) is defined as the following sum of the similarity scores:

$$\text{TotalScore} = \text{total metabolism similarity} + \text{alert similarity} + \text{physical chemical similarity} \quad (4)$$

While weighting the individual components to TotalScore results in improved correlation with analogue rating, we have found that a simple 1:1:1 weighting of total metabolism similarity:alert similarity:physical chemical similarity is sufficient for the limited number of case studies considered here. As we populate our database of similarity scores with more SOI/analogue pairs, the optimum weighting for each component may change. Pipeline pilot software (Biovia, version 2020) was used to calculate all nonstructural similarity scores and TotalScores.

2.5. Data Analysis. *2.5.1. Structural Similarity and Separability Analysis.* Structural similarity scores were computed for comparison with analogue ratings. Structural fingerprint sets were obtained using open-source python libraries (PyFingerprint,[20] and RDKit[21]). A diverse set of structural fingerprints was selected to maximize the chemical information content extracted from the structural similarity data. The fingerprint types comprised the following: (1) circular count-based extended connectivity fingerprints (ECFC4, 2048 substructure keys), (2) circular binary ECFP4 (2048 substructure keys), (3) Molecular ACCess System public keys (MACCS, 166 substructure keys), (4) RDKit (2048 substructure keys), (5) path-based Avalon (512 substructure keys), (6) pharmacophore-based extended reduced-graph (ErG, 315 features), and (7) Saagar structural feature-based fingerprint (834 features—v1).[22] The structural similarity score was computed using the Tanimoto coefficient for all molecular fingerprints, except for the pharmacophore-based ErG fingerprints for which an algebraic form of Tanimoto coefficients was used. The structural fingerprints and their respective similarity scores were computed in a jupyter notebook[23] using python (3.6).[24]

To evaluate and quantify predictivity of analogue rating using all types of similarity scores, a categorical statistical metric was computed called degree of separability, S_d, which represents the effectiveness of a similarity metric in distinguishing analogues considered "not suitable" from those rated either "suitable" or "suitable with interpretation". The criterion for separation per case study is defined with the following expression:

https://doi.org/10.1021/acs.chemrestox.2c00311
Chem. Res. Toxicol. 2023, 36, 230−242

Table 2. Fingerprint keys and similarity scores for case study 1 (benzyl alcohol). Analogue ratings are as follows: SOI = structure of interest, S = suitable, SI = suitable with interpretation, and NS = not suitable

	SOI	589-18-4	89-95-2	536-60-7	100-39-0
Metabolism Keys					
aliphatic_hydroxylation_(primary_carbon)	0	0	0	1	0
primary_alcohol_oxidation_(benzylic)	1	1	1	1	0
glycination_(aromatic_carboxyl)	1	1	1	1	0
O-glucuronidation_(aromatic_carboxyl)	1	1	1	1	0
GSH conjugation	0	0	0	0	1
metabolism similarity score (MSS)	SOI	1	1	0.75	0
Reactive Metabolite/Species Keys					
benzyl_halide	0	0	0	0	1
reactive metabolism similarity score (RMSS)	SOI	1	1	1	0
total metabolism similarity score (MSS) × (RMSS)	SOI	1	1	0.75	0
Alert Keys					
Mutagenicity in vitro bacterium_Alkylating Agent	0	0	0	0	1
Skin sensitization mammal_Benzylic, allylic, or propargylic halide	0	0	0	0	1
Carcinogenicity mammal_Alkylating Agent	0	0	0	0	1
Chromosome damage in vitro _mammal_alkylating agent	0	0	0	0	1
DART_DT	0	0	0	0	0
alert similarity score	SOI	1	1	1	0.11
Physical Chemical Property Keys					
DeltaLogPbin 1 ($0 \leq \Delta\text{LogP} < 2$)	1	1	1	1	1
MWbin2 ($100 \leq \text{MW} < 300$)	1	1	1	1	1
charge—neutral	1	1	1	1	1
VOC	0	0	0	0	1
Physchem similarity score	SOI	1	1	1	0.6
total similarity score	SOI	3	3	2.75	0.7
analogue rating	SOI	S	S	SI	NS

$$S_d = \frac{1}{N_{S+SI} + N_{NS}} \left(\sum_{i=1}^{N_{S+SI}} [\text{score}_i > \max(\text{score}_{\{NS\}})] + \sum_{j=1}^{N_{NS}} [\text{score}_j < \min(\text{score}_{\{S,SI\}})] \right) \quad (5)$$

where score_i refers to a similarity score of an analogue in either the "suitable" or "suitable with interpretation" rating groups, and score_j refers to a similarity score of a "not suitable" analogue. $\max(\text{score}_{\{NS\}})$ is the maximum score for all analogues rated "not suitable", and $\min(\text{score}_{\{S,SI\}})$ is the minimum similarity score for the "suitable/suitable with interpretation" analogue bin. N_{NS} is the total number of "not suitable" analogues, and N_{S+SI} is the total number of "suitable/suitable with interpretation" analogues. S_d for each approach is calculated as the fraction of analogues that satisfy the above condition. For example, for an SOI with 10 analogues, where 9 analogues can be distinguished with a similarity score and 1 analogue has the same score as an analogue from a different analogue group (borderline), the degree of separability would be 90%. Separability analyses were performed using R (4.1.0)[25] and python (3.6)[23] in an Rmarkdown notebook[26] and jupyter notebook,[22] respectively.

2.5.2. Machine Learning (ML) Modeling. Binary classification models were built using python package Scikit-learn[7] to estimate the relative feature importance or coefficients of the similarity components contributing to the qSIM (quantitative similarity) score (eq 6) for predicting the analogue ratings shown in Figure 4 where two classes of ratings are considered (class 1 = S or SI versus class 2 = NS). A classification-based machine learning pipeline was constructed to explore how well the different machine learning models predicted the expert-judgment-based analogue ratings using the similarity components. Multinomial logistic regression was found to be the best-performing model for predicting the analogue rating class, with feature importance computed using Scikit-learn built-in libraries.[27] The inputs of the machine learning models are the three qSIM components, and the output of the model is a two-class categorical prediction (class 1 = S or SI versus class 2 = NS).

Weighted averages of the similarity components were computed using the ML-derived coefficients (a, b, c) in eq 6. The coefficients represent normalized feature importance values. The weighted averages were multiplied by 3, the maximum qSIM score, to scale the data for comparing different forms of the qSIM score. The computed rescaled weighted averages were compared against the base qSIM score consisting of a linear sum of the components where $a = b = c = 1$, to identify coefficients that best fit the expert-judgment-based analogue rating assignments. The rescaled weighted average scores can be expressed with the following equation:

$$\text{qSIM} = \frac{a \times \text{AlertSim} + b \times \text{PhysChemSim} + c \times \text{TotalMetabSim}}{a + b + c} \times 3 \quad (6)$$

where AlertSim (alert similarity), PhysChemSim (physical chemical similarity), and TotalMetabSim (total metabolism similarity) are the similarity components, and a, b, and c are the component coefficients, respectively. TotalScore (eq 4) corresponds to the base score (eq 6) with equal weighting of the individual similarity scores ($a = b = c = 1$). The derived component coefficients from machine learning modeling optimize the qSIM score to improve the separability of the qSIM score between the different analogue rating groups.

3. RESULTS

For all case study chemicals, analogues requiring metabolism to justify their use in read-across are excluded from the studies presented here. These analogues are labeled "suitable with precondition" and require metabolism data and extrapolation to in vivo concentrations to justify their use.

3.1. Case Study 1—Benzyl Alcohol (CAS 100-51-6). Relevant fingerprint keys and similarity scores are presented in Table 2 for case study 1 which is benzyl alcohol (CAS 100-51-

https://doi.org/10.1021/acs.chemrestox.2c00311
Chem. Res. Toxicol. 2023, 36, 230−242

6). The metabolism of benzyl alcohol (SOI) is well-known in humans where it has been shown that the alcohol is rapidly oxidized to benzaldehyde and then to benzoic acid which undergoes phase 2 conjugation with glycine to form hippuric acid or with glucuronic acid to form the O-glucuronide.[28] The metabolism fingerprint for benzyl alcohol contains 3 active biotransformation keys representing oxidation of the alcohol to the acid (aldehyde oxidation is not considered as a separate step) and the phase 2 transformations glycination and O-glucuronidation of the benzoic acid group with no reactive metabolism transformations. Analogues *p*-methylbenzyl alcohol (CAS 589-18-4) and *o*-methylbenzyl alcohol (CAS 89-95-2) possess benzoic acid scaffolds with an extra methyl group at the p- or o-positions. These analogues are known metabolites of *p*- and *o*-xylene, respectively, and it has been demonstrated that the alcohol group undergoes oxidation followed by phase 2 conjugation with no biotransformation of the methyl groups.[29] Therefore, these two analogues possess the same nonzero biotransformation keys as the SOI with no reactive metabolite keys identified resulting in a TMSS of 1. Analogue 4-isopropylbenzyl alcohol (CAS 536-60-7) has an additional isopropyl group at the p-position, which would yield hydroxylation of the isopropyl group as an additional metabolic pathway. Alcohol oxidation, glucuronidation, and glycine conjugation of CAS 536-60-7 would still occur; the MSS of this analogue is 0.75, and in the absence of reactive metabolite formation the TMSS is 0.75. Analogue benzyl bromide (CAS 100-39-0) is an electrophilic benzyl halide with glutathione conjugation being the major metabolic transformation.[30] No common metabolic pathways can be found between the SOI and this analogue; therefore, the MSS = 0. Because benzyl bromide itself is a reactive species exemplified by the reactive metabolite/species key for a benzyl halide, the RMSS = 0 for this analogue resulting in a TMSS = 0.

Benzyl alcohol and the three alkyl derivatives fire no alerts in DEREK or the DART DT resulting in a reactivity similarity score of 1.0 for these SOI/analogue pairs. Benzyl bromide, on the other hand, fires four structural alerts in DEREK due to the key functional group alkyl bromide which produces DEREK alerts for carcinogenicity, chromosome damage, and mutagenicity in vitro in mammals in addition to alerts for eye irritation, respiratory track irritation, and skin sensitization. As such, the alert similarity score for this SOI/analogue pair is 0.11.

Benzyl alcohol and the four analogues possess similar physical chemical properties and all fall into the same ΔLogP bin and MW bin and are neutral in charge. Volatility is the only key for which differences exist since the SOI and its alkyl derivatives are not considered volatile; however, the benzyl bromide has a much higher predicted vapor pressure resulting in a positive VOC classification. The similarity score for physical chemical properties is then equal to 1 for the three analogues that are alkyl derivatives of the SOI while benzyl bromide results in a PhysChem similarity score of 0.6 with the SOI.

A sum of all similarity scores results in a TotalScore for each analogue shown in the second to last row of Table 2 and correlates well with the expert-judgment-based analogue rating assigned for these compounds and listed in the last row of the table. The correlation with analogue rating is more evident in the stacked bar plot for case study 1, shown in Figure 1, where the two "suitable" analogues receive a TotalScore of 3; the "suitable with interpretation" analogue has a score of 2.75

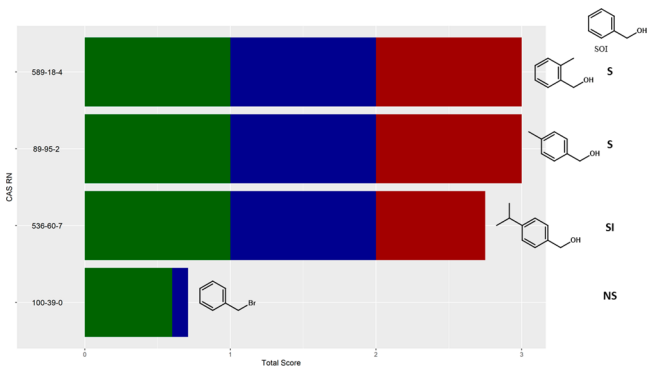


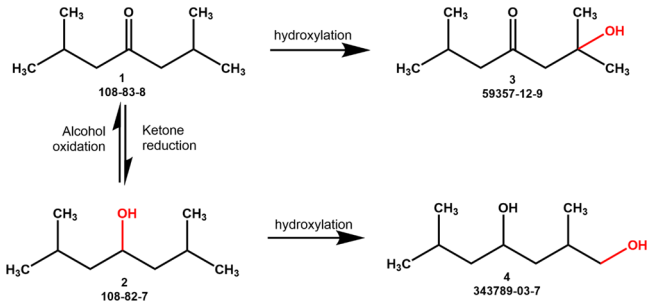
Figure 1. Horizontal stacked bar plot of the TotalScore for case study 1 (benzyl alcohol). Each bar is the sum of the following components: total metabolism similarity (red), alert similarity (blue), and physical chemical similarity (green). Analogue ratings correspond to S (suitable), SI (suitable with interpretation), and NS (not suitable).

while the TotalScore for "not suitable" analogue is much lower at 0.7.

3.2. Case Study 3—Diisobutyl Ketone (CAS 108-83-8).

The phase I metabolism of diisobutyl ketone (SOI) was studied in the rat, is shown in Scheme 1, and includes

Scheme 1. Metabolic Pathways for Case Study 3, Diisobutyl Ketone (CAS 108-83-8)



reduction of the carbonyl group (1) to the alcohol (2), oxidation of the alcohol group back to the carbonyl (1), and hydroxylation of the alkyl chain at the primary (ω-position, 4) or tertiary carbon (ω-1-position, 3).[31] Fingerprints for metabolism, structural alerts, and physical chemical properties are summarized in Table 3 along with similarity scores which are shown graphically in Figure 2 with chemical structures. Analogue diisoamyl ketone (CAS 2050-99-9) contains the same structural features and has the same metabolism fingerprint as the SOI; therefore, the TMSS for this analogue is 1. Analogue 2,6,8-trimethyl-4-nonanone (CAS 123-18-2) possesses an extra branched methyl group at the 6-position, which may lead to an additional hydroxylation biotransformation key resulting in an MSS of 0.8. Compounds 3-ocatanone (CAS 106-68-3), 3-nonanone (CAS 925-78-0), 4-heptanone (CAS 123-19-3), 3-hexanone (CAS 589-38-8), 6-undecanone (CAS 927-49-1), and 7-tridecanone (CAS 462-18-0) are straight-chain monoketones. Because of the absence of chain branching, the ω-1 hydroxylation of these analogues occurs at a secondary carbon (vs a tertiary carbon in the SOI), while the other three metabolic pathways match those for the SOI. The MSS for each of these analogues is 0.6, and because no reactive metabolites will form for these analogues, the RMSS for these analogues is 1 resulting in a TMSS = 0.6.

https://doi.org/10.1021/acs.chemrestox.2c00311
Chem. Res. Toxicol. 2023, 36, 230−242

Table 3. Fingerprint Keys and Similarity Scores for Case Study 3 (Diisobutyl Ketone)[a]

	SOT	205 0-99-9	123-18-2	106-68 3	925-78-0	123-19-3	589-38-8	448 5-09-0	927-49-1	96-22-0	462-18-0	106-35-4	541-85-5	123-54-6	504-20-1	78-85-3
Metabolism Keys																
carbonyl_reduction_(aliphatic)	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	0
aliphatic_hydroxylation_primary_carbon	1	1	2	1	1	1	1	1	1	0	1	1	2	0	1	0
aliphatic_hydroxylation_(tert_carbon)	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
aliphatic_hydroxylation_(sec_carbon)	0	0	0	1	1	1	1	1	1	0	1	1	1	0	0	0
secondaiy_alcohol_oxidation_(aliphatic)	1	1	1	1	1	1	1	1	1	1	1	1	1	2	0	0
aldehyde_oxidation_(aliphatic)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
O-glucuronidation_ (aromatic_carboxyl)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
glutathionation_alphabetaunsat_carbonyl	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1
oxidation to gamma-diketone	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
metabolism similarity score (MSS)	SOI	1	0.80	0.60	0.60	0.60	0.60	0.60	0.60	0.50	0.60	0.50	0.43	0.33	0.33	0
Reactive Metabolite/Species Keys																
gamma-diketonc	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
alpha-beta unsaturated carbonyl	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
reactive metabolism similarity score (RMSS)	SOI	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0
(TMSS) = MSS × RMSS	SOI	1	0.80	0.60	0.60	0.60	0.60	0.60	0.60	0.50	0.60	0	0	0.33	0	0
	SOI	2050-99-9	123-18-2	106-68-3	925-78-0	123-19-3	589-38-8	4485-09-0	927-49-1	96-22-0	462-18-0	106-35-4	541-85-5	123-54-6	504-20-1	78-85-3
Alert Keys																
Mutagenicity in vitro bacterium_alpha, beta-Unsaturated aldehyde or precursor	0	0	0	0	a	0	0	0	0	0	0	0	0	0	0	1
Skin sensitization mammal Tautomer 1_1,3-Diketone	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Skin sensitization mammal alpha,beta-Unsaturated aldehyde or precursor	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Neurotoxicity mammal_garnma- Diketone or precursor	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
Chromosome damage in vitro mammal_alpha,beta-Unsaturated ketone	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
PG DART Decision Tree_Positive	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
PG DART Decision Tree_Scaffold	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
AlertSim	SOI	1	1	1	1	1	1	1	1	1	1	0.71	0.71	0.71	0.5	0.33
	SOI	2050-99-9	123-18-2	106-68-3	925-78-0	123-19-3	589-38-8	4485-09-0	927-49-1	96-22-0	462-18-0	106-35-4	541-85-5	123-54-6	504-20-1	78-85-3
Physical Chemical Property Keys																
DeltaLogPbin 1 (0 ≤ ALogP < 2)	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0
DeltaLogPbin 2 (2 ≤ ALogP < 3)	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1
MWbinl (MW < 100)	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
MWbin2 (100 ≤ MW < 300)	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0
charge—neutral	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
VOC	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1
physical chemical similarity	SOI	1	1	1	1	1	1	1	0.60	0.60	0.33	1	1	0.60	1	0.33
TotalScore	SOI	3	2.80	2.60	2.60	2.60	2.60	2.60	2.20	2.10	1.93	1.71	1.71	1.64	1.50	0.66
analogue rating	SOI	S	SI	SI	SI	SI	SI	SI	SI	NS	NS	NS	NS	NS	NS	NS

[a]Analogue ratings are as follows: SOI = structure of interest, S = suitable, SI = suitable with interpretation, and NS = not suitable.

https://doi.org/10.1021/acs.chemrestox.2c00311
Chem. Res. Toxicol. 2023, 36, 230−242

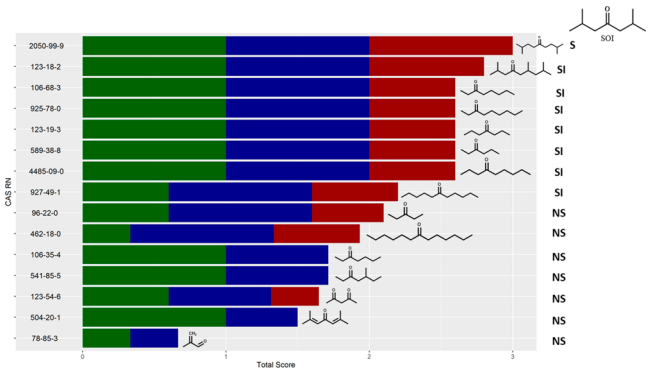


Figure 2. Horizontal stacked bar plot of the TotalScore for case study 3 (diisobutyl ketone). Each bar is the sum of the following components: total metabolism similarity (red), alert similarity (blue), and physical chemical similarity (green). Analogue ratings correspond to S (suitable), SI (suitable with interpretation), and NS (not suitable).

Analogue diethyl ketone (CAS 96-22-0) is a short-chain monoketone for which alkyl chain hydroxylation is unlikely to occur as observed in the rat resulting in an MSS and TMSS of 0.5.[32] Analogue acetylacetone (CAS 123-54-6) contains two carbonyl groups, which would have a count of 2 for the carbonyl reduction and secondary alcohol oxidation biotransformation keys. Because of the shorter alkyl chain for this analogue, alkyl chain hydroxylation is not likely to occur resulting in an MSS and TMSS of 0.33.[32]

Analogues 3-heptanone (CAS 106-35-4) and 5-methyl-3-heptanone (CAS 541-85-5) are linear and methyl-branched ketones, respectively, and because of the position of the carbonyl group, ω-1 hydroxylation followed by alcohol oxidation would lead to the formation of a γ-diketone reactive metabolite that may cause neurotoxicity.[33] The MSSs for 3-heptanone (CAS 106-35-4) and 5-methyl-3-heptanone (CAS 541-85-5) are 0.5 and 0.43, respectively. Due to the reactive metabolite formation, the RMSSs are 0 for these two analogues. Analogues methacrolein (CAS 78-85-3) and diisobutenyl ketone (CAS 504-20-1) contain one or more α,β-unsaturated ketone group resulting in different metabolic pathways from the monoketone, but it is also a reactive functional group. Therefore, analogues methacrolein and diisobutenyl ketone result in an MSS of 0 and 0.33, respectively, but with RMSSs = 0, the resulting total metabolism similarity TMSS is 0.

The alert fingerprint and similarity scores for analogues of diisobutyl ketone (CAS 108-83-8) also are shown in Table 3. For this structural class the structural alerts are reflective of reactive metabolite formation with no alerts found for the first 10 analogues consisting of straight-chain or branched alkyl ketones that form no reactive metabolites. The two analogues, 3-heptanone (CAS RN 106-35-4) and 5-methyl-3-heptanone (CAS RN 541-85-5), are γ-diketone precursors and fire a DEREK alert for neurotoxicity leading to an alert similarity score of 0.71. Acetylacetone (CAS RN 123-54-6) also has an alert similarity score of 0.71, but this results from the presence of two ketone groups (diketone) instead of two alkyl groups (dialkyl ketone) resulting in a skin sensitization alert in DEREK. Diisobutenyl ketone (CAS RN 504-20-1) is an α,β-unsaturated dialkyl ketone resulting in a DEREK alert for chromosome damage. In addition, it fires a Scaffold Match alert in the P&G DART DT. All together this leads to an alert similarity score of 0.50. Methacrolein is an α,β-unsaturated aldehyde and results in the lowest alert similarity score of 0.33. The structure fires DEREK alerts for both mutagenicity and skin sensitization along with a positive alert in the P&G DART DT.

Table 3 and Figure 2 show that most analogues for this case study are similar in physical chemical properties to the SOI. 10 out of the 15 analogues have a physical chemical similarity score of 1 with the SOI because all compounds possess similar alkyl chain length, occupy the lowest ΔLogP bin and the same MW bin as the SOI, and are all classified as VOCs. The longer alkyl chain length for 6-undecanone (CAS 927-49-1) results in a negative VOC classification and a physical chemical similarity of 0.60. The even longer alkyl chain length for 7-tridecanone (CAS 462-18-0) results in a logP that is greater than 2 units away from that for the SOI in addition to a negative VOC classification producing a physical chemical similarity score of 0.33. Diethyl ketone (CAS 96-22-0) falls into a lower MW bin than the SOI and receives a physical chemical similarity score of 0.60. Acetylacetone (CAS RN 123-54-6) has a much lower LogP than the SOI and falls into ΔLogP bin 2 resulting in a physical chemical similarity score of 0.60, and a much lower LogP and MW for methacrolein produce a physical chemical property similarity of 0.33.

As shown in Figure 2, the suitable analogue receives the highest TotalScore which is equal to 3 corresponding to similarity scores of 1 for each of the nonstructural attributes. All suitable with interpretation analogues have TotalScores less than 3 resulting from small changes in metabolism as reflected by TMSSs < 1 and TotalScores ranging from 2.2 to 2.8. 6-Undecanone (CAS 927-49-1) is the "suitable with interpretation" analogue with a TotalScore of 2.2 resulting from changes in physical chemical properties from the longer alkyl chain length as well as small differences in metabolism. It is interesting to note that diethyl ketone (CAS 96-22-0) receives a rating of "not suitable" with a similarity score of 2.1 resulting from differences in metabolism and physical chemical properties. The remaining "not suitable" analogues display TotalScores less than 2 resulting from large differences in metabolism due to the formation of reactive metabolites also captured in the alert similarity formation for 3-heptanone (CAS 106-35-4), 5-methyl-3-heptanone (CAS 541-85-5), diisobutenyl ketone (CAS 504-20-1), and methacrolein (CAS 78-85-3). Meanwhile, the much longer alkyl chain for 7-tridecanone (CAS 462-18-0) results in a much lower physical chemical similarity score. Finally, acetyl acetone (CAS 123-54-6) is a diketone, and its low TotalScore results from low metabolism similarity and alert similarity.

3.3. Case Study 10—4-Nitrotoluene (CAS 99-99-0). 4-Nitrotoluene (CAS 99-99-0) is the SOI for case study 10 and is an interesting case of a small alkyl nitrobenzene. Computed TotalScores with analogue structures are displayed in Figure 3 (see the Supporting Information for contributing fingerprints and similarity scores), and metabolism similarity is the most important attribute contributing to the TotalScore for the first 8 analogues. 4-Nitrotoluene (CAS 99-99-0) is extensively metabolized in the rat producing major metabolites consistent with oxidation of the methyl group to produce 4-nitrobenzoic acid (CAS 62-23-7) and reduction of the nitro group of the 4-nitrobenzoic acid (CAS 62-23-7) metabolite followed by N-acetylation to form *p*-acetaminobenzoic acid.[34] What is most important about the metabolism of this compound is that oxidation of the methyl group is a primary biotransformation

https://doi.org/10.1021/acs.chemrestox.2c00311
Chem. Res. Toxicol. 2023, 36, 230−242

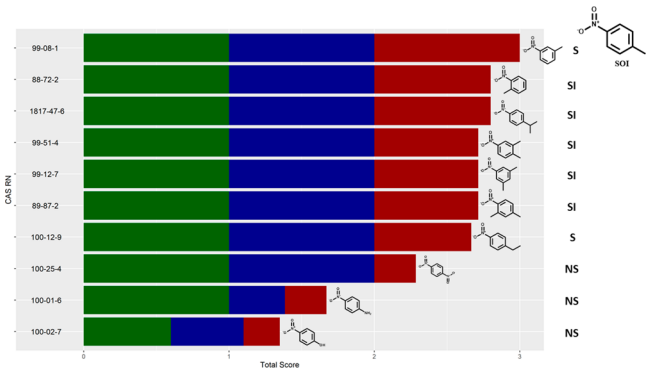


Figure 3. Horizontal stacked bar plot of the TotalScore for case study 10 (4-nitrotoluene). Each bar is the sum of the following components: total metabolism similarity (red), alert similarity (blue), and physical chemical similarity (green). Analogue ratings correspond to S (suitable), SI (suitable with interpretation), and NS (not suitable).

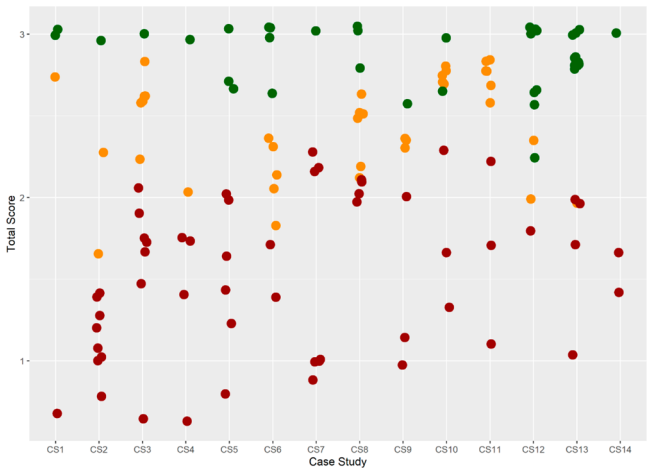


Figure 4. Total similarity score (TotalScore) versus case study analogue for all case studies with slight jittering to view overlapping points. Marker color corresponds to analogue rating as follows: green (suitable), orange (suitable with interpretation), and red (not suitable).

pathway in contrast to the methyl derivatives of benzyl alcohol in case study 1 where no metabolism was observed for these substituents. Because the methyl group of the nitrotoluenes undergoes hydroxylation followed by oxidation as well as reduction of the nitro group followed by acetylation, the metabolic pathway for 3-nitrotoluene is similar to that of the SOI. For 2-nitrotoluene, the primary pathway involves oxidation of the methyl group followed by benzyl alcohol oxidation to form 2-nitrobenzoic acid; however, the proximity of the acid group to the nitro group results in a much lower extent of nitro group reduction to aniline for 2-nitrotoluene relative to 4-nitrotoluene with no N-acetylation of aniline. Biotransformation trends observed for the nitrotoluenes are extended to the dimethyl nitrobenzene derivatives, and slightly lower metabolism similarity scores of 0.71 result because similarity is calculated on a counts basis with two counts for benzylic hydroxylation and carboxylic acid formation. The "suitable" assignment for 4-ethylbenzene reflects "rules of thumb" that have evolved as the P&G framework has been applied; however, the TotalScore is only 2.67 because of the difference in metabolism similarity for biotransformation of the ethyl group relative to the methyl group of the SOI. Replacement of the methyl group of 4-nitrotoluene with an amine 4-nitroaniline (CAS 100-01-6), a second nitro group 1,4-dinitrobenzene (CAS 100-25-4), and 4-nitrophenol (CAS 100-02-7) results in significantly lower similarity scores for metabolism and alerts similarity due to the different biotransformation pathways and structural alerts for these compounds relative to the SOI.

3.4. Results for All Case Studies. TotalScores for all SOI/analogue pairs for the 14 case study chemicals are presented in Figure 4 with analogue rating assignments (see the Supporting Information for chemical structures, fingerprints, and similarity scores for metabolism, alerts, and physical chemical properties). Analogue rating assignments reflect a combination of assignments discussed in Blackburn et al., 2011, but with updates reflecting changes warranted by closer scrutiny of each analogue, new metabolism and alerts data, as well as evolution and standardization of the approach over the past 10 years of implementation.[3] For all case studies, analogues rated "suitable" display the highest TotalScore with values ranging from 2.3 to 3.0. Analogues rated as "suitable with interpretation" range in TotalScore from 1.7 to 2.8 with a small number of analogues overlapping others rated both "suitable" and "not suitable". Analogues with a TotalScore <1.5 were rated as "not suitable", and most with a score between 1.5 and 2.0 are considered "not suitable" with some overlap with "suitable with interpretation" analogues. In general, for every case study, there is a consistent progression of TotalScore values from "not suitable" to "suitable with interpretation" to "suitable" analogues. Case study 10 (SOI 4-nitrotoluene CAS 99-99-0) is an interesting example with a "suitable" analogue with a TotalScore of 2.7 (4-ethyl-nitrobenzene CAS 100-12-9). This rating is based on internal "rules of thumb" which have evolved over the past 12 years of applying the framework where the change in alkyl chain length for the alkyl aromatic derivative results in little change in toxicity for most compounds of this type. The SOI for case study 12 is lauryl alcohol which is a fatty alcohol containing 12 carbon atoms. The ratings for this case study are based on the template for rating fatty acids/fatty alcohols in Lester et al., 2018, and are supported by the similarity scores presented here.[2] As shown in Table 1, case study chemicals represent a 50:50 mix of aliphatic and aromatic compounds. Except for case study 13, for which there are analogues rated "suitable with interpretation" with the same scores as those considered "not suitable", TotalScore can be used to distinguish "not suitable" analogues from those considered "suitable" or "suitable with interpretation" defined as separability which is explored to a greater extent in Figure 5.

3.5. Results from ML Modeling Data Analysis. Results from separability analysis of the chemical similarity data per case study are summarized in Figure 5. The degree of separability, S_d, ranging from 0 to 1, has been computed for structural similarity (7 descriptor/fingerprint sets), for individual similarity components (e.g., alert similarity, physical chemical similarity, total metabolism similarity), and for results from fitting the qSIM score (eq 6) to the expert-judgment-based analogue rating scores. S_d reflects the effectiveness of similarity scores for distinguishing "not suitable" analogues from all other analogues. For the structural similarity scores, the fraction of case studies with complete separation (S_d = 1) ranged from 3/14 (ECFP4, ECFC4, and Avalon) to 7/14 (MACCS). For the individual component similarity scores, the fraction of case studies with complete separation ranged from

https://doi.org/10.1021/acs.chemrestox.2c00311
Chem. Res. Toxicol. 2023, 36, 230−242

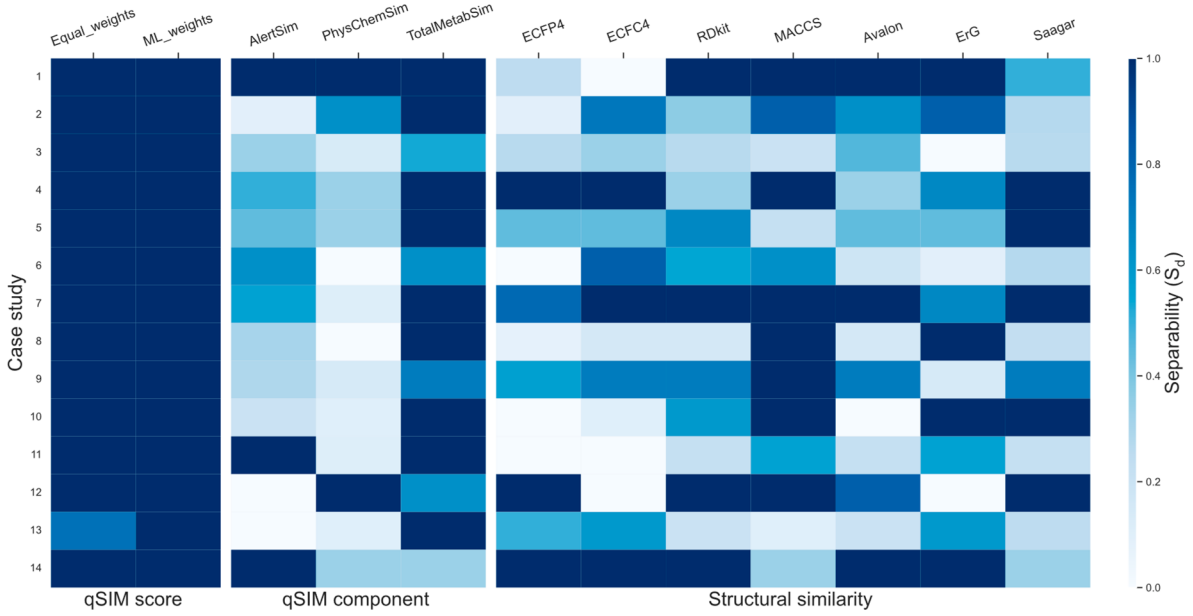


Figure 5. Heatmap of degree of separability (S_d) calculated for qSIM scores, individual similarity components, and structural similarities. Each row represents data per case study as labeled on the vertical axis. Each column is a separability score as labeled on the horizontal axis. Separability analysis was performed using two bins for analogue rating (not suitable and suitable/suitable with interpretation). Equal_weights correspond to a qSIM score (eq 6) calculated with $a = b = c = 1$. ML_weights correspond to eq 6 with $a = 0.98$, $b = 0.85$, and $c = 2.94$ for alert similarity, physical chemical similarity, and total metabolism similarity, respectively.

3/14 (alert similarity) to 2/14 (physical chemical similarity) to 9/14 (total metabolism similarity) demonstrating that metabolism similarity is the most important attribute contributing to the analogue rating for these 14 case studies. All components, however, are important for the analogue ratings determined for case study 1; both metabolism and alert similarity are important for case study 11; physical chemical properties dominated the prediction for case study 12; and alert similarity is most important for case study 14.

The linear sum of the similarity components (qSIM score with equal weights) or with weightings revealed from machine learning exhibits a significantly higher fraction of analogues displaying complete separability of the two analogue rating bins: 13/14 and 14/14, respectively. The weights identified from the classification modeling are $a = 0.98$, $b = 0.85$, and $c = 2.94$ for alert similarity, physical chemical similarity, and total metabolism similarity, respectively. The calculation reveals that metabolism similarity is approximately 3 times more important than alert and physical chemical similarities for predicting analogue rating.

4. DISCUSSION

Comparison of the similarity between an SOI and potential analogue is required for justifying the use of the analogue for read-across. Structural similarity using Tanimoto comparisons of molecular fingerprints is not always consistent with suitability for read-across and strongly depends on the structural class and molecular fingerprint selected. In this study, we demonstrate that it is more important to consider and quantify similarity in biologically relevant attributes including metabolism, reactivity, and physical chemical properties. Methods for calculating similarity in these biologically relevant attributes involve a detailed tabulation of the keys defining the fingerprints for each attribute promoting a systematic process that will increase transparency and consistency among experts.

Summing the component similarity scores produces a TotalScore that can be used to support the analogue rating assignment. In all case studies shown in Figure 4, a TotalScore of 3 is consistent with a "suitable" analogue and easily justified, because the analogue is expected to display the same metabolic pathway, toxicity alerts, and physical chemical properties as the SOI. If the TotalScore < 1.5, justification of the use of the analogue for read-across would be difficult, and the analogue would receive a rating of "not suitable". There is overlap among analogues receiving TotalScores between 1.5 and 3, but except for case studies 8 and 13, all "not suitable" analogues possess a lower TotalScore from the "suitable" and "suitable with interpretation" analogues. These TotalScores facilitate automation of the rating process, and as we apply the methods to more SOI/analogue pairs of structures in our extensive database, we will determine a more accurate correlation between TotalScore and rating. As implied by the calculation of a separability score, the most important distinction in rating analogues for read-across is distinguishing between those considered "not suitable" from those considered "suitable" or "suitable with interpretation". In fact, most of the ambiguity among chemists rating analogues on our team occurs in distinguishing "suitable" analogues from those considered "suitable with interpretation".

Quantification of similarity for each attribute transparently reveals the origin of differences between the SOI and analogue as shown in Figures 1, 3, and 4 for 3 of the case study compounds. A low similarity score for metabolism exposes differences in the biotransformation pathways between the two compounds, or the reactive metabolism similarity uncovers reactive metabolite formation that may occur for one compound but not the other. A low similarity in physical chemical properties is indicative of differences in bioavailability, and differences in alert similarity are related to differences in reactivity. The technique is particularly useful when you have a series (or category) of analogues, and you

https://doi.org/10.1021/acs.chemrestox.2c00311
Chem. Res. Toxicol. 2023, 36, 230−242

need to differentiate among them. This may occur if there are toxicological data for the same end point for more than one analogue. In this case, the TotalScore is indicative of the quality of an analogue and could be used to justify the use of data from an analogue with a higher TotalScore from those for a second analogue with the lower score, thus, streamlining and simplifying the read-across process.

Analysis of the separability for the 14 case study chemicals (Figure 5) is particularly informative demonstrating the importance of nonstructural similarity for demonstrating analogue suitability and providing additional evidence supporting the insufficiency of structural similarity alone for predicting analogue suitability. While some structural similarity scores discriminate between "not suitable" and "suitable/suitable with interpretation" analogues, no one structural fingerprint distinguished between the two bins for all case studies. Molecular fingerprints resulting in separability $S_d = 1$ for most case studies include the MACCS (7 case studies) and Saagar (5 case studies) fingerprints. The relatively high performance for MACCS fingerprints may result from the fact that the case study chemicals are all relatively small molecules that tend to be well characterized by the small structural fragments in this fingerprint of only 166 structural keys. Saagar molecular fingerprints were developed to cover both small and large molecules for building interpretable models or for read-across and have been shown to provide higher scaffold similarities among analogues selected based on activity toward the aromatic hydrocarbon receptor.[22]

Analysis of the separability of the individual similarity components for metabolism, alerts, and physical chemical properties shown in Figure 5 demonstrates that total metabolism similarity is the most important attribute for predicting suitability for most case studies. For case study 1 there are only 4 analogues, and all three similarity components are shown to contribute to the separability with contributions presented in Figure 1. Alert similarity and metabolism similarity are important for case study 6 for which the SOI is dimethyl terephthalate (CAS 120-61-6). For this case study, the alert similarity reflects the potential for DART toxicity associated with dialkyl orthophthalates, and total metabolism similarity reflects differences in metabolism associated with analogues possessing alcohol groups with longer straight chain or branched alkyl chains (Supporting Information). Because the SOI for case study 11 is 2-nitroaniline (CAS 88-74-4), alert similarity and total metabolism similarity both produce low similarity scores for analogues lacking the aromatic amine group. Physical chemical similarity is the most important attribute for determining analogue suitability for case study 12 for which the SOI is lauryl alcohol, and the analogues consist of straight-chain alcohols with chain lengths raging from C5 to C22 which populate several different ΔLogP and MW bins. Case study 14 corresponds to 4,4′-diamino-2,2′-stilbenedisulfonic acid (CAS 81-11-8) with only three analogues selected. Alert similarity is most important for this class of compounds because the presence of the sulfonate groups for the SOI removes the alerts present for not suitable analogue 4,4′-diaminostilbene (CAS 621-96-5), and the presence of a different scaffold for not suitable analogue 1,1′-benzidinedisulfonic acid (CAS 117-61-3) results in a different DART alert from that for the SOI.

Figure 5 displays the calculated separability when the qSIM score (eq 6) is computed using an equal weighting of each of the similarity components $a = b = c = 1$ or $a = 0.98$, $b = 0.85$, and $c = 2.94$ which are derived from machine learning. Except for case study 13, equal weighting of each similarity component performs well providing accurate prediction of analogue rating with a separability ($S_d = 1$) for 13 out of 14 case studies while the coefficients from the ML treatment of the data performed slightly better ($S_d = 1$) for 14 out of 14 case studies. Since the ML weights were derived from the limited number of case studies considered here, we will continue to use an equal weighting of the similarity components in eq 6 resulting in calculation of the TotalScore (eq 5). However, the ML-derived coefficients illustrate the importance of metabolism for determining analogue suitability where metabolism similarity was found to be approximately 3 times more important than alert or physical chemical similarities. Selection of fingerprint keys for metabolism and reactive metabolism requires searching the literature for metabolism data for the SOI and analogues or for data on compounds possessing the same functional groups in similar scaffolds when metabolism data are missing to identify relevant biotransformation keys. Calculation of such a metabolism similarity score is not new; however, some implementations rely on in silico prediction of pathways which often select all possible transformations for the structural features present, thus diluting the biological relevance of the attribute.[8] Another approach for assessing similarity in metabolism controls the propagation of in silico predicted metabolites with metabolic data and expert judgment and involves a side-by-side comparison of metabolic maps for SOI/analogue pairs of structures.[11]

Next steps for this work involve the development of a database to house the fingerprints for each compound considered. Metabolism is the most labor-intensive similarity metric involving some expert judgment and literature review to select the most relevant biotransformation keys. A metabolism database has been developed and will be populated with metabolism fingerprints for all compounds in the repository so the metabolism similarity of any pair of compounds may be considered. As the approach is applied to more internal SAR-based toxicological assessments, a database of TotalScores with analogue ratings will be compiled providing a large data set for modeling permitting the reduction of the reliance on expert judgment for the analogue rating process. The approach also neglects to account for differences in the rate and extent of biotransformations. Determining how best to include these data into a metabolism fingerprint also will be explored.

While we share the fingerprints and similarity scores generated for all 14 case study chemicals in the Supporting Information, we are automating this quantitative similarity approach using open-source programming languages and plan to include these tools in later publications.

5. CONCLUSIONS

These data contribute to evidence supporting the insufficiency of structural similarity for justifying a read-across prediction demonstrating that biologically and toxicologically relevant information must be included.[5,35] It is unfortunate that many of the tools for assisting a read-across assessment select analogues using only structural similarity potentially missing many suitable analogue data sources and choosing others that would be considered not suitable. As the P&G read-across framework has evolved, analogue selection relies predominantly on the use of an MMP approach followed by justification of suitability by assessing similarity in metabolism, reactivity (alerts), and physical chemical properties.[4] In this

https://doi.org/10.1021/acs.chemrestox.2c00311
Chem. Res. Toxicol. 2023, 36, 230−242

paper, a systematic process for defining fingerprints for metabolism, reactivity, and physical chemical properties is presented. Incorporation of these biologically and toxicologically relevant comparisons into a total similarity score that is concordant with suitability for a read-across prediction greatly enhances transparency and increases consistency among experts. The information provided will facilitate implementation by others ultimately increasing chances for regulatory acceptance.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.chemrestox.2c00311.

Tables listing fingerprint keys for computing metabolism, alert, and physical chemical similarity (PDF)

Tables listing fingerprints and similarity scores for all case study chemicals (XLS)

■ AUTHOR INFORMATION

Corresponding Author

Cathy Lester − *The Procter & Gamble Company, Mason, Ohio 45040, United States;* orcid.org/0000-0001-5695-0986; Email: lester.cc@pg.com

Authors

ElLantae Byrd − *The Procter & Gamble Company, Mason, Ohio 45040, United States*

Mahmoud Shobair − *The Procter & Gamble Company, Mason, Ohio 45040, United States*

Gang Yan − *The Procter & Gamble Company, Mason, Ohio 45040, United States;* orcid.org/0000-0001-5648-7653

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chemrestox.2c00311

Author Contributions

The manuscript was written with contributions from all authors. All authors have given approval to the final version of the manuscript. CRediT: **Cathy Lester** conceptualization, formal analysis, methodology, software, writing-original draft, writing-review & editing; **ElLantae Byrd** data curation, methodology, writing-original draft, writing-review & editing; **Mahmoud Shobair** software, visualization, writing-original draft, writing-review & editing; **Gang Yan** data curation, methodology, writing-original draft, writing-review & editing.

Funding

This work was funded by Procter & Gamble, by whom the authors are employed.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We gratefully acknowledge Xiaoling Zhang and Shaoying Zhou for critical review of this manuscript and for many insightful scientific discussions.

■ ABBREVIATIONS

SOI = structure of interest
MMP = mapped molecular pair
SyGMa = Systematic Generation of possible Metabolites
NOAEL = no adverse effect limit
S = suitable
NS = not suitable
SI = suitable with interpretation
MSS = metabolism similarity score
RMSS = reactive metabolism similarity score
TMSS = total metabolism similarity score
DART DT = developmental and reproductive toxicity decision tree
qSIM = quantitative similarity

■ REFERENCES

(1) Wu, S.; Blackburn, K.; Amburgey, J.; Jaworska, J.; Federle, T. A. A framework for using structural, reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogs for SAR-based toxicological assessments. *Regul. Toxicol. Pharmacol.* **2010**, *56*, 67−81.

(2) Lester, C.; Reis, A.; Laufersweiler, M.; Wu, S.; Blackburn, K. Structure activity relationship (SAR) toxicological assessments: The role of expert judgment. *Regul. Toxicol. Pharmacol.* **2018**, *92*, 390−406.

(3) Blackburn, K.; Bjerke, D.; Daston, G.; Felter, S.; Mahony, C.; Naciff, J.; Robison, S.; Wu, S. Case studies to test: A framework for using structural, reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogs for SAR-based toxicological assessments. *Regul. Toxicol. Pharmacol.* **2011**, *60*, 120.

(4) Lester, C. C.; Yan, G. A matched molecular pair (MMP) approach for selecting analogs suitable for structure activity relationship (SAR)-based read across. *Regul. Toxicol. Pharmacol.* **2021**, *124*, 104966.

(5) Mellor, C. L.; Marchese Robinson, R. L.; Benigni, R.; Ebbrell, D.; Enoch, S. J.; Firman, J. W.; Madden, J. C.; Pawar, G.; Yang, C.; Cronin, M. T. D. Molecular fingerprint-derived similarity measures for toxicological read-across: Recommendations for optimal use. *Regul. Toxicol. Pharmacol.* **2019**, *101*, 121−134.

(6) Ball, N.; Cronin, M. T.; Shen, J.; Blackburn, K.; Booth, E. D.; Bouhifd, M.; Donley, E.; Egnash, L.; Hastings, C.; Juberg, D. R.; Kleensang, A.; Kleinstreuer, N.; Kroese, E. D.; Lee, A. C.; Luechtefeld, T.; Maertens, A.; Marty, S.; Naciff, J. M.; Palmer, J.; Pamies, D.; Penman, M.; Richarz, A. N.; Russo, D. P.; Stuard, S. B.; Patlewicz, G.; van Ravenzwaay, B.; Wu, S.; Zhu, H.; Hartung, T. Toward Good Read-Across Practice (GRAP) guidance. *Altex* **2016**, *33*, 149−166.

(7) Patlewicz, G.; Cronin, M. T. D.; Helman, G.; Lambert, J. C.; Lizarraga, L. E.; Shah, I. Navigating through the minefield of read-across frameworks: A commentary perspective. *Computational Toxicology* **2018**, *6*, 39−54.

(8) Gadaleta, D.; Golbamaki Bakhtyari, A.; Lavado, G. J.; Roncaglioni, A.; Benfenati, E. Automated integration of structural, biological and metabolic similarities to improve read-across. *ALTEX* **2020**, *37*, 469−481.

(9) Ridder, L.; Wagener, M. SyGMa: combining expert knowledge and empirical scoring in the prediction of metabolites. *ChemMedChem.* **2008**, *3*, 821−832.

(10) Boyce, M.; Meyer, B.; Grulke, C.; Lizarraga, L.; Patlewicz, G. Comparing the performance and coverage of selected in silico (liver) metabolism tools relative to reported studies in the literature to inform analogue selection in read-across: A case study. *Computational Toxicology* **2022**, *21*, 100208.

(11) Yordanova, D. G.; Schultz, T. W.; Kuseva, C. D.; Mekenyan, O. G. Assessing metabolic similarity for read-across predictions. *Computational Toxicology* **2021**, *18*, 100160.

(12) Yang, C.; Rathman, J. F.; Magdziarz, T.; Mostrag, A.; Kulkarni, S.; Barton-Maclaren, T. S. Do Similar Structures Have Similar No Observed Adverse Effect Level (NOAEL) Values? Exploring Chemoinformatics Approaches for Estimating NOAEL Bounds and Uncertainties. *Chem. Res. Toxicol.* **2021**, *34*, 616−633.

(13) Low, Y.; Sedykh, A.; Fourches, D.; Golbraikh, A.; Whelan, M.; Rusyn, I.; Tropsha, A. Integrative Chemical-Biological Read-Across

Approach for Chemical Hazard Classification. *Chemical Reserach in Toxicology* **2013**, *26*, 1199−1208.
(14) De Abrew, K. N.; Shan, Y. K.; Wang, X.; Krailler, J. M.; Kainkaryam, R. M.; Lester, C. C.; Settivari, R. S.; LeBaron, M. J.; Naciff, J. M.; Daston, G. P. Use of connectivity mapping to support read across: A deeper dive using data from 186 chemicals, 19 cell lines and 2 case studies. *Toxicology* **2019**, *423*, 84−94.
(15) Bajusz, D.; Rácz, A.; Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of cheminformatics* **2015**, *7*, 20−20.
(16) Marchant, C. A.; Briggs, K. A.; Long, A. In silico tools for sharing data and knowledge on toxicity and metabolism: derek for windows, meteor, and vitic. *Toxicol Mech Methods* **2008**, *18*, 177−187.
(17) Wu, S.; Fisher, J.; Naciff, J.; Laufersweiler, M.; Lester, C.; Daston, G.; Blackburn, K. Framework for Identifying Chemicals with Structural Features Associated with the Potential to Act as Developmental or Reproductive Toxicants. *Chem. Res. Toxicol.* **2013**, *26*, 1840−1861.
(18) Gleeson, M. P. (2008) Generation of a Set of Simple, Interpretable ADMET Rules of Thumb, In *J. Med. Chem.* pp 817−834, American Chemical Society.
(19) Gaohua, L.; Miao, X.; Dou, L. Crosstalk of physiological pH and chemical pKa under the umbrella of physiologically based pharmacokinetic modeling of drug absorption, distribution, metabolism, excretion, and toxicity. *Expert Opinion on Drug Metabolism & Toxicology* **2021**, *17*, 1103−1124.
(20) *PyFingerprint (2.0): Python tool to generate chemical fingerprints*. https://github.com/hcji/PyFingerprint.
(21) *RDKit: Open-source cheminformatics*. http://www.rdkit.org.
(22) Sedykh, A. Y.; Shah, R. R.; Kleinstreuer, N. C.; Auerbach, S. S.; Gombar, V. K. Saagar−A New, Extensible Set of Molecular Substructures for QSAR/QSPR and Read-Across Predictions. *Chem. Res. Toxicol.* **2021**, *34*, 634−640.
(23) Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B. E.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J. B.; Grout, J.; Corlay, S.; Ivanov, P.; Avila, D.; Abdalla, S.; Willing, C.; Team, J. D. Jupyter Notebooks - a publishing format for reproducible computational workflows. *ELPUB* **2016**, 87−90.
(24) Rossum, G. V.; Drake, F. L.*Python 3 Reference Manual*; CreateSpace; 2009.
(25) *R. C. T. R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
(26) JJ Allaire, Y. X.; Jonathan, M.; Javier, L.; Kevin, U.; Aron, A.; Hadley, W.; Joe, C.; Winston, C.; Richard, I.*rmarkdown: Dynamic Documents for R. R package version 2.5*; 2020.
(27) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(28) Nair, B.Final report on the safety assessment of Benzyl Alcohol, Benzoic Acid, and Sodium Benzoate. *Int. J. Toxicol20 Suppl 3***2001**, 23−50.
(29) *Toxicological review of xylenes*; U.S. EPA, 2003.
(30) Uetrecht, J. P., Trager, W., Eds. *Drug Metabolism Chemical and Enzymatic Aspects*; Informa Healthcare: New York, 2007.
(31) Zhang, F.; Bartels, M. J.; Clark, A. J.; Staley, J. L.; Lardie, T. S.; Markham, D. A.; Hughes, B. J.; Ball, N. S. Comparative metabolism and pharmacokinetics of diisobutyl ketone and diisobutyl carbinol in male SD rats. *Toxicol. Lett.* **2015**, *232*, 175−181.
(32) Kleber, A.; Maurer, F.; Lorenz, D.; Wolf, B.; Albrecht, F.; Shopova, T.; Sessler, D. I.; Volk, T.; Kreuer, S.; Fink, T. Metabolism of 3-pentanone under inflammatory conditions. *Journal of breath research* **2016**, *10*, 047101.
(33) O'Donoghue, J. L.Ketones of Six to Thirteen Carbons. In *Patty's Toxicology*; John Wiley & Sons, Inc., 2012, pp 807−914.
(34) Chism, J. P.; Turner, M. J., Jr.; Rickert, D. E. The metabolism and excretion of mononitrotoluenes by Fischer 344 rats. *Drug Metab. Dispos.* **1984**, *12*, 596−602.
(35) Schultz, T. W.; Cronin, M. T. D. Lessons learned from read-across case studies for repeated-dose toxicity. *Regul. Toxicol. Pharmacol.* **2017**, *88*, 185−191.

https://doi.org/10.1021/acs.chemrestox.2c00311
Chem. Res. Toxicol. 2023, 36, 230−242