LIVERPOOL JOHN MOORES UNIVERSITY

LJMU Research Online

Thompson, CV, Webb, SD, Leedale, JA, Penson, PE, Paini, A, Ebbrell, D and Madden, JC

Using read-across to build physiologically-based kinetic models: Part 1. Development of a KNIME workflow to assist analogue selection for PBK modelling

http://researchonline.ljmu.ac.uk/id/eprint/22143/

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

Thompson, CV, Webb, SD, Leedale, JA, Penson, PE, Paini, A, Ebbrell, D and Madden, JC (2023) Using read-across to build physiologically-based kinetic models: Part 1. Development of a KNIME workflow to assist analogue selection for PBK modelling. Computational Toxicology, 29. p. 100292. ISSN

LJMU has developed **LJMU Research Online** for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

http://researchonline.ljmu.ac.uk/

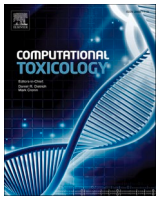
http://researchonline.ljmu.ac.uk/

Computational Toxicology 29 (2024) 100292

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.sciencedirect.com/journal/computational-toxicology



Full Length Article

Using read-across to build physiologically-based kinetic models: Part 1. Development of a KNIME workflow to assist analogue selection for PBK modelling

Courtney V. Thompson [a,b], Steven D. Webb [b], Joseph A. Leedale [b], Peter E. Penson [a], Alicia Paini [c], David Ebbrell [a], Judith C. Madden [a,*]

[a] School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, England
[b] Syngenta, Jealott's Hill International Research Centre, Bracknell, RG42 6EY, England
[c] esqLABS, 34, Hambierich, 26683 Saterland, Germany

ARTICLE INFO

Keywords:
PBK
PBPK
Read-across
KNIME
Similarity
NAM

ABSTRACT

Read-across refers to the process by which information from one (source) chemical is used to infer information about another similar (target) chemical. This method can be used to fill data gaps and so inform safety assessment where data are lacking for chemicals of interest. As one chemical cannot be considered as absolutely similar to another, only similar with respect to a given property, it is essential to justify the selection of similar chemicals (analogues) for the purposes of read-across. A previously created dataset of available physiologically-based kinetic (PBK) models (referred to as the PBK modelling dataset or PMD) was used in the development of a KNIME workflow. KNIME is a freely-available, open-source analytics platform that allows users to create workflows to analyse and visualise data. The KNIME workflow described here was designed to identify chemical analogues with a corresponding model in the PMD. The PMD combined with the KWAAS enables PBK model information from source chemical(s) to be used in a read-across approach to help develop new PBK models for target chemicals. This KNIME workflow was applied to six chemicals, representing different types of chemical classes (drugs, cosmetics, botanicals, industrial chemicals, pesticides, and food additives) to assess its applicability across various industries. Information acquired from these PBK models can be used to support safety assessment of chemicals and reduce reliance on animal testing.

1. Introduction

Safety assessment of chemicals is an essential process across multiple sectors, including the cosmetic, food additive, pharmaceutical, plant protection product industries, and regulatory authorities. Assessment requires knowledge of the potential of the chemical to elicit an effect, combining information regarding both intrinsic activity and kinetics. As many chemicals, including those new to the market, have little or no safety data, and the testing of all such chemicals would neither be practicable nor ethically acceptable, New Approach Methodologies (NAMs), such as predictive computational methods are essential. Physiologically-based kinetic (PBK) models are used to simulate organ-level concentration–time profiles, enabling the relevant dose metrics, associated with an effect, to be determined. In PBK modelling the body is represented as a series of compartments, linked by blood flow. The model structure comprises physiological and anatomical parameters (such as blood flow and organ volume) combined with chemical-specific factors (such as solubility, partitioning behaviour, and plasma-protein binding) to enable the simulation of the concentration–time profile in the relevant organs. However, these models are costly, in terms of time and animal resources, to develop *de novo*. Hence, relatively few models are available in comparison to the number of chemicals in use.

Such gaps in knowledge, can however be filled using a read-across approach, wherein information from a data-rich (source) chemical is used to fill a gap in knowledge for a similar, data-poor (target) chemical. Read-across is commonly used in submissions for regulatory submissions under the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulation and guidance on performing and evaluating read-across predictions is available [1,2]. Guidance on read-across emphasises the importance of ascertaining the similarity of the

* Corresponding author.
E-mail address: j.madden@ljmu.ac.uk (J.C. Madden).

https://doi.org/10.1016/j.comtox.2023.100292
Received 3 July 2023; Received in revised form 18 October 2023; Accepted 29 November 2023
Available online 1 December 2023
2468-1113/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

source and target chemicals, with respect to relevant characteristics, i.e., that the source chemical can be considered a suitable "analogue" of the target. Previously, it has been shown that a read-across approach can be applied successfully to the development of PBK models [3,4]. In this case information, such as equations and structure, from an existing PBK model, developed for a source chemical (analogue), can be used to inform the development of a PBK model for a "similar" target chemical. To apply this approach successfully, two critical pieces of information are required: details of existing PBK models and a method to identify suitable analogues. In 2021 a readily searchable dataset of 7,541 existing PBK models was published containing information on the chemical modelled, administration route, species, whether full equations were reported, etc [5]. This Physiologically-Based Kinetic Modelling Dataset (PMD) serves as a valuable resource for identifying existing PBK models. The original version of the PMD is downloadable from Thompson et al (2021) [5]. Researchers are able to adapt the PMD to suit their needs, for example by updating with additional models of interest. It is anticipated that future versions of the PMD will be published where a substantial number of additional models have been curated. Herein, we describe the development of a complementary workflow for finding models for target chemicals within the PMD, or for selecting analogues from the PMD for the purpose of read-across, if no models are available for the target.

The selection of analogues is predicated on identifying a chemical that can be considered similar to the target with respect to relevant properties. There are many ways in which one chemical can be considered as being similar to another, e.g., in terms of structural or physico-chemical properties, mechanism of action, or metabolite formation [4]. Similarity in chemical structure can be assessed using fingerprints. Many methods for generating and comparing chemical fingerprints are available; however, each method will give a different result and there is no consensus as to which one is the most appropriate [6]. Examples of different similarity metrics and their calculation have been described previously [7]. Briefly, key structural features are ascribed a "bit" value of '1' or '0', for their presence or absence respectively, within a chemical structure. The number of "bits" that two structures have in common can be used to generate their similarity score. A Tanimoto score allows for the degree of commonality in the presence of key structural features (bits) to be determined, where a Tanimoto score between zero (not similar) and one (highly similar) is calculated. A Tanimoto score of 0.6 or greater has previously been applied as a cut-off value when searching for toxicologically similar chemicals, although this figure is arbitrary and higher or lower values may be suitable for different scenarios [8].

The degree of similarity between chemicals can also be determined in relation to their physico-chemical properties. In terms of kinetics, relevant physico-chemical properties include the logarithm of the octanol:water partition coefficient (log P), pKa (degree of ionisation), molecular weight and aqueous solubility. These chemical-specific parameters can be derived from experimental measurements or predicted using software. Collations of computational resources for predicting physico-chemical properties and absorption, distribution, metabolism, and excretion (ADME) properties, with a focus on those relevant for PBK modelling have been published previously [9,10]. Note that defining the similarity of chemicals in terms of physico-chemical properties alone is usually not appropriate as many chemicals exhibiting significantly different kinetic behaviour may have similar physico-chemical properties.

The present study describes the development of a KNIME workflow, specifically designed to be used in conjunction with the PMD to identify analogues that may be used in a read-across approach for PBK model development. KNIME (https://www.knime.com) is a freely-available, open-source analytics software that allows for a user to undertake data analysis through building visual workflows. Tutorials, including worked examples and links to training videos for using KNIME for a range of purposes are available at: https://www.knime.com/learning. Creating the automated workflow using the KNIME software provides flexibility for subsequent users to adapt the workflow to specific requirements. Chemical similarity is assessed using information from nine chemical fingerprints and readily calculable physico-chemical properties. Guidance for how best to use the KNIME workflow and the PMD for selecting the most appropriate PBK model to use for different scenarios is described here.

2. Methods

2.1. Using KNIME to identify existing models for target chemicals within the PMD

When developing a new PBK model for a target chemical, the first step is to identify whether or not there are existing PBK models for that chemical that may inform the development of the new model. Hence the ability to search the PMD for models for a specific target chemical makes it a highly valuable and effective tool. Searching by chemical name, SMILES string, or CAS number are all problematic, due to multiple variations being possible (e.g. different spellings or SMILES representations). Therefore, a node was included in the workflow that enabled searching of the PMD for a target chemical using its unique international chemical identifier (InChiKey). For cases where no suitable existing PBK model was available in the PMD for the target chemical, a KNIME workflow for identifying suitable analogues was devised.

2.2. Development of the KNIME workflow for assisting analogue selection (KWAAS)

The KNIME workflow for assisting analogue selection (KWAAS) was developed using an iterative design and test cycle. This approach helped to determine that multiple methods to generate chemical fingerprints, assess structural similarity, and compare key physico-chemical properties were needed to rationally identify suitable analogues. Chemical fingerprints were used to determine how similar one chemical is to another. The presence or absence of key structural features in chemicals are used to calculate the similarity of one chemical to another using a mathematical formula based on the number of features the chemicals have in common. This can be translated into a Tanimoto score; these scores range from zero to one. For chemicals that are highly similar or the same, a Tanimoto score close to or equal to one is assigned, whilst dissimilar chemicals will have a score closer to zero. Tanimoto similarity scores, using nine different fingerprints, are generated by the workflow, in order to be able to capture a wide range of chemicals that could be considered similar. As each method generates fingerprints using a unique approach, each type of fingerprint may identify different analogues. The nine fingerprint types that were used within the workflow to generate a list of similar chemicals to a target chemical were Morgan, FeatMorgan, AtomPair, Torsion, RDKit, Avalon, Layered, MACCS, and Pattern. The results are ranked from most similar to least similar and the top five most similar analogues, as calculated by each of the nine different types of fingerprints, is provided as an output. Theoretically, a list with a maximum of 45 chemicals (analogues) could be generated, although duplications in chemicals identified as similar reduces the list considerably. Analogues that are identified as similar by only one type of fingerprint are rejected, as this indicates that the similarity is being assessed using a unique criterion that does not represent more general similarity.

Further refinement based on Tanimoto score and physico-chemical properties was developed to reduce the number of potential analogues being proposed. A minimum Tanimoto score of 0.6, a commonly used cut-off value [8] was used (this value can be adjusted by the user, if required). Molecular weight, log P (predicted logarithm of the octanol: water partition coefficient), TPSA (topological polar surface area), hydrogen bond donors/acceptors and Lipinksi rule violations for each chemical were obtained directly from the PMD created by Thompson et al. (2021) [5]. In the analysis undertaken here, molecular weight and

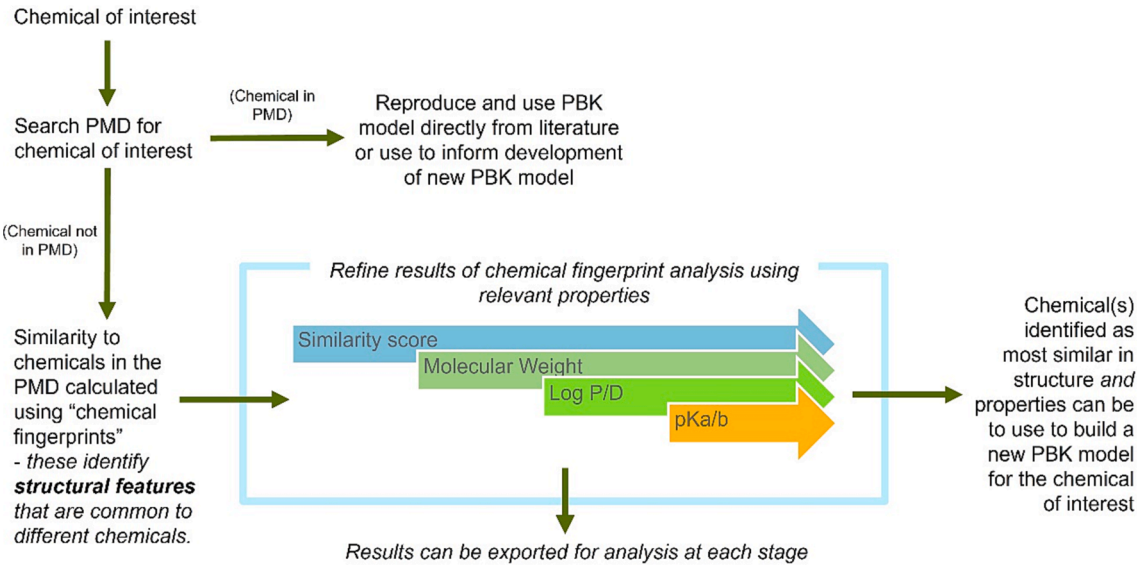


Fig. 1. A summary of the KWAAS process. The main stages of the workflow include, searching PBK model dataset (PMD) for an available model for the chemical of interest (target chemical), determining similar analogues to the target (i.e., potential source chemicals) based on structure and relevant characteristics such as physico-chemical properties, and refining the selection of source chemicals as necessary.

log P were included in the workflow refinement options to identify the most similar chemicals (inclusion ranges can be set and adjusted by the user, as required). However, a user can also select different physico-chemical properties. In order to test the workflow, the properties pKa and log D at pH 7.4 and 5.5 were used. These were predicted using the OPERA software (https://ntp.niehs.nih.gov/whatwestudy/niceatm/comptox/ct-opera/opera.html; accessed October 2023). OPERA was unable to calculate physico-chemical properties for certain chemicals, (for example, unusually large chemicals, organometallics, mixtures, and inorganics); therefore, these properties were not available for all chemicals. The physico-chemical properties (molecular weight, log P, log D, and pKa) were integrated into the KNIME workflow to use throughout the refinement and selection of appropriate chemicals using the selected nodes in KNIME. Early iterations of workflow development considered only molecular weight and log P for refining the selection of chemicals identified using structural fingerprints. Subsequent iterations demonstrated that refinement using log D and pKa enabled the selection to be optimised.

The workflow has been designed such that a user can adapt the properties employed by the KWAAS to those best suited to their needs. For example, the user could generate physico-chemical properties or properties related to ADME characteristics for all chemicals in the PMD using in-house or external tools. These can be integrated into the workflow by the user to refine analogue selection, either instead of, or in addition to, properties described here. The flexibility and adaptability of the KNIME workflow is one of its key benefits.

The suitability of a model for a source chemical (analogue) to be used in a read-across approach can also be influenced by the characteristics of the model. For example, a model developed using a specific type of software or using a specific route of administration may be more relevant. Therefore, the ability to refine analogue selection based on such model characteristics is essential. For this reason, an additional node was incorporated that enabled selection to be refined based on model characteristics such as the species or sex of the animal used, software employed to generate the model, availability of equations, etc. The workflow was designed such that the user could export the results for analysis at each stage, enabling expert judgement to be incorporated into the selection process.

2.3. Using the KWAAS

A workflow of the similarity tool created in KNIME, an open-source analytics platform, is summarised in Fig. 1. The KNIME workflow developed has three main stages as listed below.

1. Search PBK model dataset (PMD) for an available model for the chemical of interest (target chemical).
2. Determine similar analogues to the target (i.e., potential source chemicals) based on structure and relevant characteristics such as physico-chemical properties.
3. Refine the selection of source chemicals as necessary.

A guide on how to use the similarity tool is available in the supplementary material. The KNIME workflow can be freely downloaded from GitHub (https://github.com/courtneythomp/KWAAS; accessed October 2023) or obtained by contacting the corresponding author.

If a user has additional PBK models that they want to include in the workflow (e.g., models not captured in the PMD or in-house models), this can be done in two ways. Firstly, by downloading the PMD and adding additional PBK models to the Excel spreadsheet before re-uploading into the workflow, or secondly, by adding chemicals to the table within the workflow itself.

2.3.1. Stage 1

The first stage is to use the workflow to search the PMD for existing models for the chemical of interest. Whilst this may be performed by searching for a chemical name, CAS number, or SMILES, searching using the InChIKEY of the target chemical, as a unique chemical identifier, is recommended. There are multiple websites and webtools available to obtain InChiKeys for example PubChem (https://pubchem.ncbi.nlm.nih.gov; accessed October 2023) or ChemSpider (https://www.chemspider.com; accessed October 2023). Once the target chemical has been inputted (e.g., by InChiKey) information on the model, including all model characteristics as recorded in the PMD, will be output in the form of an Excel spreadsheet.

2.3.2. Stage 2

If no existing models for the target chemical are available in the PMD, or the models were deemed unsuitable by the user, the second stage of the workflow can be used to identify similar chemicals (analogues) from the dataset for which PBK models are available. Similarity

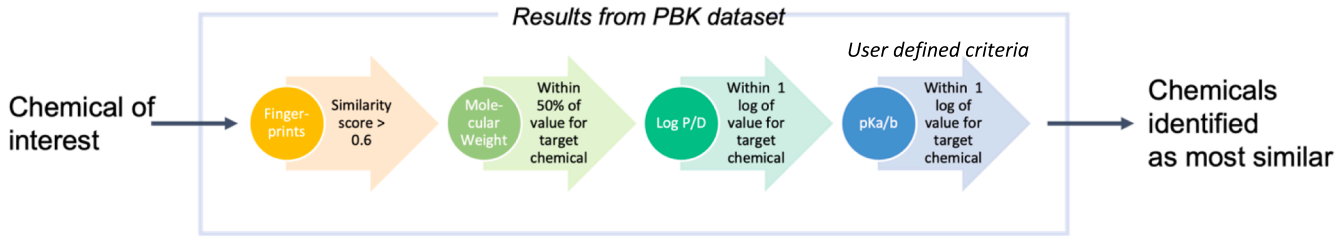


Fig. 2. Stages of refinement and the recommended ranges to use at each stage of the KWAAS – the user may adapt any of these as required. These are preliminary recommendations for using the KWAAS. The user can add any other properties for selection that they can obtain and identify as being useful.

Table 1
Results of the analogue search and refinement for genistein showing the inclusion ranges and number of chemicals identified at each step of the refinement process.

Features for searching and refining	Target - Genistein	Inclusion range for refining results	Number of analogues identified by KWAAS filter at each step
Fingerprints			11*
Tanimoto score		> 0.6	9*
Molecular weight	270.24 Da	135.12 – 405.36 Da	7*
Log P	2.58	1.58 – 3.58	5*

* results include the target chemical, genistein.

Table 2
The target chemical, bicyclopyrone, the inclusion range at each step of the refinement process, when applying the KWAAS, and the number of chemicals identified at each stage.

Features for searching and refining	Target - Bicyclopyrone	Inclusion range for refining results	Number of analogues identified by KWAAS filter at each step
Fingerprints			42
Tanimoto score		> 0.6	15
Molecular weight	399.4	199.7 – 599.1 Da	12
Log P	1.45	0.45—2.45	3

Table 3
The target chemical, benzoic acid, the inclusion range at each step of the refinement process, when applying the KWAAS, and the number of chemicals identified at each stage.

Features for searching and refining	Target - Benzoic acid	Inclusion range for refining results	Number of analogues identified by KWAAS filter at each step
Fingerprints			12*
Tanimoto score		> 0.6	9*
Molecular weight	122.12 Da	61.06 – 183.18 Da	7*
Log P	1.38	0.38 – 2.38	6*

* results include the target chemical, benzoic acid.

of chemicals within the PMD is calculated using nine fingerprinting methods (as described above). Theoretically a maximum of 45 potential chemical analogues can be generated, although in many cases the same chemical will be ranked in the top five by multiple fingerprints, so reducing the number of potential analogues. An Excel file of the results including chemical name, SMILES, and structures of the analogues can be obtained at this stage. Refinements to the analogue selection can be made based on Tanimoto score and physico-chemical properties.

Physico-chemical properties including molecular weight, log P, pKa/pKb, and log D can all be used to further determine how similar one chemical is to another. Output in the form of an Excel spreadsheet containing information on the chemical name, structure, SMILES and physico-chemical properties can be obtained at any point of the refinement process. The different stages of refinement and recommended refinement ranges are summarised in Fig. 2. Fingerprints, molecular weight and log P are recommended to be used initially as these are the most universal parameters. In addition, it is recommended that at every stage of the workflow the user reviews the structures of the chemicals identified as similar, to assess their suitability, using expert judgement.

Table 4
The target chemical, methyleugenol, the inclusion range at each step of the refinement process, when applying the KWAAS, and the number of chemicals identified at each stage.

Features for searching and refining	Target - Methyleugenol	Inclusion range for refining results	Number of analogues identified by KWAAS filter at each step
Fingerprints			11*
Tanimoto score		> 0.6	11*
Molecular weight	178.23 Da	89.12 – 267.35 Da	10*
Log P	2.43	1.43 – 3.43	9*

* results include the target chemical, methyleugenol.

Table 5
The target chemical, loxapine, the inclusion range at each step of refining the results from the KWAAS and the number of chemicals identified at each stage.

Features for searching and refining	Target - loxapine	Inclusion range for refining results	Number of analogues identified by KWAAS filter at each step
Fingerprints			18
Tanimoto score		> 0.6	13
Molecular weight	327.8 Da	163.9 – 491.7 Da	7
Log P	3.6	2.6 – 4.6	5

Table 6
The target chemical, benzyl butyl phthalate, the inclusion range at each step of the refinement process, when applying the KWAAS, and the number of chemicals identified at each stage.

Features for searching and refining	Target - Benzyl butyl phthalate	Inclusion range for refining results	Number of analogues identified by KWAAS filter at each step
Fingerprints			11
Tanimoto score		> 0.6	8
Molecular weight	312.4 Da	156.2 – 468.6 Da	8
Log P	4.82	3.82 – 5.82	0
Log D (pH 7.4)	4.68	3.68 – 5.68	1

2.3.3. Stage 3

Once suitable analogues have been determined in Stage 2, the PMD can again be searched to identify all associated PBK models in the PMD.

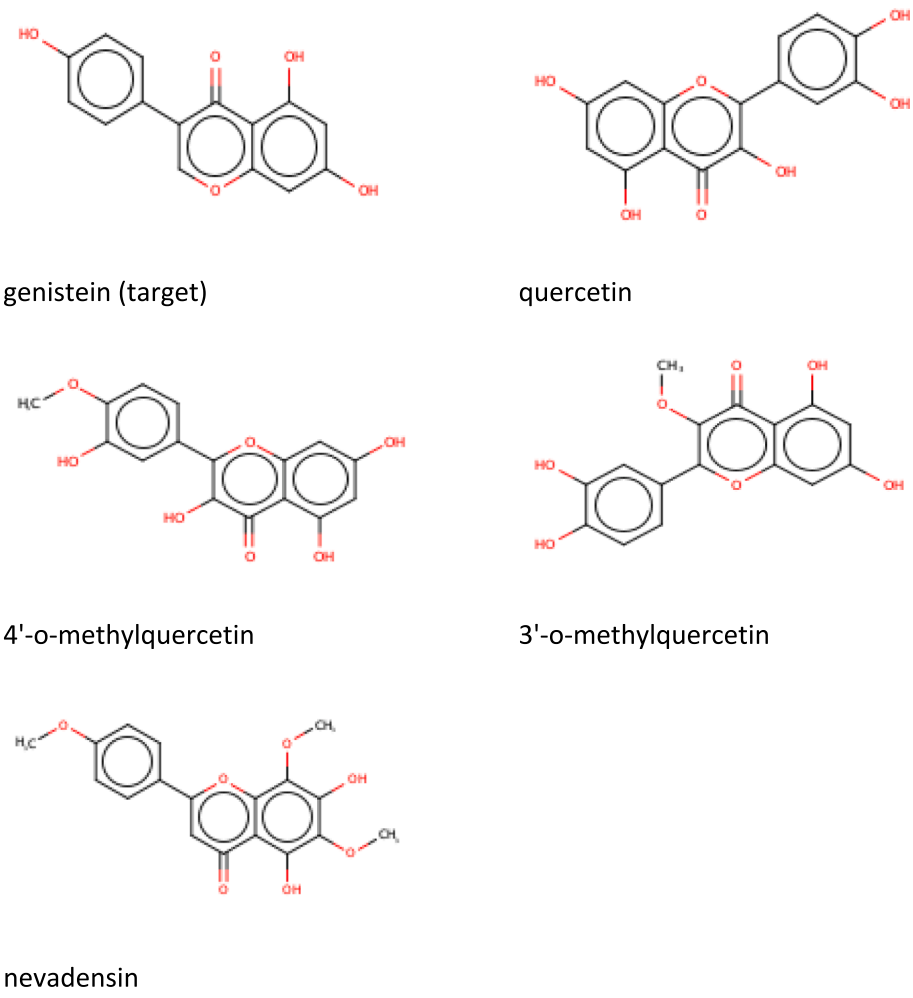


Fig. 3. Structure of target chemical, genistein, and the chemicals identified as similar after refining by log P.

A corresponding Excel file containing all the information recorded in the PMD for this PBK model dataset can be generated. The PBK model dataset can be further filtered by species, availability of equations, software used in modelling, sex, life-stage, etc. The results at each step (i.e. after the application of each filtering node) can be exported to Excel and examined as described in Stage 2.

2.4. Testing the KWAAS

In order to test the applicability and relevance of the KWAAS, as well as its limitations, a test series of six chemicals were used in a pilot study. One chemical from each of six in-house datasets (comprising botanicals, plant protection products, cosmetic ingredients, food additives, pharmaceuticals, and REACH (Registration, Evaluation Authorisation, and restriction of Chemicals) chemicals) was selected.

Genistein was chosen as an example of a commonly used botanical - its molecular weight and log P were obtained from the PMD. Bicyclopyrone was selected to represent plant protection products - molecular weight and log P were obtained from PubChem (https://pubchem.ncbi.nlm.nih.gov; accessed October 2023) and ChemSpider (https://www.chemspider.com; accessed October 2023). Benzoic acid was selected as a commonly used cosmetic ingredient - molecular weight and log P were obtained from the PMD. Methyleugenol was chosen to represent food additives as this chemical had been part of a previous investigation into read-across of PBK models by Paini et al. [4] and could be used to confirm the results from the KWAAS - molecular weight and log P were obtained from the PMD. Loxapine was chosen at random to represent pharmaceuticals - molecular weight and log P obtained from PubChem. A REACH chemical, benzyl butyl phthalate, identified by ECHA as a chemical of concern https://echa.europa.eu/substance-information/-/substanceinfo/100.001.475; accessed October 2023), was selected to represent industrial chemicals - molecular weight and log P were obtained from the CompTox Chemicals Dashboard.. In this case log D was also obtained using ADMETlab 2.0 (https://admet.scbdd.com/; accessed October 2023) for prediction.

The name of each of the six chemicals was entered into PubChem to obtain the respective InChiKey. This was used for the initial search of the PMD. The availability of any PBK models for the selected chemicals was recorded (i.e., Stage 1 of the KWAAS). The selected chemicals were then inputted into Stage 2 of the KWAAS to identify analogues that had one or more existing PBK models. In this pilot study all six chemicals progressed into Stage 2, even if at Stage 1 it was determined that PBK models were available for the chemical of interest itself. This was to determine how well the KWAAS performed in selecting analogues for the different types of chemicals.

The molecular weight of analogue chemicals had to be within ± 50 % of the molecular weight of the target to be included. Inclusion ranges for log P, and log D (at pH 7.4) were selected to be within ± 1 log unit of the target (Fig. 2).

3. Results

The applicability of the KWAAS across different types of chemicals was tested using exemplar chemicals from six different datasets

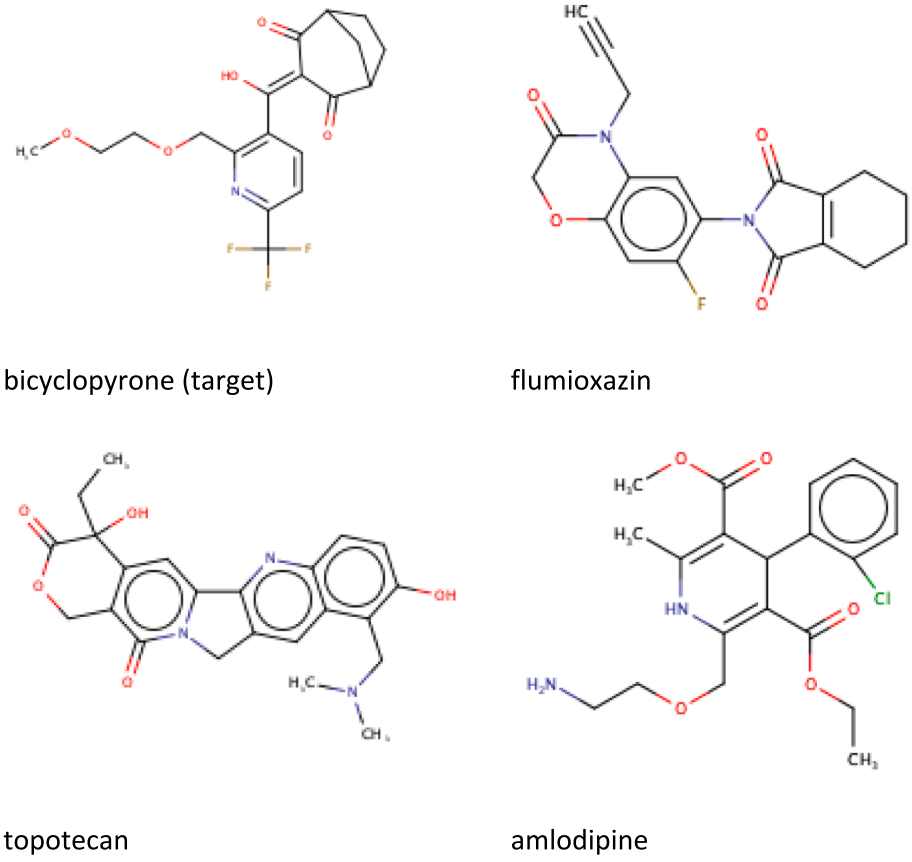


Fig. 4. Structure of target chemical, bicyclopyrone, and the chemicals identified as similar after refining by log P.

representing plant protection products, botanicals, drugs, cosmetics, food additives and industrial chemicals. The property information for the target chemical, the inclusion ranges used within the KWAAS process, and the number of chemical analogues identified as similar at each stage is summarised for each target chemical in Tables 1-6. The results at each step of the workflow, using genistein as an illustrative example, are provided in the supplementary material. The structure of the target chemical and its identified chemical analogues are depicted for each target chemical in Figs. 3-8.

A total of five PBK models for the botanical chemical genistein were found in the PMD. Stage 2 of the KWAAS was used to identify similar chemicals to genistein (Table 1, Fig. 3). Refining by molecular weight, log P, and pKa made the similarity search too specific, resulting only in genistein itself being identified by the KWAAS. However, by using fingerprints, molecular weight and log P alone within the KWAAS, a reasonable number of analogues (four) were identified, and this was considered a reasonable point at which to stop refining the selection. Information on whether or not the identified analogue is a metabolite of a parent compound can be found in the column entitled 'Additional Information on Chemical' given in the PMD.

Bicyclopyrone (Table 2, Fig. 4) had no available PBK models in the PMD. Initially, 42 unique chemicals were identified as similar to bicyclopyrone using fingerprints alone in the KWAAS, hence results were refined using molecular weight and log P. Four PBK models were initially identified in the PMD for benzoic acid (Table 3, Fig. 5). Using the chemical methyleugenol (Table 4, Fig. 6) for the food additives sector, a total of 11 chemicals, including methyleugenol, were initially identified as similar. Note that some of chemicals identified as similar were metabolites of the target. In this case the most relevant PBK model would be the one representing the parent compound. The pharmaceutical sector was represented using the chemical loxapine (Table 5, Fig. 7). The industrial chemical sector was represented by benzyl butyl phthalate (Table 6, Fig. 8). Filtering using log P resulted in no similar chemicals being identified and so for this example log D was used as a filter.

3.1. *Applying knowledge from existing PBK models*

The results of the KWAAS pilot study show that for each of the six chemicals, a range of PBK models can be identified for either the target or its analogues. These PBK models offer a rich source of kinetic information that may be used in a variety of ways.

During stage 1 of the KWAAS, a PBK model(s) may be identified for the target itself. If the PBK model meets the relevant suitability assessment criteria (see below), then the information from the existing model may provide sufficient data to inform safety assessment without generating a new model. If there is no model for the target, but a suitable model is available for an analogue (see below for considerations of suitability), then information from this model may be used in different ways. One possibility is to use the analogue PBK model structure and information to inform the development of a new PBK model for the target. In this scenario, the same model structure is used but the chemical-specific parameters are adjusted to those of the target chemical. This approach has been successfully applied in previous studies [3,4].

Results generated by the KWAAS may alternatively be used to optimise existing, generic PBK models. Historically, the use of PBK modelling software required specialist knowledge; however, more recently, simplified generic PBK models have become publicly available, e.g., PK-Sim (https://www.open-systems-pharmacology.org; accessed October 2023). Whilst generic models can be useful, their lack of chemical-specific parameterisation can result in inaccurate predictions. Where PBK model information is available for either the target chemical or for analogues, this information could be used to optimise a generic model,

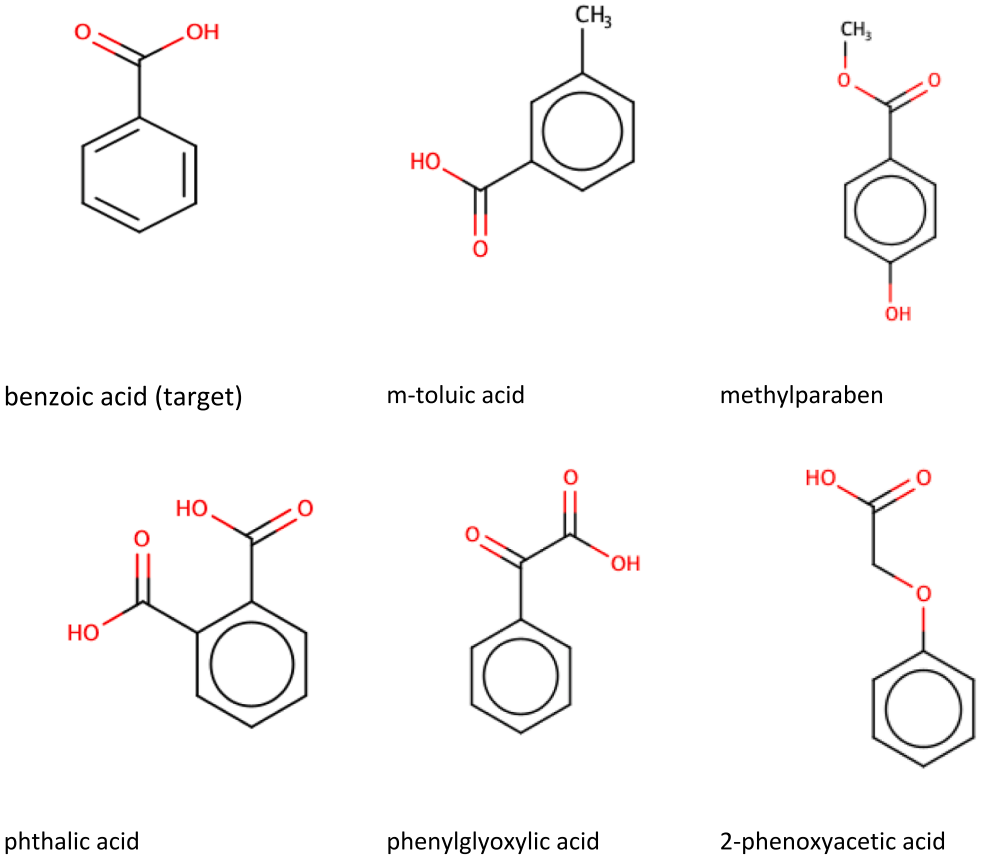


Fig. 5. Structure of target chemical, benzoic acid, and the chemicals identified as similar after refining by log P.

for example, by identifying model components that are relevant to a particular chemical type, enabling the generic model to be rationally adapted. The advantage of this approach is that it can be applied where a complete PBK model is not available or is not reproducible.

3.2. Assessing the suitability of a PBK model for a given purpose

In 2021 the Organisation for Economic Cooperation and Development produced "a guidance document on the characterisation, validation and reporting of PBK models for regulatory purposes" [11]. This document updated previous guidance from the World Health Organisation on the "Characterization and Application of Physiologically Based Pharmacokinetic Models in Risk Assessment" [12]. The focus of the OECD document was on the use of NAMs and includes guidance on selecting analogues for the purpose of read-across (as applied to PBK modelling). The document from WHO proposed guidance on assessing the suitability of PBK models for a given purpose.

Model quality is often associated with key metrics (such as the area under the concentration–time curve (AUC), maximum concentration (C_{max}), time to reach maximum concentration (T_{max}) or half-life ($t_{½}$)) being within 2-fold error of known values. Other metrics include visual inspection of goodness-of-fit and lineshape, or a correlation coefficient between predicted and observed values (e.g., R^2 of ≥ 0.75). The KWAAS identifies PBK models from the PMD; however, no assessment of the quality of these models has been undertaken, other than that inferred from their inclusion in peer-reviewed publications. There are different criteria to consider when selecting the most appropriate PBK model to use from those proposed by the KWAAS. For example, only selecting models where assessment of model quality is assured (e.g., determined by lineshape analysis or key metrics being within two-fold) or more prosaically, that a relevant species/sex/route of administration was used, that the relevant equations were reported, the software used is widely available, or the model was reported in sufficient detail to be readily reproducible. All of these factors may influence which PBK model, identified by the KWAAS, is ultimately utilised. These considerations can be summarised in a flow diagram for selecting PBK models in different scenarios (Fig. 9).

4. Discussion

Through an iterative process of design and testing, a KNIME workflow to aid the selection of chemicals that are similar to a given target, for which existing PBK model(s) are available, has been developed. Identifying appropriate PBK models enables relevant kinetic information to be obtained, new models to be built using a read-across approach, or generic models to be optimised. The KWAAS successfully identified similar chemicals, with available PBK models in the literature, for six example chemicals, as well as identifying existing models for the target chemicals themselves where available. Initial visual inspection of the analogues selected for five of the chemicals (genistein, benzoic acid, methyleugenol, loxapine and benzyl butyl phthalate) indicate that one or more of these may be suitable starting points for developing new PBK models in a read-across approach or for obtaining relevant kinetic data. However, the chemicals identified for bicyclopyrone were more diverse and hence may not be suitable analogues to use for this purpose.

More potential analogues were identified for methyleugenol (a food additive), than were identified for genistein (a botanical). Methyleugenol was the subject of an earlier study by Paini et al (2021) [4] who also used a read-across approach to develop a PBK model for this chemical. In their approach, the authors searched the literature and the OECD QSAR Toolbox to find analogues of methyleugenol (whether or not these analogues had an existing PBK model). Whilst Paini et al identified nine potential analogues, only two of these (estragole and safrole) were associated with valid and reproducible PBK models. In

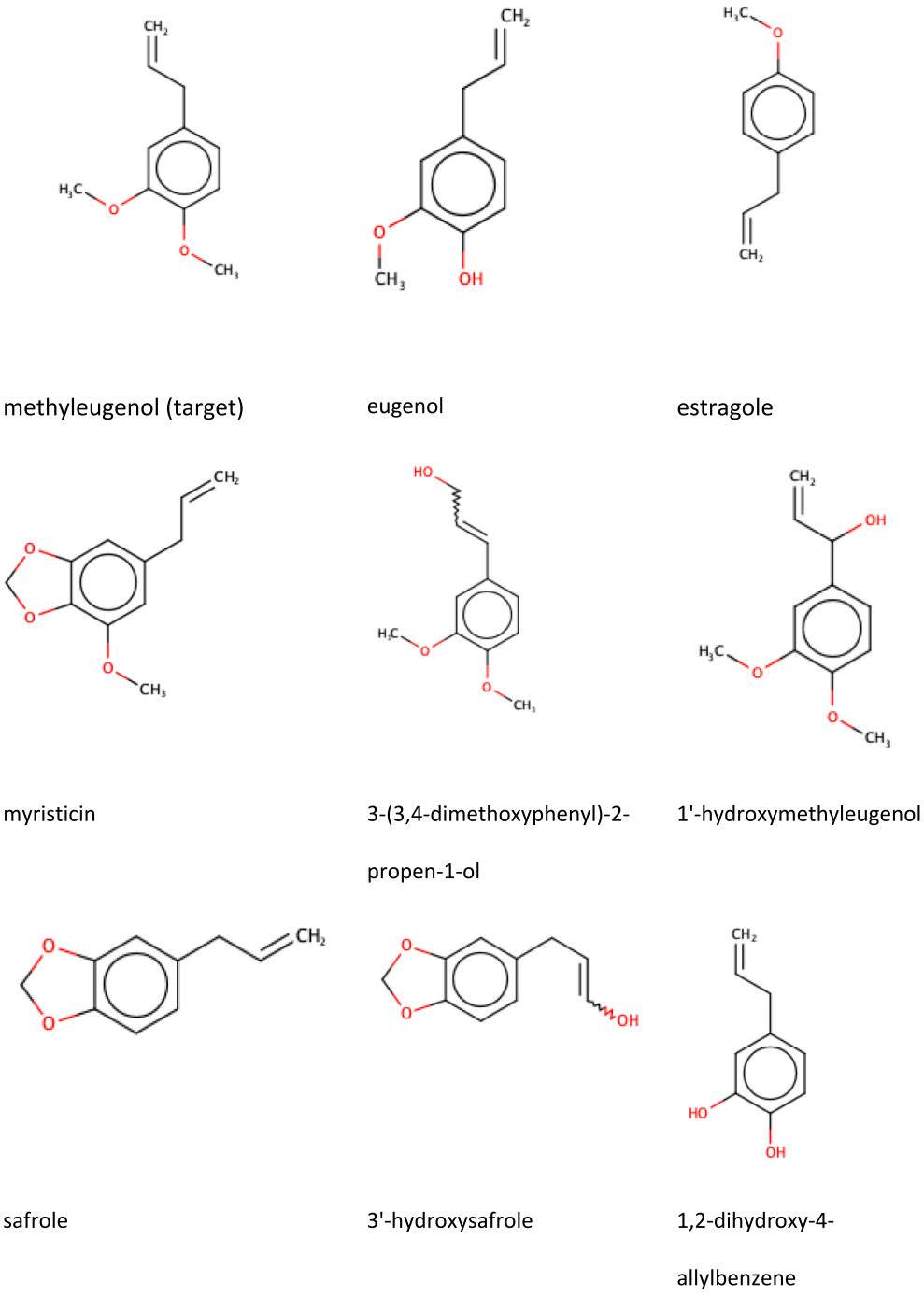


Fig. 6. Structure of target chemical, methyleugenol, and the chemicals identified as similar after refining by log P.

contrast, the present study identified nine potential analogues, with existing PBK models, these did include both estragole and safrole. The largest number of potential analogues (42) were identified for bicyclopyrone (a plant protection product). This is not unexpected as many plant protection products are present in the PMD. However, upon visual inspection, the structures appear dissimilar. For each of the target chemicals assessed, there were at least ten chemicals identified as being similar using fingerprints alone. However, this number of potential analogue chemicals decreased once the results were filtered by physicochemical properties, and quite significantly in the case of bicyclopyrone.

In contrast to the approach taken by Sweeney (2022) [13], which uses a single fingerprint method (Saagar) to determine similarity of chemicals from the PMD, the KWAAS uses nine different fingerprints to identify similar chemicals. Metabolites are commonly highlighted as similar to given parent chemical targets by the KWAAS, particularly when using chemical fingerprints as the basis for similarity. This is because both the parent and metabolite will have some common structural features that would be identified by the fingerprints. Therefore, there is a need to use expert judgement when determining the most useful chemical, with a PBK model, for a given purpose. Molecular weight and log P are universal and readily available properties; hence they are readily incorporated as filters when assessing the chemicals identified as similar using chemical fingerprints. A key advantage of this workflow is that additional properties (e.g., ADME measurements) can be added by the user to further refine the results.

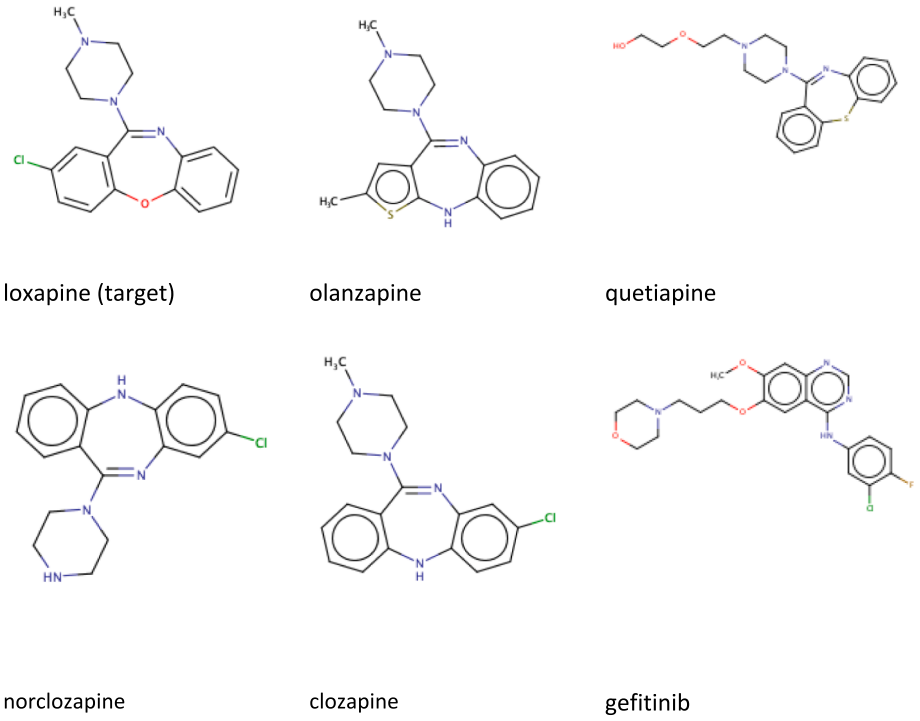


Fig. 7. Structure of target chemical, loxapine, and the chemicals identified as similar after refining by log P.

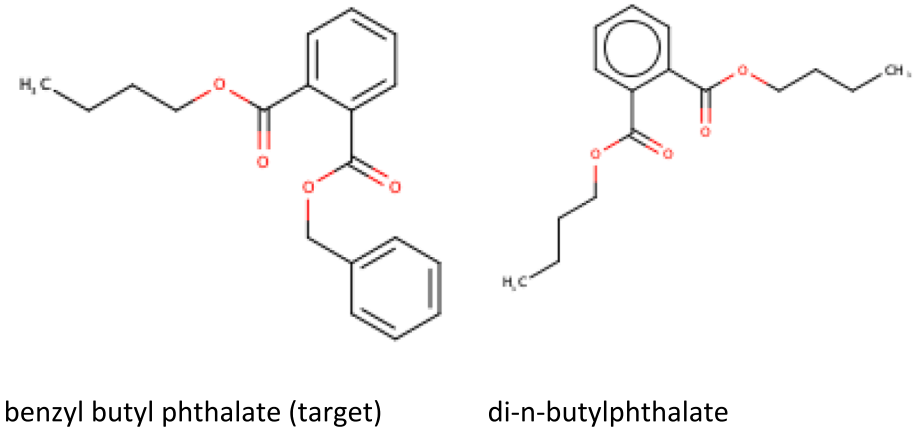


Fig. 8. Structure of target chemical, benzyl butyl phthalate, and the chemical identified as similar after refining by log D (at pH 7.4).

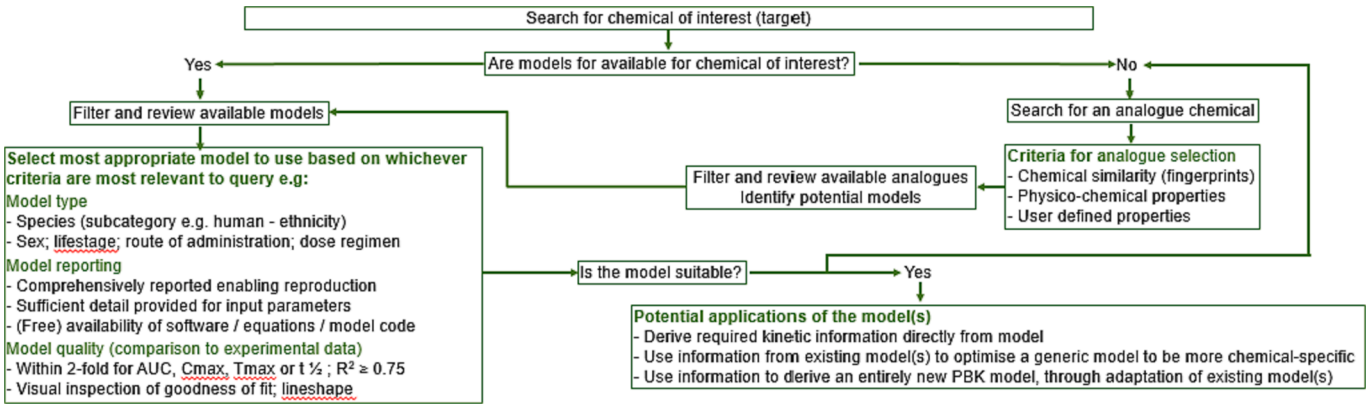


Fig. 9. Flow diagram showing criteria for model assessment and potential applications of the models.

5. Conclusions

Data required for safety assessment of chemicals is lacking in many cases, and computational methods are therefore needed to predict both effects and kinetic data. The KWAAS is a flexible tool, adaptable for individual user needs, that can help to automate and rationalise the process of analogue selection for read-across. Although applied here to PBK modelling information, similar principles can be applied to other read-across scenarios. The scope of the KWAAS could be further improved by populating the PMD with additional PBK models and including (accurate) predicted ADME properties that are used as input parameters in PBK modelling. It is essential that further case studies are carried out to demonstrate the utility and validity of the KWAAS approach in selecting analogues for read-across in PBK modelling. Two such case studies are described by Thompson et al. in a linked publication (Using Read-Across to Build Physiologically-Based Kinetic Models: Part 2. Case Studies for Atenolol and Flumioxazin; *submitted to Computational Toxicology)* [14]. The KWAAS, is freely available from (https://github.com/courtneythomp/KWAAS; accessed October 2023). A detailed user guide is given in the Supplementary Information for this article.

Funding.

Judith Madden and Courtney Thompson gratefully acknowledge the funding and scientific advice of the European Partnership for Alternative Approaches to animal testing (EPAA) with respect to the funding of a studentship for Courtney Thompson.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.comtox.2023.100292.

References

[1] T.W. Schultz, P. Amcoff, E. Berggren, F. Gautier, M. Klaric, D.J. Knight, C. Mahony, M. Schwarz, A. White, M.T. Cronin, A strategy for structuring and reporting a read-across prediction of toxicity, Regul Toxicol Pharmacol 72 (2015) 586–601.

[2] ECHA 2017. Read-across assessment framework (RAAF). Helsinki: European Chemicals Agency.

[3] J. Lu, M.R. Goldsmith, C.M. Grulke, D.T. Chang, R.D. Brooks, J.A. Leonard, M. B. Phillips, E.D. Hypes, M.J. Fair, R. Tornero-Velez, J. Johnson, C.C. Dary, Y. M. Tan, Developing a physiologically-based pharmacokinetic model knowledgebase in support of provisional model construction, PloS Comput Biol 12 (2016) e1004495.

[4] A. Paini, A. Worth, S. Kulkarni, D. Ebbrell, J. Madden, Assessment of the predictive capacity of a physiologically based kinetic model using a read-across approach, Comput Toxicol 18 (2021), 100159.

[5] C.V. Thompson, J.W. Firman, M.R. Goldsmith, C.M. Grulke, Y.M. Tan, A. Paini, P. E. Penson, R.R. Sayre, S. Webb, J.C. Madden, A systematic review of published physiologically-based kinetic models and an assessment of their chemical space coverage, Altern Lab Anim 49 (2021) 197–208.

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

[7] J.C. Madden, S.J. Enoch, A. Paini, M.T.D. Cronin, A review of in silico tools as alternatives to animal testing: principles, resources and applications, Altern Lab Anim 48 (2020) 146–172.

[8] S.J. Enoch, M.T.D. Cronin, J.C. Madden, M. Hewitt, Formation of structural categories to allow for read-across for teratogenicity, QSAR Comb Sci 28 (6–7) (2009) 696–708.

[9] J.C. Madden, G. Pawar, M.T.D. Cronin, S. Webb, Y.-M. Tan, A. Paini, In silico resources to assist in the development and evaluation of physiologically-based kinetic models, Comput Toxicol 11 (2019) 33–49.

[10] J.C. Madden, C.V. Thompson, in: In Silico Methods for Predicting Drug Toxicity, Benfenati E. (Ed) 2nd ed.:, Springer US, 2022, pp. 57–83.

[11] OECD. Guidance document on the characterisation, validation and reporting of Physiologically Based Kinetic (PBK) models for regulatory purposes. *OECD Series on Testing and Assessment.* No. 331, Environment, Health and Safety, Environment Directorate, OECD, Paris, 2021.

[12] WHO. 2010. *Characterization And Application of Physiologically Based Phamacokinetic Models in Risk Assessment* [Online]. Geneva, Switzerland: World Health Organisation. Available: https://apps.who.int/iris/handle/10665/44495 [Accessed October 2023].

[13] I.M. Sweeney, Case study on the impact of the source of metabolism parameters in next generation physiologically based pharmacokinetic models: Implications for occupational exposures to trimethylbenzenes, Regul. Toxicol. Pharmacol. 134 (2022), 105238.

[14] C. V. Thompson, S.D. Webb, J.A. Leedale, A. Paini, P. Penson, D. Ebbrell & J.C. Madden, 2023. Using Read-Across to Build Physiologically-Based Kinetic Models: Part 2. Case Studies for Atenolol and Flumioxazin (*submitted to Computational Toxicology*).