Sustainable Chemistry and Pharmacy 51 (2026) 102387

Contents lists available at ScienceDirect

Sustainable Chemistry and Pharmacy

journal homepage: www.elsevier.com/locate/scp

Predicting the biodegradability of active pharmaceutical ingredients using freely available *in silico* models: new application area or impossible?

Sabrina Groth, Klaus Kümmerer, Oliver Olsson *

Institute of Sustainable Chemistry, Leuphana University Lüneburg, 21335, Lüneburg, Germany

ARTICLE INFO

Keywords:
QSAR
In silico
Environmental biodegradability
Active pharmaceutical ingredients (API)
Closed-bottle-test (CBT)
OECD 301D

ABSTRACT

The development of environmentally friendly alternatives to pharmaceuticals and active pharmaceutical ingredients (APIs) is increasingly necessary, because pharmaceutical agents can persist in the environment, cause ecotoxicological effects, and require high biodegradability due to increasing regulatory requirements. Experimental classification into biodegradable or non-biodegradable according to OECD Guideline 301 is reliable but very time- and resource-intensive, hindering efficient pharmaceutical development. *In silico* models offer a promising alternative by predicting environmental biodegradability computationally. However, their suitability for APIs remains uncertain. This study compares the performance of different *in silico* models for determining biodegradability of APIs. using a test data set of 50 different pharmaceuticals and APIs with experimentally validated biodegradability data (OECD 301D). The applicability domains (AD) of the models were examined, and the performance parameters - accuracy, sensitivity, and specificity - were calculated.

Results: show that models predict non-biodegradable substances more reliably than biodegradable ones (higher specificity), indicating greater reliability in identifying persistent compounds. This suggests current models are better suited for early detection of potentially problematic, persistent APIs than for confidently confirming environmental safety. However, none of the models achieved sufficient accuracy for reliable predictions: all performance levels fell below the 80% threshold considered the minimum requirement for reliable *in silico* predictions under OECD Guideline 174 and ECHA practice. The current limitations highlight the urgent need for improved models with enhanced predictive power, particularly in accurately identifying biodegradable APIs. Future model development should focus on increasing accuracy, sensitivity, and robustness to support sustainable pharmaceutical design and regulatory decision-making.

1. Introduction

The principles of green chemistry and sustainability play an increasingly important role in the design and development of new pharmaceuticals (Anastas and Eghbail, 2010; Bertram et al., 2025). According to the "European Union Strategic Approach to Pharmaceuticals in the Environment", pharmaceuticals should be designed to be environmentally friendly (European Commission, 2019).

* Corresponding author.
E-mail addresses: sabrina.groth@leuphana.de (S. Groth), klaus.kuemmerer@leuphana.de (K. Kümmerer), oliver.olsson@leuphana.de (O. Olsson).

https://doi.org/10.1016/j.scp.2026.102387
Received 24 February 2026; Accepted 16 March 2026
Available online 18 March 2026

This primarily includes the Bening by Design (BbD) principle, in which environmental behavior, (eco)toxicity, and stability during storage and use should be taken into account during the development of new pharmaceuticals and new APIs (Kümmerer et al., 2020; Maertens and Plugge, 2018). BbD aims to minimize the environmental burden of pharmaceuticals by guiding molecular design toward inherent biodegradability, low bioaccumulation potential, and reduced ecotoxicity. Recent studies highlight that structural features such as aromatic rings, halogen substitutions, and high molecular weight are often associated with poor biodegradability and environmental persistence (Suk et al., 2023).

Accordingly, APIs should be designed from scratch or redesigned for full mineralization in the environment after they have delivered the wanted service and effects and following introduction into the environment. This means ensuring that APIs undergo complete mineralization either during wastewater treatment processes or within the natural environment. This aligns with the goal of creating safe and inherently sustainable chemical compounds, as outlined in the "Chemicals Strategy for Sustainability towards a Toxic-Free Environment" by the European Commission (European Commission, 2020; Kümmerer, 2019).

Biodegradability is a critical property influencing environmental fate and potential persistence of chemicals. Conducting experimental tests to determine biodegradability during the development phase of new APIs is very labor- and time-intensive, primarily due to the lack of standardized, high-throughput test systems. As a result, these assays are costly and significantly prolong development timelines (Cronin, 2004; Cronin and Madden, 2010; Dobbelaere et al., 2024). By using quantitative structure-activity analysis (QSAR), relevant endpoints, in this case biodegradability, can be predicted, eliminating time-and resources consuming laboratory work. They offer a very good opportunity to minimize in vitro tests. QSAR models assume that physio-chemical and environmental properties are linked to the chemical structure and molecular properties of a compound and are widely used to fill regulatory data gaps in registration dossiers (Pizzo et al., 2013). In the European Union, QSARs are explicitly permitted under the REACH regulation (EC No, 1907/2006; European Commission, 2006) for hazard and risk assessment, provided they are properly validated and applied within their applicability domain (AD). Similarly, the FDA and EMA recognize QSARs as integral components of the drug development process, especially in the context of environmental risk assessment and the implementation of the "Benign by Design" principle (OECD, 2024).

In silico calculated properties can inform chemical design by prioritizing the most promising or excluding most not wanted structures for further testing (Burello, 2015). Substances with complete biodegradability can thus be selected for further synthesis and testing (Lorenz et al., 2021; Amsel et al., 2024). Such (Quantitative) structure–biodegradability relationship models ((Q)SBR) can be applied to support the design of readily biodegradable and mineralizing APIs. (Q)SBR models help make better informed decisions in the design process prior to the synthesis of APIs, potentially saving time and resources (Puhlmann et al., 2021; van Dijk et al., 2022).

Currently, several different *in silico* tools for predicting the biodegradability of chemical substances are freely available online. Their performance has been evaluated across various chemical categories, including pharmaceuticals, cosmetics, and pesticides. For example, Sulmona et al. (2025) assessed biodegradability predictions for cosmetic ingredients, Zacarias et al. (2024) studied cosmetic formulations, and Muhammad et al. (2025) analyzed banned pesticides. However, the predictive accuracy of these models for APIs — a chemically diverse and environmentally relevant class — remains insufficiently validated, highlighting a critical gap in current *in silico* approaches.

Rücker and Kümmerer (2012) provided a comprehensive review of various *in silico* approaches for predicting aquatic aerobic biodegradation from a user's perspective. While these tools are occasionally applied in studies to assess biodegradability, their performance has not been systematically evaluated, particularly regarding their suitability for predicting ready biodegradability of APIs. As a result, there is currently no robust scientific evidence to support their use as reliable screening tools in the early design and environmental assessment of new APIs. This gap in validation is critical, as *in silico* models can only contribute meaningfully to sustainable pharmaceutical development if their predictive accuracy and AD are rigorously established (Rücker and Kümmerer, 2012).

Since then, new software packages and algorithms have been introduced. However, neither these newer tools nor the established models have been systematically evaluated for their suitability, validity, and performance in predicting the environmental biodegradability of APIs. This study aims to close this knowledge gap by conducting a comprehensive evaluation of available *in silico* models with respect to their applicability and predictive performance for APIs from selected, structurally and environmentally relevant classes. The focus is on non-steroidal anti-inflammatory drugs (NSAIDs), antibiotics, and key building blocks and intermediates of APIs—substance classes that are frequently detected in the environment, exhibit diverse chemical structures, and are of high regulatory and ecological concern. While data from other API classes are available, they have not yet been analyzed in this context. The selected classes serve as a representative and strategically relevant subset to assess model performance under real-world conditions and to inform future model development and application in pharmaceutical design.

2. Material and methods

2.1. General approach

First, a comprehensive literature and online search was conducted to identify freely available *in silico* models capable of predicting environmental biodegradability. This search identified nine publicly accessible tools that are currently used in environmental risk assessment and pharmaceutical development.

A test dataset was compiled comprising 50 APIs with diverse chemical structures and functional groups, for which experimentally determined biodegradability data were available according to OECD Guideline 301D (Closed-Bottle-Test; CBT). Biodegradability predictions for all 50 substances were then generated using each of the nine models. Prior to performance evaluation, the AD of each model was assessed to ensure predictions were made within the model's valid chemical space.

The predictive performance of each model was evaluated using standard metrics: accuracy, sensitivity, and specificity. These

parameters were calculated based on the comparison between the experimental (OECD 301D) and predicted biodegradability classifications. The results were then compared across models to assess their relative strengths and limitations in predicting the biodegradability of APIs.

2.2. Description of assessed in silico tools models

The comprehensive search focused on tools that are publicly accessible, validated in peer-reviewed studies, and used in regulatory or pharmaceutical contexts. From the initial pool, nine models were selected based on the following criteria: (1) availability as open-source or free-to-use software, (2) documentation of validation and AD, (3) use in environmental risk assessment or pharmaceutical development, (4) ability to predict ready biodegradability (OECD 301D-like endpoint), and (5) binary classification output (readily biodegradable vs. not readily biodegradable). The selected models—BIOWIN 1/2/5/6/Consensus, VEGA, OCHEM, OPERA, TOXTREE, and the model published by Lunghini et al.—represent a diverse range of approaches, including QSAR, machine learning, and expert systems, ensuring a comprehensive evaluation of current *in silico* capabilities. Further information on the parameters of the selected *in silico* models such as model type, source of training data, and availability can be found in the Supplementary data Table S1.

This study focuses on the prediction of aerobic biodegradability of APIs in the aquatic environment, specifically in water, activated sludge, and sediment. To assess the performance of *in silico* models, this study evaluates nine freely available tools that predict aerobic biodegradability under conditions relevant to these environments. The models were selected based on their applicability to aqueous systems and their use of OECD 301 series test data (e.g., OECD 301D, 301B, 301C). The BIOWIN 7 model, which predicts anaerobic biodegradability, is excluded as it does not align with the study's focus on aerobic processes. Table 1 provides an overview of this.

2.2.1. BIOWIN (EPI Suite™ 4.11)

BIOWIN is part of EPI (Estimation Programs Interface) Suite™, a suite of physical and chemical property and environmental fate estimation programs, which was developed by EPA's and Syracuse Research Corporation (SRC). BIOWIN is a screening-level tool and estimates aerobic and anaerobic biodegradability of organic chemicals using seven different models (BIOWIN 1 - 7) (Boethling et al., 2003).

The BIOWIN 1 and 2 models make predictions about aerobic biodegradability based on multiple linear or non-linear regression analyses (Howard et al., 1992). The data for the experimental determination of biodegradability were taken from the SRC database (Howard et al., 1987). The data set contains 295 chemicals whose biodegradability was assessed in binary terms: 1 (the substance is rapidly biodegradable, probability 1.0) or 0 (the substance is not rapidly biodegradable, probability 0.0). Thus, 186 chemicals were classified as “biodegrades fast” and 109 as “does not biodegrade fast”. From the data set, 36 chemical structural features/functional groups were identified that have a potential effect on biodegradability. To determine the probability of degradation, coefficients were determined for the individual fragments using several linear or non-linear regression analyses. A degradation probability of more than 0.5 is considered "rapidly biodegradable". A degradation probability of less than 0.5 is considered "not rapidly biodegradable". Molecular weight (MW) was defined as another descriptor. Apart from the minimum and maximum value for the MW of a chemical substance, a formal AD is not defined. Crucially, the validation was performed on an independent test set of industrial chemicals, with no inclusion of APIs. The model was last updated in 2003 and has not been revised since, despite significant advances in computational modeling and the increasing importance of pharmaceuticals in environmental risk assessment. Given the structural complexity, high molecular weight, and functional group diversity of APIs, it is highly likely that many pharmaceuticals fall outside the model's AD (EPA – United States Environmental Protection Agency, 2012; Howard et al., 1992).

In contrast to the BIOWIN 1 and 2 models, which provide a binary classification (i.e., a yes/no answer: “readily biodegradable” or “not readily biodegradable”), the BIOWIN 3 and 4 models predict the time required for biodegradation (e.g., in days, weeks, or months) or the degree of recalcitrance (resistance to degradation). This means they do not assign a clear category but instead return a continuous value (e.g., “degrades in 14 days” or “highly recalcitrant”) (Boethling et al., 1994). Because these models do not produce a binary outcome, they are not directly comparable to the experimental data (OECD 301D). Furthermore, the results from BIOWIN 3 and 4 are highly variable and not standardized, making it difficult to assess their performance consistently across different substances. For these reasons, BIOWIN 3 and 4 were excluded from this study, which focuses on binary classification models that can be directly compared to the OECD 301D test.

The BIOWIN 5 and 6 models use a regression-based approach to predict ready biodegradability, similar to BIOWIN 1 and 2.

Table 1
Overview of selected models, predictions, relevant environment, primary test method and source.

Model	Predictions	Relevant Environment	Primary Test Method (OECD)	Source
BIOWIN 1 & 2	Aerobic biodegradability	Water, Activated sludge	OECD 301D (Closed Bottle Test)	Boethling et al. (2003)
BIOWIN 3 & 4	Time to biodegradation	Water, Activated sludge	OECD 301D, 301B	Boethling et al. (1994)
BIOWIN 5 & 6	Aerobic biodegradability	Activated sludge	OECD 301C (MITI)	Tunkel et al. (2000)
BIOWIN Consensus	Aerobic biodegradability	Water, Activated sludge	OECD 301D, 301C	Boethling et al. (2003)
VEGA	Ready biodegradability	Water, Activated sludge	OECD 301C (MITI)	Ferrari et al. (2013)
OCHEM (Consensus)	Ready biodegradability	Water, Activated sludge	OECD 301 series	Vorberg and Tetko (2014)
OPEn (q)saR App (RBioDeg)	Ready biodegradability	Water, Activated sludge	OECD 301 series	Mansouri et al. (2018)
TOXTREE	Ready biodegradability	Water, Activated sludge	OECD 301 series (301D, 301C)	Patlewicz et al. (2008)
Lunghini et al. (2020)	Ready biodegradability	Water, Activated sludge	OECD 301 series	Lunghini et al. (2020)

However, the model type (linear or non-linear) is not chosen by the user, but determined during model development based on the best fit to the training data. In this case, non-linear regression was used to capture complex structure–activity relationships (Tunkel et al., 2000). The difference is in the experimental test used to measure the biodegradability data. This was determined for 884 chemicals using the Japanese MITI (Ministry of International Trade and Industry) OECD 301C and ready biodegradation tests. This dataset consists of 385 chemicals that were classified as "readily degradable" and 499 chemicals that were classified as "not readily degradable". A substance was classified as readily biodegradable if it achieved ≥60 % degradation within 10 days; otherwise, it was classified as not readily biodegradable. From this data set, which was divided into a training and a test data set, the probability values of individual fragments for biodegradability were derived (1: readily degradable or 0: not readily degradable). From this, 42 structural fragments and molecular weight were used as descriptors. Notably, no formal AD is defined for the model. While a molecular weight range (100–600 g/mol) is mentioned, no chemical space or structural boundaries are specified, limiting the reliability of predictions for compounds outside this range (Tunkel et al., 2000).

In addition to the results on biodegradability according to models 1 to 7, BIOWIN also provides a statement predicting ready biodegradability (yes or no), based on the results of models 3 and 5. If the result of BIOWIN 3 is “weeks” or faster and the probability of BIOWIN 5 is ≥ 0.5, the prediction is “Readily biodegradable”. This model is called BIOWIN Consensus model. However, this criterion does not align with the OECD nomenclature, which defines ready biodegradability strictly as ≥60 % degradation within 10 days. The term “weeks” is ambiguous and not standardized in OECD guidelines (Boethling et al., 2003). Therefore, the BIOWIN Consensus model's classification may not be directly comparable to experimental results from OECD 301 tests, and its use for regulatory purposes should be interpreted with caution.

2.2.2. VEGA 1.2.4

VEGA is a more complex platform compared to BIOWIN that includes 113 QSAR models of different endpoints such as carcinogenicity, fish acute toxicity, algae acute toxicity, or persistence of chemicals in the environment. The ready biodegradability model (IRFMN) 1.0.10 provides a qualitative evaluation in the form of a binary classification to predict biodegradability and is based on the OECD 301C modified MITI test data (I): the compound is considered readily biodegradable if it achieves ≥60 % degradation within the first 10 days of a 28-day test, in accordance with OECD Guideline 301. This criterion is based on the requirement that a substance must show rapid degradation, not just overall degradation over the full test period (Ferrari et al., 2013; OECD, 1992).

For this purpose, a data set with 728 chemicals was used (Toropov et al., 2012), that were extracted from the OECD QSAR toolbox v 2.0 and from the BIOWIN 5 and 6 models and then combined. From this, a training data set with 582 chemicals and a test data set with 146 chemicals were created. From this data on biodegradability, 78 alerts, i.e. fragments that have been found to be related to biodegradability, (non-readily biodegradable, possible non-biodegradable, possible biodegradable, readily biodegradable) were identified. The structural alerts of the model were produced using the SARpy software (Ferrari et al., 2013). In addition to the prediction of biodegradability, the reliability of the prediction is given; this model tests the AD. An independent algorithm is used for this, which shows similar compounds, evaluates the QSAR results for the similar compounds and analyzes some relevant chemical features in the target compound and its related compounds. An AD index value of 0.65 or greater indicates a reliable prediction (Lombardo et al., 2014; Pizzo et al., 2013).

2.2.3. OCHEM

The Online Chemical Modeling Environment (OCHEM) is a web-based platform – available at http://www.ochem.eu and was developed by Vorberg and Tetko (2014). It consists of two main subsystems: a database with experimental measurements and a modeling framework. The database currently contains 10,817 biodegradability data - primarily derived from OECD 301 series tests, including OECD 301D (CBT), 301B (CO_2 Evolution Test), 301C (MITI (I), 301E (Modified MITI), and 301F (Manometric Respirometry; MRT) – which were entered by four different authors. There are eight chemical descriptor-based machine learning models for predicting biodegradability (classification into biodegradable/not readily biodegradable). A chemical is classified as readily biodegradable if it has a mineralization degree of over 70% within 28 days.

The Consensus Ready Biodegradability model, which combines predictions from seven individual models, was trained on a dataset of 1884 chemicals, integrating data from multiple sources to ensure broad applicability: This includes the Boethling data set (63 pharmaceuticals), and data from the CADASTER database (1400 entries), from Cheng et al. (2012), and the BIOWIN dataset (Tunkel et al., 2000), as well as data from Gramatica et al. (Gramatica and Papa, 2007; Gramatica et al., 2016).

The Boethling test dataset contains 63 drugs - such as ibuprofen, diclofenac, and ciprofloxacin - the experimental data were collected using different test methods from the OECD test series 301 (B, D, E, F), according to FDA 3.11 28d or according to ISO 9408. Importantly, the other datasets used to train the OCHEM consensus model also contain a significant proportion of APIs. The CADASTER dataset includes pharmaceuticals, pesticides, and industrial chemicals, with approximately 30–40 % being APIs. The Cheng et al. (2012) dataset, derived from the Japanese NITE database, contains numerous pharmaceuticals alongside other organic compounds. The BIOWIN dataset (Tunkel et al., 2000) is predominantly composed of APIs, including NSAIDs, antibiotics, and antidepressants. Additionally, data from Gramatica et al. include a substantial number of pharmaceuticals.

In order to be able to specify the reliability of the predictions, determination of the AD is also carried out in this model. In order to distinguish reliable from unreliable predictions the model includes an estimation whether the compounds tested fall into the AD, it is based on the standard deviation of the predictions of the individual models. If different models give significantly different outcomes for a compound, the consensus prediction is more likely to be unreliable. The particular compound is subsequently flagged as out of domain (Vorberg and Tetko, 2014).

2.2.4. OPERA 2.9

The OPEn (q)saR App (OPERA) was developed by Mansouri et al. (2018) and contains a variety of QSAR/QSPR (quantitative structure pharmacokinetic relationship) models for chemical properties like structural properties, physicochemical properties, environmental fate, toxicology endpoints, and ADME (absorption, distribution, metabolism, excretion) endpoints (Mansouri et al., 2018). The RBioDeg model is a binary model with the endpoint ready biodegradability. The descriptor-based software uses a k-nearest-neighbor (kNN) algorithm to assign tested compounds to the discrete classes ready or not-readily biodegradable. The model was trained on a curated dataset of 1609 chemicals, originally sourced from the PHYSPROP database, which compiled experimental biodegradability data from multiple sources, including OECD 301 series tests (e.g., CBT, MRT). A substance was classified as readily biodegradable if its biochemical oxygen demand (BOD) exceeded 60 % within 28 days. Although the PHYSPROP database is no longer publicly available and has since been integrated into EPI-Suite™, the model's training data and methodology are fully documented in the original publication (Mansouri et al., 2018) and the QMRF (Q17-23a-0014) (European Commission, 2017). For this study, the model's performance was evaluated using the same dataset, and its AD was assessed to ensure predictions were made within the model's valid chemical space. The model is applicable to a wide range of organic compounds, including those with diverse functional groups and molecular weights.

The software includes an automated check whether a chemical is within the AD of the model. According to the authors, the RBioDeg model is applicable to a wide range of organic compounds with diverse chemical structures and physicochemical properties. The training set includes substances with molecular weights between 100 and 600 g/mol, and a broad variety of functional groups, including aromatic rings, halogens, amides, esters, sulfonamides, and alcohols. The model was developed on a dataset encompassing pharmaceuticals, pesticides, and industrial chemicals, reflecting its suitability for structurally diverse and environmentally relevant compounds (Mansouri et al., 2018).

2.2.5. TOXTREE 3.1.0

Toxtree (Toxic Hazard Estimation by Decision Tree Approach) was developed by Molecular Networks GmbH and is an application that uses a decision tree approach to estimate toxic hazards in various modules. The START (Structural Alerts for Reactivity in Toxtree) module is a plug-in for biodegradation and persistence. It is based on a compilation of structural alerts for environmental persistence and biodegradability. These are molecular functional groups or substructures known to be associated with the biodegradability of chemicals. These structural alerts are used in logical decision trees to predict potential persistence or biodegradability. An evaluation of the implementation of the Cramer classification scheme is included in the Toxtree software (Patlewicz et al., 2008). The definition of persistence applied is according to the Persistence and Bioaccumulation Regulation of the Government of Canada from the year 2000. Thereby a chemical is defined as persistent in water when its half-life is equal or greater than 182 days (Branch, 2006). This half-life refers to the time required for the primary degradation (i.e., 50 % reduction in the concentration of the parent compound), not mineralization.

Toxtree uses 32 structural alerts, 23 of which relate to environmentally persistent and nine to readily biodegradable structural elements. Compounds are classified into one of the following categories depending on whether they contain persistent or readily biodegradable structural alerts: Class 1 – readily biodegradable chemical; Class 2 – persistent chemical; or Class 3 – unknown biodegradability if no structural alerts are detected (Molecular Networks GmbH Computerchemie, 2008).

2.2.6. Lunghini et al

The model developed by Lunghini et al. is the most recent model of the ones investigated here (Lunghini et al., 2020). It is a consensus model of 15 individual models making use of different machine learning approaches, such as support vector machine (SVM), random forest (RF), and naïve Bayesian (NB) algorithms. The models classify chemicals into readily or not readily biodegradable categories.

The data to build the models is derived from a variety of sources, including the training sets of existing tools (VEGA, EPI Suite, and OPERA), available databases (NITE, European Chemicals Agency (ECHA)), and literature sets (Cheng et al., 2012; Mansouri et al., 2018). In addition, an industrial partner provided a confidential data set. The data from all sources were merged in a “global” data set, containing 3146 chemicals. All data originated from tests of the OECD 301 series including OECD 301B, C, D, F. To represent the chemicals *in silico*, 19 descriptor sets with the best performances were selected. Detailed information on the algorithms and descriptor spaces can be found in the publication and its supplemental material (Lunghini et al., 2020).

The AD of each model was assessed using fragment-based similarity criteria: a test compound was considered to be within the AD only if all of its molecular fragments were present in at least one molecule from the training set. If a fragment was not encountered in any training compound, the molecule was classified as outside the AD. For the consensus prediction, the final classification was determined by majority voting among the individual models. A prediction was considered reliable only if the majority vote exceeded 60 % or was below 40 %. Predictions falling within the 40–60 % range were deemed unreliable, as they indicate a lack of consensus and may reflect random or ambiguous model behavior. Furthermore, predictions from models where the test compound was outside the AD were excluded from the consensus. The overall consensus result was considered unreliable if the AD was satisfied for 25 % or fewer of the individual models.

2.3. Selection of experimental ready biodegradability data for the test data set

From the Institute of Sustainable Chemistry (INSC) at Leuphana University, data (Olsson et al., 2024) based on the in-house OECD 301D biodegradation experiment was provided and used for the setup of the test data set (Table 2). The selection of APIs was based on a

systematic and transparent process to ensure a representative, diverse, and environmentally relevant test dataset. The following criteria were applied in hierarchical order: (1) availability of experimental biodegradability data, (2) structural diversity: the dataset was designed to cover a wide range of chemical structures and functional groups, such as aromatic rings, halogen substitutions (Cl, F, Br), amide, ester, and sulfonamide groups, and (3) therapeutic and environmental relevance.

The herein used dataset included 50 APIs of different classes such as antibiotics, NSAR, various agonists, sedatives, anticholinergics, cytostatics, and analgetics, as well as pharmaceuticals in the treatment of different diseases (psoriasis, Parkinson's, epilepsy, dementia, etc.), as well as building blocks of APIs and synthesis intermediates. When selecting the substances, care was taken to ensure a balanced ratio between readily biodegradable (RB) and non-biodegradable (nRB) substances. In addition, care was taken to ensure that APIs with the same structural features or functional groups are not clustered in the test data set. Compound specific information on the individual substances: name, SMILES code, CAS number, associated group and the result of the experimental determination of the CBT as a mean value as well as an indication of biodegradability (yes/no) was collected. The dataset comprises 50 APIs; however, a larger dataset could have provided more statistical power, the focus was on representative, high-quality data rather than sheer quantity. The use of 50 APIs ensures a balanced assessment of model performance across diverse chemical structures and functional groups.

2.4. Model performance assessment

For the performance assessment, the AD was first evaluated. For every substance in the test dataset, it was checked whether the compound fell within the chemical space covered by the model's training data. Only predictions made within the AD were considered valid for performance evaluation. Substances outside the AD were excluded from the calculation of accuracy, sensitivity, and specificity to avoid biased or unreliable results.

Because of the great vastness and diversity of chemical substances, a QSAR model is not expected to reliably predict the nature of every chemical. The reliability of a prediction made is estimated by its AD. The AD is the region in chemical space a QSAR model is expected to interpolate data correctly and thus make reliable predictions for. Through the model descriptors, it is defined by the nature of the chemicals in the training set, and it is assumed, that predictions for compounds not similar enough to the training set are deemed as unreliable (Rücker and Kümmerer, 2012; Gramatica and Papa, 2007). According to the OECD principle 3 for the validation for regulatory purposes of QSAR models, any QSAR model used for regulatory purposes is required to have a defined AD (OECD, 2004).

To determine the prediction of the biodegradability of the substances belonging to the validation data set, the structural formula of the substance to be analyzed was entered into the respective models as a SMILES code or sdfile. The software OpenBabel 2.3.0 was used to convert the SMILES codes into the sdfile. The predictions of the models were then summarized and compared with the test data set (Table 2). The results for each model were divided into the number of true predictions (true positive (TP) or true negative (TN)) and false predictions (false negative (FP) or false negative (FN)). Positive results in this case mean that the substance is biodegradable. Negative results therefore mean that the substance is not biodegradable.

The following parameters were determined for the validation of the individual models: Accuracy - the total number of correct predictions; Sensitivity - the number of correctly predicted positive results; and Specificity - the number of correctly predicted negative results.

The equations are defined as follows:

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3}$$

For the performance evaluation, every prediction made by the models was considered, even if it was marked as out of scope. If a model did not provide a prediction for a query compound, it was omitted.

3. Results and discussion

3.1. Testing the applicability domain and reliability of predictions

The results in Table 3 show that four of the nine tested models have both a defined AD and a mechanism that checks whether the substance to be predicted is in the AD or not. These include the models VEGA, OCHEM, OPERA, and the model described by Lunghini et al. The Toxtree model also performs a check of the AD. However, in this case, the chemicals being tested are classified as "unknown biodegradability" unless structural alerts stored in the database are detected (Molecular Networks GmbH Computerchemie, 2008). In contrast, the BIOWIN models do not describe a defined AD. It is therefore emphasized by the authors that the user of the software decides whether the prediction for the tested substance is considered reliable (Tunkel et al., 2000). For this, ECHA gives some recommendations to the user to identify if a chemical is inside the AD of a certain model (European Chemicals Agency, 2016).

The presented results also show the number of substances defined as within or outside the AD or that could not be classified by the respective models (Table 3). The prediction of biodegradability by the *in silico* models proved unreliable for some substances. These

Table 2
Detailed information about the test data set: substance names, SMILES Code, CAS number, group description, and biodegradation. RB: readily biodegradable, N: no, Y: yes.

No.	Substance	SMILES Code	CAS No.	Group	RB (CBT)	MV (CBT)
1	2-Trifluoromethylphenol	OC1=CC=CC=C1C(F)(F)F	444-30-4	building block of APIs	Y	64.9
2	2-(4-Methylphenyl)propanoic acid	Cc1ccc(cc1)C(C)C(=O)O	938-94-3	building block of APIs	Y	70.0
3	2,4-Dihydroxyquinoline	c1ccc2c(c1)c(cc(=O)[nH]2)O	86-95-3	building block of antibiotics	Y	74.6
4	2-Hydroxyphenylacetic acid	O=C(O)Cc1ccccc1O	614-75-5	intermediate in the synthesis	Y	86.0
5	2-Quinolinol	c1ccc2c(c1)ccc(n2)O	59-31-4	degradation product of quinolones (e.g. ciprofloxacin)	Y	65.0
6	3-(2-Amino-1-hydroxyethyl) phenol hydrochloride	c1cc(cc(c1)O)C(CN)O.Cl	4779-94-6	adrenergic alpha-agonist	Y	74.0
7	3,4-Dihydroxyphenylacetic acid	O=C(O)Cc1ccc(O)c(O)c1	102-32-9	metabolite of APIs	Y	80.0
8	3-Fluoroaniline	C1=CC(=CC(=C1)F)N	372-19-0	building block of APIs	N	6.8
9	4-Biphenylacetic acid	C1=CC=C(C=C1)C2=CC=C(C=C2)CC(=O)O	5728-52-9	NSAR	N	−3.2
10	4-Fluoroaniline	Nc1ccc(F)cc1	371-40-4	building block of APIs	Y	62.0
11	4-Fluorophenol	C1=CC(=CC=C1O)F	371-41-5	building block of APIs	N	46.0
12	4-Hydroxypyridine	O = c1cc[nH]cc1	626-64-2	building block of antibiotics	Y	65.0
13	4-(Trifluoromethyl)-aniline	C1=CC(=CC=C1C(F)(F)F)N	455-14-1	building block of APIs	Y	64.5
14	Allantoin	C1(C(=NC(=N1)O)O)NC(=N)O	97-59-6	keratolytic	N	39.0
15	Amoxicillin	O=C(O)[C@@H]2N3C(=O)[C@@H](NC(=O)[C@@H](c1ccc(O)cc1)N)[C@H]3SC2(C)C	26787-78-0	antibiotic	N	−1.0
16	L-Arginine	C(CC(C(=O)O)N)CNC(=N)N	74-79-3	vascular	Y	78.1
17	Aspirin	CC(=O)Oc1ccccc1C(=O)O	50-78-2	NSAR	Y	71.0
18	Atenolol	CC(C)NCC(O)COc1ccc(CC(N)=O)cc1	29122-68-7	beta-blockers	N	1.0
19	Benzylamine	C1=CC=C(C=C1)CN	100-46-9	local analgesic, anti-inflammatory	Y	82.9
20	Benzol	C1=CC=CC=C1	71-43-2	building block; intermediate product	Y	99.5
21	Butyric acid	CCCC(=O)O	107-92-6	histamine antagonist	Y	69.0
22	Carbamazepine	NC(=O)N1c2ccccc2C=Cc2ccccc21	298-46-4	anticonvulsant	N	0.0
23	Ciprofloxacin	C1CC1N2C=C(C(=O)C3=CC(=C(C=C32)N4CCNCC4)F)C(=O)O	85721-33-1	antibiotic	N	2.0
24	(−)-Cytisine	c1cc2n(c(=O)c1)C[C@H]3C[C@@H]2CNC3	485-35-8	acetylcholine receptor agonist	N	3.0
25	Diclofenac	O=C(O)Cc1ccccc1Nc2c(Cl)cccc2Cl	15307-86-5	NSAR	N	14.0
26	Disodium fumarate	C(=CC(=O)[O-])C(=O)[O-].[Na+].[Na+]	17013-01-3	API (psoriasis)	Y	87.0
27	Fosfomycin Disodium Salt	CC1C(O1)P(=O)([O-])[O-].[Na+].[Na+]	26016-99-9	antibiotic	N	−1.5
28	Genistein	O = c1c(-c2ccc(O)cc2)coc2cc(O)cc(O)c12	446-72-0	phytoestrogen	Y	76.7
29	Huperzine A	C/C=C/1\[C@@H]2CC3=C([C@]1(CC(=C2)C)N)C=CC(=O)N3	102518-79-6	antidementia	N	−11.6
30	Hydroxyurea	C(=O)(N)NO	127-07-1	cytostatic	N	7.1
31	Ibuprofen	CC(C)Cc1ccc(cc1)C(C)C(=O)O	15687-27-1	NSAR	N	7.9
32	Lamivudine	NC1=NC(=O)N(C=C1)[C@@H]2CS[C@H](CO)O2	134678-17-4	nucleoside reverse transcriptase inhibitor (NRTI)	N	−1.9
33	Levetiracetam	CCC(C(=O)N)N1CCCC1=O	102767-28-2	antiepileptic	N	−0.8
34	Menthol	C[C@@H]1CC[C@H]([C@@H](C1)O)C(C)C	1490-04-6		N	−6.1
35	Metamizole	CC1=C(C(=O)N(N1C)C2=CC=CC=C2)N(C)CS(=O)(=O)O	5907-38-0	non-opioid analgesic, antipyretic	N	11.3
36	Methacholine chloride	CC(C[N+](C)(C)C)OC(=O)C.[Cl-]	62-51-1	muscarinic receptor agonists	N	11.1
37	Methacryloylcholine methyl sulfate	CC(=C)C(=O)OCC[N+](C)(C)C·COS(=O)(=O)[O-]	6891-44-7	monomer of polyquaternium-14	Y	83.9
38	S-Nicotine	CN1CCCC1c1cccnc1	54-11-5	nicotinic acetylcholine receptor agonist	Y	75.0
39	Papaverine hydrochloride	COc1ccc(cc1OC)Cc2c3cc(c(cc3ccn2)OC)OC.Cl	61-25-6	nonxanthine phosphodiesterase inhibitor	N	41.0
40	Paracetamol	CC(=O)Nc1ccc(cc1)O	103-90-2	NSAR	N	−5.3
41	Sanilvudine	Cc1cn(c(=O)[nH]c1 = O)[C@H]2C=C[C@H](O2)CO	3056-17-5	nucleoside reverse transcriptase inhibitor (NRTI)	Y	65.7
42	(−)-Scopolamine methyl bromide	[Br-].C[N+]1(C)[C@H]2CC(C[C@H]1[C@H]3O[C@@H]23)OC(=O)C(CO)c4ccccc4	155-41-9	anticholinergic	N	6.0

(*continued on next page*)

Table 2 (*continued*)

No.	Substance	SMILES Code	CAS No.	Group	RB (CBT)	MV (CBT)
43	Sodium lauryl sulfate	CCCCCCCCCCCCOS(=O)(=O)[O-].[Na+]	151-21-3	emulsifier; antibacterial, antiviral	Y	73.9
44	Sorbic acid	C/C=C/C=C/C(=O)O	110-44-1	laxative	Y	71.0
45	Sulfamethoxazole	Nc1ccc(cc1)S(=O)(=O)Nc2cc(C)on2	723-46-6	antibiotic	N	5.1
46	Thalidomide	c1ccc2c(c1)C(=O)N(C2=O)C3CCC(=NC3=O)O	50-35-1	sedative, hypnotic, immunosuppressant	N	5.0
47	o-Tolylic acid	Cc1ccccc1C(=O)O	118-90-1	bacteriostatic	Y	84.3
48	Tropolone	c1ccc(c(=O)cc1)O	533-75-5	catechol-O-methyltransferase (COMT) inhibitor	Y	71.1
49	Valeric acid	CCCCC(=O)O	109-52-4	sedative, hypnotic	Y	79.0
50	Zidovudine	CC1=CN(C(=O)NC1=O)[C@H]2C[C@@H]([C@H](O2)CO)N = [N+] = [N-]	30516-87-1	nucleoside reverse transcriptase inhibitor (NRTI)	N	4.4

Table 3
Evaluation of the applicability domain using test data set.

Model	Check of Applicability Domain	Number of chemicals inside domain	Number of chemicals Out of domain	Number of chemicals Not classifiable
BIOWIN1	No	-	-	-
BIOWIN2	No	-	-	-
BIOWIN5	No	-	-	-
BIOWIN6	No	-	-	-
BIOWIN Consensus	No	-	-	-
VEGA	Yes	10	36	4
OCHEM	Yes	44	6	0
OPERA	Yes	49	1	0
TOXTREE	(Yes)	45	0	5
Lunghini et al.	Yes	41	9	0

were therefore marked as "out of domain". There was no agreement, i.e., the substances assessed as out of domain by each model were different. This API, used as a cytostatic, is a structural analogue of folic acid. Due to the high mass of the molecule (454 g/mol) and the nitrogen-containing aromatics, it is likely not included in the training data sets of the various *in silico* models. Substances containing oxygen, nitrogen or sulfur-containing heterocyclic components are also often classified as out of domain. Presumably because substances in the training data sets have less complex functional groups. The VEGA model names the most out of domain with 37 substances and OPERA the fewest with one substance. In addition, some substances could not be classified because the model did not provide structural information, four in VEGA and three in Toxtree, respectively. These are designated as "not classifiable." In the Toxtree model, these substances fall into the category of "unknown biodegradability".

3.2. Prediction results and performance assessment

Table 4 shows the results of the individual model performance in relation to the prediction of biodegradability. Columns 2 and 3 indicate how many of the 50 substances analyzed were classified as biodegradable or as non-biodegradable. According to experimental determination using CBT, half of these substances were biodegradable and half were non-biodegradable.

It is noteworthy that the model developed by Lunghini et al. (2020) achieves a balanced prediction, classifying approximately 50 % of the substances as readily biodegradable - matching the experimental CBT data. This balance is significant because most other models

Table 4
Listing of the models according to the prediction of biodegradability. RB: readily biodegradable, TP: true positives, FP: false positives, TN: true negatives, FN: false negatives.

Model	RB	not RB	TP	FP	TN	FN
BIOWIN1	37	13	20	17	8	5
BIOWIN2	35	15	20	15	10	5
BIOWIN5	13	35	12	1	24	13
BIOWIN6	17	33	13	4	21	12
BIOWIN Consensus	12	38	12	0	25	13
VEGA	8	2	8	0	1	1
OCHEM	17	27	14	3	19	8
OPERA	32	17	21	11	14	3
TOXTREE	17	28	11	6	18	10
Lunghini et al.	21	20	17	4	14	6

show systematic bias: BIOWIN 1 and 2 are overly optimistic (70 % predicted as biodegradable), while BIOWIN 5, 6, and the Consensus model are overly pessimistic (30 % or less). The VEGA and OPERA models also classify most substances as biodegradable, whereas the OCHEM and Toxtree models are rather conservative. All models, apart from VEGA, were able to predict the biodegradability of either all or most of the substances to be classified in the dataset.

The VEGA model, due to its strict AD, could only classify 9 of the 50 substances, limiting its utility. In contrast, the balanced performance of the Lunghini model—based on a consensus of 15 machine learning models trained on diverse OECD 301 data—suggests a more robust and reliable prediction, particularly for early-stage pharmaceutical development where unbiased screening is critical.

Columns 4 and 5 of Table 4 show how many of the substances classified as biodegradable were correctly identified (true positive) or incorrectly classified (false positive). Regarding the non-biodegradable substances, columns 6 and 7 show how many have been correctly identified (true negative) or misclassified (false negative). Based on these values, the parameters sensitivity, specificity, and accuracy are calculated (see formulas 1-3, Section 2.3). This allows statements to be made about the performance of the individual models. The results of the performance assessment of the individual *in silico* models are shown in Fig. 1.

In terms of sensitivity, i.e. the correct prediction of biodegradable substances, the values in the models ranged from 0.48 to 0.89 (Fig. 1). The models VEGA as well as OPERA have with the values 0.89 and 0.88 a higher sensitivity than the other models. This suggests that these models are good at predicting biodegradable substances.

The models exhibit a specificity, i.e. the correct prediction of non-biodegradable substances, in a wide range with values from 0.32 to 1.0. The models BIOWIN Consensus, VEGA, and BIOWIN 5 have the highest specificities in comparison. The values are 1.0, 1.0 and 0.96. While these values suggest excellent performance in identifying non-biodegradable compounds, a specificity of 1.0 should be interpreted with caution. Such a high value may reflect a small number of non-biodegradable substances in the test set, or a limited number of predictions, rather than a universally robust model. In such cases, a specificity of 1.0 does not necessarily indicate superior predictive power but may result from low statistical power or data imbalance. Therefore, while these models perform well in this dataset, their generalizability to larger, more diverse datasets remains to be validated.

The sensitivity and specificity values can be presented together in a diagram (Fig. 2), forming two clusters. This type of representation shows a clustering of models, which underscores their similar performance. One cluster consists of the BIOWIN Consensus, BIOWIN 5 and 6, OCHEM, Toxtree, and Lunghini et al. models, which are characterized by higher specificity combined with lower sensitivity. A second cluster, formed by the OPERA and BIOWIN 1 and 2 models, is characterized by higher sensitivity values and lower specificity values. The results illustrate that the models can provide either accurate predictions for readily biodegradable APIs or accurate predictions for not readily biodegradable APIs. The only model that can provide reliable predictions for both is the VEGA model. In comparison, it showed the highest levels of both sensitivity and specificity.

In terms of accuracy, the tested models performed differently with rates of correct predictions between 0.56 and 0.90. The two BIOWIN models 1 and 2 have the lowest accuracy values in comparison. This value results from the low specificity of only 0.32 or 0.40. The VEGA model has the highest accuracy compared to the other models. However, this value is not very meaningful, as it is based on the correct prediction of only ten APIs. Of the 50 APIs to be determined, 40 were outside the AD of this model or could not be classified. The significance of these results is considered low due to the small number of substances that were within the AD and could therefore be evaluated. Therefore, VEGA is only of limited use for predicting the biodegradability of APIs. When applying the BIOWIN models, an

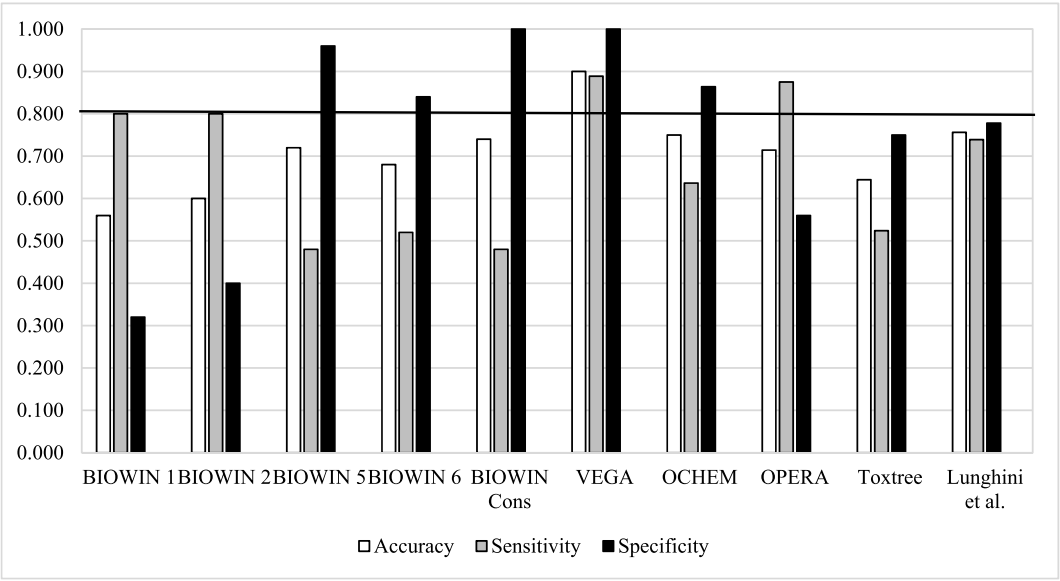

Fig. 1. Calculated accuracy, sensitivity, and specificity values. 0.8: minimum requirement for reliable *in silico* predictions under OECD Guideline 174 and ECHA practice.

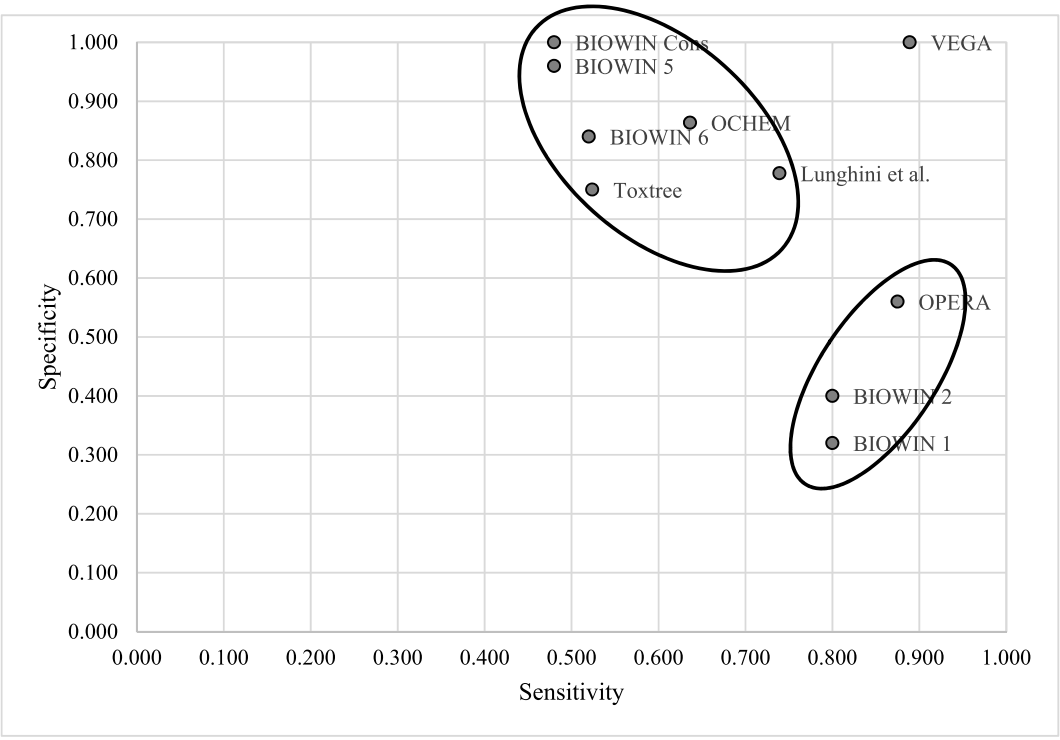

Fig. 2. Sensitivity plotted against specificity of the modelled biodegradability results of 50 test substances, using the selected biodegradation models.

AD check is not performed, thus it cannot be determined whether the APIs to be evaluated in the test data set were even within the AD of these models. These models are therefore unsuitable for the above-mentioned question.

Summarizing, a model that provides valid results for all 50 substances in the prediction of both biodegradable (positive) and non-biodegradable (negative) substances is currently not available.

The models exhibit very different performance evaluations. The range of models is very high in both sensitivity and specificity, resulting in varying levels of accuracy. It should be noted that the data for the validation dataset was collected according to OECD 301D (CBT), while the training data for the models was generated using different test methods, primarily OECD 301C (MITI-I). One advantage of the data for model validation is that the data was generated in-house, thus providing controlled conditions that lead to reliable results. Many of the existing models examined here use data from multiple sources, generated using a variety of (sometimes) unspecified tests from the OECD 301 series. Furthermore, the quality of some training data may be questionable due to the lack of necessary controls, such as toxicity controls. The degradability values of models using the same data source may vary depending on the descriptors used and the proportion of data used in the training dataset. This can also lead to different biodegradability assessments in the different models. The test datasets for the models were not generated internally but compiled from external sources. Another disadvantage of the *in silico* models analyzed is the selection of the data used. Due to the lack of a range of substances with different chemical structures, the models evaluated here lack diversity, as most of the data refer to relatively simple industrial bulk compounds, in contrast to pharmaceuticals, which are often very complex molecules, e.g., in terms of size and the functional groups hetero atoms and cyclic structures present. The models are global models that are often not applicable to specific questions, in this case, the biodegradability of APIs.

Looking at individual substances, it is noticeable that most models had difficulty correctly predicting the biodegradation behavior of chlorine-containing substances, particularly when they were in the form of salts. For example, methacholine chloride and papaverine chloride—both not readily biodegradable in the CBT—were predicted as readily biodegradable by all models. This discrepancy may be attributed to the fact that most *in silico* models operate on neutral molecular structures and do not account for ionic forms. In reality, these salts exhibit high water solubility but reduced bioavailability due to poor membrane permeability, which limits their biodegradability. In contrast, when chlorine is part of a neutral molecule—such as in diclofenac—the model correctly predicts non-readily biodegradable behavior. This suggests that the failure to account for ionization state and physicochemical properties like solubility and bioavailability may significantly impact prediction accuracy, especially for ionic compounds.

Another problem concerns the influence of fluorination. All models classified the five fluorine-containing aromatic compounds 2-trifluoromethylphenol, 4-fluoroaniline, 4-fluorophenol, 4-(trifluoromethyl)aniline, and ciprofloxacin as not readily biodegradable. This is understandable, as fluorination and halogenation are generally strongly associated with negative effects on environmental biodegradability (Straub et al., 2023; Allpress and Gowland, 1998). However, there are some exceptions to the ready biodegradability

of fluorinated aromatics, namely 2-(trifluoromethyl)phenol, 4-(trifluoromethyl)phenol, and 4-fluoroaniline, which were classified as readily biodegradable in the CBT test. None of the models could predict these exceptions. The model predictions only consider the presence of fluorine atoms, but not other factors that influence biodegradability, such as their position in the molecule or whether they are to alkyl or aromatic moieties. Furthermore, the number and diversity of fluorinated compounds in the respective training datasets is rather small. This means that exceptions to this rule cannot be identified.

4. Conclusion

This study evaluated the applicability of nine freely available *in silico* models for predicting ready environmental biodegradability of APIs. To date, freely available QSAR models have mostly been applied to bulk and fine chemicals, some cosmetic ingredients, or some pesticides. The test data set conveyed APIs, non bulk or fine chemicals, chemicals of a more complex chemical structure. Many compounds in the test dataset were outside the AD of the models based on more simple chemicals.

The performance evaluation conducted in this study showed that the QSAR software for predicting the ready biodegradability of chemicals demonstrated only limited capabilities for predicting environmental biodegradability of APIs.

Presence of fluorine is sometimes important in APIs. This demonstrates that further data and research are needed to allow a definitive conclusion on the role of fluorination, as only small molecules that can serve as building blocks for pharmaceuticals were used here. Therefore, it is generally recommended to consider several different *in silico* models in parallel when predicting biodegradability ("*in silico* testing battery"). Another point is the degree of reliability needed which might be different within the drug design process compared to a legal environment. The use of multiple models increases the reliability of the predictions (Puhlmann et al., 2021). The VEGA model is very well suited for predicting the biodegradability of APIs, provided the substance to be determined falls within the AD of the model which has been found in our study too. However, it is not just to apply more models but to know their individual performance too. All the other models are unsuitable APIs because the predictions are too imprecise and often not reliable.

New models should therefore be developed with high-quality data, preferably from a single source, and only using a test from the standardized OECD 301 series, primarily according to CBT, which is commonly used for APIs and in pharmaceuticals. Industry could share data, too. Furthermore, the selection of chemical substances in the training dataset is crucial for applicability. Depending on the availability of high-quality data it might even be an approach to go for local, individual pharmaceuticals group specific models and not just global ones. In global models a wide range of different functional groups should be considered, as pharmaceuticals vary greatly in chemical terms compared to bulk and fine chemicals. An assessment and review of the AD should also be part of the model to ensure reliable predictions. An important component in this context will be the generation of further data on the biodegradability of APIs in the environment.

CRediT authorship contribution statement

Sabrina Groth: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft. **Klaus Kümmerer:** Conceptualization, Funding acquisition, Resources, Writing – review & editing. **Oliver Olsson:** Conceptualization, Methodology, Funding acquisition, Writing – review & editing.

Funding

This work was supported by the TransPharm project, which is funded by the European Union under Horizon Europe Grant Agreement n°101057816. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scp.2026.102387.

Data availability

Data will be made available on request.

References

Allpress, J.D., Gowland, P.C., 1998. Dehalogenases: environmental defence mechanism and model of enzyme evolution. Biochem. Educ. 26 (4), 267–276. https://doi.org/10.1016/S0307-4412(98)00090-9.

Amsel, A.-K., Chakravarti, S., Olsson, O., Kümmerer, K., 2024. Modelling biodegradability based on OECD 301D data for the design of mineralising ionic liquids. Green Chem. 26 (12), 7363–7376. https://doi.org/10.1039/d4gc00889h.

Anastas, P., Eghbali, N., 2010. Green chemistry: principles and practice. Chem. Soc. Rev. 39 (1), 301–312. https://doi.org/10.1039/B918763B.

Bertram, M.G., Wong, B.B.M., Kümmerer, K., Jörg, M., 2025. Development of environmentally biodegradable drugs: what are the key challenges? Expet Opin. Drug Discov. 20 (1), 1–4. https://doi.org/10.1080/17460441.2024.2442746.

Boethling, R.S., Lynch, D.G., Thom, G.C., 2003. Predicting ready biodegradability of premanufacture notice chemicals. Environ. Toxicol. Chem. 22 (4), 837–844. https://doi.org/10.1002/etc.5620220423.

Boethling, R.S., Howard, P.H., Meylan, W., Stiteler, W., Beauman, J., Tirado, N., 1994. Group contribution method for predicting probability and rate of aerobic biodegradation. Environ. Sci. Technol. 28, 459–465.

Branch, L.S., 2006. Consolidated federal laws of Canada, persistence and bioaccumulation regulations. Available online: https://laws-lois.justice.gc.ca/eng/regulations/sor-2000-107/page-1.html.

Burello, E., 2015. Computational design of safer nanomaterials. Environ. Sci. Nano 2 (5), 454–462.

Cheng, F., Ikenaga, Y., Zhou, Y., Yu, Y., Li, W., Shen, J., Du, Z., Chen, L., Xu, C., Liu, G., Lee, P.W., Tang, Y., 2012. In silico assessment of chemical biodegradability. J. Chem. Inf. Model. 52 (3), 655–669. https://doi.org/10.1021/ci200622d.

Cronin, M.T.D., Madden, J.C., 2010. In silico toxicology: an introduction. In: Cronin, M., Madden, J. (Eds.), In Silico Toxicology, Issues in Toxicology. Royal Society of Chemistry, pp. 1–10. https://doi.org/10.1039/9781849732093-00001.

Cronin, M.T.D., 2004. Predicting chemical toxicity and fate in humans and the environment: an introduction. In: Cronin, M.T.D., Livingstone, D.J. (Eds.), Predicting Chemical Toxicity and Fate. CRC Press.

Dobbelaere, M.R., Lengyel, I., Stevens, C.V., Van Geem, K.M., 2024. Geometric deep learning for molecular property predictions with chemical accuracy across chemical space. J. Cheminf. 16, 99. https://doi.org/10.1186/s13321-024-00895-0.

EPA – United States Environmental Protection Agency, 2012. EPI Suite^{tm} Usage Guide.

European Chemicals Agency, 2016. How to Use and Report (Q)Sars.Practical Guide 5. Publications Office. LU: Available online: https://data.europa.eu/doi/10.2823/81818.

European Commission, 2020. Chemicals Strategy for Sustainability Towards a Toxic-free Environment. European Commission, Brussels.

European Commission, 2019. European Union Strategic Approach to Pharmaceuticals in the Environment. European Commission, Brussels.

European Commission, 2017. QMRF identifier (JRC inventory): Q17-23A-0014. OPERA-model for readily biodegradability. Available online: https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/EURL-ECVAM/datasets/QSARDB/LATEST/PDF/_qmrf_protocol_Q17-23a-0014_document.pdf.

European Commission, 2006. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 Concerning the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH).

Ferrari, T., Cattaneo, D., Gini, G., Golbamaki Bakhtyari, N., Manganaro, A., Benfenati, E., 2013. Automatic knowledge extraction from chemical structures: the case of mutagenicity prediction. SAR QSAR Environ. Res. 24 (5), 365–383. https://doi.org/10.1080/1062936X.2013.773376.

Gramatica, P., Papa, E., 2007. Screening and ranking of POPs for global half-life: QSAR approaches for prioritization based on molecular structure. Environ. Sci. Technol. 41 (8), 2833–2839.

Gramatica, P., Cassani, S., Sangion, A., 2016. Aquatic ecotoxicity of personal care products: QSAR models and ranking for prioritization and safer alternatives' design. Green Chem. 18 (16), 4393–4406.

Howard, P.H., Hueber, A.E., Boethling, R.S., 1987. Biodegradation data evaluation for structure/biodegradation relations. Environ. Toxicol. Chem. 6, 1–10.

Howard, P.H., Stiteler, W.M., Meylan, W.M., Hueber, A.E., Beauman, J.A., Larosche, M.E., Boethling, R., 1992. Predictive model for aerobic biodegradability developed from a file of evaluated biodegradation data. Environ. Toxicol. Chem. 11 (5), 593–603. https://doi.org/10.1002/ETC.5620110502.

Kümmerer, K., Clark, J.H., Zuin, V.G., 2020. Rethinking chemistry for a circular economy. Science 367 (6476), 369–370. https://doi.org/10.1126/science.aba4979.

Kümmerer, K., 2019. From a problem to a business opportunity-design of pharmaceuticals for environmental biodegradability. Sustain. Chem. Pharm. 12, 100136. https://doi.org/10.1016/j.scp.2019.100136.

Lombardo, A., Pizzo, F., Benfenati, E., Manganaro, A., Ferrari, T., Gini, G., 2014. A new in silico classification model for ready biodegradability, based on molecular fragments. Chemosphere 108, 10–16. https://doi.org/10.1016/j.chemosphere.2014.02.073.

Lorenz, S., Amsel, A.-K., Puhlmann, N., Reich, M., Olsson, O., Kümmerer, K., 2021. Toward application and implementation of in silico tools and workflows within benign by design approaches. ACS Sustain. Chem. Eng. 9 (37), 12461–12475. https://doi.org/10.1021/acssuschemeng.1c03070.

Lunghini, F., Marcou, G., Gantzer, P., Azam, P., Horvath, D., Van Miert, E., Varnek, A., 2020. Modelling of ready biodegradability based on combined public and industrial data sources. SAR QSAR Environ. Res. 31 (3), 171–186. https://doi.org/10.1080/1062936X.2019.1697360.

Maertens, A., Plugge, H., 2018. Better metrics for "sustainable by design": toward an in silico green toxicology for green(er) chemistry. ACS Sustainable Chemical Engineering 6 (2), 1999–2003. https://doi.org/10.1021/acssuschemeng.7b03393.

Mansouri, K., Grulke, C.M., Judson, R.S., Williams, A.J., 2018. OPERA models for predicting physicochemical properties and environmental fate endpoints. J. Cheminf. 10 (1), 10. https://doi.org/10.1186/s13321-018-0263-1.

Molecular Networks GmbH Computerchemie, 2008. START (structural alerts for reactivity. In: Toxtree) Biodegradation & Persistence Decision Tree - User Manual.

Muhammad, H.N., Yahayam, A.S., Muritala, H.F., 2025. Quantitative structure–activity relationship (QSAR) hazard assessment of the banned pesticides by the national agency for food and drugs administration and control. N. Schmied. Arch. Pharmacol. 398, 8729–8744. https://doi.org/10.1007/s00210-024-03694-4.

OECD, 2024. (Q)SAR assessment framework: guidance for the regulatory assessment of (quantitative) structure activity relationship models and predictions. OECD Series on Testing and Assessment No. 405, second ed. OECD Publishing, Paris.

OECD, 2004. OECD principles for the validation, for regulatory purposes, of (quantitative) structure-activity relationship models. Available online: https://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf.

OECD, 1992. OECD Guideline for Testing of Chemicals 301. Available online: https://www.oecd-ilibrary.org/docserver/9789264070349-en.pdf?expires=1689151797&id=id&accname=ocid56027153&checksum=94BD1B6C42538ADF4323BBDC2BFDFEE1.

Olsson, O., Logunova, E., Suk, M., Kümmerer, K., 2024. Experimental data on ready biodegradability. OECD 301D (Closed Bottle Test) 2023, V1. https://doi.org/10.48548/pubdata-77.

Patlewicz, G., Jeliazkova, N., Safford, R.J., Worth, A.P., Aleksiev, B., 2008. An evaluation of the implementation of the cramer classification scheme in the toxtree software. SAR QSAR Environ. Res. 19 (5-6), 495–524.

Pizzo, F., Lombardo, A., Manganaro, A., Benfenati, E., 2013. *In Silico* models for predicting ready biodegradability under REACH: a comparative study. Sci. Total Environ. 463–464, 161–168. https://doi.org/10.1016/j.scitotenv.2013.05.060.

Puhlmann, N., Mols, R., Olsson, O., Slootweg, J.C., Kümmerer, K., 2021. Towards the design of active pharmaceutical ingredients mineralizing readily in the environment. Green Chem. 23, 5006–5023. https://doi.org/10.1039/D1GC01048D.

Rücker, C., Kümmerer, K., 2012. Modeling and predicting aquatic aerobic biodegradation–a review from a user's perspective. Green Chem. 14 (4), 875–887.

Straub, J.O., Le Roux, J., Tedoldi, D., 2023. Halogenation of pharmaceuticals is an impediment to ready biodegradability. Water 15 (13), 2430. https://doi.org/10.3390/w15132430.

Suk, M., Lorenz, S., Kümmerer, K., 2023. Identification of environmentally biodegradable scaffolds for the benign design of quinolones and related substances. Sustain. Chem. Pharm. 31, 100947.

Sulmona, K., Louazzani, M., Ginestar, J., Lanica, P., Baleydier, A., 2025. Comparative study of (Q)SAR models for predicting the environmental fate of cosmetic ingredients: persistence, bioaccumulation, and mobility. NAM Journal 1, 100054. https://doi.org/10.1016/j.namjnl.2025.100054.

Toropov, A., Toropova, A., Lombardo, A., Roncaglioni, A., Brita, N., Stella, G., Benfenati, E., 2012. CORAL: the prediction of biodegradation of organic compounds with optimal SMILES-based descriptors. Open Chem. 10 (4), 1042–1048. https://doi.org/10.2478/s11532-012-0031-4.

Tunkel, J., Howard, P.H., Boethling, R.S., Stiteler, W., Loonen, H., 2000. Predicting ready biodegradability in the Japanese ministry of international trade and industry test. Environ. Toxicol. Chem. 19 (10), 2478–2485. https://doi.org/10.1002/etc.5620191013.

van Dijk, J., Flerlage, H., Beijer, S., Slootweg, J.C., van Wezel, A.P., 2022. Safe and sustainable by design: a computer-based approach to redesign chemicals for reduced environmental hazards. Chemosphere 296, 134050. https://doi.org/10.1016/j.chemosphere.2022.124050.

Vorberg, S., Tetko, I.V., 2014. Modeling the biodegradability of chemical compounds using the online CHEmical modeling environment (OCHEM). Molecular Informatics 33 (1), 73–85. https://doi.org/10.1002/minf.201300030.

Zacarias, C.H., Paludetti, M.F., de Oliveira, A.A.S., Federico, L.B., 2024. An integrated approach to address the biodegradability of cosmetic formulations as part of a corporate sustainability strategy. Int. J. Cosmet. Sci. 47. https://doi.org/10.1111/ics.13033, 434-361.