Environ Monit Assess (2023) 195:1535
https://doi.org/10.1007/s10661-023-12159-9

REVIEW

A review on in silico prediction of the environmental risks posed by pharmaceutical emerging contaminants

Jian Gao · Jinru Zhao · Xintong Chen · Jun Wang

Received: 11 July 2023 / Accepted: 18 November 2023 / Published online: 27 November 2023

Abstract Computer-aided (in silico) prediction has shown good potential to support the environmental risk assessment (ERA) of pharmaceutical emerging contaminants (PECs), allowing low-cost, animal-free, high-throughput screening of multiple potential risks posed by a wide variety of pharmaceuticals in the environment based on insufficient toxicity data. This review provided recent insights regarding the application of in silico approaches in prediction for environmental risks of PECs. Based on the review of 20 included articles from 8 countries published since 2018, we found that the researchers' interest and concern in this research topic were sharply aroused since 2021. Recently, in silico approaches have been widely used for the prediction of bioaccumulation and biodegradability, lethal endpoints, developmental toxicity, mutagenicity, other eco-toxicological effects such as ototoxicity and hematological toxicity, and human health hazards of exposure to PECs. Particular attention has been given to the simultaneous discernment of multiple environmental risks and health effects of PECs based on mechanistic data of pharmaceuticals using advanced bioinformatic methods such as transcriptomic analysis and network pharmacology prediction. In silico software platforms and databases used in the included studies were diversified, and there is currently no standardized and accepted in silico model for ERA of PECs. Date suggested that in silico prediction of the environmental risks posed by PECs is still in its infancy. Considerable critical challenges need to be addressed, including consideration of environmental exposure concentration for PECs, interactions among mixtures of PECs and other contaminants coexisting in environments, and development of *in silico* models specific to ERA of PECs.

Keywords In silico · Environmental risk assessment · Pharmaceutical emerging contaminants · Database · Risk evaluation platform

Introduction

Along with the increasing consumption of pharmaceutical products worldwide, the environmental pollution caused by pharmaceutical emerging contaminants (PECs) has become a global public problem attracting more and more concerns. So far, various categories of PECs including antibiotics, β-blockers, analgesics and nonsteroidal anti-inflammatory drugs (NSAIDs), antineoplastic, blood lipid-lowering agents, central nervous system acting drugs, antiviral and antiparasitic drugs, hormones, and their metabolites have been frequently reported to be detected

Jian Gao and Jinru Zhao contributed equally to this work.

J. Gao · J. Zhao · X. Chen · J. Wang (✉)
Institute of Pharmaceutical Innovation, Hubei Province Key Laboratory of Occupational Hazard Identification and Control, School of Medicine, Wuhan University of Science and Technology, Wuhan 430065, China
e-mail: wangjun@wust.edu.cn

within environmental samples collected from different countries and regions around the world (Cunha et al., 2019; Fonseca et al., 2021; He et al., 2017; Hu et al., 2021; Kar et al., 2018; Li et al., 2020). Following discharge from domestic wastewater, industry, medical institutions, etc., pharmaceuticals are exposed to complex transport and transformation processes in the environment. Owing to the generally inefficient removal for PECs by common wastewater treatment plants, the unimpeded flows and spreading of these emerging contaminants result in their wide and unquantifiable distribution in surface water, ground water, sediments, soil, organisms, and other compartments of the ecosystem (Guedes-Alonso et al., 2020; Kar et al., 2018). As a class of specially designed active compounds with potent biological activities, pharmaceuticals would yield their intrinsic toxicities to non-target organisms that are long-termly exposed to PECs in the environment even at trace or ultra-trace levels (Cunha et al., 2019; He et al., 2017; Jiao et al., 2022; Li et al., 2020; Miller et al., 2018; Wang et al., 2021). Moreover, unpredictable interactions among multitudinous PECs including pharmaceutical parent compounds, metabolites, and other contaminants coexisting in receiving environments aggravate and amplify the environmental risks posed by PECs to the ecosystem (De Vaugelade et al., 2017; Koltsakidou et al., 2019; Ofrydopoulou et al., 2021).

To address the environmental issues of PECs, the European Medicines Agency (EMA) and the United States (US) Food and Drug Administration (FDA) have developed and introduced various environmental risk assessment (ERA) guidelines to monitor the residual levels and evaluate the potential risks of pharmaceuticals in the environment in recent decades (Jose et al., 2020). For example, the ERA as a part of the registration procedure must be in place before approval of a new human pharmaceutical drug in the European Union (EU) and the USA, which is required to be addressed by pharmaceutical company for the drug's environmental fate and impact (Holm et al., 2013; Jiao et al., 2022; Kar et al., 2018). The risk quotient (RQ) value is commonly applied in the ERA for harmful effects of PECs on ecosystem, which is defined as the ratio of the maximum measured environmental concentration (MEC) to the predicted no effect concentration (PNEC) (Holm et al., 2013;Molnar et al., 2021). The determination of PNEC depends on available toxicological data, for example, ecotoxicological threshold data from experiments on representative aquatic organisms including algae, Cladocera usually *Daphnia sp*., and/or fish species (Molnar et al., 2021). In general, RQ value < 0.01 denotes a negligible risk, $RQ < 0.1$ reveals a low risk, $0.1 < RQ < 1$ represents a medium risk, and $RQ > 1$ indicates a high ecological risk to aquatic organisms (Gouveia et al., 2019; Molnar et al., 2021; Nieto-Juarez et al., 2021; Riva et al., 2019). As the simplest ERA method, the RQ-based procedure has been widely adopted to determine whether chemicals such as PECs in the environment might be posing risks to ecological systems. However, the comprehensive and reliable ERA of PECs has been recognized to be hampered by considerable inherent weaknesses of RQ method (Raimondo & Forbes, 2022), for example, the enormous diversity of PECs with different chemical, pharmacological, and toxicological properties; the limited available long-term toxicological data across the lifespan of different surrogate species; and the limited data on residual concentrations of PECs in various environmental matrices, being based solely on laboratory data. It has been considered that the risk assessment of pharmaceuticals and their metabolites based only on MEC and RQ values might underestimate their risks to the environment and humans (Wielens Becker et al., 2020). Therefore, further developing ERA tools is necessary to effectively support PEC risk management decision-making.

Driven by the advanced innovations of artificial intelligence technologies, more and more computer-aided (in silico) approaches as cost-effective, easily accessible, safe, feasible, and promising computer-assisted animal-free testing strategies provide new ideas and opportunities for the rapid prediction and comprehensive assessment of potential environmental risks posed by PECs. Here, we review reports on the application of in silico approaches to support the ERA of PECs as documented in recently published literature.

Methods

In this review, we performed a literature search on the electronic MEDLINE bibliographic database via PubMed using keywords "*risk(s)* AND (*environmental* OR *ecological)* AND (*assess* OR *assessment* OR *predict* OR *prediction* OR *predictive*) AND

(*in silico* OR *computational)* AND (*drug(s)* OR *pharmaceutical(s))*". Only peer-reviewed articles written in English fully published online in recent 5 years (between January 2018 and December 2022) were included. Reviews were excluded from the retrieved articles, and the remaining were reviewed individually to identify and select articles that met the following criteria for analysis: Research and illustrative articles about the ERA of PECs that were active pharmaceutical ingredients identified in the Drugbank database (Wishart et al., 2018); research articles based on in silico platforms, systems, or software for evaluation of study endpoints; and illustrative articles showing research platforms, systems, or software.

Results and discussion

A total of 20 studies employing the in silico predictive approaches to aid in the ERA of PECs have been found. The researchers' interest and concern in this research topic appear to be sharply aroused from 2 years ago, which is reflected by the yearly distribution of included publications as follows: 1 (2018) (Raitano et al., 2018), 2 (2019) (Miller et al., 2019; Tung et al., 2019), 4 (2020) (Della-Flora et al., 2020; Garcia-Martin et al., 2020; Wielens Becker et al., 2020; Zhang et al., 2020), 8 (2021) (Han et al., 2021; Hua et al., 2021; Huang et al., 2021; Kumar et al., 2021; Liu et al., 2021; Marmon et al., 2021; Saavedra & Duchowicz, 2021; Sanabria et al., 2021), and 5 (2022) (Badry et al., 2022; Han et al., 2022; Kumar et al., 2022; Regnery et al., 2022; Spînu et al., 2022). As an uninvited guest for human society, the COVID-19 crisis broken out in 2020 brought some opportunities regardless of enormous problems it created. In particular, facing the post-pandemic era of information explosion, the application of in silico methods has been very useful and popular in managing different projects associated with COVID-19 crisis (Moradi et al., 2022; Sharifi et al., 2021). Moreover, due to the increasing environmental loads of PECs resulting from intensive pharmaceutical consumption triggered by COVID-19, the ERA of PECs has attracted significant academic and political interest during post-pandemic period (Anand et al., 2022; Guo et al., 2021; Morales-Paredes et al., 2022). These reasons might account for the sudden increase in the in silico studies under the field of ERA for PECs since 2021.

Spatial distribution of included studies

Among the 20 included studies published in the last 5 years, 8 studies (40%) were from Asia (Han et al., 2021, 2022; Hua et al., 2021; Huang et al., 2021; Kumar et al., 2021; Liu et al., 2021; Tung et al., 2019; Zhang et al., 2020), another 8 studies (40%) were from Europe (Badry et al., 2022; Garcia-Martin et al., 2020; Kumar et al., 2022; Marmon et al., 2021; Miller et al., 2019; Raitano et al., 2018; Regnery et al., 2022; Spînu et al., 2022), and the remaining 4 studies (20%) were conducted by researchers from South America (Della-Flora et al., 2020; Saavedra & Duchowicz, 2021; Sanabria et al., 2021; Wielens Becker et al., 2020). As shown in Table 1, the distribution of retrieved articles across countries showed that the

Table 1 Number of publications per year for 20 studies employing the in silico approaches to aid in the ERA of PECs

Region	Year					Total
	2018	2019	2020	2021	2022	
Asia		1	1	5	1	8
China		1	1	4	1	7
India				1		1
Europe	1	1	1	1	4	8
UK		1		1	1	3
Germany					2	2
Spain			1		1	2
Italy	1					1
South America			2	2		4
Brazil			2	1		3
Argentina				1		1

highest number of studies were from China (7 studies, 35%), followed by Brazil (3 studies, 15%), the UK (3 studies, 15%), Spain (2 studies, 10%), Germany (2 studies, 10%), Italy (1 study, 5%), India (1 study, 5%), and Argentina (1 study, 5%). Theoretically, the highly shareable feature of digital research resources might provide equal opportunities for researchers all over the world to conduct the in silico studies. Especially for resource limited settings, the use of in silico research tools could help to reduce the consumption of laboratory materials and instructors. It can be speculated that the further expansion of computer and Internet technologies in research field might allow a wider study on in silico prediction of environmental risks posed by PECs among researchers across institutions, areas, and countries, especially for those from low-and middle-income countries.

Research topic analysis in the application of in silico techniques for prediction of environmental risks posed by PECs

Although the study purposes and in silico models of the included articles were diverse, we found that the research topics are mainly concerned with the following aspects associated to the prediction and evaluation of environmental risks posed by PECs.

Predicting bioaccumulation and biodegradability

Uptake and accumulation of active pharmaceutical ingredients in non-targeted organisms that inhabit the PECs-impacted environment would follow by the potential impairment of organ-specific functions in exposed organisms, once the accumulative levels of PECs were higher than the safety levels for organs (Kunene & Mahlambi, 2023; Nendza et al., 2018; Nozaki et al., 2023). More seriously, through bioaccumulation, PECs enter the food chain, hereby resulting in biomagnification of contaminants in the environment and thus posing potential risks to human health (Kunene & Mahlambi, 2023; Nendza et al., 2018; Nozaki et al., 2023). On the other hand, the exposed organisms have the biodegradability to metabolize the PEC compounds. However, PECs also affect this natural detoxification capability in exposed organisms by negatively impacting their metabolism processes (Kunene & Mahlambi, 2023). Currently, the assessment of bioaccumulation potential of pharmaceuticals is required as an essential and mandatory part of their regulatory ERA in the EU (Regnery et al., 2022). We found that, of the 20 papers included in this review, 35% (7 paper) (Badry et al., 2022; Della-Flora et al., 2020; Garcia-Martin et al., 2020; Miller et al., 2019; Regnery et al., 2022; Sanabria et al., 2021; Wielens Becker et al., 2020) reported the application of in silico techniques to assess the bioaccumulation and biodegradability of PECs in the environment.

The well-accepted bioaccumulation assessments are mainly based on bioconcentration factors (BCFs) (Nendza et al., 2018). However, a standard experimental determination of BCFs needs to use more than 100 fish and is very time-consuming and expensive. According to structural features and physicochemical properties related to the distribution, solubility, volatility and persistence of chemicals in water bodies and aquatic biota, the quantitative structure–activity relationship (QSAR) classifications have been considered promising candidates for the replacement and reduction of in vivo BCF testing on fish (Nendza et al., 2018;Thomas et al., 2018). In fact, the Environmental Protection Agency of the United States (US EPA) and Canadian Environmental Assessment Agency (CEAA) have routinely encouraged to use the QSAR approaches to prioritize and support new chemical registrations in the last decades (Thomas et al., 2018). In order to improve the deficiency of single RQ values in ERA of PECs and perform a more proactive prioritization of complex mixture of pharmaceuticals and metabolites in the environment, Wielens Becker et al., (2020) employed the in silico Prometheus software, in which a battery of QSAR models were used, as a complimentary tool for predicting and ranking the parent pharmaceuticals and metabolites occurring in hospital wastewater samples as possible bioaccumulative compounds in terms of biodegradability. A Brazil research team (Della-Flora et al., 2020; Sanabria et al., 2021) assessed the biodegradability of anti-cancer drugs and their environmental transformation products by in silico QSAR free software package including BIOWIN 1–7 and VEGA (IRFMN model). Prometheus software was used to rank the transformation products as possible persistent, bioaccumulative, and toxic (PBT) compounds. The in silico predictions indicated that the transformation products formed during the degradation process of anti-cancer flutamide and anastrozole were not biodegradable, while some of them were

classified near the threshold point to be considered as PBT compounds (Della-Flora et al., 2020; Sanabria et al., 2021). Another study (Regnery et al., 2022) performed the in silico QSAR-based BCF predictions using the regression-based QSAR model BCFBAF (v3.01 EpiSuite, US EPA) to conduct in vitro to in vivo extrapolation for the bioaccumulation assessment of pharmaceutical anticoagulants propranolol, phenprocoumon, and warfarin in fish, under the condition that experimental data on fish metabolism for these PECs were rarely available. Using the high resolution-mass spectrometry coupled to liquid and gas chromatography, Badry et al., (2022) found that an anthelmintic agent oxfendazole was detected in all the 30 livers of German white-tailed sea eagles, a species of apex predators foraging on fish and water birds, with the average level of 40.6 ng/g. In line with this observed finding, when predicting its PBT properties by the JANUS tool (https://www.vegahub.eu/portfolio-item/janus/) based on a battery of QSAR models, oxfendazole had a score P (persistent) of 0.712 (a score > 0.6 indicates that PBT properties are likely to be met) and thus was also shown to be persist in the environment. This finding suggested that the in silico JANUS software might be reliable for predicting PBT properties for certain PECs. Nevertheless, other PECs with a score > 0.6 including pindolol, desethylhydroxychloroquine, sulfadoxine, lidocaine, and lidocaine-N-oxide were found at very low concentrations and detection rates in liver samples, suggesting mismatches between observed exposures and in silico predicted PBT properties. There might be the complexity of exposure events to some PECs for apex predators that are not solely associated to intrinsic chemical properties.

Compared with conventional statistics-based analyses, the machine learning methods have been demonstrated better performance in QSAR modeling considering the complex relationships between structures and PBT properties of chemicals and been used for predicting bioaccumulation and biodegradability of PECs based on a series of training data (Miller et al., 2019; Xu et al., 2022). Miller et al., (2019) evaluated 24 linear and machine learning models to predict the BCFs of PECs in fish *Cyprinus carpio* and optimally selected a 4-layer multilayer perceptron machine learning algorithm using 14 molecular descriptors among them. The modelled descriptors covered 6 topological ones including radial centric information index (ICR), ramification index (Ram), Narumi harmonic topological function (Hnar), spanning tree number (STN), superpendentic index (SPI), and topological polar surface area (TPSA); 4 constitutional descriptors including number of nitrogens (nN), number of carbons (nC), number of hydrogens (nH), and molecular weight (MW); 3 electrotopological descriptors including maximal electrotopological positive variation (MAXDP), maximal electrotopological negative variation (MAXDN), and mean atomic Sanderson electronegativity (Me), as well as a physicochemical property that is the octanol-water distribution coefficient (logD). When employing this optimized model for further prediction of BCFs for PECs in fish and invertebrate *Gammarus pulex*, the machine learning was found to show good performance in cross-species prediction of bioaccumulation and thus enable rapid prioritization of PECs during ERA without the need for ethically challenging and costly animal experiments. Being different from simple hydrophobicity models which poorly account for the complexity of PECs, this multilayer perceptron model contributed to better predict the bioaccumulation and its driving molecular descriptors (Miller et al., 2019). Depending on a machine learning support vector machine (SVM) predictor that is trained to distinguish biodegradable from recalcitrant PECs using a vector representation of the compounds, an in silico system named BiodegPred (https://sysbiol.cnb.csic.es/BiodegPred/) (Garcia-Martin et al., 2020) merges computational methods to predict biodegradability of a PEC with others that assess eventual biological toxicity. Even for PECs without previously existed biodegradation data, the BiodegPred using only the chemical structure as the input could provide a prognosis of the chance that a given pharmaceutical compound can eventually be metabolized in the biosphere when it is released into the environment. Then, the blind predictions based on BiodegPred for a set of antiviral agents of medical interest showed that 3 of the 148 studied antiviral agents would be "ready biodegradable," and most of the antivirals belonged to the "non-ready biodegradable" category (Garcia-Martin et al., 2020).

Predicting the lethal endpoints

Aquatic acute and chronic toxicity based on the half-maximal lethal concentration (LC_{50}) values is a key

and common endpoint in ecological risk assessment. We found that, since 2018, there were 3 articles (Han et al., 2021, 2022; Raitano et al., 2018) reporting the in silico prediction for PECs-caused lethal endpoints in aquatic organisms.

Han et al., (2021) constructed and validated the quantitative structure-toxicity relationship (QSTR) models using genetic function approximation (GFA) algorithm based on data sets of 9 quinolone drugs and 7 impurities to build reliable in silico mathematical models for predicting aquatic toxicity of quinolone antibiotics. Results showed that the acute toxicity LC_{50} values of 11 quinolones predicted using QSTR modelling were in good agreement with the in vivo experimental findings in zebrafish embryos (*Danio rerio*), suggesting that the constructed model is accurate and has a good predictive power. Then, this research team further developed QSTR models for zebrafish toxicity prediction in the same way and applied models for β-lactam antibiotics prediction (Han et al., 2022). Moreover, this recent study (Han et al., 2022) also used the Ecological Structure–Activity Relationships (ECOSAR) in silico prediction to estimate the toxicity of insoluble penicillins and the fifth-generation cephalosporins for aquatic organisms. The LC_{50} values for acute and chronic exposure to β-lactams from the ECOSAR software prediction were similar to the QSTR results showing that eight β-lactams might exhibit low toxicity. Among them, mecillinam might have the strongest chronic toxic effect to fish, daphnia, and algae, and ceftibuten might cause acute and chronic toxicity to green algae (Han et al., 2022).

As we mentioned above, the procedure based on RQ, the MEC/PNEC ratio, currently acts as the most common tool in the ERA of PECs (Gouveia et al., 2019; Molnar et al., 2021; Nieto-Juarez et al., 2021; Riva et al., 2019). However, no adequate experimental data supporting the calculation of PNECs, the latter of which were usually calculated based on the LC_{50} values, as reported in the literature has been believed as a big challenge for the RQ assessment of various PECs (Raimondo & Forbes, 2022). The use of in silico methods contributes to fill the gaps of information about RQ-based procedure. Raitano et al., (2018) employed multiple in silico QSAR modeling platforms to obtain the required and sufficient hazard values for PECs when experimental data were missing or uncertain. In this study (Raitano et al., 2018), the EPI Suite™ (Estimation Program Interface) software based on ECOSAR model, developed by US EPA, was used for predicting acute and chronic toxicity endpoints in algae, *Daphnia magna* and fish; the T.E.S.T. (Toxicity Estimation Software Tool) and VEGA (Virtual models for property Evaluation of chemicals within a Global Architecture) provided predictive LD_{50} parameters for fish acute toxicity; the T.E.S.T. gave the prediction of *Daphnia magna* acute toxicity. Then, the obtained data were used to support the calculation of RQ values of PECs for ecological risk assessment, and results from the empirical study showed the high RQs of PECs including clarithromycin, furosemide, ranitidine, and diazinon occurring in water samples collected from Ledra River, Italy (Raitano et al., 2018).

Predicting developmental toxicity

Among the included articles, the general developmental toxicity, cardiac developmental toxicity, and developmental neurotoxicity of PECs were predicted using the in silico methods in 3 studies (Han et al., 2021; Saavedra & Duchowicz, 2021; Spînu et al., 2022), respectively.

Saavedra & Duchowicz, (2021) developed a QSAR model based on a zebrafish embryo developmental toxicity database provided by the ToxCast™phase I chemical library belonging to the US EPA Computational Toxicology Research Program, and applied it for predicting the half-maximal active concentration (AC_{50}) of 188 antimicrobial, antibacterial, and antiparasitic products in a set of 28,038 non-conformational molecular descriptors, which encode permanent structural features affecting the general developmental toxicity. This in silico method provided a promising way for the low-cost, cruelty-free, high-throughput screening and exhaustive analysis during the ERA for zebrafish embryo developmental toxicity of a structurally diverse set of PECs (especially for new, untested even hypothetical PECs) according to their predictive effects on thousands of potential molecular targets.

After treatment of 12 quinolone drugs for 6–72 hpf, more than 80% of abnormal zebrafish embryos exhibited heart malformations, suggesting quinolone PECs-induced cardiac developmental toxicity (Han et al., 2021). Being consistent with this laboratory-based finding, in silico molecular docking prediction

showed good affinity interactions between quinolones and the active binding pocket of zebrafish ERG protein, the human homolog gene of which encodes the Kv11.1 potassium ion channel responsible for quinolone-induced heart QT prolongation (Han et al., 2021), as a molecular target. This study indicated that, if there is a well-identified molecular target as ecological risk endpoint, the in silico molecular docking simulation might be a potential tool for the ERA of PECs.

Considering the susceptibility of developing nervous system to the neurotoxic effects of exogenous chemicals, testing compounds including PECs for developmental neurotoxic potential has been a significant societal and scientific goal (Atzei et al., 2021). Currently, the presence of environmentally relevant PECs such as psychoactive pharmaceuticals in the aquatic environment has been linked to the neurodevelopmental toxicity in fish species (Atzei et al., 2021). However, because the developmental neurotoxicity is a complex process with multiple cellular and molecular paths, current common methods for developmental neurotoxicity testing including measures of gross morphology in the brain, neurochemistry, a range of behavioral assays, and biomarkers of gliosis and cytotoxicity are generally complex and expensive in terms of scientific resources, animal use, and time (Spînu et al., 2022). Spînu et al., 2022 developed a hierarchical multiparameter model of adverse outcome pathway (AOP) network for developmental neurotoxicity testing based on Bayesian machine learning, to predict the probability that a PEC induced each of three upstream common key events of the AOP network and the adverse outcome of developmental neurotoxicity. The modelling workflow could deal with missing values, accommodate unbalanced and correlated data, and follow the structure of a directed acyclic graph (DAG) to simulate the biological path. The developed Bayesian model was found to predict developmental neurotoxicity potential in a data set of 88 compounds including pharmaceuticals, industrial chemicals, and pesticides with an accuracy of 76% (Spînu et al., 2022).

Predicting mutagenicity

Currently, ranging from relatively simple in silico approaches, such as linear statistical techniques, to sophisticated machine learning methods has been widely used to predict and assess the mutagenicity, the ability to cause the mutations in deoxyribonucleic acid (DNA) sequence, of compounds (Kumar et al., 2021). In particular, the mutagenic, genotoxic and carcinogenic potential of considerable pharmaceutical compounds makes them interesting candidates for study especially when entering the environment, food chain, and interfering with the ecosystems as PECs (Sharif et al., 2016). Kumar et al., (2021) developed an in silico model based on the deep neural network (DNN), the latter of which is a subset of optimized machine learning algorithms not requiring manual feature extraction, as an in-depth novel architecture for mutagenicity prediction. When comparing the developed model's performance parameters in mutagenicity prediction with traditional machine learning methods including SVM, k-nearest neighbor, and random forest, the DNN-based prediction model achieved the highest prediction accuracy of 92.95% and 83.81% with the training and test sets of thousands of compounds including PECs, respectively. In order to predict mutagenicity of anti-cancer drugs flutamide, anastrozole, and their environmental transformation products, the VEGA 1.1.4 software including 4 different carcinogenicity models (CAESAR v. 2.1.9, ISS v. 1.0.2, IRFMN/Antares v. 1.0.0, and IRFMN/ISSCAN-CGX v. 1.0.0) was used according to the Ames test (CONSENSUS model v. 1.0.2) (Della-Flora et al., 2020; Sanabria et al., 2021). Additionally, the QSAR Toolbox v. 4.3.1 software was employed to provide a series of complementary mutagenicity alerts generated for the chemical structures. Results showed that some transformation products of anastrozole, the parent compound flutamide, and most of its transformation products were predicted to give positive alerts concerning the mutagenicity and carcinogenicity endpoints (Della-Flora et al., 2020; Sanabria et al., 2021).

Predicting other toxic effects, such as ototoxicity and hematological toxicity

Considerable environmental chemicals such as carbon disulfide, butyl nitrite, trichloroethylene, styrene, and xylene have been known to result in organ specific toxicity to inner ear. It could be speculated that, once entering into the environment, the common ototoxic drugs including aminoglycoside antibiotics, loop diuretics, salicylates, quinine, and platinum-based

anticancer drugs as PECs would be likely to cause hearing loss or tinnitus in exposed animals even humans. Zhang et al., (2020) extracted 897 ototoxic pharmaceuticals and 1715 approved drugs and developed an in silico model using naïve Bayes classifier approach. This model was reported to provide an overall prediction accuracy of 88.7% for the external test set. Furthermore, Huang et al., (2021) proposed a higher-quality data set containing 1102 ototoxic pharmaceutical agents and 1705 non-ototoxic drugs and then developed in silico ototoxicity models using machine learning and deep learning algorithms on online chemical database and modeling environment (OCHEM) that could be used for the ERA of ototoxic PECs. The consensus model with high predictive accuracy and the datasets used for model development were made publicly and freely available at https://ochem.eu/model/46566321.

In addition, as an adverse effect on bone marrow or blood cells, hematotoxicity can result from environmental PECs. On the basis of a high-quality data set containing 632 hematotoxic compounds and 1525 drugs without hematotoxicity, the machine learning and deep learning methods based on structurally diverse chemicals have been used to develop in silico models for the estimation of PECs-induced hematotoxicity (Hua et al., 2021). Among the 35 machine learning models and 3 deep learning models, a model developed with random forest regression and classification algorithm (RFR) and QNPR (quantitative name property relationship) descriptors performed as the best individual one, yielding the balanced accuracy of 0.77 on external validation and the prediction accuracy of 0.83 (Hua et al., 2021).

In silico human health hazard assessment for PECs

PEC residues in the environment can not only cause toxic effects on the exposed microorganisms, plants, and animals, but also enter the human food and drinking water through migration and thus produce potential risks to human health. The current ERA of PECs is mainly based on experimental animal testing as the gold standard, which findings are not fully in good and comparable concordance with human responses (Tung et al., 2019). We found that there were 4 recent publications (Han et al., 2022; Kumar et al., 2022; Raitano et al., 2018; Tung et al., 2019) reporting the in silico human health hazard assessment for PECs based on datasets of biological or toxicological endpoints for human data.

In addition to employ in silico QSAR modeling platforms to support the RQ-based ERA of PECs in algae, *Daphnia magna* and fish, for the evaluations of consequent health risks of PECs to exposed humans, Raitano et al., (2018) developed ad hoc models through integrating multiple in silico methods and software with traditional toxicological risk analysis to predict missing reference dose (RfD) values for non-cancer health assessment and slope factors for carcinogenic toxicity and thus provided a comprehensive general picture of the (eco)toxicological maps. Similarly, based on the in silico experimental ERA in fish, daphnia, and algae, Han et al., (2022) further predicted the human health hazard values of β-lactam antibiotics by admetSAR (http://lmmd.ecust.edu.cn/admetsar2/) and found that carbenicillin, ceftaroline, ceftibuten, ceftobiprole, and ceftolozane might have hepatotoxicity, and ampicillin might have carcinogenicity for human beings. In order to achieve better predictive performance on PECs as human skin sensitizers over animal testing, Tung et al., (2019) developed a transfer learning-based multitask learning method with high prediction performance and coverage to combine the experimental results to key events of the skin sensitization AOP in humans and implement a freely accessible web server named SkinSensPred (https://cwtung.kmu.edu.tw/predict) that could be used for the identification and prioritization of possible human skin sensitizers from PECs. This in silico method used a tree-based ensemble multitask learning algorithm simultaneously dealing with multiple tasks, to achieve the goal to incorporate and leverage three relevant learning tasks corresponding to the three key events of well-defined skin sensitization AOP including keratinocyte activation, protein binding, and dendritic cell activation (Tung et al., 2019). In addition, Kumar et al., (2022) investigated the pharmacophoric modeling approach for human health risk assessment by screening the blood-brain barrier permeation of pharmaceuticals as xenobiotics. Through a computationally expensive process, an optimized 3D structure-data file library of the pharmaceutical molecules and substrates of P-glycoprotein, the latter of which acts as a gatekeeper in human blood-brain barrier, was generated. The interaction fingerprints created from generated data and combinations of docked and extended connectivity (ECFP4)

fingerprints were trained using machine learning algorithms to classify permeable and non-permeable groups. This modeling pipeline was considered a generic framework providing a path to the in silico human risk assessment of environmental chemicals such as PECs (Kumar et al., 2022).

Simultaneously discerning a broad spectrum of potential environmental risks and health effects based on mechanistic endpoints

In the "real world," each pharmaceutical often interacts with more than one targets, eliciting a cascade of complex biological responses through the related genes that may cause multiple health effect. During the age of Big Data, the advanced bioinformatic methods covering biological information systematically and holistically, for example, transcriptomic analysis and network pharmacology prediction, have enabled the development and application of in silico prediction for multiple biological/toxicological effects of PECs according to the network-based drug-gene-disease relationship. Especially under the condition that ecotoxicity data limitations are common for PECs, the prior understanding of the pharmaceuticals' biological activities in conjunction with molecular responses associated with mechanistic effects has been believed to provide a foundation for predicting potential ecologically relevant outcomes through in silico models (Ankley et al., 2022). Using mechanistic data of pharmaceuticals to predict adverse ecological effects of PECs is a novel and promising approaches that could be applied to support PEC evaluations (Ankley et al., 2022).

To adequately and comprehensively predict the sophisticated health risks of environmental PECs, Liu et al., (2021) exploited the cell-based drug transcriptomics data from a chemo-centric perspective and proposed an in silico prioritization method for novel toxic health effects of environmental chemicals. By means of an in silico algorithm termed non-negative matrix factorization (NMF) for condensing sparse high-dimensional data, both the genomic and chemotype data matrices of a pharmaceutical were mapped into a low-dimensional latent feature space, thereby allowing the association analysis between the pharmaceuticals' structural features and transcriptomic data involving multiple gene signaling pathways and different health outcomes. In this predictive model, a matrix of 953 pharmaceuticals and 20,183 genes were included in the biological space, and the chemical space was symbolized by 953 pharmaceuticals and 3534 structural fragments, yielding a total of 13 pivotal types of health effects. Furthermore, this study (Liu et al., 2021) used the potential environmental estrogens from the EPA's endocrine disruptor screening program (EDSP) as external data to verify the prediction capacity of the established model and found that the precision and recall values reached 0.76 and 0.77, respectively. Moreover, using this model allowing the simultaneous prediction of a range of health effects, more potential health impacts such as cardiovascular effect were verified for some tested endocrine disrupting pharmaceuticals (Liu et al., 2021).

Based on network pharmacology approach, Marmon et al., (2021) developed a novel in silico multiscale model for prediction of the risks to fish posed by anti-inflammatory NSAIDs and their mixtures under realistic exposure scenarios. Through the integration of pharmacological as well as network-centered and target-centered mechanistic considerations into the ERA of PECs, this model provided highly specific predictions of the adverse phenotypes potentially occurring in wild fish associated with exposure to NSAIDs. According to physiological functions of the NSAIDs-perturbing targets, a gene-phenotype anchoring analysis based on the Monarch Initiative platform (https://monarchinitiative.org/) indicated that the environmental risk-based NSAIDs bioactivity network might cause ecological effects on general development, cardiovascular and immune systems, the liver, pancreas, and kidney functions, as well as growth and reproduction in fish. It is worth mentioning that, among all the included articles in this review, this interesting study (Marmon et al., 2021) is the only one with sufficient consideration of environmental residual levels of PECs and their concentrations inside the exposed organism as a key parameter for the realistic and accurate risk predictions. Using the Fish Plasma Model, the real measured concentrations of NSAIDs in water samples such as treated waste-water treatment plant effluents and surface waters were transformed into the predicted effect plasma concentrations in the blood of exposed wild fish. Moreover, when generating hazard-based bioactivity networks of NSAIDs mixtures as PECs in fish based on the ToxCast and ChEMBL platforms, most

Table 2 An illustrative list of software platforms recently used the *in silico* prediction of environmental risks posed by PECs

Software platforms	Website	Brief description	Access (public/owner)	Ref.
BiodegPred	https://sysbiol.cnb.csic.es/BiodegPred/	A tool for predicting biodegradability and/or toxicity, based on the different definitions provided for some of the most popular compound databases	Owner	Garcia-Martin et al., (2020)
Prometheus	https://www.prometheussoftware.ca/	Prometheus software provides customers with a myriad of services, including software design and development, systems integration, network support, IT consulting, professional services, and project management	Public	Della-Flora et al., (2020) Sanabria et al., (2021) Wielens Becker et al., (2020)
OCHEM	https://ochem.eu/	OCHEM (Online Chemical Modeling Environment) is a user-friendly web-based platform that aims to automate and simplify the typical steps required for QSAR modeling	Public	Hua et al., (2021) Huang et al., (2021)
AdmetSAR	http://lmmd.ecust.edu.cn/admetsar2/	AdmetSAR creates a user-friendly interface to search for ADME/T (Absorption, Distribution, Metabolism, Excretion, and Toxicity) properties profiling by name, CASRN, and similarity search. admetSAR can predict about 50 ADMET endpoints with QSAR models	Public	Han et al., (2022)
QSAR Toolbox	https://www.qsartoolbox.org/	The toolbox is a free software application that supports reproducible and transparent chemical hazard assessment. It offers functionalities for retrieving experimental data, simulating metabolism, and profiling properties of chemicals. These information and tools can be used to find structurally and mechanistically defined analogues and chemical categories, which can serve as sources for read-across and trend analysis for data gap filling	Public	Della-Flora et al., (2020) Raitano et al., (2018) Sanabria et al., (2021)

Table 2 (continued)

Software platforms	Website	Brief description	Access (public/owner)	Ref.
ECOSAR	https://earth.gsfc.nasa.gov/bio/instruments/ecosar	ECOSAR (ECOlogical Structure Activity Relationship) is freely available from the US EPA which utilizes a number of class-specific log K_{ow}-based QSTRs to predict the toxicity (both short-term and long term) of chemicals. Hazard assessment of environmentally occurring pharmaceuticals to fish, daphnids, and green algae can be performed	Public	Han et al., (2022) Raitano et al., (2018)
T.E.S.T.	https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test	The Toxicity Estimation Software Tool (T.E.S.T) was developed to allow users to easily estimate the toxicity of chemicals using QSARs methodologies	Public	Raitano et al., (2018)
VEGA	http://www.vega-qsar.eu/	VEGA (Virtual models for property Evaluation of chemicals within a Global Architecture) freely offers tens of models for properties such as persistence, logP, BCF, carcinogenicity, mutagenicity, and skin sensitization	Public	Della-Flora et al., (2020) Raitano et al., (2018) Sanabria et al., (2021) Wielens Becker et al., (2020)
CASE Ultra	https://multicase.com/case-ultra	CASE Ultra is an easy to use and capable QSAR software. It can automatically derive alerts from data using statistical methods or use alerts obtained using expert knowledge for predicting bioactivity of untested compounds	Public	Della-Flora et al., (2020) Sanabria et al., (2021)
EPI Suite™	https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface	The EPI (Estimation Programs Interface) Suite™ is a Windows®-based suite of physical/chemical property and environmental fate estimation programs developed by EPA's and Syracuse Research Corp. (SRC)	Public	Della-Flora et al., (2020) Raitano et al., (2018) Regnery et al., (2022) Wielens Becker et al., (2020)
ToxTree	http://toxtree.sourceforge.net	Toxtree is a full-featured and flexible user-friendly open-source application, which is able to estimate toxic hazard by applying a decision tree approach	Public	Raitano et al., (2018) Tung et al., (2019)

Table 2 (continued)

Software platforms	Website	Brief description	Access (public/owner)	Ref.
JANUS	https://www.vegahub.eu/portfolio-item/janus/	JANUS is a tool that was funded by the German Federal Ministry for the Environment, Nature Conservation, and Building and Nuclear Safety to prioritize and screen substances, considering PBT, CMR (carcinogenic, mutagenic and reprotoxic) substances, as well as chemicals supposed to be endocrine disruptors	Public	Badry et al., (2022)
AOP-Wiki	https://aopwiki.org/	The AOP-Wiki provides access to the AOP information via a web interface that supports browsing and searching for AOPs, KEs, KERs, and stressors known to perturb the AOPs. The AOP-Wiki also provides programmatic access to all information via web services as well as downloadable files containing all content in XML format. In addition, the AOP-Wiki provides a user the ability to create snapshots of each AOP in html and pdf formats, which can be used for offline access	Public	Spînu et al., (2022)
ACD/Percepta	https://www.selectscience.net/products/acd+percepta-platform/?prodID=11332023tab-2	The ACD/Percepta Platform delivers full complement of physchem (Absorption, Distribution, Metabolism, and Excretion – ADME) and toxicity tools in one seamless environment	Public	Miller et al., (2019)
Schrodinger	https://www.schrodinger.com/	A physics-based computational platform leverages a deep understanding of physics, chemistry, and predictive modeling to accelerate innovation	Public	Han et al., (2021)
RDKit	http://www.rdkit.org	RDKit is an open source toolkit for chemical informatics. Based on the operations of 2D and 3D molecules of compounds, RDKit uses machine learning methods for compound descriptor generation, fingerprint generation, compound structure similarity calculation, 2D and 3D molecule display, etc.	Public	Liu et al., (2021)

Table 2 (continued)

Software platforms	Website	Brief description	Access (public/owner)	Ref.
alvaDesc	https://www.alvascience.com/	The alvaDesc is a next-generation software for the calculation of molecular descriptors. It can calculate over 5000 molecular descriptors which include RDF descriptors, 2D matrix-based descriptors, 2D autocorrelations, 3D atom pairs, 3D-MoRSE descriptors, atom-centered fragments, basic indices, burden eigenvalues, charge descriptors, connectivity indices, connectivity-like indices, constitutional indices, drug-like indices, eigenvalues, functional group counts, geometrical descriptors, getaway descriptors, information indices, molecular properties, Randic molecular profiles, ring descriptors, spectral moments, topological indices, walk and path counts, and whim descriptors	Public	Kumar et al., (2021)
JSME molecular editor	http://peter-ertl.com/jsme/	A molecule editor, i.e., a program facilitating graphical input and interactive editing of molecules, is an indispensable part of every cheminformatics or molecular processing system	Public	Garcia-Martin et al., (2020)
PaDEL	http://www.yapcwsoft.com	PaDEL-Descriptor is a software for calculating molecular descriptors and fingerprints. The software currently calculates 797 descriptors (663 1D, 2D descriptors, and 134 3D descriptors) and 10 types of fingerprints. Some additional descriptors and fingerprints were added, which include atom type electrotopological state descriptors, McGowan volume, molecular linear free energy relation descriptors, ring counts, count of chemical substructures identified by Laggner, and binary fingerprints and count of chemical substructures identified by Klekota and Roth	Public	Saavedra & Duchowicz, (2021)

Table 2 (continued)

Software platforms	Website	Brief description	Access (public/owner)	Ref.
Mold2	https://www.fda.gov/science-research/bioinformatics-tools/mold2	A free software for fast-calculating descriptors from a two-dimensional chemical structure that is suitable for small and large datasets	Public	Saavedra & Duchowicz, (2021)
QuBiLs-MAS	http://tomocomd.com/software/qubils-mas	QuBiLs-MAS software is for the calculation of 2D molecular descriptors based on the bilinear, quadratic, and linear algebraic forms and thus is the unique software that compute these kinds of indices, applying matrix transformations (simple-stochastic, double-stochastic, and mutual probability), cut-offs, local calculations, and aggregation operators	Public	Saavedra & Duchowicz, (2021)
CHARMM	https://www.charmm.org/	A molecular simulation program with broad application to many-particle systems with a comprehensive set of energy functions, a variety of enhanced sampling methods, and support for multi-scale techniques including QM/MM, MM/CG, and a range of implicit solvent models	Public	Han et al., (2021)
CDK 2.0 descriptors	https://www.redhat.com/zh/about/videos/cdk-20-docker-kubernetes-and-ose-your-desk	The Chemistry Development Kit (CDK) is a widely used open source cheminformatics toolkit, providing data structures to represent chemical concepts along with methods to manipulate such structures and perform computations on them. The library implements a wide variety of cheminformatics algorithms ranging from chemical structure canonicalization to molecular descriptor calculations and pharmacophore perception. It is used in drug discovery, metabolomics, and toxicology	Public	Hua et al., (2021) Huang et al., (2021)
Chemaxon	https://chemaxon.com/	As the world's leading professional chemical software, ChemAxon is a Java-based cross-platform comprehensive software package for pharmaceutical informatics	Public	Huang et al., (2021)

Table 2 (continued)

Software platforms	Website	Brief description	Access (public/owner)	Ref.
Dragon software	http://www.talete.mi.it/products/dragon_description.htm	Dragon is an application for the calculation of molecular descriptors. These descriptors can be used to evaluate molecular structure-activity or structure-property relationships, as well as for similarity analysis and high-throughput screening of molecule databases. Actually Dragon is widely used in scientific studies as well as part of several QSAR suites	Public	Hua et al., (2021) Huang et al., (2021) Raitano et al., (2018)
Discovery Studio 3.5	http://accelrys.com/products/discovery-studio/	Discovery Studio, a comprehensive suite of modeling and simulation solutions for protein modeling and computational chemistry from Accelrys Inc., includes a diverse collection of sophisticated software applications and provides advanced software solutions for life science researchers	Public	Han et al., (2022) Zhang et al., (2020)

of the drug targets were found to be shared by two or more NSAIDs, providing a mechanistic rationale for the assessment of potential NSAIDs-mixture effects. Therefore, this study (Marmon et al., 2021) paved a way to explore the potential target-mediated effects of mixtures of PECs co-occurring in the same environmental matrix using *in silico* tools.

Software platforms and databases used in the included studies

Our review summarized the software platforms (Table 2) and databases (Table 3) used in the included studies. Results showed that a total of 26 software platforms and 15 databases have been employed in the included 20 studies reported the in silico prediction of the environmental risks posed by PECs.

As shown in Table 2, the included studies used different software platforms for different study purposes. In the field of in silico prediction for bioaccumulation and biodegradability of PECs, the EPI Suite™ (Della-Flora et al., 2020; Raitano et al., 2018; Regnery et al., 2022; Wielens Becker et al., 2020), VEGA (Della-Flora et al., 2020; Raitano et al., 2018; Sanabria et al., 2021; Wielens Becker et al., 2020), Prometheus (Della-Flora et al., 2020; Sanabria et al., 2021; Wielens Becker et al., 2020), and QSAR Toolbox (Della-Flora et al., 2020; Raitano et al., 2018; Sanabria et al., 2021) were relatively more commonly employed. A vast majority of these tools are public-access and available for free. Although the included publications are addressing this subject, there is no software platform where all components for ERA of PECs can be seen as a whole.

The success of an in silico predictive model is highly dependent on the datasets or databases from which the related information is extracted. In addition to few studies (Han et al., 2022; Huang et al., 2021; Kumar et al., 2021; Zhang et al., 2020) exclusively using the self-developed data sets unavailable via the Internet, major data sources supporting prediction for the environmental risks of PECs are open to public availability (Table 3). The availability of these public datasets facilitates the collection of a large number of pharmaceuticals/PECs' structural substances and the associated experimental data for in silico modeling purposes. However, we found that, in general, most of in silico models in the included studies are built on different datasets; therefore, their performances are

Table 3 An illustrative list of databases recently used the *in silico* prediction of environmental risks posed by PECs (the self-developed data sets unavailable via the Internet were excluded)

Database	Data size (compounds)	Website	Brief description	Ref.
EAWAG-BBD	1396	http://eawag-bbd.ethz.ch/	This database contains information on microbial biocatalytic reactions and biodegradation pathways for primarily xenobiotic, chemical compounds. The goal of the EAWAG-BBD is to provide information on microbial enzyme-catalyzed reactions that are important for biotechnology	Garcia-Martin et al., (2020)
PPDB	1888	http://sitem.herts.ac.uk/aeru/ppdb/en/index.htm	The University of Hertfordshire Pesticide Properties DataBase (PPDB) is a comprehensive relational database of pesticide chemical identity, physicochemical, human health, and ecotoxicological data. It has been developed by the Agriculture & Environment Research Unit (AERU) at the University of Hertfordshire for a variety of end users to support risk assessments and risk management	Garcia-Martin et al., (2020) Raitano et al., (2018)
Drugbank 5.1.10	15,419	https://go.drugbank.com/	DrugBank Online is a comprehensive, free-to-access, online database containing information on drugs and drug targets. As both a bioinformatics and a cheminformatics resource, this database combine detailed drug (i.e., chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e., sequence, structure, and pathway) information	Garcia-Martin et al., (2020) Hua et al., (2021) Kumar et al., (2022)
ChEMBL	2,331,700	https://www.ebi.ac.uk/chembl/	ChEMBL is a manually curated database of bioactive molecules with drug-like properties. It brings together chemical, bioactivity, and genomic data to aid the translation of genomic information into effective new drugs	Marmon et al., (2021)

Table 3 (continued)

Database	Data size (compounds)	Website	Brief description	Ref.
CompTox Chemicals Dashboard	1,200,059	https://comptox.epa.gov/dashboard	The CompTox Chemicals Dashboard is a widely used resource for chemistry, toxicity, and exposure information of hundreds of thousands of chemicals. The available data includes, but is not limited to, chemical properties, environmental fate and transport, hazard (e.g., point of departure, legacy toxicity values, screening levels, exposure limits), in vitro to in vivo extrapolation (IVIVE), exposure (e.g., predictions associated with the ExpoCast program, use categories, toxic release inventory, monitoring data), and bioactivity (e.g., high-throughput data from ToxCast and Tox21 efforts)	Saavedra & Duchowicz, (2021) Spînu et al., (2022)
EURAS	863	https://ambit.sourceforge.net/euras/	EURAS bioconcentration factor (BCF) Gold Standard Database included 863 biological concentration factor records	Miller et al., (2019)
METLIN	> 960,000	https://metlin.scripps.edu/landing_page.php?pgcontent=mainPage	METLIN provides links and information for every one of its 960,000 compounds. These include name, systematic name, structure, elemental formula, mass, CAS number, KEGG ID and link, HMDB ID and link, PubChem ID and link, commercial availability, and direct search options on the molecule itself	Wielens Becker et al., (2020)
Chemspider	115 million	http://www.chemspider.com/	ChemSpider is a free chemical structure database providing fast text and structure search access to over 100 million structures from hundreds of data sources	Miller et al., (2019) Raitano et al., (2018)
SIDER	1430	http://sideeffects.embl.de/	SIDER database contains information on marketed medicines and their recorded adverse drug reactions. The information is extracted from public documents and package inserts. The available information include side effect frequency, drug and side effect classifications as well as links to further information, for example drug–target relations	Hua et al., (2021)
PubChem	113,979,487	https://pubchem.ncbi.nlm.nih.gov/	PubChem is the world's largest collection of freely accessible chemical information, allows to search chemicals by name, molecular formula, structure, and other identifiers	Han et al., (2021)

Table 3 (continued)

Database	Data size (compounds)	Website	Brief description	Ref.
ECOTOX	12,714	https://cfpub.epa.gov/ecotox/	ECOTOX database integrates three previously independent databases - AQUIRE, PHYTOTOX, and TERRETOX - into a unique system which includes toxicity data derived predominately from the peer-reviewed literature, for aquatic life, terrestrial plants, and terrestrial wildlife, respectively	Raitano et al., (2018)
IRIS	572	https://www.epa.gov/iris	IRIS database provides information on how chemicals affect human health and is a primary source of EPA risk-assessment information on chemicals of environmental concern	Raitano et al., (2018)
RAIS	2871	https://rais.ornl.gov/cgi-bin/tools/TOX_search?select=chem	RAIS database includes Chemical Toxicity Metadata, Chemical Toxicity Values, Chemical Specific Parameters, etc.	Raitano et al., (2018)
ToxCast	> 4500	https://www.epa.gov/chemical-research/exploring-toxcast-data-downloadable-data	ToxCast has data for chemicals from a broad range of sources including industrial and consumer products, food additives, and potentially green chemicals that could be safer alternatives to existing chemicals	Marmon et al., (2021) Saavedra & Duchowicz, (2021)
UFZ-LSER	3700	https://www.ufz.de/index.php?en=31698&contentonly=1&m=0&lserd_data[mvc]=Public/start	This database contains experimentally determined Linear Solvation Energy Relationship (LSER) descriptors, E (excess molar refraction), S (polarizability/dipolarity), A (solute hydrogen bond acidity), B, and/or B0 (solute hydrogen bond basicity), and L (logarithmic gas-hexadecane partition coefficient) of about 3700 compounds reported in the literature. The database also stores the McGowan's molar volume V calculated for each compound from its molecular structure. These solute descriptors can be used, in combination with LSER equations or other types of polyparameter linear free energy relationships (pp-LFERs), to calculate partition coefficients for various systems	Regnery et al., (2022)
NORMAN	Not available	https://www.norman-network.com/nds/	NORMAN organizes the development and maintenance of various web-based databases for the collection and evaluation of data/information on emerging substances in the environment	Badry et al., (2022)

Table 3 (continued)

Database	Data size (compounds)	Website	Brief description	Ref.
IPCHEM	Not available	https://ipchem.jrc.ec.europa.eu	IPCHEM aims to support a more coordinated approach for collecting, storing, accessing and assessing data related to the occurrence of chemicals and chemical mixtures, in relation to humans and the environment	Badry et al., (2022)
ISS-INAIL	Not available	https://www.iss.it/en/basi-di-dati	A large database containing Remediation database, Carcinogenic database, Safety data sheet template database, Sensitizing database, Chemical Labeling, ISSTOX Chemical Toxicity Databases, Exempt rare diseases database, etc.	Raitano et al., (2018)
CMap database	Not available	https://cmap.ihmc.us/	CMap is an expression profile database based on intervention gene expression developed by the Broad Institute. It is widely used in pharmacogenomics research by computational biologists and drug screening researchers to reveal functional links between small molecule compounds, genes, and disease states. CMap not only provides gene expression profiling data before and after drug treatment in human cell lines, but also provides online tools based on pattern matching algorithms	Liu et al., (2021)

not comparable with each other. Therefore, it might be necessary to develop standardized and accepted in silico models specially designed to meet key case scenarios (e.g., bioaccumulation and biodegradability) within the prediction of environmental risks associated with PECs, through integration of pharmaceutical data and specialized tool sets for ERA.

Conclusions

As high throughput screening tools for potential environmental risks posed by a wide range of PEC candidates in an economical, fast, and time-saving manner, the *in silico* approaches have been considered to be promising to efficiently and comprehensively evaluate drug-related data sets/databases for potential ecological risks posed by pharmaceuticals when occurring in the environment. This review focused on the current knowledge on the application of *in silico* approaches to support ERA of PECs reported in 20 articles that have appeared since 2018. We found that the researchers' interest and concern appeared to be sharply aroused by the outbreak of COVID-19 crisis, but the included studies were from only 8 countries around the world. Recently, a great diversity of *in silico* techniques have been widely employed for prediction of endpoints for ERA of PECs, including bioaccumulation and biodegradability, lethality, developmental toxicity, mutagenicity, and other toxic effects, such as ototoxicity and hematological toxicity. Moreover, several researchers have *in silico* assessed the potential human health hazards of exposure to PECs based on human data. In addition, transcriptomic analysis and network pharmacology approaches have been used to simultaneously discern a broad spectrum of potential environmental risks and health effects of PECs based on the network-based drug-gene-disease relationship. However, considerations of environmental exposure concentration for PECs and interactions among mixtures of PECs were not sufficiently addressed in the included studies. Despite a great diversity of software platforms and databases that the scientific community recently used for *in silico* prediction of the environmental risks posed by PECs, hardly any among them is specific to ERA of PECs. In conclusion, *in silico* prediction of the environmental risks posed by PECs is still in its infancy with limited examples, thus creating both opportunities and challenges for researchers.

Author contributions J.G.: literature search, data analysis, and drafting. J.Z.: literature search, data analysis, and drafting. X.C.: data analysis, drafting and editing. J.W.: conceptualization, supervision, and critical review.

Funding This work was supported by the National Natural Science Foundation of China (No. 71974153), Humanities and Social Science Foundation from Hubei Provincial Department of Education (No.22D025) and the Provincial University Students' Innovative Entrepreneurial Training Program in Hubei (No. S202310488166).

Data availability Not applicable.

Declarations

Ethics approval Not applicable.

Consent for publication All authors agree with publishing.

Competing interests The authors declare no competing interests.

References

Anand, U., Adelodun, B., Cabreros, C., Kumar, P., Suresh, S., Dey, A., Ballesteros, F., Jr., & Bontempi, E. (2022). Occurrence, transformation, bioaccumulation, risk and analysis of pharmaceutical and personal care products from wastewater: A review. *Environmental Chemistry Letters, 20*(6), 3883–3904. https://doi.org/10.1007/s10311-022-01498-7

Ankley, G. T., Berninger, J. P., Maloney, E. M., Olker, J. H., Schaupp, C. M., Villeneuve, D. L., & Lalone, C. A. (2022). Linking mechanistic effects of pharmaceuticals and personal care products to ecologically relevant outcomes: A decade of progress. *Environmental Toxicology and Chemistry*. https://doi.org/10.1002/etc.5416

Atzei, A., Jense, I., Zwart, E. P., Legradi, J., Venhuis, B. J., Van Der Ven, L. T. M., Heusinkveld, H. J., & Hessel, E. V. S. (2021). Developmental neurotoxicity of environmentally relevant pharmaceuticals and mixtures thereof in a zebrafish embryo behavioural test. *International Journal of Environmental Research and Public Health, 18*(13), 6717. https://doi.org/10.3390/ijerph18136717

Badry, A., Treu, G., Gkotsis, G., Nika, M. C., Alygizakis, N., Thomaidis, N. S., Voigt, C. C., & Krone, O. (2022). Ecological and spatial variations of legacy and emerging contaminants in white-tailed sea eagles from Germany: Implications for prioritisation and future risk management. *Environment International, 158*, 106934. https://doi.org/10.1016/j.envint.2021.106934

Cunha, D. L., Mendes, M. P., & Marques, M. (2019). Environmental risk assessment of psychoactive drugs in the aquatic environment. *Environmental Science and Pollution Research International, 26*(1), 78–90. https://doi.org/10.1007/s11356-018-3556-z

De Vaugelade, S., Nicol, E., Vujovic, S., Bourcier, S., Pirnay, S., & Bouchonnet, S. (2017). UV-vis degradation of alpha-tocopherol in a model system and in a cosmetic emulsion-Structural elucidation of photoproducts and toxicological consequences. *Journal of Chromatography. A, 1517*, 126–133. https://doi.org/10.1016/j.chroma.2017.08.015

Della-Flora, A., Wilde, M. L., Pinto, I. D. F., Lima, E. C., & Sirtori, C. (2020). Degradation of the anticancer drug flutamide by solar photo-Fenton treatment at near-neutral pH: Identification of transformation products and in silico (Q)SAR risk assessment. *Environmental Research, 183*, 109223. https://doi.org/10.1016/j.envres.2020.109223

Fonseca, V. F., Duarte, I. A., Duarte, B., Freitas, A., Pouca, A. S. V., Barbosa, J., Gillanders, B. M., & Reis-Santos, P. (2021). Environmental risk assessment and bioaccumulation of pharmaceuticals in a large urbanized estuary. *Science of the Total Environment, 783*, 147021. https://doi.org/10.1016/j.scitotenv.2021.147021

Garcia-Martin, J. A., Chavarria, M., De Lorenzo, V., & Pazos, F. (2020). Concomitant prediction of environmental fate and toxicity of chemical compounds. *Biology Methods and Protocols, 5*(1), bpaa025. https://doi.org/10.1093/biomethods/bpaa025

Gouveia, T. I. A., Alves, A., & Santos, M. S. F. (2019). New insights on cytostatic drug risk assessment in aquatic environments based on measured concentrations in surface waters. *Environment International, 133*(Pt B), 105236. https://doi.org/10.1016/j.envint.2019.105236

Guedes-Alonso, R., Montesdeoca-Esponda, S., Pacheco-Juarez, J., Sosa-Ferrera, Z., & Santana-Rodriguez, J. J. (2020). A survey of the presence of pharmaceutical residues in wastewaters. Evaluation of their removal using conventional and natural treatment procedures. *Molecules, 25*(7), 1639. https://doi.org/10.3390/molecules25071639

Guo, J., Liao, M., He, B., Liu, J., Hu, X., Yan, D., & Wang, J. (2021). Impact of the COVID-19 pandemic on household disinfectant consumption behaviors and related environmental concerns: A questionnaire-based survey in China. *Journal of Environmental Chemical Engineering, 9*(5), 106168. https://doi.org/10.1016/j.jece.2021.106168

Han, Y., Ma, Y., Chen, B., Zhang, J., & Hu, C. (2022). Hazard assessment of beta-lactams: Integrating in silico and QSTR approaches with in vivo zebrafish embryo toxicity testing. *Ecotoxicology and Environmental Safety, 229*, 113106. https://doi.org/10.1016/j.ecoenv.2021.113106

Han, Y., Ma, Y., Yao, S., Zhang, J., & Hu, C. (2021). In vivo and in silico evaluations of survival and cardiac developmental toxicity of quinolone antibiotics in zebrafish embryos (Danio rerio). *Environmental Pollution, 277*, 116779. https://doi.org/10.1016/j.envpol.2021.116779

He, B. S., Wang, J., Liu, J., & Hu, X. M. (2017). Eco-pharmacovigilance of non-steroidal anti-inflammatory drugs: Necessity and opportunities. *Chemosphere, 181*, 178–189. https://doi.org/10.1016/j.chemosphere.2017.04.084

Holm, G., Snape, J. R., Murray-Smith, R., Talbot, J., Taylor, D., & Sorme, P. (2013). Implementing ecopharmacovigilance in practice: Challenges and potential opportunities. *Drug Safety, 36*(7), 533–546. https://doi.org/10.1007/s40264-013-0049-3

Hu, L. X., Olaitan, O. J., Li, Z., Yang, Y. Y., Chimezie, A., Adepoju-Bello, A. A., Ying, G. G., & Chen, C. E. (2021). What is in Nigerian waters? Target and non-target screening analysis for organic chemicals. *Chemosphere, 284*, 131546. https://doi.org/10.1016/j.chemosphere.2021.131546

Hua, Y., Shi, Y., Cui, X., & Li, X. (2021). In silico prediction of chemical-induced hematotoxicity with machine learning and deep learning methods. *Molecular Diversity, 25*(3), 1585–1596. https://doi.org/10.1007/s11030-021-10255-x

Huang, X., Tang, F., Hua, Y., & Li, X. (2021). In silico prediction of drug-induced ototoxicity using machine learning and deep learning methods. *Chemical Biology & Drug Design, 98*(2), 248–257. https://doi.org/10.1111/cbdd.13894

Jiao, B., Zhu, Y., Xu, J., Dong, F., Wu, X., Liu, X., & Zheng, Y. (2022). Identification and ecotoxicity prediction of pyrisoxazole transformation products formed in soil and water using an effective HRMS workflow. *Journal of Hazardous Materials, 424*(Pt A), 127223. https://doi.org/10.1016/j.jhazmat.2021.127223

Jose, J., Sandra Pinto, J., Kotian, B., Mathew Thomas, A., & Narayana Charyulu, R. (2020). Comparison of the regulatory outline of ecopharmacovigilance of pharmaceuticals in Europe, USA. *Japan and Australia. Sci Total Environ, 709*, 134815. https://doi.org/10.1016/j.scitotenv.2019.134815

Kar, S., Roy, K., & Leszczynski, J. (2018). Impact of pharmaceuticals on the environment: Risk assessment using QSAR modeling approach. *Methods in Molecular Biology, 1800*, 395–443. https://doi.org/10.1007/978-1-4939-7899-1_19

Koltsakidou, A., Antonopoulou, M., Epsilonvgenidou, E., Konstantinou, I., & Lambropoulou, D. (2019). A comparative study on the photo-catalytic degradation of cytarabine anticancer drug under Fe(3+/)H(2)O(2), Fe(3+/)S(2)O(8)(2-), and [Fe(C(2)O(4))(3)](3-)/H(2)O(2) processes. Kinetics, identification, and in silico toxicity assessment of generated transformation products. *Environmental Science and Pollution Research International, 26*(8), 7772–7784. https://doi.org/10.1007/s11356-018-4019-2

Kumar, R., Khan, F. U., Sharma, A., Siddiqui, M. H., Aziz, I. B., Kamal, M. A., Ashraf, G. M., Alghamdi, B. S., & Uddin, M. S. (2021). A deep neural network-based approach for prediction of mutagenicity of compounds. *Environmental Science and Pollution Research International, 28*(34), 47641–47650. https://doi.org/10.1007/s11356-021-14028-9

Kumar, S., Deepika, D., & Kumar, V. (2022). Pharmacophore modeling using machine learning for screening the blood-brain barrier permeation of xenobiotics. *International Journal of Environmental Research and Public Health, 9*(20), 13471. https://doi.org/10.3390/ijerph192013471

Kunene, P. N., & Mahlambi, P. N. (2023). Case study on antiretroviral drugs uptake from soil irrigated with contaminated water: Bio-accumulation and bio-translocation to roots, stem, leaves, and fruits. *Environmental Pollution, 319*, 121004. https://doi.org/10.1016/j.envpol.2023.121004

Li, S., Wen, J., He, B., Wang, J., Hu, X., & Liu, J. (2020). Occurrence of caffeine in the freshwater environment: Implications for ecopharmacovigilance. *Environmental Pollution, 263*(Pt B), 114371. https://doi.org/10.1016/j.envpol.2020.114371

Li, S., He, B., Wang, J., Liu, J., & Hu, X. (2020). Risks of caffeine residues in the environment: Necessity for a targeted ecopharmacovigilance program. *Chemosphere, 243*, 125343. https://doi.org/10.1016/j.chemosphere.2019.125343

Liu, X., Zhang, H., Xue, Q., Pan, W., & Zhang, A. (2021). In silico health effect prioritization of environmental chemicals through transcriptomics data exploration from a chemo-centric view. *Science of the Total Environment, 762*, 143082. https://doi.org/10.1016/j.scitotenv.2020.143082

Marmon, P., Owen, S. F., & Margiotta-Casaluci, L. (2021). Pharmacology-informed prediction of the risk posed to fish by mixtures of non-steroidal anti-inflammatory drugs (NSAIDs) in the environment. *Environment International, 146*, 106222. https://doi.org/10.1016/j.envint.2020.106222

Miller, T. H., Bury, N. R., Owen, S. F., Macrae, J. I., & Barron, L. P. (2018). A review of the pharmaceutical exposome in aquatic fauna. *Environmental Pollution, 239*, 129–146. https://doi.org/10.1016/j.envpol.2018.04.012

Miller, T. H., Gallidabino, M. D., Macrae, J. I., Owen, S. F., Bury, N. R., & Barron, L. P. (2019). Prediction of bioconcentration factors in fish and invertebrates using machine learning. *Science of the Total Environment, 648*, 80–89. https://doi.org/10.1016/j.scitotenv.2018.08.122

Molnar, E., Maasz, G., & Pirger, Z. (2021). Environmental risk assessment of pharmaceuticals at a seasonal holiday destination in the largest freshwater shallow lake in Central Europe. *Environmental Science and Pollution Research International, 28*(42), 59233–59243. https://doi.org/10.1007/s11356-020-09747-4

Moradi, M., Golmohammadi, R., Najafi, A., Moosazadeh Moghaddam, M., Fasihi-Ramandi, M., & Mirnejad, R. (2022). A contemporary review on the important role of in silico approaches for managing different aspects of COVID-19 crisis. *Informatics in Medicine Unlocked, 28*, 100862. https://doi.org/10.1016/j.imu.2022.100862

Morales-Paredes, C. A., Rodriguez-Diaz, J. M., & Boluda-Botella, N. (2022). Pharmaceutical compounds used in the COVID-19 pandemic: A review of their presence in water and treatment techniques for their elimination. *Science of the Total Environment, 814*, 152691. https://doi.org/10.1016/j.scitotenv.2021.152691

Nendza, M., Kuhne, R., Lombardo, A., Strempel, S., & Schuurmann, G. (2018). PBT assessment under REACH: Screening for low aquatic bioaccumulation with QSAR classifications based on physicochemical properties to replace BCF in vivo testing on fish. *Science of the Total Environment, 616-617*, 97–106. https://doi.org/10.1016/j.scitotenv.2017.10.317

Nieto-Juarez, J. I., Torres-Palma, R. A., Botero-Coy, A. M., & Hernandez, F. (2021). Pharmaceuticals and environmental risk assessment in municipal wastewater treatment plants and rivers from Peru. *Environment International, 155*, 106674. https://doi.org/10.1016/j.envint.2021.106674

Nozaki, K., Tanoue, R., Kunisue, T., Tue, N. M., Fujii, S., Sudo, N., Isobe, T., Nakayama, K., Sudaryanto, A., Subramanian, A., Bulbule, K. A., Parthasarathy, P., Tuyen, L. H., Viet, P. H., Kondo, M., Tanabe, S., & Nomiyama, K. (2023). Pharmaceuticals and personal care products (PPCPs) in surface water and fish from three Asian countries: Species-specific bioaccumulation and potential ecological risks. *Science of the Total Environment, 866*, 161258. https://doi.org/10.1016/j.scitotenv.2022.161258

Ofrydopoulou, A., Evgenidou, E., Nannou, C., Vasquez, M. I., & Lambropoulou, D. (2021). Exploring the phototransformation and assessing the in vitro and in silico toxicity of a mixture of pharmaceuticals susceptible to photolysis. *Science of the Total Environment, 756*, 144079. https://doi.org/10.1016/j.scitotenv.2020.144079

Raimondo, S., & Forbes, V. E. (2022). Moving beyond risk quotients: Advancing ecological risk assessment to reflect better, more robust and relevant methods. *Ecologies (Basel), 3*(2), 145–160. https://doi.org/10.3390/ecologies3020012

Raitano, G., Goi, D., Pieri, V., Passoni, A., Mattiussi, M., Lutman, A., Romeo, I., Manganaro, A., Marzo, M., Porta, N., Baderna, D., Colombo, A., Aneggi, E., Natolino, F., Lodi, M., Bagnati, R., & Benfenati, E. (2018). (Eco)toxicological maps: A new risk assessment method integrating traditional and in silico tools and its application in the Ledra River (Italy). *Environment International, 119*, 275–286. https://doi.org/10.1016/j.envint.2018.06.035

Regnery, J., Riegraf, C., Jacob, S., & Friesen, A. (2022). New insights on in vitro biotransformation of anticoagulant rodenticides in fish. *Chemosphere, 294*, 133727. https://doi.org/10.1016/j.chemosphere.2022.133727

Riva, F., Zuccato, E., Davoli, E., Fattore, E., & Castiglioni, S. (2019). Risk assessment of a mixture of emerging contaminants in surface water in a highly urbanized area in Italy. *Journal of Hazardous Materials, 361*, 103–110. https://doi.org/10.1016/j.jhazmat.2018.07.099

Saavedra, L. M., & Duchowicz, P. R. (2021). Predicting zebrafish (Danio rerio) embryo developmental toxicity through a non-conformational QSAR approach. *Science of the Total Environment, 796*, 148820. https://doi.org/10.1016/j.scitotenv.2021.148820

Sanabria, P., Scunderlick, D., Wilde, M. L., Ludtke, D. S., & Sirtori, C. (2021). Solar photo-Fenton treatment of the anti-cancer drug anastrozole in different aqueous matrices at near-neutral pH: Transformation products identification, pathways proposal, and in silico (Q)SAR risk assessment. *Science of the Total Environment, 754*, 142300. https://doi.org/10.1016/j.scitotenv.2020.142300

Sharif, A., Ashraf, M., Anjum, A. A., Javeed, A., Altaf, I., Akhtar, M. F., Abbas, M., Akhtar, B., & Saleem, A. (2016). Pharmaceutical wastewater being composite mixture of environmental pollutants may be associated with mutagenicity and genotoxicity. *Environmental Science and Pollution Research International, 23*(3), 2813–2820. https://doi.org/10.1007/s11356-015-5478-3

Sharifi, A., Ahmadi, M., & Ala, A. (2021). The impact of artificial intelligence and digital style on industry and energy post-COVID-19 pandemic. *Environmental Science and Pollution Research International, 28*(34), 46964–46984. https://doi.org/10.1007/s11356-021-15292-5

Spînu, N., Cronin, M. T. D., Lao, J., Bal-Price, A., Campia, I., Enoch, S. J., Madden, J. C., Mora Lagares, L., Novič, M., Pamies, D., Scholz, S., Villeneuve, D. L., & Worth, A. P. (2022). Probabilistic modelling of developmental neurotoxicity based on a simplified adverse outcome pathway

network. *Computational Toxicology, 21*, 100206. https://doi.org/10.1016/j.comtox.2021.100206

Thomas, P. C., Bicherel, P., & Bauer, F. J. (2018). How in silico and QSAR approaches can increase confidence in environmental hazard and risk assessment. *Integrated Environmental Assessment and Management, 15*(1), 40–50. https://doi.org/10.1002/ieam.4108

Tung, C. W., Lin, Y. H., & Wang, S. S. (2019). Transfer learning for predicting human skin sensitizers. *Archives of Toxicology, 93*(4), 931–940. https://doi.org/10.1007/s00204-019-02420-x

Wang, J., Li, S., Zhu, Y., Guo, J., Liu, J., & He, B. (2021). Targeted eco-pharmacovigilance as an optimized management strategy for adverse effects of pharmaceuticals in the environment. *Environmental Toxicology and Pharmacology, 82*, 103565. https://doi.org/10.1016/j.etap.2020.103565

Wielens Becker, R., Ibanez, M., Cuervo Lumbaque, E., Wilde, M. L., Flores Da Rosa, T., Hernandez, F., & Sirtori, C. (2020). Investigation of pharmaceuticals and their metabolites in Brazilian hospital wastewater by LC-QTOF MS screening combined with a preliminary exposure and in silico risk assessment. *Science of the Total Environment, 699*, 134218. https://doi.org/10.1016/j.scitotenv.2019.134218

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research, 46*(D1), D1074–D1082. https://doi.org/10.1093/nar/gkx1037

Xu, M., Yang, H., Liu, G., Tang, Y., & Li, W. (2022). In silico prediction of chemical aquatic toxicity by multiple machine learning and deep learning approaches. *Journal of Applied Toxicology, 42*(11), 1766–1776. https://doi.org/10.1002/jat.4354

Zhang, H., Liu, C. T., Mao, J., Shen, C., Xie, R. L., & Mu, B. (2020). Development of novel in silico prediction model for drug-induced ototoxicity by using naive Bayes classifier approach. *Toxicology In Vitro, 65*, 104812. https://doi.org/10.1016/j.tiv.2020.104812