Trends in Food Science & Technology 148 (2024) 104513

Contents lists available at ScienceDirect

Trends in Food Science & Technology

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

Unveiling the chemical complexity of food-risk components: A comprehensive data resource guide in 2024

Dachuan Zhang [a,*,1], Dongliang Liu [b,1], Jiayi Jing [c], Bingxuan Jia [b], Ye Tian [b], Yingying Le [b], Yaochun Yu [d], Qian-Nan Hu [b]

[a] Institute of Environmental Engineering, ETH Zurich, Laura-Hezner-Weg 7, 8093, Zurich, Switzerland
[b] CAS Key Laboratory of Computational Biology, CAS Key Laboratory of Nutrition, Metabolism and Food Safety, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, 200031, China
[c] Bioinformatics Group, Wageningen University and Research, 6708PB, Wageningen, the Netherlands
[d] Department of Environmental Chemistry, Swiss Federal Institute of Aquatic Science and Technology (Eawag), 8600 Dübendorf, Switzerland

ARTICLE INFO

Handling Editor: Dr I Oey

Keywords:
Food informatics
Big data
Food safety
Food components
Mycotoxins
Pesticides

ABSTRACT

Background: With the influence of climate change, environmental pollution, global industrialization, and new agricultural practices, increasing amounts of chemical substances with potential risks—both anthropogenic and biogenic—enter the food supply chain, entailing new challenges to food safety and security. Although some food-risk components (FRCs) have been accessed and regulated, the toxicity and exposure level of the numerous components detected in food remain unknown, leaving questions on their effect on food safety. Therefore, multiple databases on emerging FRCs have been constructed to aid in food safety assessment, regulation, and communication; however, their focus areas, data content, quality, and accessibility have not been systematically summarized, which hinders their applications and the development of data-driven methods in the food safety field.

Scope and approach: The major objective of this review is to introduce representative FRC databases with different focus areas, along with their chemical presentation, data quality and availability, and successful applications.

Key findings and conclusions: Over the past decades, over 50 FRC databases have been released, contributing significantly to scientific research, policymaking, and education. However, our analysis unveils persistent challenges such as delayed updates, accessibility concerns, reproducibility issues, suboptimal data quality, and inadequate coverage in underdeveloped regions. To address these shortcomings, we propose an initiative aimed at enhancing future FRC databases, prioritizing the principles of findability, accessibility, interoperability, and reusability. Additionally, we highlight the potential of future strategies, e.g., natural language processing, cheminformatics-enpowered suspect and non-targeted analysis, and genome mining, for the detection and analysis of emerging new FRCs outside of existing databases. By embracing these initiatives and strategies, we lay the groundwork for a robust framework facilitating enhanced food safety assessment and informed decision-making in the face of evolving challenges.

1. Introduction

Food-risk components (FRCs) pose global concerns and trade obstacles, jeopardizing public health and food systems resilience (Jin et al., 2020; Zhang et al., 2020). FRCs include both anthropogenic and biogenic chemicals, such as pesticide residues, environmental pollutants, illegal or risky additives, and biogenic food toxins (Shi, et al., 2022). According to estimates from the World Health Organization (WHO), FRC-contaminated food causes disease in 1 in 10 people, resulting in 420,000 deaths annually (WHO, 2023).

With the process of global industrialization and subsequent environmental pollution, such as metal mining and smelting, industrial wastewater discharge, solid waste stacking, and pesticide use, the number of detected anthropogenic pollutants in food continues to rise

* Corresponding author.
E-mail addresses: dachuan.zhang@ifu.baug.ethz.ch (D. Zhang), liudongliang2018@sinh.ac.cn (D. Liu), jiayi.jing@wur.nl (J. Jing), jiabingxuan2017@sinh.ac.cn (B. Jia), tianye@sibs.ac.cn (Y. Tian), yyle@sibs.ac.cn (Y. Le), yaochun.yu@eawag.ch (Y. Yu), qnhu@sibs.ac.cn (Q.-N. Hu).
[1] Contributed equally.

https://doi.org/10.1016/j.tifs.2024.104513
Received 1 January 2024; Received in revised form 23 March 2024; Accepted 24 April 2024
Available online 26 April 2024

(Z. Wang, Walker, Muir, & Nagatani-Yoshida, 2020). One primary route through which anthropogenic pollutants enter food crops and livestock is when agricultural practices take place in proximity to pollution sources, such as when soil is contaminated by nearby industrial activity or when irrigation relies on water tainted with pollutants (Adnan, et al., 2022). Large-scale investigations have shown the negative impacts of pesticide residues on soil, terrestrial, and aquatic ecosystems, and positive correlations with the risk of neurological and reproductive diseases and cancer, and veterinary drug residues have been proven to cause drug-induced deafness, liver and kidney damage, bacterial imbalance, and carcinogenic effects (Bacanli & Basaran, 2019; Carvalho, 2017; Thompson et al., 2017). Infants and toddlers, in particular, are susceptible to these contaminants present in food (Mielech, Puscion-Jakubik, & Socha, 2021). Prolonged exposure to these substances can result in adverse health effects, such as developmental delays (Mielech, et al., 2021). Older adults and immunocompromised patients also represent a high-risk group for foodborne illnesses due to several factors, including weakened immune systems, chronic illnesses, and the resulting increased likelihood of complications (Thaivalappil, Young, Paco, Jeyapalan, & Papadopoulos, 2020).

Besides anthropogenic pollutants, biogenic toxins that exist in plants and animals or are produced by fungi and bacteria also adversely affect human health. Mycotoxins are secondary toxic fungal metabolites, mainly produced by fungi of the genus *Aspergillus*, *Fusarium*, *Alternaria*, and *Penicillium* that cause contamination of food, vegetables, and fruits (Tian, et al., 2022). Bacterial toxins are bacteria-produced toxic molecules, polypeptides, and proteins that can affect the extracellular matrix and cell membranes (Babiychuk & Draeger, 2015). Editable plants may also contain certain toxic metabolites, such as glycoalkaloids in potatoes and aristolochic acid in fish mint, that can cause liver and kidney damage and other health risks (Kemprai, et al., 2023). In addition, some peptides and proteins that exist in animal products can interfere with the activities of key enzymes, receptors, and ion channels in human metabolism, and cause damage to homeostasis and the nervous and cardiovascular systems (Kuzmenkov, Krylov, Chugunov, Grishin, & Vassilevski, 2016).

Furthermore, food adulteration and fraud also affect the safety of food products. The US Grocery Manufacturers Association estimates global food fraud and adulteration costs to range between $10 billion and $15 billion annually, accounting for approximately 10% of all commercial food sales (Johnson, 2014)). In adulteration activities, low-quality raw materials and illegal additives that are outside the allowed list are often added to food to extend shelf life, enhance color, mask unpleasant flavors, and improve appearance and texture illegally (Peng, et al., 2017). In addition, the improper use of food additives can also introduce potential risks. Although food additives have undergone strict toxicological and safety evaluation during approval processes, some may still pose potential health risks to certain communities. For instance, children's attention deficit hyperactivity disorder, colitis, metabolic disorders, and obesity have been linked to the improper use of certain food additives (McCann, et al., 2007; Neltner, Alger, Leonard, & Maffini, 2013; Roca-Saavedra, et al., 2018). Because of these potential health risks, the European Food Safety Authority (ESFA) has begun to re-evaluate the safety of food additives since 2008 (EFSA, et al., 2021).

Over time, as regulations and evaluations of FRCs have evolved, a substantial amount of data have been accumulated, leading to the establishment of FRC databases (Arvidson, 2008; EFSA et al., 2019; Manning & Soon, 2016). These databases mainly include three types: *standard*, *large-scale general,* and *domain-specific scientific* databases. *Standard databases*, such as the WHO/FAO Codex General Standard for Food Additives online database, usually led by international organizations or regulatory agencies, record regulations and standards of FRCs. *Large-scale general databases*, such as ToxCast (Richard, et al., 2016), contain data on numerous chemical substances, some of which are related to food safety. *Domain-specific scientific databases*, such as AdditiveChem (D. Zhang, Cheng, et al., 2020) provide high-quality data on specific fields and customized functions for data query and analysis. These databases have been widely used for policymaking, scientific research, and education and have promoted the development of data-driven approaches, including computational toxicology, food informatics, and emerging artificial intelligence (AI)-based approaches. Conventional risk assessment and analysis often struggle to keep pace with the gradually increasing complexity of FRCs, while data-driven approaches offer a systematic and efficient means of processing, analyzing, and interpreting this vast array of information. For instance, data-driven approaches have been successfully used to predict FRCs' toxicological properties, masked forms and metabolites during manufacturing, and migration from the environment to food, thereby bridging data gaps in food safety assessment due to the high diversity and complexity of FRCs (Shi et al., 2022; Wang, Lin, Wang, Lin, & Tung, 2023; Zhang et al., 2021). Moreover, AI-driven approaches can facilitate real-time monitoring of food production processes, enabling early detection of contamination events and proactive intervention measures. By harnessing the potential of AI, stakeholders in the food industry can enhance their ability to mitigate risks, uphold regulatory compliance, and ultimately safeguard public health (Mu, et al., 2024). However, the rapid development of data-driven approaches has also raised higher requirements for the quality and comprehensiveness of FRC data, as erroneous or biased data can lead to unreliable models and research conclusions. Despite the wide discussion on AI and other data-driven approaches (Deng, Cao, & Horn, 2021; Tao, Yang, & Feng, 2020), no reviews have explored FRC-related data resources and their achievements as well as limitations that slow the application of AI in food safety systems.

Besides well-studied FRCs that have been included in FRC regulation and scientific databases, emerging FRCs have been detected in food and reported in publications with the rapid development of multi-omics approaches (Geueke et al., 2022; Medema et al., 2015). Those previously overlooked FRCs may also lead to negative health effects. For instance, masked mycotoxins, which are derivatives of native mycotoxins, can be generated under a series of chemical and biological transformations during food storage and processing (Ji, et al., 2023). Due to the change in molecular structures, they had usually been neglected in targeted detection, thus affecting the results of dietary exposure assessment. These masked FRCs usually have higher absorption and bioavailability than their native forms and might be transformed to more toxic forms under the action of gut microbes, leading to health implications (Ji, et al., 2023). Data-driven methods, such as natural language processing (NLP) (Xing, Zhang, Cai, Zhang, & Hu, 2023), cheminformatics (Ji, et al., 2023; D. Zhang, Tian, et al., 2021), and genome mining (Medema, et al., 2015; D. Zhang, Zhang, et al., 2021), have been used to identify emerging FRCs and fill data gaps in food safety databases, achieving better efficiency than conventional methods; however, no review has systematically summarized their methodology and successful applications in FRC detection and analysis, which hinders their wide application.

In this review, we present a thorough compilation of FRC-related data resources designed to assist researchers in accessing data, foster the digital transformation of the food industry, and support intelligent food safety supervision (Fig. 1). We outline the current limitations of existing FRC databases and suggest directions for future investigations. Additionally, we delve into the emergence of novel methodologies leveraging NLP, cheminformatics, and genome mining to pinpoint overlooked FRCs, aiming to offer innovative strategies for enhancing food safety assessment and decision-making.

2. Food-risk component databases and their applications

To comprehensively collect existing FRC-related data resources, we retrieved publications related to FRC data from Web of Science, PubMed, and Google Scholar. For those who did not generate scientific publications, we retrieved data from Google and Bing. Finally, we

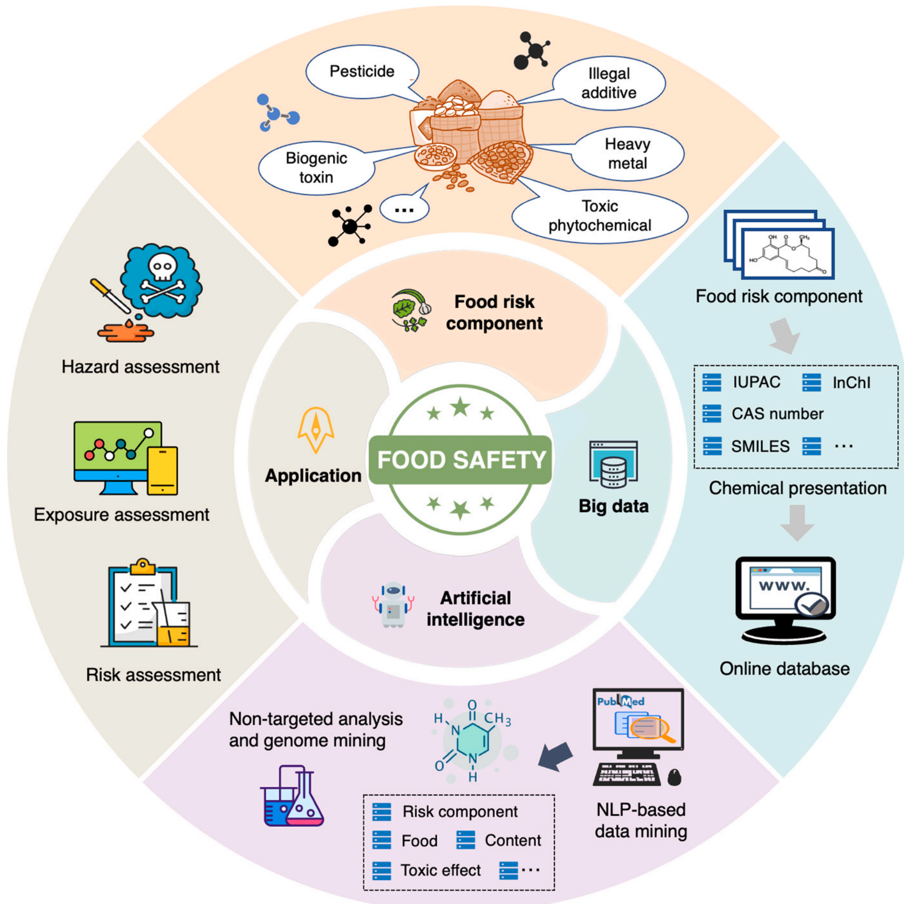


Fig. 1. Development and application of food-risk component databases. Increasing amounts of chemical substances with potential risks enter the food supply chain and affect human health, inspiring the development and application of food-risk component databases for their assessment and regulation. Chemical presentations have been used for food-risk component databases to promote standard data sharing, including the International Union of Pure and Applied Chemists (IUPAC) names, the International Chemical Identifier (InChI), the Chemical Abstracts Service (CAS) registry number, and the simplified molecular input line entry specification (SMILES). Emerging strategies, such as natural language processing (NLP), cheminformatics-enpowered suspect and non-targeted analysis, and genome mining, have facilitated the detection and analysis of emerging new FRCs outside of existing databases. These databases and new strategies have been extensively used for the hazard, exposure, and risk assessment of those components.

retrieved 50 FRC databases, of which 11 were on pesticides and veterinary drug residues; 5 were on food adulteration and illegal additives; 10 were on food additives; 4 were on food contact materials; 8 were on biogenic toxins; and 12 were integrated databases that included more than one type of FRC.

In this section, we discuss representative FRC databases with a focus ranging from anthropogenic chemicals to biogenic food toxins in the following order: (1) databases on environmental contaminants, pesticides, and veterinary drug residues, (2) food adulteration and illegal additives databases, (3) food additive databases, (4) food contact material databases, and (5) biogenic toxin databases. The databases discussed in this section are shown in Table 1 and Fig. 2.

2.1. Databases of environmental contaminants, pesticides, and veterinary drug residues

Multiple databases have been constructed, e.g., by the ESFA, US Environmental Protection Agency (EPA), US National Institutes of Health (NIH), and US Food and Drug Administration (FDA), to strengthen the compliant use of chemicals while controlling pollutants. Such databases are, for example, the Global Environment Monitoring System/Food Contamination Monitoring and Assessment Program (GEMS/Food), OpenFoodTox, ToxCast, Joint FAO/WHO Meeting on Pesticide Residues (JMPR) database, and Food Contaminant and Residue Information System (FCRIS).

2.1.1. GEMS/Food

The GEMS/Food project, established in 1976, is a collaboration between the WHO and over 30 global institutions (Y. Zhang, Yang, Fang, & Wei, 1997). Its goal is to inform governments, agencies, and the public about food contaminant levels and trends. The project has collected over seven million data records of 687 FRCs, including regions, pollutants, food categories, WHO food identifiers, codes, analyzed foods' status, test results, detection and quantitation limits, sampling cycle start and end times, sampling methods, food origin, and analytical quality assurance. GEMS/Food has been widely used for risk assessment of various FRCs (Vasconcelos Neto, Quintal, Porto, & Vitorino Carvalho de Souza, 2021). Using GEMS/Food data, Armaroli et al. proposed an approach to predict the distribution of chemical substances in foodstuffs to evaluate consumer risks (Armaroli, et al., 2020).

2.1.2. OpenFoodTox

Since 2002, the ESFA has conducted risk assessments for over 5712 substances based on over 2400 scientific opinions, statements, and conclusions, on which the OpenFoodTox database has been constructed (Dorne, et al., 2021). To date, the OpenFoodTox covers data on 2666 food ingredients, 1320 pesticides, 986 feed substances, 397 containments, and 343 food contact materials, and their chemical

Table 1
Representative food-risk component databases. Note: URL, Uniform Resource Locator; FRC, food risk component; ToxCast, Toxicity ForeCaster; DSSTox, Distributed Structure-Searchable Toxicity; Tox21, The Toxicology in the 21st Century; GEMS, Global Environment Monitoring System; CONT11, 11 food contaminants; MRL, Maximum Residue Limits; RASFF, The Rapid Alert System for Food and Feed; EMA, Economically Motivated Adulteration; FADB, Food Adulteration Database; FOCUS-DB, Food-Additive-Consumption-Safety Database; FCCdb, The Food Contact Chemicals database; FCCmigex, Migrating and Extractable Food Contact Chemicals; DBETH, Database for Bacterial ExoToxins; TPPT, Toxic Plants – Phytotoxin; ATDB, the Animal Toxin Database; KEGG, Kyoto Encyclopedia of Genes and Genomes; T3DB, The Toxin and Toxin Target Database; FRCD, Food Risk Component Database; JMPR, The Joint FAO/WHO Meeting on Pesticide Residues; MSDVDR, Mass Spectral Database of Veterinary Drug Residues; CAC, Codex Alimentarius Commission; GRAS, Generally Recognized as Safe; SCOGS, Select Committee on GRAS Substances; FEMA, The Flavor and Extract Manufacturers Association of the United States; JECFA, The Joint FAO/WHO Expert Committee on Food Additives; GSFA, Codex General Standard for Food Additives; USP, US Pharmacopeia; FDA, Food and Drug Administration; EURL-FCM, The European Union Reference Laboratory for Food Contact Materials; "/" no validated information available.

Name	Type	URL	Release date	Last update	Accessibility	Num. of FRCs
GEMS/Food	Integrated	https://extranet.who.int/gemsfood/	1976	2024	free download	687
RASFF	Integrated	https://webgate.ec.europa.eu/rasff-window/screen/search	1979	2024	free browse	/
DSSTox	Integrated	https://www.epa.gov/chemical-research/distributed-structure-searchable-toxicity-dsstox-database	2004	2024	free download	/
ToxCast	Integrated	https://www.epa.gov/chemical-research/exploring-toxcast-data	2007	2023	free download	~8500
Tox21	Integrated	https://comptox.epa.gov/dashboard/chemical-lists/tox21sl	2008	2023	free download	~13,000
T3DB	Integrated	http://www.t3db.ca/	2010	2015	free download	3678
FCRIS	Integrated	https://nucleus.iaea.org/sites/fcris/Pages/Home.aspx	2013	2023	free browse	/
OpenFoodTox	Integrated	https://www.efsa.europa.eu/en/microstrategy/openfoodtox	2017	2023	free download	5712
invitroDB	Integrated	https://epa.figshare.com/articles/dataset/ToxCast_Database_invitroDB_/6062623/1	2018	2023	free download	~8500
Tox21BodyMap	Integrated	https://sandbox.ntp.niehs.nih.gov/bodymap/	2020	2020	free download	9270
FRCD	Integrated	http://www.rxnfinder.org/frcd/	2020	2020	free browse	12,018
CONT11	Integrated	https://doi.org/10.1016/j.fct.2023.113843	2023	2023	Not publicly available	11
CAC Pesticide Residue Database	Pesticide and veterinary drug	https://www.fao.org/fao-who-codexalimentarius/codex-texts/dbs/pestres/en/	1963	2022	free browse	251
CAC Veterinary Drug Residue Database	Pesticide and veterinary drug	https://www.fao.org/fao-who-codexalimentarius/codex-texts/dbs/vetdrugs/en/	1963	2022	free browse	82
JMPR Database	Pesticide and veterinary drug	https://apps.who.int/pesticide-residues-jmpr-database/Home/Range/All	1963	2022	free download	363
Pesticide Info	Pesticide and veterinary drug	https://www.pesticideinfo.org/	2000	2020	free browse	~15,300
Phish-Pharm	Pesticide and veterinary drug	https://www.fda.gov/animal-veterinary/tools-resources/phish-pharm	2005	2022	free download	/
Japan Database of Maximum Pesticide Residue Limits in Food	Pesticide and veterinary drug	https://db.ffcr.or.jp/front/	2006	2023	free browse	/
Canada Pesticide Product Information Database	Pesticide and veterinary drug	https://pest-control.canada.ca/pesticide-registry/en/index.html	2006	2021	free browse	/
European Union Pesticide Database	Pesticide and veterinary drug	https://food.ec.europa.eu/plants-old/pesticides/eu-pesticides-database_en	2009	2022	free download	693
What's On My Food?	Pesticide and veterinary drug	https://www.whatsonmyfood.org/	2012	2014	free browse	296
The Pesticide Properties Database	Pesticide and veterinary drug	https://sitem.herts.ac.uk/aeru/ppdb/en/index.htm	2020	2023	free browse	2300
MSDVDR	Pesticide and veterinary drug	https://doi.org/10.1016/j.ijms.2018.11.014	2019	2019	Not publicly available	~400
USP Food Fraud Database	Food adulteration and illegal additive	https://www.foodchainid.com/food-fraud-database/	2012	2023	Not publicly available	15,500 incident records
Univ Minnesota EMA database	Food adulteration and illegal additive	https://doi.org/10.4315/0362-028X.JFP-12-399	2013	2013	Not publicly available	137 incident records
FADB-China	Food adulteration and illegal additive	http://www.rxnfinder.org/FADB-China/	2019	2020	free browse	564
China food fraud database	Food adulteration and illegal additive	https://doi.org/10.1007/s00003-021-01361-x	2022	2022	Not publicly available	454 incident records
Meat EMA database	Food adulteration and illegal additive	https://doi.org/10.1038/s41538-023-00189-z	2023	2023	Not publicly available	1214 incident records
FDA food additive list	Food additive	https://www.cfsanappsexternal.fda.gov/scripts/fdcc/?set=IndirectAdditives	1938	2024	free download	3647
JECFA database	Food additive	https://apps.who.int/food-additives-contaminants-jecfa-database/	1956	2023	free browse	/
SCOGS Database	Food additive	https://www.fda.gov/food/generally-recognized-safe-gras/gras-substances-scogs-database	1972	1980	free download	381
GSFA	Food additive	https://www.fao.org/gsfaonline/index.html	1995	2023	free browse	/
GRAS Notice Inventory	Food additive	https://www.fda.gov/food/generally-recognized-safe-gras/gras-notice-inventory	1998	2024	free browse	1116

(*continued on next page*)

Table 1 (*continued*)

Name	Type	URL	Release date	Last update	Accessibility	Num. of FRCs
EU food additive database and flavor substance database	Food additive	https://ec.europa.eu/food/food-feed-portal/screen/food-additives/search	2008	2023	free browse	410
FADB	Food additive	https://www.tandfonline.com/doi/abs/10.1080/19440049.2014.888784	2014	2014	Not publicly available	1227
FEMA GRAS List	Food additive	https://www.femaflavor.org/fema-gras	2015	2024	free browse	/
AdditiveChem	Food additive	http://www.rxnfinder.org/additivechem/	2018	2019	free browse	9064
FOCUS-DB	Food additive	http://14.139.62.46:8888/focus/	2020	2020	free browse	2885
FDA Food Contact Substance List Notification	Food contact material	https://www.cfsanappsexternal.fda.gov/scripts/fdcc/?set=FCN	2000	2024	free download	1662
FCCmigex	Food contact material	https://www.foodpackagingforum.org/fccmigex	2018	2022	free browse	~4000
FCCdb	Food contact material	https://www.foodpackagingforum.org/fccdb	2019	2020	free download	12,285
EURL-FCM - Substance Repository	Food contact material	https://joint-research-centre.ec.europa.eu/eurl-food-contact-materials/eurl-fcm-reference-substances-database_en	2021	2022	free browse	~400
ATDB	Biogenic toxin	http://protchem.hunnu.edu.cn/toxin	2007	2010	Not publicly available	2132
DBETH	Biogenic toxin	http://www.hpppi.iicb.res.in/btox/	2011	2012	free download	229
TPPT	Biogenic toxin	https://www.agroscope.admin.ch/agroscope/de/home/services/apps/tppt.html	2018	2023	free download	1586
KEGG Natural Toxin	Biogenic toxin	https://www.kegg.jp/brite/br08009	2018	2018	free browse	117
ToxinDB	Biogenic toxin	http://www.rxnfinder.org/toxindb/	2020	2020	free browse	4836
CyanoMetDB	Biogenic toxin	https://doi.org/10.1016/j.watres.2021.117017	2021	2021	free download	~2100
MycotoxinDB	Biogenic toxin	http://www.mycotoxin-db.com/	2023	2023	free download	189
MicotoXilico	Biogenic toxin	https://chemopredictionsuite.com/MicotoXilico	2023	2023	free browse	4360

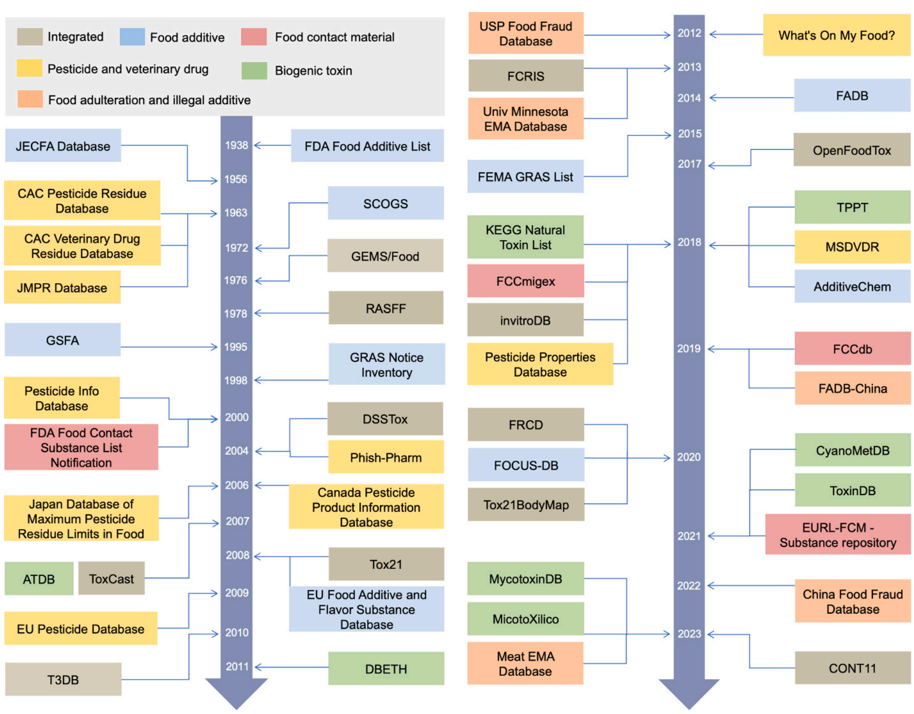


Fig. 2. Timeline of the establishment of representative FRC databases. The colors represent the types of databases. Note: ToxCast, Toxicity ForeCaster; DSSTox, Distributed Structure-Searchable Toxicity; Tox21, The Toxicology in the 21st Century; GEMS, Global Environment Monitoring System; CONT11, 11 food contaminants; MRL, Maximum Residue Limits; RASFF, The Rapid Alert System for Food and Feed; EMA, Economically Motivated Adulteration; FADB, Food Adulteration Database; FOCUS-DB, Food-Additive-Consumption-Safety Database; FCCdb, The Food Contact Chemicals database; FCCmigex, Migrating and Extractable Food Contact Chemicals; DBETH, Database for Bacterial ExoToxins; TPPT, Toxic Plants – Phytotoxin; ATDB, the Animal Toxin Database; KEGG, Kyoto Encyclopedia of Genes and Genomes; T3DB, The Toxin and Toxin Target Database; FRCD, Food Risk Component Database; JMPR, The Joint FAO/WHO Meeting on Pesticide Residues; MSDVDR, Mass Spectral Database of Veterinary Drug Residues; CAC, Codex Alimentarius Commission; GRAS, Generally Recognized as Safe; SCOGS, Select Committee on GRAS Substances; FEMA, The Flavor and Extract Manufacturers Association of the United States; JECFA, The Joint FAO/WHO Expert Committee on Food Additives; GSFA, Codex General Standard for Food Additives; USP, US Pharmacopeia; FDA, Food and Drug Administration; EURL-FCM, The European Union Reference Laboratory for Food Contact Materials.

characterization, regulations, toxicity, reference points (e.g., no observed adverse effect level, benchmark dose limits, LD50), reference values (e.g., acceptable daily intake and environmental standards like predicted no effect concentration), uncertainty factors, and EFSA scientific outputs. It is also part of the EFSA Data Warehouse to improve public access to data. Thanks to its good data quality and accessibility, OpenFoodTox has been widely used to develop computational toxicology models for chemicals as alternatives to animal testing. For example, based on data from OpenFoodTox, Toropov et al. developed several quantitative structure-activity relationship models to predict the toxicity of pesticides (Toropov, et al., 2017).

2.1.3. Tox21

The Toxicology in the 21st Century (Tox21) database (Berg, et al., 2011) is a collaborative project of the EPA, NIH, and FDA, with the goal of developing toxicity assessment models to test if given chemical substances can cause adverse health effects. To date, Tox21 has tested over 13,000 chemical substances, including commercial chemicals, pesticides, food additives/contaminants, and medical compounds, through quantitative high-throughput screening experiments and has yielded millions of concentration-response curves. Its main goals also include prioritizing substances for further in-depth toxicological evaluation, identifying mechanisms of action for further investigation (e.g., toxicity-associated and disease-associated pathways), and developing models that better predict how substances will affect biological responses as well as replacing animals used in toxicity testing and reduce time and costs (Tox21, 2024). Based on Tox21 data, many tools have been developed to assist in chemical/biological functional annotation of chemicals and identify their target organs in the human body. For instance, Tox21 Enricher was built based on Tox21 to identify over-represented chemical/biological annotations among lists of chemicals (Hur, et al., 2018), and a new database, Tox21BodyMap, was constructed to visually display chemical biological activity patterns by mapping experimental results to human organs, using organ-specific gene expression data (Borrel, Auerbach, Houck, & Kleinstreuer, 2020).

2.1.4. ToxCast, invitroDB, and DSSTox

The EPA established the Toxicity Forecaster (ToxCast) database to support the development of rapid and effective chemical safety assessment methods (Richard, et al., 2016). The ToxCast Pipeline utilizes open-source R packages, namely tcpl and tcplfit2, to store, manage, curve-fit, and visualize data and populate its relational database, invitroDB. The most recent ToxCast data is available in the invitroDBv4.1 database. It contains data on approximately 8500 chemicals, including industrial chemicals, consumer chemicals, and food additives. The EPA has made all ToxCast chemical data publicly available as part of its commitment to gather and share its chemical data in open and transparent ways. Based on exposure and toxicity data from ToxCast, João Barbosa et al. developed a Concern Index for prioritization of chemicals detected in the North Sea (Barbosa, De Schamphelaere, Janssen, & Asselman, 2021). The Distributed Structure-Searchable Toxicity (DSSTox) database is another EPA-led database aiming to improve toxicological prediction by providing high-quality public chemical resources (Grulke, Williams, Thillanadarajah, & Richard, 2019). The DSSTox database is known for its precise alignment of bioassay and physicochemical data with chemical structures. This capability allows researchers to access comprehensive information about chemical substances, facilitating various modeling and research endeavors, particularly for computational toxicology. DSSTox currently serves as the cornerstone of the EPA's CompTox Chemicals Dashboard, a pivotal platform offering public access to DSSTox content.

2.1.5. JMPR

The Joint FAO/WHO Meeting on Pesticide Residues (JMPR) has met annually since 1963 to conduct scientific evaluations of pesticide residues in food, and the JMPR database has served as a high-quality repository containing basic information on pesticides scrutinized by the JMPR (Evans, Scholze, & Kortenkamp, 2015). Its primary aim is to facilitate the harmonization of requirements and risk assessments concerning pesticide residues. Moreover, the database plays a crucial role in furnishing independent scientific expertise, offering invaluable advice to regulatory agencies tasked with overseeing pesticide usage and safety standards. To date, it has included data on 363 common pesticides and their chemical identifiers (e.g., CAS numbers), usage, acceptable daily intake (ADI), environmental health criteria, health and safety guides, carcinogenicity, and evaluation history.

2.1.6. Pesticide Info and What's On My Food?

The Pesticide Info database, created by the Pesticide Action Network (PAN), contains public data related to pesticides. It includes human toxicity, ecotoxicity, and regulations for approximately 15,300 pesticide active ingredients and their transformation products, including solvents and other additives used in pesticide products (PestcideInfo, 2024). A distinguishing feature of the PAN Pesticide Info database is that it integrated California pesticide use data directly from the Department of Pesticide Regulation and US Geological Survey data on annual agricultural use of more than 500 pesticides in the United States; it also has performed a number of data processing steps to present the data and summarize the information. Additionally, the database offers an interactive map enabling users to explore global pesticide bans. Users can search by chemical names to discover which countries have imposed restrictions on a specific pesticide, or by country names to ascertain the pesticides prohibited within each nation. The PAN also created an interactive database, named "What's On My Food?", which shows the potential pesticide residues in specific foods and their harm to humans.

2.1.7. FCRIS

The Food Contaminant and Residue Information System (FCRIS), pioneered by the International Atomic Energy Agency, is a unique resource dedicated to chemical food contaminants and their analytical methods (FCRIS, 2024). With a primary focus on Codex food standards/guidelines, its mission is to mitigate the occurrence of food trade rejections stemming from chemical contaminants and unacceptable levels of residues. FCRIS includes information on pesticides, veterinary drugs, toxic metals, and mycotoxins, incorporating both chemical composition and toxicological profiles. A distinguishing feature of the FCRIS is its specific focus on analytical methods for determination of chemical food contaminants. To date, approximately 200 analytical methods have been included in FCRIS along with their original publications.

2.1.8. Maximum residue level databases

Several databases provide maximum residue levels (MRLs) of environmental contaminants, pesticides, and veterinary drugs, such as the Codex Alimentarius Commission pesticide and veterinary drug residue databases, the European Union (EU) pesticide database, the Japan Food Chemistry Research Foundation Database of Maximum Pesticide Residue Limits in Food, and the Canada Pesticide Product Information Database. Cuarental et al. comprehensively summarized data sources containing maximum residue and contaminant limits for food contaminants, including relevant records from international sources, the EU, and several countries in the Americas (Cuarental, 2022).

2.1.9. Other databases

Many small-scale FRC databases have also been built. For instance, the Pesticide Properties Database (Lewis, Tzilivakis, Warner, & Green, 2016) integrates chemical identity, physicochemical properties, human health, and ecotoxicology data on approximately 2300 pesticides and their reported metabolites. The CONT11 database collects information on 11 types of contaminants, such as 5-hydroxymethyl-2-furfural, pyrraline, and furosine from 35 data sources, and covers 220 types of food (Hinojosa-Nogueira, et al., 2023). Lene et al. established a database on dietary exposure of chemical contaminants in fish for the Danish population (Duedahl-Olesen, et al., 2020).

Less effort has been devoted to developing databases on veterinary drug residues, unlike pollutants and pesticides. To prevent or cure disease, food-producing animals are commonly treated with veterinary drugs, which can leave residues in the food products originating from these animals. Besides various lists provided by regulatory agencies, we rarely found scientific databases specifically focusing on veterinary drug residues, except Phish-Pharm and Chen's mass spectral database tailored to veterinary drug residues. Phish-Pharm (Crosby, Kittel, & Gieseker, 2022) includes over 700 pharmacokinetics and drug residue literature on fish. Chen et al. developed a mass spectral database tailored to veterinary drug residues (D. Chen et al., 2019), encompassing over 400 veterinary drugs, providing both molecular and mass details. Users can access mass spectra based on ionization mode, precursor, and product ion peaks and conduct compound searches using advanced and fuzzy search functions (D. Chen et al., 2019).

2.2. Food adulteration and illegal additive databases

The food industry has gradually established digital fraud prevention measures to cope with the increasingly complex global food supply chain. Researchers are aiming to determine the authenticity of food by analyzing structured and unstructured data from various sources to protect consumers' health while minimizing economic losses caused by food adulteration. To support digital supply chain management and food adulteration risk alert, several databases have been established, including the EU Rapid Alert System for Food and Feed (RASFF), the US Pharmacopeial Convention (USP) Food Fraud Database, the Economically Motivated Adulteration (EMA) database, and food additive database (FADB)-China.

2.2.1. RASFF

RASFF is a representative digital platform for detecting food adulteration risks (Krisztina, Zsolt, & Peter, 2005). The system was established in 1979 to effectively share information among its members by providing 24-h service to ensure efficient sending, receiving, and response of emergency notifications (Krisztina, et al., 2005). Through the RASFF's important information exchange mechanism, members can avoid food safety risks by recalling problematic products from the market before they cause harm to consumers. Based on RASFF, many studies have been conducted to identify potentially emerging food safety issues. For instance, using RASSF, D.'Amico et al. analyzed the possible relationship between notified products based on the notifications related to seafood from 2011 to 2015 (D.'Amico, et al., 2018). Lüth et al. analyzed RASFF notifications on food products contaminated with *Listeria monocytogenes* and proposed several options for improving the RASFF (Lüth, Boone, Kleta, & Al Dahouk, 2019).

2.2.2. USP food fraud database

To increase consumer trust, provide brand protection, and support food safety regulations promulgated by the FDA, in 2012, the US USP established a food fraud database to help food manufacturers and retailers identify and analyze ingredients in their products that may be adulterated (Moore, Spink, & Lipp, 2012). The USP food fraud database includes thousands of food ingredients and their associated adulterants, as well as related scientific publications, regulations, judicial records, incident reports, surveillance records, and analytical methods. The database allows users to perform automatic alerts on food fraud records and automatic analysis of specific ingredient-related data through a custom dashboard. It also can generate hazard reports for specific substances that have been adulterated, allowing manufacturers and retailers to quickly identify food products that contain detected substances. By analyzing the database, Moore et al. found that olive oil, milk, honey, and saffron were mentioned most frequently in academic papers; high-performance liquid chromatography and infrared spectroscopy were the analytical methods most used for food-adulteration detection; and chemometric data analysis has been used in several studies (Moore, et al., 2012). They also identified a range of potential food fraud issues, including the addition of spices with lead chromate and lead tetroxide, the substitution of poisonous Japanese star anise for Chinese star anise, and the adulteration of melamine in food (Moore, et al., 2012).

2.2.3. The University of Minnesota EMA database

The National Center for Food Protection and Defense at the University of Minnesota created an EMA database to help governments, agencies, and businesses assess food authenticity risks (Everstine, Spink, & Kennedy, 2013). It contains 37 unique incidents in 11 food categories: fish and seafood, dairy products, fruit juices, oils and fats, grain products, honey and other natural sweeteners, spices and extracts, wine and other alcoholic beverages, infant formula, plant-based proteins, and other food products collected from both scientific literature and media sources (Everstine, et al., 2013). Via analyzing the database, Everstine et al. identified common characteristics among the incidents that may help to better evaluate and reduce the risk of EMA (Everstine, et al., 2013).

2.2.4. FADB-China

FADB-China is a molecular-level food-adulteration database (D. Zhang, Ouyang, et al., 2020). It includes 961 food adulteration cases that occurred in China between 1998 and 2019. A unique feature of FADB-China is that critical molecules involved in food adulteration have been manually annotated by food chemists, providing detailed information to investigate the potential toxic effects and health risks. Based on FADB-China, Zhang et al. developed a prediction algorithm for screening potentially illegal food additives based on molecular fingerprints and similarity (D. Zhang, Ouyang, et al., 2020). This algorithm was successfully used to predict 1919 chemical substances with the potential to be illegally added to food.

2.2.5. Other databases

Besides the databases mentioned above, several databases with a more specific focus have also been developed to assist the fight against food adulteration and fraud (European Commission, 2024). For instance, Li et al. created a database on meat-food fraud risks based on official reports and media coverage in China (X. Li et al., 2023). This database includes information on various types of meat, including livestock, poultry, by-products, and processed meat products. It records details of food fraud incidents, such as the types of adulterants involved, as well as the categories and subcategories of food, risk links, and locations (X. Li et al., 2023). Li et al. collected data from online news and literature to create a food fraud database comprising 454 news sources in China from 2001 to 2019 (H. Li, Cheng, Luo, Li, & Wu, 2022). They analyzed the sources of food fraud reports, their temporal and spatial distribution, the types of fraud involved, and the classification of the implicated food items. These databases play an important role in improving the authenticity of food products by providing valuable information, analytical methods, and reference data to stakeholders in the industry and regulatory authorities.

2.3. Food additive databases

Based on the definition and regulation by the Joint FAO/WHO Expert Committee on Food Additives (JECFA), national agencies usually independently formulates the definition, scope, and usage standards of food additives based on local dietary patterns and dietary exposure assessments (C. Wang, Wu, & Gao, 2014; Zhang et al., 2020). Therefore, there are differences in the definitions of food additives in various countries. For example, the scope of food additives defined in China and the United States includes nutritional supplements, but this category is not included in the EU. Indirect food additives defined by the United States, such as packaging materials, do not fall into the category of food additives in China and the EU. The different definitions of food additives in different countries have resulted in diverse food additive databases.

2.3.1. JECFA and GSFA

The JECFA has been evaluating the safety of food additives, pollutants, natural toxins, and pesticide and veterinary drug residues since 1956. The JECFA database contains basic information of over 2000 food additives and flavorings, their ADI, evaluation history, related reports, and monographs. Another initiative proposed by the FAO/WHO is the Codex Alimentarius Commission online database of the Codex General Standard for Food Additives (GSFA). The GSFA database contains the usage specifications, including the food (or food category), for each food additive, the maximum allowable usage amount, and related processes, and supports users to browse by food additive names, synonyms, functional categories, and food categories.

2.3.2. China food additive databases

China national standards stipulate that food additives refer to synthetic or natural substances added to food to improve the quality, color, aroma and taste, as well as for the needs of antisepsis, preservation and processing (C. Wang et al., 2014). China mainly uses "Food Additives Usage Standards" (GB2760-2024) and "Nutrient Supplements Usage Standards" (GB14880-2012) to regulate food additives. To date, approximately 2300 types of food additives have been approved in China.

2.3.3. EU food additive databases

Regulation (EC) No 1333/2008 of the European Parliament and of the Council of 16 December 2008 on food additives stipulates that food additives are substances that are not normally consumed as food itself but are added to food intentionally for a technological purpose described in this regulation, such as the preservation of food (EUR-Lex, 2008). The EU divides food additives into 26 categories, ~2600 types, which are mainly included in the EU food additive database and the flavor substance database.

2.3.4. US food additive databases

The United States defines the scope of food additives more broadly. Part 170 of the US Code of Federal Regulations Title 21 (21 CFR) stipulates that food additives include all substances not exempted by section 201(s) of the act, the intended use of which results or may reasonably be expected to result, directly or indirectly, either in their becoming a component of food or otherwise affecting the characteristics of food (CFR, 2024). There are ~4000 types of food additives approved for use in the United States, such as natural flavors, synthetic flavors, and enzyme preparations, which are primarily listed in the FDA Food Additive List, the Indirect Food Contact Additive Database, the Generally Recognized as Safe (GRAS) Announcement Database, the Food Contact Substance List Notification, the Select Committee on GRAS Substance Database, and the Flavor & Extract Manufacturers Association GRAS List.

2.3.5. Other databases

In recent years, food additive databases with detailed molecular information (such as molecular structure, biological activity, etc.) have been established. To provide a database for *in silico* screening, Ginex et al. constructed a FADB (Ginex, Spyrakis, & Cozzini, 2014) that provides the three-dimensional structure and physical and chemical properties of 2540 food additives. Zhang et al. developed AdditiveChem (D. Zhang, Cheng, et al., 2020), a database that contains over 9064 types of food additives worldwide, and their molecular structure, toxicity properties, ADI, and usage specifications. Chauhan et al. developed FOCUS-DB (Chauhan, Sachan, & Parthasarathi, 2021), which includes 2885 food additives and their approval status in various regulatory authorities, physicochemical properties, and biological pathways. These databases have become useful tools for understanding food additives' functionality, toxicity, and effects on human health as well as designing new healthier food additives.

2.4. Food contact material databases

Food contact materials (FCMs) are substances that come into contact with food during production, processing, and storage, mainly including heavy metals, plasticizers, and additives (Eti, et al., 2023; Gerassimidou et al., 2023). The lists of FCMs are explicitly provided by EU and US food safety agencies, providing basic information on FCMs. The databases hosted by the ESFA and the FDA provide information on 392 and 3647 substances in FCMs, respectively. In addition, databases, namely the food contact chemicals database (FCCdb) and the database on migrating and extractable food contact chemicals (FCCmigex), provide additional high-quality data on FCMs.

2.4.1. FCCdb

FCMs might introduce chemicals into food, known as food contact chemicals (FCC), causing worrying health risks. However, information on the chemical structures, use patterns, and migration potential of FCCs is often absent. To address data deficiency, FCCdb (Groh, Geueke, Martin, Maffini, & Muncke, 2021) was created by consolidating information from 67 FCC lists obtained from global regulatory bodies and industry inventories. It encompasses 12,285 unique FCCs, with the primary objective of evaluating their diversity and potential hazards. Groh et al. reviewed all substances in FCCdb and delved deeper into the hazards associated with FCCdb chemicals by consulting various authoritative sources (Groh, et al., 2021). They prioritized 608 hazardous FCCs for further assessment and potential substitution in food contact materials. Furthermore, they conducted an evaluation based on non-authoritative predictive hazard data, such as from *in silico* modeling and literature analysis, and identified an additional 1411 FCCdb substances that could potentially pose concerns.

2.4.2. FCCmigex

The FCCmigex database (Geueke, et al., 2022) evaluates the chemicals present in FCMs, based on 1210 studies, increasing the known range of FCCs to 14,153 compounds. Based on FCCmigex, it was discovered that 2881 FCCs showed analytical evidence of the presence in food or FCM extractions, most of them being poorly studied and observed. Based on the database, computational methods have been established and applied to evaluate the chemical migration potential of FCMs from packaging materials to food (S. S. Wang, Lin, Wang, Lin, & Tung, 2023).

2.5. Biogenic toxin databases

Biogenic FRCs mainly include mycotoxins, bacterial toxins, phytotoxins, and animal toxins. Owing to their threat to human health, several databases have been constructed, including MycotoxinDB, MicotoXilico, Database for Bacterial Exotoxins (DBETH), Toxic Plants-PhytoToxins (TPPT), Animal Toxin Database (ATDB), and CyanoMetDB.

2.5.1. MycotoxinDB and MicotoXilico

Mycotoxins can contaminate crops, vegetables, and fruits. Vomiting, nausea, diarrhea, abdominal pain, and hematopoietic tissue damage are common symptoms of mycotoxin poisoning (Yao & Long, 2020). To assist in mycotoxin research, especially the identification of masked forms of mycotoxins, Ji et al. developed MycotoxinDB (Ji, et al., 2023). MycotoxinDB includes 189 parent forms of mycotoxins, and their molecular structures, physicochemical properties, and known and predicted masked forms. Tolosa et al. classified 4360 mycotoxins into 170 categories and developed MicotoXilico (Tolosa, et al., 2023), a comprehensive mycotoxin database. Based on the data in MicotoXilico, specific quantitative structure-activity relationship models have been proposed for estimating the mutagenicity, genotoxicity, and carcinogenicity of mycotoxins. Thanks to the tailored data on algorithms, those models showed good accuracy (Tolosa, et al., 2023).

2.5.2. DBETH

To address the need for systematic compilation and analysis of known bacterial toxins, DBETH was developed (Chakraborty, Ghosh, Chowdhary, Maulik, & Chakrabarti, 2012). The database includes 229 toxins from 26 bacterial genera categorized into 24 mechanistic and activity types, providing information on their sequence, structure, and interaction network. Furthermore, it integrates a prediction server to identify bacterial toxin-like sequences based on homology and machine learning techniques. DBETH, together with its related prediction server, offers researchers useful resources to gain insight into the characteristics and mechanisms of bacterial toxins.

2.5.3. TPPT

Phytotoxins, such as pyrrolizidine and aristolochic acid, are widely found in food and herbal medicines that may cause liver and kidney damage when ingested by humans (Mol, Van Dam, Zomer, & Mulder, 2011). However, these phytotoxins are often overlooked due to the lack of systematic data resources. Gunthardt et al. developed TPPT (Gunthardt, Hollender, Hungerbuhler, Scheringer, & Bucheli, 2018) that contains information on 1586 phytotoxins from 844 plant species in Switzerland and Central Europe. Using the TPPT database, researchers evaluated phytotoxins in terms of their occurrence and environmental behaviors, and revealed that more than 34% of phytotoxins are potential aquatic micropollutants and have high risk of contaminating aquatic products (Gunthardt, et al., 2018).

2.5.4. ATDB

ATDB is a comprehensive database with >3235 animal toxins and >54,000 toxin-channel (T-C) interactions (He, et al., 2010). To facilitate data search, exchange, and comparison, ATDB uses a new ontology to standardize toxin annotations and has manually assigned over 8423 toxin ontology terms to 2132 toxins by trained biologists.

2.5.5. CyanoMetDB

CyanoMetDB is a carefully organized database of cyanobacterial secondary metabolites gathered from 850 peer-reviewed articles published between 1967 and 2020 (Jones, et al., 2021). It contains 2010 cyanobacterial metabolites and 99 structurally related compounds (Jones, et al., 2021). This database has been used for identifying and confirming known cyanobacterial toxins and secondary metabolites, discovering new natural products derived from cyanobacteria, supporting research on cyanobacterial secondary metabolite biosynthesis, and facilitating investigations into the prevalence, longevity, and harmful effects of these substances in natural settings (Jones, et al., 2021).

2.5.6. Other databases

Several integrated databases covering more than one category of biogenic FRCs have also been developed. The Kyoto Encyclopedia of Genes and Genomes natural toxin dataset contains 354 toxins, covering mycotoxins, phytotoxins, cyanotoxins, marine toxins, and animal toxins (Kanehisa, Furumichi, Tanabe, Sato, & Morishima, 2017). The Toxin and Toxin Target Database (T3DB) contains >900 food-related toxins and their chemical properties, targets of action, toxicity mechanisms, and references (Lim, et al., 2010; Wishart et al., 2015). ToxinDB, a biogenic toxin database, contains >4836 biogenic toxins extracted from publications, along with 21,848 experimentally measured toxicity properties (D. Zhang, Tian, et al., 2021).

3. Chemical identifiers and representations used by food-risk component databases

Images of FRC structure are commonly presented on online FRC databases. However, the excessive amount of data makes such images unsuitable for sharing, transferring, or retrieving information. Various chemical identifiers/representations have been put forward to expedite the sharing of information. Here, we present four types of chemical presentations widely used by FRC databases to assist researchers in organizing and using data as well as to evaluate the data standardization and quality of FRC databases. These presentations include names from the International Union of Pure and Applied Chemists (IUPAC), the IUPAC International Chemical Identifier (InChI), the Chemical Abstracts Service (CAS) registry number, and the simplified molecular input line entry specification (SMILES). Table 2 presents a list of common FRCs and their identifiers as well as the basic usage of SMILES.

3.1. IUPAC names

In 1958, the IUPAC presented a hierarchical nomenclature for organic compounds relying on names of alkanes that have all carbon atoms in a straight chain. This is precisely associated with the names and structures of organic compounds; however, considering that some names are too intricate for recall and reuse, simple names (such as dicloran, instead of 2,6-dichloro-4-nitroaniline) are often favored over IUPAC names in FRC databases.

3.2. SMILES

SMILES is a line notation that describes the structure of chemical species using short ASCII strings (Weininger, 1990). It is widely used because of its simplicity. In SMILES, the symbols -, = , #, and: represent single, double, triple, and aromatic bonds, respectively. Adjacent atoms are connected by single or aromatic bonds. It is a more compact representation than most other structural representations, making it suitable for storing large amounts of data. Structural data of FRCs can also be saved in structure-data files, such as SDF and mol2, which include information on connectivity, coordinates in 2D or 3D space, bond orders, and other relevant details.

3.3. InChI

InChI is a non-proprietary identifier for chemical substances that can be used in electronic or print data sources, facilitating the linking of diverse data collections. Identifiers describe chemical substances with respect to a variety of information layers, such as atoms, bond connections, tautomeric information, isotope information, stereochemistry, and electronic charge information. The straightforward representation frequently serves as a search operator in databases.

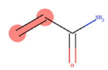
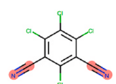
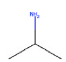
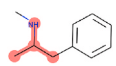
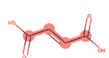
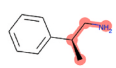
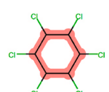
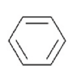
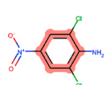
3.4. CAS registry number

The CAS assigns its registry numbers as identifiers for chemical substances mentioned in scientific publications. These numbers typically comprise serial numbers with a check digit that lacks structural information. For example, the CAS number for acrylamide is 79–06-1: three numbers separated by two dashes. It is critical to note that CAS numbers do not correlate one-to-one with chemical compounds. Although each CAS number corresponds to a unique chemical substance, in practical settings, a single compound may have multiple CAS numbers. Such cases arise due to different isomeric forms, diverse hydrates or salts, or even for historical reasons. Therefore, it is crucial to recognize that a compound may be linked with several CAS numbers when using them for compound identification or referencing in the scientific literature.

3.5. Chemical identifiers used in food-risk component databases

By summarizing the frequency of chemical identifiers used in FRC databases, we found that, besides common FRC names, CAS numbers are the most commonly used representations in FRC databases (used by 46% of FRC databases), followed by SMILES (36%), InChI (22%), and IUPAC names (20%) (see supplementary materials). Approximately 35% of databases provide detailed information on FRC molecular structures, such as SMILES or InChI. Among the different types of FRC databases, databases on biogenic toxins most comprehensively utilize chemical identifiers - nearly all provide information on molecular structures.

Table 2
Chemical representations of food risk components. Note: SMILES, simplified molecular-input line-entry system; FRC, food risk component; InChI, International Chemical Identifier; IUPAC, International Union of Pure and Applied Chemistry; CAS, Chemical Abstracts Service.

Type	SMILES	Structure	FRC example	FRC name	FRC SMILES	InChI	IUPAC name	CAS registry number
Double bond	C=C			Acrylamide	C=CC(=O)N	InChI = 1 S/C3H5NO/c1-2-3 (4)5/h2H,1H2, (H2,4,5)	prop-2-enamide	79-06-1
Triple bond	C#N			Chlorothalonil	C (#N)C1=C(C(=C(C(=C1Cl)Cl) Cl)C#N)Cl	InChI = 1 S/C8Cl4N2/c9-5-3 (1–13)6 (10)8 (12)7 (11)4 (5) 2-14	2,4,5,6-tetrachlorobenzene-1,3-dicarbonitrile	1897-45-6
Branch	CC(C)N			Methamphetamine	CC(CC1=CC=CC=C1)NC	InChI = 1 S/C10H15 N/c1-9 (11–2)8-10-6-4-3-5-7-10/h3-7,9,11H,8H2,1–2H3/t9-/m0/s1	(2 S)–*N*-methyl-1-phenylpropan-2-amine	537-46-2
E-Z configuration	C/C=C\C			Fumaric acid	C (=C/C (=O)O)\C (=O)O	InChI = 1 S/C4H4O4/c5-3 (6)1-2-4 (7)8/h1-2H, (H,5,6) (H,7,8)/b2-1+	(E)-but-2-enedioic acid	110-17-8
R–S configuration	C [C@@H] (C)C			B-methylphenethylamine	C [C@H](CN)C1=CC=CC=C1	InChI = 1 S/C9H13 N/c1-8 (7–10)9-5-3-2-4-6-9/h2-6,8H,7,10H2,1H3/t8-/m1/s1	(2 S)-2-phenylpropan-1-amine	17596-79-1
Common ring	C1CCCCC1			Lindane	C1(C(C(C(C(C1Cl)Cl)Cl)Cl)Cl)Cl	InChI = 1 S/C6H6Cl6/c7-1-2 (8)4 (10)6 (12)5 (11)3 (1)9/h1-6H	1,2,3,4,5,6-hexachlorocyclohexane	319-84-6
Aromatic ring	c1ccccc1			Dicloran	C1=C(C=C(C(=C1Cl)N)Cl)[N+] (=O)[O-]	InChI = 1 S/C6H4Cl2N2O2/c7-4-1-3 (10 (11)12)2–5 (8)6 (4)9/h1-2H,9H2	2,6-dichloro-4-nitroaniline	99-30-9

Databases on environmental pollutants, pesticides, and veterinary drug residues provide less information on chemical identifiers, potentially because many of those databases mainly focus on regulations and standards. Among all food adulteration databases, FADB-China is the only database that includes information on critical chemicals/illegal additives involved in food adulteration incidents, while others mainly focus on regulations, judicial records, incident reports, surveillance records, or analytical methods. Compared with regulation and standard databases, scientific domain-specific databases better manage FRC chemical identifiers. More than 80% of scientific domain-specific databases use at least one type of chemical identifier, while the number for regulation and standard databases is approximately 50%.

4. Limitations of existing FRC databases and recommendations for improvement

In the past decades, over 50 FRC databases have been established (Fig. 2), and many of them have been updated more than once. With time, newly proposed databases or new versions of previously established databases have demonstrated many advantages, such as more comprehensive data coverage, more centralized access, advanced search and query capabilities, and more comprehensive integration of tools and predictive models. For instance, new proposed databases compile data from more sources, including scientific literature, regulatory agencies, industry reports, and research studies, allowing for a more thorough understanding of FRCs. In addition, unlike previous fragmented approaches in which relevant information may be scattered across different sources or not easily accessible, recently established/updated FRC databases, such as OpenFoodTox 2.0 (Dorne, et al., 2021) and AdditiveChem (D. Zhang, Cheng, et al., 2020), provide a centralized platform for accessing and retrieving data. These databases often come with advanced search and query capabilities, allowing users to filter, sort, and analyze data based on various criteria. This enables researchers to extract meaningful insights and identify patterns or trends more efficiently. Furthermore, some databases also include built-in tools and models for data analysis, visualization, and predictive modeling, facilitating data interpretation and decision-making by providing users with actionable insights and recommendations. Despite many achievements, we also observed some persistent challenges, such as delayed updates, accessibility concerns, reproducibility issues, suboptimal data quality, and inadequate coverage in underdeveloped regions.

4.1. Accessibility

An overarching challenge with existing FRC databases lies in their accessibility. Notably, approximately 15% of the databases examined in this review are not accessible to the public, rendering them unavailable for reproduction or utilization in research. Furthermore, more than 60% of the databases do not facilitate batch downloads. While computational approaches, such as *in silico* screening, hold promise for enhancing food safety assessment and regulation, the limitations in existing data-sharing strategies may hinder their wide application.

4.2. Data quality and traceability

The quality of FRC data across various resources can vary significantly due to differences in data collection, cleaning, and processing methods. Notably, approximately 20% of databases solely comprise first-hand data with their own quality control standard, while the remaining 80% incorporate information sourced from publications and other data repositories. However, we found insufficient disclosure of original data sources in most databases, which causes low confidence in data quality and inconsistency between databases.

4.3. Timeliness and updates

The research findings highlight a notable disparity in the timeliness of database updates between those hosted by international organizations and those managed by research groups. Specifically, nearly all databases hosted by international organizations feature records of updates, whereas only 40% of those maintained by research groups do. However, the absence of current scientific findings can undermine the strength of conclusions drawn from these databases.

4.4. Inconsistency

Overall, approximately 30% of FRC databases, particularly those concerning regulations and standards, do not provide any information on chemical identifiers, while approximately 65% do not offer detailed molecular information, such as SMILES and InChI. This deficiency poses a significant challenge as it may lead to inconsistencies across databases and studies (Grulke, et al., 2019; Neltner et al., 2013; D. Zhang, Cheng, et al., 2020). Urgent action is needed to transition towards more standardized and transparent data annotation and management practices.

4.5. Data coverage bias

Data coverage bias, especially the lack of food safety data for underdeveloped regions, is another significant concern. Existing databases have mainly been established by national agencies and researchers in developed countries and a few large developing countries such as China and India (see supplementary materials), and tend to focus on food safety status within their respective borders. Databases established by international organizations, such as GEMS/Food, provide platforms for sharing food safety data globally. However, even these efforts are not fully comprehensive and suffer from biases. For instance, in GEMS/Food, the ratio of records from the Americas and European regions to those from the rest of the world is approximately 8 to 1, underscoring the deficiency in food safety data for developing countries and regions.

4.6. Recommendations for improvement

Despite great achievements in the past decades, existing FRC databases still face shortcomings in terms of accessibility, data quality, timely updates, and inconsistency. As short term improvements, we suggest that databases appropriately include chemical identifiers of FRCs, e.g., CAS registry number, InChI, or more accurately, SMILES or structure data files whenever possible to help researchers and other stakeholders recognize the chemical nature of FRCs. This may also avoid the severe inconsistency across different databases that has been found previously (Grulke, Williams, Thillanadarajah, & Richard, 2019; Neltner, Alger, Leonard, & Maffini, 2013; Zhang et al., 2020).

Furthermore, we encourage FRC databases to commit to better data availability and provide user-centered interfaces for data acquisition. This will promote their wide application while considerably promoting the future development of big data and AI-based approaches in food science. Moreover, we encourage FRC databases to enhance data quality by appropriately stating original data resources, for example, by presenting whether the data have been directly collected from in-house experiments, extracted from publications, or integrated from other databases, the data cleaning procedure, and the standard for quality checks.

In the long term, enhancing standardized data-sharing strategies should be continuously encouraged. The findability, accessibility, interoperability, and reuse (FAIR) guideline (Wilkinson, et al., 2019) has provided a reusable standard for FRC databases to manage and share data, which should be precisely implemented. Moreover, establishing FRC databases specifically focusing on cases in developing countries and regions should be continuously encouraged, and FRC databases hosted by international organizations are also expected to focus more on food safety status in such regions. Finally, new approaches for data annotation and management, including emerging NLP-based automatic data extraction and large language models, should be investigated and customarily used in the food safety field. Thus far, a few databases have benefited from these approaches. For example, Zhang et al. developed a Food Risk Component Database (FRCD), which contains 12,018 FRCs from 152,737 publications (D. Zhang, Gong, et al., 2020). However, the application of NLP is only restricted to annotating FRC-named entities, failing to process information on exposure level, the content in respective food categories, and detailed information related to molecular structure and properties. With the rapid development of AI, we foresee that such approaches will considerably enhance the efficiency of identification and organization of FRC data from the large body of literature and media reports, and alleviate issues of late updates and weak confidence in data quality.

As a first step to overcome current limitations and improve the use of digital assets in food safety according to the FAIR guideline, we propose five general initiatives for the future development and management of FRC online databases.

- Transparent data sourcing: Databases should state data resources whenever possible. For data from in-house experiments, detailed experiment procedures should be introduced to enhance reproducibility. For data sourced from publications and other databases, the original source should be clearly mentioned, and cross-references and URL links should be provided for verification.
- Free accessibility: Databases should be freely available with no restrictions, including the need to register or login. A copy preserved in permanent data repositories, e.g., Zenodo, is highly encouraged.

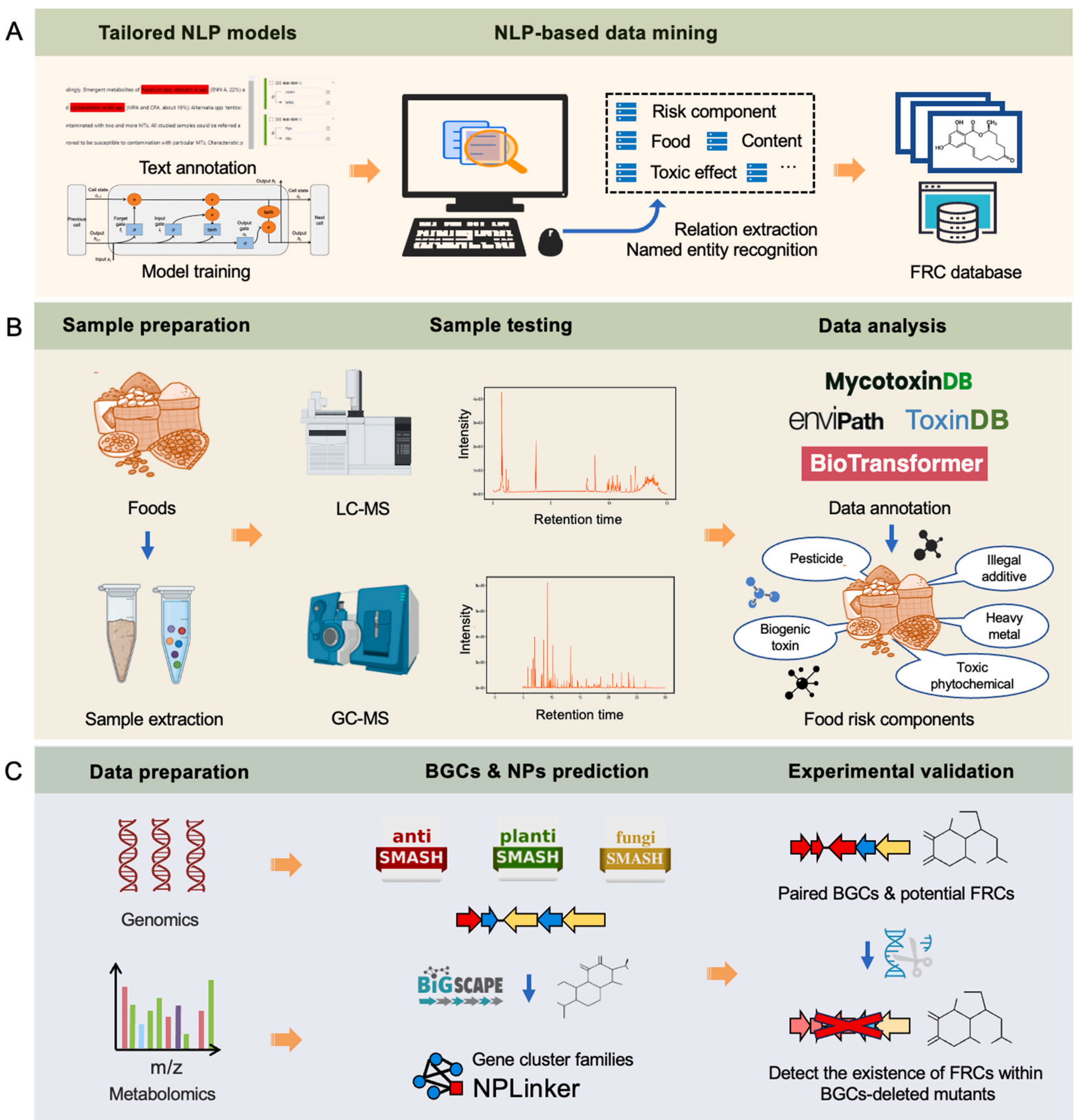


Fig. 3. Data-driven strategies to identify less studied and unknown food-risk components. (A) Mining of food-risk component data from literature and public opinion using natural language processing (NLP). (B). Suspect screening and non-targeted analysis of FRC with cheminformatics tools. (C). Genome mining for identifying biogenic food-risk components. After prediction, the relationship between toxic natural products (NPs) and biosynthetic gene clusters (BGCs) can be confirmed in gene knockout experiments by suppressing the expression of target BGCs, followed by observation of changes in the corresponding NPs.

- Long-term maintenance and stable URLs: Maintain databases at the same URL for at least 3 years, especially if the database results in a scientific publication.
- Security and privacy measures: Databases are encouraged to use the HTTPS protocol, especially if the website accepts or transmits sensitive data.
- Data availability and formats: Databases should address the availability of the underlying data, including the formats and terms for data download.

5. Data-driven identification of less studied or unknown food-risk components

Despite the inclusion of well-studied FRCs in databases, new FRCs found in food, such as masked forms of biogenic toxins and emerging new anthropogenic pollutants, have largely been neglected, posing new challenges to food safety (Barabási, Menichetti, & Loscalzo, 2020; Fu et al., 2018). In recent years, data-driven strategies have been used to identify emerging FRCs and demonstrate their promising potential. This discussion focuses on three powerful tools for identifying less studied or unknown FRCs beyond the coverage of existing FRC databases, namely NLP, cheminformatics-empowered suspect screening and non-targeted analysis, and genome mining.

5.1. Mining food-risk components in literature and public opinion based on natural language processing

Literature and public opinion are valuable sources that contain information on FRCs and their health risks; however, they cannot be fully utilized, given the low efficiency of manual information processing. Recently, NLP-based approaches have been increasingly used to identify FRCs and consumer's perceptions toward them from the literature and media, providing new insights into FRC regulation and control (Fig. 3A).

Massive FRCs have been investigated and mentioned in publications, yet only a small fraction has been included in public databases (Hooton, Menichetti, & Barabasi, 2020). Combining NLP-based name entity recognition (NER) and manual data clearing, comprehensive FRC databases such as FRCD have been proposed, demonstrating higher efficiency (D. Zhang, Gong, et al., 2020). Likewise, Hooton et al. developed FoodMine to identify the garlic and cocoa components from PubMed publications and discovered that many food components have been reported in scientific publications but have not been integrated into any databases (Hooton, et al., 2020). Despite the improvement in efficiency and coverage, the development of FRCD and FoodMine still relies on manual data cleaning because the non-standard name of FRCs used in the literature may harm the accuracy of NER. To address the limited availability of annotated corpora related to food, FoodBase, a corpus of food entities, was developed, which serves as a benchmark for food-related NLP tasks (Popovski, Seljak, & Eftimov, 2019). Despite its great potential, the application of NLP-based approaches in food science is still in the early stages. In the future, models should be explored to extract the relationship between FRCs and their related food, contamination content, and toxic effects to provide more systematic data.

Besides scientific literature, public opinion on the Internet also contains a wealth of information on food safety, such as views on foodborne diseases. It can serve as an invaluable resource in the mining of FRCs. By understanding the concerns, preferences, and insights of the public, researchers and policymakers can prioritize certain FRCs for investigation, evaluate the perceived risks associated with various chemicals, and tailor their communication strategies to address prevalent misconceptions. Engaging with the public ensures that the investigation of FRCs is aligned not just with scientific objectives, but also with society's expectations. Given this, Zhang et al. developed IFoodCloud to collect data from >3100 public sources and constructed sentiment prediction models to analyze public sentiment toward specific food safety issues (H. Zhang, Jia, et al., 2023). Bouzembrak et al. established a food-adulteration monitoring system, named MedISys-FF, based on social media information to collect, process, and present food fraud reports published in global media (Bouzembrak, et al., 2018). Chen et al. established a food safety information database in Greater China based on information from approximately 100 websites regarding food safety (S. Chen, Huang, Nong, & Kwan, 2016). Tiozzo et al. analyzed data related to food risks from online sources using web monitoring, content analysis, and visualization (Tiozzo, et al., 2020).

Using NLP-based approaches, FRCs were comprehensively identified from the literature and public opinion, providing a comprehensive overview of potential risks associated with the food supply chain. This information will be valuable to policymakers, regulators, and stakeholders in the food industry in designing effective risk management strategies. Furthermore, the sentiment analysis results will allow us to understand the public perception and concerns about specific food risks. This understanding can help develop targeted communication strategies and public awareness campaigns to address these concerns and enhance consumer confidence in the food industry.

5.2. Suspect screening and non-targeted analysis of food-risk components with cheminformatics tools

Traditional methods for detecting FRCs are limited by commercially available FRC standards, which restrict the detection of unknown FRCs and their transformation products within food matrixes. Given this, suspect screening and non-targeted analysis approaches are proposed based on high-resolution mass spectrometry (HRMS) to accelerate the detection of unknown FRCs (Bletsou, Jeon, Hollender, Archontaki, & Thomaidis, 2015).

Suspect screening remains within the domain of target screening but distinguishes itself by utilizing molecular formulas of unknown chemicals as inputs. These formulas can be sourced from various databases or generated by cheminformatics tools, such as EnviPath (Wicker, et al., 2016), ToxinDB (D. Zhang, Tian, et al., 2021), Biotransformer (Djoumbou-Feunang, et al., 2019), and MycotoxinDB (Ji, et al., 2023). For instance, ToxinDB predicts FRC's transformation products based on approximately 8000 reaction rules extracted from known biotransformation reactions. Using ToxinDB, approximately 550,000 putative toxin metabolites were predicted based on 4836 known toxins (D. Zhang, Tian, et al., 2021). During food processing and storage, FRCs can be transformed into masked forms due to mechanical or thermal energies. Given this, MycotoxinDB was designed to predict the products of chemical transformation and masked forms of FRCs. Using MycotoxinDB, the 48 masked forms of deoxynivalenol, a representative FRC mainly contaminated in crops, were predicted, and seven were experimentally verified with LC-HRMS (Ji, et al., 2023).

However, limited by the extent of the suspect list and the availability of reference data, suspect screening may overlook novel or unexpected FRCs and/or their uncharacterized transformation products. To address these limitations, non-targeted analysis emerges as an advancing approach that spans a broader chemical space compared to targeted methods, which often search for limited, pre-defined sets of chemicals. The nontargeted screening-based mass spectrometry Platform has been widely used in the detection of FRCs. For instance, Wang et al. identified five mycotoxin metabolites in rice using a non-targeted screening method based on LC-HRMS (T. Wang, Duedahl-Olesen, & Lauritz Frandsen, 2021). Xu et al. developed a non-targeted screening approach for risk components in animal-derived foods based on an LC-HRMS platform and screened six risk substances, such as sulfamethazine, in meat samples (Fu, et al., 2018). Liang et al. identified seven veterinary drugs, such as metronidazole, in eggs using a nontargeted screening, suggesting potential food safety issues (Liang, et al., 2022).

Non-targeted analysis is useful for exploratory research and novel chemical fingerprint identification, albeit being more time-consuming and resource-intensive compared to targeted suspect screening. Therefore, the choice among different methods should depend on the research objectives and available resources of the specific study. By combining different analytical approaches with cheminformatics tools, researchers are making progress in uncovering the complexity of FRCs, offering further insights and fundamental knowledge toward enhancing food safety assessment.

5.3. Identifying potential food-risk components using genome mining

Besides anthropogenic FRCs, the risks of biogenic toxins that contaminate or originally exist in food are receiving more attention (Tian, et al., 2022). Biogenic toxins are a special type of natural products (NPs), that are usually produced by fungi, bacteria, and plants, and can enter the food chain during planting, food storage, and processing. For example, fermented foods can be contaminated by pathogenic microorganisms during fermentation, posing unneglected food safety risks (D. Zhang, Jia, et al., 2023). When planting crops, mycotoxins produced by fungi can be introduced, whereas warm and wet weather promotes the infection process (Tian, et al., 2022). In addition, some phytotoxins existing in eatable plants can cause serious toxic effects (Logesh, Das, Sellappan, Piesik, & Mondal, 2023). With advances in the field of genome biology, it has been revealed that groups of certain genes, termed biosynthetic gene clusters (BGCs), often appear together that collaborate to synthesize specific biogenic toxins (Medema, et al., 2015). Via identification of BGCs within genomes, genome mining (i.e., the identification of BGCs) has evolved into an efficient way to reveal potential biogenic toxins contaminated in food.

In 2011, the initial version of antiSMASH (Medema, et al., 2011) was introduced as a novel approach to predict NPs by identifying specific patterns within bacterial genomes and retrieving BGCs that encode the synthesis of NPs. Subsequently, multiple tools, such as fungiSMASH (Blin, et al., 2021) and plantiSMASH (Kautsar, Suarez Duran, Blin, Osbourn, & Medema, 2017), have emerged to accelerate the discovery of new NPs and have substantially expanded their application scope from bacteria to plants and fungi. The recent development of tools such as BiG-SCAPE (Navarro-Munoz, et al., 2020) that aid in the classification and comparison of BGC and NPs across diverse microbial genomes has further enriched the toolbox available to researchers, allowing them to explore, classify, and use BGC data effectively. These analytical tools allow researchers to better understand the presence of FRCs and their potential risks across different food sources; for example, exogenous BGCs found in contaminated food matrices may suggest biogenic FRCs that pose health risks. These tools can also be used to predict molecular structures of biogenic toxins, but this prediction usually has high uncertainty. Given this, complementary tools such as NPlinker (Louwen, Medema, & van der Hooft, 2023) integrate genomic data with metabolomic data for FRC detection, offering new insights to identify unknown biogenic FRCs in foodborne pathogens and plants.

Here, we propose a conceptual workflow that integrates data preparation, BGC and NP prediction, and experimental validation to systematically identify potential biogenic FRCs (Fig. 3C). It begins with data preparation, which encompasses the compilation of microbial and/or plant genomic and metabolomic datasets. Following this phase, bioinformatic tools, such as antiSMASH (Medema, et al., 2011), plantiSMASH (Kautsar, et al., 2017), and fungiSMASH (Blin, et al., 2021), can be used to identify BGCs based on genomic or metagenomic data. Subsequently, NPlinker (Louwen, et al., 2023) can link the genomic information to metabolic data to screen BGCs that result in specific FRCs. In parallel, experimental verification, such as MS analysis, can further validate the predictions. Using this strategy, Bignell et al. analyzed the BGCs that may enhance the virulence of opportunistic human pathogens (Bignell, Cairns, Throckmorton, Nierman, & Keller, 2016). Furthermore, illustrative examples of plant BGCs offer valuable insight into forecasting plant-originated toxins within food systems (Polturak, Liu, & Osbourn, 2022).

6. Conclusions and perspectives

In the age of big data, advances in digitization in the food industry and its supply chain as well as extensive food and chemical safety assessments have generated massive amounts of data, leading to the establishment of FRC databases. To guide food scientists in selecting and employing those data resources, this review summarizes 50 FRC databases with different objectives, including their URL, focus, content, accessibility, timeliness, and their applications. By providing centralized access and retrieval of scientific evidence, these FRC databases have played a crucial role in food safety risk assessment, regulation, and communication.

FRC databases compile data from various sources, including scientific literature, regulatory agencies, and industry reports, allowing for a thorough understanding of food-risk components, their properties, toxicity profiles, exposure levels, and regulatory standards. These custom FRC databases facilitate easy access to relevant information, eliminating the need to search through multiple resources or databases, thus saving time and effort for researchers, policymakers, and other stakeholders. Furthermore, the comprehensively compiled toxicity data also provides a foundation for the development of computational toxicology models to bridge data gaps on new substances that generally lack experimental information. Moreover, FRC databases, such as MycotoxinDB and CyanoMetDB, provide data on masked forms and transformation products of FRCs, which promotes accurate and unbiased exposure assessment.

FRC databases also facilitate global collaboration and knowledge sharing among stakeholders in the food safety community. For instance, the GEMS/Food database provides more than 8 million analytical results based on the collaboration between the WHO and over 30 institutions. RASFF ensures the exchange of information between member countries to support swift reaction by food safety authorities in case of risks to public health resulting from the food chain. The newly proposed IFoodCloud tool compiles information from more than 3100 websites, providing comprehensive insights on the public's awareness of food safety. By sharing insights, best practices, and lessons learned, stakeholders can collectively address emerging food safety challenges and develop effective risk regulation and communication strategies that are tailored to local contexts and cultural norms.

Despite the widespread use of FRC databases, we have identified persistent issues with delayed updates and an inadequate commitment to accessibility, reproducibility, and data quality. To address these limitations and promote studies following the FAIR guidelines, we propose initiatives for the future development and management of FRC databases with specific focuses on transparent data sourcing, free accessibility, long-term maintenance, security and privacy measures, and data availability and formats.

Furthermore, we summarize three promising approaches to boost the identification of less studied or unknown FRCs outside the coverage of existing FRC databases, including mining FRCs from literature and public opinion based on NLP, suspect screening and non-targeted analysis of FRCs with cheminformatics tools, and identifying biogenic FRCs with genome mining. We anticipate that data-driven approaches will greatly advance our understanding of "dark matter" chemicals in food and provide more detailed insights for food safety assessment.

Several aspects of the current review could be further expanded upon. Although we have made an effort to comprehensively gather FRC online databases, there may still be omissions within the current review. Our plan is to create a continuously updated portal website containing FRC-related data resources that will enable researchers to access targeted data easily. This will also provide a platform for database users to provide feedback, report bugs, and suggest improvements. Actively engaging with food scientists will not only enhance the usability of databases but also foster a sense of community among researchers in the field. Moreover, data quality can be assessed from various aspects, such as data resources, experiment methods, used equipment, and repeatability reported in the literature. In this assessment, we focus on data resources only due to limited data availability. A more systematic evaluation will be conducted once the accessibility and data traceability of FRC databases improves. This could also result in a standardized protocol or guideline for evaluating the quality of new FRC databases, ensuring the field toward a future in which the FAIR guidelines are applied more thoroughly.

Although many avenues are worth further exploring, we are confident that the present investigation establishes a comprehensive handbook for food scientists looking for high quality data, a robust basis for establishing FRC databases in the future, as well as identifying previously disregarded FRCs with computational techniques. We anticipate widespread use of FRC databases and computational approaches, considerably reducing knowledge gaps in food safety assessment and contributing to the creation of new healthy foods.

CRediT authorship contribution statement

Dachuan Zhang: Conceptualization, Data curation, Writing – original draft, Writing – review & editing, Resources, Supervision, Project administration, Funding acquisition. **Dongliang Liu:** Data curation, Writing – original draft. **Jiayi Jing:** Writing – original draft. **Bingxuan Jia:** Writing – original draft. **Ye Tian:** Writing – review & editing. **Yingying Le:** Supervision. **Yaochun Yu:** Writing – original draft, Writing – review & editing. **Qian-Nan Hu:** Supervision.

Declaration of competing interest

The authors declare no conflicts of interest related to this study.

Data availability

Data will be made available on request.

Acknowledgements

This project received funding from the Applied Research Partnership Grants [Grant number: ARPG_112023_14] funded by the ETH Zurich, the Leading House Asia mandated by Switzerland's State Secretariat for Education, Research and Innovation. The authors thank anonymous reviewers for their insightful and detailed suggestions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tifs.2024.104513.

References

Adnan, M., Xiao, B., Xiao, P., Zhao, P., Li, R., & Bibi, S. (2022). Research progress on heavy metals pollution in the soil of smelting Sites in China. *Toxics, 10*, 231. https://doi.org/10.3390/toxics10050231

Armaroli, E., Bechaux, C., Delagrave, R., Ingenbleek, L., Leblanc, J. C., & Verger, P. (2020). A method to prioritize the surveillance of chemicals in food commodities to access international market and its application to four countries in Sub-Saharan Africa. *Environment International, 135*, Article 105386. https://doi.org/10.1016/j.envint.2019.105386

Arvidson, K. B. (2008). FDA toxicity databases and real-time data entry. *Toxicology and Applied Pharmacology, 233*, 17–19. https://doi.org/10.1016/j.taap

Babiychuk, E. B., & Draeger, A. (2015). Defying death: Cellular survival strategies following plasmalemmal injury by bacterial toxins. *Seminars in Cell & Developmental Biology, 45*, 39–47. Elsevier.

Bacanli, M., & Basaran, N. (2019). Importance of antibiotic residues in animal food. *Food and Chemical Toxicology, 125*, 462–466. https://doi.org/10.1016/j.fct.2019.01.033

Barabási, A.-L., Menichetti, G., & Loscalzo, J. (2020). The unmapped chemical complexity of our diet. *Nature Food, 1*, 33–37. https://doi.org/10.1038/s43016-019-0005-1

Barbosa, J., De Schamphelaere, K., Janssen, C., & Asselman, J. (2021). Prioritization of contaminants and biological process targets in the North Sea using toxicity data from ToxCast. *Science of the Total Environment, 758*, Article 144157. https://doi.org/10.1016/j.scitotenv.2020.144157

Berg, N., De Wever, B., Fuchs, H. W., Gaca, M., Krul, C., & Roggen, E. L. (2011). Toxicology in the 21st century–working our way towards a visionary reality. *Toxicology in Vitro, 25*, 874–881. https://doi.org/10.1016/j.tiv.2011.02.008

Bignell, E., Cairns, T. C., Throckmorton, K., Nierman, W. C., & Keller, N. P. (2016). Secondary metabolite arsenal of an opportunistic pathogenic fungus. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences, 371*, Article 20160023. https://doi.org/10.1098/rstb.2016.0023

Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van Wezel, G. P., Medema, M. H., et al. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research, 49*, W29–W35. https://doi.org/10.1093/nar/gkab335

Borrel, A., Auerbach, S. S., Houck, K. A., & Kleinstreuer, N. C. (2020). Tox21BodyMap: a webtool to map chemical effects on the human body. *Nucleic Acids Res, 48*, W472–W476. https://doi.org/10.1093/nar/gkaa433

Bouzembrak, Y., Steen, B., Neslo, R., Linge, J., Mojtahed, V., & Marvin, H. J. P. (2018). Development of food fraud media monitoring system based on text mining. *Food Control, 93*, 283–296. https://doi.org/10.1016/j.foodcont.2018.06.003

Carvalho, F. P. (2017). Pesticides, environment, and food safety. *Food and Energy Security, 6*, 48–60. https://doi.org/10.1002/fes3.108

CFR. Website. https://www.ecfr.gov/current/title-21/chapter-I/subchapter-B/part-170/subpart-A/section-170.3.

Chakraborty, A., Ghosh, S., Chowdhary, G., Maulik, U., & Chakrabarti, S. (2012). Dbeth: A database of bacterial Exotoxins for human. *Nucleic Acids Research, 40*, D615–D620. https://doi.org/10.1093/nar/gkr942

Chauhan, S. S., Sachan, D. K., & Parthasarathi, R. (2021). FOCUS-DB: An online comprehensive database on food additive safety. *Journal of Chemical Information and Modeling, 61*, 202–210. https://doi.org/10.1021/acs.jcim.0c01147

Chen, S., Huang, D., Nong, W., & Kwan, H. S. (2016). Development of a food safety information database for Greater China. *Food Control, 65*, 54–62. https://doi.org/10.1016/j.foodcont.2016.01.002

Chen, D., Pei, X., Wu, M., Xie, S., Pan, Y., Huang, L., et al. (2019). Development of a networked mass spectral database for veterinary drug residues. *International Journal of Mass Spectrometry, 439*, 1–12. https://doi.org/10.1016/j.ijms.2018.11.014

Crosby, T. C., Kittel, E. C., & Gieseker, C. M. (2022). Phish-pharm: A searchable database of pharmacokinetics and drug residue literature in fish - 2022 update. *The AAPS Journal, 24*, 105. https://doi.org/10.1208/s12248-022-00750-w

Cuarental, A. S.n. (2022). Maximum residue and contaminant limits in food: Databases. *Revista del Comité Científico de la AESAN, 36*, 237–259. https://doi.org/10.5281/zenodo.8118892

'Amico, P. D., Nucera, D., Guardone, L., Mariotti, M., Nuvoloni, R., & Armani, A. (2018). Seafood products notifications in the EU rapid alert system for food and feed (RASFF) database: Data analysis during the period 2011–2015. *Food Control, 93*, 241–250. https://doi.org/10.1016/j.foodcont.2018.06.018

Deng, X., Cao, S., & Horn, A. L. (2021). Emerging applications of machine learning in food safety. *Annual Review of Food Science and Technology, 12*, 513–538. https://doi.org/10.1146/annurev-food-071720-024112

Djoumbou-Feunang, Y., Fiamoncini, J., Gil-de-la-Fuente, A., Greiner, R., Manach, C., & Wishart, D. S. (2019). BioTransformer: A comprehensive computational tool for small molecule metabolism prediction and metabolite identification. *Journal of Cheminformatics, 11*, 2. https://doi.org/10.1186/s13321-018-0324-5

Dorne, J., Richardson, J., Livaniou, A., Carnesecchi, E., Ceriani, L., Baldin, R., et al. (2021). EFSA's OpenFoodTox: An open source toxicological database on chemicals in food and feed and its future developments. *Environment International, 146*, Article 106293. https://doi.org/10.1016/j.envint.2020.106293

Duedahl-Olesen, L., Cederberg, T. L., Christensen, T., Fagt, S., Fromberg, A., Granby, K., et al. (2020). Dietary exposure to selected chemical contaminants in fish for the Danish population. *Food Additives & Contaminants Part A, Chemistry, Analysis, Control, Exposure & Risk Assessment, 37*, 1027–1039. https://doi.org/10.1080/19440049.2020.1743374

EFSA, Younes, M., Aquilina, G., Castle, L., Engel, K. H., Fowler, P., et al. (2019). Scientific opinion on the proposed amendment of the EU specifications for titanium dioxide (E 171) with respect to the inclusion of additional parameters related to its particle size distribution. *EFSA Journal, 17*, Article e05760. https://doi.org/10.2903/j.efsa.2019.5760

EFSA, Younes, M., Aquilina, G., Castle, L., Engel, K. H., Fowler, P., et al. (2021). Safety assessment of titanium dioxide (E171) as a food additive. *EFSA Jorunal, 19*, Article e06585. https://doi.org/10.2903/j.efsa.2021.6585

Eti, S. A., Islam, M. S., Shourove, J. H., Saha, B., Ray, S. K., Sultana, S., et al. (2023). Assessment of heavy metals migrated from food contact plastic packaging: Bangladesh perspective. *Heliyon, 9*, Article e19667. https://doi.org/10.1016/j.heliyon.2023.e19667

EUR-Lex. Website. https://eur-lex.europa.eu/eli/reg/2008/1333/oj.

European Commission. (2024). Website. https://knowledge4policy.ec.europa.eu/publication/food-fraud-data-bases_en.

Evans, R. M., Scholze, M., & Kortenkamp, A. (2015). Examining the feasibility of mixture risk assessment: A case study using a tiered approach with data of 67 pesticides from the Joint FAO/WHO Meeting on pesticide residues (JMPR). *Food and Chemical Toxicology, 84*, 260–269. https://doi.org/10.1016/j.fct.2015.08.015

Everstine, K., Spink, J., & Kennedy, S. (2013). Economically motivated adulteration (EMA) of food: Common characteristics of EMA incidents. *Journal of Food Protection, 76*, 723–735. https://doi.org/10.4315/0362-028X.JFP-12-399

FCRIS. (2024). Website. https://nucleus.iaea.org/sites/fcris/Pages/Home.aspx.

Fu, Y., Zhang, Y., Zhou, Z., Lu, X., Lin, X., Zhao, C., et al. (2018). Screening and determination of potential risk substances based on liquid chromatography-high-resolution mass spectrometry. *Analytical Chemistry, 90*, 8454–8461. https://doi.org/10.1021/acs.analchem.8b01153

Gerassimidou, S., Geueke, B., Groh, K. J., Muncke, J., Hahladakis, J. N., Martin, O. V., et al. (2023). Unpacking the complexity of the polyethylene food contact articles value chain: A chemicals perspective. *Journal of Hazardous Materials, 454*, Article 131422. https://doi.org/10.1016/j.jhazmat.2023.131422

Geueke, B., Groh, K. J., Maffini, M. V., Martin, O. V., Boucher, J. M., Chiang, Y. T., et al. (2022). Systematic evidence on migrating and extractable food contact chemicals: Most chemicals detected in food contact materials are not listed for use. *Critical Reviews in Food Science and Nutrition, 63*, 9425–9435. https://doi.org/10.1080/10408398.2022.2067828

Ginex, T., Spyrakis, F., & Cozzini, P. (2014). Fadb: A food additive molecular database for in silico screening in food toxicology. *Food Additives & Contaminants Part A, Chemistry, Analysis, Control, Exposure & Risk Assessment, 31*, 792–798. https://doi.org/10.1080/19440049.2014.888784

Groh, K. J., Geueke, B., Martin, O., Maffini, M., & Muncke, J. (2021). Overview of intentionally used food contact chemicals and their hazards. *Environment International, 150*, Article 106225. https://doi.org/10.1016/j.envint.2020.106225

Grulke, C. M., Williams, A. J., Thillanadarajah, I., & Richard, A. M. (2019). EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research. *Computational Toxicology, 12*. https://doi.org/10.1016/j.comtox.2019.100096

Gunthardt, B. F., Hollender, J., Hungerbuhler, K., Scheringer, M., & Bucheli, T. D. (2018). Comprehensive toxic plants-phytotoxins database and its application in assessing aquatic micropollution potential. *Journal of Agricultural and Food Chemistry, 66*, 7577–7588. https://doi.org/10.1021/acs.jafc.8b01639

He, Q. Z., Han, W. J., He, Q. Y., Huo, L. J., Zhang, J. J., Lin, Y., et al. (2010). Atdb 2.0: A database integrated toxin-ion channel interaction data. *Toxicon, 56*, 644–647. https://doi.org/10.1016/j.toxicon.2010.05.013

Hinojosa-Nogueira, D., Muros, J. J., Navajas-Porras, B., Delgado-Osorio, A., Perez-Burillo, S., Pastoriza, S., et al. (2023). Development of a food composition database of different food contaminants CONT11 and estimation of dietary exposure in children of southern Spain. *Food and Chemical Toxicology, 177*, Article 113843. https://doi.org/10.1016/j.fct.2023.113843

Hooton, F., Menichetti, G., & Barabasi, A. L. (2020). Exploring food contents in scientific literature with FoodMine. *Scientific Reports, 10*, Article 16191. https://doi.org/10.1038/s41598-020-73105-0

Hur, J., Danes, L., Hsieh, J. H., McGregor, B., Krout, D., & Auerbach, S. (2018). Tox21 enricher: Web-based chemical/biological functional annotation analysis tool based on Tox21 toxicity screening platform. *Molecular Informatics, 37*, Article e1700129. https://doi.org/10.1002/minf.201700129

Ji, J., Zhang, D., Ye, J., Zheng, Y., Cui, J., & Sun, X. (2023). MycotoxinDB: A data-driven platform for investigating masked forms of mycotoxins. *Journal of Agricultural and Food Chemistry, 71*, 9501–9507. https://doi.org/10.1021/acs.jafc.3c01403

Jin, C., Bouzembrak, Y., Zhou, J., Liang, Q., van den Bulk, L. M., Gavai, A., et al. (2020). Big Data in food safety- A review. *Current Opinion in Food Science, 36*, 24–32. https://doi.org/10.1016/j.cofs.2020.11.006

Johnson, R. (2014). *Food Fraud and "Economically Motivated Adulteration" of Food and Food Ingredients*. Congressional Research Service Report.

Jones, M. R., Pinto, E., Torres, M. A., Dorr, F., Mazur-Marzec, H., Szubert, K., et al. (2021). CyanoMetDB, a comprehensive public database of secondary metabolites from cyanobacteria. *Water Research, 196*, Article 117017. https://doi.org/10.1016/j.watres.2021.117017

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). Kegg: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research, 45*, D353–D361. https://doi.org/10.1093/nar/gkw1092

Kautsar, S. A., Suarez Duran, H. G., Blin, K., Osbourn, A., & Medema, M. H. (2017). plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Research, 45*, W55–W63. https://doi.org/10.1093/nar/gkx305

Kemprai, P., Bora, P. K., Saikia, S. P., & Haldar, S. (2023). Chavibetol: Major and potent phytotoxin in betel (piper betle L.) leaf essential oil. *Pest Management Science, 79*, 4451–4462. https://doi.org/10.1002/ps.7645

Krisztina, B., Zsolt, H., & Peter, B. (2005). Rapid alert system for food and feed (RASFF) in the European union. *Magyar Allatorvosok Lapja, 127*, 299–306.

Kuzmenkov, A. I., Krylov, N. A., Chugunov, A. O., Grishin, E. V., & Vassilevski, A. A. (2016). Kalium: A database of potassium channel toxins from scorpion venom. *Database, 2016*. https://doi.org/10.1093/database/baw056. baw056.

Lewis, K. A., Tzilivakis, J., Warner, D. J., & Green, A. (2016). An international database for pesticide risk assessments and management. *Human and Ecological Risk Assessment, 22*, 1050–1064. https://doi.org/10.1080/10807039.2015.1133242

Li, H., Cheng, Y., Luo, J., Li, L., & Wu, Y. (2022). Establishment of a food fraud database and analysis of fraud information based on network data in China. *Journal of Consumer Protection and Food Safety, 17*, 177–181. https://doi.org/10.1007/s00003-021-01361-x

Li, X., Zang, M., Li, D., Zhang, K., Zhang, Z., & Wang, S. (2023). Meat food fraud risk in Chinese markets 2012-2021. *NPJ Science of Food, 7*, 12. https://doi.org/10.1038/s41538-023-00189-z

Liang, W., Zheng, F., Chen, T., Zhang, X., Xia, Y., Li, Z., et al. (2022). Nontargeted screening method for veterinary drugs and their metabolites based on fragmentation characteristics from ultrahigh-performance liquid chromatography-high-resolution mass spectrometry. *Food Chemistry, 369*, Article 130928. https://doi.org/10.1016/j.foodchem.2021.130928

Lim, E., Pon, A., Djoumbou, Y., Knox, C., Shrivastava, S., Guo, A. C., et al. (2010). T3DB: A comprehensively annotated database of common toxins and their targets. *Nucleic Acids Research, 38*, D781–D786. https://doi.org/10.1093/nar/gkp934

Logesh, R., Das, N., Sellappan, G., Piesik, D., & Mondal, A. (2023). Unripe fruits of Litchi chinensis (Gaertn.) Sonn.: An overview of its toxicity. *Annales Pharmaceutiques Françaises, 81*, 925–934. https://doi.org/10.1016/j.pharma.2023.07.004

Louwen, J. J. R., Medema, M. H., & van der Hooft, J. J. J. (2023). Enhanced correlation-based linking of biosynthetic gene clusters to their metabolic products through chemical class matching. *Microbiome, 11*, 13. https://doi.org/10.1186/s40168-022-01444-3

Lüth, S., Boone, I., Kleta, S., & Al Dahouk, S. (2019). Analysis of RASFF notifications on food products contaminated with Listeria monocytogenes reveals options for improvement in the rapid alert system for food and feed. *Food Control, 96*, 479–487. https://doi.org/10.1016/j.foodcont.2018.09.033

Manning, L., & Soon, J. M. (2016). Food safety, food fraud, and food Defense: A Fast evolving literature. *Journal of Food Science, 81*, R823–R834. https://doi.org/10.1111/1750-3841.13256

McCann, D., Barrett, A., Cooper, A., Crumpler, D., Dalen, L., Grimshaw, K., et al. (2007). Food additives and hyperactive behaviour in 3-year-old and 8/9-year-old children in the community: A randomised, double-blinded, placebo-controlled trial. *The Lancet, 370*, 1560–1567. https://doi.org/10.1016/S0140-6736(07)61306-3

Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., et al. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research, 39*, W339–W346. https://doi.org/10.1093/nar/gkr466

Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., et al. (2015). Minimum information about a biosynthetic gene cluster. *Nature Chemical Biology, 11*, 625–631. https://doi.org/10.1038/nchembio.1890

Mielech, A., Puscion-Jakubik, A., & Socha, K. (2021). Assessment of the risk of contamination of food for infants and toddlers. *Nutrients, 13*, 2358. https://doi.org/10.3390/nu13072358

Mol, H., Van Dam, R., Zomer, P., & Mulder, P. P. (2011). Screening of plant toxins in food, feed and botanicals using full-scan high-resolution (Orbitrap) mass spectrometry. *Food Additives & Contaminants Part A, Chemistry, Analysis, Control, Exposure & Risk Assessment, 28*, 1405–1423. https://doi.org/10.1080/19440049.2011.603704

Moore, J. C., Spink, J., & Lipp, M. (2012). Development and application of a database of food ingredient fraud and economically motivated adulteration from 1980 to 2010. *Journal of Food Science, 77*, R118–R126. https://doi.org/10.1111/j.1750-3841.2012.02657.x

Mu, W., Kleter, G. A., Bouzembrak, Y., Dupouy, E., Frewer, L. J., Radwan Al Natour, F. N., et al. (2024). Making food systems more resilient to food safety risks by including artificial intelligence, big data, and internet of things into food safety early warning and emerging risk identification tools. *Comprehensive Reviews in Food Science and Food Safety, 23*, Article e13296. https://doi.org/10.1111/1541-4337.13296

Navarro-Munoz, J. C., Selem-Mojica, N., Mullowney, M. W., Kautsar, S. A., Tryon, J. H., Parkinson, E. I., et al. (2020). A computational framework to explore large-scale biosynthetic diversity. *Nature Chemical Biology, 16*, 60–68. https://doi.org/10.1038/s41589-019-0400-9

Neltner, T. G., Alger, H. M., Leonard, J. E., & Maffini, M. V. (2013). Data gaps in toxicity testing of chemicals allowed in food in the United States. *Reproductive Toxicology, 42*, 85–94. https://doi.org/10.1016/j.reprotox.2013.07.023

Peng, G.-J., Chang, M.-H., Fang, M., Liao, C.-D., Tsai, C.-F., Tseng, S.-H., et al. (2017). Incidents of major food adulteration in Taiwan between 2011 and 2015. *Food Control, 72*, 145–152. https://doi.org/10.1016/j.foodcont.2016.07.043

PestcideInfo. Website. https://www.pesticideinfo.org/.

Polturak, G., Liu, Z., & Osbourn, A. (2022). New and emerging concepts in the evolution and function of plant biosynthetic gene clusters. *Current Opinion in Green and Sustainable Chemistry, 33*, Article 100568. https://doi.org/10.1016/j.cogsc.2021.100568

Popovski, G., Seljak, B. K., & Eftimov, T. (2019). *FoodBase corpus: a new resource of annotated food entities. Database*. https://doi.org/10.1093/database/baz121. baz121.

Richard, A. M., Judson, R. S., Houck, K. A., Grulke, C. M., Volarath, P., Thillainadarajah, I., et al. (2016). ToxCast chemical Landscape: Paving the Road to 21st century toxicology. *Chemical Research in Toxicology, 29*, 1225–1251. https://doi.org/10.1021/acs.chemrestox.6b00135

Roca-Saavedra, P., Mendez-Vilabrille, V., Miranda, J. M., Nebot, C., Cardelle-Cobas, A., Franco, C. M., et al. (2018). Food additives, contaminants and other minor components: Effects on human gut microbiota-a review. *Journal of Physiology & Biochemistry, 74*, 69–83. https://doi.org/10.1007/s13105-017-0564-2

Shi, X. X., Wang, F., Wang, Z. Z., Huang, G. Y., Li, M., Simal-Gandara, J., et al. (2022). Unveiling toxicity profile for food risk components: A manually curated toxicological databank of food-relevant chemicals. *Critical Reviews in Food Science and Nutrition*, 1–16. https://doi.org/10.1080/10408398.2022.2152423

Tao, D., Yang, P., & Feng, H. (2020). Utilization of text mining as a big data analysis tool for food science and nutrition. *Comprehensive Reviews in Food Science and Food Safety, 19*, 875–894. https://doi.org/10.1111/1541-4337.12540

Thaivalappil, A., Young, I., Paco, C., Jeyapalan, A., & Papadopoulos, A. (2020). Food safety and the older consumer: A systematic review and meta-regression of their knowledge and practices at home. *Food Control, 107*, Article 106782. https://doi.org/10.1016/j.foodcont.2019.106782

Thompson, L. A., Darwish, W. S., Ikenaka, Y., Nakayama, S. M. M., Mizukawa, H., & Ishizuka, M. (2017). Organochlorine pesticide contamination of foods in Africa: Incidence and public health significance. *Journal of Veterinary Medical Science, 79*, 751–764. https://doi.org/10.1292/jvms.16-0214

Tian, Y., Zhang, D., Cai, P., Lin, H., Ying, H., Hu, Q.-N., et al. (2022). Elimination of Fusarium mycotoxin deoxynivalenol (DON) via microbial and enzymatic strategies:

Current status and future perspectives. *Trends in Food Science & Technology, 124*, 96–107. https://doi.org/10.1016/j.tifs.2022.04.002

Tiozzo, B., Ruzza, M., Rizzoli, V., D'Este, L., Giaretta, M., & Ravarotto, L. (2020). Biological, chemical, and nutritional food risks and food safety issues from Italian online information sources: Web monitoring, content analysis, and data visualization. *Journal of Medical Internet Research, 22*, Article e23438. https://doi.org/10.2196/23438

Tolosa, J., Serrano Candelas, E., Valles Pardo, J. L., Goya, A., Moncho, S., Gozalbes, R., et al. (2023). MicotoXilico: An interactive database to predict mutagenicity, genotoxicity, and carcinogenicity of mycotoxins. *Toxins, 15*, 355. https://doi.org/10.3390/toxins15060355

Toropov, A. A., Toropova, A. P., Marzo, M., Dorne, J. L., Georgiadis, N., & Benfenati, E. (2017). QSAR models for predicting acute toxicity of pesticides in rainbow trout using the CORAL software and EFSA's OpenFoodTox database. *Environmental Toxicology and Pharmacology, 53*, 158–163. https://doi.org/10.1016/j.etap.2017.05.011

Tox21. (2024). Website. https://ntp.niehs.nih.gov/whatwestudy/tox21.

Vasconcelos Neto, M., Quintal, A. P. N., Porto, L. B. G., & Vitorino Carvalho de Souza, S. (2021). Lead in Brazilian food: Exposure assessment and risk characterization. *Food Additives & Contaminants Part A Chem Anal Control Expo Risk Assess, 38*, 315–325.

Wang, T., Duedahl-Olesen, L., & Lauritz Frandsen, H. (2021). Targeted and non-targeted unexpected food contaminants analysis by LC/HRMS: Feasibility study on rice. *Food Chemistry, 338*, Article 127957. https://doi.org/10.1016/j.foodchem.2020.127957

Wang, S. S., Lin, P., Wang, C. C., Lin, Y. C., & Tung, C. W. (2023). Machine learning for predicting chemical migration from food packaging materials to foods. *Food and Chemical Toxicology, 178*, Article 113942. https://doi.org/10.1016/j.fct.2023.113942

Wang, Z., Walker, G. W., Muir, D. C. G., & Nagatani-Yoshida, K. (2020). Toward a global understanding of chemical pollution: A first comprehensive analysis of national and regional chemical Inventories. *Environmental Science and Technology, 54*, 2575–2584. https://doi.org/10.1021/acs.est.9b06379

Wang, C., Wu, J., & Gao, X. (2014). History,reality and future of food additives. *Food Additives & Contaminants Part A, Chemistry, Analysis, Control, Exposure & Risk Assessment, 38*, 315–325. https://doi.org/10.1080/19440049.2020.1852321

Weininger, D. (1990). Smiles .3. Depict - Graphical Depiction of chemical structures. *Journal of Chemical Information and Computer Sciences, 30*, 237–243. https://doi.org/10.1021/ci00067a005

WHO. (2023). Website https://www.who.int/activities/assessing-chemical-risks-in-food.

Wicker, J., Lorsbach, T., Gutlein, M., Schmid, E., Latino, D., Kramer, S., et al. (2016). enviPath–The environmental contaminant biotransformation pathway resource. *Nucleic Acids Research, 44*, D502–D508. https://doi.org/10.1093/nar/gkv1229

Wilkinson, M. D., Dumontier, M., Jan Aalbersberg, I., Appleton, G., Axton, M., Baak, A., et al. (2019). Addendum: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 6*, 6. https://doi.org/10.1038/s41597-019-0009-6

Wishart, D., Arndt, D., Pon, A., Sajed, T., Guo, A. C., Djoumbou, Y., et al. (2015). T3DB: The toxic exposome database. *Nucleic Acids Research, 43*, D928–D934. https://doi.org/10.1093/nar/gku1004

Xing, H., Zhang, D., Cai, P., Zhang, R., & Hu, Q. N. (2023). RDBridge: A knowledge graph of rare diseases based on large-scale text mining. *Bioinformatics, 39*, Article btad440. https://doi.org/10.1093/bioinformatics/btad440

Yao, Y., & Long, M. (2020). The biological detoxification of deoxynivalenol: A review. *Food and Chemical Toxicology, 145*, Article 111649. https://doi.org/10.1016/j.fct.2020.111649

Zhang, D., Cheng, X., Sun, D., Ding, S., Cai, P., Yuan, L., et al. (2020). AdditiveChem: A comprehensive bioinformatics knowledge-base for food additive chemicals. *Food Chemistry, 308*, Article 125519. https://doi.org/10.1016/j.foodchem.2019.125519

Zhang, D., Gong, L., Ding, S., Tian, Y., Jia, C., Liu, D., et al. (2020). Frcd: A comprehensive food risk component database with molecular scaffold, chemical diversity, toxicity, and biodegradability analysis. *Food Chemistry, 318*, Article 126470. https://doi.org/10.1016/j.foodchem.2020.126470

Zhang, D., Jia, C., Sun, D., Gao, C., Fu, D., Cai, P., et al. (2023). Data-driven prediction of molecular biotransformations in food fermentation. *Journal of Agricultural and Food Chemistry, 71*, 8488–8496. https://doi.org/10.1021/acs.jafc.3c01172

Zhang, D., Ouyang, S., Cai, M., Zhang, H., Ding, S., Liu, D., et al. (2020). FADB-China: A molecular-level food adulteration database in China based on molecular fingerprints and similarity algorithms prediction expansion. *Food Chemistry, 327*, Article 127010. https://doi.org/10.1016/j.foodchem.2020.127010

Zhang, D., Tian, Y., Tian, Y., Xing, H., Liu, S., Zhang, H., et al. (2021). A data-driven integrative platform for computational prediction of toxin biotransformation with a case study. *Journal of Hazardous Materials, 408*, Article 124810. https://doi.org/10.1016/j.jhazmat.2020.124810

Zhang, Y., Yang, D. J., Fang, C. R., & Wei, K. K. (1997). Monitoring of organochlorine pesticide residues–the GEMS/Food program in China. *Biomedical and Environmental Sciences, 10*, 102–106.

Zhang, D., Zhang, T., Liu, S., Sun, D., Ding, S., Cheng, X., et al. (2021). SARS2020: An integrated platform for identification of novel coronavirus by a consensus sequence-function model. *Bioinformatics, 37*, 1182–1183. https://doi.org/10.1093/bioinformatics/btaa767

Zhang, H., Zhang, D., Wei, Z., Li, Y., Wu, S., Mao, Z., et al. (2023). Analysis of public opinion on food safety in Greater China with big data and machine learning. *Current Research in Food Science, 6*, Article 100468. https://doi.org/10.1016/j.crfs.2023.100468