

DrugBank 6.0: the DrugBank Knowledgebase for 2024

Craig Knox¹, Mike Wilson¹, Christen M. Klinger¹, Mark Franklin¹, Eponine Oler², Alex Wilson¹, Allison Pon¹, Jordan Cox¹, Na Eun (Lucy) Chin¹, Seth A. Strawbridge¹, Marysol Garcia-Patino¹, Ray Kruger², Aadhavya Sivakumaran², Selena Sanford², Rahil Doshi², Nitya Khetarpal², Omolola Fatokun², Daphnee Doucet², Ashley Zubkowski², Dorsa Yahya Rayat², Hayley Jackson², Karxena Harford², Afia Anjum², Mahi Zakir², Fei Wang², Siyang Tian², Brian Lee², Jaanus Liigand^{2,3}, Harrison Peters², Ruo Qi (Rachel) Wang¹, Tue Nguyen¹, Denise So¹, Matthew Sharp¹, Rodolfo da Silva¹, Cyrella Gabriel¹, Joshua Scantlebury¹, Marissa Jasinski¹, David Ackerman¹, Timothy Jewison¹, Tanvir Sajed², Vasuk Gautam² and David S. Wishart ^{©2,4,5,6,*}

¹OMx Personal Health Analytics, Inc., 700–10130 103 St NW, Edmonton, AB T5J 1B9, Canada

²Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada

⁴Department of Computing Science, University of Alberta, Edmonton, AB T6G 2E8, Canada

⁵Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, AB T6G 2H1, Canada

⁶Department of Laboratory Medicine and Pathology, University of Alberta, Edmonton, AB T6G 1C9, Canada

*To whom correspondence should be addressed. Tel: +1 780 492 8574; Email: david.wishart@ualberta.ca

Abstract

First released in 2006, DrugBank (https://go.drugbank.com) has grown to become the 'gold standard' knowledge resource for drug, drug-target and related pharmaceutical information. DrugBank is widely used across many diverse biomedical research and clinical applications, and averages more than 30 million views/year. Since its last update in 2018, we have been actively enhancing the quantity and quality of the drug data in this knowledgebase. In this latest release (DrugBank 6.0), the number of FDA approved drugs has grown from 2646 to 4563 (a 72% increase), the number of investigational drugs has grown from 3394 to 6231 (a 38% increase), the number of drug-drug interactions increased from 365 984 to 1 413 413 (a 300% increase), and the number of drug-food interactions expanded from 1195 to 2475 (a 200% increase). In addition to this notable expansion in database size, we have added thousands of new, colorful, richly annotated pathways depicting drug mechanisms and drug metabolism. Likewise, existing datasets have been significantly improved and expanded, by adding more information on drug indications, drug-drug interactions, drug-food interactions and many other relevant data types for 11 891 drugs. We have also added experimental and predicted MS/MS spectra, 1D/2D-NMR spectra, CCS (collision cross section), RT (retention time) and RI (retention index) data for 9464 of DrugBank's 11 710 small molecule drugs. These and other improvements should make DrugBank 6.0 even more useful to a much wider research audience ranging from medicinal chemists to metabolomics specialists to pharmacologists.

Graphical abstract



Received: September 16, 2023. Revised: October 9, 2023. Editorial Decision: October 10, 2023. Accepted: October 16, 2023 © The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

³Institute of Chemistry, University of Tartu, Tartu, Estonia

Introduction

Drug research has increasingly become a field of data science. High-throughput technologies such as robotic drug screening assays, in silico screening/docking, automated compound library generation, high throughput genotyping and various combinations of multi-omics studies are generating huge amounts of drug-related or drug-target data. The amount of drug data being generated is leading to greater use of artificial intelligence (AI) and machine learning (ML), both of which use 'big data' to identify novel patterns and trends. Datadriven AI has already made significant contributions to earlystage drug discovery ranging from target identification to lead identification/optimization to drug repurposing to chemical synthesis optimization, and even to the prediction of important drug properties such as efficacy, toxicity and drug-drug interactions (1-3). Data science is also impacting other areas of pharmaceutical practice. Clinicians, pharmacists and other healthcare professionals (HCPs) require accurate, up-todate drug data to make informed decisions. Often, the sheer volume of drug-gender, drug-response, drug-genotype, drugdrug and drug-food interaction data is overwhelming, and the increasing use of automated clinical decision support (CDS) platforms is reflecting the data challenges faced by HCPs (4). As a result, there is a continued trend toward the development of increasingly sophisticated data-dependent rule-based or AI-based algorithms for drug informatics (5).

It is because of these emerging trends in drug informatics that we established DrugBank in 2006 (6). DrugBank was originally designed to be a comprehensive, up-to-date, freely available web resource containing detailed drug, drug–target, drug mechanism of action and drug interaction information for both approved and experimental drugs. By making this rich, high quality, primary-sourced content freely available, we hoped it would make the discovery of new drugs, the repurposing of old drugs, the understanding of drug mechanisms and the tracking of drug interactions easier for academics, medicinal chemists, pharmacists and pharmaceutical companies. This concept appears to have been very appealing and as a result, DrugBank has become one of the world's most widely used reference drug resources, receiving >30 million web accesses each year and averaging >5000 citations/year.

Throughout the past 18 years, DrugBank's content, accessibility and layout have continuously evolved, reflecting advances in hardware and internet technology, requests from its diverse user community and specific needs identified by the DrugBank curation and programming teams. DrugBank 1.0, released in 2006, provided modest amounts of data on selected FDA-approved drugs and their drug-targets in a convenient web-based format (6). Over the following two decades, DrugBank has gone through multiple updates, that have included important additions such as pharmacogenomic data, drug metabolism data and comprehensive ADMET (absorption, distribution, metabolism, excretion and toxicity) information. DrugBank 5.0, the last update which appeared in 2018, added more investigational drug data, new drug-drug interaction data, new types of pharmaco-omic data and a significant amount of drug spectral (MS and NMR) information (7).

Herein we provide an overview of the most recent, openaccess version of the DrugBank knowledgebase. DrugBank 6.0 incorporates numerous enhancements and improvements over the previous version, including significantly more FDA-

approved drugs (a 72% increase), many more investigational drugs (a 38% increase), a massive increase in catalogued drug-drug interactions (nearly 300%) and the number of tracked drug-food interactions (a 200% increase). In addition to this notable expansion in database size (Table 1), we have added thousands of new, colorful, richly annotated pathways depicting drug mechanisms and drug metabolism. These have been complemented by thousands of newly (and more accurately) predicted spectral data for 1D ¹H and ¹³C NMR, LC-MS, GC-MS and related chromatographic data. Drug-Bank's layout has also been modified to incorporate new data fields and new kinds of data searches. Additionally, the Drug-Bank curation team has worked to improve data quality, by developing and refining data entry SOPs, improving training and expanding the number of curators. These updates and enhancements will be described in more detail under the following four headings: (i) Frontend Layout, Backend and Curation Protocols; (ii) Data Additions; (iii) Data Quality Enhancements and (iv) Interface Improvements. Overall, we believe these advancements will ensure DrugBank remains a vital data resource in a landscape that is increasingly reliant on data-driven, in silico approaches.

Layout, backend and curation protocols

Frontend layout

DrugBank's web layout has evolved significantly since the last release in 2018. The new homepage (https://go.drugbank. com) has been redesigned to be more aesthetically pleasing and functional, with a large centrally located search bar that allows users to query (by name) for 'Drugs', 'Targets', 'Pathways' and 'Indications' (Figure 1A). Below the search bar are several display panels with links to DrugBank's restrictedaccess (i.e. commercial) offerings, followed by a 'What's trending' section that capture frequently viewed drugs and products. As noted in our last update from 2018 (7), DrugBank has evolved to become a public-private partnership (P3), with revenue from sales of its commercial products (via OMx Personal Health Analytics Inc.) to the pharma industry being used, in partnership with grant funding from the University of Alberta and other public Canadian funding agencies, to support the regular maintenance and updating of the open-access version of DrugBank.

As before, the DrugBank homepage offers dropdown menu tabs through a navigation bar to support 'Browse', 'Search', 'Downloads' and 'About'. Under 'Browse', users may browse DrugBank by 'Drugs', 'Starred Drugs', 'Categories', 'Pathways', 'Drug Reactions', 'Drug Classifications' and 'Drug Targets'. 'Starred Drugs' are a new feature offered in Drug-Bank 6.0 that were developed, based on user feedback, to allow users to customize their viewing preferences. Customized viewing requires that users must login so that the website can track specific user preferences. Once logged in, a user may 'star' one or more drugs so that they will appear in user-specified list for ease of revisiting. This is similar to the mechanism by which browsers or online vendors remember user preferences for customized web viewing. The 'Categories' page allows users to search DrugBank categories by either drug category or description. This option also shows users the number of drugs and associated drug targets within each drug category. Similarly, 'Pathways' can be searched by the pathway name, the pathway category (such

Table 1.	Comparison	of key data	between DrugBank 5.0) and DrugBank 6.0
----------	------------	-------------	----------------------	--------------------

Category	DrugBank 5	DrugBank 6	% Increase	
# of search types	20	23	15%	
# of illustrated drug-action pathways	319	404	26.7%	
# of illustrated drug metabolism pathways	64	2721	4151%	
# of drug metabolites with structures	1360	3037	123.3%	
# of drug metabolism reactions	1530	3703	142.0%	
# of drugs with drug transporter data	1954	3408	74.4%	
# of drugs with taxonomic classification information	7387	12723	72.2%	
# of food-drug interactions	1195	2475	107.1%	
# of drug-drug interactions	365 984	1 413 413	286.2%	
# of drugs with experimental NMR spectra	922	1822	97.6%	
# of drugs with experimental MS spectra	2521	2888	15%	
# of approved small molecule drugs	2110	2751	30.4%	
# of approved drugs with product ingredient structures	1551	4030	159.8%	
# of approved biotech drugs	555	1601	188.5%	
# of nutraceutical drugs	97	134	38.1%	
# of withdrawn drugs	209	317	51.7%	
# of illicit drugs	202	205	1.5%	
# of experimental drugs	4964	6722	35.4%	
# of investigational drugs (phase I/II/III trials)	4501	6231	38.4%	
# of all drug targets (unique)	4563	4939	8.2%	
# of approved-drug enzymes/carriers (unique)	479	526	9.8%	
# of all drug enzymes/carriers (unique)	497	556	11.9%	
# of linked drug indications	3024	3820	26.3%	
# of clinical trials	245 356	464 870	89.5%	

as 'Metabolic', 'Physiological', 'Signaling', 'Drug metabolism', 'Drug action' or 'Disease'), or by any of the drugs associated with that pathway. 'Drug Reactions' refers primarily to drug metabolic reactions corresponding to a substrate, an associated enzyme and a product (all of which can be searched). The 'Drug Targets' page includes options that allow users to search by drug, type of interaction (by filtering for enzyme, blood protein carrier and membrane transporter interactions) and biological entity. Four additional browse options, 'Pharmaco-genomics', 'Pharmaco-metabolomics', 'Pharmacotranscriptomics' and 'Pharmaco-proteomics' are available; each containing relevant omic data linked to drugs. All browse pages also contain options to filter by drug group (such as 'approved') or market availability across the US, Canada and the EU.

Under 'Search', users may search DrugBank by Chemical Structure (via drawn structure images, Figure 1B), Molecular Weight (including MW ranges), Drug/Food Interactions, Target (protein) Sequences, Pharmaco-omics and Advanced Text searches (by specifying predicates and inclusion fields, Figure 1C and D). A separate search option is available for eight different spectral/chromatographic searches (MS, MS/MS, GC/MS, 1D NMR, 2D NMR, CCS, RI and RT). New tabs have been added to the menu bar, including 'Interaction Checker' to support drug-drug and food-drug interaction queries (described later in this manuscript) and 'Products,' which covers licensable Drug-Bank products available primarily to the pharmaceutical industry. The 'About' dropdown contains FAQs and guidelines on how to cite DrugBank, as well as general information about the Wishart research group at the University of Alberta.

To address the needs of mobile users, we also undertook extensive development of a mobile-friendly interface for Drug-Bank. Users visiting the homepage (https://go.drugbank.com) on a mobile device such as a smartphone or tablet, are greeted with a similar view as on desktop, but with a condensed single dropdown menu box in the top right corner denoted by three horizontal lines (Figure 1E). Clicking this menu shows a series of additional dropdown menus that match those displayed in the desktop version. When viewing a DrugCard, a web page that displays a specific drug's information within DrugBank, the user is presented with a similar layout as seen on a desktop. However, the side navigation bar is replaced by a floating 'NAV' bar that pops out when the user clicks on it (Figure 1F). This navigation bar behaves in the same manner as on desktop; clicking a section will scroll the drug card to that section and expands any subsections within the navigation bar itself; which can also be clicked (Figure 1G).

Backend design

DrugBank 6.0 was implemented using a newly updated version of Ruby on Rails (http://rubyonrails.org, version 6.1.0 running on Ruby 3.2) web framework incorporating a MariaDB SQL relational database (https://mariadb.com, version 10.4) and a Neo4j graph database (https://neo4j.com, version 3.5) to manage all drug data, including entity relationships, external references, descriptions, visualization specifications and chemical structures. DrugBank utilizes a modular monolithic architecture, in which various systems are isolated based on clear domain boundaries. The structured knowledge stored in both its relational and graph databases is dynamically extracted and rendered into web pages by DrugBank's HTML interface responder. Search functionality across the website and its application programming interfaces (APIs) is provided by a custom search mapping software powered by an Elasticsearch cluster (http://elastic.co, version 6.5). Drug-Bank services (website, chemical structure handling, APIs and data processing pipelines) are distributed and vertically scaled across a series of virtual servers (UpCloud, http://upcloud. com) equipped in total with 138 CPU cores, 9.6 TB of disk space, 600 GB of RAM and an additional 15 TB of storage space on an Amazon S3 storage facility.



Figure 1. Highlights of the current DrugBank Online interface. This figure shows screenshots highlighting different parts of the current DrugBank Online user interface. (A) The main search bar present on the homepage that allows users to search by Drugs, Targets, Pathways and Indications. (B) DrugBank's chemical search page; users can draw arbitrary chemical structures as search inputs to the known structures present in DrugBank. Selecting 'Molecular Weight' at the top will instead allow users to supply a range of molecular weights with which to filter DrugBank drugs by (C) DrugBank's advanced search function, which allows users to design complex queries using predicates. Here, for example, the name of any returned entry must contain the string 'acet' and the drug must be approved. Users can also specify one or more additional fields to show up in the results; these fields, however, have no effect on the actual search. (D) The first result of the query shown in (C). Users who are logged in to a free account can also export the small floating 'NAV' bar on the left hand side of the screen. (G) The navigation bar in action on mobile; in this case, the user has selected 'Categories', revealing the subsections present in the drug card for this section.

Knowledgebase curation protocols

The DrugBank curation team (comprising bioinformaticians, trained pharmacists and individuals with undergraduate or graduate degrees in pharmacology or biochemistry) is jointly managed by the University of Alberta and by OMx Personal Health Analytics Inc. Both sites have implemented a rigorous curator training program spanning a period of roughly two months. During this time, trainees progress from dedicated learning of DrugBank's data and platforms, through to small curation projects under the supervision of a senior curator, to independent curation work (which is al-

ways peer reviewed by members of the independent curator team).

Throughout the entire process, significant emphasis is placed on using a robust set of standard operating protocols (SOPs) developed by DrugBank's senior curators. Currently, this set of SOPs spans hundreds of specific activities, describing in detail how to find, format and enter information into the DrugBank knowledgebase. New SOPs are added to mirror new data types or processes, and existing SOPs are kept up to date on a standard verification interval. All information entered by DrugBank curators at both the private (OMx Inc.) and public (University of Alberta) sites undergoes a secondary review by the curation team lead or other senior curators. Likewise, all changes made to the DrugBank knowledgebase are tracked in DrugBank's annotation system, and tagged to specific curators so that feedback can be provided in the case of error reports. Regular audits are applied to various aspects of the knowledgebase to ensure information currency.

Data additions

New approved drugs

Since the publication of DrugBank 5.0, there have been 919 newly approved drugs and drug salts reported across 11 different countries or regions. Of these, 393 are small molecule drugs and 526 are biotech (protein, DNA or cell-based) drugs. The biotech drugs can be further broken down into allergenics [numbering 51], gene therapies [21], cell transplant therapies [11], protein-based therapies [89] and vaccines [249]. Aside from vaccines (which are mostly protein), protein-based therapies represent the largest increase, of which 51/89 were monoclonal antibodies.

Since 2018, DrugBank has expanded its reported drug coverage from just three geographic regions [the US (FDA), Canada (Health Canada) and the EU (European Medicines Agency)] to 11 geographic regions (Table 2). For each supported region, DrugBank matches the tempo of the public release of new information by the relevant drug agency, aligning them to existing DrugCards (if possible). If no DrugCard exists, a new one is created after confirmation that the product ingredient represents a new active ingredient not previously captured within DrugBank. Collectively, these eight new regions account for 223 127 of the 570 091 drug products in DrugBank (40% of DrugBank's total list of drug entities).

For the US, Canada and EU, the DrugBank team manually monitors new drug approvals and adds or updates existing DrugCard entries. New drug approvals and updates to existing product labels including indications, contraindications and pharmaceutical forms are tracked daily. Emergency use authorizations, withdrawals and the availability of new biosimilars are also tracked. Finally, for the US only, safetyrelated labelling changes (SrLCs), such as new adverse effects, boxed warnings and interactions are monitored. Once identified, these updates are passed to senior curators who review the updated product label and make any necessary changes. This daily tracking ensures that complete information for all approved drugs and drug products across all supported geographic regions are included within DrugBank in a timely manner.

New drug metabolites

Since the publication of DrugBank 5.0, there have been 1677 new drug metabolites added for which structural information is available, bringing the total of such entries to 3037. This corresponds to 2173 new metabolism reaction entries (with data on substrates and products, involved enzyme(s), reaction type, metabolite activity and other details). Metabolism information is manually curated based on product monographs and published literature, often synthesizing data from multiple sources to recapitulate the entire pathway for a given drug.

New investigational drugs

Since the publication of DrugBank 5.0, there have been 3161 new investigational/experimental (phase III or earlier) drugs added across all supported geographic regions, bringing the total number of such drugs in DrugBank to 11 262. Of these, 1747 are small molecule drugs while the other 1414 are biotech drugs. These drugs are largely added based on imports from clinicaltrials.gov, wherein drug-based interventions that do not have an existing match in DrugBank prompt the creation of new DrugCards. Additionally, the DrugBank curation team routinely monitors upcoming potential approvals based on positive clinical trial results and expend additional effort to curate information for these entries ahead of formal regulatory approval.

New drug-drug and drug-food interactions

Since the publication of DrugBank 5.0, the number of drugdrug interactions has grown from 365 984 to 1 413 791, nearly a 300% increase. This large number of interactions is possible in part due to the newly developed ability to create interactions between drug categories and between an individual drug and drug category. However, these less specific categorybased interactions are subservient to specific drug-drug interactions catalogued for any given pair of drugs, ensuring the most accurate information on drug-drug interactions is provided.

All interactions have a 'subject' drug and an 'affected' drug, allowing for a notion of causality within interactions. Each interaction also has a 'type,' which succinctly captures the nature of the interaction; there are 12 types of interactions within DrugBank (https://dev.drugbank.com/guides/ drug_interactions/parameters). Of these, all types with the exception of 'increase risk of hypersensitivity' can either be associated with 'increase' or 'decrease.' This captures whether the subject drug increases or decreases the relevant property of the affected drug. Additional types, including 'change dynamics,' 'change specific active metabolite,' 'change specific adverse effects,' and 'change specific effects' also allow related entities to be specified such as a metabolite or adverse effect. These interaction types are present as separate entries, but also appear in a standardized 'summary' string describing the interaction.

Other important information captured within drug-drug interactions entries include the evidence level, severity, description and management information. The evidence level is either encoded as 1 or 2, depending on if the information comes from a product monograph (encoded as 1) or other credible sources showing in vitro or clinical evidence for the interaction (encoded as 2). Within DrugBank, there are three severity levels for an interaction: 'minor,' 'moderate,' and 'major.' By our definition, minor interactions exist but are generally not clinically relevant, moderate interactions may or may not result in any substantial changes for a patient, and major interactions should prompt consideration of additional monitoring or treatment alteration by a HCP. The description and management fields are authored by curators to describe the interaction in more detail and steps that can assist in clinically managing the interaction, respectively.

In addition to interactions with other drugs, drug-food interactions represent an important data type for both patients and HCPs alike. The number of drug-food interactions has grown from 1195 in DrugBank 5.0 to 2475 in DrugBank 6.0. This was due to a focused review conducted in 2020 and on-

Geographical region	Responsible regulatory authority	# of products	
United States of America	Food and Drug Administration (USFDA)	240 511	
Canada	Health Canada	40 922	
European Union	European Medicines Agency (EMA)	14 629	
Austria	Federal Office for Safety in Healthcare (BASG)	9516	
Italy	Italian Medicines Agency (AIFA)	124 943	
Turkey	Turkish Medicines and Medical Devices Agency (TITCK)	21 596	
Colombia	National Institute for Food and Drug Surveillance (INVIMA)	22 452	
Indonesia	BADAN POM	12 690	
Malaysia	National Pharmaceutical Regulatory Agency (NPRA)	9428	
Thailand	Thai Health Information Standards Development Center (THIS)	16 448	
Singapore	Health Sciences Authority (HSA)	6115	

Table 2. Supported regulatory regions in DrugBank 6.0

going efforts by the curation team to capture these data. Similar to drug–drug interactions, drug–food interactions in Drug-Bank possess an interaction type, of which there are 68. The most common types (accounting for 66% of all interactions) are 'take with or without food', 'avoid alcohol', 'take with food', 'avoid St. John's Wort', 'avoid grapefruit products' and 'take on an empty stomach'. These interaction types form the basis for the interaction description, which can also be augmented by additional information, when available. Both types of drug–drug and drug–food interactions come primarily from prescribing information, but are also extracted from peer reviewed articles, clinical guidelines and other reputable sources. New interactions are added based on updates to existing drug labels, as monitored by the curation team.

New drug-entity interactions

Since the publication of DrugBank 5.0, there have been 3446 new drug-target entries added, with 546 distinct new bioentities (proteins, genes, other cellular components) identified. This has led to a total of 4939 distinct bioentities representing drug targets, which is an 8% increase compared to the 4563 in DrugBank 5.0. For new protein entities, we use UniProt keywords associated with each UniProt identifier to annotate the protein targets. Roughly 33% of these are kinases, suggesting that kinases have represented a particularly valuable target class in the last six years. Certain cellular functions were also highly represented, with nucleotide binding (44%) and transferase (40%) activities being common among newly added target entities. In addition to drug targets, DrugBank 6.0 has added 2550 new drug-enzyme entries, 1560 new drugtransporter entries and 550 new drug-carrier entries. These include 65, 60 and 14 new distinct bioentities, respectively. The total number of distinct enzymes, transporters and carriers has increased to 836.

These entries have been added as a result of routine curation efforts, including those for both new drugs and updates to existing drugs. Additionally, the DrugBank team has built several natural language processing (NLP) models capable of scraping PubMed abstracts to identify potential drug–entities and drug–target relationships. These NLP models are integrated into curator workflows in a way that increases the speed with which curators can review predicted and established relationships as well as the associated reference(s).

New drug mechanism pathways

The curation team has worked particularly hard over the last three years to generate 1057 drug mechanism-of-action

(MOA) pathways using the PathWhiz pathway drawing system (8). These new pathways are of much higher quality (in terms of content and detail) than previous pathways, and are also more aesthetically pleasing. To generate these pathways, drugs were classified based on known interactions contained within DrugBank and then further reviewed via additional searches in PubMed (9). In most MOA pathways, information relating to the known route of administration and corresponding path the drug takes to reach the intended site of action are included. Most MOA pathways show direct interactions between the drugs, their receptors, the affected cells, relevant organs and organelles. Detailed text descriptions for each pathway have also been added along with citations. Currently, 1706 approved small molecule and biologic drugs in DrugBank lack well-defined or confirmed drug-target actions (inducer/inhibitor/agonist/etc.); as such, these drugs do not have MOA pathways.

New drug metabolism pathways

The curation team also generated 2721 drug metabolism (ADME) pathways based on existing (manually curated) DrugBank drug metabolism data. These pathways are intended to cover the entire journey of each drug from ingestion, distribution to the site of metabolism, the metabolic reactions the drug undergoes and finally culminating in drug/metabolite excretion. Data for the manually generated ADME pathways was gathered via the information contained in DrugBank's extensive (manually annotated) drug metabolism datasets. The ADME pathway was then manually rendered using the same SOPs as with the MOA pathways, including depictions of the organs, organelles, the metabolic reactions as well as a clear indication of how the drug and metabolites are excreted. ADME pathways showing the degradation of DrugBank's protein or nucleic acid (biotech) drugs to individual nucleotides or amino acids have also been added for all biotech drugs.

New spectral and chromatographic data

For DrugBank 6.0, significant numbers of experimental MS/MS and EI-MS data for purified reference drug compounds have also been added to support analytical studies of drugs and drug metabolites. These include 34 069 experimental MS/MS spectra for 1943 drug compounds and 4666 experimental MS/MS spectra for 240 drug metabolites. These updates are visible under the 'Spectra' field in each DrugCard. We have also added 3390 experimental EI-MS spectra for 857 drugs and another 524 EI-MS spectra for 122 metabolites, respectively. Unfortunately, very little experimental reference

NMR, retention index (RI) or retention time (RT) data exists for most drugs or drug metabolites. Likewise, not all drugs have high-quality experimental MS/MS or EI-MS data. As a result, for DrugBank 6.0 significant resources were placed into generating: (i) accurate MS/MS spectral predictions; (ii) accurate 1D ¹H and ¹³C NMR spectral predictions; (iii) accurate retention index (RI) data and (iv) accurately predicted collision cross section (CCS) data for analyzing ion mobility spectroscopy (IMS) data (Table 3).

The MS/MS predictions for DrugBank 6.0 were performed by the latest version of the competitive fragment modeling tool called CFM-ID version 4.0 (10). The performance of CFM-ID 4.0 is approximately 30% better than previous versions of CFM-ID. CFM-ID 4.0 was used to predict both the positive ion and negative ion mode MS/MS spectra (at collision energies 10, 20 and 40 eV) for ~10 000 small molecule drug compounds in DrugBank 6.0. This led to the generation of 59 721 MS/MS predicted spectra for drug compounds and 15 144 predicted MS/MS spectra for 3132 drug metabolites. These MS/MS data have also been incorporated into Drug-Bank's new MS/MS search functions.

In addition to the MS/MS predicted data, we also generated 18 370 ¹H and ¹³C NMR spectral predictions for 9185 drug compounds using NMR spectral prediction tools described previously for HMDB (11) and NP-MRD (12). Recent advances in NMR theory along with continuing innovations in computing techniques are allowing remarkably accurate NMR spectral simulations and NMR parameter predictions to be made for many small molecules, with ¹H and ¹³C shifts having errors as small as <0.15 ppm for ¹H shifts and <1.5 ppm for ¹³C shifts. The NMR data in DrugBank covers 1D ¹H and ¹³C NMR spectra for all drugs dissolved in H₂O at 500 MHz. These NMR data have also been incorporated into DrugBank's NMR search function.

Retention indices (RI) are another useful set of observables that can be used to identify molecules via GC-MS. RIs are essentially adjusted retention times used in gas chromatography that allow nearly universal comparisons of retention times across GC platforms. The DrugBank team used a machine learning algorithm called RI-Pred (13) that has an RI error of <2% (Table 3). Using a cut-off mass of 900 Daltons (the upper mass limit for most GC-MS instruments), a total of \sim 4000 compounds were selected from the DrugBank as being 'GC-MS' compatible. These compounds were then computationally derivatized with TMS and TBDMS to generate ~100k derivatized structures. RI-Pred was then used to predict the retention indices for all ~100k derivative structures across three standard types of GC columns (semi-standard non-polar, standard non-polar and standard polar). This led to the generation of ~300k predicted column-specific retention indices-all of which have been entered in the 'Predicted Spectral Properties' subsection (under the 'Properties' field) of every GC-MS compatible drug. These RI data have been incorporated into Drug-Bank's GC-MS search function as described later.

The development of ion mobility spectroscopy (IMS) and the appearance of tandem IMS-MS systems has led to a growing interest in using IMS for compound identification. IMS retention values are related to the average collision cross section (CCS) of the molecule, which can be accurately predicted based on a compound's 3D structure. For DrugBank 6.0 we used several published CCS predictors (Table 3), including All-CCS (14) and DeepCCS (15) to generate the CCS values for 11 417 small molecule drugs and 2852 metabolites. Most of these CCS predictors report errors of <3–4%. Using these predictors, a total of 34 251 predicted CCS values for drugs and 8556 predicted CCS for drug metabolites have been added to the DrugBank. All predicted CCS values have been entered in the 'Predicted Spectral Properties' subsection (under the 'Properties' field) of the relevant DrugCard. These CCS values have also been incorporated into DrugBank's new 'LC– MS Search' and 'LC–MS/MS Search' functions, which are described later.

Finally, we added another important observable: retention time (RT) to DrugBank's drug data. RT corresponds to the time taken for a given compound to elute from a liquid chromatography (LC) column (usually a high performance or HPLC system). The RT value is highly dependent on the system, solvents, column type and chromatographic method (CM). We used a machine learning based RT predictor called RTPred predictor version 1.0 (Table 3) that was trained on the METLIN (16) small molecule dataset. Using this method, we have generated 38 244 and 9672 RT predictions corresponding to 9561 drugs and 2418 metabolites, respectively. These are being generated for four different chromatographic methods, described within each DrugCard. All predicted RT values have been entered in the 'Predicted Spectral Properties' subsection (under the 'Properties' field) of the relevant Drug-Card.

Improved spectral search functions

The addition of many new or newly predictable spectral observables (CCS, RI, RT, MS, NMR chemical shifts, etc.) also necessitated a substantial upgrade to the spectral search functions for DrugBank 6.0. Furthermore, improvements in our spectral visualization program (as explained in the next section) also allowed us to undertake improvements in the graphical display of the spectral match output. Both the 'LC-MS Search' and 'LC-MS/MS Search' functions now support IMS and RT data as an additional search constraint. Both have an option to input a CCS value or an RT value with a default 5% tolerance. If no CCS or RT input value is provided, the search functions will still perform their regular MS or MS/MS searches without the CCS constraint. Matched compounds for new 'LC-MS Search' are ranked according to their m/z and CCS or RT matches (using a combined weight of 90% for delta m/z and 10% for delta CCS/RT).

The output table from 'LC–MS Search' provides a browsable list that contains information on the matching compound names, the DrugBank links (IDs), their m/z values, the CCS/RT matches (if a CCS or RT value was provided) and the overall score. Matched compounds for this improved 'LC– MS/MS Search' are ranked according to their spectral similarity and CCS/RT matches. The output table from 'LC–MS/MS Search' provides a browsable list with similar information as the LC–MS search. Clicking on the 'Show Spectrum' produces a mirror plot with the input spectrum shown at the top (in red) and the matching spectrum at the bottom (in blue).

The improved 'GC–MS Search' has now been modified to support RI data as an additional search constraint. Users may input an RI value with a default 3% tolerance. If the RI option is chosen, users must also choose any one of three types of GC columns as the RI values are specific to the column type. Additionally, the type of chemical derivatization(s) used must be provided. Users have the option to indicate no derivatization, TMS derivatization, or TBDMS derivatization. If no RI input

Predictor/ method	Tool/ algorithm	Version	Reference
NMR Predictor	NmrShiftDB	2.0	Jonas and Kuhn 2019 (20)
NMR Predictor	NmrPred	1.0	unpublished
MS-MS Predictor	CFMID 4.0	4.5	Wang <i>et al.</i> 2021 (10)
RI Predictor	RI Kovats	1.0	Anjum <i>et al.</i> 2023 (13)
RT Predictor	de novo ML-based	N/A	Unpublished
CCS Method(s)	All CCS	1.10	Zhou <i>et al.</i> 2020 (14)
CCS Method(s)	DarkChem	N/A	Colby <i>et al.</i> 2020 (21)
CCS Method(s)	DeepCCS	N/A	Plante et al. 2019 (15)

value is provided, the GC–MS search function will still perform its regular EI-MS-only search without the RI constraint. Matched compounds for DrugBank's new 'GC–MS Search' are ranked according to their spectral similarity and RI similarity.

DrugBank's 'NMR Search' has been simplified and it now allows users to enter lists of ¹H or ¹³C chemical shifts to search for spectral matches against experimental NMR spectra, predicted NMR spectra or both. Users must provide a chemical shift list (relative intensities are optional), select the nucleus (¹H or ¹³C) of interest and choose a chemical shift tolerance (default of 0.2 ppm for ¹H and 2.0 ppm for ¹³C) before pressing the 'Search' button. A typical query produces a browsable table of hits showing the compound name, the DrugBank ID, the structure, the chemical formula, the molecular weight, the chemical shift dot-product-score (a measure of weighted dot product between query peaks and library peaks), the fraction of peak matches and a colored 'Show Spectrum' button. Clicking on 'Show Spectrum' produces a JSV mirror plot with the input NMR spectrum shown at the top (in red) and the database NMR spectrum shown at the bottom (in blue).

Improved data visualization

Both NMR and MS spectra can be viewed through this 'View Spectrum' page via a locally developed JavaScript spectral viewer called JSpectraViewer or JSV (17). For predicted MS data, JSV allows users to mouse over each peak to interactively view the predicted mass and fragment ion structure. The MS data for both experimental and predicted spectra are available and downloadable as lists of m/z values and intensities in *.txt and mzML format. For NMR data, JSV is somewhat more sophisticated and now supports the display of both 1D and 2D NMR spectra. JSV displays NMR peak/chemical shift assignments both on the NMR spectrum and on the molecule itself, which is shown as a thumbnail image with numbered atoms and an assignment table. In the spectral view window JSV displays blue traces, which correspond to the predicted/simulated NMR spectra while the black traces correspond to the experimental NMR spectra. JSV for NMR also supports interactive spectral zooming, moving, gridding, scaling and image saving/downloading. Each NMR spectrum of a pure compound (experimental or predicted) in the Drug-Bank has downloadable information in the form of a set of peak lists (CSV format), peak assignments (CSV), spectral images (PNG), a spectral and/or assignment validation report and the actual or simulated NMR data in the form of nmrML (18) and JCAMPDX files (19). Wherever experimental data is available, DrugBank provides native free-induction-decay (FID) or time-domain data in the original depositor format (Bruker, Varian, Agilent, JEOL).

Data quality enhancements

Enhanced drug interaction data

Drug-drug interactions within DrugBank have been greatly improved through better curator training, the use of SOPs and the use of drug categories to generate interactions. In addition, our regular monitoring of the US, Canada and EU for labeling updates now includes new drug-drug and drug-food interactions, ensuring better drug interaction coverage. In addition to these efforts, curators have also undertaken several large-scale projects that either directly or indirectly improved the quality of our interactions data. One such initiative (conducted in 2020–2021) addressed the accuracy of therapeutic categorizations, some of which are used to generate drug-drug interactions. This included an initial review of roughly 1200 therapeutic drug-category relations associated to 819 distinct drugs. After further review, a total of 901 relations for drugs were changed to no longer be therapeutic and were instead replaced by a different drug-category therapeutic relation. Currently, 151 therapeutic categories generate drug-drug interactions within DrugBank. As such, the improved accuracy of their drug membership represents a substantial improvement to the overall quality of DrugBank's drug interaction dataset.

Another initiative was undertaken in 2020 to improve the capture of narrow therapeutic index (NTI) drugs in Drug-Bank. As different sources capture this concept differently, we developed an internal definition of 'NTI' that includes professional consensus on NTI designation, required therapeutic monitoring and available overdose outcomes. For this review, 425 drugs deemed potentially NTI based on various external sources were reviewed, leading to 254 drugs that fit our internal definition. A set of 24 categories that can participate in interactions was then checked to include all relevant NTI drugs. All remaining interactions, not specifically linked to the NTI status of the drug, were also reviewed and new interactions were added where necessary. These efforts ensure accurate drug-drug interaction information for NTI drugs, whose serum levels must be maintained within a narrow range to avoid unacceptable dose-related toxicity.

Data gap filling

In addition to the large number of drugs added since the previous DrugBank version, 11 891 existing drug entries have also been thoroughly updated. This gap filling was possible due to the increase in the number of permanent curation team members, but also facilitated by our rigorous update tracking system for previously approved drugs. Input from the broader community into inaccuracies and omissions from existing entries is another key reason, as curators will routinely check all fields of an existing drug when addressing such issues.

Α	DRIG INTERACTIONS	В			
	ADD DRUG TO CHECK FOR INTERACTIONS Tylenol Add at least two and up to five drugs. Check Interactions		ADD DRUG	to CHECK FOR INTERACTIONS	
	Warning: If no interactions are found between two drugs, it does not necessarily mean that no interactions exist. Always consult with a healthcare professional.		CH Warning: If no mean that no inte	eck Interactions CLEAR > Interactions are found between two drugs, it does not rractions exist. Always consult with a healthcare profess	LOAD EXAMPLE necessarily ional.
С	Interactions Found	D	Clarithro	mycin IX Warfarin (WAOR	The metabolism of Warfarin can be decreased when combined with Cambrowyncin.
	Clarithromycin Dit Warfarin Muon EXTENDED DESCRIPTION The subject drug is a strong CYP3A4 enzyme inhibitor; and the affected drug Concomitant administration of these agents will decrease the metabolism of	The metabolism of Warfarin can be decreased when combined with classification of the combined with is metabolized by the CYP3A4 enzyme. the CYP3A4 substrate, increasing READ MORE	EXTENDED	The subject drug is a strong CVP3A4 enzyme inhibitor, and the affected Concombant administration of these agents will decrease the metabolic concentration and therapeutic effect. Drugs with a narrow therapeutic concentration range in order to be start additional discuss. An increased o index can lead to significant adverse effects and toxicity ^{CALD}	I drug is metabolized by the CMP3A4 enzyme. Isin of the CMP3A4 substrate, increasing its serum index must be maintained within a specific concentration of a drug with a narrow therepeutic READ LISS
E	REFERENCES 1. Dresser GX, Spence JD, Balley DG: Pharmacokinetic pharmacodynamic con P450 344 inhibition. Clin Pharmacokinet. 2000 Jan,38(1):41-57. [Article] Hover over products below to view reaction partners Hover over (Valproic acid) (Valproic 1)	requences and clinical relevance of cytochrome BEAD MORE products below to view reaction partners	REFERENCES	 Dresser GX, Spence JD, Balley DG. Pharmacokinetic-pharmacolynami P403 DAI inhibition. Clin Pharmacokinet. 2003 Jan;98(1):4157, [Keidel 2: Doui SF) Ongol Sheeha si skottetsking. Hishborta and induces of hum May(94(3):1052; Lenticia) 3: uprich T, Vine A. The effect of cytochrome P403 metabolism on drug Physician. 2007 Aug 1:-96(3):914. [Lenticia] A. Keine Y, Zanger U. Miramacogenomics of Cytochrome P403 DAI: Ree Publism. Troot Genet. 2013 P40 ZiAr12. doi: 10.1016/j. Balis Ki, Vikida Y, Moger TA, Neuham A. Drugs with narrow that Detail toppilating patients. Pharm Priot Elinabal. 2010; Aug(8):055. Epui 4: TAN Amore Therappeolic clinks drugs prest [Link] 	c consequences and clinical relevance of systechrome of an systechrome P460 JAA. Curr Drug Metab. 2008 response, interactions, and adverse effects. Am Fam cent Progress Toward the "Missing Heritability" 2. collifiction JOTS (Inrice) clinical and JACob to the risk management of 2010 Mar YS. (Inrice)
	(4-ene-Valproic acid) (2,4-Oiene-VPA) (5+Hydroxyvalproic acid) (3-Hydroxyvalproic acid) (3-Hydroxyvalproic acid) (3-Xovvalproic acid) (3-Oxvvalproic acid) (3-Oxvvalproic acid)	ne-Valproic acid (2.4-Diene-VPA) hydroxyvalproic acid (2.2-Propylglutaric acid) hydroxyvalproic acid) (3.0xovalproic acid) (3.0xovalproic acid)	Reaction: 2-n-Pr	opyl-4-oxopentanoic acid to 1 product	Seepond of a section
	(4+Hydroxyvalproic acid) (4+Hydroxyvalproic acid) (2:n-Propyl-4-oxopentanoic acid) (2:n-Propylsuccinic acid) (Valproic acid [:-o-glucuronide) (valproic acid) (2:nevalproic acid) (2:nevalproic acid)	Verforgenzalprois acid 2n-Propyl-t-oxopentanois acid 2Propylsuscinis acid prois acid (>o-glucuronide) me-Valprois acid 3Hydroxyvalprois acid	Rexcise type Direction Major directiong metabolice? Spontaneeus? Activity Activity	No Autom	~~i
	(32)-2-Propylpent-3-enoic acid (32-Ene-VPA) (33)-2-Propylpent-3-enoic acid (3E-Ene-VPA) (34) (35)-2-Propylpent-3-enoic acid (3E-Ene-VPA) (37) (38) (38) (39) (31) (32) (32) (32) (33) (34) (34) (35) (36) (37) (38) (38) (38) (38) (39) (31) (31) (32) (32) (32) (33) (34) (34) (35) (36) (37) (38) (38) (39) (31) (32) (32) (33) (34) (35) (35) (36) (37) (38) (38) (39) (31) (31) (32) (32) (33) (34) (35) (35) (36) <tr< th=""><th>2)-2-Propylpent-3-enoic acid (32-Ene-VPA)) =)-2-Propylpent-3-enoic acid (3E-Ene-VPA)) proic acid CoA</th><th>Comments Exzymes</th><th>1. Guidea R. Galla, Gaussiana, Y. Ghara, Y. Andre Y. Hallo, "Beneficial and the Bella" modern for any defension for general. Next Available:</th><th>A NUM AN ANY ANY ADDRESS AND ADDRESS AND ADDRESS AND ADDRESS ADDRES</th></tr<>	2)-2-Propylpent-3-enoic acid (32-Ene-VPA)) =)-2-Propylpent-3-enoic acid (3E-Ene-VPA)) proic acid CoA	Comments Exzymes	1. Guidea R. Galla, Gaussiana, Y. Ghara, Y. Andre Y. Hallo, "Beneficial and the Bella" modern for any defension for general. Next Available:	A NUM AN ANY ANY ADDRESS AND ADDRESS AND ADDRESS AND ADDRESS ADDRES

Figure 2. Improved user interfaces in DrugBank 6.0. This figure outlines some of the interface improvements in the latest version of DrugBank. (A) An overview of the new interaction checker interface, showing the search bar. (B) The same view as (A), but with three drugs added. Clicking the pink button with a minus sign removes the drug from the selection. (C) One of three interactions found upon clicking 'Check Interactions,' which populate below the main interface. Note that this interaction is between Clarithromycin and Warfarin and is considered 'major' in terms of severity. (D) The same view as (C) showing the result of expanding both the extended description and references sections. (E) An overview of the new visual metabolism interface present in drug cards. (F) Appearance of the interface when a user mouses over 2-propylsuccinic acid. Note that this metabolite and its direct parent, 2-n-propyl-4-oxopentanoic acid, are coloured pink while all previous metabolites and the starting compound are coloured grey. (G) The reaction overview screen reached by clicking on 2-propylsuccinic acid.

Many updates were also due to large-scale curation projects aimed at improving specific sets of drugs or data. In terms of drugs, many updates were added as part of a project conducted during 2019–2020 aimed at extensive curation of the top 300 drugs in global clinical use, as ascertained from the WHO list of globally prescribed medications (https://www.who.int/publications/i/ item/WHOMVPEMPIAU2019.06). Other gap fill-in efforts included improvements to drug-food interactions, chemical structures and drug vocabularies (synonyms, names and codes) as well as increased drug coverage for clinical trials.

Interface improvements

New drug-drug and drug-food interaction checkers

Although the addition of large volumes of interaction data for drugs, both with food and other drugs (see section and Table 1), is highly beneficial, the average user may find these large

datasets difficult to navigate. The most common reason for accessing DrugBank's drug interaction data is to cross-reference foods/drugs that are known to interact with a given drug. It is with this use case in mind that we developed and implemented a web-based interaction checker. While there was a drug interaction checker available in DrugBank 5.0, this has been greatly enhanced in both form and functionality for DrugBank 6.0. This tool can be accessed under 'Interaction Checker' or by navigating to the URL (https://go.drugbank.com/drug-interaction-checker).

Users are met with a page that has a tab-based division between drug-drug and drug-food interactions. On either tab, users are prompted to add drugs through the use of a search bar. The new search now accepts drug synonyms and common branded product names (Figure 2A). Both types of interaction checker accept up to five drugs at a time (Figure 2B). For drug-drug interactions, at least two drugs must be entered before clicking 'check interactions.' Food interactions are reported in a simple table with one interaction description per row. Drug interactions are reported in a series of card-like entries, with each entry containing the names of the interacting drugs, the severity of the interaction, a brief description, an extended description containing more details and any references associated with the interaction (Figure 2C). For concise display, only the first reference is shown by default, but the others can be viewed by clicking the pink 'READ MORE' text (Figure 2D).

Improved drug metabolism view

Drug metabolism can be difficult to visually capture when provided only in terms of structured data entries, especially for complex metabolic schemes involving multiple pathways from drug to end metabolites. To assist users in understanding drug metabolism, we have added an interactive 'tree-like' graphical interface to drug card entries, under the text metabolism description (Figure 2E). Individual molecules are represented as labelled ellipses connected by arrows, with each arrow pointing from the substrate to the product in an individual metabolic reaction. A single substrate can give rise to multiple possible products. When a user places their cursors over an individual ellipse, both this ellipse and its direct parent (together with the arrow connecting them) will be highlighted in pink. Any more distant ancestors, if present, will be highlighted in grey (Figure 2F). These visual cues make it easier for users to follow individual reaction pathways. Clicking on any of the metabolites will open a page detailing the metabolic reaction giving rise to the metabolite, including the involved enzyme(s), if known (Figure 2G).

Conclusion and future plans

DrugBank has continued to expand since its original inception in 2006, with each release bringing more data and more features. DrugBank 6.0 is no different, with this year's release adding significantly more data as well as a variety of new data types (Table 1). Many of these datasets have grown several fold in just the past six years, and it is likely this pace will continue moving forward as more and more relevant drug information enters the public domain.

We believe the public-private partnership (P3) between the University of Alberta and OMx Personal Health Analytics Inc. has helped to greatly improve the breadth, currency and quality of data in DrugBank. It is becoming exceedingly difficult to maintain active, heavily used open-access databases for an extended period of time, especially with current public funding models. However, through this partnership DrugBank is now supported by a much larger team of curators, developers and other supporting staff who are dedicated to improving and maintaining an open-access version DrugBank. As such, we believe this rather unique arrangement substantially benefits both the academic research community and the general public. Indeed, this may be a useful model for other data resources wishing to ensure long-term sustainability, accuracy and currency.

Looking ahead, there are several areas of possible focus for the next update of DrugBank. As previously mentioned, we continue to develop ML-powered tools to process and structure publicly available data at scale and deliver targeted details to curators for manual approval. Combined with potential improvements to the detail captured for each drug-target interaction, this could represent an invaluable resource to those working in the drug development space. Other areas for continued advancement include those that are more focused on DrugBank's architecture and user interfaces. Although we have made great strides in terms of tracking data lineage, there is still more development needed to enable a truly robust tracking system. Similarly, new MLpowered tools to automatically track and report issues within our systems would be welcome additions to avoid interruptions in uptime, erroneous alterations of data and other such concerns.

DrugBank continues to grow and evolve over time thanks in large part to continued interest from the community, thoughtful user feedback and the dedicated individuals who work every day to improve it. While the current version has introduced a number of improvements in terms of the volume of new data, the variety of new data types and a series of user interface additions, there is much more to be done. Nevertheless, we are hopeful that this version of Drug-Bank will continue to serve its users for many years to come.

Data availability

DrugBank 6.0 adheres to FAIR guiding principles (see https:// go.drugbank.com/fair). An extensive and well-annotated data download section is also provided with most drug and drugtarget information available in standard XML and CSV formats. There are several interoperability datasets released under a Creative Commons (CC) CC0 (public domain) International License, as well as additional datasets released under a Creative Commons (CC) 4.0 License Suite according to the Attribution BY and Non-Commercial NC licensing conditions. The CC0 datasets include: (i) a 'DrugBank Vocabulary', to enable easy linking to DrugBank concepts by ID, name or structural identifier and (ii) 'DrugBank Structures' which includes the chemical structures for drugs, as well as names and accession numbers. The CC BY-NC 4.0 datasets include an XML export that includes many details sub-modules such as drug targets, interactions, sequences, chemical properties and metabolism information. There are additional CC BY-NC 4.0 datasets that are for more specific use cases (e.g. protein sequence in FASTA format).

Acknowledgements

The authors would like to thank ChemAxon, Inc. for their continued support as well as the many users of DrugBank for their valuable feedback and suggestions.

Funding

Canadian Institutes of Health Research (CIHR); Natural Sciences and Engineering Research Council (NSERC) Alliance Program; Alberta Innovates (via the Campus Alberta Small Business Engagement Program - CASBE); Genome Alberta, a division of Genome Canada. Funding for open access charge: Research grants from CIHR, NSERC, Genome Canada and Alberta Innovates.

Conflict of interest statement

None declared.

References

- 1. Pun,F.W., Ozerov,I.V. and Zhavoronkov,A. (2023) AI-powered therapeutic target discovery. *Trends Pharmacol. Sci.*, 44, 561–572.
- 2. Tran,T.T.V., Tayara,H. and Chong,K.T. (2023) Artificial intelligence in drug metabolism and excretion prediction: recent advances, challenges, and future perspectives. *Pharmaceutics*, **15**, 1260.
- 3. Chan,H.C.S., Shan,H., Dahoun,T., Vogel,H. and Yuan,S. (2019) Advancing drug discovery via artificial intelligence. *Trends Pharmacol. Sci.*, **40**, 592–604.
- White,N.M., Carter,H.E., Kularatna,S., Borg,D.N., Brain,D.C., Tariq,A., Abell,B., Blythe,R. and McPhail,S.M. (2023) Evaluating the costs and consequences of computerized clinical decision support systems in hospitals: a scoping review and recommendations for future practice. *J. Am. Med. Inform. Assoc.*, 30, 1205–1218.
- 5. Hassan, N., Slight, R., Morgan, G., Bates, D.W., Gallier, S., Sapey, E. and Slight, S. (2023) Road map for clinicians to develop and evaluate AI predictive models to inform clinical decision-making. *BMJ Health Care Informatics*, **30**, e100784.
- 6. Wishart,D.S., Knox,C., Guo,A.C., Shrivastava,S., Hassanali,M., Stothard,P., Chang,Z. and Woolsey,J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.*, 34, D668–D672.
- Wishart,D.S., Feunang,Y.D., Guo,A.C., Lo,E.J., Marcu,A., Grant,J.R., Sajed,T., Johnson,D., Li,C., Sayeeda,Z., et al. (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res., 46, D1074–D1082.
- 8. Pon,A., Jewison,T., Su,Y., Liang,Y., Knox,C., Maciejewski,A., Wilson,M. and Wishart,D.S. (2015) Pathways with PathWhiz. *Nucleic Acids Res.*, 43, W552–W559.
- 9. Wishart,D.S., Li,C., Marcu,A., Badran,H., Pon,A., Budinski,Z., Patron,J., Lipton,D., Cao,X., Oler,E., *et al.* (2020) PathBank: a comprehensive pathway database for model organisms. *Nucleic Acids Res.*, **48**, D470–D478.
- Wang,F., Liigand,J., Tian,S., Arndt,D., Greiner,R. and Wishart,D.S. (2021) CFM-ID 4.0: more accurate ESI-MS/MS spectral prediction and compound identification. *Anal. Chem.*, 93, 11692–11700.

- D622–D631.
 12. Wishart,D.S., Sayeeda,Z., Budinski,Z., Guo,A.C., Lee,B.L., Berjanskii,M., Rout,M., Peters,H., Dizon,R., Mah,R., *et al.* (2022) NP-MRD: the natural products magnetic resonance database. *Nucleic Acids Res.*, 50, D665–D677.
- Anjum,A., Liigand,J., Milford,R., Gautam,V. and Wishart,D.S. (2023) Accurate prediction of isothermal gas chromatographic Kováts retention indices. J. Chromatogr. A, 1705, 464176.
- Zhou,Z., Luo,M., Chen,X., Yin,Y., Xiong,X., Wang,R. and Zhu,Z.-J. (2020) Ion mobility collision cross-section atlas for known and unknown metabolite annotation in untargeted metabolomics. *Nat. Commun.*, 11, 4334.
- Plante,P.L., Francovic-Fontaine,É., May,J.C., McLean,J.A., Baker,E.S., Laviolette,F., Marchand,M. and Corbeil,J. (2019) Predicting ion mobility collision cross-sections using a deep neural network: DeepCCS. *Anal. Chem.*, 91, 5191–5199.
- 16. Domingo-Almenara,X., Guijas,C., Billings,E., Montenegro-Burke,J.R., Uritboonthai,W., Aisporna,A.E., Chen,E., Benton,H.P. and Siuzdak,G. (2019) The METLIN small molecule dataset for machine learning-based retention time prediction. *Nat. Commun.*, 10, 5811.
- Wishart,D.S., Feunang,Y.D., Marcu,A., Guo,A.C., Liang,K., Vázquez-Fresno,R., Sajed,T., Johnson,D., Li,C., Karu,N., *et al.* (2018) HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.*, 46, D608–D617.
- Schober, D., Jacob, D., Wilson, M., Cruz, J.A., Marcu, A., Grant, J.R., Moing, A., Deborde, C., De Figueiredo, L.F., Haug, K., *et al.* (2018) nmrML: a community supported open data standard for the description, storage, and exchange of NMR data. *Anal. Chem.*, 90, 649–656.
- Davies, A.N. and Lampen, P. (1993) JCAMP-DX for NMR. Appl. Spectrosc., 47, 1093–1099.
- Jonas, E. and Kuhn, S. (2019) Rapid prediction of NMR spectral properties with quantified uncertainty. J. Cheminform., 11, 50.
- 21. Colby,S.M., Nuñez,J.R., Hodas,N.O., Corley,C.D. and Renslow,R.R. (2020) Deep learning to generate in silico chemical property libraries and candidate molecules for small molecule identification in complex samples. *Anal. Chem.*, **92**, 1720–1729.

Received: September 16, 2023. Revised: October 9, 2023. Editorial Decision: October 10, 2023. Accepted: October 16, 2023 © The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

(http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com