frontiers | Research Topics

Drug discovery and development explained: introductory notes for the general public

Edited by
Bruno Villoutreix, Jean-Luc Poyet and Katya Tsaioun

Published in
Frontiers in Drug Discovery

ISSN 1664-8714
ISBN 978-2-8325-5774-7
DOI 10.3389/978-2-8325-5774-7

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Drug discovery and development explained: introductory notes for the general public

Topic editors

Bruno Villoutreix — Hôpital Robert Debré, France

Jean-Luc Poyet — INSERM U976 Immunologie, Dermatologie, Oncologie, France

Katya Tsaioun — Johns Hopkins University, United States

Citation

Villoutreix, B., Poyet, J.-L., Tsaioun, K., eds. (2024). *Drug discovery and development explained: introductory notes for the general public*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5774-7

Table of contents

frontiers | Frontiers in Drug Discovery

TYPE Editorial
PUBLISHED 29 November 2024
DOI 10.3389/fddsv.2024.1528467

OPEN ACCESS

EDITED AND REVIEWED BY
Liang Xu,
University of Kansas, United States

*CORRESPONDENCE
Bruno O. Villoutreix,
bruno.villoutreix@inserm.fr

RECEIVED 14 November 2024
ACCEPTED 18 November 2024
PUBLISHED 29 November 2024

CITATION
Villoutreix BO, Poyet J-L and Tsaioun K (2024) Editorial: Drug discovery and development explained: introductory notes for the general public.
Front. Drug Discov. 4:1528467.
doi: 10.3389/fddsv.2024.1528467

Editorial: Drug discovery and development explained: introductory notes for the general public

Bruno O. Villoutreix[1,2]*, Jean-Luc Poyet[3] and Katya Tsaioun[2,4]

[1]Inserm U1141, Hospital Robert Debré, University Paris Cité, Paris, France, [2]Akttyva Therapeutics, Inc., Boston, MA, United States, [3]Onco-Dermatology and Therapies, Inserm UMRS976, Hospital Saint Louis, France Institut de Recherche Saint Louis, University Paris Cité, Paris, France, [4]Department of Environmental Health Sciences, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

KEYWORDS

small molecules, peptides, antibody, therapeutic agents, animal model, drug discovery, artificial intelligence

Editorial on the Research Topic
Drug discovery and development explained: introductory notes for the general public

Finding new medications is a complex and costly process. Yet, as we work towards creating better, safer, and faster treatments, it is essential to make this process more understandable and accessible to everyone. This Research Topic is dedicated to introducing the main concepts and methods in a way that is accessible to all. With contributions from experts across the field, this Research Topic aims to demystify the drug development pipeline and address some of the its most pressing challenges (Figure 1).

The article by Singh et al. provides a comprehensive overview of the entire drug discovery process, explaining the key steps from basic research to the final stages of clinical trials and post-market surveillance. By laying out the five main stages of drug discovery: pre-discovery, discovery, preclinical development, clinical trials, and approval, this article serves as a primer for readers unfamiliar with the field. The review also explains what are the main therapeutic agents (e.g., small molecules, peptides, biologics like antibodies. . .), the pros and cons of drug repurposing and highlights the high cost, long timelines and high attrition rates associated with drug discovery and development. The integration of artificial intelligence (AI) with traditional or novel experimental technologies, offers promising avenues to eventually streamline the process. Yet, many obstacles remain, including the lack of high quality data and the difficulty in understanding the disease state and human biology.

Chavez-Hernandez et al. explore the critical role of chemical and biological data in drug discovery. Their review underscores the importance of balancing the quantity and quality of data, especially as AI and machine learning methods become integral to the drug design process. The authors advocate for a better reporting of both active and inactive compounds to foster a more comprehensive understanding of bioactivity, emphasizing the need for balanced datasets that should drive more accurate predictions and hopefully lead to better treatments.

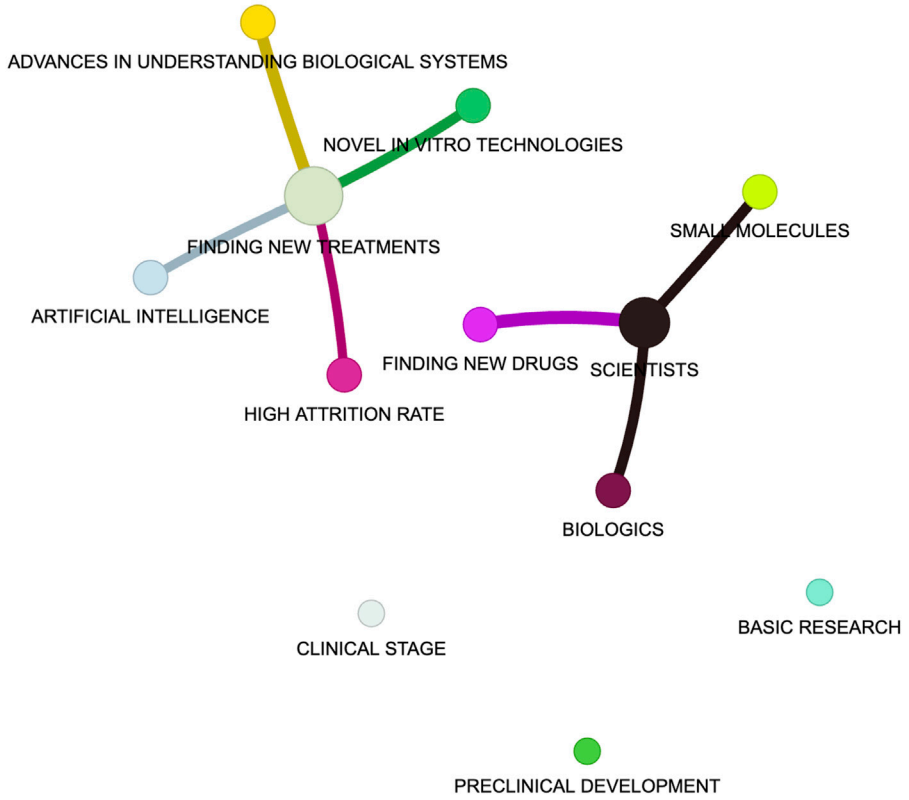


FIGURE 1
The graph illustrates a semantic network constructed automatically (i.e., without human-in-the-loop) using the LightRAG framework (https://arxiv.org/abs/2410.05779) and the abstracts of this Research Topic. LightRAG builds upon existing retrieval-augmented generation (RAG) methodologies by integrating graph-based retrieval in a more efficient, context-aware, and dynamic manner. It uses advanced language processing and a large language model (LLM, here we used Mistral-7B Instruct) to identify key concepts (nodes) and their relationships (edges) within a selected document. This process involves breaking down the text, embedding it into a high-dimensional space, and extracting meaningful entities and their interactions, resulting in a visual map that highlights the main themes and connections across the analyzed text. Such AI methods (and many others), used on scientific results and data, may assist the drug discovery process in a near future.

Gadiya et al. explain the concept of FAIR (Findable, Accessible, Interoperable, Reusable) data management. In an era where data is often siloed within institutions and companies, the lack of accessible datasets hinders progress. The authors first suggest that embracing FAIR principles across the drug discovery pipeline can enhance collaboration and thus improve the process. Then they further mention that the FAIR approach should help drug developers to learn from past efforts, thus reducing redundancy and accelerating the development of new therapies.

Two articles discuss specifically the most commonly used types of therapeutics: small molecules. Southey and Brunavs explore small molecule drug discovery, outlining the various steps and the challenges in the field. They note that despite well-established protocols and novel knowledge, the process still remains very complex. The authors then present emerging technologies aimed at overcoming current limitations, hopefully making the path to new drug approvals more efficient. In a related discussion, Giraud provides a mini-review explaining how high-throughput screening and biophysical methods are used in the early stages of drug discovery. The article highlights the two main strategies used in the field: target-based and phenotypic-based discovery.

Munsier et al. present a compelling and timely review on the transformative potential of AI in the discovery of biologics, with a particular focus on antibodies, a major therapeutic area driving innovation in drug development. Traditionally, antibody discovery relied heavily on animal models and lengthy experimental processes. The authors highlight the advancements in in silico approaches, which are now capable of accelerating antibody design while reducing reliance on animal testing. AI-driven approaches are presented to showcase the shift towards more efficient and de-risked antibody discovery processes, marking the beginning of an exciting new chapter in developing biologic treatments.

Public engagement and participation are crucial components of advancing drug development. Wang et al. present the results of a survey on public awareness and willingness to participate in drug clinical trials (DCTs) in China. Their findings reveal significant gaps in knowledge and highlight the demographic factors influencing participation rates. The authors call for improved public outreach and communication strategies to foster greater understanding and involvement in DCTs. This could have a significant impact on the success of treatments.

The focus of several articles is about the growing concern of using animal models in drug discovery. Marshall and Conlee discuss the limitations of animal testing, noting the high failure rates of drug candidates that appear safe in animals but prove ineffective or toxic in humans. The article suggests to move towards human biology-

based testing methods, which not only promise to be more predictive of human responses but also align with ethical imperatives to reduce animal use. Krebs and Herrmann provide an overview of the international movement towards reducing animal testing in biomedical research. They review new non-animal research approaches that mimic human physiology. Despite of the emergence of these new approaches, the authors acknowledge the persistence of an animal methods bias. They call for a cultural shift in the scientific community, supported by changes in regulatory policies and funding incentives. Hartung offers a balanced perspective on the role of animal models in medical research. This author acknowledges their historical contributions but also highlight their limitations. The article argues for better, more humane alternatives that use novel methods, envisioning a future where drug development is both more effective and ethical.

Together, these articles offer a comprehensive yet accessible overview of the drug discovery process. This Research Topic aims to help the public and patient communities better understand the world of drug discovery, empowering them to engage in the dialogue surrounding drug development. As the field continues to evolve, informed public involvement will be key to shaping a more transparent, efficient, and patient-centered approach to discovering new medicines.

Author contributions

BV: Conceptualization, Funding acquisition, Investigation, Methodology, Software, Visualization, Writing–original draft. J-LP: Writing–review and editing. KT: Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The Research Topic Editors express their gratefulness to all the contributors for submitting their work to this Research Topic, to the Review Editors, and external Reviewers who participated in the review process, and to the Editorial and Production teams of Frontiers for their valuable assistance through the various stages of the publication process.

Conflict of interest

Authors BV and KT were employed by Akttyva Therapeutics, Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers | Frontiers in Drug Discovery

TYPE Mini Review
PUBLISHED 24 May 2023
DOI 10.3389/fddsv.2023.1201419

OPEN ACCESS

EDITED BY
Vasanthanathan Poongavanam,
Uppsala University, Sweden

REVIEWED BY
Duc Duy Vo,
Uppsala University, Sweden
Priyanka Banerjee,
Charité University Medicine Berlin, Germany

*CORRESPONDENCE
Bruno O. Villoutreix,
Bruno.villoutreix@inserm.fr

†PRESENT ADDRESS
Natesh Singh, Evotec SE, Molecular Architects, Integrated Drug Discovery, Campus Curie, 195 Rte d'Espagne, 31100 Toulouse, France

RECEIVED 06 April 2023
ACCEPTED 08 May 2023
PUBLISHED 24 May 2023

CITATION
Singh N, Vayer P, Tanwar S, Poyet J-L, Tsaioun K and Villoutreix BO (2023), Drug discovery and development: introduction to the general public and patient groups. *Front. Drug Discov.* 3:1201419. doi: 10.3389/fddsv.2023.1201419

Drug discovery and development: introduction to the general public and patient groups

Natesh Singh[1,2†], Philippe Vayer[2], Shivalika Tanwar[2], Jean-Luc Poyet[3], Katya Tsaioun[4,5] and Bruno O. Villoutreix[2,5*]

[1]Evotec SE, Molecular Architects, Integrated Drug Discovery, Campus Curie, Toulouse, France, [2]NeuroDiderot Department, Inserm UMR 1141, Robert-Debré Hospital, Université Paris Cité, Paris, France, [3]Onco-Dermatology and Therapies, Inserm UMRS976, Hôpital Saint Louis, France Institut de Recherche Saint Louis, Université Paris Cité, Paris, France, [4]Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States, [5]Akttyva Therapeutics, Inc., Mansfield, MA, United States

Finding new drugs usually consists of five main stages: 1) a pre-discovery stage in which basic research is performed to try to understand the mechanisms leading to diseases and propose possible targets (e.g., proteins); 2) the drug discovery stage, during which scientists search for molecules (two main large families, small molecules and biologics) or other therapeutic strategies that interfere or cure the investigated disease or at least alleviate the symptoms; 3) the preclinical development stage that focuses on clarifying the mode of action of the drug candidates, investigates potential toxicity, validates efficacy on various *in vitro* and *in vivo* models, and starts evaluate formulation; 4) the clinical stage that investigates the drug candidate in humans; 5) the reviewing, approval and post-market monitoring stage during which the drug is approved or not. In practice, finding new treatments is very challenging. Despite advances in the understanding of biological systems and the development of cutting-edge technologies, the process is still long, costly with a high attrition rate. New approaches, such as artificial intelligence and novel *in vitro* technologies, are being used in an attempt to rationalize R&D and bring new drugs to patients faster, but several obstacles remain. Our hope is that one day, it becomes possible to rapidly design inexpensive, more specific, more effective, non-toxic, and personalized drugs. This is a goal towards which all authors of this article have devoted most of their careers.

KEYWORDS

drug discovery, drug development, therapeutic agent, biologics, small molecules, artificial intelligence (AI)

GRAPHICAL ABSTRACT

Introduction

Drug discovery has a long history and dates back to the early days of human civilization. In those ancient times, treatments were often discovered by chance or resulted from observation of nature, typically but not exclusively, using ingredients extracted from plants/animals, and not just used for physical remedy but also for spiritual healing. Modern drug discovery research started to being performed around the early 1900s. Nowadays, the development of a new medicine usually starts when basic research, often performed in academia, identifies a macromolecule (i.e., a molecule with a large molecular weight like genes/proteins), or a dysfunctional signaling pathway or a molecular mechanism apparently linked to a disease condition (pre-discovery stage) (Figure 1; Table 1) (Hefti, 2008; Hughes et al., 2011; Mohs and Greig, 2017; Villoutreix, 2021). In general, at this stage, research teams attempt to identify the so-called therapeutic targets (often a protein) that are linked to the disease state (Gashaw et al., 2012). To be nominated therapeutic target, scientists will also have to find therapeutic agents that modify the function of the perturbed target and restore health or alleviate symptoms. Finding the right target is however extremely challenging. Further, drugs are efficient in humans because of specific actions on the intended therapeutic target but also due to interactions with other, unintended (often unknown) targets! The process continues with the search of therapeutic agents followed by a preclinical phase, during which potential drugs are tested in a battery of animal models, to demonstrate safety and select drug candidates (novel strategies to avoid animal testing are being developed, see below). Clinical studies in humans can then get started to establish safety and efficacy of the drugs in patients with the highest benefit-to-risk ratio (Kandi and Vadakedath, 2023). The studies are then submitted to regulatory agencies, which review the documents and decide about market approval. If the review is positive, the drug can then be released to the market and be administrated to patients. Once a drug has been approved, investigations continue to monitor putative side effects that could be caused, over time, by the new treatment. This last step is often referred to as pharmacovigilance studies (or real-world evidence), generally dubbed "phase 4" clinical trial. The entire drug discovery and development process involves many disciplines, years of efforts and is very expensive. It also implies the generation and use of vast amount of data usually obtained via different types of high-throughput technologies. Many of these experiments and the analysis of the results can be automated via computer-assisted methods to speed-up some steps of the process, gain knowledge and reduce mistakes.

As mentioned above, to act on a disease, the problematic target(s) have to be modulated by a therapeutic agent (or several). There is a wide variety of agents that traditionally fits into two major classes, the so-called "small molecules" (small chemical compounds, some modified short peptides. . .) and the "biologics" (typically macromolecules such as recombinant proteins, antibodies, siRNAs, long peptides, cells, genes . . . and vaccines). There are major differences between biologics and small molecules (Figure 2; Table 2) and we will essentially focus here on small molecules. It is also important to note that gene therapy is different

from the other types of therapeutic agents because it is a technique that modifies a person's genes to treat or cure a disease. In this case, the target is a disease-causing gene which has to be modified with a healthy copy of the gene, or the disease causing gene could be inactivated. Thus, beside technical issues, there are a number of ethical questions surrounding gene therapy and genome editing strategies that are not easy to answer. Further, some therapeutic agents are not acceptable to some parts of the population, as seen during the COVID-19 crisis and vaccine hesitancy. This is often due to misunderstanding of the biological processes and misinformation, resulting in fears, but yet this has to be considered. Also, about 5%–10% of the population are non-responders and have to receive other medications than vaccines. The division into small molecules and biologics is far from being perfect as some therapeutic agents combine a small molecule grafted onto a biologic (e.g., tisotumab vedotin is an antibody-drug conjugate used to treat cervical cancer). Therapeutic agents can be administrated to patients via different routes, called "routes of administration". Small molecules can in general be administrated orally (the most convenient route for patients), while biologics usually need to be injected. The choice of a route of administration is also governed by the patient's condition, for instance, in acute situations in hospitals, drugs are most often given intravenously. Other critical medical interventions that will not be discussed here are surgery, radiotherapy and psychological support.

Drug discovery and development: overview of the process

There are several stages in the drug discovery process that require numerous skills and the use of various advanced technological platforms (often a combination of computational and experimental approaches) to validate targets and search for therapeutic agents. When initial experimental compounds have been sufficiently optimized to be selective, potent and safe in preliminary *in vitro* experiments and animal models, they can be nominated as drug candidates. At this stage, the project focus shifts from drug discovery to drug development to enable human clinical trials. If the therapeutic agent is successful in all three phases of the clinical trials, it goes through regulatory registration and the drug can be marketed (Hefti, 2008; Hughes et al., 2011; Mohs and Greig, 2017).

Now, we will take a closer look at the process with the discovery of small molecules as an example. The process usually begins by focusing on a disease and the search of possible targets, often proteins, that can be modulated by small compounds (Hughes et al., 2011) (Figure 1). These compounds are expected to interfere or prevent the disease or at least limit the development of symptoms. These targets can be identified using cellular assays, genomic studies, proteomic studies, among many others. Then, thousands (to millions or even billions when using computer-aided drug design approaches prior to vitro assays) of small molecules have to be tested in various types of assays and a few promising molecules are then evaluated in animal models (and in alternative *in vitro* models) of human diseases. It is worth mentioning here that animal models can be misleading (e.g., a drug found toxic in animal models may not be toxic to humans or the opposite) (Pognan et al., 2023). At the same time, absorption, distribution and elimination studies (ADME) are conducted. After years of research, a few compounds will hopefully be safe and effective enough to take forward to trials in patients. The different stages can have different names in the scientific literature, often they are referred to as: the pre-discovery and basic research stage (around 5–6 years) in which

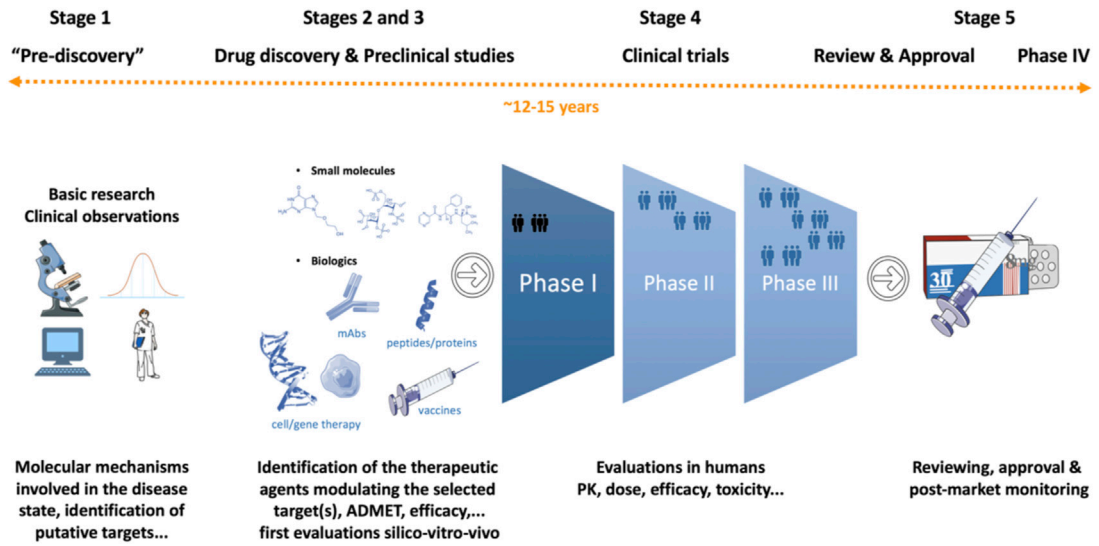


FIGURE 1
Drug discovery and development. The main stages are represented in a highly simplified manner. The process varies depending on the molecular mechanisms expected to be linked to the disease and the type of therapeutic agents that needs to be developed. The approximate cost is around US $2.8 billion and the time needed to complete the entire process is around 12–15 years.

TABLE 1 Glossary.

Targets or drug targets	A therapeutic target is in general a macromolecule (typically a protein), which may cause or be associated with a particular disease, that can be modulated by a therapeutic agent in a measurable way
Genes	Genes are macromolecules, made up of deoxynucleic acid (DNA) bases. In humans, genes vary in size from a few hundred DNA bases to more than 2 million bases and about 20,000 to 25,000 genes have been identified
Proteomics	Proteomics is the process of separation and characterization of all the proteins of a biological system. Target identification with proteomics can be performed by comparing the protein expression levels in normal and diseased tissues
Biologics	Diverse molecules from biological origins that include, nucleic-acids, various (recombinant) proteins, antibodies . . . some types of peptides. Biologics typically have a high molecular weight
Proteins	Proteins are large biomolecules (up to thousands of atoms) that comprise one or more chains of amino acids. Proteins perform a vast array of functions within organisms, often through interactions with other macromolecules. Proteins are products of genes; they generally fold into a specific three-dimensional (3D) structure that determines their activities. There are different types of proteins with different functions and locations in the body
Peptides	Peptides are short chains of amino acids. They can be modified to include non-natural amino acids (up to hundreds of atoms). Some peptides belong to the category of glycopeptide or lipopeptide, among others. Very short and modified peptides can behave like small molecules while longer peptides (e.g., insulin which is used to manage diabetes) fit in the category of biologics. At present, there are very few approved peptides that can be given by oral route but important work is ongoing in this field to enable oral delivery
Small molecules	Any organic compound with around 80–100 atoms. Most are made synthetically (aspirin), while others can be derived from natural product (e.g., morphine, which is used to relieve moderate to severe pain)
Drug candidates	A molecule suitable for clinical testing. The molecule is expected to bind selectively to a target involved in the disease process, to elicit the desired functional responses *in vivo*, often in animal models of the human disease, to have adequate bioavailability and bio-distribution within the body to reach the intended target and to pass formal toxicity evaluation in various *in vitro* and animal models
Bioinformatics, Chemoinformatics, Artificial Intelligence (AI)	Bioinformatics is a branch of molecular biology that involves extensive analysis of biological data using computers
	Chemoinformatics is a field that attempts to solve chemical problems on the computer, including chemical structure coding, properties modeling and development of databases
	Artificial Intelligence (AI), as used today (the so-called weak-AI), combines computer sciences and mathematics and uses (large) datasets to enable problem-solving. It includes various learning approaches, natural language processing, knowledge representation and reasoning, among others
ADMET	Absorption, distribution, metabolism, excretion, and toxicity. A drug has to reach the intended target(s), elicit the desired functional response with no or limited toxicity and be eliminated from the body (typically via the liver or kidneys). These are critical properties of the drug candidates that are commonly investigated at various stages of the process
PB/PK	Physiologically-based pharmacokinetic modeling and simulation (PB/PK) is a computer modeling approach that incorporates blood flow and tissue composition of organs to define the pharmacokinetics (PK) of drug candidates
PK	Pharmacokinetics (PK) is the time-concentration profile of drugs administered *in vivo* to living organisms. PK parameters include clearance, volume of distribution, peak plasma concentration . . . PK is sometimes described as "what the body does to a drug"
PD	Pharmacodynamics (PD) refers to the relationship between drug concentration at the site of action and the resulting effect, including the time course and intensity of therapeutic and adverse effects. PD parameters include minimum effective concentration, maximum safe concentration, onset of action, therapeutic range and therapeutic index. PD describes how biological processes in the body respond to or are impacted by a drug
PK/PD	Relationship of the drug effect (pharmacodynamics) to the drug concentrations in the body compartments (e.g., blood, organs) as a function of time after drug administration
Off-target activity	Action of a drug on targets other than the intended biological target. Such events commonly contribute to adverse effects or toxicity, however, in some cases, off-target activity can be valuable for therapeutic purposes
On-target toxicity	A drug is usually designed to interact with its intended target. In some situations, the drug induces exaggerated and adverse pharmacological effects at the target of interest. This is commonly referred in the literature to as on target toxicity
Adverse events	Unintended pharmacological effects that occur when a medication is administered correctly. There are different types of reactions (mild, moderate or severe) that can be dose-dependent or not
Side effects	Secondary unwanted effects that occurs due to the drug therapy. Side effects are usually known and patients are informed about such effects

(Continued on following page)

TABLE 1 (*Continued*) Glossary.

Preclinical development	Preclinical studies are a stage of research that precedes clinical trials (testing in humans). The therapeutic agents are tested in animal models of human diseases or in systems that simulate human diseases. The main goals are to determine a starting, safe dose for first-in-human study and assess potential toxicity. Research into early formulations (e.g., tablet, capsule, intramuscular injection, intravenous, sublingual. . .) is also performed
Clinical trials	Research studies performed in humans aiming at evaluating the efficacy (does the drug cure or slow the progression of a disease?) and safety (does the drug cause undesired effects, or toxicity?) of drug candidates. Human clinical research is tightly regulated by authorities around the world (e.g., US Food and drug Administration or US FDA and European Medicines Agency or EMA). Pharmaceutical companies and other organizations developing drugs have to conduct extensive preclinical evaluations, propose the design of clinical trials and formally submit these data and a clinical plan to regulatory authorities. If regulatory authorities approve the proposed strategy, Phase I (first-in-humans) clinical trial can start. Each study has its own pre-defined rules about which patients can or cannot participate, which is called eligibility
Phase I	Aka "first-in-humans" trial. Test on 20–80 healthy volunteers to assess the safety and pharmacokinetics, absorption, metabolism, and elimination, actions on the body, as well as possible side effects, formulation, and dose. In some cases, a placebo can be used. For some drugs, a phase 0 can be sometime performed before phase I to evaluate some properties of the drug on few patients or on healthy individuals
Phase II	Assesses drug safety and efficacy on about 100–500 patients (suffering from a specific disease), some of which may receive a placebo or an approved drug for that disease, called "standard of care". Analysis of optimal dose is performed while adverse events and risks are recorded
Phase III	Phase III enrolls numerous patients (e.g., 1,000–5,000), enabling medication labeling and instructions for proper drug use. Efficacy, dose, and toxicity are observed and adjustments to the final medication label are being made based on such information
Phase IV or pharmacovigilance or "real world evidence"	Following drug approval and manufacturing, regulatory agencies require companies to monitor the safety of the approved drug. Drug makers, health professionals, hospitals and patients report adverse events occurring when taking the approved drug
Therapeutic window	The dosage (a range of concentrations) of a drug that provides efficacious therapy and is safe (without serious side effects)
Drug formulation	The process in which the therapeutic agent is combined with different substances to produce a final medicinal product (e.g., a tablet, infusion solution, etc.). Formulation optimization is ongoing throughout pre-clinical and clinical stages. It ensures drugs are absorbed into the body and delivered to the proper organ at the right time and in the right amount
Patent	A patent is an exclusive right granted by the governments for an invention. Patents give an inventor (academic group or a private company) the exclusive right to prevent others from making, using, selling, or importing a product or process based on the patented invention without the inventor's prior permission, such as through a patent license. Patent protection is limited to the country or region where it was issued and limited in time, typically 20 years from the date of patent application filing. Pharmaceutical patents can be extended for new indications or novel formulations

targets and modifying small molecules are searched *in silico* (i.e., using a computer), *in vitro* (i.e., in the test tube), *ex vivo* (e.g., on tissues or organs) and *in vivo* using simple animal models (i.e., in a living organism, typically rats or mice) and a preclinical stage (2–3 years) during which the best small molecules are selected using various *in silico*, *in vitro* and *in vivo* experiments. In general, after all these steps, only a few compounds progress to the next stage. Toxicity is investigated further on at least two animal models [one rodent (e.g., rat) and one non-rodent (e.g., dogs, mini-pigs)] often using different administration routes before they become nominated clinical candidates and get a regulatory permission to proceed to human clinical trials. Prior to starting clinical trials, a so-called Investigational New Drug (IND) application is submitted to regulatory agencies (e.g., the Food and Drug Administration in the United States). Such documents, at least up to now (see below), usually include animal efficacy data and toxicity (Good Laboratory Practice (GLP)-compliant animal toxicology data are performed supporting the dose, dosing schedule, administration), manufacturing information, clinical protocols (e.g., patient population, number of patients, duration of the study) proposed for the clinical trials and information about the investigators of the study.

If the IND is approved, then clinical trials start (4–7 years) (Kandi and Vadakedath, 2023). In some specific cases such as cancer, a so-called phase 0 may get started, which involves the use of very small doses of the new drug in a limited number of people and sometimes in patients. This is an exploratory study with the goal of quickly exploring if and how the drug may work. In Phase I, the safety, and tolerability of the therapeutic agent (usually a single dose at first and then short-term multi-dose studies) is tested in a small number of healthy individuals (e.g., 20–80 people). Other parameters are investigated including the dose. Phase II typically involves 100–500 patients and the study can take place in several hospitals located in different countries. The study is designed to determine whether or not the therapeutic agent provides the desired therapeutic effect. Safety studies continue through the phase II trials. In the first part of phase II, referred to as phase IIa, the goal is to further refine the dose required to provide the desired therapeutic impact or monitored endpoints for the clinical candidate. Once the proper dose levels are determined, phase IIb studies can be initiated. The goal of the phase IIb is to determine the overall efficacy of the

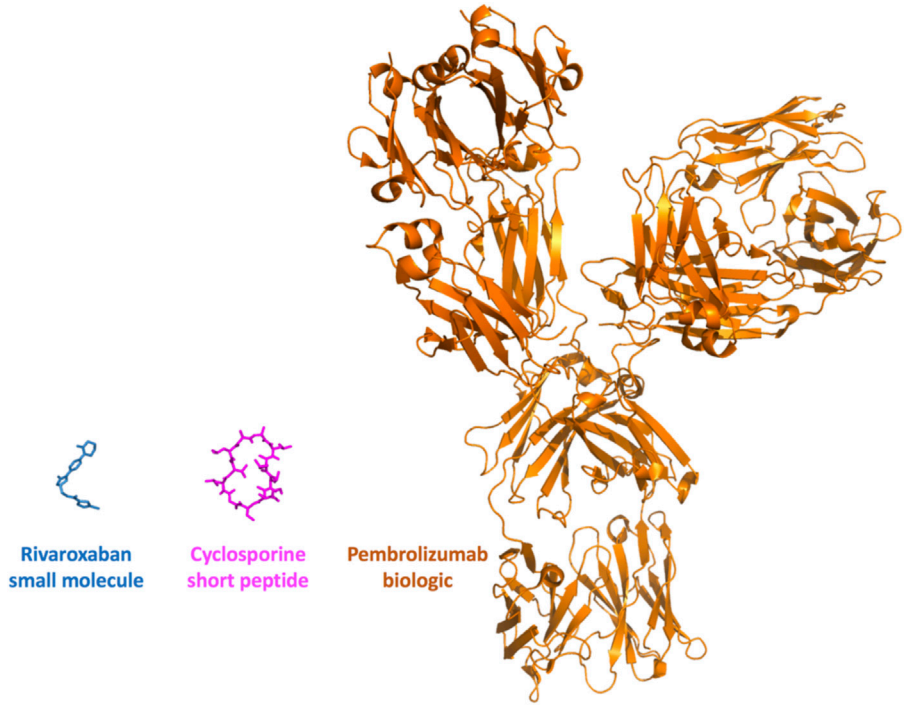


FIGURE 2
Small molecules, peptides and biologics. The properties and sizes of the therapeutic agents vary greatly. Three molecules are presented at the same scale, these involve rivaroxaban, a small chemical molecule used to treat thrombosis and pulmonary embolism, cyclosporine, a short immunosuppressive cyclic peptide (11 amino-acids, a biologic that still resembles to a certain extent to a small molecule) used to treat post-transplant organ rejection and a biologic, pembrolizumab (antibody, over 1300 amino-acids), used to treat various types of cancer.

TABLE 2 General characteristics of small molecule drugs and biologics.

Property	Small molecules	Biologics
Size	Low molecular weight (around 80–100 atoms)	High molecular weight (hundreds to several thousand atoms)
Stability	Usually stable at room temperature	Usually unstable at room temperature (need to be stored in refrigerators and freezers)
Three-dimensional structure	Relatively simple	Complex
Route of administration	Often oral	Typically, via injection or infusion
Cell membrane permeability	High	Low
Tissue distribution	Easily distributed via circulation	Limited distribution via circulation and lymphatics
Immunogenicity	Limited	Possible
Treatment cost	Relatively low	Relatively high
Attrition (business aspect)	Relatively high	Relatively low
Competition (business aspect)	Very high (after patent expiration or before)	Less severe competition (after patent expiration or before)

candidate drugs in a limited population of subjects. Numerous drug candidates fail in phase II due to safety issues or lack of efficacy. In phase III, the efficacy of the drug candidate is evaluated in a larger patient population. These studies are typically randomized and involve 1,000–5,000 patients at multiple clinical trial centers and are designed to determine the efficacy of the candidate compound relative to the current standard of care or a placebo, possible interactions with other medications and re-assess different doses (optimal dose is important for medication effectiveness). When neither the clinicians nor the patients know which of the treatments the patient is getting, the study is said to be double-blind. The cost and time associated with this phase can vary dramatically depending on the disease and the clinical endpoint under investigation. Phase III clinical trials are the most expensive part of drug discovery and

development as it has a complex design and requires a large number of patients. Last but not least, formulation and stability studies are performed during the development stage to characterize the impurities present (either in batches or during storage conditions worldwide), and to determine the best formulation. Upon completion of the phase III trial, a New Drug Application (NDA) is submitted to the regulatory agencies to demonstrate drug safety and efficacy. Regulatory reviews can lead to requests for additional information, or even additional clinical trials to further establish either safety or efficacy. Ideally, these reviews lead to regulatory approval, including labelling requirements, and approval to market (review and approval ~1–2 years). For approval, the drug must have adequate pharmaceutical quality, therapeutic effectiveness, and safety. It has to have a favorable "risk-benefit ratio". Drugs offering important advances in treatment of a condition are given priority. Approval of regulatory bodies does not, however, signal the end of clinical trials. In many cases, regulatory agencies will require additional follow-up studies, often referred to as phase IV or post-marketing surveillance ("real-world evidence" trials) with infinite duration. In general, these studies are designed to detect rare adverse effects across a much larger population of patients or long-term adverse effects. The impact of phase IV studies can include alterations to labelling based on safety observations, contraindications for use of the new drug in combination with other medications, or even the withdrawal of marketing approval if the findings are severe enough.

Drug repurposing: challenges and opportunities

Drug repurposing or repositioning aims to take a drug (approved or in advanced clinical stages or even a drug that has been withdrawn from the market, most of the time it involves small molecules but biologics like antibodies are also explored), thus a molecule that has undergone extensive safety and efficacy testing, and use it for an additional or unrelated indication (van den Berg et al., 2021; Roessler et al., 2021; Schipper et al., 2022). In some situations, even a withdrawn drug can be repurposed like thalidomide, originally intended as a sedative and then used for treating a wide range of other conditions, including morning sickness in pregnant women. Thalidomide was then withdrawn due to causing birth defects but then was approved to treat leprosy (in 1998) and multiple myeloma (in 2006) (Begley et al., 2021). Drug repurposing approach can be very valuable in most cases including emergency situation like a pandemic, for rare and neglected diseases [for which specific drug developments are in general missing in pharmaceutical companies (Scherman and Fetro, 2020; Roessler et al., 2021)]. This strategy is promoted as a cost- and time-effective approach for providing novel medicines. It is often claimed that repurposing drugs can be faster, more economical, less risky, and carry higher success rates as compared to traditional approaches, primarily because it is in theory possible to bypass early stages of development such as establishing drug safety. Other benefits that come with this approach include readily available products and manufacturing supply chains. Drug repurposing can be very profitable as in the case of fenfluramine (in 2022, acquisition of Zogenix by UCB Pharma for about US$ 1.9 billion, https://www.ucb.com/stories-media/Press-Releases/article/UCB-Completes-Acquisition-of-Zogenix-Inc), a drug initially developed for weight loss, withdrawn and now used in several countries for the treatment of some forms of epilepsy (Odi et al., 2021). Yet, despite advantages, drug repurposing suffers from several issues. One problem is that there are no possibilities for optimization of the therapeutic molecule without losing the repurposing potential because any small change in the structure of the therapeutic agent means a new full manufacture process validation and preclinical safety development. Identifying an optimal dosage and formulation for the new disease indication can also be time consuming and requires novel investigations while side effects can indeed arise due to the new indication or in cases doses need to be changed. Also, assessing the patent status of the drug to repurpose requires very specific skills. The molecules that are investigated for repurposing are either patented or off-patent, and in some cases the intellectual property protection for the new indication may not be strong enough to engage in such project. Overall, while drug repurposing is intuitively attractive as it offers shorter routes to the clinic, challenges throughout the entire process are usually substantial. Investigating molecular mechanisms behind repurposing can however be very valuable as it can help identifying novel targets and as the repurposed drugs could be considered as starting point for the development of novel compounds (e.g., lenalidomide and pomalidomide are superior molecules derived from thalidomide) and as such emerge as breakthrough innovation in a reduced amount of time and still reduced cost compared to starting from scratch. It could also be of interest to combine several approved drugs (in some cases with a newer drug) to increase effectiveness.

Artificial intelligence: trust, but verify

Providing efficient and safe drug to patients is a long and complex process. The amount of data generated during this process or that can be collected from various sources is massive. It is thus necessary to integrate as much as possible quality data so as to be able to make decision in real time. Artificial Intelligence (AI or indeed, most of the time, machine learning) can definitely contribute here as it involves the use of powerful computers and efficient program algorithms to integrate large volume of data to train expert systems to perform a complex task (Brogi and Calderone, 2021; Ruffolo et al., 2021; Jayatunga et al., 2022; Sadybekov and Katritch, 2023). During the early discovery phases, AI is used to rationalize processes, and to assist in project management (e.g., definition of a target product profile that allows to locate each compound with regard to the expected final drug specifications in a complex multi-dimensional space), to summarize information, to understand better complex biological systems (e.g., using for instance system biology and chemogenomics approaches), or to propose original compounds or biologics (e.g., small molecules, peptides) generated by the machine under various types of constraints (e.g., ADMET constraints or affinity to the target) (Lambert, 2010; Gupta et al., 2021; Paul et al., 2021; Kontoyianni, 2022; Vijayan, et al., 2022). Most of the well-known success stories of AI have been in image recognition (e.g., in the early days, the approach was trained to for instance recognize cat and dog images,

but today the method can be used to analyze biopsies or guide surgery) while also advertised in reducing time to reach phase I clinical trial. In the latter case, one can site the story of compound DSP-1181, developed by Exscientia and Sumitomo Dainippon Pharma, intended to treat obsessive compulsive disorder where time from first screening to the development stage was 4 time faster than using a conventional approach (although, unfortunately, the molecule failed in phase I, for numerous reasons including a difficult target while it was also observed that the molecules generated by AI were not novel) (Santa Maria Jr et al., 2023) (https://www.science.org/content/blog-post/another-ai-generated-drug; https://www.cas.org/resources/cas-insights/drug-discovery/ai-designed-drug-candidates). Similar observations have been posted by hundreds of financial analysts and research scientists about results obtained by other AI companies. In other words, the AI predictions are not perfect and indeed cannot be perfect at present (Bajorath, 2021; Bender and Cortés-Ciriano, 2021). This situation reflects the dependency of AI/machine learning to quality, size and diversity of the data used to train the mathematical models. There are millions of compounds (most will never be a drug) tested via standard experiments available in various databases, but there are only a few thousand approved in humans that are annotated on which to learn from, highlighting the so-called data gap (i.e., there are billions of pictures of dogs and cats to learn from, but a limited amount of quality data is available in the field of drug discovery despite the use of numerous the high-throughput approaches). The predictions can thus be misleading, because we do not have enough quality data as input and/or because we do not understand enough the complexity of the biological systems (Moingeon et al., 2022). During the drug development phases, in human, AI is associated to data-mining to for instance model some properties (e.g., PB/PK, PK/PD or population-based simulations and analysis, prediction of drug-drug interactions . . .). At this stage, these computer approaches can also be used to select the most informative population profile to be included in clinical trials or to explain the variability of effects, or provide « virtual » patients or populations, and applied to, for example, pediatric formulation using as input data collected on adults (Lang et al., 2021). Related to these, the concept of digital twins (which has been around for a while in other areas of research), now starts to be explored in the context of drug discovery and development. The overall idea would be to collect data about a particular disease, how it progresses, about the current treatments, about specific patients, and about a whole population, encapsulate all these data into a computer model so as to create a digital representation of a biological system or of a person and be able to simulate, for example, what might happen if one were to take a novel drug. While the concept is attractive, there are still major challenges and obstacles ahead but progresses are being made (An and Cockrell, 2022). Overall, AI, in the field of drug discovery and development, is still in the infancy stage and it will take time to fully integrate the technology into the R&D process (Hillisch et al., 2015). AI-discovered drugs do not guarantee success in clinical trials. The understanding of the data used as well as the critical mind of the scientists are key points that lead to the success or failure of AI-assisted drug research and development processes. The technology, in some circumstances, can make the process faster and more cost-effective, however, AI needs quality data to produce meaningful results and still today requires significant experimental validation. As such, it is important to trust AI, but verify the predictions (Schneider et al., 2020; Bajorath, 2021).

Rising cost: from drug discovery to new treatments

Analyses across all therapeutic areas indicate that the development of a new medicine, from target identification through approval for marketing, takes around 12–15 years and often longer. The cost to develop a new drug is very high, in part because failure is endemic in drug discovery, and success is rare. While various numbers have been reported, the latest formal assessment is around US $2.8 billion (DiMasi, 2020). There are many factors that contribute to this situation: the lack of understanding of what causes the disease can lead to the selection of the wrong therapeutic target; the impossibility of reaching the target with a sufficient concentration of drug *in vivo* without leading to adverse effects; no formulation compatible with the use of the drug in human; the therapeutic agent developed during years is found in phase III to have very low efficacy; the therapeutic agents or a metabolite (e.g., case of a small molecule) can interacts, specifically or not, with other drugs or with hundreds of molecules in the body, these interactions are usually not known in details and can lead to numerous adverse effects; animal experiments that are used to evaluate potency, selectivity, and toxicity during the different stages of the process can be highly misleading; stricter regulatory guidelines; duration of patents; the identified therapeutic molecule can be toxic in some patients but this could not be anticipated during the clinical trials due to the relatively small number of patients treated. Next and related to the cost of R&D, comes the cost of the treatments. Although there is a very complex protocol to determine the price tag of a drug (it varies from country to country, it can consider the insurance system, whether the drug is curative and represents a major advance to both patients and the health system or it has a minor effect on the disease), but in the end, biologics are generally much more expensive than small molecules, in part due to the complex manufacturing process. Studies suggest that on average, the daily dose of biologics costs 22 times more than a small molecule (Makurvet, 2021). It is important to keep in mind that the healthcare systems, in many countries, are about to collapse and that about half of the world population cannot get access to basic treatments (Ozawa et al., 2019). Biologics have been here for several decades already and are becoming increasingly important in several therapeutic areas. For example, cancer checkpoint inhibitors (e.g., the antibody ipilimumab and about 4–5 others at the time of writing) have received considerable and broad interest because of their ability to generate responses in many hitherto intractable malignant tumors. Yet, many recent studies suggest that such molecules lead to responses in less than 10%–15% of patients with cancer. Clearly, such molecules offer hope but also rise many questions (Fojo et al., 2014; Kantarjian and Rajkumar, 2015). That is, in some cases, biologics are real innovative breakthroughs, but in other situations, the strategy is pursued only for commercial reasons and alternative molecules such as small molecules are not even considered. These questions are, in theory, investigated by regulatory agencies [The United States Food and Drug

Administration (FDA), European Medicines Agency (EMA), Pharmaceuticals and Medical Devices Agency (PMDA)] so as to try to avoid speculative drugs but more transparent processes would certainly be beneficial to patients and the general population. Although finding new treatments is very difficult, it is a profitable market, with global drug sales expected to grow to US$ 1.9 trillion by 2027 (Mullard, 2023).

Innovation in regulatory science and methodologies

It is important to note that, in step with the scientific progress in human tissue models research in the past decades, in the US, new medicines may not have to be tested in animals, according to legislation signed by the President Joe Biden in late December 2022 ("Text–S.5002–117th Congress (2021–2022): FDA Modernization Act 2.0." 29 September 2022. https://www.congress.gov/bill/117th-congress/senate-bill/5002/text).
Accordingly, US FDA is already accepting data from *in vitro* studies as part of the formal submission to the Agency (Wadman, 2023). Additionally, at the same time, following the leadership of some academic researchers (e.g., Guzelian et al., 2005; Hoffmann and Hartung, 2006), major European and US agencies started using evidence-based methodologies, such as systematic reviews and systematic maps, in toxicological assessment. These methodologies were developed and tested over the last 40 years in clinical research, spearheaded by Cochrane Collaboration (www.Cochrane.org) to compare the effectiveness of treatments, and have been applied to toxicological assessment of data-rich substances by the European Food Safety Authority (EFSA, 2017) and US Environmental Protection Agency's (US EPA, https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=356370). While some of the aspects of these methodologies are not entirely applicable to drug-discovery because of the proprietary nature of the work, the main principles of evidence-based approaches, which encourage pre-publishing the methodologies before the research is conducted, comprehensiveness and transparency in data selection, minimization of bias (or systematic error), are in line with basic principles of the scientific method, and are applicable to drug discovery. Programs and drug candidates are all too frequently selected based on a biased opinion of a few scientists who are bound by similar training, scientific methodologies and beliefs. Opening-up drug discovery to scrutiny by other scientists with different training and opinions may lead to more failures in the earlier discovery stage, but less failures in the clinic, resulting in enhanced efficiency and more successes, benefiting the patients who need new treatments, first and foremost.

Concluding remarks

Drug discovery and development is a long and difficult endeavor; all novel ideas and strategies that can improve the process are valuable to explore. It is interesting to note that despite the steady increase in research and development expenditure, and major scientific advances in proteomics and genomics, the discovery of new drugs either seems to be drying-up some years or to remain essentially stable (Laermann-Nguyen and Backfisch, 2021). This situation has various origins (e.g., many diseases with no treatment are extremely difficult to study), while, certainly, industry scientists would benefit from greater exposure to new ideas from public research and public researchers would benefit from the private sector to move beyond exploration of molecular mechanisms towards the end goal of efficient development of candidate therapeutic agents. Along these lines, some countries like the United States and United Kingdom have been working extensively at improving academic drug discovery (e.g., all the skills and platforms connected via open research networks with rational protocols) but in the others, the process is fragmented (no coordination, no intent, duplication of efforts and inefficient investments ...) and, thus, not capable of producing desired results compared to the time, energy and money spent. A first step could be to develop strong academic drug discovery networks in countries where this type of activity is not coordinated or not considered. Strong collaborations between the private sector, academia and not-for-profit institutions are clearly of major importance and have led to some successes in the past but such partnerships can be difficult to maintain over a long period of time (Yildirim et al., 2016; Takebe et al., 2018). The rationale being that open interconnections between the different scientific disciplines involved in drug research allow a "cross-fertilization", each of them benefiting from the advances of the other fields. Obviously, such collaborations tend to be easier when academic and private research teams are located on the same campus, with possibilities of sharing ideas or technologies. Other types of collaboration imply building consortia, often for around 4–5 years, with research teams located in different cities or countries (unfortunately, most of the time, when the consortia have been built, they function as closed systems not allowing new scientists or novel research teams to join). Therefore, novel strategies need to be pursued, and among the novel public-private models that are being investigated, open science partnerships, could be of interest, if correctly implemented (e.g., the system must be open to all interested scientists, teams and relevant disciplines) (Gold and Edwards, 2022). Open science projects (Chodera et al., 2020), like the consortia models discussed above, are built on the differential expertise of the various partners, with generally academic and governmental partners taking on a larger role in the earlier stages and big pharmas leading in the later stages (e.g., advanced preclinical investigations, product development, manufacturing, and distribution). But in open science projects, results, publications, data, tools, and materials are open without regard for intellectual property. At some points, the various partners are free to use the results and develop their own proprietary products if deemed appropriate.

Next, novel technologies including AI could be a game changer in the years to come, even more so once we get past the hype stage. Novel approaches to replace animal models by more efficient, ethical, human-biology-based *in vitro* approaches could also play a significant role this next decade. Indeed, new tools and understanding, in, for instance, the area of investigative toxicology, are continually being implemented to reduce safety-related attrition in drug development (Aleo et al., 2020). Combining all these strategies, methods and know-how should definitively facilitate the design of more specific, effective, non-toxic, and

patient-tailored drugs, thereby, providing a more optimistic outlook to the field. As a last note, we encourage the general public and patients to become more curious about the process of finding novel therapies, from the pre-discovery to the post-marketing stages. Further, crowd-funded citizen science initiatives are emerging in various areas of drug discovery and development (e.g.,https://www.clinicaltrialsarena.com/news/citizen-science-as-an-open-trials-tool-for-post-marketing-and-drug-repurposing-5909331-2/; see also the CTSA program at NIH), these projects are definitively valuable to the field.

Author contributions

NS, PV, and BV conceptualized the topic and drafted the first version of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Support from the INSERM institute is appreciated. This article is based upon work supported by the National Science Foundation under Grant No. 2136307.

Acknowledgments

We thank Stephane Auvin, Epileptologist and Child Neurologist, Head of the Pediatric Neurology Department, Robert Debré Hospital, for interesting discussions about the use of fenfluramine for the treatment of some forms of epilepsy.

Conflict of interest

Authors KT and BV were employed by the company Akttyva Therapeutics, Inc. Author NS is employed by Evotech Se.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Aleo, M. D., Shah, F., Allen, S., Barton, H. A., Costales, C., Lazzaro, S., et al. (2020). Moving beyond binary predictions of human drug-induced liver injury (DILI) toward contrasting relative risk potential. *Chem. Res. Toxicol.* 33, 223–238. doi:10.1021/acs.chemrestox.9b00262

An, G., and Cockrell, C. (2022). Drug development digital twins for drug discovery, testing and repurposing: A schema for requirements and development. *Front. Syst. Biol.* 2, 928387. doi:10.3389/fsysb.2022.928387

Bajorath, J. (2021). State-of-the-art of artificial intelligence in medicinal chemistry. *Future Sci. oa.* 7, FSO702. doi:10.2144/fsoa-2021-0030

Begley, C. G., Ashton, M., Baell, J., Bettess, M., Brown, M. P., Carter, B., et al. (2021). Drug repurposing: Misconceptions, challenges, and opportunities for academic researchers. *Sci. Transl. Med.* 13, eabd5524. doi:10.1126/scitranslmed.abd5524

Bender, A., and Cortés-Ciriano, I. (2021). Artificial intelligence in drug discovery: What is realistic, what are illusions? *Drug Discov. Today.* 26, 511–524. doi:10.1016/j.drudis.2020.12.009

Brogi, S., and Calderone, V. (2021). Artificial intelligence in translational medicine. *Int. J. Transl. Med.* 1, 223–285. doi:10.3390/ijtm1030016

Chodera, J., Lee, A. A., London, N., and von Delft, F. (2020). Crowdsourcing drug discovery for pandemics. *Nat. Chem.* 12, 581. doi:10.1038/s41557-020-0496-2

DiMasi, J. A. (2020). Research and development costs of new drugs. *JAMA* 324, 517. doi:10.1001/jama.2020.8648

EFSA (2017). Protocol for a systematic review on health outcomes related to the age of introduction of complementary food for the scientific assessment of the appropriate age of introduction of complementary feeding into an infant's diet. *EFSA J.* 15, e04969. doi:10.2903/j.efsa.2017.4969

Fojo, T., Mailankody, S., and Lo, A. (2014). Unintended consequences of expensive cancer therapeutics—the pursuit of marginal indications and a me-too mentality that stifles innovation and creativity: The john conley lecture. *JAMA Otolaryngol. Head. Neck Surg.* 140, 1225–1236. doi:10.1001/jamaoto.2014.1570

Gashaw, I., Ellinghaus, P., Sommer, A., and Asadullah, K. (2012). What makes a good drug target? *Drug Discov. Today* Suppl, S24–S30. doi:10.1016/j.drudis.2011.12.008

Gold, E. R., and Edwards, A. M. (2022). Overcoming market failures in pandemic drug discovery through open science: A Canadian solution. *Front. Drug. Discov.* 2. doi:10.3389/fddsv.2022.898654

Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R., and Kumar, P. (2021). Artificial intelligence to deep learning: Machine intelligence approach for drug discovery. *Mol. Divers.* 25, 1315–1360. doi:10.1007/s11030-021-10217-3

Guzelian, P. S., Victoroff, M. S., Halmes, N. C., James, R. C., and Guzelian, C. P. (2005). Evidence-based toxicology: A comprehensive framework for causation. *Hum. Exp. Toxicol.* 24, 161–201. doi:10.1191/0960327105ht517oa

Hefti, F. F. (2008). Requirements for a lead compound to become a clinical candidate. *BMC Neurosci.* 9, S7. doi:10.1186/1471-2202-9-S3-S7

Hillisch, A., Heinrich, N., and Wild, H. (2015). Computational chemistry in the pharmaceutical industry: From childhood to adolescence. *ChemMedChem* 10, 1958–1962. doi:10.1002/cmdc.201500346

Hoffmann, S., and Hartung, T. (2006). Toward an evidence-based toxicology. *Hum. Exp. Toxicol.* 25, 497–513. doi:10.1191/0960327106het648oa

Hughes, J. P., Rees, S., Kalindjian, S. B., and Philpott, K. L. (2011). Principles of early drug discovery. *Br. J. Pharmacol.* 162, 1239–1249. doi:10.1111/j.1476-5381.2010.01127.x

Jayatunga, M. K. P., Xie, W., Ruder, L., Schulze, U., and Meier, C. (2022). AI in small-molecule drug discovery: A coming wave? *Nat. Rev. Drug Discov.* 21, 175–176. doi:10.1038/d41573-022-00025-1

Kandi, V., and Vadakedath, S. (2023). Clinical trials and clinical research: A comprehensive review. *Cureus* 15, e35077. doi:10.7759/cureus.35077

Kantarjian, H., and Rajkumar, S. V. (2015). Why are cancer drugs so expensive in the United States, and what are the solutions? *Mayo Clin. Proc.* 90, 500–504. doi:10.1016/j.mayocp.2015.01.014

Kontoyianni, M. (2022). Library size in virtual screening: Is it truly a number's game? *Expert Opin. Drug Discov.* 17, 1177–1179. doi:10.1080/17460441.2022.2130244

Laermann-Nguyen, U., and Backfisch, M. (2021). Innovation crisis in the pharmaceutical industry? A survey. *SN Bus. Econ.* 1, 164. doi:10.1007/s43546-021-00163-5

Lambert, W. (2010). Considerations in developing a target product profile for parenteral pharmaceutical products. *AAPS Pharm. Sci. Tech.* 11, 1476–1481. doi:10.1208/s12249-010-9521-x

Lang, J., Vincent, L., Chenel, M., Ogungbenro, K., and Galetin, A. (2021). Impact of hepatic CYP3A4 ontogeny functions on drug-drug interaction risk in pediatric physiologically-based pharmacokinetic/pharmacodynamic modeling: Critical literature review and ivabradine case study. *Clin. Pharmacol. Ther.* 109, 1618–1630. doi:10.1002/cpt.2134

Makurvet, F. D. (2021). Biologics vs. small molecules: Drug costs and patient access. *Med. Drug Discov.* 9, 100075. doi:10.1016/j.medidd.2020.100075

Mohs, R. C., and Greig, N. H. (2017). Drug discovery and development: Role of basic biological research. *Alzheimers Dement. (NY)* 3, 651–657. doi:10.1016/j.trci.2017.10.005

Moingeon, P., Kuenemann, M., and Guedj, M. (2022). Artificial intelligence-enhanced drug design and development: Toward a computational precision medicine. *Drug Discov. Today.* 27, 215–222. doi:10.1016/j.drudis.2021.09.006

Mullard, A. (2023). Drug sales to reach $1.9 trillion within 5 years? *Nat. Rev. Drug. Discov.* 22, 172. doi:10.1038/d41573-023-00026-8

Odi, R., Invernizzi, R. W., Gallily, T., Bialer, M., and Perucca, E. (2021). Fenfluramine repurposing from weight loss to epilepsy: What we do and do not know. *Pharmacol. Ther.* 226, 107866. doi:10.1016/j.pharmthera.2021.107866

Ozawa, S., Shankar, R., Leopold, C., and Orubu, S. (2019). Access to medicines through health systems in low- and middle-income countries. *Health Policy Plan.* 34, iii1–iii3. doi:10.1093/heapol/czz119

Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., and Tekade, R. (2021). Artificial intelligence in drug discovery and development. *Drug Discov. Today.* 26, 80–93. doi:10.1016/j.drudis.2020.10.010

Pognan, F., Beilmann, M., Boonen, H. C. M., Czich, A., Dear, G., Hewitt, P., et al. (2023). The evolving role of investigative toxicology in the pharmaceutical industry. *Nat. Rev. Drug. Discov.* 1, 317–335. doi:10.1038/s41573-022-00633-x

Roessler, H. I., Knoers, N. V. A. M., van Haelst, M. M., and van Haaften, G. (2021). Drug repurposing for rare diseases. *Trends Pharmacol. Sci.* 42, 255–267. doi:10.1016/j.tips.2021.01.003

Ruffolo, J. A., Sulam, J., and Gray, J. J. (2021). Antibody structure prediction using interpretable deep learning. *Patterns (N Y).* 3, 100406. doi:10.1016/j.patter.2021.100406

Sadybekov, A. V., and Katritch, V. (2023). Computational approaches streamlining drug discovery. *Nature* 616, 673–685. doi:10.1038/s41586-023-05905-z

Santa Maria, J. P., Jr, Wang, Y., and Camargo, L. M. (2023). Perspective on the challenges and opportunities of accelerating drug discovery with artificial intelligence. *Front. Bioinform.* 3, 1121591. doi:10.3389/fbinf.2023.1121591

Scherman, D., and Fetro, C. (2020). Drug repositioning for rare diseases: Knowledge-based success stories. *Therapie* 75, 161–167. doi:10.1016/j.therap.2020.02.007

Schipper, L. J., Zeverijn, L. J., Garnett, M. J., and Voest, E. E. (2022). Can drug repurposing accelerate precision oncology? *Cancer Discov.* 12, 1634–1641. doi:10.1158/2159-8290.CD-21-0612

Schneider, P., Walters, W. P., Plowright, A. T., Sieroka, N., Listgarten, J., Goodnow, R. A., et al. (2020). Rethinking drug design in the artificial intelligence era. *Nat. Rev. Drug Discov.* 19, 353–364. doi:10.1038/s41573-019-0050-3

Takebe, T., Imai, R., and Ono, S. (2018). The current status of drug discovery and development as originated in United States academia: The influence of industrial and academic collaboration on drug discovery and development. *Clin. Transl. Sci.* 11, 597–606. doi:10.1111/cts.12577

van den Berg, S., de Visser, S., Leufkens, H. G. M., and Hollak, C. E. M. (2021). Drug repurposing for rare diseases: A role for academia. *Front. Pharmacol.* 12, 746987. doi:10.3389/fphar.2021.746987

Vijayan, R. S. K., Kihlberg, J., Cross, J. B., and Poongavanam, V. (2022). Enhancing preclinical drug discovery with artificial intelligence. *Drug Discov. Today.* 27, 967–984. doi:10.1016/j.drudis.2021.11.023

Villoutreix, B., O. (2021). Post-pandemic drug discovery and development: Facing present and future challenges. *Front. Drug. Discov.* 1. doi:10.3389/fddsv.2021.728469

Wadman, M. (2023). FDA no longer has to require animal testing for new drugs. *Science* 379, 127–128. doi:10.1126/science.adg6276

Yildirim, O., Gottwald, M., Schüler, P., and Michel, M. C. (2016). Opportunities and challenges for drug development: Public-private partnerships, adaptive designs and big data. *Front. Pharmacol.* 7, 461. doi:10.3389/fphar.2016.00461

frontiers | Frontiers in Drug Discovery

TYPE Review
PUBLISHED 21 June 2023
DOI 10.3389/fddsv.2023.1222655

OPEN ACCESS

EDITED BY
Ho Leung Ng,
Atomwise Inc., United States

REVIEWED BY
Rodolpho C. Braga,
InsilicAll, Brazil
Jeremy J. Yang,
University of New Mexico, United States

*CORRESPONDENCE
José L. Medina-Franco,
✉ medinajl@unam.mx

RECEIVED 15 May 2023
ACCEPTED 12 June 2023
PUBLISHED 21 June 2023

CITATION
Chávez-Hernández AL, López-López E and Medina-Franco JL (2023), Yin-yang in drug discovery: rethinking *de novo* design and development of predictive models.
Front. Drug Discov. 3:1222655.
doi: 10.3389/fddsv.2023.1222655

Yin-yang in drug discovery: rethinking *de novo* design and development of predictive models

Ana L. Chávez-Hernández[1], Edgar López-López[1,2] and José L. Medina-Franco[1]*

[1]Department of Pharmacy, DIFACQUIM Research Group, School of Chemistry, Universidad Nacional Autónoma de México, Avenida Universidad, Mexico City, Mexico, [2]Department of Chemistry and Graduate Program in Pharmacology, Center for Research and Advanced Studies of the National Polytechnic Institute, Mexico City, Mexico

Chemical and biological data are the cornerstone of modern drug discovery programs. Finding qualitative yet better quantitative relationships between chemical structures and biological activity has been long pursued in medicinal chemistry and drug discovery. With the rapid increase and deployment of the predictive machine and deep learning methods, as well as the renewed interest in the *de novo* design of compound libraries to enlarge the medicinally relevant chemical space, the balance between quantity and quality of data are becoming a central point in the discussion of the type of data sets needed. Although there is a general notion that the more data, the better, it is also true that its quality is crucial despite the size of the data itself. Furthermore, the active versus inactive compounds ratio balance is also a major consideration. This review discusses the most common public data sets currently used as benchmarks to develop predictive and classification models used in *de novo* design. We point out the need to continue disclosing inactive compounds and negative data in peer-reviewed publications and public repositories and promote the balance between the positive (Yang) and negative (Yin) bioactivity data. We emphasize the importance of reconsidering drug discovery initiatives regarding both the utilization and classification of data.

KEYWORDS

big data, chemoinformatics, chemical libraries, data quality, *de novo* design, drug discovery, machine learning, negative results

Abbreviations: 2D/3D, two-dimensional/three-dimensional; ADMET, absorption, distribution, metabolism, excretion, and toxicity; AI, artificial intelligence; CADD, computer-aided drug design; COCONUT, Collection of Open NatUral ProdUcTs; DNN, deep neural networks; HBA, hydrogen bond acceptors; HBD, hydrogen bond donors; IMPPAT, A curated database of Indian Medicinal Plants, Phytochemistry And Therapeutics; MW, molecular weight; LBDD, ligand-based drug design; log P, octanol-water partition coefficient; NCE, new chemical entities; NIH(US), National Institutes of Health; PAINS, pan-assay interference compounds; Peru NPDB, Peruvian Natural Products Database; QSAR, quantitative structure-activity relationships; REAL, Enamine's REadily AccessibLe; RNNs, recurrent neural networks; SBDD, structure-based drug design; TCM, Traditional Chinese Medicine; TPSA, topological surface area; UNPD, Universal Natural Product Database.

1 Introduction

Data and the increasing role of predictive models, including machine and deep learning (Mouchlis et al., 2021; Bajorath et al., 2022), are the cornerstone of modern drug discovery programs (Zhang et al., 2022). The increasing use of computational methods that recently included deep learning is reducing the time and financial costs of finding drug candidates (Zhang et al., 2022). For instance, computer-aided drug design (CADD) has led to the discovery of more than seventy approved drugs (Sabe et al., 2021) including remdesivir as an emergency treatment against SARS-CoV-2 in 2021 (Dos Santos Nascimento et al., 2021).

CADD methods are typically divided into two main categories, structure-based drug design (SBDD) and ligand-based drug design (LBDD) that rely on the three-dimensional (3D) structure data available for one or more molecular targets, or the structure-activity data of ligands, respectively. Examples of deep learning applications in SBDD include AlphaFold to assist in homology modeling, and DiffDock in molecular docking. AlphaFold predicts 3D protein structures according to their amino acid sequences (Jumper et al., 2021), and DiffDock predicts the binding mode between the ligand and specific protein target (Corso et al., 2022). One of the most notable approaches in LBDD are quantitative structure-activity relationships (QSAR) (Dos Santos Nascimento et al., 2021). Current QSAR methods use machine learning and deep learning (Soares et al., 2022) that can be divided into linear methods and nonlinear methods (Patel et al., 2014; Greener et al., 2022). Linear methods include linear regression, multiple linear regression, partial least squares, and principal component analysis (Patel et al., 2014). Nonlinear methods include artificial neural networks, k-nearest neighbors, and Bayesian neural nets, to name a few examples (Patel et al., 2014; Greener et al., 2022).

Advances in deep learning models have a significant progress in molecule generation, representing a big step forward in bridging the gap between chemical entities and drug-like properties (Krishnan et al., 2021). Deep learning algorithms are currently used in the renewed interest in the *de novo* design of chemical libraries. In 2020, the successful application of deep learning in drug discovery, that included the *de novo* design using deep learning, was selected by the Massachusetts Institute of Technology Technology Review as one of the top ten breakthrough technologies (Juskalian et al., 2023).

De novo design is aimed at generating new chemical entities (NCE) with desired properties (Palazzesi and Pozzan, 2022). *De novo* design based on deep learning algorithms (Palazzesi and Pozzan, 2022) requires a large number of compounds that may demand significant computational resources. However, bioactivity data for a biological endpoint is not always sufficient. The lack of data has led to the development of new methods for compound selection and applications for deep learning algorithms are being developed (Guo M et al., 2021).

Knowledge-based drug design frequently involves quality data (Perron et al., 2022b) to develop models with useful predictions (Schneider et al., 2020). To this end, rethinking the methodologies used for drug discovery and development campaigns is crucial. The quality of data sets, decoy data sets and inactive compounds used in predictive models, and *de novo* design models need to be reviewed and discussed.

The main purpose of this manuscript is discussing the importance of quality data, decoy data sets, and the balance needed between inactive (i.e., "Yin") and active ("Yang") compounds currently employed in *de novo* design and developing predictive models of biological activity to generate NCE. Following up on previous studies (Schneider et al., 2020; Bajorath et al., 2022; Cherkasov, 2023), we comment on the need to rethink the way to drug design and develop campaigns. The manuscript is organized into four main sections. After this Introduction, Section 2 presents an overview of *de novo* design. Section 3 discusses the main public data sources used to develop predictive models. Section 4 discusses criteria to generate quality data sets. The last section presents a summary of conclusions and perspectives.

2 *De novo* design overview

De novo design aims to generate new chemical structures from scratch with desired predicted properties, e.g., absorption, distribution, metabolism, excretion, toxicity (ADMET), other drug-likeness properties, and biological activities (Palazzesi and Pozzan, 2022). The two main strategies for *de novo* design can be classified into SBDD and LBDD (*vide supra*) (Zhang et al., 2022). A recent example of a structured-based *de novo* design is the RELATION model that learns from the desired geometric features of protein-ligand complexes to generate new molecules (Wang et al., 2022). The generation process applies a fragment-based strategy given an initial chemical scaffold embedded in the binding site of the target protein. The pre-trained model generates molecules iteratively by sequentially adding, deleting, inserting, or replacing and linking fragments (Zhang et al., 2022).

In contrast, ligand-oriented *de novo* design focuses on the ligands themselves, thereby generating compounds with new chemical structures with novel scaffolds from active compounds while optimizing the desired properties (Xie et al., 2022). A general workflow is schematically summarized in Figure 1 which has seven main steps (Krishnan et al., 2021; Zhang et al., 2022): 1) Selecting compound data sets from public or in-house sources (further discussed in Section 3); 2) Filtering molecular data sets with desired properties such as drug-likeness. In the example of Figure 1 a data set with three subsets of compounds is represented with a star, triangle, and circle, respectively. The compounds represented with a star have drug-like properties (Lipinski et al., 2001; Veber et al., 2002); those represented with triangles comply with some of the drug-likeness properties, and those represented with circles are not compliant. Other approaches to select compounds from the data sets use molecular fingerprints (Kadurin et al., 2017) or filter compounds directly via similarity-based virtual screening instead of designing NCE from scratch (Tong et al., 2021). 3) Selecting the molecular representation as a basis to learn and represent the structures and properties of molecules, e.g., SMILES (Weininger, 1988), SELFIES (Krenn et al., 2020) or molecular graphs (Simonovsky and Komodakis, 2018). 4) Developing and validating the model for molecule generation using metrics such as the operating characteristic curve. 5) Optimizing the model by combining reinforcement learning and property prediction (Olivecrona et al., 2017). 6)

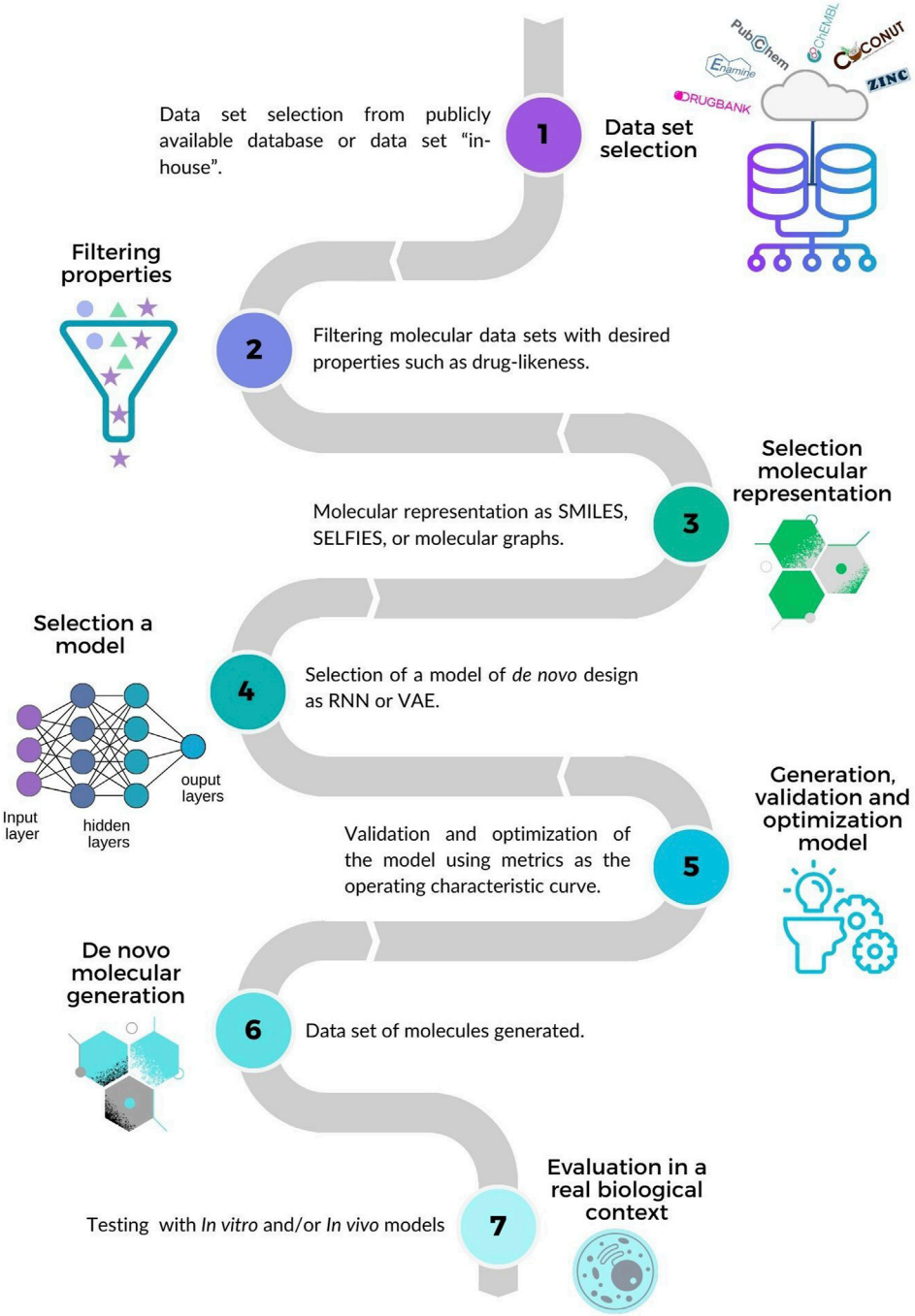


FIGURE 1
Overview of ligand-based *de novo* design. 1) Selecting data sets. 2) Filtering molecular data sets with desired properties such as drug-likeness. In this example, compounds represented with stars comply with drug-likeness properties (Lipinski et al., 2001; Veber et al., 2002). 3) Choosing a molecular representation. 4) Selecting a *de novo* design model. 5) Developing, validating and optimizing the model. 6) Generating molecules *de novo*. 7) Testing the compounds in a relevant biological experiment.

Generating molecules *de novo*, 7) Assessing the biological activity of the compounds designed in relevant *in vitro* or *in vivo* models.

Deep learning, currently used in ligand-based *de novo* design, learns the probability distribution of molecular data and generates continuous or discrete latent representations for molecules with property optimization (Gómez-Bombarelli et al., 2018). The algorithms map the learned probability distribution and molecule representation into novel molecules while optimizing molecular properties (Bilodeau et al., 2022) through the tuning of hyperparameters (Perron et al., 2022a; Bender et al., 2022). Advances in deep learning are significantly advancing molecule generation, representing a big step forward in bridging the gap

between chemical entities and drug-like properties (Krishnan et al., 2021).

Ligand´s properties can be optimized in two steps: 1) property-based generation, wherein models would learn the chemical space of molecules with desirable properties; and 2) novel molecules are generated within a desired property space (Bilodeau et al., 2022). Examples of ligand-based *de novo* design are deep neural networks (DNN), recurrent neural networks (RNNs) (Olivecrona et al., 2017), and variational autoencoders (VAE) (Gómez-Bombarelli et al., 2018). Olivercroma *et al.* (Olivecrona et al., 2017) proposed the REIVENT model that uses RNN for *de novo* design. They introduced a reinforcement learning method to fine-tune the pre-trained RNN so the model could generate structures with desirable properties. Recently, Blaschke et al. released REINVENT 2.0 (Blaschke et al., 2020) making the code freely accessible in Github.

Ligand-based *de novo* design using DNN (Palazzesi and Pozzan, 2022) requires a large number of compounds that demand more computational resources. The DNN architecture is prone to problems because of fitting numerous parameters. For this reason, a large training data set is needed to reduce the risk of overfitting. However, sufficient bioactivity data for a biological endpoint is not always available (Wu et al., 2018). The lack of sufficient data has led to using methods for compound selection or the development of new methods for compound selection. Altae-Tran et al. (Altae-Tran et al., 2017) demonstrated how the one-shot learning paradigm can be used to address the overfitting problem; they used DNN to transform small molecules into embedding vectors in a continuous feature space whose similarity measures are then iteratively learned. They showed that this DNN architecture offers convincing performance in many activity prediction tasks given limited amounts of training. On the other hand, computer scientists advise using algorithms that can detect meaningful patterns in small data sets, which is a typical case in the early stage of drug discovery (Schneider and Clark, 2019). For instance, an initial approach to *de novo* design is to start from small data sets of compounds with diverse structures and diverse properties of pharmaceutical relevance (Chávez-Hernández and Medina-Franco, 2023).

The availability of gold standard datasets as well as independently generated data sets are valuable in generating well-performing models (Vamathevan et al., 2019). Dissimilarity-based compound selection could be improved if one focused the selection on a structural diverse dataset (for instance derived from natural products). Some approaches proposed suggest using quality data sets using a dissimilarity-based compound selection method such as the MaxMin or MaxSum algorithms (Leach and Gilleteds, 2007). Recently, we reported the use of the MaxMin algorithm for the selection of natural product subsets (Chávez-Hernández and Medina-Franco, 2023) using the Universal Natural Product Database (UNPD) (Gu et al., 2013). In that study, the natural product subsets generated had the most diverse chemical structures with physicochemical properties of pharmaceutical interest similar to the original data set. Chemical structures in the natural product subsets were represented with SMILES encoding chirality, an important feature of natural products.

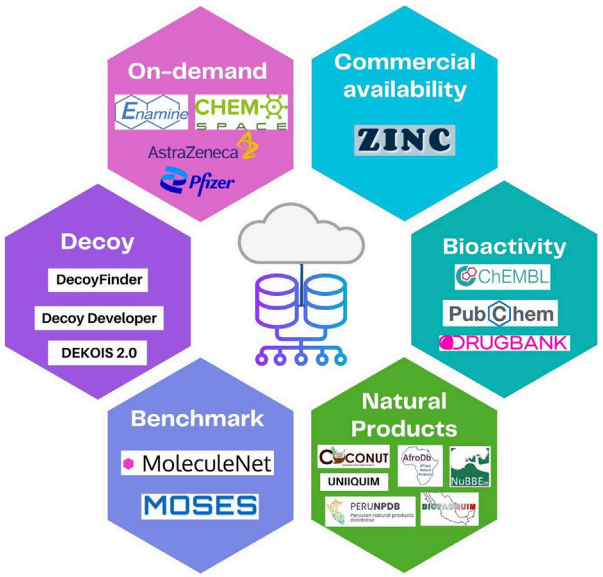


FIGURE 2
Classification of compound databases and representative examples of each one. For the discussion of this manuscript, databases are split into six main categories: on-demand, commercial availability, bioactivity, natural products, benchmark and decoys.

3 Main sources of data sets used to develop generative and predictive models

3.1 Current status of reference and benchmark datasets

The first step in *de novo* design is to select, from the vast chemical space, the appropriate subset of all possible molecules for a desired biological activity (Schneider et al., 2000). To have an idea, the size of the chemical space has been estimated at around 10^{60} small molecules and between 10^{20}–10^{24} for all molecules up to 30 atoms that comply with Lipinski's rule-of-five (Reymond, 2015). According to Yang *et al.* compound data sets can be classified into on-demand databases, collections containing bioactivity data, compounds databases commercially available, and natural products databases (Yang et al., 2019). Herein, we include benchmark, decoy and inactive compounds data sets as others categories as illustrated in Figure 2. In this figure, on-demand databases are further divided into commercially available (e.g., Enamine-REAL, CHEMriya and Freedom Space) (Chemspace, 2023) and in-house (e.g., Pfizer and AstraZeneca). The figure shows examples of compound databases in other categories which are discussed in the remainder of this section.

Among the different types of chemical databases, *de novo* design employs libraries from different categories outlined in Figure 2. Specific examples are ChEMBL (Davies et al., 2015; Mendez et al., 2019), PubChem (Kim et al., 2023), DrugBank (Wishart et al., 2006; Wishart et al., 2008; Wishart et al., 2018), Enamine´s REadily AccessibLe (REAL) (Enamine, 2023), CHEMriya (CHEMriya, 2023), Freedom Space (Chemspace, 2023), ZINC-22 (Tingle

TABLE 1 Main sources of public molecular data sets used in *de novo* design.

Data sets	Category	Description	Ref.
ChEMBL	Bioactivity	Database with 2,354,965 bioactive drug-like small molecules with 2D structures and calculated properties.	Davies et al. (2015), Mendez et al. (2019)
PubChem	Bioactivity	Database at the US National Institutes of Health with 115 million compounds. It includes names, molecular formulas, structures, physical properties, and biological activities.	Kim et al. (2023)
DrugBank	Bioactivity	Version 5.1.10 contains 15,448 drug entries including 2,740 approved small molecule drugs.	Wishart et al. (2006)
ZINC-22	Commercial	Database with over 37 billion enumerated, searchable, commercially available compounds in 2D.	Tingle et al. (2023)
CHEMriya	On-demand	Database with 12 billion novel and synthetically feasible small molecules.	CHEMriya (2023)
Freedom Space (Chemspace)	On-demand	Database with 201 million molecules; 73% of its compounds comply with drug-likeness properties.	Chemspace (2023)
Enamine-REAL	On-demand	Database with 6 billion synthetic compounds that comply with drug-likeness properties.	Enamine (2023)
MoleculeNet	Benchmark	Compilation of 17 datasets with over 700,000 compounds in total used for comparison of different machine learning algorithms.	Wu et al. (2018)
MOSES	Benchmark	Dataset with 1,936,962 molecules from ZINC Clean Lead suitable for hit identification and ADMET optimization. It does have metrics to detect common issues in generative models such as overfitting or if the model does not limit to producing only a few typical molecules.	Polykovskiy et al. (2020)

et al., 2023), and MoleculeNet (Wu et al., 2018) which more details for each one are provided in Table 1 and further commented in the next sections.

3.2 On-demand databases

Early approaches to ligand-based *de novo* design involved fragment compounds into unique building blocks which could be recombined to make new molecules. A number of commercial suppliers of chemical samples offer large make-on-demand collections that can be reliably synthesized because the building blocks are available as well as the synthetic routes and methods (Warr et al., 2022; Korn et al., 2023). There are also large collections of fragments or building blocks commercially available. Examples of on-demand compound databases and suppliers are REAL (Enamine) (Enamine, 2023), CHEMriya (OTAVA) (CHEMriya, 2023), and Freedom Space (Chemspace) (Chemspace, 2023) (Table 1). REAL database (Enamine, 2023) comprises over 6 billion molecules that comply with the traditional drug-likeness criteria. CHEMriya (CHEMriya, 2023) contains 12 billion novel and synthetically feasible small molecules whose molecules are not explicitly listed in the public domain. Freedom Space (Chemspace, 2023) contains 201 million molecules and 73% of its compounds are drug-like (as assessed with the "rule of five"). Examples of on-demand in-house databases from the pharmaceutical industry are 10^{15} compounds of AZ Space (AstraZeneca) (Grebner, 2022), 10^{19} compounds of JFS (Johnson & Johnson) (Warr, 2021), 10^{18} compounds of PGVL (Pfizer) (Hu et al., 2012), 10^{17} compounds BICLAIM (Boehringer Ingelheim) (Korn et al., 2023), and 10^{20} compounds MASSIV (Merck/EMD) (Korn et al., 2023).

3.3 Commercially available databases

One of the largest and long-standing compendiums of commercially available compounds in ZINC. The most recent version, ZINC-22 (Tingle et al., 2023) contains over 37 billion enumerated, searchable, commercially available compounds in 2D. Over 4.5 billion have been built in biologically relevant ready-to-dock 3D formats (Tingle et al., 2023). Some examples of *de novo* design using ZINC include the design of inhibitors of DDR1 (discoidin domain receptor 1, a kinase target implicated in fibrosis and other diseases) (Zhavoronkov et al., 2019) and compounds with activity towards the dopamine receptor D2 (Liu et al., 2019; Maziarka et al., 2020).

3.4 Bioactivity databases

De novo design based on deep learning algorithms frequently use PubChem, ChEMBL, and DrugBank to select subsets of compounds focused on a biological target or biological endpoint as the design of ligands (Li et al., 2018; Li et al., 2022; Liu et al., 2019). PubChem (Kim et al., 2023) is a freely accessible database from the US National Institutes of Health (NIH) with over 115 million compounds. At the time of writing, the most recent version release of ChEMBL is 32 (Davies et al., 2015; Mendez et al., 2019) and contains 2,354,965 compounds bioactive drug-like small molecules with 2D structures and calculated properties. DrugBank (Wishart et al., 2006; Wishart et al., 2008; Wishart et al., 2018) version 5.1.10 (released 2023-01-04) contains 15,448 drug entries including 2,740 approved small molecule drugs, 1,577 approved biologics (proteins, peptides, vaccines, and allergens), 134 nutraceuticals and over 6,717 experimental (discovery-phase) drugs. Some applications include the *de novo* design of SARS-CoV-2 Mpro inhibitors (Li et al., 2022), the design of ligands against the adenosine receptor ($A_{2A}R$) (Liu et al., 2019), and the generation of compounds analogs to celecoxib (used to manage symptoms of various types of arthritis pain and reduce precancerous polyps in the colon) (Li et al., 2018; DRUGBANK, 2023).

TABLE 2 Examples of natural product databases in the public domain.

Data sets	Description	Ref.
COCONUT	Extensive database with 406,076 unique structures.	Sorokina et al. (2021)
SuperNatural 3.0	A database with 449 058 natural compounds and derivatives. It includes chemical structure, physicochemical information, information on pathways, mechanism of action, toxicity, vendor information if available, drug-like chemical space prediction for several diseases such as antiviral, antibacterial, antimalarial, anticancer, and target-specific cells.	Gallo et al. (2023)
UNPD	Second-largest database with around 229,000 natural products that contain chirality information.	Gu et al. (2013)
TCM Database@Taiwan	Database with more than 20,000 pure compounds isolated from 453 TCM ingredients.	Chen (2011)
IMPPAT	Database of 9,596 phytochemicals from 1,742 Indian medicinal plants.	Mohanraj et al. (2018)
AfroDB	Compound collection with more than 1,000 compounds from African medicinal plants.	Ntie-Kang et al. (2013)
$NuBBE_{DB}$	Brazilian database with 2,223 natural products encoding as SMILES, InChI, and InChIKey strings, Ro5 and Veber descriptors, source, therapeutic effect, and reference.	Valli et al. (2013), Pilon et al. (2017), Saldívar-González et al. (2019)
SistematX	Brazilian database with 9,514 unique secondary metabolites encoding as SMILES, InChI, and InChIKey strings, and include physicochemical drug-like descriptors, predicted biological activities, and reference.	Scotti et al. (2018), Costa et al. (2021)
CIFPMA	Database developed at the University of Panama. It contains natural products that have been tested in over 25 *in vitro* and *in vivo* bioassays, for different therapeutic targets.	Olmedo et al. (2017), Olmedo and Medina-Franco (2020)
PeruNPDB	Peru database developed at the Catholic University of Santa Maria. The current version has 280 natural products from animals and plants.	Barazorda-Ccahuana et al. (2023)
BIOFACQUIM	Mexican database with structures of 531 natural products isolated and characterized at UNAM and other Mexican institutions.	Pilón-Jiménez et al. (2019), Sánchez-Cruz et al. (2019)
UNIIQUIM	Mexican database with 1,112 plant natural products mostly isolated and characterized at the Institute of Chemistry of the UNAM.	*UNIIQUIM* (2015)

Other libraries of natural products with an emphasis on commercial availability are listed on the NIH website (NIH, 2023).

3.5 Natural product databases

Natural product databases (Gómez-García and Medina-Franco, 2022; Saldívar-González et al., 2022) are important in drug discovery. From drugs approved by 2020 about 23% are natural products or derivatives (Newman and Cragg, 2020). Natural products have a diversity of privileged scaffolds (Atanasov et al., 2021; Grigalunas et al., 2022) and molecular fragments (Chávez-Hernández et al., 2020a; Chávez-Hernández et al., 2020b) that depend on the particular source (Medina-Franco et al., 2022b); a diversity of chiral centers; and a larger fraction of sp^3 carbon atoms and functional groups (Atanasov et al., 2021; Grigalunas et al., 2022).

Privileged structures were defined by Evans et al. (Evans et al., 1988) as *chemical structures capable of providing useful ligands for more than one receptor judicious modification of such structures could be a viable alternative in the search for new receptor agonists and antagonists*. Schneider and Schneider (2017) define a privileged structure as a chemical structure that may be considered to possess geometries suitable for decoration with side chains, such that the resulting products bind to different target proteins or a ligand that potently interacts with one (selective binder) or many target receptors (promiscuous binder). To this end, natural products are used in the development of pseudo-natural products, compounds that are generated through a *de novo* combination of natural product fragments, allowing the exploration of uncharted areas of biologically relevant chemical space that are different from the chemical space covered by the compounds from which they are derived (Grigalunas et al., 2022).

Representative natural product datasets that can be used in *de novo* design are Collection of Open NatUral ProdUcTs (COCONUT) (Sorokina et al., 2021), SuperNatural 3.0 (Gallo et al., 2023), UNPD (Gu et al., 2013), $NuBBE_{DB}$ (Pilon et al., 2017; Saldívar-González et al., 2019), SistematX (Scotti et al., 2018; Costa et al., 2021), CIFPMA (Olmedo et al., 2017; Olmedo and Medina-Franco, 2020), PeruNPDB (Barazorda-Ccahuana et al., 2023), BIOFACQUIM (Pilón-Jiménez et al., 2019; Sánchez-Cruz et al., 2019), UNIIQUIM(UNIIQUIM, 2015), and are summarized in Table 2.

SuperNatural 3.0, COCONUT and UNPD are the most extensive natural product databases. SuperNatural 3.0 (Gallo

et al., 2023) is arguably the most extensive natural product database with 449,058 natural compounds and derivatives; followed by COCONUT (Sorokina et al., 2021) with 406,076 unique structures (no encoding stereochemistry) and UNPD (Gu et al., 2013) with 197,201 natural products that contain chirality information.

Several public natural products databases compile the compounds isolated and characterized from a geographical region or the country of origin as China, India and Africa. For instance, Chinese Traditional Medicine (TCM) Database@Taiwan (Chen, 2011) is a non-commercial TCM database with more than 20,000 pure compounds isolated from 453 TCM ingredients; A curated database of Indian Medicinal Plants, Phytochemistry And Therapeutics (IMPPAT) (Mohanraj et al., 2018) is a manually curated database of 9,596 phytochemicals from 1,742 Indian medicinal plants; and AfroDB (Ntie-Kang et al., 2013) with more than 1,000 small and structural diversity compounds from African medicinal plants.

Representative Latin American databases (Gómez-García and Medina-Franco, 2022) are $NuBBE_{DB}$ (Pilon et al., 2017; Saldívar-González et al., 2019), SistematX (Scotti et al., 2018; Costa et al., 2021) from Brazil; CIFPMA (Olmedo et al., 2017; Olmedo and Medina-Franco, 2020) from Panama; PeruNPDB (Barazorda-Ccahuana et al., 2023) from Peru; BIOFACQUIM (Pilón-Jiménez et al., 2019; Sánchez-Cruz et al., 2019) and UNIIQUIM (*UNIIQUIM*, 2015) from Mexico. The current version of $NuBBE_{DB}$ (Pilon et al., 2017; Saldívar-González et al., 2019) contains 2,223 natural products encoding as linear notations as SMILES. SistematX (Scotti et al., 2018; Costa et al., 2021) has 9,514 unique secondary metabolites arising from 20,934 botanical occurrences across five families. Other natural product collections from Latin America are CIFPMA, the Natural Products Database from the University of Panama, Republic of Panama (Olmedo et al., 2017; Olmedo and Medina-Franco, 2020) with 354 compounds. CIFPMA molecules have the potential to show target selectivity in biochemical assays and are useful molecules to identify reference compounds for virtual screening campaigns (Olmedo et al., 2017; Olmedo and Medina-Franco, 2020). The first version of the Peruvian Natural Products Database (PeruNPDB) had 280 natural products isolated from plants and animal sources (Barazorda-Ccahuana et al., 2023). BIOFACQUIM (Pilón-Jiménez et al., 2019; Sánchez-Cruz et al., 2019) contains 531 natural products isolated and characterized at the School of Chemistry of the National Autonomous University of Mexico (UNAM) and other Mexican institutions. UNIIQUIM (UNIIQUIM, 2015) with 1,112 plant natural products mostly isolated and characterized at the Institute of Chemistry of the UNAM.

3.6 Benchmark databases

The development of reliable machine learning algorithms has been limited due to the lack of standard benchmark datasets to compare the efficacy of the methods proposed (Jain and Nicholls, 2008). Furthermore, machine learning in chemistry compared with other areas such as computer speech and vision has a main disadvantage, the data recovery (Wu et al., 2018; Guo et al., 2022), because of measuring chemical properties often requires specialized instruments; as a result, datasets with experimentally determined results are small and often not sufficiently large to cover the high-demanding needs of machine-learning tasks (Wu et al., 2018). Another challenge is data splitting (the way in which datasets are split into training data and testing data). Some are random selection and rational selection. The former is randomly extracting a compound's fraction from the data set. In contrast to rational selection, training and testing are selected from the same clusters of compounds. Random selection is common in machine learning but is often not correct for chemical data (Sheridan, 2013). In response to these challenges, standard benchmark data sets are being developed to evaluate *de novo* design protocols [(Wu et al., 2018; Brown et al., 2019; Polykovskiy et al., 2020). One example is MoleculeNet (Wu et al., 2018), a large-scale data set built upon multiple public databases. MoleculeNet is organized into regression and classification datasets and has over 700,000 compounds tested on a range of different properties subdivided into four categories (quantum mechanics, physical chemistry, biophysics, and physiology). Another example is the Molecular Sets (MOSES) (Polykovskiy et al., 2020) that contains 1,936,962 molecules (split into training, testing and scaffold datasets) and a set of metrics to evaluate the quality and diversity of generated structures. Metrics detect common issues in generative models such as overfitting or if the *de novo* design model just generates fairly common (not novel) structures (Brown et al., 2019; Polykovskiy et al., 2020). The developers of MOSES implemented and compared several molecular generation models and suggested using the results as reference points for further advancements in generative chemistry research.

3.7 Current decoy data sets and inactive compounds

Accuracy of predictive models depends on data quality and quantity. Also, the balance between active and inactive compounds is important, which remains an issue to resolve. Historically, the publication of active compounds in a given assay or with a particular endpoint has been prioritized over inactive molecules. For example, a recent comprehensive analysis of published screening bioactivity data shows that in ChEMBL V.29 (release in 2022) there is a large number of active compounds (*ca. 71%*) with respect to the inactive ones (*ca.* 31%); contrary to what it would be expected (López-López et al., 2022). These results highlight the relevance of changing the mindset about the importance and utility of inactive or negative data (keeping in mind that the definition of "inactive" is subjective as it depends on the particular biological assay and the predefined threshold to deem a compound inactive).

Decoy data sets have been developed in an attempt to reduce the gap between inactive (or negative) and active compounds. Decoy molecules are assumed non-active but have high physicochemical property similarity (but not topologically) to reference compounds (Réau et al., 2018). Decoys are useful to evaluate benchmark models that were assembled in the absence of inactive compounds experimentally measured (Irwin, 2008) and can be used to enrich *de novo* design models. Table 3 summarizes examples of large databases of experimentally tested active or inactive compounds, decoy datasets, and tools to generate decoys for specific projects.

Decoy compounds have been used to describe, explore, and expand the knowledge of active molecules. For example,

TABLE 3 Examples of potential inactive and decoy resources for enriching *de novo* design models.

Datasets with active and inactive compounds	Criteria to select inactive data	Ref.
ChEMBL	Reported activity data.	Davies et al. (2015), Mendez et al. (2019)
PubChem		Kim et al. (2023)
Binding DB	Reported ligand-receptor affinity.	Chen et al. (2002)
Decoy datasets	Common decoy selection criteria	
ZINC	Compounds that share drug-like properties with the reference (active) compounds.	Tingle et al. (2023)
DUD-E		Mysinger et al. (2012)
DUD	Database with 2950 annotated ligands and 95,316 property-matched decoys for 40 targets.	Irwin (2008)
MUV	Compounds that share structural similarity with active reported compounds.	Rohrer and Baumann (2009)
DEKOIS 2.0	Compounds that share drug-like properties and structural similarity with the reference (active) compounds.	Bauer et al. (2013)
Decoy tools	Common decoy compound selection criteria	
DecoyFinder	Allows the automatic creation of datasets of compounds with physicochemical similarity and without structural similarity respect to the reference (active) compounds.	Cereto-Massagué et al. (2012)
RADER	Allows the automatic generation of datasets of compounds with physicochemical and structural similarity with respect to the reference (active) compounds.	Wang et al. (2017)
ZINC pharmer	Enables the automatic identification of compounds with pharmacophore similarity with respect to the reference (active and inactive) compounds.	Koes and Camacho (2012)
Decoy Developer	Allows the automatic generation of peptides decoys.	Shipman et al. (2019)

TABLE 4 Examples of applications of decoys in *de novo* design.

Approach	Purpose of using decoy sets	Ref.
Ligand-based	• Validation of new protocols and scoring functions based on similarity metrics and 3D shape.	(Arús-Pous et al. (2020); Awale and Reymond. (2015); Cao et al. (2020); Medina-Franco et al. (2019); Norinder et al. (2019); Papadopoulos et al. (2021); Skalic et al. (2019b); Skalic et al. (2019a); Ullanat (2020)
	• Improvement of the accuracy of AI-based models.	
	• Improvement of the accuracy of QSAR models.	
	• Enrichment of inactive "dark regions" in chemical space.	
Structure-based	• Validation of new protocols and scoring functions based in docking, molecular dynamics, and pharmacophore modeling.	Balius et al. (2013); Beato et al. (2013); Guo J et al. (2021); Ma et al. (2021); Niitsu and Sugita (2023)
	• Peptide and protein design.	

rationalizing the physicochemical, chemical, biological, and clinical data of active compounds (López-López et al., 2021a). Recently, decoys can be employed in several *de novo* protocols based on ligand or structure as summarized in Table 4.

4 Criteria to generate compound datasets with high quality

The quality of a data set is multifaceted. Commonly, it is associated with the experimental reproducibility of each data point and the experimental similarities between the protocols used to derive such data. Another important aspect of data quality is the balance between active and inactive compound. The latter is specially a challenge in public data sets due to the overall lack of published negative data. Finding qualitative yet better quantitative relationships between chemical structures and biological activity has been long pursued in medicinal chemistry and drug discovery. With the rapid increase and deployment of the predictive machine and deep learning methods, as well as the increased interest in the *de novo* design of chemical libraries (Mouchlis et al., 2021), the quantity and quality of data are

TABLE 5 Overview of suggested general criteria to generate quality datasets useful in *de novo* design.

Criteria	Brief description	Ref.
Balance	• Quality and quantity data allow the exploration of substantial regions of chemical space.	Scannell et al. (2022); Yang et al. (2023)
Quality (confidence) data	• The reliability of the activity data (active or inactive) is crucial to develop predictive models. This is the activity data reproducibility.	Kumar et al. (2022)
Diversity	• Datasets with a high chemical and structural diversity improve the generation of novel molecules.	Saldívar-González and Medina-Franco (2022)
Preparation or curation	• Dataset curation must be focused on one or multiple drug targets. Therefore, molecular descriptors and the cut-off threshold used for the curated must be properly selected. • Dataset should be oriented to resolve specific outcomes and avoid Pan-Assay Interference Compounds (PAINS) structures or chemical structures related to side effects. • In small datasets it is very important to have as much accurate data as possible. The maximum observable accuracy of classification models also depends on the experimental uncertainty and the distribution of the measured values. For instance, datasets with large noise are not recommended for the comparison of different models.	Fourches et al. (2016); Kramer and Lewis (2012)
Complete information	• According to the main objective of each project, the dataset used must contain reliable data related to the project's objective. For example, structure containing chemical and physicochemical information, bioactivity data for the related biological endpoint, or outcomes from clinical trials, etc.	López-López et al. (2021b); López-López and Medina-Franco. (2023); Wu et al. (2023a); Wu et al. (2023b)

becoming a central point in the discussion of the type of data sets needed (Schneider et al., 2020). While the more data (Cherkasov, 2023), the better, it is also true that the quality of the data available (that might not be quite large) is also crucial. Furthermore, the balance between active and inactive compounds is also a major consideration (López-López et al., 2022). Table 5 summarizes criteria for generating quality data sets. The list is not exhaustive but covers what the authors consider key points based on experience and what has been discussed extensively in the literature. Each point is supported by the references indicated in the table and further commented in the next subsections.

4.1 Balance

As discussed previously, several current data sets in the public domain are unbalanced due to the infrequent practice of reporting inactive compounds and negative data in general. Historically, the negative and inactive data of preclinical compounds has been ignored by most journals that favor the publication of most active compounds and positive results (Medina-Franco and López-López, 2022). However, inactive and negative data are essential in drug design and development. For example, the analysis of high-quality inactive and negative data improves clinical success rate, reduces costs associated with drug development, and reduces the side effects rates (Hayes and Hunter, 2012; López-López and Medina-Franco, 2023). Moreover, data mining and AI approaches are largely benefitted from inactive compounds (Yu, 2021; López-López et al., 2022). The use of inactive and negative data allows real data augmentation to develop AI models, improve their accuracy, and reduce the rate of false-positive cases (Korkmaz, 2020; *IBM*, 2022). Also, the inactive and negative data facilitates the generation of QSPRs models that allows the rationalization of basically any property (Kramer and Lewis, 2012; Norinder et al., 2019).

4.2 Confidence of the activity data

An unwritten rule on AI and computational projects in general is "garbage in, garbage out". This perspective has direct implications in drug design (Bajorath et al., 2022). Recent studies have demonstrated that the use of quality data allows generating of AI models with higher accuracy than the AI models generated from larger datasets but with low-quality.

4.3 Chemical and structural diversity

In general, a compound dataset with a large or broad applicability domain, as captured by the diversity of the contents, can give rise to predictive models with a large coverage. This is, molecules from diverse chemical structures could be conveniently interpolated in those models. As a comparison in an experimental setting, high-throughput screening of chemical diverse libraries increases the chances to find hit compounds for targets for which no hit compounds have been previously identified.

Due to the rapid expansion of the chemical universe, recently called the 'Big Bang' of the chemical universe (Cherkasov, 2023) it is relatively easy to have access to large and diverse regions of the chemical space. However, a practical challenge is to manage such large compound data sets computationally while developing and testing new models. A similar practical problem emerged when combinatorial chemistry was at its peak: it was challenging to design rationally novel large and diverse combinatorial libraries. To tackle this problem numerous diversity selection algorithms have been developed (Leach and Gillet, 2007). We recently applied a dissimilarity-based compound selection method to obtain three diverse subsets of natural products (with 14,994, 7,497, and 4,998 compounds, respectively) from the UNP. The subsets, that are freely available, can be readily used for *the novo* design

applications and as benchmarks for similarity/diversity analysis (Chávez-Hernández and Medina-Franco, 2023).

4.4 Preparation or curation

A general curation protocol used on drug discovery datasets is to eliminate duplicate structures, canonize their SMILES representation, eliminate salts, and metals. However, according to the main goal of the *de novo* design model, additional steps to prepare a dataset could be taking into account, for example: 1) eliminating compounds with structural PAINS to reduce the rate of false-positive compounds prediction; 2) deleting compounds reported with side effects and/or ADMET deficiencies, to prioritize the generation of safe and optimization compounds.; or 3) making sure to keep in the dataset compounds with high activity confidence to improve the quality of predicted outputs. This list must be adapted according to the main goal of the *de novo* design model. It is also noted the need to develop robust and consistent protocols that take into scout metal-containing compounds as they have a major role in medicinal inorganic chemistry (Medina-Franco et al., 2022a).

4.5 Completeness

Chemical structures should contain the required or relevant information for the goals of the study. For instance, compounds should be annotated with stereochemistry information if the 3D structure and conformation is critical; electronic density and quantum chemical data if the reactivity is key point to predict; the type of the biological activity data such as biochemical, cell-based or functional assays; drug-drug interaction data, pharmacogenomics, or post-marketing annotations; should be aligned with the type of outcome to be predicted and later validated experimentally.

5 Perspectives of *de novo* design

One of the major perspectives of the *de novo* design is using balanced data sets (as much as experimental data is available) to build reliable models. Similar to QSAR predictive models, it is also crucial the validation of *de novo* protocols using standard and well-curated benchmark datasets (discussed in Section 3.6). With the increasing data availability to generate and train new models, it is becoming increasingly easy to explore regions of chemical space previously uncharted and continue contributing to the so-called "big bang" expansion of the chemical space. A major perspective in this direction is to explore biologically relevant compounds but outside the traditional small molecule chemical space (Medina-Franco et al., 2014). For instance, exploring metallodrugs (Medina-Franco et al., 2022a), macrocycles (Liang et al., 2022), peptides, or the combination of commonly explored chemical spaces, e.g., pseudo-natural products (discussed in Section 3.5).

6 Conclusion

Among the main types of datasets used in *the novo* design are on-demand collections, compounds annotated with biological activity, commercially available libraries, and natural products. More recently, a large benchmark data set was developed for machine learning applications. Although there is a general agreement in machine learning that the more data, the better, it is becoming more and more evident to consider the reliability and the quality of the data sets as critical features of the data. Part of the quality is associated with the balance between inactive and active compounds (in a rough analogy with the Yin-Yang concept), tasks that are not always feasible due to the general scarcity of negative (inactive compounds). The later point further emphasizes the continued need to publish and disclose negative results. Due to the fact that the experimental data of inactive compounds are not common, the community is using decoy data sets that by themselves are subject to design and refining using rational approaches. Decoy data sets try to fill the void of experimentally determined inactive molecules. Major criteria to take into account to generate compound data sets with high quality include balanced data sets in terms of active and inactive compounds (when the experimental information is available), structural and chemical diversity, curation or preparation according to the goals of the project, and complete information. All these together contribute to the perspectives of *de novo* design that foresees a continued and rapid expansion of molecules with the potential to become drugs.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

Authors are grateful to DGAPA, UNAM, Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), grant no. IN201321. We also thank the Dirección General de Cómputo y de Tecnologías de Información y Comunicación (DGTIC), UNAM, for the computational resources to use Miztli supercomputer at UNAM under project LANCAD-UNAM-DGTIC-335; and the innovation space UNAM-HUAWEI the computational resources to use their supercomputer under project-7 "*Desarrollo y aplicación de algoritmos de inteligencia artificial para el diseño de fármacos aplicables al tratamiento de diabetes mellitus y cáncer*".

Acknowledgments

AC-H and EL-L are thankful to CONACyT, Mexico, for the Ph.D. scholarships number 847870 and 894234, respectively.

Conflict of interest

The author JLM-F declared that he was an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Altae-Tran, H., Ramsundar, B., Pappu, A. S., and Pande, V. (2017). Low data drug discovery with one-shot learning. *ACS central Sci.* 3 (4), 283–293. doi:10.1021/acscentsci.6b00367

Arús-Pous, J., Patronov, A., Bjerrum, E. J., Tyrchan, C., Reymond, J. L., Chen, H., et al. (2020). SMILES-based deep generative scaffold decorator for de-novo drug design. *J. cheminformatics* 12 (1), 38. doi:10.1186/s13321-020-00441-8

Atanasov, A. G., Zotchev, S. B., Dirsch, V. M., and Supuran, C. T.International Natural Product Sciences Taskforce (2021). Natural products in drug discovery: Advances and opportunities. *Nat. Rev. Drug Discov.* 20 (3), 200–216. doi:10.1038/s41573-020-00114-z

Awale, M., and Reymond, J-L. (2015). Similarity mapplet: Interactive visualization of the directory of useful decoys and ChEMBL in high dimensional chemical spaces. *J. Chem. Inf. Model.* 55 (8), 1509–1516. doi:10.1021/acs.jcim.5b00182

Bajorath, J., Chávez-Hernández, A. L., Duran-Frigola, M., Fernández-de Gortari, E., Gasteiger, J., López-López, E., et al. (2022). Chemoinformatics and artificial intelligence colloquium: Progress and challenges in developing bioactive compounds. *J. cheminformatics* 14 (1), 82. doi:10.1186/s13321-022-00661-0

Balius, T. E., Allen, W. J., Mukherjee, S., and Rizzo, R. C. (2013). Grid-based molecular footprint comparison method for docking and de novo design: Application to HIVgp41. *J. Comput. Chem.* 34 (14), 1226–1240. doi:10.1002/jcc.23245

Barazorda-Ccahuana, H. L., Ranilla, L. G., Candia-Puma, M. A., Cárcamo-Rodriguez, E. G., Centeno-Lopez, A. E., Davila-Del-Carpio, G., et al. (2023). PeruNPDB: The Peruvian natural products database for *in silico* drug screening. *Sci. Rep.* 13 (1), 7577. doi:10.1038/s41598-023-34729-0

Bauer, M. R., Ibrahim, T. M., Vogel, S. M., and Boeckler, F. M. (2013). Evaluation and optimization of virtual screening workflows with DEKOIS 2.0-a public library of challenging docking benchmark sets. *J. Chem. Inf. Model.* 53 (6), 1447–1462. doi:10.1021/ci400115b

Beato, C., Beccari, A. R., Cavazzoni, C., Lorenzi, S., and Costantino, G. (2013). Use of experimental design to optimize docking performance: The case of LiGenDock, the docking module of LiGen, a new de novo design program. *J. Chem. Inf. Model.* 53 (6), 1503–1517. doi:10.1021/ci400079k

Bender, A., Schneider, N., Segler, M., Patrick Walters, W., Engkvist, O., and Rodrigues, T. (2022). Evaluation guidelines for machine learning tools in the chemical sciences. *Nat. Rev. Chem.* 6 (6), 428–442. doi:10.1038/s41570-022-00391-9

Bilodeau, C., Jin, W., Jaakkola, T., Barzilay, R., and Jensen, K. F. (2022). Generative models for molecular discovery: Recent advances and challenges. *Comput. Mol. Sci.* 12 (5), e1608. doi:10.1002/wcms.1608

Blaschke, T., Arús-Pous, J., Chen, H., Margreitter, C., Tyrchan, C., Engkvist, O., et al. (2020). Reinvent 2.0: An AI tool for de novo drug design. *J. Chem. Inf. Model.* 60 (12), 5918–5922. doi:10.1021/acs.jcim.0c00915

Brown, N., Fiscato, M., Segler, M. H. S., and Vaucher, A. C. (2019). GuacaMol: Benchmarking models for de Novo molecular design. *J. Chem. Inf. Model.* 59 (3), 1096–1108. doi:10.1021/acs.jcim.8b00839

Cao, L., Goreshnik, I., Coventry, B., Case, J. B., Miller, L., Kozodoy, L., et al. (2020). De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science* 370 (6515), 426–431. doi:10.1126/science.abd9909

Cereto-Massagué, A., Guasch, L., Valls, C., Mulero, M., Pujadas, G., and Garcia-Vallvé, S. (2012). DecoyFinder: An easy-to-use python GUI application for building target-specific decoy sets. *Bioinformatics* 28 (12), 1661–1662. doi:10.1093/bioinformatics/bts249

Chávez-Hernández, A. L., and Medina-Franco, J. L. (2023). Natural products subsets: Generation and characterization. *Artif. Intell. Life Sci.* 3, 100066. doi:10.1016/j.ailsci.2023.100066

Chávez-Hernández, A. L., Sánchez-Cruz, N., and Medina-Franco, J. L. (2020a). A fragment library of natural products and its comparative chemoinformatic characterization. *Mol. Inf.* 39 (11), e2000050. doi:10.1002/minf.202000050

Chávez-Hernández, A. L., Sánchez-Cruz, N., and Medina-Franco, J. L. (2020b). Fragment library of natural products and compound databases for drug discovery. *Biomolecules* 10 (11), 1518. doi:10.3390/biom10111518

Chemriya (2023). CHEMriya. Available at: https://chemriya.com/ (accessed May 13, 2023).

Chemspace (2023). Freedom space. Available at: https://chem-space.com/compounds/freedom-space (accessed May 13, 2023).

Chen, C. Y-C. (2011). TCM Database@Taiwan: The world's largest traditional Chinese medicine database for drug screening *in silico*. *PloS one* 6 (1), e15939. doi:10.1371/journal.pone.0015939

Chen, X., Lin, Y., Liu, M., and Gilson, M. K. (2002). The binding database: Data management and interface design. *Bioinformatics* 18 (1), 130–139. doi:10.1093/bioinformatics/18.1.130

Cherkasov, A. (2023). The 'Big Bang' of the chemical universe. *Nat. Chem. Biol.* 19, 667–668. doi:10.1038/s41589-022-01233-x

Corso, G., Stärk, H., Jing, B., et al. (2022). *DiffDock: Diffusion steps, twists, and turns for molecular docking*. arXiv [q-bio.BM]. Available at: http://arxiv.org/abs/2210.01776.

Costa, R. P. O., Lucena, L. F., Silva, L. M. A., Zocolo, G. J., Herrera-Acevedo, C., Scotti, L., et al. (2021). The SistematX web portal of natural products: An update. *J. Chem. Inf. Model.* 61 (6), 2516–2522. doi:10.1021/acs.jcim.1c00083

Davies, M., Nowotka, M., Papadatos, G., Dedman, N., Gaulton, A., Atkinson, F., et al. (2015). ChEMBL web services: Streamlining access to drug discovery data and utilities. *Nucleic acids Res.* 43 (W1), W612–W620. doi:10.1093/nar/gkv352

Dos Santos Nascimento, I. J., de Aquino, T. M., and da Silva-Júnior, E. F. (2021). Drug repurposing: A strategy for discovering inhibitors against emerging viral infections. *Curr. Med. Chem.* 28 (15), 2887–2942. doi:10.2174/0929867327666200812215852

DRUGBANK (2023). Celecoxib. Available at: https://go.drugbank.com/drugs/DB00482 (accessed May 13, 2023).

Enamine (2023). Real database. Available at: https://enamine.net/compound-collections/real-compounds/real-database (accessed May 13, 2023).

Evans, B. E., Rittle, K. E., Bock, M. G., DiPardo, R. M., Freidinger, R. M., Whitter, W. L., et al. (1988). Methods for drug discovery: Development of potent, selective, orally effective cholecystokinin antagonists. *J. Med. Chem.* 31 (12), 2235–2246. doi:10.1021/jm00120a002

Fourches, D., Muratov, E., and Tropsha, A. (2016). Trust, but verify II: A practical guide to chemogenomics data curation. *J. Chem. Inf. Model.* 56 (7), 1243–1252. doi:10.1021/acs.jcim.6b00129

Gallo, K., Kemmler, E., Goede, A., Becker, F., Dunkel, M., Preissner, R., et al. (2023). SuperNatural 3.0-a database of natural products and natural product-based derivatives. *Nucleic acids Res.* 51 (D1), D654–D659. doi:10.1093/nar/gkac1008

Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., et al. (2018). Automatic chemical design using a data-driven continuous representation of molecules. *ACS central Sci.* 4 (2), 268–276. doi:10.1021/acscentsci.7b00572

Gómez-García, A., and Medina-Franco, J. L. (2022). Progress and impact of Latin American natural product databases. *Biomolecules* 12 (9), 1202. doi:10.3390/biom12091202

Grebner, C. (2022). Webinar: "exploration and mining of large virtual chemical spaces. Available at: https://youtu.be/fMrI11SXwpU (accessed May 13, 2023).

Greener, J. G., Kandathil, S. M., Moffat, L., and Jones, D. T. (2022). A guide to machine learning for biologists. *Nat. Rev. Mol. Cell Biol.* 23 (1), 40–55. doi:10.1038/s41580-021-00407-0

Grigalunas, M., Brakmann, S., and Waldmann, H. (2022). Chemical evolution of natural product structure. *J. Am. Chem. Soc.* 144 (8), 3314–3329. doi:10.1021/jacs.1c11270

Gu, J., Gui, Y., Chen, L., Yuan, G., Lu, H. Z., and Xu, X. (2013). Use of natural products as chemical library for drug discovery and network pharmacology. *PloS one* 8 (4), e62839. doi:10.1371/journal.pone.0062839

Guo J, J., Janet, J. P., Bauer, M. R., Nittinger, E., Giblin, K. A., Papadopoulos, K., et al. (2021). DockStream: A docking wrapper to enhance de novo molecular design. *J. cheminformatics* 13 (1), 89. doi:10.1186/s13321-021-00563-7

Guo M, M., Thost, V., Li, B., et al. (2021). "Data-efficient graph grammar learning for molecular generation," in *International conference on learning representations*, 9. February 2021. Available at: https://research.ibm.com/publications/data-efficient-graph-grammar-learning-for-molecular-generation (accessed May 13, 2023).

Guo, M., Thost, V., Li, B., et al. (2022). Data-efficient graph grammar learning for molecular generation. arXiv [cs.LG]. Available at: http://arxiv.org/abs/2203.08031.

Hayes, A., and Hunter, J. (2012). Why is publication of negative clinical trial data important? *Br. J. Pharmacol.* 167 (7), 1395–1397. doi:10.1111/j.1476-5381.2012.02215.x

Hu, Q., Peng, Z., Sutton, S. C., Na, J., Kostrowicki, J., Yang, B., et al. (2012). Pfizer global virtual library (PGVL): A chemistry design tool powered by experimentally validated parallel synthesis information. *ACS Comb. Sci.* 14 (11), 579–589. doi:10.1021/co300096q

IBM (2022). How to use AI to discover new drugs and materials with limited data. Available at: https://research.ibm.com/blog/ai-discovery-with-limited-data#fnref-1 (accessed April 16, 2023).

Irwin, J. J. (2008). Community benchmarks for virtual screening. *J. computer-aided Mol. Des.* 22 (3-4), 193–199. doi:10.1007/s10822-008-9189-4

Jain, A. N., and Nicholls, A. (2008). Recommendations for evaluation of computational methods. *J. computer-aided Mol. Des.* 22 (3-4), 133–139. doi:10.1007/s10822-008-9196-5

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596 (7873), 583–589. doi:10.1038/s41586-021-03819-2

Juskalian, R., Regalado, A., Orcutt, M., et al. (2023). *10 breakthrough technologies 2020*. Available at: https://www.technologyreview.com/10-breakthrough-technologies/2020/ (accessed February 26, 2020).

Kadurin, A., Aliper, A., Kazennov, A., Mamoshina, P., Vanhaelen, Q., Khrabrov, K., et al. (2017). The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology. *Oncotarget* 8 (7), 10883–10890. doi:10.18632/oncotarget.14073

Kim, S., Chen, J., Cheng, T., Gindulyte, A., et al. (2023). PubChem 2023 update. *Nucleic acids Res.* 51 (D1), D1373–D1380. doi:10.1093/nar/gkac956

Koes, D. R., and Camacho, C. J. (2012). ZINCPharmer: Pharmacophore search of the ZINC database. *Nucleic acids Res.* 40, W409–W414. Web Server issue). doi:10.1093/nar/gks378

Korkmaz, S. (2020). Deep learning-based imbalanced data classification for drug discovery. *J. Chem. Inf. Model.* 60 (9), 4180–4190. doi:10.1021/acs.jcim.9b01162

Korn, M., Ehrt, C., Ruggiu, F., Gastreich, M., and Rarey, M. (2023). Navigating large chemical spaces in early-phase drug discovery. *Curr. Opin. Struct. Biol.* 80, 102578. doi:10.1016/j.sbi.2023.102578

Kramer, C., and Lewis, R. (2012). QSARs, data and error in the modern age of drug discovery. *Curr. Top. Med. Chem.* 12 (17), 1896–1902. doi:10.2174/156802612804547380

Krenn, M., Häse, F., Nigam, A., Friederich, P., and Aspuru-Guzik, A. (2020). Self-referencing embedded strings (SELFIES): A 100% robust molecular string representation. *Mach. Learn. Sci. Technol.* 1 (4), 045024. doi:10.1088/2632-2153/aba947

Krishnan, S. R., Bung, N., Bulusu, G., and Roy, A. (2021). Accelerating de novo drug design against novel proteins using deep learning. *J. Chem. Inf. Model.* 61 (2), 621–630. doi:10.1021/acs.jcim.0c01060

Kumar, S. A., Ananda Kumar, T. D., Beeraka, N. M., Pujar, G. V., Singh, M., Narayana Akshatha, H. S., et al. (2022). Machine learning and deep learning in data-driven decision making of drug discovery and challenges in high-quality data acquisition in the pharmaceutical industry. *Future Med. Chem.* 14 (4), 245–270. doi:10.4155/fmc-2021-0243

Leach, A. R., and Gillet, V. J. (2007). "Selecting diverse dets of compounds," in *An introduction to chemoinformatics* (Dordrecht: Springer Netherlands), 119–139. doi:10.1007/978-1-4020-6291-9_6

Li, S., Wang, L., Meng, J., Zhao, Q., Zhang, L., and Liu, H. (2022). De Novo design of potential inhibitors against SARS-CoV-2 Mpro. *Comput. Biol. Med.* 147, 105728. doi:10.1016/j.compbiomed.2022.105728

Li, Y., Zhang, L., and Liu, Z. (2018). Multi-objective de novo drug design with conditional graph generative model. *J. cheminformatics* 10 (1), 33. doi:10.1186/s13321-018-0287-6

Liang, Y., Fang, R., and Rao, Q. (2022). An insight into the medicinal chemistry perspective of macrocyclic derivatives with antitumor activity: A systematic review. *Molecules* 27 (9), 2837. doi:10.3390/molecules27092837

Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. drug Deliv. Rev.* 46 (1-3), 3–26. doi:10.1016/s0169-409x(00)00129-0

Liu, X., Ye, K., van Vlijmen, H. W. T., Ijzerman, A. P., and van Westen, G. J. P. (2019). An exploration strategy improves the diversity of de novo ligands using deep reinforcement learning: A case for the adenosine A2A receptor. *J. cheminformatics* 11 (1), 35. doi:10.1186/s13321-019-0355-6

López-López, E., Bajorath, J., and Medina-Franco, J. L. (2021a). Informatics for chemistry, biology, and biomedical sciences. *J. Chem. Inf. Model.* 61 (1), 26–35. doi:10.1021/acs.jcim.0c01301

López-López, E., Cerda-García-Rojas, C. M., and Medina-Franco, J. L. (2021b). Tubulin inhibitors: A chemoinformatic analysis using cell-based data. *Molecules* 26 (9), 2483. doi:10.3390/molecules26092483

López-López, E., Fernández-de Gortari, E., and Medina-Franco, J. L. (2022). Yes SIR! On the structure-inactivity relationships in drug discovery. *Drug Discov. today* 27 (8), 2353–2362. doi:10.1016/j.drudis.2022.05.005

López-López, E., and Medina-Franco, J. L. (2023). Towards decoding hepatotoxicity of approved drugs through navigation of multiverse and consensus chemical spaces. *Biomolecules* 13 (1), 176. doi:10.3390/biom13010176

Ma, B., Terayama, K., Matsumoto, S., Isaka, Y., Sasakura, Y., Iwata, H., et al. (2021). Structure-based de novo molecular generator combined with artificial intelligence and docking simulations. *J. Chem. Inf. Model.* 61 (7), 3304–3313. doi:10.1021/acs.jcim.1c00679

Maziarka, L., Pocha, A., Kaczmarczyk, J., Rataj, K., Danel, T., and Warchoł, M. (2020). Mol-CycleGAN: A generative model for molecular optimization. *J. cheminformatics* 12 (1), 2. doi:10.1186/s13321-019-0404-1

Medina-Franco, J. L., Flores-Padilla, E. A., and Chávez-Hernández, A. L. (2022b). "Chapter 23 - discovery and development of lead compounds from natural sources using computational approaches," in *Evidence-based validation of herbal medicine*. Editor P. K. Mukherjee Second Edition (Elsevier), 539–560. doi:10.1016/B978-0-323-85542-6.00009-3

Medina-Franco, J. L., López-López, E., Andrade, E., Ruiz-Azuara, L., Frei, A., Guan, D., et al. (2022a). Bridging informatics and medicinal inorganic chemistry: Toward a database of metallodrugs and metallodrug candidates. *Drug Discov. today* 27 (5), 1420–1430. doi:10.1016/j.drudis.2022.02.021

Medina-Franco, J. L., and López-López, E. (2022). The essence and transcendence of scientific publishing. *Front. Res. metrics Anal.* 7, 822453. doi:10.3389/frma.2022.822453

Medina-Franco, J. L., Martinez-Mayorga, K., and Meurice, N. (2014). Balancing novelty with confined chemical space in modern drug discovery. *Expert Opin. drug Discov.* 9 (2), 151–165. doi:10.1517/17460441.2014.872624

Medina-Franco, J. L., Naveja, J. J., and López-López, E. (2019). Reaching for the bright StARs in chemical space. *Drug Discov. today* 24 (11), 2162–2169. doi:10.1016/j.drudis.2019.09.013

Mendez, D., Gaulton, A., Bento, A. P., Chambers, J., De Veij, M., Félix, E., et al. (2019). ChEMBL: Towards direct deposition of bioassay data. *Nucleic acids Res.* 47 (D1), D930–D940. doi:10.1093/nar/gky1075

Mohanraj, K., Karthikeyan, B. S., Vivek-Ananth, R. P., Chand, R. P. B., Aparna, S. R., Mangalapandi, P., et al. (2018). Imppat: A curated database of indian medicinal plants, phytochemistry and therapeutics. *Sci. Rep.* 8 (1), 4329. doi:10.1038/s41598-018-22631-z

Mouchlis, V. D., Afantitis, A., Serra, A., Fratello, M., Papadiamantis, A. G., Aidinis, V., et al. (2021). Advances in de novo drug design: From conventional to machine learning methods. *Int. J. Mol. Sci.* 22 (4), 1676. doi:10.3390/ijms22041676

Mysinger, M. M., Carchia, M., Irwin, J. J., and Shoichet, B. K. (2012). Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *J. Med. Chem.* 55 (14), 6582–6594. doi:10.1021/jm300687e

Newman, D. J., and Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* 83 (3), 770–803. doi:10.1021/acs.jnatprod.9b01285

NIH (2023). Natural product libraries. Available at: https://www.nccih.nih.gov/grants/natural-product-libraries.

Niitsu, A., and Sugita, Y. (2023). Towards de novo design of transmembrane α-helical assemblies using structural modelling and molecular dynamics simulation. *Phys. Chem. Chem. Phys. PCCP* 25 (5), 3595–3606. doi:10.1039/d2cp03972a

Norinder, U., Naveja, J. J., López-López, E., Mucs, D., and Medina-Franco, J. L. (2019). Conformal prediction of HDAC inhibitors. *SAR QSAR Environ. Res.* 30 (4), 265–277. doi:10.1080/1062936X.2019.1591503

Ntie-Kang, F., Zofou, D., Babiaka, S. B., Meudom, R., Scharfe, M., Lifongo, L. L., et al. (2013). AfroDb: A select highly potent and diverse natural product library from african medicinal plants. *PloS one* 8 (10), e78085. doi:10.1371/journal.pone.0078085

Olivecrona, M., Blaschke, T., Engkvist, O., and Chen, H. (2017). Molecular de-novo design through deep reinforcement learning. *J. cheminformatics* 9 (1), 48. doi:10.1186/s13321-017-0235-x

Olmedo, A. D., and Medina-Franco, J. L. (2020). "Chemoinformatic approach: The case of natural products of Panama," in *Cheminformatics and its applications* (IntechOpen). doi:10.5772/intechopen.87779

Olmedo, D. A., González-Medina, M., Gupta, M. P., and Medina-Franco, J. L. (2017). Cheminformatic characterization of natural products from Panama. *Mol. Divers.* 21 (4), 779–789. doi:10.1007/s11030-017-9781-4

Palazzesi, F., and Pozzan, A. (2022). "Deep learning applied to ligand-based de novo drug DesignDe novo drug design," in *Artificial intelligence in drug design*. Editor A. Heifetz (New York, NY: Springer US), 273–299. doi:10.1007/978-1-0716-1787-8_12

Papadopoulos, K., Giblin, K. A., Janet, J. P., Patronov, A., and Engkvist, O. (2021). De novo design with deep generative models based on 3D similarity scoring. *Bioorg. Med. Chem.* 44, 116308. doi:10.1016/j.bmc.2021.116308

Patel, H. M., Noolvi, M. N., Sharma, P., Jaiswal, V., Bansal, S., Lohan, S., et al. (2014). Quantitative structure–activity relationship (QSAR) studies as strategic approach in drug discovery. *Med. Chem. Res. Int. J. rapid Commun. Des. Mech. action Biol. Act. agents* 23 (12), 4991–5007. doi:10.1007/s00044-014-1072-3

Perron, Q., da Silva, V. B. R., Atwood, B., and Gaston-Mathé, Y. (2022b). Key points to succeed in Artificial Intelligence drug discovery projects. *Chem. Int.* 44 (1), 19–21. doi:10.1515/ci-2022-0106

Perron, Q., Mirguet, O., Tajmouati, H., Skiredj, A., Rojas, A., Gohier, A., et al. (2022a). Deep generative models for ligand-based de novo design applied to multi-parametric optimization. *J. Comput. Chem.* 43 (10), 692–703. doi:10.1002/jcc.26826

Pilon, A. C., Valli, M., Dametto, A. C., Pinto, M. E. F., Freire, R. T., Castro-Gamboa, I., et al. (2017). NuBBEDB: An updated database to uncover chemical and biological information from Brazilian biodiversity. *Sci. Rep.* 7 (1), 7215. doi:10.1038/s41598-017-07451-x

Pilón-Jiménez, B. A., Saldívar-González, F. I., Díaz-Eufracio, B. I., and Medina-Franco, J. L. (2019). Biofacquim: A Mexican compound database of natural products. *Biomolecules* 9 (1), 31. doi:10.3390/biom9010031

Polykovskiy, D., Zhebrak, A., Sanchez-Lengeling, B., Golovanov, S., Tatanov, O., Belyaev, S., et al. (2020). Molecular sets (MOSES): A benchmarking platform for molecular generation models. *Front. Pharmacol.* 11, 565644. doi:10.3389/fphar.2020.565644

Réau, M., Langenfeld, F., Zagury, J-F., Lagarde, N., and Montes, M. (2018). Decoys selection in benchmarking datasets: Overview and perspectives. *Front. Pharmacol.* 9, 11. doi:10.3389/fphar.2018.00011

Reymond, J-L. (2015). The chemical space project. *Accounts Chem. Res.* 48 (3), 722–730. doi:10.1021/ar500432k

Rohrer, S. G., and Baumann, K. (2009). Maximum unbiased validation (MUV) data sets for virtual screening based on PubChem bioactivity data. *J. Chem. Inf. Model.* 49 (2), 169–184. doi:10.1021/ci8002649

Sabe, V. T., Ntombela, T., Jhamba, L. A., Maguire, G. E. M., Govender, T., Naicker, T., et al. (2021). Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review. *Eur. J. Med. Chem.* 224, 113705. doi:10.1016/j.ejmech.2021.113705

Saldívar-González, F. I., Aldas-Bulos, V. D., Medina-Franco, J. L., and Plisson, F. (2022). Natural product drug discovery in the artificial intelligence era. *Chem. Sci.* 13 (6), 1526–1546. doi:10.1039/d1sc04471k

Saldívar-González, F. I., and Medina-Franco, J. L. (2022). Approaches for enhancing the analysis of chemical space for drug discovery. *Expert Opin. drug Discov.* 17 (7), 789–798. doi:10.1080/17460441.2022.2084608

Saldívar-González, F. I., Valli, M., Andricopulo, A. D., da Silva Bolzani, V., and Medina-Franco, J. L. (2019). Chemical space and diversity of the NuBBE database: A chemoinformatic characterization. *J. Chem. Inf. Model.* 59 (1), 74–85. doi:10.1021/acs.jcim.8b00619

Sánchez-Cruz, N., Pilón-Jiménez, B. A., and Medina-Franco, J. L. (2019) Functional group and diversity analysis of BIOFACQUIM: A Mexican natural product database. *F1000Research* 8, Chem Inf Sci-2071. doi:10.12688/f1000research.21540.2

Scannell, J. W., Bosley, J., Hickman, J. A., Dawson, G. R., Truebel, H., Ferreira, G. S., et al. (2022). Predictive validity in drug discovery: What it is, why it matters and how to improve it. *Nat. Rev. Drug Discov.* 21 (12), 915–931. doi:10.1038/s41573-022-00552-x

Schneider, G., and Clark, D. E. (2019). Automated de novo drug design: Are we nearly there yet? *Angew. Chem.* 58 (32), 10792–10803. doi:10.1002/anie.201814681

Schneider, G., Clément-Chomienne, O., Hilfiger, L., SchneiderKirschBöhm, et al. (2000). Virtual screening for bioactive molecules by evolutionary de novo design. *Angew. Chem.* 39 (22), 4130–4133. doi:10.1002/1521-3773(20001117)39:22<4130:aid-anie4130>3.0.co;2-e

Schneider, P., and Schneider, G. (2017). Privileged structures revisited. *Angew. Chem.* 56 (27), 7971–7974. doi:10.1002/anie.201702816

Schneider, P., Walters, W. P., Plowright, A. T., Sieroka, N., Listgarten, J., Goodnow, R. A., et al. (2020). Rethinking drug design in the artificial intelligence era. *Nat. Rev. Drug Discov.* 19 (5), 353–364. doi:10.1038/s41573-019-0050-3

Scotti, M. T., Herrera-Acevedo, C., Oliveira, T. B., Costa, R. P. O., Santos, S. Y. K. d. O., Rodrigues, R. P., et al. (2018). SistematX, an online web-based cheminformatics tool for data management of secondary metabolites. *Molecules* 23 (1), 103. doi:10.3390/molecules23010103

Sheridan, R. P. (2013). Time-split cross-validation as a method for estimating the goodness of prospective prediction. *J. Chem. Inf. Model.* 53 (4), 783–790. doi:10.1021/ci400084k

Shipman, J. T., Su, X., Hua, D., and Desaire, H. (2019). DecoyDeveloper: An on-demand, de novo decoy glycopeptide generator. *J. proteome Res.* 18 (7), 2896–2902. doi:10.1021/acs.jproteome.9b00203

Simonovsky, M., and Komodakis, N. (2018). "GraphVAE: Towards generation of small graphs using variational autoencoders," in *Artificial neural networks and machine learning – icann 2018* (Springer International Publishing), 2018, 412–422. doi:10.1007/978-3-030-01418-6_41

Skalic, M., Jiménez, J., Sabbadin, D., and De Fabritiis, G. (2019b). Shape-based generative modeling for de novo drug design. *J. Chem. Inf. Model.* 59 (3), 1205–1214. doi:10.1021/acs.jcim.8b00706

Skalic, M., Sabbadin, D., Sattarov, B., Sciabola, S., and De Fabritiis, G. (2019a). From target to drug: Generative modeling for the multimodal structure-based ligand design. *Mol. Pharm.* 16 (10), 4282–4291. doi:10.1021/acs.molpharmaceut.9b00634

Soares, T. A., Nunes-Alves, A., Mazzolari, A., Ruggiu, F., Wei, G. W., and Merz, K. (2022). The (Re)-evolution of quantitative structure-activity relationship (qsar) studies propelled by the surge of machine learning methods. *J. Chem. Inf. Model.* 62 (22), 5317–5320. doi:10.1021/acs.jcim.2c01422

Sorokina, M., Merseburger, P., Rajan, K., Yirik, M. A., and Steinbeck, C. (2021). COCONUT online: Collection of open natural products database. *J. cheminformatics* 13 (1), 2. doi:10.1186/s13321-020-00478-9

Tingle, B. I., Tang, K. G., Castanon, M., Gutierrez, J. J., Khurelbaatar, M., Dandarchuluun, C., et al. (2023). ZINC-22─A free multi-billion-scale database of tangible compounds for ligand discovery. *J. Chem. Inf. Model.* 63 (4), 1166–1176. doi:10.1021/acs.jcim.2c01253

Tong, X., Liu, X., Tan, X., Jiang, J., Xiong, Z., et al. (2021). Generative models for de novo drug design. *J. Med. Chem.* 64 (19), 14011–14027. doi:10.1021/acs.jmedchem.1c00927

Ullanat, V. (2020). "Variational autoencoder as a generative tool to produce de-novo lead compounds for biological targets," in *2020 14th international conference on innovations in information Technology (IIT)*, 102–107. doi:10.1109/IIT50501.2020.9299078

UNIIQUIM (2015). Uniiquim. Available at: https://uniiquim.iquimica.unam.mx/ (accessed May 13, 2023).

Valli, M., dos Santos, R. N., Figueira, L. D., Nakajima, C. H., Castro-Gamboa, I., Andricopulo, A. D., et al. (2013). Development of a natural products database from the biodiversity of Brazil. *J. Nat. Prod.* 76 (3), 439–444. doi:10.1021/np3006875

Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., et al. (2019). Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* 18 (6), 463–477. doi:10.1038/s41573-019-0024-5

Veber, D. F., Johnson, S. R., Cheng, H-Y., Smith, B. R., Ward, K. W., and Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* 45 (12), 2615–2623. doi:10.1021/jm020017n

Wang, L., Pang, X., Li, Y., Zhang, Z., and Tan, W. (2017). Rader: A RApid DEcoy retriever to facilitate decoy based assessment of virtual screening. *Bioinformatics* 33 (8), 1235–1237. doi:10.1093/bioinformatics/btw783

Wang, M., Hsieh, C-Y., Wang, J., Wang, D., Weng, G., Shen, C., et al. (2022). Relation: A deep generative model for structure-based de novo drug design. *J. Med. Chem.* 65 (13), 9478–9492. doi:10.1021/acs.jmedchem.2c00732

Warr, W. A., Nicklaus, M. C., Nicolaou, C. A., and Rarey, M. (2022). Exploration of ultralarge compound collections for drug discovery. *J. Chem. Inf. Model.* 62 (9), 2021–2034. doi:10.1021/acs.jcim.2c00224

Warr, W. (2021). Report on an NIH workshop on ultralarge chemistry databases. Chemrxiv: 43. Available at: https://chemrxiv.org/engage/chemrxiv/article-details/60c75883bdbb89984ea3ada5.

Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* 28 (1), 31–36. doi:10.1021/ci00057a005

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic acids Res.* 46 (D1), D1074–D1082. doi:10.1093/nar/gkx1037

Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., et al. (2008). DrugBank: A knowledgebase for drugs, drug actions and drug targets. *Nucleic acids Res.* 36, D901–D906. doi:10.1093/nar/gkm958

Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., et al. (2006). DrugBank: A comprehensive resource for *in silico* drug discovery and exploration. *Nucleic acids Res.* 34, D668–D672. doi:10.1093/nar/gkj067

Wu, A., Ye, Q., Zhuang, X., Chen, Q., Zhang, J., Wu, J., et al. (2023a). Elucidating structures of complex organic compounds using a machine learning model based on the 13C NMR chemical shifts. *Precis. Chem.* 1 (1), 57–68. doi:10.1021/prechem.3c00005

Wu, J., Xiao, Y., Cai, H., Zhao, D., Li, Y., et al. (2023b). DeepCancerMap: A versatile deep learning platform for target- and cell-based anticancer drug discovery. *Eur. J. Med. Chem.* 255, 115401. doi:10.1016/j.ejmech.2023.115401

Wu, Z., Ramsundar, B., Feinberg, E. N., Gomes, J., Geniesse, C., Pappu, A. S., et al. (2018). MoleculeNet: A benchmark for molecular machine learning. *Chem. Sci.* 9 (2), 513–530. doi:10.1039/c7sc02664a

Xie, W., Wang, F., Li, Y., Lai, L., and Pei, J. (2022). Advances and challenges in de novo drug design using three-dimensional deep generative models. *J. Chem. Inf. Model.* 62 (10), 2269–2279. doi:10.1021/acs.jcim.2c00042

Yang, J., Wang, D., Jia, C., Wang, M., Hao, G., and Yang, G. (2019). Freely accessible chemical database resources of compounds for *in silico* drug discovery. *Curr. Med. Chem.* 26 (42), 7581–7597. doi:10.2174/0929867325666180508100436

Yang, X., Yang, G., and Chu, J. (2023). *The balanced matrix factorization for computational drug repositioning*. arXiv [cs.CE]. Available at: http://arxiv.org/abs/2301.06448.

Yu, H. (2021). Responsible use of negative research outcomes-accelerating the discovery and development of new antibiotics. *J. antibiotics* 74 (9), 543–546. doi:10.1038/s41429-021-00439-w

Zhang, Y., Luo, M., Wu, P., Wu, S., Lee, T. Y., and Bai, C. (2022). Application of computational biology and artificial intelligence in drug design. *Int. J. Mol. Sci.* 23 (21), 13568. doi:10.3390/ijms232113568

Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., et al. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* 37 (9), 1038–1040. doi:10.1038/s41587-019-0224-x

frontiers | Frontiers in Drug Discovery

TYPE Mini Review
PUBLISHED 12 July 2023
DOI 10.3389/fddsv.2023.1226727

OPEN ACCESS

EDITED BY
Bruno Villoutreix,
Institut National de la Santé et de la Recherche Médicale (INSERM), France

REVIEWED BY
Rodolpho C. Braga,
InsilicAll, Brazil
Mahesh Koirala,
Akttyva Therapeutics, Inc., United States

*CORRESPONDENCE
Wei Gu,
wei.gu@elixir-luxembourg.org

RECEIVED 22 May 2023
ACCEPTED 30 June 2023
PUBLISHED 12 July 2023

CITATION
Gadiya Y, Ioannidis V, Henderson D, Gribbon P, Rocca-Serra P, Satagopam V, Sansone S-A and Gu W (2023), FAIR data management: what does it mean for drug discovery?
Front. Drug Discov. 3:1226727.
doi: 10.3389/fddsv.2023.1226727

FAIR data management: what does it mean for drug discovery?

Yojana Gadiya[1,2,3], Vassilios Ioannidis[4,5], David Henderson[6], Philip Gribbon[1,2], Philippe Rocca-Serra[7,8], Venkata Satagopam[9], Susanna-Assunta Sansone[7] and Wei Gu[9,10]*

[1]Fraunhofer Institute for Translational Medicine and Pharmacology (ITMP), Hamburg, Germany, [2]Fraunhofer Cluster of Excellence for Immune-Mediated Diseases (CIMD), Frankfurt, Germany, [3]Bonn-Aachen International Center for Information Technology (B-IT), University of Bonn, Bonn, Germany, [4]Vital-IT Group, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland, [5]UNIRIS, University of Lausanne, Lausanne, Switzerland, [6]Bayer AG, Business Development & Licensing & OI, Pharmaceuticals, Berlin, Germany, [7]Oxford e-Research Centre, Department of Engineering Science, University of Oxford, Oxford, United Kingdom, [8]AstraZeneca, Data Office, Data Science and AI Unit R&D, Cambridge, United Kingdom, [9]Luxembourg Centre for Systems Biomedicine, ELIXIR Luxembourg, University of Luxembourg, Esch-sur-Alzette, Luxembourg, [10]Luxembourg National Data Service, Esch-sur-Alzette, Luxembourg

The drug discovery community faces high costs in bringing safe and effective medicines to market, in part due to the rising volume and complexity of data which must be generated during the research and development process. Fully utilising these expensively created experimental and computational data resources has become a key aim of scientists due to the clear imperative to leverage the power of artificial intelligence (AI) and machine learning-based analyses to solve the complex problems inherent in drug discovery. In turn, AI methods heavily rely on the quantity, quality, consistency, and scope of underlying training data. While pre-existing preclinical and clinical data cannot fully replace the need for *de novo* data generation in a project, having access to relevant historical data represents a valuable asset, as its reuse can reduce the need to perform similar experiments, therefore avoiding a "reinventing the wheel" scenario. Unfortunately, most suitable data resources are often archived within institutes, companies, or individual research groups and hence unavailable to the wider community. Hence, enabling the data to be Findable, Accessible, Interoperable, and Reusable (FAIR) is crucial for the wider community of drug discovery and development scientists to learn from the work performed and utilise the findings to enhance comprehension of their own research outcomes. In this mini-review, we elucidate the utility of FAIR data management across the drug discovery pipeline and assess the impact such FAIR data has made on the drug development process.

KEYWORDS

drug discovery, FAIR principles, data management, data sharing, machine learning

Introduction

Ensuring effective exploitation of experimental and computational data resources is a major issue within the drug discovery community, which faces rising costs in bringing safe and effective medicines to market. As part of the search for new medicines, large amounts of data are generated in order to support decision-making on the efficacy, safety, and developability of a potential new drug as it progresses along the discovery pipeline. These new data are generated on a daily basis as a part of *in silico*, laboratory, or clinical studies, and the high cost incurred directly impacts the overall capacity of the

pharmaceutical and biotech industries to bring treatments to the clinic. The average cost of research and development (R&D) to bring a new drug to market is estimated to be around 900 million to 2.8 billion dollars (Wouters et al., 2020; Simoens and Huys, 2021). Research expenditure is eventually transferred to the price of treatments and represents a significant part of healthcare spending. To add a further burden, in recent years, the volume and complexity of data generated by scientists involved in research and development have increased exponentially, creating what has been termed a "Big Data" challenge. This has followed the increased adoption of large-scale automated experimentation methods. For example, it is routine to sequence cancer patients' tumour biopsies to identify which specific genetic mutations are associated with their individual tissue malignancies. As part of drug research efforts, these same tumour-derived tissues can then be analysed using powerful high-resolution imaging microscopes to help identify prototype drugs which kill the tumour cells and have the potential to be further developed into new medicines. The challenge scientists now must face in the light of economic constraints is to make the data which has been expensively generated within their studies reusable so that the entire community has the chance to learn from the work performed and, ideally, apply the results to understand the results of their own studies better. It is far more cost-effective to reuse well-validated results from a trusted database rather than repeat the same experimental study again. This situation has led to the previously "un-exciting" process of data management becoming increasingly important in drug discovery, as it directly supports the use of artificial intelligence (AI) and machine learning (ML) based analyses. Such advanced analyses are highly dependent on the quality, consistency, and scope of the training data upon which predictive models are built. In situations where effective data management and quality assessments are not prioritised, then there is a risk of low-quality, poorly controlled or out-of-scope training data emerging, which in the worst case can lead to a counter-productive "garbage-in garbage-out" scenario.

The costs associated with data generation are distributed across the pre-clinical and clinical stages of drug discovery. In the preclinical stage, complex and diverse data are generated, mainly on cellular or *in-vivo* models, to establish the development and toxicity profile of potential drug candidates. In clinical stages, where the major costs of a development programme are incurred, drug candidates are tested for safety and then efficacy in humans, resulting in large amounts of electronic health record-type data. These clinical trial data may be simple numerical results, for example, the level of a diagnostic marker in a blood sample, or highly complex data, which require additional analysis tools such as a low-dose CT image of a patient's lung. Although existing preclinical and clinical data cannot fully replace the need to generate new data in clinical trials, especially when developing a new drug that has not been tested in the clinic before, they are very valuable as they can help to reduce the need to perform redundant research. An additional potential strategy is the usage of "virtual clinical cohorts", created based on information in electronic health records (Tan et al., 2021). Electronically assembled cohorts can act as placebo or control arms in both Phase 2 and 3 trials (wherein the drug is administered to a larger diseased population and observed for long-term effects) creating a situation where all trial participants have the chance to benefit from the therapeutic, as well as reducing the total number of individuals involved. At this point, it is important to highlight that up to 90% of the cost of bringing a drug to market is incurred when conducting clinical trials. In most cases, these cannot be replaced by accessing existing data because the drug being developed is novel and has not been in the clinic previously, rather, the existing data can enable directed decision-making for novel drugs (for, e.g., drugs with active scaffolds). Nevertheless, it has been estimated that the availability of high-quality data could reduce the capitalised R&D costs by about 200 million dollars for each new drug brought to the clinic (Simoens and Huys, 2021). On the other hand, it has been estimated that a high quality data platform in neurology could bring more efficient research and development of new drugs with an annual value of 2.8 billion dollars (https://www.mckinsey.com/industries/life-sciences/our-insights/better-data-for-better-therapies-the-case-for-building-health-data-platforms).

Despite the value represented by large data resources, many are often archived within institutes, companies, or individual research groups and hence effectively unavailable to the wider community. As a consequence, they are in practice "invisible" to the wider community and in some cases even divisions within the same company. This leads to the need for data to be Findable, Accessible, Interoperable, and Reusable (FAIR) (Wilkinson et al., 2016). Each FAIR aspect can be tackled individually. Associating standardised metadata (i.e., information that describes the data) to globally unique and persistent identifiers can then readily ensure the findability of the data it describes. Data needs to be accessible and should be made available via repositories (which are storage spaces for researchers to deposit data sets associated with their research) with a clearly-defined access protocol potentially integrating an authentication and authorisation procedure to control access. Overall FAIR data should be "as open as possible and as close as necessary" (Collins et al., 2018): "open" in order to foster the reusability, or, if relevant, "closed" to safeguard the privacy of the information. This is very important for commercial organisations seeking to generate intellectual property, as they can protect their data and control its sharing for instance during a patent deposition or for collaborations (van Vlijmen, 2020). Similarly, it is important to protect sensitive personal data, such as patients' medical records and to ensure compliance with data protection regulations. Then is the interoperability factor, which involves adopting standards using consistent models, formats, dictionaries (ontologies) and vocabularies for the terms and documentation of the data, including the methods used to generate the data. Several standards exist with their applicability to the Life Sciences (https://fairsharing.org/search?fairsharingRegistry=Standard).
Failure to ensure data are interoperable can lead to extensive time and resource expenditure since additional curation must occur before data can be used. Finally, information about the restrictions defined in consent, local and international laws and rules, or user licences for the data collected ensures that a firm legal framework exists to support the eventual reuse of the data by others. Academic and industry research groups have acknowledged the need to drive reusability and have adopted changes to working practices, for example, collaborating with scientific journals to implement better documentation and deposition of research data in public repositories (McNutt, 2014; van Vlijmen, 2020). Furthermore, pharmaceutical industries have adopted data

standards aligned with FAIR principles to strengthen cross-collaborations with academic and industry partners in the research years. Roche and AstraZeneca have provided a holistic overview of their FAIRification pipelines alongside their downstream impact (Harrow et al., 2022). Despite these efforts, there's still a considerable need to regularly improve the state of FAIR data (Begley and Ioannidis, 2015; Baker, 2016). This simply indicates that FAIR is a journey and needs to be re-visited at specific time points during data evolution to ensure the data follows a FAIR path as addressed by Harrow et al. (2022).

In the following part of this mini-review, we will illustrate with examples the application of FAIR data at various stages within the drug discovery pipeline, starting from the preclinical through to the clinical stages. Beyond these applications, FAIR data is a valuable resource supporting research across multiple scientific and non-scientific fields.

Preclinical applicability of FAIR data

As mentioned, large efforts have been initiated to organise and structure data commonly used in research and development. These involve the establishment of large-scale open-source repositories such as UniProt (UniProt Consortium, 2023) which reports data related to the proteins potentially involved in disease processes, ChEMBL (Gaulton et al., 2012) which includes results on drug-like compounds which are investigated in the early discovery phase, and SureChEMBL (Papadatos et al., 2016) which covers patent-related data. Such repositories serve two main functions within the FAIR context: first, the formalisation of a structure for storing domain-specific information, and second, the open source feature of the repositories allow researchers across the globe to store, access, and interpret the underlying data. As machine-readable and interpretable resources, the data stored in these repositories can become training data for advanced machine algorithms such as artificial intelligence (AI). A compelling example of the impact of data reuse is provided by AlphaFold, an AI model developed by DeepMind (Jumper et al., 2021). The model can predict protein 3-D organisation, thus expanding the repertoire of knowledge from the existing "known" protein structures (which had been solved experimentally) to now include previously "unknown" protein structures. In the drug discovery field, such predictive models play a role in identifying protein-protein and drug-protein interactions that contribute to our understanding of how drugs act at a molecular level. An important aspect of such modelling systems is that they allow computational assessment of the binding efficiency of a molecule to a protein of interest for which an experimentally derived 3-D structure is not available. This can save costs when identifying new compounds which bind proteins and also creates new ways to help understand how the function of the protein can be modulated to change a disease process in a beneficial way. The model owes its success to the presence of open-access and FAIR data repositories and infrastructures. AlphaFold has been trained on data available in UniProt for sequence-based similarity and Protein Data Bank (PDB) for computation of the 3D structure of the model (Berman et al., 2000). Without such repositories supported by machine-interpretable data formats, the training and building of a groundbreaking AI model such as AlphaFold would not have been possible.

It is, unfortunately, the case that only a limited subset of data in the drug discovery field is FAIR and efforts to mobilise the community to implement FAIR-compliant systems need to be initiated (Wise et al., 2019). One prominent effort leading the way in bringing FAIR into practice is the IMI Innovative Medicines Initiative (IMI) FAIRplus project (https://fairplus-project.eu/). FAIRplus was established with the aim to generate reproducible workflows for data FAIRification in the life science field and promoting the FAIR principles among academic and industrial researchers. One project, focussed on reducing drug-associated toxicology, is a useful example of how FAIR data can be leveraged to enable automated downstream tasks. For each potential compound, toxicity data associated with specific chemical structural features can be identified and act as a guide when designing novel compounds with fewer or less acute safety issues. Acknowledging the importance of effectively reusing toxicology data, the project IMI eTOX (http://www.etoxproject.eu/) was established. Within eTOX, a database of preclinical toxicity data from participating pharmaceutical companies was created. After the completion of the project, the FAIR pipelines built by IMI FAIRplus for eTOX were provided to the IMI eTRANSAFE project for further reuse (Custers et al., 2021). Similarly, the IMI CARE project was initiated in response to the COVID-19 pandemic, and as part of the project, ~5,500 FDA-approved drugs and clinical candidates were screened *in vitro* for anti-SARS-CoV-2 activity. Therefore, IMI FAIRplus project assisted in disseminating these data into the ChEMBL public repository (Custers et al., 2022). While these data did not lead to the discovery of an eligible compound for further development to treat COVID-19, they are still very valuable information for informing community-wide COVID-19 drug development efforts. The eTRANSAFE (https://etransafe.eu/) project also developed predictive models for translational clinical research. A common tool, known as FLAME, was published in the project, which reused the bioactivity data within ChEMBL and assisted in activity prediction, specifically toxicity, for compound libraries of interest (Pastor et al., 2021). A key advantage of the tool is its ability to be repurposed for datasets not available in public repositories, such as in-house pharmaceutical company databases (Steger-Hartmann et al., 2018; Sanz et al., 2023). Thus, researchers can re-use the tool for proprietary data by simply harmonising the data format for in-house generated bioassay data to a ChEMBL-compliant format.

FAIR data in clinical studies

During the latter clinical phases of drug development, testing of candidate drugs in patients is done to assess the drug's efficacy for the intended indication. Furthermore, an investigational drug's short- and long-term effects are measured to confirm the safety and tolerability profile of the drug. A recently proposed alternative approach to the design of clinical trials involves generating synthetic patients in the form of virtual cohorts. Such virtual cohorts can represent the diverse human population that differs across ethnicity, anatomy, genetics, environmental, and lifestyle factors, and can be constructed using access to standardised, anonymised FAIR clinical

data. Of particular utility is the potential to replace control cohort participants in trials, patients who normally receive a placebo or comparator drug treatments (Azizi et al., 2021). This diverse population representation allows for two significant advantages: first, the ability to evaluate virtually large patient groups irrespective of geographic location or condition; second, it is relatively cost-efficient since analyses are computational in nature.

Two fundamental ingredients are needed to generate useful synthetic data that can mimic the features of a real dataset: advanced algorithms/methods and access to high-quality clinical data and healthcare records. Many ML-based methods have been derived for the method aspect, acknowledging the interest of the drug discovery industry in synthetic patient generators. Models such as Synthea (Walonoski et al., 2018) and SASC (Khorchani et al., 2022) leverage statistical rules defined on real-world healthcare data to generate the synthetic patient cohort. On the other hand, deep neural network-based models like autoencoder-based VAMBN (Gootjes-Dreesbach et al., 2020) or an agent-based simulation model (Popper et al., 2021) have accelerated the field with virtual patient simulation being closer to the real patient. With respect to the data ingredient, resources have been built towards different types of data related to biomedical research. The clinicaltrials.gov is a large open-access database for clinical trial data. The European Health Data & Evidence Network (EHDEN, www.ehden.eu) has built a federated network to enable FAIRness of electronic health record data. A broader list of synthetic data resources has been summarised in the FAIR Cookbook (https://w3id.org/faircookbook/FCB069). Overall, there are ongoing efforts to improve and automate the process of cohort generation, given the benefits which can be accrued in terms of flexibility to share virtual clinical data, lower costs, and reduced data privacy needs relative to real-world clinical data. In summary, it is essential to note that although synthetic data is closed aligned with FAIR principles (given its seamless data sharing and reuse without infringing on privacy), the importance of this data is mainly in building ML/AI algorithms that can mimic real-world scenarios. Consortiums like Common Infrastructure for National Cohorts in Europe, Canada, and Africa (CINECA, https://www.cineca-project.eu/) have aligned their mission in this direction.

FAIR data and drug repurposing

In the scenarios discussed above, we have examined the role played by the analysis of FAIR data in the classical drug discovery process, in which the goal is the identification of new drug candidates for the disease in question. Equally, however, re-use of data can be applied in the search among existing marketed drugs for new therapeutic purposes. This approach is referred to as "drug repurposing" or "repositioning" and is of particular interest in the search for treatment for rare diseases, where the very small number of patients hampers the conduct of clinical trials (Whicher et al., 2018; Pushpakom et al., 2019). The identification of repurposed drugs is supported by resources such as the Drug Repurposing Hub (https://clue.io/repurposing) that comprehensively aggregate pre-clinical and clinical data to assist in decision-making (Corsello et al., 2017). Furthermore, the open resources with curated data generated during pre-clinical and clinical drug discovery pipelines like Open Targets (Koscielny et al., 2017), SureCHEMBL (Papadatos et al., 2016), PubChem (Kim et al., 2016), allow for tools such as Swiss Target Prediction (Gfeller et al., 2014), COVID-19 Pharamcome (Schultz et al., 2021), Patent EnrichMent Tool PEMT (Gadiya et al., 2023a), and others to access, extract, evaluate, and predict patterns in the underlying data. The COVID-19 Pharamcome based approach by Schultz et al. (2021) enabled the integration of existing data (both literature and experimental) on Sars-CoV-2 allowing for the identification of synergistic drug combinations like remdesivir-thioguanosine and nelfinavir-raloxifene. On the other hand, the applicability of the PEMT tool by Gadiya et al. (2023B) focused on retrospective analysis of patent documents to identify the reasoning behind existing drug repurposing cases like Cleave Biosciences's CB-5083, from cancer to rare diseases, for its target specificity. Both these approaches emphasise the significance of adopting a legacy data perspective to inform future decisions in drug discovery. Furthermore, these endeavours have garnered recognition from European communities resulting in the launch of drug repurposing initiatives such as REPO4EU (https://repo4.eu/) and REMEDI4ALL (https://remedi4all.org/).

Discussion

There is an urgent need to lower the costs and accelerate the process of drug discovery. To help achieve these necessary improvements, access to FAIR data can make a major contribution to community-wide learning of lessons from past failures and successes. FAIR data can also support ML predictions based on well-curated findings from past experiences. The increased adoption of ML methods also drives the further adoption of FAIR principles. FAIR data management involves ensuring that data is easily located, accessible to all who need it (and by machines/automated access and analyses), structured in a way that allows it to be used with other data, and accompanied by sufficient metadata to make it understandable and interpretable. The implementation of FAIR principles in data management comes with an initial cost but has the potential to significantly accelerate scientific discovery by enabling the effective use of data across a range of domains and disciplines. Given the benefits of following the FAIR data principles, it is clear that the effort of making data FAIR is considerable. Any attempt to implement FAIR should be carefully planned, and its benefits should be evaluated prior to starting. Obstacles, as well as potential solutions or strategies to overcome them, have been reviewed in recent works (Gu et al., 2021; Alharbi et al., 2022). Through the journey of making data FAIR, maintaining a close watch on FAIR pipelines' reusability is always encouraged by FAIR Doers. This has led to the establishment of practical recipes on how to implement FAIR in practices such as the FAIR Cookbook (https://faircookbook.elixir-europe.org, Rocca-Serra et al., 2023) created by biopharmaceutical and academic professionals and guidance on data management practices, such as the RDMKit (https://rdmkit.elixir-europe.org/); both community-driven resources welcome contributions from knowledgeable individuals to share examples and showcase resources that help researchers in their FAIR journey.

Author contributions

YG, VI, DH, PG, and WG contributed to the conception, refinement of the framework and wrote the manuscript. PR-S, VS, and S-AS, along with the other authors, critically revised the paper for intellectual content and approved the final version of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The authors received funding from the Innovative Medicines Initiative Joint Undertaking (FAIRplus project, grant agreement number IMI 802750).

Acknowledgments

The authors thank reviewers for their comments and suggestions on the manuscript.

Conflict of interest

The authors PR-S, VPS, S-AS, and WG declared that they were editorial board members of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision. Author DH was employed by the Company Bayer AG. Author PR-S was employed by the company AstraZeneca, Data Office, Data Science and AI Unit R&D.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Alharbi, E., Gadiya, Y., Henderson, D., Zaliani, A., Delfin-Rossaro, A., Cambon-Thomsen, A., et al. (2022). Selection of data sets for FAIRification in drug discovery and development: Which, why, and how? *Drug Discov. today* 27, 2080–2085. doi:10.1016/j.drudis.2022.05.010

Alharbi, E., Skeva, R., Juty, N., Jay, C., and Goble, C. (2021). Exploring the current practices, costs and benefits of FAIR implementation in pharmaceutical research and development: A qualitative interview study. *Data Intell.* 3 (4), 507–527. doi:10.1162/dint_a_00109

Azizi, Z., Zheng, C., Mosquera, L., Pilote, L., and El Emam, K.GOING-FWD Collaborators (2021). Can synthetic data be a proxy for real clinical trial data? A validation study. *BMJ open* 11 (4), e043497. doi:10.1136/bmjopen-2020-043497

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature* 533 (7604), 452–454. doi:10.1038/533452a

Begley, C. G., and Ioannidis, J. P. (2015). Reproducibility in science: Improving the standard for basic and preclinical research. *Circulation Res.* 116 (1), 116–126. doi:10.1161/CIRCRESAHA.114.303819

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic acids Res.* 28 (1), 235–242. doi:10.1093/nar/28.1.235

Collins, S., Genova, F., Harrower, N., Hodson, S., Jones, S., Laaksonen, L., et al. (2018). Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data. Available at: https://op.europa.eu/s/yBY0.

Corsello, S. M., Bittker, J. A., Liu, Z., Gould, J., McCarren, P., Hirschman, J. E., et al. (2017). The drug repurposing Hub: A next-generation drug library and information resource. *Nat. Med.* 23 (4), 405–408. doi:10.1038/nm.4306

Custers, Ilse, Boutsma, Erwin, Boiten, Jan-Willem, Duyndam, Alexander, Xu, Fuqi, Juty, Nick, et al. (2022). FAIRplus use case IMI CARE: Quick-response COVID-19 effort opens FAIR data on ~5,500 compounds. *Zenodo*. doi:10.5281/zenodo.7441699

Custers, I., Weitenberg, E., Duyndam, A., Boiten, J. W., Pérez, S., XèniaWillighagen, E., et al. (2021). FAIRplus: eTOX case study - opening up toxicology data about candidate drugs. *Zenodo*. doi:10.5281/zenodo.5786675

Gadiya, Y., Gribbon, P., Hofmann-Apitius, M., and Zaliani, A. (2023b). Pharmaceutical patent landscaping: A novel approach to understand patents from the drug discovery perspective. *Artif. Intell. Life Sci.* 3, 100069. doi:10.1016/j.ailsci.2023.100069

Gadiya, Y., Zaliani, A., Gribbon, P., and Hofmann-Apitius, M. (2023a). Pemt: A patent enrichment tool for drug discovery. *Bioinformatics* 39 (1), btac716. doi:10.1093/bioinformatics/btac716

Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., et al. (2012). ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic acids Res.* 40 (D1), D1100–D1107. doi:10.1093/nar/gkr777

Gfeller, D., Grosdidier, A., Wirth, M., Daina, A., Michielin, O., and Zoete, V. (2014). SwissTargetPrediction: A web server for target prediction of bioactive small molecules. *Nucleic acids Res.* 42 (W1), W32–W38. doi:10.1093/nar/gku293

Gootjes-Dreesbach, L., Sood, M., Sahay, A., Hofmann-Apitius, M., and Fröhlich, H. (2020). Variational autoencoder modular Bayesian networks for simulation of heterogeneous clinical study data. *Front. big Data* 3, 16. doi:10.3389/fdata.2020.00016

Gu, W., Hasan, S., Rocca-Serra, P., and Satagopam, V. P. (2021). Road to effective data curation for translational research. *Drug Discov. Today* 26 (3), 626–630. doi:10.1016/j.drudis.2020.12.007

Harrow, I., Balakrishnan, R., McGinty, H. K., Plasterer, T., and Romacker, M. (2022). Maximizing data value for biopharma through FAIR and quality implementation: FAIR plus Q. *Drug Discov. Today* 27 (5), 1441–1447. doi:10.1016/j.drudis.2022.01.006

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596 (7873), 583–589. doi:10.1038/s41586-021-03819-2

Khorchani, T., Gadiya, Y., Witt, G., Lanzillotta, D., Claussen, C., and Zaliani, A. (2022). Sasc: A simple approach to synthetic cohorts for generating longitudinal observational patient cohorts from COVID-19 clinical data. *Patterns* 3 (4), 100453. doi:10.1016/j.patter.2022.100453

Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., et al. (2016). PubChem substance and compound databases. *Nucleic acids Res.* 44 (D1), D1202–D1213. doi:10.1093/nar/gkv951

Koscielny, G., An, P., Carvalho-Silva, D., Cham, J. A., Fumis, L., Gasparyan, R., et al. (2017). Open targets: A platform for therapeutic target identification and validation. *Nucleic acids Res.* 45 (D1), D985–D994. doi:10.1093/nar/gkw1055

McNutt, M. (2014). Journals unite for reproducibility. *Science* 346 (6210), 679. doi:10.1126/science.aaa1724

Papadatos, G., Davies, M., Dedman, N., Chambers, J., Gaulton, A., Siddle, J., et al. (2016). SureChEMBL: A large-scale, chemically annotated patent document database. *Nucleic acids Res.* 44 (D1), D1220–D1228. doi:10.1093/nar/gkv1253

Pastor, M., Gómez-Tamayo, J. C., and Sanz, F. (2021). Flame: An open source framework for model development, hosting, and usage in production environments. *J. Cheminformatics* 13 (1), 31–15. doi:10.1186/s13321-021-00509-z

Popper, N., Zechmeister, M., Brunmeir, D., Rippinger, C., Weibrecht, N., Urach, C., et al. (2021). Synthetic reproduction and augmentation of COVID-19 case reporting data by agent-based simulation. *Data Sci. J.* 20 (1), 16. doi:10.5334/dsj-2021-016

Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., et al. (2019). Drug repurposing: Progress, challenges and recommendations. *Nat. Rev. Drug Discov.* 18 (1), 41–58. doi:10.1038/nrd.2018.168

Rocca-Serra, P., Gu, W., Ioannidis, V., Abbassi-Daloii, T., Capella-Gutierrez, S., Chandramouliswaran, I., et al. (2023). The FAIR Cookbook - the essential resource for and by FAIR doers. *Sci. data* 10, 292. doi:10.1038/s41597-023-02166-3

Sanz, F., Pognan, F., Steger-Hartmann, T., Díaz, C., Asakura, S., Amberg, A., et al. (2023). eTRANSAFE: data science to empower translational safety assessment. *Nat. Rev. Drug Discov*. doi:10.1038/d41573-023-00099-5

Schultz, B., Zaliani, A., Ebeling, C., Reinshagen, J., Bojkova, D., Lage-Rupprecht, V., et al. (2021). A method for the rational selection of drug repurposing candidates from multimodal knowledge harmonization. *Sci. Rep.* 11 (1), 11049. doi:10.1038/s41598-021-90296-2

Simoens, S., and Huys, I. (2021). R&D costs of new medicines: A landscape analysis. *Front. Med.* 8, 760762. doi:10.3389/fmed.2021.760762

Steger-Hartmann, T., and Pognan, F. (2018). Improving the safety assessment of chemicals and drug candidates by the integration of bioinformatics and chemoinformatics data. *Basic & Clin. Pharmacol. Toxicol.* 123, 29–36. doi:10.1111/bcpt.12956

Tan, K., Bryan, J., Segal, B., Bellomo, L., Nussbaum, N., Tucker, M., et al. (2022). Emulating control arms for cancer clinical trials using external cohorts created from electronic health record-derived real-world data. *Clin. Pharmacol. Ther.* 111 (1), 168–178. doi:10.1002/cpt.2351

The UniProt Consortium (2023). UniProt: The universal protein knowledgebase in 2023. *Nucleic Acids Res.* 51 (D1), D523–D531. doi:10.1093/nar/gkac1052

van Vlijmen, H., Mons, A., Waalkens, A., Franke, W., Baak, A., Ruiter, G., et al. (2020). The need of industry to go FAIR. *Data Intell.* 2 (1-2), 276–284. doi:10.1162/dint_a_00050

Walonoski, J., Kramer, M., Nichols, J., Quina, A., Moesel, C., Hall, D., et al. (2018). Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *J. Am. Med. Inf. Assoc.* 25 (3), 230–238. doi:10.1093/jamia/ocx079

Whicher, D., Philbin, S., and Aronson, N. (2018). An overview of the impact of rare disease characteristics on research methodology. *Orphanet J. rare Dis.* 13 (1), 14–12. doi:10.1186/s13023-017-0755-5

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. data* 3 (1), 160018–160019. doi:10.1038/sdata.2016.18

Wise, J., de Barron, A. G., Splendiani, A., Balali-Mood, B., Vasant, D., Little, E., et al. (2019). Implementation and relevance of FAIR data principles in biopharmaceutical R&D. *Drug Discov. today* 24 (4), 933–938. doi:10.1016/j.drudis.2019.01.008

Wouters, O. J., McKee, M., and Luyten, J. (2020). Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *Jama* 323 (9), 844–853. doi:10.1001/jama.2020.1166

frontiers | Frontiers in Drug Discovery

TYPE Mini Review
PUBLISHED 30 November 2023
DOI 10.3389/fddsv.2023.1314077

OPEN ACCESS

EDITED BY
Bruno Villoutreix,
Institut National de la Santé et de la Recherche Médicale (INSERM), France

REVIEWED BY
Taisiia Feoktistova,
Merck, United States
Daniela Trisciuzzi,
University of Bari Aldo Moro, Italy

*CORRESPONDENCE
Michelle W. Y. Southey,
michelle.southey@evotec.com

RECEIVED 10 October 2023
ACCEPTED 14 November 2023
PUBLISHED 30 November 2023

CITATION
Southey MWY and Brunavs M (2023), Introduction to small molecule drug discovery and preclinical development. *Front. Drug Discov.* 3:1314077. doi: 10.3389/fddsv.2023.1314077

Introduction to small molecule drug discovery and preclinical development

Michelle W. Y. Southey* and Michael Brunavs

Molecular Architects, Drug Discovery Chemistry, Evotec (UK) Ltd., Abingdon, United Kingdom

Over 90% of marketed drugs are small molecules, low molecular weight organic compounds that have been discovered, designed, and developed to prompt a specific biological process in the body. Examples include antibiotics (penicillin), analgesics (paracetamol) and synthetic hormones (corticosteroids). On average, it takes 10–15 years to develop a new medicine from initial discovery through to regulatory approval and the total cost is often in the billions. For every drug that makes it to the market, there are many more that do not, and it is the outlay associated with abortive efforts that accounts for most of this expense. The discovery of new drugs remains a significant challenge, involving teams of researchers from chemistry, biology, drug development, computer science and informatics. In this article we will discuss the key concepts and issues encountered in small molecule preclinical drug discovery and introduce some of the emerging technologies being developed to overcome current obstacles.

KEYWORDS

drug discovery, small molecules, compound screening, artificial intelligence, multimodalities

1 Introduction

Small molecule drugs are synthetic medicinal chemicals designed to mimic, enhance, or diminish the behaviour of natural substances or products within the body. They have relatively simple structures, customizable to meet specific therapeutic goals. They are generally stable and rarely need specialized storage conditions. Their behaviour in the body, or *in vivo*, is usually predictable, leading to straightforward, often oral, dosing protocols that patients find easy to manage. They can treat a wide variety of diseases because they can move through the body easily, transferring from the gut *via* the blood stream to the site of action, permeating through cell membranes to reach intracellular targets. They can be administered as pills, inhalers, suppositories or injectables, making them very flexible.

The chemical structure of small molecules can be designed to interact selectively with specific biological targets. By altering the atomic composition of small molecules, their overall properties can be fine-tuned to a particular purpose, eliciting only the desired response. The flexibility afforded by being able to explore all "chemical space" in this way, offers small molecule approaches a marked advantage over other modalities. The process of inventing a small molecule drug and ensuring that it performs precisely as it should, minimizing unwanted side effects, involves meticulous design and synthetic mastery from researchers, often over several years.

Compared to therapeutic proteins, or biologics, they are also easier to develop (Makurvet, 2021). Once optimized, small molecule drugs can be manufactured very

reproducibly, an advantage for researchers seeking a return on their investment. Once patent life expires, non-branded generic forms of the medicine will increase availability to patients.

In this mini-review we will describe the key concepts and considerations involved in the discovery of small molecule drugs, covering traditional approaches, and discuss how recent advances such as the rise of artificial intelligence and innovative new modalities have reinvigorated the field.

1.1 How do small molecules interact with biological targets?

Proteins are the most common therapeutic targets; they are large complex molecules that play important roles in the body. Proteins are comprised of small building blocks known as amino acids. The sequence in which these amino acids are arranged determines the precise shape and function of the protein.

There are many ways in which small molecules work to elicit a therapeutic response in the human body (Silverman, 1992; Patrick, 2001; Young, 2009). Three of the most common are listed below:

1. Enzyme inhibitors—enzymes are proteins that catalyze biochemical reactions. By blocking the activity of these proteins, small molecules can interfere with disease processes to provide therapeutic benefits. Statins are a class of enzyme inhibitor drugs; they work by inhibiting the activity of an enzyme involved in the production of cholesterol in the liver. By reducing overall cholesterol levels in the body, they reduce the risk of heart disease and stroke.
2. Receptor agonist/antagonists—small molecules that can interact with proteins that exist on the surface of cells, usually in one of two ways: *agonists* which activate the receptor, mimicking the natural signaling molecule or *antagonists* which block the receptor, inhibiting the binding of the natural signaling molecule and reducing activation. Albuterol is a receptor agonist prescribed for the treatment of asthma which activates the receptor responsible for opening the airways in the lung.
3. Ion channel modulators—ion channels are proteins embedded in cell membranes which are responsible for regulating the flow of ions into and out of those cells. They play a key role in a wide variety of physiological processes including regulation of heartbeat and neurotransmission. Small molecule drugs can modulate the opening and closing of these channels to treat diseases such as epilepsy.

Many of the mechanisms of action described above involve a well-defined region on the protein into which a small molecule can fit and bind. These regions are known as *active sites*, and the geometric arrangement of their amino acids is such that they only have affinity for a few naturally occurring molecules, or *substrates*, within the body. This mechanism is often referred to as the "lock and key" theory. By understanding the requirements of the lock, researchers can create the best small molecule "key" to fit it and thereby generate the desired response. The stronger and the more specific the compound interaction with the amino acids of the targeted active site is, the less likely that compound will be to bind to different proteins. In turn, the fewer side effects it will have. Thus, the design of small molecule drugs is highly specialized.

As a result of better disease understanding and the development of innovative technologies, more diverse approaches for disease modulation by small molecules have evolved that exploit different mechanisms of action. Examples include the modulation of protein-protein interactions (PPI) (Wells and McClendon, 2007) (Trisciuzzi, et al., 2023), bifunctional protein degraders (Sun, et al., 2019) and stabilizers (Dong, et al., 2021; Henning, et al., 2022; Mullard, 2023). Some of these other modalities are explored later in the article. Table 1 shows some of the small molecule drugs that have reached the clinic over the last century, highlighting the evolution of their structural complexity as well as their mechanisms of action.

2 The preclinical drug discovery process

Paths towards the identification of a preclinical drug candidate that successfully reaches the market are complicated, and depend upon a variety of factors including the complexity of the disease and the rigorous validation and testing required to meet regulatory approval requirements. The general process is outlined in Figure 1. In the following sections we will discuss each stage of the process in more detail.

2.1 Target discovery

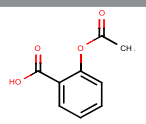
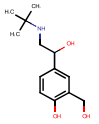
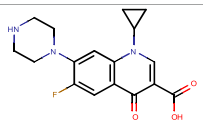
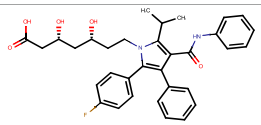
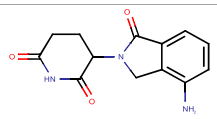
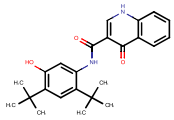
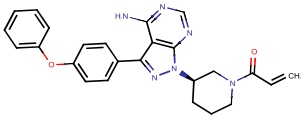
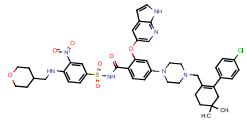
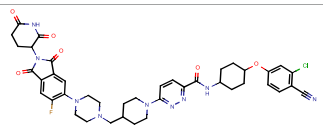
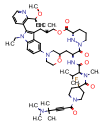
Before drug discovery programs are prosecuted, the chosen biological target must be validated as relevant. This can be a time-consuming process as researchers try to demonstrate the role of the target in a particular biological pathway, process, or disease of interest (Schenone, et al., 2013). Upon validation, the assumption is made that modulation of that target will elicit the desired effect. Proteins remain the most represented class of therapeutic targets (Santos, et al., 2017), but other types of biological molecules can also be targeted by drugs, such as nucleic acids (DNA and RNA) (Kulkarni, et al., 2021). Therapeutic targets are chosen based on the disease being treated and the potential to interfere with the mechanisms of disease progression. For example, many anti-cancer drugs target proteins responsible for abnormal cell growth and division whilst for Alzheimer's disease, a common target is amyloid beta, a protein that forms plaques (Ramanan and Day, 2023). Drugs can be developed to prevent plaque formation or degrade those that have already formed.

2.2 Screening and lead identification

Once the research team understands the physiological role played by the target, they can assess how its modulation will affect the disease state and begin their search for a chemical agent to achieve the desired outcome.

If researchers know little or nothing about a target at the outset of their work (maybe the utility of the target has only just been discovered, and the team is seeking to be *first-in-class*) then the simplest approach is a *random high throughput screen (HTS)*. In this

TABLE 1 Examples of small molecule drugs that have reached the clinic and the mechanisms of action associated with them.

Name	Molecular structure	Smiles	Mechanism of action	Clinical stage	Indication
Aspirin		CC(=O)Oc1ccccc1C(O)=O	Enzyme Inhibitor	Approved 1915	Pain
Salbutamol (Albuterol)		CC(C)(C)NCC(O)C1=CC(CO)=C(O)C=C1 |c:15,t:8,12|	Receptor Agonist	Approved 1974	Asthma
Ciprofloxacin		OC(=O)C1=CN(C2CC2)c2cc(N3CCNCC3)c(F)cc2C1=O |t:3|	Enzyme Inhibitor	Approved 1987	Antibiotic
Atorvastatin		CC(C)c1c(C(=O)Nc2ccccc2)c(c(-c2ccc(F)cc2)n1CC[C@@H](O)C[C@@H](O)CC(O)=O)-c1ccccc1	Enzyme Inhibitor	Approved 1996	High Cholesterol
Lenalidomide		Nc1cccc2C(=O)N(Cc12)C1CCC(=O)NC1=O	Degrader	Approved 2005	Multiple Myeloma
Ivacaftor		CC(C)(C)c1cc(c(NC(=O)C2=CNc3ccccc3C2=O)cc1O)C(C)(C)C |t:11|	Ion Channel Potentiator	Approved 2012	Cystic Fibrosis
Ibrutinib		NC1=NC=NC2=C1C(=NN2[C@@H]1CCCN(C1)C(=O)C=C)C1=CC=C(OC2=CC=CC=C2)C=C1 |c:3,5,8,30,32,35,t:1,23,25,28|	Covalent	Approved 2013	Mantle Cell Lymphoma
Venetoclax		CC1(C)CCC(CN2CCN(CC2)c2ccc(C(=O)NS(=O)(=O)c3ccc(NCC4CCOCC4)c(c3)[N+]([O-])=O)c(Oc3cnc4[nH]ccc4c3)c2)=C(C1)c1ccc(Cl)cc1 |c:57|	PPI modulator	Approved 2016	Chronic Myeloid Leukaemia
Bavdegalutamide (ARV-110)		FC1=C(C=C2C(=O)N(C3CCC(=O)NC3=O)C(=O)C2=C1)N1CCN(CC2CCN(CC2)C2=NN=C(C=C2)C(=O)NC2CCC(CC2)OC2=CC(Cl)=C(C=C2)C#N)CC1 |c:20,37,39,56,58,t:1,3,35,53|	PROTAC	Phase II (NCT05177042)	Prostate Cancer
RM-6291		CCN1C2=C3C=C(C=C2)N2CCO[C@H](C2)C[C@H](NC(=O)[C@H](C(C)C)N(C)C(=O)C2(F)CCN(CC2)C(=O)C#CC(C)(C)N(C)C)C(=O)N2CCC[C@H](N2)C(=O)OCC(C)(C)CC3=C1C1=C(N=CC=C1)[C@H](C)OC |c:5,7,66,71,73,t:3,69|	Molecular Glue	Phase I (NCT05462717)	Non-Small Cell Lung Cancer

approach a large library of structurally diverse compounds (usually over 100 K) will be tested against the target, in the hope that some will prove active. These can be compounds that the researchers have purchased or made before, or purified natural products. Hit rates are typically low (~1%, or sometimes lower!) but the goal is to find something, and build the project up from the actives, or *seeds*, identified.

Medicinal chemists continually wrestle to ensure that the compound collections they use for screening are fit for purpose.

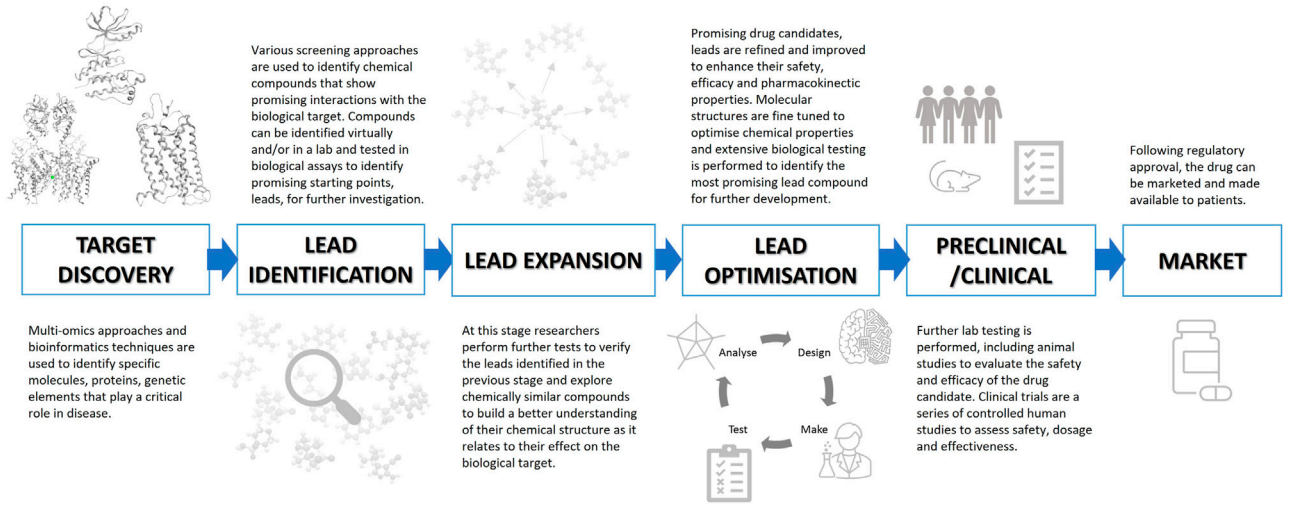


FIGURE 1
An overview of the drug discovery process from target discovery to market approval.

It is estimated that the total number of compounds that make up oral druglike chemical space is 10^{60} (Reymond, et al. 2010). Even a screening set containing a million compounds—which is rare—barely scratches the surface of available druglike space. In cases where more is known about a target (perhaps the binding site of the agents has already been ascertained), the screening set can be more focused. To create that focus, *virtual screening* can be undertaken. Docking experiments, in which compounds are fitted into the known binding site using a software package, can be carried out ahead of the physical testing. In this way researchers can eliminate compounds that have no hope of binding and promote those that have a better chance. This process is known as *structure-based virtual screening* because it is enabled by knowledge of the protein structure. This type of screening is designed to evaluate specific compound binding hypotheses, unlike the random HTS experiments described earlier.

Where the binding site is not known, it may still be possible to create a focused screening set if agents active against the target have been identified in previous efforts. In what is known as a *ligand-based virtual screen*, compounds with similar structures to the known actives are selected, in the hope that these will provide a *best-in-class* solution. These might contain motifs structurally similar to those in the comparator actives, or different while still bearing similar protein interaction properties. These *scaffold hopping* searches can be conducted in two- or three-dimensions, and success often opens new intellectual property space.

The concept of the *fast follower* takes the focus concept even further. Here a synthetic effort will start from a validated active discovered during a previous effort, to address residual target selectivity, drug stability or toxicological concerns as rapidly as possible. Often, this can result in the discovery of one or more new chemical series.

Having identified a suitable set of compounds for test, whether that be a random HTS with many compounds, or a smaller focused set informed by prior knowledge, the search is then performed. Lots of preparation goes into this: an assay must be created, expressing the target protein in relevant cell systems, validated using appropriate control compounds, and the correct concentrations at which to test compounds determined through pilot work.

2.3 Lead expansion

Hopefully, the researchers will get some hits. These may all come from the same structural class of compounds, or *chemotype*, or from several different ones. Either way, attention will now focus strongly on the active chemotypes, and the elucidation of the *structure-activity relationship* (SAR).

In a manner akin to the ligand-based virtual screening described above, researchers will seek to augment their knowledge around hit chemotypes by searching compound databases for additional, structurally related test materials. Such databases could be corporate, in-house compound collections, or the catalogues of commercial suppliers. If key compounds prove unavailable, the team will synthesize them.

Armed with the additional information this augmentation provides, researchers will first verify that there really is an SAR. A group of structurally related compounds which are all active is a potential red flag: genuinely bioactive compounds interact with the target protein and cases should exist where those interactions are suboptimal, and activity is reduced. If a chemotype is active across all its members (a "flat" SAR) this may indicate false positivity, arising from an artefactual phenomenon in the assay and not the desired interaction. Such chemotypes would drop out of contention.

Having removed these duds, the surviving chemotypes are compared to see which ones are most likely to generate a marketable clinical candidate down the line. Several factors inform the decision. For instance, can new members of the chemotype be synthesized quickly so that its optimization will be cost effective? Are the actives already known in the literature, and/or the intellectual property of competitors? Can the probable side effects

of the compounds or their stability *in vivo* be anticipated, through modelling and testing, and plans created to discharge such risks?

Researchers will execute a *design-make-test-analyse (DMTA)* cycle to explore a chemotype through synthesis and assay. Computational scientists help to design the compounds best able to complete the study in the fewest number of cycles, deploying models to assess risks and predict activity, refining their forecasts and guiding the research ever more accurately over time as further data are acquired. In this way, the team builds a data package for each chemotype, allowing them to make an informed choice between them and create an issue-focused plan as they move forward into the optimization phase.

2.4 Lead optimization to preclinical development

At this stage of the process the objective is to gain as much knowledge as possible about the lead compounds' efficacy and safety before they are tested in humans during clinical trials. This usually involves a multitude of laboratory tests, or *assays*, to assess both the drug's action on the body and the body's action on the drug.

A successful drug needs to be able to reach its target and exert a medicinal effect for the required length of time. Depending on the indication, this timeframe can vary markedly. A sleeping pill must reach its site of action quickly and be eliminated from the body by morning, whereas a drug designed to alleviate more chronic indications, such as cancer or dementia, will ideally last much longer. While target interaction, or *engagement*, is essential for a small molecule to have the desired pharmacological effect, it is equally vital that drugs be able to reach the location of the target efficiently and in sufficient concentration to effect that target engagement. It must also do so as selectively as possible to minimize potential toxicity and undesired side effects.

The engagement of a compound with the target of interest is investigated using structural biology techniques, through the computational creation and exploration of 3D models of the target. These structural models are central to the further design and optimization of compounds during the lead optimization phase, as researchers endeavour to improve the "fit" of the compound to the active site.

DMPK (drug metabolism and pharmacokinetics) studies allow researchers to understand how their compounds are absorbed into, distributed around, and eliminated from the body whilst pharmacodynamics (PD) studies the interactions of the drug with the body and tells the team about the potency and effectiveness of the compound. These assessments drive decisions during the discovery process, combining lab-based experimentation with the use of computational modelling and machine learning methods to make early predictions of various DMPK outcomes (Obrezanova, 2023). They help to determine the optimal dosage, side effect profile and toxicity risks. Toxicity can arise directly, through the action of the compound itself (insufficient selectivity for the required target) or indirectly, by interfering with the action of other drugs. Such drug-drug interactions (DDIs) are often problematic: if one drug inhibits critical enzymes then another might endure in the body to a much greater extent than is safe. This is a major concern in geriatric care for instance, where frail patients often take many medicines at once.

Thus, there are a vast number of consider researchers must think through at this stage to ensure their agents are safe and effective, and the DMTA cycle continues until this is achieved. Researchers will make modifications to the structures of their leading compounds to arrive at a better balance of DMPK properties. They will seek to bring about the desired clinical outcome from the smallest possible dose—another good strategy for the elimination of side effects. At the same time, they will seek to ensure that any structural changes they make will neither diminish potency against their target nor reduce selectivity. Drug discovery researchers often find themselves optimizing multiple, often competing parameters to find the "sweet spot" that will deliver the optimal properties for a given indication.

Prior to entering clinical trials, compounds are assessed both *in vitro* and *in vivo*. The latter involves the use of animal models, usually but not exclusively rodents, to represent the systems and functions of the human body. Such studies reveal important pharmacological and toxicological information, and until such time as computer modelling becomes so accurate that we can turn confidently to human volunteers when testing new drugs *in vivo* for the first time, their use will always have a place. At the same time, ethical concerns clearly have a huge role in driving the future of drug discovery, and pharmaceutical companies continue to drive animal testing down to the bare minimum needed, cognizant of the fact that *in vivo* responses in such testing do not always translate effectively to humans. There is continued heavy investment in more and better *in vitro* and *ex vivo* testing, and computer modelling, to meet this important challenge (Powell, 2018).

Many times, despite months of costly effort, it proves impossible to design a drug which is both safe and efficacious (DiMasi et al., 2003). At other times though, success is achieved. A drug candidate is developed with the activity, safety and DMPK profile needed to combat the targeted disease with a dose regimen that best suits the lifestyle of the patient. These compounds move forward to clinical trials.

3 Emerging technologies in small molecule drug discovery

Despite the large amount of money invested in drug discovery, there are still only around 500 treatments but over 7,000 human diseases (Austin, 2021). Drug discovery is expensive and time-consuming, with high rates of late-stage attrition due to lack of efficacy or compound related safety issues. The high failure rate underlines the complexity of drug discovery; learning from past mistakes and exploring new technologies is crucial if the industry is to improve its success rate. The implementation of innovative methodologies such as artificial intelligence (AI) and the development of new modalities to target "undruggable" targets are two ways the industry is evolving to improve the chances of producing new therapies and better patient outcomes.

3.1 Artificial intelligence (AI)

AI techniques offer the potential to improve the speed, efficiency and therefore cost of the drug discovery process on the basis that computers are more efficient at analyzing and processing large

amounts of data to help researchers make less biased, more informed decisions (Brown, et al., 2020). A current example of AI application is in compound screening (Sadybekov and Katritch, 2023). Traditional high throughput screening is limited to hundreds of thousands of compounds at best, but AI can be used to screen millions of compounds virtually, solving the chemical space conundrum in a much more cost-effective way to identify potential novel drug candidates. Some commercial vendors have enormous compound libraries of synthetically tractable compounds which can be made to order if a virtual screen suggests they will be active. The ability of AI to explore broader chemical space improves the chances of success in finding potential leads and is especially pertinent given the rise of clinically approved compounds that reside outside traditional "oral druglike chemical space" (Doak, 2014). In other areas, advanced machine learning models are trained utilizing neural networks, imitating the biological network of the human brain, and used to predict the efficacy of compounds and safety endpoints (Cavasatto and Scardino, 2022). In turn, better informed decisions are made at an earlier stage in the process, mitigating the risk of more costly late-stage failure (Vamathevan, et al., 2019). Natural language processing (Gruetzemacher, 2022) can mine and process millions of scientific research articles to reveal insights into disease mechanisms and biology. AI tools now commonly used in drug discovery can design molecules from first principles (Vanhaelen et al., 2020), repurpose (Prasad and Kumar 2021; Roessler, et al., 2021) known drugs for the treatment of other diseases and even help plan synthetic routes (Thakkar, et al., 2021) based on prior knowledge.

This all highlights one of the key challenges with AI in drug discovery (Bender and Cortés-Ciriano, 2021). To be well enough informed to make accurate decisions, AI tools require large volumes of data from which they are "trained". In the absence of sufficient quality and volume of data, these tools can be unreliable and inaccurate.

One of the most cited examples of how transformative AI can be in the field of life sciences is the release of DeepMind's AlphaFold, which uses deep learning technology to predict three-dimensional protein structures from amino acid sequences (Jumper, et al., 2021). The neural network system used was trained on a vast dataset of protein structures and sequences, learning the intricate relationship between amino acid sequences and their preferred spatial arrangements with high accuracy. Whilst this has revolutionized the field of protein structure prediction, its full impact on the field of drug discovery is only just starting to be realized (Arnold, 2023).

AI holds much promise and has achieved some notable successes so far but human researchers are clearly far from being replaced; their experience is key to the validation of AI tools and the interpretation of their predictions (Jiménez-Luna et al., 2020). The synergistic integration of AI technologies, traditional screening approaches and human wisdom remains essential for modern drug discovery programs to successfully deliver new treatments (Griffen et al., 2020).

3.2 New modes of action for small molecules

The occupancy-based mode of action for many small molecule drugs has historically precluded a significant portion of the genome. For example, non-enzymatic and/or intrinsically disordered proteins do not possess well-defined active sites to which small molecules can bind. Such targets have been considered "undruggable", but innovative strategies to chemically modulate these proteins are emerging and have attracted huge interest in drug research and development, offering the potential to treat many more diseases and address at least some of the gaps in unmet patient need (Blanco and Gardinier, 2020).

For instance, certain amino acids on the surface of a target protein, even in otherwise featureless active sites, are capable of bonding chemically to covalent drugs. Since such compounds directly attach to target proteins, rather than interact transiently, they offer benefits such as increased efficacy and specificity as well as extended duration of action. An example of a marketed covalent drug is ibrutinib, approved by the FDA in 2013 for the treatment of various blood cancers, which works by blocking the activity of a protein involved in the growth and survival of cancer cells (Shaywitz, 2013).

Another new modality for small molecules is their use in the disruption/modulation of the interactions between two or more target proteins. Biological processes are frequently regulated through such protein-protein interactions (PPIs), so directly targeting them can bring about therapeutic benefits. Venetoclax is an approved PPI drug prescribed for the treatment of acute myeloid leukemia (AML) and chronic lymphocytic leukemia (CLL). Its mode of action involves disrupting the interaction of two proteins that would otherwise work together to promote cancer cell survival (Roberts and Huang, 2017).

Protein degrader therapies take advantage of the body's natural processes to schedule disease-causing proteins for degradation. The degrader recruits undesirable proteins to the cellular machinery responsible for protein breakdown, leading to their elimination. Lenalidomide (Armoiry et al., 2008) is a degrader that promotes the removal of a protein promoting cancer cell growth, FDA-approved for the treatment of multiple myeloma in 2005. This approach has been developed further in recent years, giving rise to a new class of drugs called PROTACs (proteolysis-targeting chimeras). Bavdegalutamide is an example of a PROTAC drug currently in Ph II trials for the treatment of prostate cancer (Gao, et al., 2022).

4 Conclusion

The field of small molecule drug discovery continues to evolve and remains a highly promising path in the pursuit of novel therapeutics (Härter, et al., 2022; Howes, 2023). Through meticulous design and rigorous testing in both preclinical and clinical settings, researchers in both the pharmaceutical industry and academia are advancing our understanding of disease biology. Compounds entering the clinic in recent years have successfully challenged previous dogma regarding the molecular properties required for an oral drug and the methodology by which disease progression can be arrested. Ongoing investment and advances in biology, computational technologies and innovative synthetic chemistry are providing researchers with increasingly efficient and precise tools for small molecule discovery and design. The synergistic partnership between scientific expertise and technological progress still holds great promise for the discovery

and development of small molecule medications and treatments in the future.

Author contributions

MS: Writing–original draft, Writing–review and editing. MB: Writing–original draft, Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

Authors MS and MB were employed by the company Evotec (UK) Ltd.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adams, D., Gonzalez-Duarte, A., O'Riordan, D., Yang, C.-C., Ueda, M., Kristen, V., et al. (2018). Patisiran, an RNAi therapeutic, for hereditary transthyretin amyloidosis. *N. Engl. J. Med.* 379, 11–21. doi:10.1056/NEJMoa1716153

Armoiry, X., Aulagner, G., and Facon, T. (2008). Lenalidomide in the treatment of multiple myeloma:a review. *J. Clin. Pharm. Ther.* 33, 219–226. doi:10.1111/j.1365-2710.2008.00920.x

Arnold, C. (2023). AlphaFold touted as next big things for drug discovery - but is it? *Nature* 622, 15–17. doi:10.1038/d41586-023-02984-w

Austin, C. P. (2021). *National Center for advancing translational sciences.* 28 january. Available at: https://ncats.nih.gov/director/january-2021 (Accessed October 30, 2023).

Bender, A., and Cortés-Ciriano, I. (2021). Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: ways to make an impact, and why we are not there yet. *Drug Discov. Today* 26, 511–524. doi:10.1016/j.drudis.2020.12.009

Blanco, M.-J., and Gardinier, K. M. (2020). New chemical modalities and strategic thinking in early drug discovery. *ACS Med. Chem. Lett.* 228–231. doi:10.1021/acsmedchemlett.9b00582

Brown, N., Ertl, P., Lewis, R., Luksch, T., Reker, D., and Schneider, N. (2020). Artificial intelligence in chemistry and drug design. *J. Comput. Aided Mol. Des.* 34, 709–715. doi:10.1007/s10822-020-00317-x

Cavasatto, C. N., and Scardino, V. (2022). Machine learning toxicity prediction: latest advances by toxicity end point. *ACS Omega* 7, 47536–47546. doi:10.1021/acsomega.2c05693

DiMasi, J. A., Hansen, R. W., and Grabowski, H. G. (2003). The price of innovation: new estimates of drug development costs. *J. Health Econ.* 22, 151–185. doi:10.1016/S0167-6296(02)00126-1

Doak, B. C., Over, B., Giordanetto, F., and Kihlberg, J. (2014). Oral druggable space beyond the rule of 5: insights from drugs and clinical candidates. *Chem. Biol.* 21, 1115–1142. doi:10.1016/j.chembiol.2014.08.013

Dong, G., Ding, Yu, He, S., and Sheng, C. (2021). Molecular glues for targeted protein degradation: from serendipity to rational discovery. *J. Med. Chem.* 64, 10606–10620. doi:10.1021/acs.jmedchem.1c00895

Gao, X., Howard, A., Vuky, J., Dreicer, R., Oliver Sartor, A., Sternberg, C. N., et al. (2022). Phase 1/2 study of ARV-110, an androgen receptor (AR) PROTAC degrader, in metastatic castration-resistant prostate cancer (mCRPC). *J. Clin. Oncol.* 17–17. doi:10.1200/JCO.2020.38.15_suppl.3500

Griffen, E. J., Dossetter, A. G., and Leach, A. G. (2020). Chemists: AI is here; unite to get the benefits. *J. Med. Chem.* 63, 8695–8704. doi:10.1021/acs.jmedchem.0c00163

Gruetzemacher, R. (2022). The power of natural language processing. Available at: https://hbr.org/2022/04/the-power-of-natural-language-processing.

Härter, M., Beck, H., Haß, B., Schmeck, C., and Baerfacker, L. (2022). Small molecules and their impact in drug discovery: a perspective on the occasion of the 125th anniversary of the Bayer Chemical Research Laboratory. *Drug Discov. Today* 27, 1560–1574. doi:10.1016/j.drudis.2022.02.015

Henning, N. J., Boike, L., Spradlin, J. N., Ward, C. C., Liu, G., Zhang, E., et al. (2022). Deubiquitinase-targeting chimeras for targeted protein stabilization. *Nat. Chem. Biol.* 18, 412–421. doi:10.1038/s41589-022-00971-2

Howes, L. (2023). Chemical and engineering news. Available at: https://cen.acs.org/pharmaceuticals/drug-discovery/small-molecule-drug-discovery-having/101/i36 (Accessed November 3, 2023).

Jiménez-Luna, J., Grisoni, F., and Schneider, G. (2020). Drug discovery with explainable artificial intelligence. *Nat. Mach. Intell.* 2, 573–584. doi:10.1038/s42256-020-00236-4

Jumper, J., Evans, R., Alexander, P., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2

Kulkarni, J. A., Witzigmann, D., Thomson, S. B., Chen, S., Leavitt, B. R., Cullis, R., et al. (2021). The current landscape of nucleic acid therapeutics. *Nat. Nanaotechnology* 16, 630–643. doi:10.1038/s41565-021-00898-0

Makurvet, F. D. (2021). Medicine in drug discovery *1-8*. Available at: https://pharmanewsintel.com/news/key-differences-in-small-molecule-biologics-drug-development (Accessed October 27, 2023).Biologics vs. small molecules: drug costs and patient access

Mullard, A. (2023). Glue-based KRAS inhibitors make their debut cancer trial mark. *Nat. Rev. Drug Discov.* doi:10.1038/d41573-023-00169-8

Obrezanova, O. (2023). Artificial intelligence for compound pharmacokinetics prediction. *Curr. Opin. Struct. Biol.* 79, 102546–102547. doi:10.1016/j.sbi.2023.102546

Patrick, G. (2001). *Instant notes medicinal chemistry.* London and New York: Taylor & Francis.

Powell, K. (2018). Replacing the replacements: animal model alternatives. *Science* 12 (October), 243–245. doi:10.1126/science.362.6411.246-b

Prasad, K., and Kumar, V. (2021). Artificial intelligence-driven drug repurposing and structural biology for. *Curr. Res. Pharmacol. Drug Discov.*, 2590–2571. doi:10.1016/j.crphar.2021.100042

Ramanan, V. K., and Day, G. S. (2023). Anti-amyloid therapies for Alzheimer disease: finally, good news for patients. *Mol. Neurodegener.* 18, 42–43. doi:10.1186/s13024-023-00637-0

Reymond, J.-L., Van Deursen, R., Blum, L. C., and Ruddigkeit, L. (2010). Chemical space as a source for new drugs. *Med. Chem. Comm.* 30-38. doi:10.1002/wcms.1104

Roberts, A. W., and Huang, D. C. S. (2017). Targeting BCL2 with BH3 mimetics: basic science and clinical application of venetoclax in chronic lymphocytic leukemia and related B cell malignancies. *Clin. Pharmacol. Ther.* 89–98. doi:10.1002/cpt.553

Roessler, H. I., Knoers, N. V. A. M., M van Haelst, M., and Gijs van Haaften, (2021). Drug repurposing for rare diseases. *Trends Pharmacol. Sci.* 42, 255–267. doi:10.1016/j.tips.2021.01.003

Sadybekov, A. V., and Katritch, V. (2023). Computational approaches streamlining drug discovery. *Nature* 616, 673–685. doi:10.1038/s41586-023-05905-z

Santos, R., Ursu, O., Gaulton, A., Bento, A. P., Donadi, S., Bologa, C. G., et al. (2017). A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* 19–34. doi:10.1038/nrd.2016.230

Schenone, M., Dančík, V., Wagner, K., and Clemons, P. A. (2013). Target Identification and mechanism of action in chemical biology and drug discovery. *Nat. Chem. Biol.* 9, 232–240. doi:10.1038/nchembio.1199

Shaywitz, D. (2013). *Forbes.* 5 april. Available at: https://www.forbes.com/sites/davidshaywitz/2013/04/05/the-wild-story-behind-a-promising-experimental-cancer-drug/ (Accessed November 3, 2023).

Silverman, R. B. (1992). *The organic chemistry of drug design and drug action*. California: Academic Press.

Sun, X., Gao, H., Yang, Y., He, M., Yue, Yu, Song, Y., et al. (2019). *PROTACs: great opportunities for academia and industry."* Signal Transduction and Targeted Therapy.

Thakkar, A., Chadimova, V., Engkvist, O., Reymond, J.-L., and Reymond, J. L. (2021). Retrosynthetic accessibility score (RAscore) – rapid machine learned synthesizability classification from AI driven retrosynthetic planning. *Chem. Sci.* 12, 3339–3349. doi:10.1039/d0sc05401a

Vamathevan, J., Clark, D., Paul, C., Ian Dunahm, H., Ferran, E., Lee, G., et al. (2019). Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* 463–477. doi:10.1038/s41573-019-0024-5

Vanhaelen, Q., Lin, Y.-C., and Zhavoronkov, A. (2020). The advent of generative chemistry. *ACS Med. Chem. Lett.* 11, 1496–1505. doi:10.1021/acsmedchemlett.0c00088

Wells, J. A., and McClendon, C. L. (2007). Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* 450, 1001–1009. doi:10.1038/nature06526

Young, D. C. (2009). *Computational drug design*. New Jersey: John Wiley & Sons, Inc.

frontiers | Frontiers in Drug Discovery

TYPE Mini Review
PUBLISHED 25 January 2024
DOI 10.3389/fddsv.2024.1342866

OPEN ACCESS

EDITED BY
Bruno Villoutreix,
Institut National de la Santé et de la Recherche Médicale (INSERM), France

REVIEWED BY
Antoine Gedeon,
Sorbonne Université, France

*CORRESPONDENCE
Stéphane Giraud,
stephane.giraud@lyon.unicancer.fr

RECEIVED 22 November 2023
ACCEPTED 05 January 2024
PUBLISHED 25 January 2024

CITATION
Giraud S (2024), Identification of first active compounds in drug discovery. how to proceed? *Front. Drug Discov.* 4:1342866. doi: 10.3389/fddsv.2024.1342866

Identification of first active compounds in drug discovery. how to proceed?

Stéphane Giraud*

Center for Drug Discovery and Development (C3D Platform), Centre de Recherche en Cancérologie de Lyon INSERM U1052-CNRS UMR5286, Université de Lyon, Université Claude Bernard Lyon1, Centre Léon Bérard, France. Synergie Lyon Cancer Foundation, Lyon, France

In the quest for the discovery of new therapies, the identification of the initial active molecules is a major challenge. Although significant progress in chemistry and biology has been made in recent years, the process remains difficult. In this mini-review, we will explain the major approaches and experimental methods that can be used to identify these molecules. Two main approaches are described, target-based and phenotypic-based and a focus is made on some high throughput technologies and biophysical methods.

KEYWORDS

drug discovery, small organic compounds, screening methods, HTS, biophysics

1 Introduction

Despite scientific and methodological advances made over the last 20 years, identifying and developing new therapies remain long and costly processes (Wouters et al., 2020). Indeed, the marketing and distribution to the general public of small molecules, antibodies or therapeutic proteins take several years; for instance, a small chemical molecule is commercially available after 12–15 years. As the duration of certain developmental phases, such as clinical trials, can hardly be reduced, a great effort has been made to spend less time on the first steps of the process (pre-discovery, drug discovery and preclinical). New biophysical, biochemical, biological and *in silico* technologies have emerged to accelerate the discovery stage.

In this review we will focus on small organic compounds; vaccines, cellular therapy, therapeutic antibodies and other biologics will not be addressed. We will particularly discuss the main methods for identifying the first bioactive small molecules, also known as "hits". The definition of a hit can vary across the scientific community, but in this article a hit will be considered as a molecule whose activity is confirmed in one or several primary biological and/or biophysical assays. These hits will then be optimized through an iterative cycle involving biology, biophysics, chemistry and AI-based methods (Vemula et al., 2023) to obtain a new drug, displaying a high efficacy and a low or even no toxicity.

The identification of new small organic molecules-based therapies requires a set of molecules to be tested and a robust validated assay. These molecules can be obtained in different ways, which will not be detailed herein, however here are some commons sources:

(1) Natural products: nature has always been a source of valuable bioactive molecules. Natural products and their derived molecules have been used since ancient times to treat diseases. These molecules are found in plants, microbes, aquatic organisms, animals, fungi and insects. Many drugs, such as antibiotics and anticancer agents, were originally derived from natural sources. (Newman and Cragg, 2020; Naeem et al., 2022).

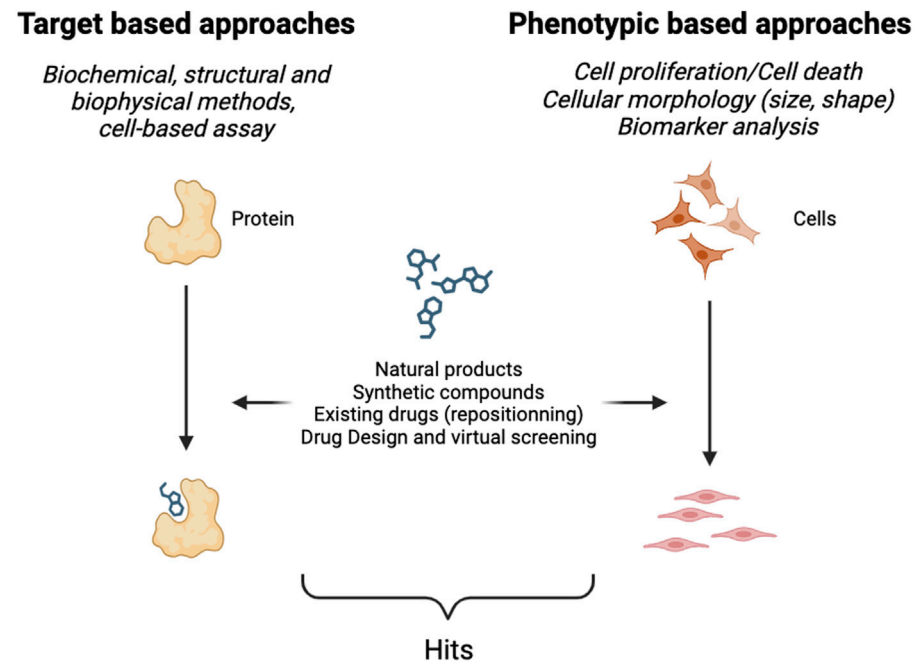


FIGURE 1
Schematic representation of hit identification.

(2) Synthetic compounds: pharmaceutical companies and research institutions often maintain libraries of synthetic chemical compounds. Combinatorial and parallel chemistry have been used to generate thousands of molecules by systematically varying chemical structures. Today more rational approaches are used to design and synthesized specific molecules intended to inhibit particular targets like kinases, ion channels, GPCR or biological mechanisms like protein-protein interactions or DNA methylation. Artificial intelligence and machine learning algorithms are also used to identify potential drug candidates by analyzing huge datasets and predicting the biological activity of molecules. (Yu and MacKerell, 2017).

(3) Repurposing of existing drugs: sometimes existing drugs that were developed for one indication can be repurposed for treating different diseases. This approach emerged in the early 1990s and has been proven to be a viable alternative to the identification of new drugs. (Gns et al., 2019).

(4) Drug design: this rational methodology consists in designing potential active compounds, i.e., compounds that bind to a particular target, based on structural data of the target or based on data of the ligand. Many computational techniques have recently emerged that help researchers identify innovative compounds. (Hoffer et al., 2018; Singh et al., 2020).

2 How to identify new bioactive molecules?

To identify new hits, a screening strategy (or method) must be adopted. A set of specific assays must be carried out to identify and optimize potential drugs that then become drug candidates for clinical trials. Two major kinds of approaches exist (Figure 1), namely, those that require the identification of a target and validation of the relationship between that target and a particular disease–called target-based approaches–and those that work in a target agnostic fashion known as phenotypic approaches. The latter consist in observing the effect(s) of a new potential therapy at the level of cells or whole organisms. Phenotypic approaches require an experimental model as close as possible of the pathology and symptoms observed in human.

2.1 Target-based screening

Target-based screening relies primarily on the identification of a disease-relevant target; typically, for example, proteins and nucleic acids. This type of screening can be performed *in vitro* using biochemical and biophysical methods, or *in cellulo* using cellular models to assess the activity of the compound towards the target in a cellular context. The assays developed to perform the screening are designed either to measure the interaction between a potential drug and the chosen target, or the ability of a drug to modulate a cellular function through its interaction with the target. The aim of these methods is to develop an assay that produces a detectable signal in order to visualize, primarily through the emission of light (in the visible or fluorescence spectra), the activity of a given compound towards the target. The development of an assay is not trivial and can take time as it must be sensitive, reliable and reproducible enough to provide comparative results when thousands of compounds are screened.

2.1.1 High Throughput Screening (HTS)

High Throughput Screening consists in the screening of large libraries of compounds (from thousands to sometime millions) in

order to identify hits (Blay et al., 2020). This approach is based on the automation of biological and biophysical assays that can be miniaturized and must be a highly sensitive method to identify active compounds. At its inception, HTS screening campaigns were carried out using 96-well plates, now screens are conducted in 384-well or 1,536-well plates of the same dimension as their predecessor. This means that the number of reactions that can be performed in parallel has significantly increased over the last 2 decades, and that the time needed to screen large libraries has thus been considerably reduced. However, it still takes several weeks to months to complete a screen, and the typical hit rate is around 1%. The success of an HTS screening largely depends on the design of the assay and the achieved statistical performance. The robustness of the assay can be assessed using a statistical parameter like the Z' score (Zhang, 1999). This parameter has been widely used to determine the suitability of an HTS assay but other parameters like the distribution of standard deviation have been described (Hanley, 2019)

HTS was the gold standard in the 1990s and gave good results. Today it is used alongside other approaches, like structure-based drug design or other computational techniques (Macarrón and Hertzberg, 2011). In order to be robust and not too expensive, an HTS assay should not comprise too many steps. To that effect, several homogeneous-phase assays have been developed, in which all reagents are mixed and no washing step is required. Among the homogeneous-phase techniques, HTRF (Homogeneous Time Resolved Fluorescence) is widely used (Gotoh et al., 2010; Shin et al., 2023). This technique is based on the transfer of energy between two fluorophores, a donor and an acceptor. This transfer occurs when both fluorophores are in close proximity, resulting in a measurable fluorescent signal. This kind of methodology is used for a number of applications, such as the detection of protein-protein interactions (each fluorophore being linked to one of the proteins), enzymatic activities or receptor binding. Another technology named ALPHAscreen (Amplified Luminescent Proximity Homogenous Assay) is also based on a signal obtained when two entities are in close proximity or linked. In this assay the donor and the acceptor are microbeads that are brought together by the molecular interaction of the binding partners that are linked to these beads. Fluorescence polarization which measures the rate of rotation of a fluorescent-labeled ligand is also a powerful method to identify hits and to obtain information during the optimization process (Lea and Simeonov, 2011; Hua et al., 2023).

2.1.2 Cell-based assay

Cell-based assays also played a crucial role in the identification and validation of bioactive compounds serving as versatile tools to assess cellular responses to various stimuli and compounds. These assays use living cells to investigate drug efficacy on cell viability, proliferation or specific cellular functions. These assays differ from phenotypic screening in terms of complexity of the readouts. While phenotypic assays generally involve the simultaneous analysis of multiple cellular parameters (shape, size, surface, biomarkers of specific pathways), cell-based assays focus on a single parameter. One notable example is the MTT (3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide) assay, where the reduction of a yellow tetrazolium salt to purple formazan product by metabolically active cells is measured, providing an indirect assessment of cell viability.

Specific target-based screens like protein-protein interactions can also be conducted using cell-based assays. For example, bimolecular complementation assays, also known as PCA (Protein-fragments Complementation assays) have been developed in the last decade (Kodama and Hu, 2012; Sharma and Anand, 2016; Bellón-Echeverría et al., 2018). In these assays, a fluorescent reporter protein is divided into two non-functional fragments. Each fragment is then fused to two proteins of interest. When these proteins are in close proximity, the split fragments of the reporter protein reassemble, restoring its functionality and resulting in a measurable signal. One of the limitations of the system is that complex reassembly is irreversible; hence, more dynamic systems have recently been developed (Tebo and Gautier, 2019).

Protein-protein interactions within cells can also be monitored, for instance using energy transfer as in FRET (Fluorescence Resonance Energy Transfer) (Song et al., 2011) or BRET (Bioluminescence Resonance Energy Transfer) (Machleidt et al., 2015; Cho and Dalby, 2021). Both methodologies rely on the transfer of energy between a donor and an acceptor, one being a bioluminescent protein in the case of BRET. The major advantage of cell-based assays is that it addresses the activity of a candidate compound in a specific cellular context. If an effect is detected this means that the compound is able to cross the cellular membrane and to reach a target. Additionally, if cells tolerate the compound, this is a first indication that the compound is not toxic for the cell.

2.1.3 Structural and biophysical methods

Structural an biophysical methods are now systematically integrated into the hit identification and validation process, as well as in the subsequent steps of candidate molecule optimization. Their use for hit identification dates back to the 1990s, via Nuclear Magnetic Resonance (NMR) and X-ray crystallography coupled with computational analyses. Since then, several techniques have been developed or adapted particularly in terms of throughput, and they have become complementary to biochemical or cellular biology methods fostering a positive selection of the compounds. These technologies have provided scientists with important information for the development of compounds, such as evidence that the compound binds to the target, the kinetics of the binding, the affinity (measurement of the strength) of the binding, or thermodynamic parameters. In addition, these techniques can also help to identify the binding mode and the binding pocket of molecules. This information is essential for developing a molecule with the right mode of action. Indeed, as an example an enzyme can be inhibited by a molecule that binds in the active site or at a distal site (allosteric inhibition).

A wide range of techniques is now available and their use in drug discovery has been reviewed elsewhere (Renaud et al., 2016; Holdgate and Bergsdorf, 2021). A majority of these methods focus on analyzing isolated targets, which implies producing and purifying the target, albeit more recent methods can now be performed using cellular extracts. Some of them require labeling of the target with a fluorochrome or use a native unmodified target, the ultimate goal for all these techniques being to demonstrate that a candidate compound binds to the target.

Among the most frequently employed methods, we can mention:

(i) Calorimetry techniques (like Isothermal Titration Calorimetry (ITC) or Differential Scanning Calorimetry (DSC)

provides thermodynamic data about the protein-ligand complex. For example, ITC measures the consumption or generation of heat when a compound binds to the protein (Falconer et al., 2021). (ii) Temperature-related intensity change that measures the modification of fluorescence intensity of a fluorochrome when the target and the compound are bound (Jerabek-Willemsen et al., 2014). (iii) NMR that relies on the behavior of certain atomic nuclei when placed in a strong magnetic field and exposed to a specific frequency of radiofrequency radiation (Shimada et al., 2019). (iv) Surface plasmon resonance detects changes in the refractive index near a metal surface. (v) Mass spectrometry that determines the mass-to-charge ratio of ions (Gavriilidou et al., 2022) (vi) X-ray diffraction that measures the diffraction angles and changes of intensities of X-rays can be applicable to crystals (X-ray crystallography) (Maveyraud and Mourey, 2020) but also to proteins in solution like enzymes (Byer et al., 2023) (vii) cryo-electron microscopy (cryo-EM) is a powerful technique used for imaging macromolecules at near atomic resolution. This technique is now complementary to NMR or X-ray diffraction in small molecule drug design ((Vénien-Bryan et al., 2017; Renaud et al., 2018)

These techniques all rely on high-standard equipment, and depending on the method employed the throughput can vary from a few compounds a day to a few compounds a week or month. These methodologies are part of the drug discovery process from the early phases to the selection of the preclinical candidate therapy.

2.2 Phenotypic-based screening

Historically, the discovery of medicines relied on phenotypic approaches, however with the advent of genomics in the 1980s and the sequencing of the human genome in 2001, these approaches were neglected. Nevertheless, over the last decade there has been renewed interest in phenotypic approaches, as they are valuable at identifying novel therapeutic agents (Ege et al., 2021; Vincent et al., 2022). One of the advantages of phenotypic assays is that they explore a broader spectrum of biological responses than target-based approaches, elucidating complex biological pathways and uncovering unforeseen interactions, thus offering a holistic perspective of the potential effect of a new agent. Technological advances have played a pivotal role in boosting phenotypic screening. Assay miniaturization, development of high-throughput screening platforms (gathering automated equipment to rapidly test a huge quantity of samples rapidly), automated imaging (microscopy technology) and data analysis systems have opened new avenues to perform phenotypic analyses. One such technique is fluorescence imaging, which enables scientists to visualize and quantify various biological processes at the cellular and subcellular levels. This technique uses fluorescent probes, markers or genetically-encoded fluorescent proteins to highlight specific cellular structures, proteins, or functional activities. High content screening (a combination of powerful imaging tools and biochemical/molecular biology assays) captures dynamic cellular events in real-time; for instance, the monitoring of processes like cell migration, proliferation, modification of the cellular morphology (shape, size…) and cell death. Additionally, it allows the concomitant assessment of multiparametric data including protein localization, analysis of subcellular organelles or responses to external stimuli. Numerous approved therapies for cardiovascular diseases, viral infections, neurodegenerative disorders and cancers originate from phenotypic screening (Blay et al., 2020). Despite, many advantages that have led to the identification of innovative therapies that could not have been identified without this approach, phenotypic screening has one major drawback–this global approach makes it difficult to decipher the molecular mechanisms of action of a drug and to identify its target(s), both necessary to optimize the potency of a drug and for its further development in the clinic.

3 Conclusion

Drug discovery is a long and challenging process which involves various fields of expertise. A crucial step of the development of a new small organic-based therapy is the identification of hits. Target-based, phenotypic-based and biophysical methods can be employed throughout the process to identify these hits and to participate in the optimization and development process of a new promising therapy. Despite significant advances in scientific and technological methods, the identification and development of new therapies remain arduous and resource-intensive. The journey from identifying bioactive molecules to developing a marketable drug involves an intricate interplay of biology, biophysics, chemistry and cutting-edge technologies. As science continues to advance, the hope is to streamline this process, making drug discovery more efficient and accessible for the benefit of patients.

Author contributions

SG: Writing–original draft, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The author received funding from the Synergy Lyon Cancer foundation, the Centre Léon Bérard, the Lyrican+ (grant INCa-DGOS-INSERM-ITMO cancer_18003) and the Institut Convergence Plascan (Grant Number ANR-17-CONV-0002).

Acknowledgments

The author thanks reviewers for their comments and suggestions on the manuscript and Brigitte Manship for her help in proofreading the manuscript. Figure 1 was Created with BioRender.com.

Conflict of interest

SG is co-founder of TheraPPI Bioscience, a spin-off of his host academic institutions.

The handling editor BOV declared a shared affiliation with the author SG at the time of the review.

Publisher's note

References

Bellón-Echeverría, I., Carralot, J.-P., Del Rosario, A. A., Kueng, S., Mauser, H., Schmid, G., et al. (2018). MultiBacMam Bimolecular Fluorescence Complementation (BiFC) tool-kit identifies new small-molecule inhibitors of the CDK5-p25 protein-protein interaction (PPI). *Sci. Rep.* 8, 5083. doi:10.1038/s41598-018-23516-x

Blay, V., Tolani, B., Ho, S. P., and Arkin, M. R. (2020). High-Throughput Screening: today's biochemical and cell-based approaches. *Drug Discov. Today* 25, 1807–1821. doi:10.1016/j.drudis.2020.07.024

Byer, A. S., Pei, X., Patterson, M. G., and Ando, N. (2023). Small-angle X-ray scattering studies of enzymes. *Curr. Opin. Chem. Biol.* 72, 102232. doi:10.1016/j.cbpa.2022.102232

Cho, E. J., and Dalby, K. N. (2021). Luminescence energy transfer–based screening and target engagement approaches for chemical biology and drug discovery. *SLAS Discov.* 26, 984–994. doi:10.1177/24725552211036056

Ege, N., Bouguenina, H., Tatari, M., and Chopra, R. (2021). Phenotypic screening with target identification and validation in the discovery and development of E3 ligase modulators. *Cell Chem. Biol.* 28, 283–299. doi:10.1016/j.chembiol.2021.02.011

Falconer, R. J., Schuur, B., and Mittermaier, A. K. (2021). Applications of isothermal titration calorimetry in pure and applied research from 2016 to 2020. *J. Mol. Recognit.* 34, e2901. doi:10.1002/jmr.2901

Gavriilidou, A. F. M., Sokratous, K., Yen, H.-Y., and De Colibus, L. (2022). High-throughput native mass spectrometry screening in drug discovery. *Front. Mol. Biosci.* 9, 837901. doi:10.3389/fmolb.2022.837901

Gns, H. S., Gr, S., Murahari, M., and Krishnamurthy, M. (2019). An update on Drug Repurposing: Re-written saga of the drug's fate. *Biomed. Pharmacother.* 110, 700–716. doi:10.1016/j.biopha.2018.11.127

Gotoh, Y., Nagata, H., Kase, H., Shimonishi, M., and Ido, M. (2010). A homogeneous time-resolved fluorescence-based high-throughput screening system for discovery of inhibitors of IKKbeta-NEMO interaction. *Anal. Biochem.* 405, 19–27. doi:10.1016/j.ab.2010.05.028

Hanley, Q. S. (2019). The distribution of standard deviations applied to high throughput screening. *Sci. Rep.* 9, 1268. doi:10.1038/s41598-018-36722-4

Hoffer, L., Muller, C., Roche, P., and Morelli, X. (2018). Chemistry-driven hit-to-lead optimization guided by structure-based approaches. *Mol. Inf.* 37, 1800059. doi:10.1002/minf.201800059

Holdgate, G. A., and Bergsdorf, C. (2021). Applications of biophysics in early drug discovery. *SLAS Discov.* 26, 945–946. doi:10.1177/24725552211035123

Hua, L., Wang, D., Wang, K., Wang, Y., Gu, J., Zhang, Q., et al. (2023). Design of tracers in fluorescence polarization assay for extensive application in small molecule drug discovery. *J. Med. Chem.* 66, 10934–10958. doi:10.1021/acs.jmedchem.3c00881

Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H. M., Duhr, S., Baaske, P., et al. (2014). MicroScale thermophoresis: interaction analysis and beyond. *J. Mol. Struct.* 1077, 101–113. doi:10.1016/j.molstruc.2014.03.009

Kodama, Y., and Hu, C.-D. (2012). Bimolecular fluorescence complementation (BiFC): a 5-year update and future perspectives. *BioTechniques* 53, 285–298. doi:10.2144/000113943

Lea, W. A., and Simeonov, A. (2011). Fluorescence polarization assays in small molecule screening. *Expert Opin. Drug Discov.* 6, 17–32. doi:10.1517/17460441.2011.537322

Macarrón, R., and Hertzberg, R. P. (2011). Design and implementation of high throughput screening assays. *Mol. Biotechnol.* 47, 270–285. doi:10.1007/s12033-010-9335-9

Machleidt, T., Woodroofe, C. C., Schwinn, M. K., Méndez, J., Robers, M. B., Zimmerman, K., et al. (2015). NanoBRET—a novel BRET platform for the analysis of protein–protein interactions. *ACS Chem. Biol.* 10, 1797–1804. doi:10.1021/acschembio.5b00143

Maveyraud, L., and Mourey, L. (2020). Protein X-ray crystallography and drug discovery. *Molecules* 25, 1030. doi:10.3390/molecules25051030

Naeem, A., Hu, P., Yang, M., Zhang, J., Liu, Y., Zhu, W., et al. (2022). Natural products as anticancer agents: current status and future perspectives. *Molecules* 27, 8367. doi:10.3390/molecules27238367

Newman, D. J., and Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* 83, 770–803. doi:10.1021/acs.jnatprod.9b01285

Renaud, J.-P., Chari, A., Ciferri, C., Liu, W., Rémigy, H.-W., Stark, H., et al. (2018). Cryo-EM in drug discovery: achievements, limitations and prospects. *Nat. Rev. Drug Discov.* 17, 471–492. doi:10.1038/nrd.2018.77

Renaud, J.-P., Chung, C., Danielson, U. H., Egner, U., Hennig, M., Hubbard, R. E., et al. (2016). Biophysics in drug discovery: impact, challenges and opportunities. *Nat. Rev. Drug Discov.* 15, 679–698. doi:10.1038/nrd.2016.123

Sharma, H., and Anand, B. (2016). Fluorescence bimolecular complementation enables facile detection of ribosome assembly defects in *Escherichia coli*. *RNA Biol.* 13, 872–882. doi:10.1080/15476286.2016.1207037

Shimada, I., Ueda, T., Kofuku, Y., Eddy, M. T., and Wüthrich, K. (2019). GPCR drug discovery: integrating solution NMR data with crystal and cryo-EM structures. *Nat. Rev. Drug Discov.* 18, 59–82. doi:10.1038/nrd.2018.180

Shin, Y. H., Kim, D.-E., Yu, K. L., Park, C. M., Kim, H. G., Kim, K.-C., et al. (2023). A novel time-resolved fluorescence resonance energy transfer assay for the discovery of small-molecule inhibitors of HIV-1 tat-regulated transcription. *Int. J. Mol. Sci.* 24, 9139. doi:10.3390/ijms24119139

Singh, N., Chaput, L., and Villoutreix, B. O. (2020). Virtual screening web servers: designing chemical probes and drug candidates in the cyberspace. *Brief. Bioinform.* 22, 1790–1818. doi:10.1093/bib/bbaa034

Song, Y., Madahar, V., and Liao, J. (2011). Development of FRET assay into quantitative and high-throughput screening technology platforms for protein–protein interactions. *Ann. Biomed. Eng.* 39, 1224–1234. doi:10.1007/s10439-010-0225-x

Tebo, A. G., and Gautier, A. (2019). A split fluorescent reporter with rapid and reversible complementation. *Nat. Commun.* 10, 2822. doi:10.1038/s41467-019-10855-0

Vemula, D., Jayasurya, P., Sushmitha, V., Kumar, Y. N., and Bhandari, V. (2023). CADD, AI and ML in drug discovery: a comprehensive review. *Eur. J. Pharm. Sci.* 181, 106324. doi:10.1016/j.ejps.2022.106324

Vénien-Bryan, C., Li, Z., Vuillard, L., and Boutin, J. A. (2017). Cryo-electron microscopy and X-ray crystallography: complementary approaches to structural biology and drug discovery. *Acta Crystallogr. Sect. F. Struct. Biol. Commun.* 73, 174–183. doi:10.1107/S2053230X17003740

Vincent, F., Nueda, A., Lee, J., Schenone, M., Prunotto, M., and Mercola, M. (2022). Phenotypic drug discovery: recent successes, lessons learned and new directions. *Nat. Rev. Drug Discov.* 21, 899–914. doi:10.1038/s41573-022-00472-w

Wouters, O. J., McKee, M., and Luyten, J. (2020). Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *JAMA* 323, 844–853. doi:10.1001/jama.2020.1166

Yu, W., and MacKerell, A. D. (2017). "Computer-aided drug design methods," in *Antibiotics, methods in molecular biology*. Editor P. Sass (New York, NY: Springer New York), 85–106. doi:10.1007/978-1-4939-6634-9_5

Zhang, J.-H., Chung, T. D., and Oldenburg, K. R. (1999). A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screen.* 4, 67–73. doi:10.1177/108705719900400206

frontiers | Frontiers in Drug Discovery

TYPE Mini Review
PUBLISHED 31 January 2024
DOI 10.3389/fddsv.2024.1347246

OPEN ACCESS

EDITED BY
Bruno Villoutreix,
Institut National de la Santé et de la Recherche Médicale (INSERM), France

REVIEWED BY
Jeremy J. Yang,
University of New Mexico, United States

*CORRESPONDENCE
Lindsay Marshall,
lmarshall@hsi.org

RECEIVED 30 November 2023
ACCEPTED 15 January 2024
PUBLISHED 31 January 2024

CITATION
Marshall LJ and Conlee KM (2024), The case of the missing mouse—developing cystic fibrosis drugs without using animals.
Front. Drug Discov. 4:1347246.
doi: 10.3389/fddsv.2024.1347246

The case of the missing mouse—developing cystic fibrosis drugs without using animals

Lindsay J. Marshall* and Kathleen M. Conlee

Animal Research Issues, The Humane Society of the United States, Gaithersburg, MD, United States

Creating and developing new drugs can take decades, costs millions of dollars, requires untold human effort and usually, takes thousands of animal lives. Despite regulators professing confidence in non-animal approaches and guidance documents that permit submission of non-animal data, toxicity testing is routinely carried out in animals, employing rodents (invariably mice) and non-rodents. However, extensive preclinical testing in animals is still no guarantee that drugs will be safe and/or effective. In fact, more than nine out of every ten drugs that appear safe from animal trials will fail when tested in people, often due to unexplained toxicity or a lack of efficacy. This paper will describe recent advances in drug development where non-animal approaches have been used, to explore how and where these could be applied more widely to revolutionize the drug development pipeline and accelerate the creation of safe and effective medicines. As one case study, we look at the small molecule channel modifiers developed to address the consequences of the mutated chloride channel in the fatal genetic condition, cystic fibrosis. We then take a closer look at where drug development could be accelerated by focusing on innovative, human biology-based testing methods. Finally, we put forward recommendations, targeting all stakeholders, including the public, that will be needed to put this into practice and enable drug development to become more efficient - focusing on human-biology based testing and cutting out the middle-mouse.

KEYWORDS

cystic fibrosis, *in vitro*, drug development, cystic fibrosis transmembrane conductance regulator, new approach methodologies

1 Introduction

"The child with the salty brow shall die." This was the prognosis for people born with cystic fibrosis (CF), before the condition had even been named. There was terrible recognition that tasting salt when you kissed your baby was a harbinger of doom. Decades of tireless research by (among others) Dorothy Andersen and Paul di Sant'Agnese provided the key observations that children dying of this condition were not producing digestive enzymes from their pancreas (Andersen, 1938); that the levels of salt (sodium chloride) in their sweat was much higher than healthy children (Di Sant'Agnese et al., 1953) and this led to the condition being named as cystic fibrosis. In 1989, an important breakthrough occurred when the gene responsible for CF disease was identified (Rommens et al., 1989). This study revealed that the "CF gene," encoded a protein called the cystic fibrosis transmembrane conductance regulator (CFTR), responsible for regulating chloride ion movement out of cells.

1.1 Understanding cystic fibrosis

Identification of the gene was a crucial advance that was hailed as the last piece of the puzzle in the search for a cure for CF. However, as is customary for modelling human genetic diseases, the identification of a specific, causative gene leads to the creation of multiple "artificial" animal models, since animals do not naturally have CF. Identification of the gene associated with CF was confirmed in 1989 and the first animal model was created a couple of years later (Snouwaert et al., 1992). Animal models of CF initially used mice, but since these manipulated mice failed to fully recapitulate the human condition, and the animals did not have symptoms of the lung disease, this expanded to other animals including rats, ferrets, sheep and pigs. There are now more than 750 different genetic animal models of CF, of which more than 690 use mice (Leenaars et al., 2020). Since animals do not have CF naturally, there has to be some manipulation of the animal in order to generate symptoms of the disease and this requires either genetic mutation techniques, used for the genetic models, or non-genetic approaches, including the use of drugs, deliberate infection with pathogens, or grafting human tissues into the animals' lungs. There are more than 220 non-genetic animal models for CF, again the majority of these use mice, but there are also rats, pigs, monkeys, and rabbits (Leenaars et al., 2021).

The Canadian database which logs the different mutations in CFTR that give rise to CF currently stands at 2,114 (http://www.genet.sickkids.on.ca/StatisticsPage.html). Even before considering the ethics and scientific relevance of this exercise, the costs and time required to develop an animal model for each of these mutations would be exorbitant. One commercial service offering development of genetically modified animals charges a basic fee of around $4,000 for a single mutation and this does not include technical support (such as animal care and surgery) or consumables (e.g., surgical equipment) (https://brcf.medicine.umich.edu/cores/transgenic-animal-model/fees/; https://www.umassmed.edu/globalassets/transgenic-animal-modeling-core/documents/whycostsomuch2015.pdf). To create one animal version of each CF mutation found in people would therefore cost more than 8 million US dollars. This is a very conservative estimate since costs would be greater for larger animals (rabbits costs at least twenty times more than mice and mini pigs are over one hundred times more costly than mice https://minipigs.dk/products-services/enquiries) and it is more expensive and time consuming to create the complex models which would more accurately represent the patient population (almost half of the UK CF population have two different mutations https://www.cysticfibrosis.org.uk/sites/default/files/2020-12/2019%20Registry%20Annual%20Data%20report_Sep%202020.pdf). Developing pig models of the spectrum of CF mutations found in humans would cost upwards of 190 million US dollars.

However, CF is a great example of how human-centred interventions (as opposed to animal-based research) have played a great role in improving understanding of CF such that the bleak prognosis for the child with the salty brow has improved over time. In the 1960s and '70s, increased survival was associated with the creation of dedicated CF centers, these were crucial for sharing best practices for nutrition and physiotherapy and for fast, aggressive treatment of infections. With these interventions and the knowledge sharing offered by the dedicated networks of CF physicians, parents and families, the median survival age for babies born with CF reached 11 years in the 70s and is still on the increase (Matthews et al., 1964; Doershuk et al., 1965) (Figure 1).

1.2 The drug development pipeline today

Traditionally, animals are used in drug development (Figure 2) to demonstrate the safety of the preparation (for a comprehensive review of the drug development process, see the paper by Singh et al. in this edition (Singh et al., 2023)). As stated by Dr van Norman in her 2019 article about animal use for drug safety testing, "[t]here is no doubt that the use of animals in science and medicine has significantly benefitted human beings." However, despite an historical reliance on animals for medical advances, drug attrition rates of more than 90% (Sun et al., 2022) indicate that animals are not predictive for humans and a drug that is safe for animals is not always safe (or effective) in people (Van Norman, 2019; Van Norman, 2020). Unfortunately, it is not possible to carry out a stringent comparison of the possible failure rates if drugs were tested using the non-animal approaches. The data submitted by the pharmaceutical companies as part of the regulatory approval are not publicly available and whilst the current legislative requirements do not demand animal data, many of the drug safety guidelines to which pharmaceutical companies refer make reference to submission of preclinical pharmacology data from animals. It is therefore not possible to calculate the likely "failure rate" of a drug tested solely on human based methods or using human data. Despite a mounting body of evidence that the human based tools are (and will be) more predictive than animals, we cannot simply assume that the non-animal tests will always accurately predict human responses. However, as data from human relevant tools accumulate, this may change. Indeed, the case study of CF presented here indicates that patient-derived samples and developing tests using human cells offer insight into the biological effects of potential drugs and can revolutionise treatments.

The new approach methodologies (NAMs) are innovative methods that no longer rely on the use of live animals, these include human cells, tissues and organs, organ-chips or microphysiological systems (MPS), use of human data or human volunteers, and computer modelling. When considering the uniquely human nature of a disease, these methods offer a more relevant and physiologically accurate approach to understanding the disease features and therefore developing an effective treatment. They also provide more confidence in the predictivity of the data obtained—recent research has shown that MPS can recreate species-specific effects (Jang et al., 2019) and that MPS using human liver cells are better able to predict toxic drugs than animal testing (Ewart et al., 2022). In fact, analysis of possible savings that could be realised if human liver MPS were included for safety testing within the drug development pipeline revealed that these might reach around 3 billion USD in drug development costs (Ewart et al., 2022).

For drug safety testing, there are national and international guidance documents that have to be adhered to, and these indicate that animal studies should be undertaken (e.g., the International Council on Harmonisation guidance M3 (R2) for "Non-clinical safety studies for the Conduct of Human Clinical Trials and Marketing Authorization for Pharmaceuticals" states that "The development of a pharmaceutical is a stepwise process involving an evaluation of both animal and human efficacy and safety information" and goes on to detail what number and species of animals should be used throughout the process U. S. Food and Drug

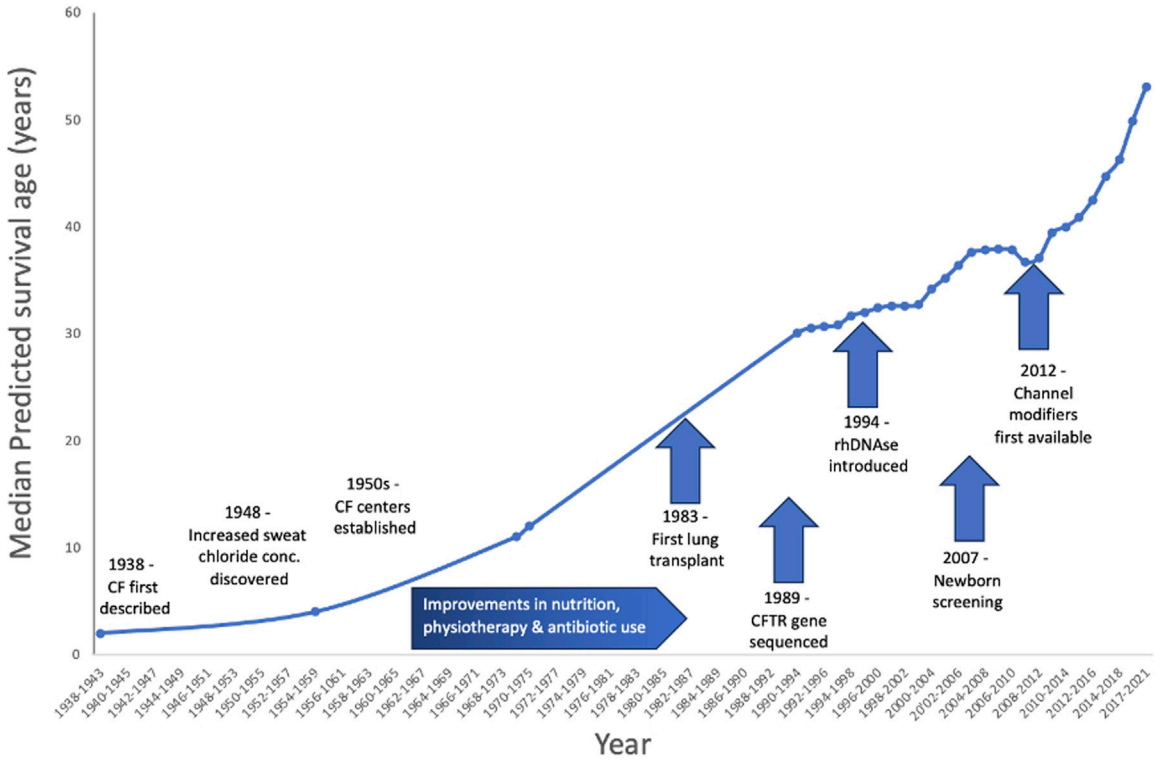


FIGURE 1
Increase in predicted survival over time for people born with CF. Improvements and advances in physiotherapy, antibiotic use and even lung transplantation have all contributed to this increase. However, aside from drugs to reduce the thick, sticky mucus and improvements in antibiotics, there were no specific treatments for CF - until 2012, when the first CFTR channel modifiers were approved for clinical use–note the steep increase in life expectancy since that time.

Administration). However, many of the regulatory agencies claim that they do not require data from animals and that they would be willing to examine data from the non-animal methods and the UK government recently declared that ["t]here is no United Kingdom legislation that mandates animal testing" (UK Parliament, 2023). Indeed, the Food and Drug Administration Modernization Act 2.0, which was signed into US law in 2022, permits drug developers to make use of "certain alternatives to animal testing, including cell-based assays and computer models, to obtain an exemption from the Food and Drug Administration to investigate the safety and effectiveness of a drug" (One hundred and seventeenth Congress of the United States, 2023). Likewise, for drug efficacy testing, (these are studies undertaken to evaluate whether a treatment will be effective), the regulatory agencies do not insist on animal data although animal models of disease are often used here, and so it seems that the door is open for non-animal data to be submitted to support product registration.

1.3 *In vitro* breakthroughs for CF show the way forwards

In 2006, researchers at Vertex first used human cell cultures with bronchial epithelial cells (taken from the upper airways of people) to develop a method to measure the activity of the CFTR channel (Van Goor et al., 2006). They used this human-relevant tool to screen hundreds of thousands of different chemicals and identify compounds that would increase CFTR activity. This assay revealed many hundreds of compounds with the desired activity, and the researchers then got rid of the ones that were unlikely to be successful drugs based on chemistry (through comparing structures to drugs that have been recognised as toxic). Fifty-three compounds were selected for further investigation using intestinal organoid models—note that organoids are tiny, cell-based models that retain the structure and function of the "parent" organ (Dekkers et al., 2013). Accessing bronchial epithelial cells is quite difficult and can require sedation, making this a complicated procedure to carry out in very young children (the intended patient population), but the cells to create intestinal organoids are more easily accessible and patients have reported limited discomfort with this technique (Servidoni et al., 2013). Thus, intestinal organoids offer a robust method for screening hundreds (or even thousands) of potential drugs. Using this approach, several likely compounds can be rapidly and efficiently assessed using the intestinal organoids, and the activity can be verified with the airways model so that the most promising candidate(s) can quickly move to the clinic.

1.4 Personalising medicine with personalised tools

Another advantage to using intestinal organoids from people with CF is that this offers a simple system with which to reveal

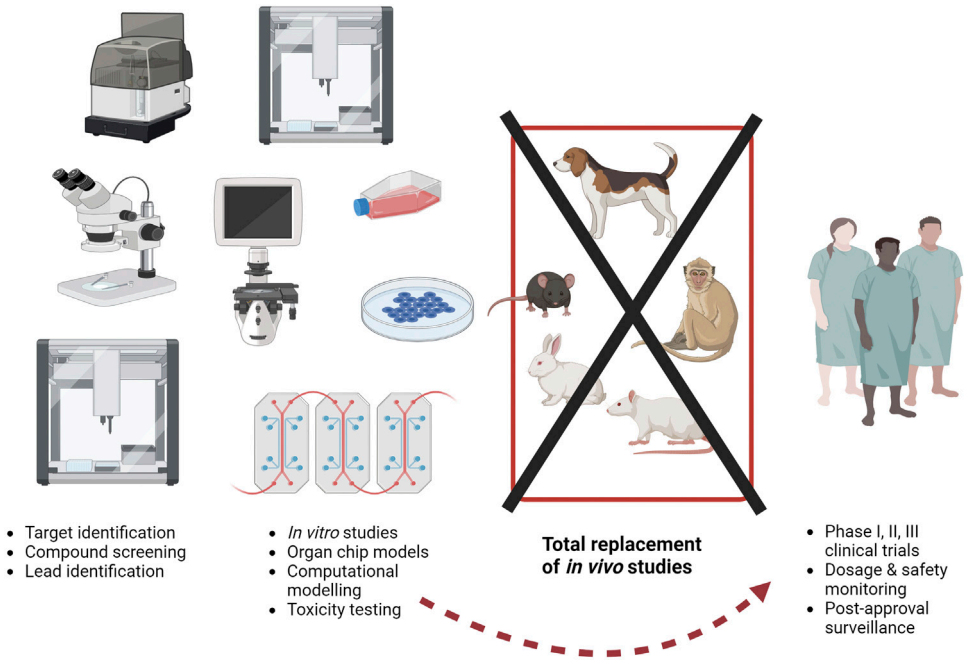


FIGURE 2
The current drug development paradigm is time consuming and costly, partly due to a reliance on animals. Estimates suggest around 10 to 5 years and 518 million to tens of billions of US dollars are required to progress from target identification (at the left of the figure) to regulatory approval (at the right pf the figure). As of today, animals may be used throughout this process, but the use of animal models is no guarantee of success, as even after approval, severe adverse effects in the extended patient population can result in drug withdrawal. More sustained, dedicated funding to the methods described on the left here (*in vitro* studies etc.) could help to bypass animal studies completely. Image created with BioRender.com.

patient specific treatments. Organoids are three dimensional structures, so that when CFTR is active, fluid influx leads to swelling of the organoid and measuring the volume of the organoid is a simple way to know whether the drug applied to the organoid is activating CFTR (where increased volume indicates active CFTR). Using this approach, patient-specific theratyping is possible, where organoids from patients can be screened against multiple compounds, or combinations of compounds, to find the optimal dose and identity of drugs to give each individual patient (Clancy et al., 2019; Conti et al., 2022). The major advantage is that the organoid retains the exact same genetic mutation in CFTR as the person, and so "treating" the organoid reflects what will happen in that individual. This testing is just not possible using animal models, or at least it would require that an animal model was available for every single individual with CF, which would take time and money and thousands of animal lives (as described above), and still does not guarantee success, given that animals do not exhibit the lung disease that we are trying to treat! Thus, adoption of the human cell-based intestinal organoid models as a screening platform makes this a very efficient way of screening individual people with CF, to ensure that the medication given will be effective, and therefore can be used to personalise treatment.

1.5 Non-animal methods could indicate where drugs won't work for patients

The example of curcumin demonstrates another way in which *in vitro*, human cell-based tools can be used to help advance drug discovery, or at least to prevent false hope. Curcumin is a derivative of the spice turmeric; this was force-fed to genetically modified CF mice (who bore the mutation most frequently found in people with CF) for 3 days, before the electrical activity of their airways and intestines was measured, as an indicator of CFTR activity. The study also used cultures of hamster kidney cells with mutated CFTR inserted in them and tracked the processing of the CFTR protein, showing that curcumin enhanced the insertion of CFTR in the appropriate cellular localisation (Egan et al., 2004). These results suggested that curcumin could correct the CFTR-dependent deficit but unfortunately, this was not recapitulated for people with CF. When the studies were repeated with airways epithelial cells isolated from someone with CF, curcumin did not increase the electrical activity (Song et al., 2004). These data are evidence of the issue with translational failures–where data in animals are not recapitulated in people and indicate that caution is needed in interpreting data from non-human model systems. We suggest that the incorporation of more human-relevant tools, such as the airways models or intestinal organoids, in the drug discovery pipeline, would help reduce these failures.

1.6 Comparing patients and mutations allows wider drug use

The non-animal, human relevant tools can be used to reveal insights into the association between physiology and disease symptoms, that are not possible in animals, and that allow

researchers to estimate the likely clinical effect of a drug. For example, clinical data comparing the electrical potential difference across the airways in people with or without CF indicates that increasing the amount of CFTR in the cell membrane by around 15%–30% would help to "normalise" this electrical activity (which is associated with ion movement, and therefore CFTR activity) in CF. Researchers can then test potential drugs, existing drugs and combinations of these, in human airways cell models to find those compounds that increase membrane CFTR and enhance CFTR activity. An additional advantage of using human cell-based systems is the ability to carry out direct comparisons. Models developed using CF cells can be compared with non-CF cultures, but importantly models from people with CF who have different mutations in their CFTR can be compared. As we have already mentioned above, research with genetically engineered animal models is almost prohibitively expensive, time consuming, and cannot accurately model human responses, whereas the human cell-based approach allows researchers to clarify the relationship between the mutation in CFTR and its activity, and therefore to make intelligent decisions regarding which CFTR modifiers may be helpful for individual people with CF. Although there is one specific mutation that affects the majority of people with CF (Kerem et al., 1989), as described above, there are over 2000 different mutations. *In vitro*, cell-based methods offer a more efficient manner with which to test the effects of drugs on different mutations and combinations of mutations (since people with CF may have two different mutations). It is very gratifying to see this potential utility of non-animal models reflected at some level of regulations, as the FDA granted a "label extension" for one of the CFTR modifiers, based purely on data from human cell models (Ratner, 2017), enabling the broader use of this drug for many more people with CF.

1.7 Drug repurposing with cell-based tools

The realisation that patient-derived intestinal organoids could be used to filter hundreds or compounds to detect possible CFTR modifiers for treating CF led to another important advance - that of the application of human relevant tools in drug repurposing, an approach suitable for applications beyond CF. Drug repurposing is an efficient way of finding new treatments as it employs existing, approved drugs for a purpose other than that for which they were approved, and so this bypasses the lengthy, expensive safety testing that is needed for entirely new compounds.

This extends beyond CF: there are other examples of drug repurposing where the efficacy data were obtained using human cell-based approaches. For example, for SARS-CoV2, researchers tested seven clinically approved drugs on airways models made of human cells, measuring the ability of these drugs to prevent the virus getting inside the cells (Si et al., 2020). When the drugs were tested at concentrations and flow rate equivalent to those found in human blood, they found that only two of them showed great promise in terms of preventing infection and, soon after, a clinical trial was set up to assess the effects of one of these drugs in people with COVID. Additionally, a human stem cell-based system showed that a biological therapeutic could be effective in a rare neuropathy disease (Rumsey et al., 2022) and data from an organ chip demonstrated that a medication used in Type 2 diabetes could prevent chemotherapy-induced kidney toxicity (Cohen et al., 2021). This is hopefully the start of a shift in the regulatory paradigm and is indicative of enhancing flexibility to enable accelerated access to safe and effective drugs for all patients.

2 Discussion

This paper uses CF as an example to show how the incorporation of non-animal tools into the drug development can be transformational. Two examples of human cell-based systems presented here -namely the use of intestinal organoids and airways epithelial cell cultures—revealed the CFTR modifying activity of small molecule therapeutics that have gone on to revolutionise life for people with CF. Given the current regulatory requirements (detailed in Figure 2) the CFTR modifier drugs were tested (for toxicity) in animals, but there was no way that animal models of CF, given the lack of respiratory involvement in mice, for example, could show the efficacy that was needed to give the confidence that these drugs would be disruptive for people living with CF and for parents of babies born with CF.

3 Recommendations

Finally, by using cystic fibrosis as a case study, we offer a few brief recommendations for some of the points where we see implementation or action is needed to enable wider use of the non-animal methods for drug development. For the purpose of this review, we have focused on the development and application of the non-animal approaches. We appreciate the historical advances that occurred as a result of animal use, including, for example, the discovery of insulin in 1921 and the advent of the polio vaccine in 1953. However, we also believe that the successful, historical use of animals, particularly in the face of rapidly evolving non-animal technologies, does not scientifically justify their continued use. This reflects the viewpoints reported in notable reviews such as the report on the use of dogs as subjects of biomedical research (The National Academies of Sciences, Engineering, and Medicine, 2020) and the general public, who are invested in the use of non-animal technologies to replace animals (Savanta, 2022). Therefore, the recommendations below are directed at accelerating the replacement of animals rather than considering how and where animal use should be continued.

Regulatory transparency is key-the agencies should be encouraged to publicly publish where they have accepted data form non-animal methods, to ensure that animal use is not duplicated.

Biobanking (where patient samples or volunteer biological material is curated and stored for widespread use by researchers) should be incentivized. This is particularly important for rare disease like CF, where clinical trials are limited by the number of patients with the same mutation in a specific geographic location. Providing global access to the tissue, and developing methods for developing assays based on these tissues, would help to address this disparity. It is also important to ensure that the biological material deposited in

the biobanks fully reflects the differences in the human population (Ghosh et al., 2022).

Funding agencies need to analyse where their funding is not providing the expected, or acceptable, return on investment. This may require retrospective reviews of projects but would be invaluable in identifying where projects or topics are failing to deliver and therefore could inform future funding strategies. Presently, over 90% of drugs fail in people (Sun et al., 2022) and a proportion of this failure is directly related to the use of animals as models (Van Norman, 2019). Understanding where the animal models continually fail enables the decision to no longer fund this sort of research, and could allow diversion of the award money to *in vitro*, epidemiological, or computer modelling based research that could help to advance the field. This is an issue that should resonate with the public, given their role in research. In the UK in 2018, £1.2billion funding originated from medical charities, representing around 14% of all health related research funding in the UK (Fraser of Allander Institute, 2021). In the US, the largest funder of biomedical research is the National Institutes of Health (NIH), which relies on taxpayer dollars. The NIH awards an estimated half of its total budget of over 47 billion dollars to animal-based research (U.S. Government Accountability Office, 2019), yet our analysis has revealed that less than 1% of this is dedicated to organ chips, for example. The public are therefore heavily invested in research, either voluntarily through charitable donations, or through their taxes and the same public is vocal in its desire to move away from using animals in medical research (IPSOS MORI, 2018; Savanta, 2022), although this depends on several factors, including the type of animal and purpose of research (Brunt and Weary, 2021).

At the Humane Society of the United States, we are keen to prioritise funding to development and use of the non-animal methods. In 2023 we introduced a bill in Maryland that requires that laboratories using animals have to contribute to a research fund which is available for non-animal method developers. This legislation creates a precedent for the transition toward the non-animal methods of the future. The subsequent funding shift will help to accelerate scientific discovery by allowing for early adoption of promising non-animal methods. However, this shift impacts just one state of the US and we need more. If the NIH and other government funding agencies could make the commitment to shift an annual 5%–10% of their budgets to non-animal research, we might see more advances, like the CFTR modifier drugs, more quickly.

Author contributions

LJM: Conceptualization, Writing–original draft, Writing–review and editing. KMC: Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Andersen, D. H. (1938). CYSTIC FIBROSIS OF THE PANCREAS AND ITS RELATION TO CELIAC DISEASE: A CLINICAL AND PATHOLOGIC STUDY. *Am. J. Dis. Child.* 56, 344–399. doi:10.1001/archpedi.1938.01980140114013

Brunt, M. W., and Weary, D. M. (2021). Public consultation in the evaluation of animal research protocols. *PLoS One* 16, e0260114. doi:10.1371/journal.pone.0260114

Clancy, J. P., Cotton, C. U., Donaldson, S. H., Solomon, G. M., Vandevanter, D. R., Boyle, M. P., et al. (2019). CFTR modulator theratyping: current status, gaps and future directions. *J. Cyst. Fibros.* 18, 22–34. doi:10.1016/j.jcf.2018.05.004

Cohen, A., Ioannidis, K., Ehrlich, A., Regenbaum, S., Cohen, M., Ayyash, M., et al. (2021). Mechanism and reversal of drug-induced nephrotoxicity on a chip. *Sci. Transl. Med.* 13, eabd6299. doi:10.1126/scitranslmed.abd6299

Conti, J., Sorio, C., and Melotti, P. (2022). Organoid technology and its role for theratyping applications in cystic fibrosis. *Child. (Basel)* 10, 4. doi:10.3390/children10010004

Dekkers, J. F., Wiegerinck, C. L., De Jonge, H. R., Bronsveld, I., Janssens, H. M., De Winter-De Groot, K. M., et al. (2013). A functional CFTR assay using primary cystic fibrosis intestinal organoids. *Nat. Med.* 19, 939–945. doi:10.1038/nm.3201

Di Sant'Agnese, P. A., Darling, R. C., Perera, G., and Shea, E. (1953). Abnormal electrolyte composition of sweat in cystic fibrosis of the pancreas; clinical significance and relationship to the disease. *Pediatrics* 12, 549–563. doi:10.1542/peds.12.5.549

Doershuk, C., Matthews, L., and Tucker, A. (1965). Evaluation of a prophylactic and therapeutic program for patients with cystic fibrosis. *Pediatrics* 36, 675–688. doi:10.1542/peds.36.5.675

Egan, M. E., Pearson, M., Weiner, S. A., Rajendran, V., Rubin, D., Glockner-Pagel, J., et al. (2004). Curcumin, a major constituent of turmeric, corrects cystic fibrosis defects. *Science* 304, 600–602. doi:10.1126/science.1093941

Ewart, L., Apostolou, A., Briggs, S. A., Carman, C. V., Chaff, J. T., Heng, A. R., et al. (2022). Performance assessment and economic analysis of a human Liver-Chip for predictive toxicology. *Commun. Med.* 2, 154. doi:10.1038/s43856-022-00209-1

Fraser of Allander Institute (2021). *The contribution of medical funding by charities to the UK economy*. Strathclyde: Univerity of Strathclyde Business School.

Ghosh, S., Nehme, R., and Barrett, L. E. (2022). Greater genetic diversity is needed in human pluripotent stem cell models. *Nat. Commun.* 13, 7301. doi:10.1038/s41467-022-34940-z

IPSOS MORI (2018). *Public attitudes to animal research in 2018*. UK: IPSOS.

Jang, K. J., Otieno, M., Ronxh, I. J., Lim, H., Ewart, L., Kodella, K., et al. (2019). Reproducing human and cross-species drug toxicities using a Liver-Chip. *Sci. Transl. Med.* 11, eaax5516. doi:10.1126/scitranslmed.aax5516

Kerem, B., Rommens, J. M., Buchanan, J. A., Markiewicz, D., Cox, T. K., Chakravarti, A., et al. (1989). Identification of the cystic fibrosis gene: genetic analysis. *Science* 245, 1073–1080. doi:10.1126/science.2570460

Leenaars, C. H., De Vries, R. B., Heming, A., Visser, D., Holthaus, D., Reijmer, J., et al. (2020). Animal models for cystic fibrosis: a systematic search and mapping review of the literature - Part 1: genetic models. *Lab. Anim.* 54, 330–340. doi:10.1177/0023677219868502

Leenaars, C. H., Vries, R. B., Reijmer, J., Holthaus, D., Visser, D., Heming, A., et al. (2021). Animal models for cystic fibrosis: a systematic search and mapping review of the literature. Part 2: nongenetic models. *Lab. Anim.* 55, 23677221990688. doi:10.1177/0023677221990688

Matthews, L. W., Doershuk, C. F., Wise, M., Eddy, G., and Nudelman, H. (1964). A therapeutic regimen for patients with cystic fibrosis. *J. Pediatr.* 65, 558–575. doi:10.1016/s0022-3476(64)80290-0

National Academies of Sciences, Engineering, and Medicine (2020). *Necessity, use, and care of laboratory dogs at the U.S. Department of Veterans Affairs*. Washington DC: National Academies Press.

One Hundred and Seventeenth Congress of the United States (2023). Consolidated appropriations Act, 2023. Avaliable at: https://www.congress.gov/117/plaws/publ328/PLAW-117publ328.pdf.

Ratner, M. (2017). FDA deems *in vitro* data on mutations sufficient to expand cystic fibrosis drug label. *Nat. Biotechnol.* 35, 606. doi:10.1038/nbt0717-606

Rommens, J. M., Iannuzzi, M. C., Kerem, B.-S., Drumm, M. L., Melmer, G., Dean, M., et al. (1989). Identification of the CysticFibrosis gene: chromosome walking and jumping. *Science* 245, 1059–1065. doi:10.1126/science.2772657

Rumsey, J. W., Lorance, C., Jackson, M., Sasserath, T., Mcaleer, C. W., Long, C. J., et al. (2022). Classical complement pathway inhibition in a "human-on-A-chip" model of autoimmune demyelinating neuropathies. *Adv. Ther.* 5, 2200030. doi:10.1002/adtp.202200030

Savanta (2022). *Eurogroup for Animals – use of animals in research, testing and education*. EU.

Servidoni, M. F., Sousa, M., Vinagre, A. M., Cardoso, S. R., Ribeiro, M. A., Meirelles, L. R., et al. (2013). Rectal forceps biopsy procedure in cystic fibrosis: technical aspects and patients perspective for clinical trials feasibility. *BMC Gastroenterol.* 13, 91. doi:10.1186/1471-230X-13-91

Si, L., Bai, H., Rodas, M., Cao, W., Oh, C. Y., Jiang, A., et al. (2020). Human organs-on-chips as tools for repurposing approved drugs as potential influenza and COVID19 therapeutics in viral pandemics. BioRxv.

Singh, N., Vayer, P., Tanwar, S., Poyet, J.-L., Tsaioun, K., and Villoutreix, B. O. (2023). Drug discovery and development: introduction to the general public and patient groups. *Front. Drug Discov.* 3. doi:10.3389/fddsv.2023.1201419

Snouwaert, J. N., Brigman, K. K., Anne, M., Malouf, N. N., Boucher, R. C., Smithies, O., et al. (1992). An animal model for cystic fibrosis made by gene targeting. *Science* 257, 1083–1088. doi:10.1126/science.257.5073.1083

Song, Y., Sonawane, N. D., Salinas, D., Qian, L., Pedemonte, N., Galietta, L. J., et al. (2004). Evidence against the rescue of defective DeltaF508-CFTR cellular processing by curcumin in cell culture and mouse models. *J. Biol. Chem.* 279, 40629–40633. doi:10.1074/jbc.M407308200

Sun, D., Gao, W., Hu, H., and Zhou, S. (2022). Why 90% of clinical drug development fails and how to improve it? *Acta Pharm. Sin. B* 12, 3049–3062. doi:10.1016/j.apsb.2022.02.002

UK Parliament (2023). Question for department of health and social care. Avaliable at: https://questions-statements.parliament.uk/written-questions/detail/2023-09-19/200502.

U. S. Food and Drug Administration (2009). *Guidance on non-clinical safety studies for the Conduct of human clinical trials and marketing authorization for pharmaceuticals M3*. USA: FDA.

U.S. Government Accountability Office (2019). "Animal use in research: federal agencies should assess and report on their efforts to develop and promote alternatives," in *United States government accountability office*. USA: GAO.

Van Goor, F., Straley, K. S., Cao, D., Gonzalez, J., Hadida, S., Hazlewood, A., et al. (2006). Rescue of DeltaF508-CFTR trafficking and gating in human cystic fibrosis airway primary cultures by small molecules. *Am. J. Physiol. Lung Cell Mol. Physiol.* 290, L1117–L1130. doi:10.1152/ajplung.00169.2005

Van Norman, G. A. (2019). Limitations of animal studies for predicting toxicity in clinical trials. *JACC Basic Transl. Sci.* 4, 845–854. doi:10.1016/j.jacbts.2019.10.008

Van Norman, G. A. (2020). Limitations of animal studies for predicting toxicity in clinical trials: Part 2: potential alternatives to the use of animals in preclinical trials. *JACC Basic Transl. Sci.* 5, 387–397. doi:10.1016/j.jacbts.2020.03.010

frontiers | Frontiers in Drug Discovery

TYPE Mini Review
PUBLISHED 05 March 2024
DOI 10.3389/fddsv.2024.1339697

OPEN ACCESS

EDITED BY
Bruno Villoutreix,
Hôpital Robert Debré, France

REVIEWED BY
Taisiia Feoktistova,
Merck, United States
Juan Fernandez-Recio,
Spanish National Research Council (CSIC), Spain

*CORRESPONDENCE
Anne Poupon,
anne.poupon@mabsilico.com

†These authors have contributed equally to this work

RECEIVED 16 November 2023
ACCEPTED 13 February 2024
PUBLISHED 05 March 2024

CITATION
Musnier A, Dumet C, Mitra S, Verdier A, Keskes R, Chassine A, Jullian Y, Cortes M, Corde Y, Omahdi Z, Puard V, Bourquard T and Poupon A (2024), Applying artificial intelligence to accelerate and de-risk antibody discovery. *Front. Drug Discov.* 4:1339697. doi: 10.3389/fddsv.2024.1339697

Applying artificial intelligence to accelerate and de-risk antibody discovery

Astrid Musnier[1], Christophe Dumet[1], Saheli Mitra[1], Adrien Verdier[1], Raouf Keskes[1], Augustin Chassine[1], Yann Jullian[1], Mélanie Cortes[1], Yannick Corde[1], Zakaria Omahdi[1], Vincent Puard[1], Thomas Bourquard[1†] and Anne Poupon[1,2*†]

[1]MAbSilico, Tours, France, [2]Physiologie de la Reproduction et des Comportements, Centre National de la Recherche Scientifique, INRAe, Nouzilly, France

As in all sectors of science and industry, artificial intelligence (AI) is meant to have a high impact in the discovery of antibodies in the coming years. Antibody discovery was traditionally conducted through a succession of experimental steps: animal immunization, screening of relevant clones, *in vitro* testing, affinity maturation, *in vivo* testing in animal models, then different steps of humanization and maturation generating the candidate that will be tested in clinical trials. This scheme suffers from different flaws, rendering the whole process very risky, with an attrition rate over 95%. The rise of *in silico* methods, among which AI, has been gradually proven to reliably guide different experimental steps with more robust processes. They are now capable of covering the whole discovery process. Amongst the players in this new field, the company MAbSilico proposes an *in silico* pipeline allowing to design antibody sequences in a few days, already humanized and optimized for affinity and developability, considerably de-risking and accelerating the discovery process.

KEYWORDS

antibody, artificial intelligence, discovery, manufacturability, affinity

1 Introduction

Antibodies have long been irreplaceable tools for research. They have more recently emerged as powerful drugs, allowing considerably higher specificity than traditional chemicals, and offering new treatment options in a growing number of pathologies (Lu et al., 2020). Thanks to their half-lives, antibody drugs also have long-lasting effects as compared to small-molecules, rendering them more adapted to chronic pathologies. Many new formats derived from antibodies have been designed allowing to exploit their exquisite specificity (Vega et al., 2022). Antibody-drug conjugates can be used to bring chemical drugs to the precise location where their action is to take place, which is particularly useful for chemotherapies involving very toxic molecules (Jin et al., 2022). Using an antibody to bring the chemotherapy to the tumor allows to increase the doses, rendering cancer therapies more efficient, and decreasing the side-effects. T-cells expressing chimeric antigen receptors (CAR-T cells), recognize their target cells through antibody-like receptors (Mehrabadi et al., 2022), for example, binding to markers of cancer, and destroy them. Bispecifics recognize two different targets and can, for example, activate immune cells in the micro-tumoral environment (Bejarano et al., 2021).

As currently performed, antibody discovery starts by immunizing animals: the target is injected into an animal (mostly mice or rabbits), together with an immune booster. The immune system of the animal reacts by producing antibodies against this molecule. The second step is the screening, which consists in finding, amongst the antibodies of the animal, those binding to the target of interest. Successive rounds of selections, mainly based on refined binding assays (*in vitro* cross-species binding, paralogs binding) and both *in vitro* and *in vivo* functional assays, are applied to downsize the number of initial hit molecules and to identify the final "leads", resulting in the well-known funnel-shaped process of antibody discovery (Hoover et al., 2021). These successive elimination steps are highly empirical, and depend more on the scalability of wet-lab techniques than on the importance of the information provided. Epitope mapping is a very good example of that. Antibody/antigen complex is highly useful for further engineering, and mandatory for IP protection, but requires time-consuming and low throughput methods, like the gold-standards X-ray crystallography or NMR. As a result, epitope mapping is carried out very late in the cycle, as a check prior patenting, whereas it would have made much sense at the very beginning of the project, as a decision-making element (Bauer et al., 2023).

After these first selections, only a few leads actually display the suitable physico-chemical properties to be qualified as candidate molecules that could be moved to preclinical and clinical trials, and ultimately become therapeutics. Maturation steps are hence often engaged to optimize the affinity and the developability properties (low immunogenicity, solubility at high concentrations, manufacturability at large scale). Sequences are herein modified, meaning that the number of molecules to test is increased back, and that the final molecules are, strictly speaking, different from the originally characterized ones. A new round of validations aiming at requalifying the matured molecule is hence necessary, hoping that the biological properties are retained along the process.

Artificial intelligence methods are gradually replacing all these experimental steps, lowering the attrition rate and shortening the whole process. This technological transition happened a decade earlier for small chemical molecules, but the complexity of biologics prevented any transfer of technology from one area to the other and specific methods had to be designed. Here are described some of the AI-based innovations dedicated at antibody discovery.

2 Methods and datasets in AI-based antibody discovery methods

Artificial intelligence, theorized by Alan Turing in the 50 s, was born with the description of genetic algorithms by J.H. Holland in 1975 (Holland, 1992). However, computers were not powerful enough for these methods to be useful, and the real takeoff happened 15 years later with the publication of David Golberg *Genetic Algorithms in Search, Optimization, and Machine Learning* (Golberg, 1989). AI methods have considerably diversified and can be divided in two main categories: machine-learning and knowledge-based methods (Figure 1). Machine-learning methods are, by far, the most used, among which neural networks. There are again many categories within neural networks, the most popular being deep-learning. Once a model has been trained or learned, for example, using a neural network, it allows to either evaluate examples not present in the training stage, or even generate new ones (generative AI). Language models are another popular application of AI which bloomed after the arrival of the iconic transformer paper "Attention is all you need" in 2017 (Vaswani et al., 2017). The model, often a deep neural network, is fed with a corpus of texts, and it learns the meaning of word ensembles in a context. This type of model has been generalized to many types of objects (apart from texts), such as images, molecules, etc.

A very important aspect of machine-learning methods in general, is that they need to be trained on a set of data called *learning set*. The result of a machine-learning campaign certainly depends as much on the quality of this learning dataset than on the detailed implementation. Many databases related to antibodies have emerged these last years (Khetan et al., 2022), that can be used to train machine-learning methods. However, most of these databases have been themselves built using automated methods and are lacking one or the other essential pieces of information like affinity, aggregation parameters, or the epitope. One crucial piece of information is the pairing between heavy and light chains, which is missing in all the large databases. For this reason, we have developed our own database, which contains more than 80,000 well-characterized antibodies: heavy and light chain pairing, but also epitopes, affinities, *in vitro* and *in vivo* data, cross-species reactivity, etc. This database is accessible through a software platform, MAbFactory[1].

3 Automatizing the different steps of antibody discovery

3.1 Epitope mapping

The first area in which AI has been used in the context of antibody discovery has been epitope and paratope prediction, which consists in predicting the regions of each protein (the region on the antibody side is called paratope and the region on the antigen side is called epitope) involved in their interaction. Whereas initial trials at tackling this problem only allowed to predict linear epitopes (which represent only 10% of antibody epitopes (Rubinstein et al., 2008)), gradual introduction of more complex algorithms, such as docking and machine-learning trained scoring functions allowed to reach useful accuracy levels (Zeng et al., 2023), such as epitope3D (da Silva et al., 2022), RosettaDock (Lyskov and Gray, 2008) or MAbTope (Bourquard et al., 2018; Tahir et al., 2021). MAbTope is docking-based and uses a coarse-grained formalism, which requires only the antibody sequences and allows high-throughput epitope mapping. It allows identifying a correct epitope region in more than 80% of cases. This method has been successfully applied to many examples (Kizlik-Masson et al., 2017; Ashraf et al., 2019; Neiveyans et al., 2019; Granel et al., 2020; Trilleaud et al., 2021; Vayne et al., 2021; Ugamraj et al., 2022), including when the crystal structure of the target is unknown, and a 3D homology model has to be built.

1 https://app-publicdemo-mabfactory-97288.azurewebsites.net/

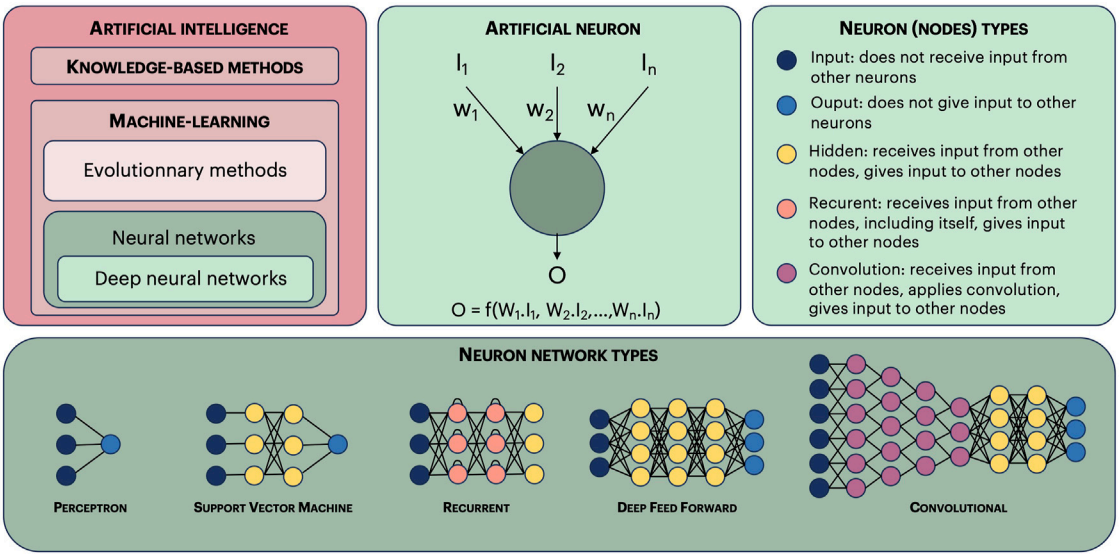


FIGURE 1
Artificial intelligence methods. Artificial intelligence methods can be divided in two main categories: knowledge-based methods and machine-learning methods. Machine-learning methods can be further divided in evolutionary methods and neural networks. This last category contains deep-learning methods. An artificial neuron (or node) receives input values (I_1, I_2, I_n), and computes an output value O, using a function and weights (w_1, w_2, w_n). Learning consists in optimizing these weights using input values for which the output value is known (learning set). The nodes are classified in five main categories. There are many types of neuron networks, we show here only the most common ones. Finally, a neuron network is qualified as "deep", allowing to make deep-learning, if it has three or more layers of hidden nodes.

3.2 Screening clones

Whether working from immune animal or from already established antibody banks, the first set of hits is mainly selected on the recombinant target using classical biology approaches based on hybridomas or display technologies, either in bacteriophages or yeasts (Köhler and Milstein, 1975; Clackson et al., 1991). High affinity is the main success criterion. This approach has three major limitations: i/many leads displaying sub-optimal affinity or less represented molecules are below the threshold of such approaches and are *de facto* excluded from the selection, and ii/ the epitope cannot be selected, meaning that selected hits binding to different places on the target molecule. Experimentally determining the epitopes of these hits, or at least knowing which ones are in competition (epitope binning) is far from trivial. The third limitation is due to the process used for transferring the animal immune repertoire to either bacteriophages or yeasts. Heavy and light chain pairing is not maintained, and the resulting antibodies are largely non-natural.

More recently, single B cells technologies have greatly improved the process of this initial clone selection (Pedrioli and Oxenius, 2021). Instead of building a bank from the immune repertoire, the B-cells of the animal, which each express a unique antibody in their membrane, can be directly selected on their affinity for the target using single-cell technologies. The antibodies coded by the retained B-cells can then be sequenced individually, resulting in natural-paired sequences. However, this technology is also relying on high affinity selection, and leads displaying sub-optimal affinity or less represented molecules are again eliminated. Moreover, even within a few thousand clones, experimentally characterization remains a problem.

Today, no published *in silico* method allows find leads against a selected target while fully exploring the sequential space of a natural repertoire, diverse both on the frameworks and CDRs (~10^9–10^{12} in diversity). State-of-the-art methods still require a seed antibody to guide the search. Deep-learning language models have had nice successes in finding novel and better leads in very large artificial library of CDR-degenerated parental antibodies, paving the way to future extension to antibody repertoires (Liu et al., 2020; Mason et al., 2021; Saka et al., 2021; Bachas et al., 2022). Examples are mentioned in the affinity maturation section.

3.3 Affinity evaluation and optimization

The first step in antibody characterization is often affinity evaluation, since the experimental technologies allow reasonably high-throughput as compared to other *in vitro* assays. Rough but large-scale evaluation is often performed in ELISA, while more precise but low throughput evaluation is performed in SPR to provide the ground-truth K_D. However, these technologies require the production of both antibody and antigen, limiting the number of clones that can be evaluated. Affinity prediction from the sequences and structures of antibody and antigen would therefore allow the evaluation of much larger ensembles. Many computational methods have been proposed for this task, and benchmarks collected, but the models still have limited efficacy (Guest et al., 2021). Moreover, many of these methods rely on the knowledge of the accurate structural assembly of antibody and target, which is generally not available, and certainly not for very large collections of antibodies.

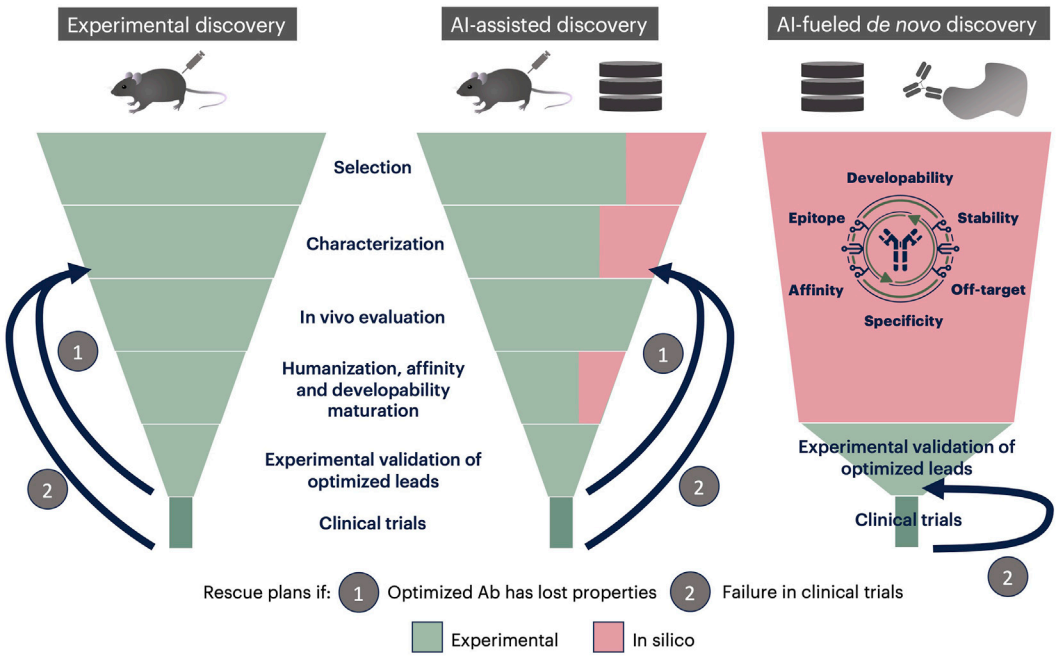


FIGURE 2
Classical, AI-assisted and AI-fueled *de novo* discovery pipelines. In the classical discovery pipeline (left), initial candidates (hundreds to thousands) are selected within the immune repertoire of an immunized animal, *in vitro* characterized (a few tens are retained) then *in vivo* evaluated, resulting in a few leads. These leads are humanized and optimized for affinity and developability. The resulting antibodies are evaluated *in vitro* and *in vivo* to verify that activity has been maintained. One of those is then selected for clinical trials. If antibodies have lost their activity during humanization and optimization (rescue plan 1), or if the chosen candidate fails in clinical trials (rescue plan 2), new candidates have to be selected within those characterized *in vitro*, and later steps gone through again. In the AI-assisted discovery pipeline (center), selection, characterization and optimization steps are partially conducted *in silico*, which accelerates the process, without changing its general organization. AI-based selection procedures allow to start from databases rather than physical antibody banks. In the AI-fueled *de novo* discovery pipeline, starting point is a database. Moreover, some technologies, such as those developed by MAbSilico, allow choosing the targeted epitope at the beginning. In this pipeline, all the antibodies of the initial database are fully evaluated in parallel, resulting in a few hundred well qualified, humanized and optimized antibodies. These candidates are then evaluated *in vitro* and *in vivo* to choose one lead for clinical trials. If this candidate fails, the next one can be chosen, without the need to repeat the whole process.

Antibodies obtained either through immunization or by screening existing antibody banks, often have insufficient affinities. Experimental methods to enhance affinity rely on random mutagenesis, usually restricted to the CDRs, and require intensive wet-lab labor. Deep-learning language models proved themselves successful at finding better binders than a parental antibody. Using the same principle as the experimental approach, language models start by building a library of the parental antibody which CDR residues are degenerated and substituted in all 20 or selected amino-acids. But the theoretical diversity to explore, even considering only the CDRs, remains largely beyond the interrogation by any wet-lab or computational means, and maturation methods are constrained to consider only a few mutated positions. As a matter of dimension, considering the CDRH3 is 10 aa-long on average, testing its full theoretical mutational space only raises the library to 10^{20}. Bachas and Mason (Bachas et al., 2022) used degenerated Trastuzumab libraries, cloned either in bacteriophages or hybridomas, and used their binding to a fluorescent HER2 (in FACS) to train models which allowed them to retrieve better binders than the parentals. They included up to 3 mutations on respectively 10 and 17 positions. Saka et al. (Saka et al., 2021) and Liu et al. (Liu et al., 2020) created degenerated libraries of an anti-kinurenin and an anti-VEGF-A (Rabinizumab), respectively, and trained a directed evolution-based model from the enriched sequences along panning rounds. The major limitation of such models, beside the restricted number of mutations, is that they are learnt on a given antibody-antigen pair, and that the resulting training set is not target-agnostic. The whole procedure is not applicable to the next target.

With the improvement of structure determination methods, rational design of mutants has significantly increased the success rate of affinity maturation (Li et al., 2023). Although rational design leads to testing a much lower number of mutants than random mutagenesis, it also requires to have precise structural data, which is in itself a difficult task. To tackle this problem, many computational methods allowing affinity prediction of mutants have appeared recently (Li et al., 2023) with various success rates. RosettaAntibodyDesign (Adolf-Bryfogle et al., 2018) is one of the most successful.

3.4 Off-targets prediction

One parameter often underestimated during antibody discovery is off-target binding. Indeed, if selectivity for the target is commonly verified by evaluating the absence of binding to close homologs, binding to unrelated proteins is usually not addressed, or very late in

the discovery process. Yet, there is now ample evidence that this phenomenon, called cross-reactivity, is far from anecdotal, as it can lead to auto-immune diseases (Cusick et al., 2012), and is most-probably also responsible for some failures in clinical trials (Lecerf et al., 2019; Cunningham et al., 2021; Loberg et al., 2021). However, cross-reactivity can also be an advantage, as in the case of rituximab. Indeed, rituximab not only binds its cognate target CD20, but also the sphingomyelin-phosphodiesterase-acid-like-3b (SMPDL-3b), and offers a treatment option for follicular segment glomerulosclerosis (FSGS). Some experimental methods exist to evaluate cross-reactivity, like tissue cross reactivity, or protein arrays, but are lengthy and expensive. MAbSilico has developed a computational method allowing to predict off-target binding with good accuracy (Musnier et al., 2022). In this method, both sequence and predicted 2D structure of antibodies are used to encode the CDRs of the antibodies. These encodings can then be compared using a specific score, based on the similarity of itemsets (Egho et al., 2015). This method allowed us predicting that 238D2 (Jähnichen et al., 2010), an anti-CXCR4 antibody, also binds hemagglutinin, and 6 human proteins. We were able to experimentally validate these predictions (Musnier et al., 2022). Using this method and our database of more than 80.000 antibodies having known targets, we are able to identify off-targets as soon as the sequences are known, and this does not require the knowledge of the antigen's 3D structure.

3.5 Developability prediction and optimization

The last step of antibody discovery is the evaluation of developability. The term *developability* generally covers different aspects: (1) immunogenicity: will this antibody elicit immune reaction when injected into human? (2) Producibility: will this antibody have high production yields in bioproduction? (3) Aggregation: will it be possible to make high concentration solution, or will the antibody aggregate? The methods and databases developed to date, are largely reviewed in (Khetan et al., 2022). Briefly, for example, prediction of immunogenicity is largely based on *humanness* scores, such as the OASis score (Prihoda et al., 2022). These scores evaluate how close the antibody of interest is to known human sequences, and are correlated with the levels of anti-drug antibodies (ADA) observed in clinical trials. Optimization of one antibody's immunogenicity starts with its humanization, which consists in modifying patterns to go back to the closest human germline. MAbSilico's CDR similarity measure (see above) allows to performed humanization. In fact, since it can identify the human antibody having the most similar CDRs, it can be considered that the frameworks of the retrieved human antibody constitute an optimal scaffold to support the CDRs. The CDRs of the animal antibody can then be grafted into the human frameworks, leading to a fully human candidate.

More general evaluation of developability can be obtained through the Therapeutic Antibody Profiler (TAP) tool (Raybould et al., 2019). This method allows to anticipate expression or aggregation issues of antibodies based on characteristics such as CDRH3 length, hydrophobicity within the CDRs or canonical forms. Gentiluomo et al. (Gentiluomo et al., 2019) use interpretable neural networks to successfully predict aggregation, together with melting temperature. Hou et al. (Hou et al., 2020) have developed the SOLart software, which uses both sequence and structure, and is based on a random-forest algorithm.

Producibility prediction seems to be an even more difficult challenge. Different studies show a correlation between the production titer and the stability of the antibody (Goldenzweig et al., 2016; Jain et al., 2017), especially the melting temperature and solubility. Harmalkar et al. (Harmalkar et al., 2023) use pre-trained language models and convolutional neural networks to predict melting temperature. Avoiding antibodies predicted to have low melting temperature or poor solubility is thus desirable, but is not a guarantee of good production titers.

4 Chaining them all: *de novo* antibody discovery

De novo antibody design holds the hope of being able to generate a highly affine, soluble, non-immunogenic, and epitope-directed antibody starting only from the name of the target. It implies mastering, at least, affinity prediction, structural characterization, and developability assessment. Solutions aiming at solving each pitfall are developed, as mentioned above, but they are still used individually along the funnel-shaped process dictated by the classical biological pipeline. Chaining them all together, in a virtuous circle, is certainly one key to success.

The whole design process must start by creating candidates, either randomly which would imply subsequent rational selection, or rationally, by "walking" on the target structure. Language-based approaches were expected to fulfill the first approach at high throughput but they are, as described above, still highly limited on the antibody diversity that can be injected in the computations (Liu et al., 2020; Mason et al., 2021; Saka et al., 2021; Lim et al., 2022). The approach proposed by Aguilar Rangel et al. (Aguilar Rangel et al., 2022) is based on a structural approach, computing CDR and epitope peptide complementarity. Authors show that the method can design *de novo* CDR peptides, which can then be grafted into nanobodies binding to three different targets (human serum albumin, SARS-CoV-2 spike protein, and trypsin), although with limited affinities. The method proposed by Anishchenko et al. (Anishchenko et al., 2021), which computes a structural evaluation of randomly generated and modeled peptides, proved accurate for protein design, but has not yet been applied to antibodies.

At MAbSilico, we have designed our own algorithms for the different steps, which allowed us to finally chain them all, unlocking our ability to *de novo* design antibodies. Our new *in silico* pipeline is target-agnostic and epitope-driven, and was successful at designing binders against the immune checkpoint inhibitor TIGIT (T-cell immunoreceptor with Ig and ITIM domains, unpublished) and against the Receptor-Binding Domain of SARS-CoV-2 (data presented at the Antibody Engineering and Therapeutics 2023; Amsterdam). In the latter project, thousands of paired VH/VL sequences were obtained from COVID-19 vaccinated patients, modeled and selected against chosen epitopes of the RBD. We identified 5 candidates, displaying nM and sub-nM affinities, and cross-neutralizing several viral strains (pre- and post- Omicron

lineage emergence). Our method was successfully scaled. In fact, starting from a collection of 4.25×10^{12} VH/VL pairs (artificially reconstituted from 1.7×10^{6} VH, and 2.5×10^{6} VL sequences obtained by NGS of a human scFv library), 16 VHs and 22 VLs were predicted as affine binders on a specified epitope of TIGIT. Amongst the 352 possible pairings, 94% were binding in an ELISA assay, and after developability optimization, the best binder had sub-nanomolar affinity in BLI. We were also able to *de novo* design binders against a GPCR, whose 3D structure has not yet been determined and for which we built several homology models. This demonstrates that our method does not require an experimental structure of the target.

5 Concluding remarks

In silico methods are being developed to replace or support antibody selection and their molecular characterization and optimization. As shown Figure 2, AI-based methods covering one step of the classical funnel-like discovery pipeline are undoubtedly useful, but they do not change the global shaping of discovery.

De novo AI-fueled methodologies, such as the one developed by MAbSilico allow to generate a few tens to a few hundred well-qualified leads, which are predicted to have high affinity, low off-target binding and good developability (Figure 2). These candidates can then be tested *in vitro* and *in vivo*, without the need to optimize or humanize them before clinical trials, which eliminates the risk of losing activity in the process. The chances of success are consequently much higher than in the classical process. Finally, the initial *in silico* step only takes up to 21 days, considerably shortening the process, and drastically abating the costs as the number of biological assays needed is decreased and the chances of success increased.

Among all characterization steps, the prediction of one antibody's biological function remains the least amenable to *in silico* prediction, as the molecular mechanisms involved are either not fully understood, or highly complex and target-specific. Targeting a precise epitope can partially circumvent this issue. For example, targeting the interaction region of a ligand on its receptor will in most cases inhibit the action of the ligand. However, antibodies having the same epitope can have different functions as illustrated by Zaitseva et al. (Zaitseva et al., 2023). These authors have generated different variants of an anti-Fn14 (fibroblast growth factor (FGF)-inducible 14) antibody, and show that, despite all binding the same epitope, they have different biological functions.

Author contributions

AM: Writing–original draft, Writing–review and editing, Conceptualization, Data curation, Investigation, Project administration, Validation. CD: Writing–review and editing, Conceptualization, Data curation, Formal Analysis, Investigation, Software, Visualization. SM: Writing–review and editing, Formal Analysis, Investigation, Methodology, Software. AV: Writing–review and editing, Formal Analysis, Investigation, Methodology, Software, Visualization. RK: Writing–review and editing, Formal Analysis, Investigation, Methodology, Software. AC: Writing–review and editing, Data curation, Resources, Software, Supervision, Visualization. YJ: Writing–review and editing, Formal Analysis, Investigation, Methodology, Software. MC: Writing–review and editing, Investigation, Validation. YC: Writing–review and editing, Investigation, Validation. ZO: Writing–review and editing, Investigation, Validation. VP: Writing–review and editing, Conceptualization, Project administration, Resources. TB: Supervision, Writing–review and editing, Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Software. AP: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Visualization, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

Authors AC, YJ, MC, YC, ZO, VP, TB, and AP were employed by MAbSilico.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adolf-Bryfogle, J., Kalyuzhniy, O., Kubitz, M., Weitzner, B. D., Hu, X., Adachi, Y., et al. (2018). RosettaAntibodyDesign (RAbD): a general framework for computational antibody design. *PLOS Comput. Biol.* 14, e1006112. doi:10.1371/journal.pcbi.1006112

Aguilar Rangel, M., Bedwell, A., Costanzi, E., Taylor, R. J., Russo, R., Bernardes, G. J. L., et al. (2022). Fragment-based computational design of antibodies targeting structured epitopes. *Sci. Adv.* 8, eabp9540. doi:10.1126/sciadv.abp9540

Anishchenko, I., Pellock, S. J., Chidyausiku, T. M., Ramelot, T. A., Ovchinnikov, S., Hao, J., et al. (2021). *De novo* protein design by deep network hallucination. *Nature* 600, 547–552. doi:10.1038/s41586-021-04184-w

Ashraf, Y., Mansouri, H., Laurent-Matha, V., Alcaraz, L. B., Roger, P., Guiu, S., et al. (2019). Immunotherapy of triple-negative breast cancer with cathepsin D-targeting antibodies. *J. Immunother. cancer* 7, 29. doi:10.1186/s40425-019-0498-z

Bachas, S., Rakocevic, G., Spencer, D., Sastry, A. V., Haile, R., Sutton, J. M., et al. (2022). Antibody optimization enabled by artificial intelligence predictions of binding affinity and naturalness. arXiv. doi:10.1101/2022.08.16.504181

Bauer, J., Rajagopal, N., Gupta, P., Gupta, P., Nixon, A. E., and Kumar, S. (2023). How can we discover developable antibody-based biotherapeutics? *Front. Mol. Biosci.* 10, 1221626. doi:10.3389/fmolb.2023.1221626

Bejarano, L., Jordão, M. J. C., and Joyce, J. A. (2021). Therapeutic targeting of the tumor microenvironment. *Cancer Discov.* 11, 933–959. doi:10.1158/2159-8290.CD-20-1808

Bourquard, T., Musnier, A., Puard, V., Tahir, S., Ayoub, M. A., Jullian, Y., et al. (2018). MAbTope: a method for improved epitope mapping. *J. Immunol.* 201, 3096–3105. doi:10.4049/jimmunol.1701722

Clackson, T., Hoogenboom, H. R., Griffiths, A. D., and Winter, G. (1991). Making antibody fragments using phage display libraries. *Nature* 352, 624–628. doi:10.1038/352624a0

Cunningham, O., Scott, M., Zhou, Z. S., and Finlay, W. J. J. (2021). Polyreactivity and polyspecificity in therapeutic antibody development: risk factors for failure in preclinical and clinical development campaigns. *mAbs* 13, 1999195. doi:10.1080/19420862.2021.1999195

Cusick, M. F., Libbey, J. E., and Fujinami, R. S. (2012). Molecular mimicry as a mechanism of autoimmune disease. *Clin. Rev. Allergy Immunol.* 42, 102–111. doi:10.1007/s12016-011-8294-7

da Silva, B. M., Myung, Y., Ascher, D. B., and Pires, D. E. V. (2022). epitope3D: a machine learning method for conformational B-cell epitope prediction. *Briefings Bioinforma.* 23, bbab423. doi:10.1093/bib/bbab423

Egho, E., Raïssi, C., Calders, T., Jay, N., and Napoli, A. (2015). On measuring similarity for sequences of itemsets. *Data Min. Knowl. Disc* 29, 732–764. doi:10.1007/s10618-014-0362-1

Gentiluomo, L., Roessner, D., Augustijn, D., Svilenov, H., Kulakova, A., Mahapatra, S., et al. (2019). Application of interpretable artificial neural networks to early monoclonal antibodies development. *Eur. J. Pharm. Biopharm.* 141, 81–89. doi:10.1016/j.ejpb.2019.05.017

Golberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Reading, MA, USA: Addison Welssey Publishing company.

Goldenzweig, A., Goldsmith, M., Hill, S. E., Gertman, O., Laurino, P., Ashani, Y., et al. (2016). Automated structure- and sequence-based design of proteins for high bacterial expression and stability. *Mol. Cell* 63, 337–346. doi:10.1016/j.molcel.2016.06.012

Granel, J., Lemoine, R., Morello, E., Gallais, Y., Mariot, J., Drapeau, M., et al. (2020). 4C3 human monoclonal antibody: a proof of concept for non-pathogenic proteinase 3 anti-neutrophil cytoplasmic antibodies in granulomatosis with polyangiitis. *Front. Immunol.* 11, 573040. doi:10.3389/fimmu.2020.573040

Guest, J. D., Vreven, T., Zhou, J., Moal, I., Jeliazkov, J. R., Gray, J. J., et al. (2021). An expanded benchmark for antibody-antigen docking and affinity prediction reveals insights into antibody recognition determinants. *Structure* 29, 606–621.e5. doi:10.1016/j.str.2021.01.005

Harmalkar, A., Rao, R., Richard Xie, Y., Honer, J., Deisting, W., Anlahr, J., et al. (2023). Toward generalizable prediction of antibody thermostability using machine learning on sequence and structure features. *mAbs* 15, 2163584. doi:10.1080/19420862.2022.2163584

Holland, J. H. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. Cambridge, MA: MIT press. Available at: https://books.google.fr/books?hl=fr&lr=&id=5EgGaBkwvWcC&oi=fnd&pg=PR7&dq=J.+H.+Holland,+Adaptation+In+Natural+And+Artificial+Systems&ots=mJoq2YKqxq&sig=fAjjZ6AOubm3-FNuSNT_Tf-CHfU (Accessed October 25, 2023).

Hoover, J. M., Prinslow, E. G., Teigler, J. E., Truppo, M. D., and La Porte, S. L. (2021). Therapeutic antibody discovery. *Remingt. Twenty-third Ed.* 2021, 417–436. doi:10.1016/B978-0-12-820007-0.00023-4

Hou, Q., Kwasigroch, J. M., Rooman, M., and Pucci, F. (2020). SOLart: a structure-based method to predict protein solubility and aggregation. *Bioinformatics* 36, 1445–1452. doi:10.1093/bioinformatics/btz773

Jähnichen, S., Blanchetot, C., Maussang, D., Gonzalez-Pajuelo, M., Chow, K. Y., Bosch, L., et al. (2010). CXCR4 nanobodies (VHH-based single variable domains) potently inhibit chemotaxis and HIV-1 replication and mobilize stem cells. *Proc. Natl. Acad. Sci.* 107, 20565–20570. doi:10.1073/pnas.1012865107

Jain, T., Sun, T., Durand, S., Hall, A., Houston, N. R., Nett, J. H., et al. (2017). Biophysical properties of the clinical-stage antibody landscape. *Proc. Natl. Acad. Sci.* 114, 944–949. doi:10.1073/pnas.1616408114

Jin, S., Sun, Y., Liang, X., Gu, X., Ning, J., Xu, Y., et al. (2022). Emerging new therapeutic antibody derivatives for cancer treatment. *Sig Transduct. Target Ther.* 7, 39–28. doi:10.1038/s41392-021-00868-x

Khetan, R., Curtis, R., Deane, C. M., Hadsund, J. T., Kar, U., Krawczyk, K., et al. (2022). Current advances in biopharmaceutical informatics: guidelines, impact and challenges in the computational developability assessment of antibody therapeutics. *mAbs* 14, 2020082. doi:10.1080/19420862.2021.2020082

Kizlik-Masson, C., Vayne, C., McKenzie, S. E., Poupon, A., Zhou, Y., Champier, G., et al. (2017). 5B9, a monoclonal antiplatelet factor 4/heparin IgG with a human Fc fragment that mimics heparin-induced thrombocytopenia antibodies. *J. Thromb. Haemost.* 15, 2065–2075. doi:10.1111/jth.13786

Köhler, G., and Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 256, 495–497. doi:10.1038/256495a0

Lecerf, M., Kanyavuz, A., Lacroix-Desmazes, S., and Dimitrov, J. D. (2019). Sequence features of variable region determining physicochemical properties and polyreactivity of therapeutic antibodies. *Mol. Immunol.* 112, 338–346. doi:10.1016/j.molimm.2019.06.012

Li, J., Kang, G., Wang, J., Yuan, H., Wu, Y., Meng, S., et al. (2023). Affinity maturation of antibody fragments: a review encompassing the development from random approaches to computational rational optimization. *Int. J. Biol. Macromol.* 247, 125733. doi:10.1016/j.ijbiomac.2023.125733

Lim, Y. W., Adler, A. S., and Johnson, D. S. (2022). Predicting antibody binders and generating synthetic antibodies using deep learning. *mAbs* 14, 2069075. doi:10.1080/19420862.2022.2069075

Liu, G., Zeng, H., Mueller, J., Carter, B., Wang, Z., Schilz, J., et al. (2020). Antibody complementarity determining region design using high-capacity machine learning. *Bioinformatics* 36, 2126–2133. doi:10.1093/bioinformatics/btz895

Loberg, L. I., Chhaya, M., Ibraghimov, A., Tarcsa, E., Striebinger, A., Popp, A., et al. (2021). Off-target binding of an anti-amyloid beta monoclonal antibody to platelet factor 4 causes acute and chronic toxicity in cynomolgus monkeys. *mAbs* 13, 1887628. doi:10.1080/19420862.2021.1887628

Lu, R.-M., Hwang, Y.-C., Liu, I.-J., Lee, C.-C., Tsai, H.-Z., Li, H.-J., et al. (2020). Development of therapeutic antibodies for the treatment of diseases. *J. Biomed. Sci.* 27, 1. doi:10.1186/s12929-019-0592-z

Lyskov, S., and Gray, J. J. (2008). The RosettaDock server for local protein–protein docking. *Nucleic Acids Res.* 36, W233–W238. doi:10.1093/nar/gkn216

Mason, D. M., Friedensohn, S., Weber, C. R., Jordi, C., Wagner, B., Meng, S. M., et al. (2021). Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning. *Nat. Biomed. Eng.* 5, 600–612. doi:10.1038/s41551-021-00699-9

Mehrabadi, A. Z., Ranjbar, R., Farzanehpour, M., Shahriary, A., Dorostkar, R., Hamidinejad, M. A., et al. (2022). Therapeutic potential of CAR T cell in malignancies: a scoping review. *Biomed. Pharmacother.* 146, 112512. doi:10.1016/j.biopha.2021.112512

Musnier, A., Bourquard, T., Vallet, A., Mathias, L., Bruneau, G., Ayoub, M. A., et al. (2022). A New *in silico* antibody similarity measure both identifies large sets of epitope binders with distinct CDRs and accurately predicts off-target reactivity. *Int. J. Mol. Sci.* 23, 9765. doi:10.3390/ijms23179765

Neiveyans, M., Melhem, R., Arnoult, C., Bourquard, T., Jarlier, M., Busson, M., et al. (2019). A recycling anti-transferrin receptor-1 monoclonal antibody as an efficient therapy for erythroleukemia through target up-regulation and antibody-dependent cytotoxic effector functions. *mAbs* 11, 593–605. doi:10.1080/19420862.2018.1564510

Pedrioli, A., and Oxenius, A. (2021). Single B cell technologies for monoclonal antibody discovery. *Trends Immunol.* 42, 1143–1158. doi:10.1016/j.it.2021.10.008

Prihoda, D., Maamary, J., Waight, A., Juan, V., Fayadat-Dilman, L., Svozil, D., et al. (2022). BioPhi: a platform for antibody design, humanization, and humanness evaluation based on natural antibody repertoires and deep learning. *mAbs* 14, 2020203. doi:10.1080/19420862.2021.2020203

Raybould, M. I. J., Marks, C., Krawczyk, K., Taddese, B., Nowak, J., Lewis, A. P., et al. (2019). Five computational developability guidelines for therapeutic antibody profiling. *Proc. Natl. Acad. Sci.* 116, 4025–4030. doi:10.1073/pnas.1810576116

Rubinstein, N. D., Mayrose, I., Halperin, D., Yekutieli, D., Gershoni, J. M., and Pupko, T. (2008). Computational characterization of B-cell epitopes. *Mol. Immunol.* 45, 3477–3489. doi:10.1016/j.molimm.2007.10.016

Saka, K., Kakuzaki, T., Metsugi, S., Kashiwagi, D., Yoshida, K., Wada, M., et al. (2021). Antibody design using LSTM based deep generative model from phage display library for affinity maturation. *Sci. Rep.* 11, 5852. doi:10.1038/s41598-021-85274-7

Tahir, S., Bourquard, T., Musnier, A., Jullian, Y., Corde, Y., Omahdi, Z., et al. (2021). Accurate determination of epitope for antibodies with unknown 3D structures. *mAbs* 13, 1961349. doi:10.1080/19420862.2021.1961349

Trilleaud, C., Gauttier, V., Biteau, K., Girault, I., Belarif, L., Mary, C., et al. (2021). Agonist anti-ChemR23 mAb reduces tissue neutrophil accumulation and triggers chronic inflammation resolution. *Sci. Adv.* 7, eabd1453. doi:10.1126/sciadv.abd1453

Ugamraj, H. S., Dang, K., Ouisse, L.-H., Buelow, B., Chini, E. N., Castello, G., et al. (2022). TNB-738, a biparatopic antibody, boosts intracellular NAD+ by inhibiting CD38 ecto-enzyme activity. *mAbs* 14, 2095949. doi:10.1080/19420862.2022.2095949

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). "Attention is all you need," in *Advances in neural information processing systems* (Red Hook, NY: Curran Associates, Inc.). Available at: https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html (Accessed November 16, 2023).

Vayne, C., Nguyen, T.-H., Rollin, J., Charuel, N., Poupon, A., Pouplard, C., et al. (2021). Characterization of new monoclonal PF4-specific antibodies as useful tools for studies on typical and autoimmune heparin-induced thrombocytopenia. *Thromb. Haemost.* 121, 322–331. doi:10.1055/s-0040-1717078

Vega, F. A., Berraondo, P., and Galluzzi, L. (2022). *New antibody formats*. Cambridge, MA: Academic Press. Available at: https://books.google.fr/books?hl=fr&lr=&id=Er5mEAAAQBAJ&oi=fnd&pg=PP1&dq=new+antibody+formats+review&ots=a2FNJEpNaQ&sig=XVPDRssdpQUnv3IlkV7jtase8NQ (Accessed January 9, 2024).

Zaitseva, O., Hoffmann, A., Löst, M., Anany, M. A., Zhang, T., Kucka, K., et al. (2023). Antibody-based soluble and membrane-bound TWEAK mimicking agonists with FcγR-independent activity. *Front. Immunol.* 14, 1194610. doi:10.3389/fimmu.2023.1194610

Zeng, X., Bai, G., Sun, C., and Ma, B. (2023). Recent progress in antibody epitope prediction. *Antibodies* 12, 52. doi:10.3390/antib12030052

frontiers | Frontiers in Drug Discovery

TYPE Mini Review
PUBLISHED 04 April 2024
DOI 10.3389/fddsv.2024.1347798

OPEN ACCESS

EDITED BY
Bruno Villoutreix,
Hôpital Robert Debré, France

REVIEWED BY
Julia M. L. Menon,
Netherlands Heart Institute, Netherlands

*CORRESPONDENCE
Catharine E. Krebs,
ckrebs@pcrm.org

RECEIVED 01 December 2023
ACCEPTED 13 March 2024
PUBLISHED 04 April 2024

CITATION
Krebs CE and Herrmann K (2024), Confronting the bias towards animal experimentation (animal methods bias).
Front. Drug Discov. 4:1347798.
doi: 10.3389/fddsv.2024.1347798

Confronting the bias towards animal experimentation (animal methods bias)

Catharine E. Krebs[1]* and Kathrin Herrmann[2,3]

[1]Physicians Committee for Responsible Medicine, Washington, DC, United States, [2]Johns Hopkins Bloomberg School of Public Health, Center for Alternatives to Animal Testing, Baltimore, MD, United States, [3]Senate Department for Justice and Consumer Protection, Berlin, Germany

Laws and policies are in place around the world to promote the replacement and reduction of nonhuman animals in science. These principles are rooted not just in ethical considerations for animals, but also in scientific considerations regarding the limitations of using nonhuman animals to model human biology, health, and disease. New nonanimal research approaches that use human biology, cells, and data to mimic complex human physiological states and therapeutic responses have become increasingly effective and accessible, replacing the use of animals in several applications, and becoming a crucial tool for biomedical research and drug development. Despite many advantages, acceptance of these new nonanimal methods has been slow, and barriers to their broader uptake remain. One such barrier is animal methods bias, the preference for animal-based methods where they are not necessary or where animal-free methods are suitable. This bias can impact research assessments and can discourage researchers from using novel nonanimal approaches. This article provides an introductory overview of animal methods bias for the general public, reviewing evidence, exploring consequences, and discussing ongoing mitigation efforts aimed at reducing barriers in the shift away from animal use in biomedical research and testing.

KEYWORDS

animal methods bias, peer review, alternatives to animal experiments, scientific publishing, biomedical research, drug development

Introduction: animal and human-based preclinical research methods

Animal experiments are frequently performed for basic research (with the aim to gain knowledge without specific applications) and for applied research (applying knowledge, for example, to try to find new drugs for humans and to test for their toxicity or safety). According to the directive on the use of animals in science in the European Union, animal experiments must be replaced whenever possible, and EU Member States should make a substantial effort to reduce and replace animal use in science (European Parliament, 2010). A similar principle is applied in other regions, including the United States: the 3Rs principle to replace, reduce, and refine animal use in science (Russell and Burch, 1959; Office of Laboratory Animal Welfare, 2015). These principles are rooted not just in ethical considerations for animals, but also in scientific considerations regarding the limitations of using nonhuman animals to model human biology, health, and disease.

Animal tests are often expensive, take a long time to conduct, and can give misleading results (Meigs et al., 2018). Approximately 92% of drugs in

development fail to pass human clinical trials—mostly due to failures during safety and efficacy testing—despite safe and effective findings demonstrated in preclinical tests (Thomas et al., 2021). A recent economic analysis estimated that the use of more predictive preclinical nonanimal technologies instead of animal tests could save over $24 billion (Ewart et al., 2022). There is increasing recognition among government, academic, and industry scientists that nonanimal research methods have the potential to overcome some of the scientific limitations of animal-based methods (Baran et al., 2022; Gribaldo and Dura, 2022; Ingber, 2022; Advisory Committee to the Director Working Group on Catalyzing the Development and Use of Novel Alternative Methods to Advance Biomedical Research, 2023).

In recent years, *in vitro* (in a dish) and *in silico* (computational) research models have become increasingly effective and accessible, replacing the use of animals in several applications, and becoming a crucial tool for preclinical research. These promising new models use human biology, cells, and data to mimic complex human physiological states and therapeutic responses (Shaker et al., 2021; Loewa et al., 2023). Examples of innovative, human-biology based *in vitro* models are organoids, organs-on-a-chip (also called organ chips and tissue chips), and induced pluripotent stem cells. Organoids are three-dimensional cell aggregates (also called spheroids) consisting of multiple cell types and designed to imitate physiological processes. Human organ-on-a-chip systems are microdevices composed of three-dimensional cells and fluids to simulate physiological processes in human organs. Induced pluripotent stem cells are adult human-derived cells that have been genetically reprogrammed to a stem cell-like state and then further engineered to one of a variety of cell types that can be found throughout the body.

Animals have been used in research for so long that the scientific community has been slow to accept novel nonanimal methods. Some of these new methods have high entry costs and can be difficult for researchers to operationalize in their laboratories, highlighting the need for more funding and expanded infrastructure (Busek et al., 2022). Acceptance of new nonanimal methods can also be improved with good laboratory practices to ensure that high-quality experiments are performed and that findings can be reproduced by other researchers, and with thorough model evaluation to confirm that experiments are suitable for their intended use (Pamies et al., 2022; van der Zalm et al., 2022).

Other barriers to the broader use of nonanimal methods are more psychological, though, such as a bias or preference for animal-based methods. This *animal methods bias* may be especially prevalent when research that uses nonanimal methods is being assessed during subjective evaluations of studies for publication or proposals for grant funding. By impacting publications and funding awards, animal methods bias can be a barrier to the sharing and uptake of novel nonanimal approaches, standing in the way of improved preclinical predictiveness and further complicating drug development. The concept of *animal methods bias* is further expanded on below, including an overview of current evidence, how it impacts research assessments, and ongoing efforts to mitigate its harmful effects on human health research.

Animal methods bias: the bias toward animal experimentation in research and publishing

Publishing plays a crucial role in the advancement of science, helping to translate research findings into medical interventions. It also impacts researchers' careers, playing a role in hiring decisions and other evaluations. The publishing process is not without biases, though. According to the Catalogue of Bias, a database of psychological, methodological, and reporting biases created by The Center for Evidence-Based Medicine at Oxford University, publication bias is defined as "when the likelihood of a study being published is affected by the findings of the study" (DeVito and Goldacre, 2019). But what if the likelihood of a study being published is affected by the *methods* of the study, namely, animal or animal-free methods?

In his article, *Is it Time for Reviewer three to Request Human Organ Chip Experiments Instead of Animal Validation Studies?*, Dr. Donald Ingber questioned why animal data is still considered the gold standard in human health research, while presenting evidence that organ chips may better suit this purpose (Ingber, 2020). He framed this issue as a problem with peer review, describing an increasingly common anecdote about reviewer requests for animal experiments even though the author(s) explained why they did not use animals for their experiments.

Animal methods bias in publishing is thus a newly defined type of publishing bias, describing a preference for animal-based methods where they are not necessary or where nonanimal methods are suitable, which affects the likelihood of a manuscript being accepted for publication or introduces a significant delay to manuscript acceptance. Animal methods bias affects other aspects of research too, including the review of grant applications, when researchers apply for funds to enable their animal-free projects but are held back by biased assessments of their proposals. It can be likened to another kind of bias called *scholarly bias*, the favoring of perspectives, theories, or methods that align with one's own (Langfeldt et al., 2023).

To further understand when and why this occurs, which in turn informs solutions, the first author of this article and colleagues conducted a small survey to assess the experiences and perceptions of authors and reviewers related to animal- and human biology-based experiments during peer review (Krebs et al., 2023b). Respondents represented a broad range of biomedical research and related fields, primarily worked in academic (74%) and industry (10%) sectors, and in the United States (32%).

Twenty-one of the 68 total respondents indicated that they have performed animal-based experiments for the sole purpose of anticipating reviewer requests for them. In other words, they did not think the experiments were necessary outside the context of review. Thirty-one of the 68 total respondents indicated that they have been asked by reviewers to add animal experiments to a study that otherwise has no animal-based experiments. Among those 31 respondents, just three indicated that they felt the request was justified, while 14 respondents felt that it was sometimes justified, and 11 did not think the request was justified (three respondents did not provide an answer to this question).

When asked to elaborate on their perceptions of these requests, respondents expressed that reviewers ask for animal

experiments out of habit, not because it is necessary or relevant. Some respondents also indicated that more prominent journals are more likely to request or expect animal experiments, which acts as an incentive for conducting animal experiments or as a punishment for researchers who use animal-free, human biology-based methods. Overall, the survey identified the following consequences of animal methods bias during manuscript peer review: the conduct of animal experiments which would have otherwise not been performed, as well as negative career repercussions, including delays in publication, rejection, or withdrawal of papers, and being forced to publish in less-prominent journals.

The survey also asked questions about respondents' experiences as reviewers, and specifically regarding reasons for making requests for additional animal experiments. Respondents indicated that their preference for animal methods or their lack of awareness of appropriate animal-free methods were reasons for making requests for additional animal experiments.

Because of the pressures to publish, researchers may feel compelled to comply with reviewer requests for animal experiments even when they disagree with their necessity. Alternatively, failing to comply with such requests may result in negative career consequences. Altogether, animal methods bias affects how nonanimal research is published and may even discourage researchers from using these methods. In other words, it is a barrier to the uptake and dissemination of nonanimal research—important research that holds promise for improving preclinical predictiveness and rates of translation from drug discovery to clinical trial approvals.

How to mitigate animal methods bias

An April 2022 workshop to address animal methods bias in scientific publishing was convened among academic and industry researchers, journal editors, government representatives, and advocates in order to: (1) explore a range of stakeholder perspectives, (2) describe the current state of animal- and nonanimal-based experimental systems, (3) describe animal methods bias in publishing and related biases in publishing and peer review, and (4) identify potential causes, consequences, and potential mitigation strategies for animal methods bias in publishing (Krebs et al., 2022).

Barriers to addressing animal methods bias were identified, including:

- The high-pressure nature of the research environment,
- Impact factor, an index measuring the impact of scholarly literature that represents the annual average number of citations to articles published in each journal over the past 2 years,
- Financial stakes,
- Animals as the "gold standard," seen as the default method by the research community,
- Institutional inertia and psychological lock-in (see Gluck, 2019), and
- Lack of knowledge or desire to learn about animal-free methods.

Workshop attendees also identified recommendations for addressing animal methods bias geared toward the scientific community, journals and publishers, and funders, governments, and policymakers. Recommendations included the following:

- Build awareness about animal methods bias among editors, peer reviewers, and the scientific community more broadly, especially early-career researchers;
- Increase authors' confidence in their ability to challenge reviewers' requests for animal experiments, such as through the Author Guide for Addressing Animal Methods Bias (Krebs et al., 2023a);
- Provide educational materials for reviewers, as recently acknowledged by the US National Institutes of Health to ensure the better evaluation of nonanimal research (Advisory Committee to the Director Working Group on Catalyzing the Development and Use of Novel Alternative Methods to Advance Biomedical Research, 2023);
- Mandate that requests for addition of animal methods be scrutinized by other reviewers; and
- Prioritize funding for animal-free, human biology-based methods, including to improve accessibility and training for researchers.

After the workshop, attendees formed the Coalition to Illuminate and Address Animal Methods Bias (COLAAB) to continue to explore and address this issue.[1] The COLAAB is currently gathering additional evidence of animal methods bias and its consequences and developing and implementing tools for overcoming it.

Conclusion

New nonanimal methods hold great promise for advancing biomedical research and drug development. Although a lot of work remains to improve the acceptance of nonanimal methods within the scientific community, researchers are increasingly turning to them to answer their research questions. Researchers should be able to do so without unfair requests or expectations for animal experiments from reviewers who prefer their own methodologies or are ill-equipped to evaluate novel ones.

Animal methods bias is a serious issue that adds additional and unnecessary work for researchers who use animal-free approaches, and it perpetuates the idea that animal-free approaches are not sufficiently valuable on their own. Animal methods bias is a symptom of a research ecosystem that rewards animal use and disincentivizes a shift toward potentially more reliable human-biology based research methods, and is therefore a barrier to changing the *status quo* from its reliance on animals.

To advance biomedical research and get safer and better drugs to more patients, researchers, drug developers, and funding agencies must address animal methods bias. Measures that empower researchers to confront unfair requests for animal experiments, prevent reviewers from making such requests, and advance the standardization, evaluation, and infrastructure for nonanimal research approaches will all be important. The public can play a

1 www.animalmethodsbias.org

role too. Consumers and taxpayers have power in the market and with publicly funded research, and they are already helping to turn the tide by demanding cruelty-free cosmetics and supporting lawmakers' shifts toward animal-free research and testing approaches.

Author contributions

CK: Conceptualization, Project administration, Writing–original draft, Writing–review and editing. KH: Conceptualization, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research is supported in part by Lush Handmade Cosmetics Ltd. (Advancing Animal Free Research Models, CK). The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Acknowledgments

The authors would like to thank Reina Pohl, John MacInnis, and Elizabeth Baker for their thoughtful review of this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Advisory Committee to the Director Working Group on Catalyzing the Development and Use of Novel Alternative Methods to Advance Biomedical Research (2023). Catalyzing the development and use of novel alternative methods. Available at: https://www.acd.od.nih.gov/documents/presentations/Working_Group_Report.pdf.

Baran, S. W., Brown, P. C., Baudy, A. R., Fitzpatrick, S. C., Frantz, C., Fullerton, A., et al. (2022). Perspectives on the evaluation and adoption of complex *in vitro* models in drug development: workshop with the FDA and the pharmaceutical industry (IQ MPS Affiliate). *ALTEX - Altern. animal Exp.* 39, 297–314. doi:10.14573/altex.2112203

Busek, M., Aizenshtadt, A., Amirola-Martinez, M., Delon, L., and Krauss, S. (2022). Academic user view: organ-on-a-chip technology. *Biosensors* 12, 126. doi:10.3390/bios12020126

DeVito, N. J., and Goldacre, B. (2019). Catalogue of bias: publication bias. *BMJ Evidence-Based Med.* 24, 53–54. doi:10.1136/bmjebm-2018-111107

European Parliament (2010). Directive 2010/63/EU. Available at: http://data.europa.eu/eli/dir/2010/63/oj/eng (Accessed October 19, 2022).

Ewart, L., Apostolou, A., Briggs, S. A., Carman, C. V., Chaff, J. T., Heng, A. R., et al. (2022). Performance assessment and economic analysis of a human Liver-Chip for predictive toxicology. *Commun. Med.* 2, 154–216. doi:10.1038/s43856-022-00209-1

Gluck, J. P. (2019). "Afterword: evidence over interests," in *Animal experimentation working towards a paradigm change*. Editors K. Herrmann and K. Jayne (Leiden, Netherlands: Brill), 689–691. Available at: https://www.jstor.org/stable/10.1163/j.ctvjhzq0f.36 (Accessed November 28, 2023).

Gribaldo, L., and Dura, A. (2022). EURL ECVAM literature review series on advanced non-animal models for respiratory diseases, breast cancer and neurodegenerative disorders. *Anim. (Basel)* 12, 2180. doi:10.3390/ani12172180

Ingber, D. E. (2020). Is it time for reviewer 3 to request human organ chip experiments instead of animal validation studies? *Adv. Sci.* 7, 2002030. doi:10.1002/advs.202002030

Ingber, D. E. (2022). Human organs-on-chips for disease modelling, drug development and personalized medicine. *Nat. Rev. Genet.* 23, 467–491. doi:10.1038/s41576-022-00466-9

Krebs, C. E., Camp, C., Constantino, H., Courtot, L., Kavanagh, O., Leite, S. B., et al. (2022). Proceedings of a workshop to address animal methods bias in scientific publishing. *ALTEX* 40, 677–688. doi:10.14573/altex.2210211

Krebs, C. E., Camp, C., Constantino, H., Courtot, L., Kavanagh, O., McCarthy, J., et al. (2023a). Author Guide for addressing animal methods bias in publishing. *Adv. Sci. n/a* 10, 2303226. doi:10.1002/advs.202303226

Krebs, C. E., Lam, A., McCarthy, J., Constantino, H., and Sullivan, K. (2023b). A survey to assess animal methods bias in scientific publishing. *ALTEX - Altern. animal Exp.* 40, 665–676. doi:10.14573/altex.2210212

Langfeldt, L., Reymert, I., and Svartefoss, S. M. (2023). Distrust in grant peer review—reasons and remedies. *Sci. Public Policy* 51, 28–41. scad051. doi:10.1093/scipol/scad051

Loewa, A., Feng, J. J., and Hedtrich, S. (2023). Human disease models in drug development. *Nat. Rev. Bioeng.* 1, 545–559. doi:10.1038/s44222-023-00063-3

Meigs, L., Smirnova, L., Rovida, C., Leist, M., and Hartung, T. (2018). Animal testing and its alternatives - the most important omics is economics. *ALTEX* 35, 275–305. doi:10.14573/altex.1807041

Office of Laboratory Animal Welfare, (2015). PHS policy on humane care and use of laboratory animals. Available at: https://olaw.nih.gov/policies-laws/phs-policy.htm (Accessed February 7, 2022).

Pamies, D., Leist, M., Coecke, S., Bowe, G., Allen, D. G., Gstraunthaler, G., et al. (2022). Guidance document on good cell and tissue culture practice 2.0 (GCCP 2.0). *ALTEX - Altern. animal Exp.* 39, 30–70. doi:10.14573/altex.2111011

Russell, W. M. S., and Burch, R. L. (1959). *The principles of humane experimental technique*. London: Methuen. Available at: http://books.google.com/books?id=j75qAAAAMAAJ (Accessed December 16, 2022).

Shaker, B., Ahmad, S., Lee, J., Jung, C., and Na, D. (2021). *In silico* methods and tools for drug discovery. *Comput. Biol. Med.* 137, 104851. doi:10.1016/j.compbiomed.2021.104851

Thomas, D., Chancellor, D., Micklus, A., LaFever, S., Hay, M., Chaudhuri, S., et al. (2021). Clinical development success rates and contributing factors 2011-2020. Available at: https://go.bio.org/rs/490-EHZ-999/images/ClinicalDevelopmentSuccessRates2011_2020.pdf.

van der Zalm, A. J., Barroso, J., Browne, P., Casey, W., Gordon, J., Henry, T. R., et al. (2022). A framework for establishing scientific confidence in new approach methodologies. *Arch. Toxicol.* 96, 2865–2879. doi:10.1007/s00204-022-03365-4

frontiers | Frontiers in Drug Discovery

TYPE Review
PUBLISHED 08 April 2024
DOI 10.3389/fddsv.2024.1355044

OPEN ACCESS

EDITED BY
Bruno Villoutreix,
Hôpital Robert Debré, France

REVIEWED BY
Olivier Taboureau,
Université Paris Cité, France
Lindsay Marshall,
Humane Society of the United States,
United States

*CORRESPONDENCE
Thomas Hartung,
thartung@jhsph.edu

RECEIVED 13 December 2023
ACCEPTED 27 February 2024
PUBLISHED 08 April 2024

CITATION
Hartung T (2024), The (misleading) role of animal models in drug development.
Front. Drug Discov. 4:1355044.
doi: 10.3389/fddsv.2024.1355044

The (misleading) role of animal models in drug development

Thomas Hartung[1,2]*

[1]Center for Alternatives to Animal Testing (CAAT), Bloomberg School of Public Health and Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, United States, [2]CAAT-Europe, University of Konstanz, Konstanz, Germany

Animals like mice and rats have long been used in medical research to help understand disease and test potential new treatments before human trials. However, while animal studies have contributed to important advances, too much reliance on animal models can also mislead drug development. This article explains for a general audience how animal research is used to develop new medicines, its benefits and limitations, and how more accurate and humane techniques—alternatives to animal testing—could improve this process.

KEYWORDS

animal models, drug development, preclinical research, clinical trials, predictive methods, alternatives to animal testing

Abbreviations: Artificial intelligence (AI), computer programs (machine learning tools), which perform tasks, which typically require human intelligence; Attrition refers to the high rate of failure that drug candidates experience during clinical development. **Biased outcome reporting**, it is easier to publish an effect than no effect: this is a classic example of bias in the scientific literature; **Blockbuster**, a blockbuster drug is a pharmaceutical product that generates annual sales of $1 billion or more for the company that sells it; **Drug target**, it is essentially a molecule within the body that a drug interacts with to produce its therapeutic effect; **European cosmetics test ban**, EU legislation from 2003 bans animal testing, enforced by marketing bans, for finished cosmetic products, for products including ingredients tested on animals where alternatives were accepted (after 2004), for acute and topical (eye and skin) test (after 2009), and all other hazards (after 2013); **Generic drugs**, a generic drug is a medication that is created to be the same as an already marketed brand-name drug in dosage form, safety, strength, route of administration, quality, performance characteristics, and intended use; **Hazard**, adverse effect of a substance; **High-throughput screening (HTS)**, a method used in scientific discovery, particularly in drug discovery, biology, materials science, and chemistry. It involves the use of automated equipment to rapidly test thousands to millions of samples for biological activity or chemical reactions; **Immunosuppressant drugs**, drugs used to hinder transplant rejections or autoimmune diseases; **Lead optimization**, the process of iteratively synthesizing and testing chemical variants of initial hit compounds from screens to improve potency, selectivity, and drug-like properties, and develop optimized lead molecules as strong final candidates for clinical development; **LD_{50}**, or lethal dose for 50%, a way to compare the toxic potential of substances through the dose at which 50% of rats die; **Nanoparticles**, defined as particles of matter with dimensions ranging from 1 to 100 nm (nm) in diameter; they can exhibit significantly different physical and chemical properties compared to their bulk material counterparts due to their high surface area to volume ratio; **Omics technologies**, simultaneous measurement of as many active genes (transcriptomics), proteins (proteomics), or metabolites (metabolomics) changes as possible; **REACH program**, acronym for Registration, Evaluation, Authorisation, and Restriction of Chemicals, a comprehensive regulation of the European Union designed to ensure a high level of protection for human health and the environment from the risks posed by chemicals. It was enacted on 1 June 2007; **Reproducibility crisis**, also known as the replication crisis, refers to the growing concern that many scientific studies' results are difficult or impossible to reproduce; **Selective analysis**, aka subgroup analysis, focuses on part of the data, neglecting the overall results, to obtain significant results. This is a common source for irreproducible results; **Teratogenic effects**, causing birth defects.

Introduction

Developing new medications is long and challenging. Before a drug can be sold, it must proceed through preclinical studies in cells and animals and usually three phases of human clinical trials: healthy volunteers, a small group of patients to assess patient safety, who may differ greatly from healthy volunteers, and then a large patient trial to prove the beneficial effect. This helps ensure the drug is reasonably safe and effective for its intended use. Animal research in the preclinical phase and in some safety studies continuing in parallel to the clinical studies provides useful but imperfect information about how drugs will behave in people. However, overreliance on animal models results—as I will explain—in many clinical trial failures and unsafe drugs reaching patients. Nevertheless, animals remain necessary until better techniques are available and broadly accepted. This article summarizes for a general audience how animals are used in drug development, their limitations in predicting human responses, and how more accurate human-cell-based and computer models could improve this process.

Historically, from the 1920s to the 1970s, animal experiments were the predominant technology in life sciences. Figure 1 shows how the use of laboratory animals peaked in the 1970s, largely for drug development. Other methods have now begun to complement and even replace animal testing, despite its continued high regard in scientific and regulatory circles. Ethical concerns were the primary drivers for questioning the use of animal experiments; the debate over the justification of animal suffering for scientific advancement varies, but public opinion is increasingly critical. In response, the scientific community has implemented measures to make animal experiments more rigorous, requiring formal justifications, permissions, and adherence to rising standards of animal welfare. Concurrently, there has been significant support for developing alternatives.

Recent challenges to animal experiments extend beyond ethics. They are resource-intensive, costly, time-consuming, and have limited predictivity for humans—issues highlighted by the European REACH program's struggle to test thousands of industrial chemicals and by the pharmaceutical industry's crisis of low success to the market. The latter refers to the extremely high failure rate in clinical trials for drug candidates due to issues like lack of efficacy or safety problems in human testing. These problems have sparked a broader discussion on the "reproducibility crisis" in science.

The drive to find alternatives to traditional animal testing—notably in toxicology, which uses about 10% of all experimental animals (according to European statistics)—has led to significant work in this area. Reasons why most work into alternatives takes place in toxicology include government funding, legislative acts like the European cosmetics test ban and REACH chemical legislation, and the relative stability of internationally standardized guideline tests.

A simplified view of the drug development process

Despite all biomedical progress, we are far from understanding the complex networked systems of the human organism and, even farther, their perturbation in disease. Intervening in these disease mechanisms as a remedy involves much trial and error. Increasingly, identifying a certain mechanism of disease or a possible target for a drug can change the odds of finding something that ultimately works. Such so-called pharmacological "targets" can be, for example, a misbehaving cell type or a receptor protein on cells in an organ that positively influences the course of disease or ameliorates a certain symptom. These observations (on the cell types or receptors) may often occur in animal "models" of disease, and this species difference compounds the difficulty with translating observations from the laboratory bench to the clinic. It is still an enormous undertaking to develop a therapy from this and bring a successful drug to the market.

On average, drug development takes 12 years and costs $2.4 billion. This 12-year timeframe, also called time-to-market, has been quite stable over time. The main reason is that a patent's lifetime is only 20 years; when it expires, competitors can offer the

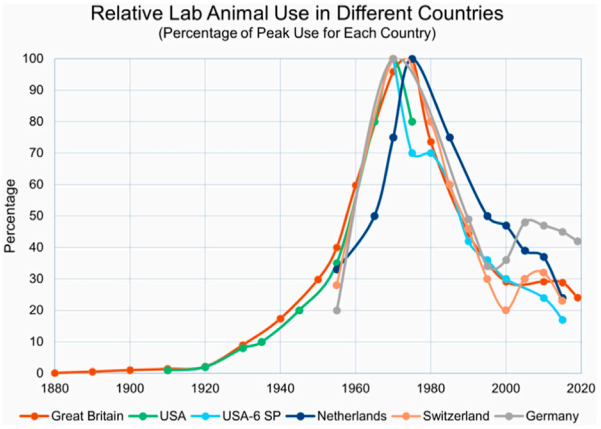


FIGURE 1
This figure, kindly provided by Dr. Andrew Rowan, shows that that all animal use peaked in the 1970s in most industrialized countries. Taken from Wellbeing International (WBI) newsletter 31 December 2021 with permission (https://wellbeingintl.org).

FIGURE 2
Visualizing the analogy of a gold rush to describe the drug discovery process using DALL-E 3.

FIGURE 3
Visualization that the "golden pill" is actually a rare find among many rocks, done with DALL-E 3.

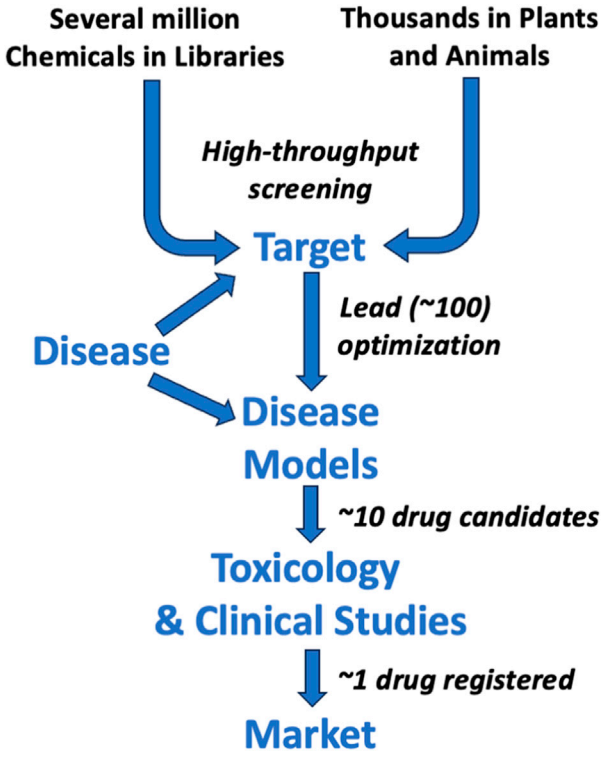


FIGURE 4
Simplified drug development process.

same drug and so prices plummet. Longer development times therefore eat into the time in which the company must at least recoup the money it has invested. If we simplify that a company has 8 years to recoup $2.4 billion, then every additional day is worth approximately $1 million. However, we must also factor in the many abandoned drug projects which never lead to marketed product. Forbes estimated that, already in 2012, about $4–11 billion was spent by the industry for a single market release. This highlights the importance of each step in the decision tree with respect to time and forgone revenue.

However, some drugs do make tens of billions of dollars per year. This creates a "goldrush" situation (Figure 2). There is in fact some similarity between drug development and a goldrush: it takes many for a few to find something—the abundance of pills on the AI-generated image in Figure 2 is actually misleading. As in a real goldrush, few get rich, and those who sell the sieves and shovels are the ones getting rich. However, "gold washing" often describes the process well, where many stones must be washed to find a rare "golden" pill (Figure 3).

Indeed, as summarized in Figure 4, the drug development process requires, as a rule of thumb, about 10,000 chemicals to enter preclinical experiments to ultimately produce one marketed drug. In recent decades, many companies start with even more molecules (sometimes several million in what is called a "chemical library") to identify some promising structures through robotized testing—so-called high-throughput screening. As no animals are used in this step, it does not change the argument of this article. However, this has also not dramatically changed development times and success rates. Some companies start the search with biological materials such as plants. They often contain several tens of thousands of molecules in these "biological libraries" with the later problem of finding out which in this mix has the desired effect or just being stuck with a "phytopharmaceutical"—essentially, a plant extract. Although many customers like such products, it often requires difficult controls, such as the following: When to harvest? How to process to maximize effects? Are all sources equivalent? How stable is the product? Are there other components in the mix which have negative effects? Such matters will not be discussed here further as the challenge is to prove safety and efficacy, and thus, the role of animal studies and their alternatives is not much different.

To stay with the goldrush metaphor, companies typically hope to make the really big find, not just "a few nuggets from the river." The dream is gold mines, not gold washing (Figure 5): companies hope for the big wins, the "blockbuster" drug or technology which brings in big money. A "blockbuster" is typically a drug which sells more than $1 billion per year. This can mean finding new important targets (how to change the course of disease), new drug entities such as genetic drugs or nano-particles in more recent years, higher throughput in drug development by faster methods (for example the current discussion around AI-generated drugs—the novel tools that employ artificial intelligence to accelerate drug discovery), and anything promising to lower "attrition," the so-called failure rate in clinical trials leading to less side effects, earlier detection, or higher efficacy of the resulting medicines. The attrition rate is really the magic number for drug companies. A 2012 study by Arrowsmith et al. showed that 95% of drug candidates failed in the clinical development stage. This means somewhere between $0.9 billion for preclinical development and $2.6 billion investment for full clinical development (using the DiMasi data again), and 19 of 20 drug development projects being abandoned. This is even lower than the rule of thumb that only 1 in 10 substances entering the clinical phase

FIGURE 5
Visualization of the difference between gold washing and gold mining as a metaphor for drug development breakthroughs, done with DALL-E 3.

FIGURE 6
Many animals are used in drug development, staying in the metaphor of a gold rush, done with DALL-E 3.

will make it to the market (Figure 4). Some 20%–40% fail because of side effects, or toxicities. Even when a drug makes it to the market, about 8% are later withdrawn, usually because of unacceptably severe or even life-threatening side effects. It has been calculated that 1 in 100 patients in hospital for any reason dies from adverse drug reactions, often from interactions between drugs that patients receive at the same time. The safety of drugs thus continues to be a concern after marketing commences. Typically, a so-called phase-IV trial monitors drugs entering the market to review their safety and efficacy under real-life conditions, and possible drug side effects are also recorded by physicians to build a knowledge base to find rare problems.

As in any gold rush, there are many animals involved (Figure 6). However, the first question for this article is how much do animals really help? They are costly, take a long time, and have limited reproducibility and predictivity for humans.

How and why are animal models used in drug development?

Animals like mice, rats, dogs, and monkeys share much biology with humans, enabling several types of preclinical studies (Box 1). About a century ago, small rodents in particular became a primary research tool in biomedicine, with a supply industry emerging. Until the 1970s, they were the almost exclusive tool for finding new drugs (Figure 2), often in the absence of any idea how they might work. Then, most animals were used in drug development; today, according to European figures, drug development is only responsible for about 20% of all animal use (plus about 5% for drug safety testing and 5% for vaccine batch control); this is an overall drastic reduction of all animal use to about 40% of 1970s numbers. And while drugs required most of animal use in 1970s, it is now about 30%. It should be noted that a culture of systematically testing candidate drugs only emerged after scandals in the 1930s. In the USA, the 1937 sulfanilamide scandal (Figure 7) killed more than 100 people (mostly children), leading to legislation that empowered the Food and Drug Administration (FDA). This disaster was pivotal in the history of drug regulation in the United States. Sulfanilamide was used to treat streptococcal infections and was effectively formulated as a tablet and powder. However, in an attempt to create a liquid formulation, the Massengill Company, a pharmaceutical manufacturer, dissolved sulfanilamide in diethylene glycol (DEG), an untested solvent. DEG is poisonous to humans, but this was not well-known at the time. The company did not conduct any safety tests on the new formulation, which was marketed as “Elixir Sulfanilamide.” The product was distributed widely and resulted in over 100 deaths, many of which were of children, due to kidney failure caused by the DEG.

Box 1. Types of biomedical studies in drug development

***In vivo*:** Studies in live animals

***In vitro*:** Cells, tissues, or embryos studied outside a living organism

Microphysiological systems: Bioengineered *in vitro* systems, which recreate aspects of organ architecture and functionality, often with perfusion as vasculature equivalent forming (multi-) organ-on-chip systems

***Ex vivo*:** Analysis of organs, tissues, or biofluids from treated animals

***In silico*:** Computational models, increasingly based on artificial intelligence (AI)

Toxicology (safety): Testing for toxic and adverse effects

Efficacy: Assessing potential treatment benefits

Pharmacokinetics: Absorption, distribution, metabolism, and excretion of the drug

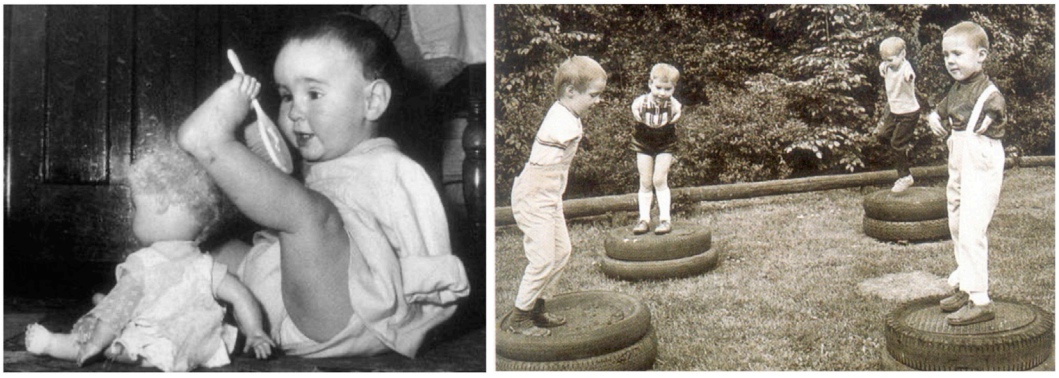


FIGURE 7
Malformations caused by thalidomide, archive of the author.

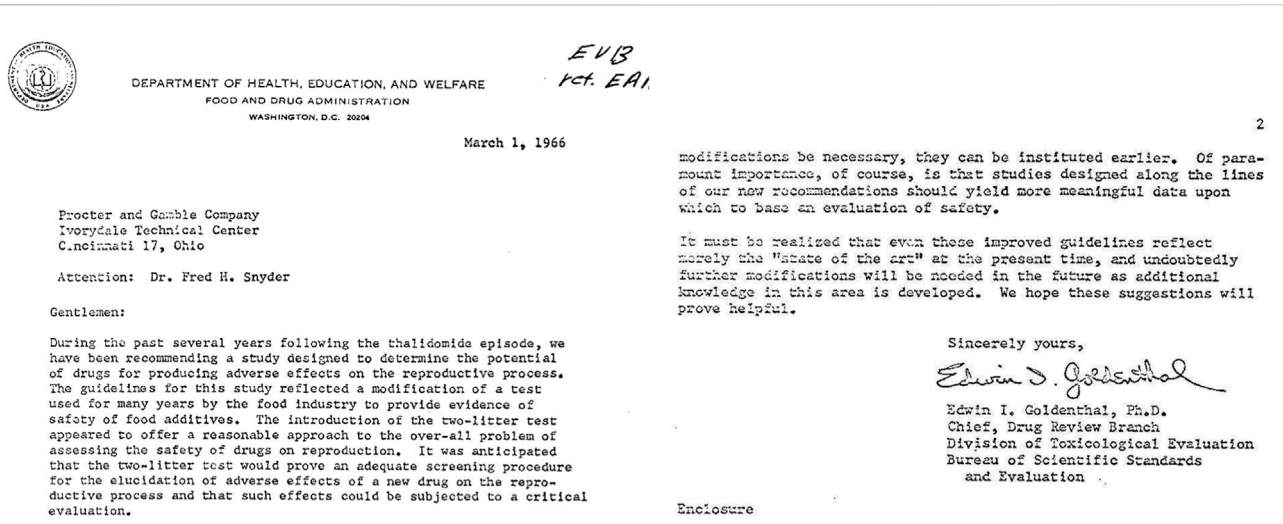
This tragedy highlighted the lack of regulation in the drug industry, particularly regarding the safety and testing of new drug formulations. At the time, the FDA had little power to regulate pharmaceuticals. The Federal *Food, Drug, and Cosmetic Act* of 1938 was passed in direct response to this incident and significantly increased the FDA's authority. The Act mandated that new drugs must be proven safe before being marketed, laying the groundwork for modern drug approval processes. This incident is often cited as a turning point in pharmaceutical regulation, demonstrating the critical need for rigorous drug testing and approval processes to ensure public safety.

The safety testing toolbox expanded continuously with problems as a patch for the future. A prominent example was the thalidomide (Contergan^{R}) scandal in the late 1950s and early 1960s, one of the most notorious medical disasters in history. Thalidomide, marketed under the brand name Contergan among others, was introduced as a sedative and later used widely to alleviate morning sickness in pregnant women.

However, thalidomide was not adequately tested for its effects during pregnancy. It was soon discovered that the drug caused severe birth defects in thousands of children (Figures 8, 9) that primarily affected limb development but also caused damage to the ears, eyes, heart, and nervous system. The drug was available in many countries, including Germany, the United Kingdom, and Australia, but was not approved in the United States. The tragedy led to a massive global overhaul of drug testing and regulatory processes. The extent of the birth defects caused by thalidomide brought to light the need for rigorous drug testing, especially for teratogenic effects (the potential to cause fetal abnormalities). In response, many countries strengthened their drug regulation laws and the processes for drug approval, making them more stringent and emphasizing the need for comprehensive clinical trials, including assessing effects on pregnancy. Since then, testing on animals has provided initial safety and efficacy data not ethically possible from humans. However, due to biological differences, small study sizes, and lack of diversity, animal research has important limitations. The thalidomide scandal remains a critical example in medical and regulatory circles of the importance of thorough drug testing and the potential consequences of inadequate drug regulation.

FIGURE 8
Title page of German weekly journal *Der Spiegel* from 5 December 1962 titled "Sick by remedies", archive of the author.

George Daston from Proctor and Gamble shared with me a letter from the FDA to their company in 1966 (Figure 10), when the "patch" for the reproductive effects of substances was created. It shows quite nicely how the increasing number of toxicity concerns led to an enlarged toolbox of safety tests. Notably, the letter ends "It must be realized that even these improved guidelines reflect merely the 'state of the art' at the present time, and undoubtedly further modifications will be needed in the future as additional knowledge in this area is developed." In fact, the very demanding animal study done on rats and rabbits did not even reliably detect the teratogenic effects (causing birth defects) of thalidomide. Several factors contribute to this discrepancy:



DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
FOOD AND DRUG ADMINISTRATION
WASHINGTON, D.C. 20204

EVB
Ref. EAI

March 1, 1966

Procter and Gamble Company
Ivorydale Technical Center
Cincinnati 17, Ohio

Attention: Dr. Fred H. Snyder

Gentlemen:

During the past several years following the thalidomide episode, we have been recommending a study designed to determine the potential of drugs for producing adverse effects on the reproductive process. The guidelines for this study reflected a modification of a test used for many years by the food industry to provide evidence of safety of food additives. The introduction of the two-litter test appeared to offer a reasonable approach to the over-all problem of assessing the safety of drugs on reproduction. It was anticipated that the two-litter test would prove an adequate screening procedure for the elucidation of adverse effects of a new drug on the reproductive process and that such effects could be subjected to a critical evaluation.

2

modifications be necessary, they can be instituted earlier. Of paramount importance, of course, is that studies designed along the lines of our new recommendations should yield more meaningful data upon which to base an evaluation of safety.

It must be realized that even these improved guidelines reflect merely the "state of the art" at the present time, and undoubtedly further modifications will be needed in the future as additional knowledge in this area is developed. We hope these suggestions will prove helpful.

Sincerely yours,

Edwin I. Goldenthal, Ph.D.
Chief, Drug Review Branch
Division of Toxicological Evaluation
Bureau of Scientific Standards
and Evaluation

Enclosure

FIGURE 9
Excerpt from 1966 letter by the FDA to Procter & Gamble on introduction of the two-generation study for reproductive toxicity, following the thalidomide scandal (courtesy of Dr. George Daston, P&G).

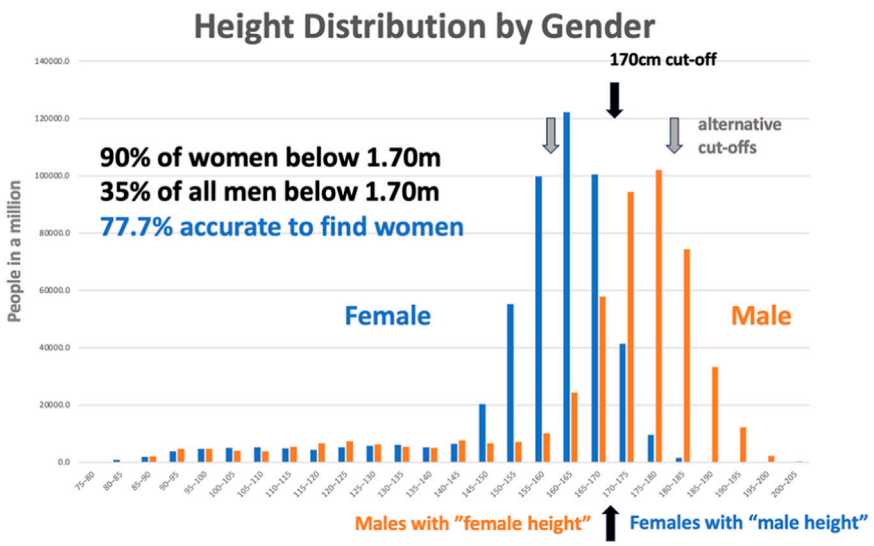


FIGURE 10
Height distribution by gender, This dataset was created by an unknown author for the blog "Why Sex Differences Don't Always Measure Up" available at https://sugarandslugs.wordpress.com/2011/02/13/sex-differences/(last accessed 12 December 2023).

- Species-specific differences in drug metabolism and sensitivity: different species metabolize drugs differently. Thalidomide's teratogenic effects are highly species-specific. It is known to cause birth defects in certain strains of mice, primates, and specific breeds of rabbit, but not in others, including the standard laboratory strains of rats and rabbits. This highlights a crucial limitation of animal studies in predicting human outcomes due to interspecies variability.
- Mechanism of action: thalidomide's teratogenic mechanism is complex and until recently not fully understood. It involves multiple pathways and is influenced by genetic and environmental factors that may not be present or that may differ significantly in animal models compared to humans.
- Dosage and exposure timing: the manifestation of thalidomide's effects is highly dependent upon the timing of exposure during pregnancy and the dosage. These factors can vary greatly between humans and animals, affecting the outcome and reliability of animal studies.
- Lack of early detection methods: when thalidomide was introduced, the methodologies for detecting teratogenic effects, especially subtle ones or those manifesting later in development, were not as advanced as today. This limited the ability to detect such effects in animal studies.

The thalidomide tragedy fundamentally changed the way drugs are tested for safety, underscoring the need for more predictive and

human-relevant models in teratology. It led to stricter regulatory requirements for drug testing, including the need for more comprehensive animal testing and the development of alternative methods to better predict human outcomes. However, the case of thalidomide remains a classic example of the limitations of animal models in accurately predicting human drug responses, especially in the context of developmental and reproductive toxicity. It is also a showcase of how, over almost 60 years, the "quick fix" of 1966 has not been replaced. The shortcomings mentioned when introducing the test have been forgotten, and it is now a standard that is difficult to replace. "We just got used to it", in the words of Petr Skrabanek and James McCormick in their wonderful book *Follies and Fallacies in Medicine* (Tarragon Press, Glasgow, 1989): "*Learning from experience may be nothing more than learning to make the same mistakes with increasing confidence*."

On the other hand, aspirin, a widely used medication, presents a notable case where animal studies (would have) made findings that are not entirely relevant or predictive of its effects in humans. In my 2009 article "*Per aspirin ad astra*," I critically examined the implications of traditional animal testing methods, underscoring the paradox of aspirin's toxicological profile—its widespread acceptance was fortunate due to the lack of stringent regulatory toxicology in 1899. In animal models, aspirin has demonstrated a range of toxic effects that are not typically observed in humans or that are observed under different conditions. These discrepancies highlight the limitations of extrapolating data from animal studies to human physiology and medicine. Aspirin when ingested is classified as harmful, with an LD_{50}, or lethal dose of 50%, to the rats used in testing, ranging from 150 to 200 mg/kg for the rodents, which is exactly the maximum daily dose used in humans. This is not a 100–1,000-fold safety factor usually suggested by toxicologists to indicate acute toxicity. Aspirin irritates the eyes, respiratory system, and skin. Although it is not directly carcinogenic, it acts as a co-carcinogen, meaning that it can promote cancer in the presence of other carcinogenic agents. Its mutagenic potential remains unclear, suggesting uncertainty about its ability to cause genetic mutations. Studies in various animal models, including cats, dogs, rats, mice, rabbits, and monkeys, have shown that it causes embryonic malformations—but not in humans, where one study analyzed 90,000 pregnancies. Due to this extensive profile of harmful effects, it is likely that such a substance would face significant challenges in the drug approval process today, making it unlikely that it would be brought to the market. In a 2009 article, I looked critically at traditional animal testing methods using the example of aspirin. I highlighted the paradox that aspirin is widely accepted and used despite results from animal tests that might have blocked its initial approval under today's strict rules.

Animal studies show that aspirin can have a range of toxic effects not typically seen in humans, or only at very high doses rarely used in patients. For example, tests suggest that aspirin is quite toxic based on lethal dose experiments in rats using the same maximum daily levels given to people. Animal studies also indicate that it may irritate eyes and airways and possibly act as a co-carcinogen—promoting cancer development alongside other chemicals. Its effects on potential gene mutations also remain unclear. Meanwhile, additional animal research implies that aspirin might cause birth defects, which over 90,000 human pregnancies that have been tracked disproved.

Due to this concerning toxicology profile from animal tests, aspirin likely would have faced major obstacles getting initially approval if today's stringent safety regulations existed back in 1899. The conflicting results between laboratory animals and human patients highlight limitations in using animal studies alone to predict safety in people. Findings in animals do not always match up with outcomes when drugs are actually given to diverse groups of people. Therefore, while useful, data from animal models have major shortcomings that impact the progress of drugs from early laboratories to patient bedsides.

Remarkably, until the 1970s, there were no efficacy data, meaning that convincing evidence that drugs are promising for curing a disease were formally required, just that they are not likely to cause harm.

Developing targeted therapeutics: the role of animal studies

Drugs today are developed to act through a defined target—a structure or component of the body to be altered by the treatment. Such targeted therapeutics are designed to specifically effect molecules associated with disease, unlike traditional chemotherapies, for example, that can also damage healthy cells. This increased specificity aims to improve treatment effectiveness and reduce side effects. However, developing a targeted therapeutic is a long, expensive, and risky process, taking on average about 12 years. Extensive testing in animals plays a crucial role in this process—typically 10-20,000 animals per drug development today. The Nuffield Council on Bioethics has estimated that 5%–15% are used to identify targets for drug action and possible medicines, 60%–80% for lead identification and optimization—choosing the optimal candidate substance—and 10%–20% for selecting candidate medicines going into clinical trials. Notably, according to European statistics, the pharmaceutical industry uses about 20% of all laboratory animals for drug development, down from about 30% in 2005, despite increasing research spending indicating that the industry is transitioning to other methods. The continuing need for animals is because cell cultures and computer models cannot replicate the full complex biology of a living organism. The traditional view is that animal testing provides invaluable data about real-world efficacy and safety that often cannot be obtained by other means. Preclinical testing with a combination of animals and, increasingly, other tools enables researchers to select the most promising candidates to move forward into clinical trials. This minimizes risks to human participants and increases the chance of success in later-stage clinical testing. Although targeted therapeutics provide exciting possibilities for treating disease, developing them often requires extensive animal research. Preclinical testing in appropriate animal models is still an essential part of bringing safe, effective targeted therapies to the clinic. The high degree of similarity between many animal species and humans leads many researchers to believe that this enables key data to be collected for guiding therapeutic development and improving human health. This rather optimistic view of the role

of animal studies in drug development is slowly being eroded, given the perceived inefficacy of the process of drug development with its many failures in the clinical (attrition) phase and increasing cases where the limitations of animal testing have been apparent. Importantly, the use of animals is also prompted by the expectations of regulators of receiving such data for decision-making and the fear of the industry that not meeting these expectations will result in delay or even refusal of registration.

Once safety and efficacy are demonstrated in animals, the most promising targeted therapeutic candidates advance to testing in humans. Clinical trials are performed sequentially in healthy volunteers and patients with disease to definitively determine overall benefit and risk. Animal research provides the foundation of knowledge necessary to justify testing new drugs in people.

Limitations of animal models

Although the historic cases of thalidomide and aspirin shed some light on how the safety testing of drugs was introduced and was flawed from start, this section will address limitations more systematically and with more recent examples. Although somewhat useful, animal models frequently fail to predict human clinical trial outcomes. Reliance on inadequate animal data results in the following:

- Many false negative errors: potentially good drugs are abandoned due to lack of efficacy or side effects in animals that would not occur in human trials (which never happened because of the animal findings).
- False positives: drugs that "work" in animals may still fail in human trials.
- Adverse events and safety issues in human volunteers and patients that were missed by prior animal testing.
- Several factors limit the accuracy of animal models, including biological differences; inbred strains vs. genetic diversity in humans; often young, healthy animals, unlike aged, sick humans; molecular differences altering drug effects; artificial experimental conditions; housing, diet, and environments that differ from human lifestyles.
- Disease that is induced artificially may differ from naturally occurring illness.
- Study design: small, short studies vs. lifelong human exposures; high doses triggering irrelevant effects; each test uses limited animal groups unlike large, diverse human trials.
- Animal research retirement is not yet feasible but should be reduced. Imperfect animal models need to be supplemented with more reliable human-based techniques such as: miniature bioengineered "organs-on-chips"; advanced computer models of human disease; big data mining of patient health records and genetic databases; small, carefully designed human clinical studies.

Used intelligently in combination, old and new methods can transform drug development to reliably predict safety and benefits for patients. Scientists have an obligation to use the most predictive tools available to efficiently develop effective medicines.

Animal testing has been an entrenched part of drug development for decades. However, there are numerous concerning examples where animal tests have misled clinical development due to inherent physiological differences between species, leading to dangerous outcomes in human trials.

The immunosuppressant drugs cyclosporine and tacrolimus, widely used today to enable organ transplantation, were almost abandoned because animal toxicities failed to predict efficacy and safety in desperate patients. Corticosteroids, in contrast, appeared beneficial in animal models of septic shock but worsened mortality rates when administered to critically ill patients.

An Alzheimer's vaccine caused severe brain swelling in early human trials despite appearing safe in animal tests. A 2006 "cytokine storm" induced by an immunomodulatory agent by Tegenero left healthy volunteers with catastrophic organ failure, despite prior animal studies being unremarkable. In 2016, one volunteer died and four suffered severe neurological damage in a French trial, although the drug showed promise and acceptable safety margins across four animal species. Severe liver injury and multiple deaths forced the termination of a hepatitis B drug trial despite earlier encouraging animal data. Differential species sensitivity to drugs like acetaminophen further highlights the pitfalls of reliance on animal models. Gene therapy vectors that have been safe in animal tests have caused liver failure and brain swelling in children. HIV vaccines, stroke treatments, inflammatory disease agents, and Alzheimer's therapies have all elicited enthusiasm in animal models yet utterly failed in human trials.

These sobering examples have played out over decades, leaving patients dead or devastated in their wake. Notably, while these included extreme examples of unanticipated side effects, many milder problems might never be detected as patients already have many health problems and the additional negative effects of drugs are not easy to identify. On the other hand, many potentially lifesaving medicines may have been lost at the same time because they performed poorly in flawed animal models. Recurrent failures speak to inherent limitations of evolving human treatments in divergent species. These cautionary tales underscore growing calls to move away from unreliable animal testing toward human-relevant alternatives for future drug development.

Systematic evaluations of animal experiments

The last chapter gave some anecdotal examples of limitations of animal tests. Over the last decades an approach, which is called a systematic review, has evolved, which defines clearly upfront the question of interest and how to find the respective evidence and analyze it. This has been applied to some extent also to the value of animal testing.

The Systematic Review Centre for Laboratory-animal Experimentation (SYRCLE) works to improve the quality and reliability of animal studies used in drug discovery. One of the main tools developed by SYRCLE is the "risk of bias" (RoB) tool, which aims to assess the methodological quality of animal studies and has been adapted for aspects of bias that play a role in animal experiments. The tool is designed to enhance transparency and applicability, and it includes signaling questions to facilitate

judgment. The widespread adoption and implementation of this tool are expected to facilitate and improve the critical appraisal of evidence from animal studies. This may subsequently enhance the efficiency of translating animal research into clinical practice and increase awareness of the necessity of improving the methodological quality of animal studies. SYRCLE identified that a significant portion of animal research is conducted at a low standard, leading to unreliable data. This includes low rates of random allocation, allocation concealment, and blinded outcome assessment, all of which contribute to an overestimation of the benefits of experimental interventions. Furthermore, animal research often suffers from selective analysis and biased outcome reporting, where only the most positive outcomes are reported. This leads to an inflated proportion of studies with positive results and an overestimation of beneficial treatment effects. Systematic reviews have also highlighted redundancy and waste in animal research, with continued experimentation even after beneficial effects were already well documented, leading to unnecessary use of animals and resources. There is evidence that shortcomings in almost every aspect of the scientific design, conduct, and reporting of animal studies contribute to their inability to translate into benefits for humans. Such findings indicate the need for improved methodological quality in animal research to ensure its clinical relevance and enhance its efficiency and reliability translating into clinical practice.

SYRCLE also advocates for the registration of all animal experiments at inception and the publication of protocols of animal studies in various journals. These practices are expected to improve the standard of research in animal sciences. However, it is important to note that animal studies have inherent limitations and can sometimes be misleading in drug discovery. For instance, a drug that shows promise in animal models may not necessarily be effective in humans due to species-specific influences and differences in biology. Importantly, SYRCLE recommends that the risk of bias assessment should be conducted by at least two independent reviewers to ensure objectivity and that any disagreements be resolved through consensus-oriented discussions or by consulting a third person. This approach underscores the need for critical and unbiased assessment in animal studies, which can significantly impact the translation of research findings from animal models to clinical applications. In summary, the work of SYRCLE, particularly through its RoB tool, has been instrumental in identifying and mitigating bias in animal studies, thereby enhancing the reliability and translatability of these studies into human clinical research—especially in the context of drug discovery. Therefore, while tools like SYRCLE's RoB tool can help improve the quality of animal studies, they cannot completely eliminate these fundamental challenges.

A review by researchers at Astra Zeneca found that over half of the protocols for forthcoming animal experiments needed amendment for proper experimental design, appropriate sample sizes, and measures to control bias. Additionally, revealing reports from pharmaceutical companies have found that much data from academia are irreproducible, indicating problems of poor experimental design and scientific conduct, as well as incomplete reporting.

The Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies (CAMARADES) is a research group that aims to improve the quality of preclinical research, particularly in the context of animal studies used in drug discovery. CAMARADES works to address these issues by promoting rigorous, high-quality, and transparent animal research. This includes advocating the use of systematic reviews and meta-analyses, improving experimental design and reporting, and developing new methodologies to assess the quality of preclinical research. CAMARADES is a database that tracks the reliability and limitations of animal research used in drug development and disease research. It was created in response to the recognition that animal studies frequently do not translate to humans, wasting resources and potentially misleading medical research. For example, one analysis found that only 37% of highly cited animal research was translated at the level of human randomized trials. Another study found that only 8% of basic science discoveries enter routine clinical use within 20 years. The reasons why animal studies can be misleading include differences in biology and physiology between species, poor experimental design and reporting, publication bias, and overinterpretation of results. CAMARADES reviews animal studies systematically and critically to assess their limitations and risk of bias. The goal is to improve the design, analysis, and reporting of preclinical animal studies so that their results are more relevant to human health and avoid wasted resources. CAMARADES has reviewed numerous animal studies of drugs and conditions like stroke, amyotrophic lateral sclerosis, and sepsis, demonstrating how animal models failed to predict human outcomes. Overall, CAMARADES aims to act as a watchdog for animal research, promoting more rigorous methodology and cautious interpretation to prevent animal studies from misleading medical research.

Why we still need animals for drug development

While flawed, animal research remains necessary for developing new medicines. Some of the reasons it persists include:

- Living systems: animals are complex living organisms that cannot yet be mimicked in the laboratory. Seeing responses across multiple organs over time requires whole-animal studies.
- Rules and expectations: regulators overseeing drug safety expect animal data before human trials can proceed. Companies must comply to keep development programs on track.
- Early safety checks: animal tests allow safety assessments at high doses so that lower, likely safe human doses can be set. Without animals as a first check, putting chemicals into people would be too risky.
- Mechanism exploration: animal tests shed light on disease mechanisms and biological pathways to help guide human research, despite not always directly predicting outcomes.

Hence, while animals are far from ideal for predicting human drug responses, they fill important gaps until technology like organ chips and computer models can provide comparable living system data (as discussed below). Animal research therefore remains ingrained in medical advancement at present.

Strategies to maximize animal data value

Such strategies include the following:

1) Assessing for flaws: tools that help analyze the quality of animal methods to improve their applicability to humans.
2) Tracking outcomes: the registration of animal trials enables the subsequent tracing of results *versus* initial expectations.
3) Documentation: cataloging animal study successes *versus* failures can inform realistic healthcare promises.

Additionally, boosting reproducibility—consistency in results—returns more knowledge per animal used. This involves careful experimental design and transparent outcome reporting, whether positive, negative, or inconclusive. In summary, wasting animal lives on poorly designed, biased research is unethical, but ensuring the thoughtful conducted of robust animal studies via quality checks while tracking outcomes will advance human medicine with care while alternatives are developed.

Reproducibility of animal studies

A helpful estimate for the accuracy of a test is its reproducibility as no test can be more accurate than it is reproducible. Reproducibility in animal testing is a significant concern in the scientific community, with many studies highlighting the challenges and proposing strategies to improve the situation. One of the key issues affecting reproducibility is biological variation, which can cause organisms' responses to experimental treatments to vary with both genetic differences and environmental conditions. Another contributing factor is the extreme standardization of trial design, which can lead to different results with slight deviations in test conditions. Even with well-planned and well-reported protocols, reproducibility is not automatically guaranteed. This is known as the "reproducibility crisis", which has led to a growing awareness that the rigorous standardization of experimental conditions may contribute to the poor reproducibility of animal studies.

Estimating the overall accuracy of animal testing in predicting efficacy and safety in human trials is challenging, based on available data. Animal studies seem to have relatively low accuracy for predicting efficacy—estimates range from about 37% to 60% correlation with human outcomes, suggesting substantial limitations. The models remain quite imperfect for their core intended purpose. There are significant inter-species differences in biology and disease progression for even highly conserved pathways. However, animal studies seem moderately accurate regarding safety, with estimates of about 70% accuracy for identifying toxic side effects that also manifest in humans. So, while still imperfect, animal testing appears, on average, better attuned to flagging potential safety issues that translate across mammals. Overall, however, predicting efficacy via animal models seems scarcely better than a coin flip based on meta-analysis. However, for safety, animal testing achieves perhaps higher accuracy under optimal conditions. Combined into an overall likelihood of success, this aligns with very high late-stage drug failure rates; animal studies do not sufficiently recapitulate human biology to reliably identify those rare winning drug candidates out of the thousands investigated. Improved models and biomarkers remain a key necessity.

These accuracy limitations highlight why robust statistics and good judgment are so crucial when interpreting pre-clinical animal research for candidate prioritization and advancement decisions—a nuanced understanding of what questions different models can actually address is essential to avoid wasting of resources by chasing false signals.

The testing challenge illustrated

Testing means that individual chemicals are subjected to a measure to classify them as belonging, in the simplest case, to either of two classes—for example, effective on a target or not, or toxic/non-toxic. The problem is that there are no perfect tests, and some misclassifications occur: "false-positives" (ascribing a property which something does not have) and "false-negatives" (missing an individual chemical that is, in fact, a property). The basis for a test is that we can measure something which distinguishes the two groups. For example, if we want to distinguish male and female individuals, we might exploit the difference in height. This might not be the best possible characteristic, as Figure 10 shows. However, if we take a cut-off of 1.70 m, we actually identify about 90% of women and include only 35% of all men—27% of the below-1.70 m group are men, or an accuracy of 77.7%. This is about the accuracy we can hope to achieve with an animal experiment when looking for a property. While this works astonishingly well, what happens when there are less women in the group to be analyzed? If we assume that the number of women is only one tenth of the actual proportion, we now still find 90% of these and the same number of men as before. The problem is that we still find the same number of small men, who are now 79% of all identified, so the false-positives (men mistaken as women based on height) now predominate. The accuracy drops to 68% if we continue and reduce the number of women in the group again to another 10th, so the real women found are less than 4% of all small people identified, corresponding to an accuracy of 66%. Why is the accuracy still that good when only one woman per 25 men is among those identified? Because the method is very good at identifying non-women: the large number of tall men with now very few tall women is making the test fairly reliable for identifying them. This is a very fundamental problem, which most people do not understand: we test our methods with more or less equal numbers of what we want to identify, and our methods do very well. Then, we move to real life, and there are very few suitable substances we want to identify among those we evaluate. This is called the prevalence problem (see next section).

This example can also serve to show how we can change the performance of the test by setting our cut-off. We can make the test more sensitive (find more of what we are looking for) or more specific (minimize the false calls). The cut-off at 1.70 m found 90% of women, but of all called women, only 72.9% were in fact women. Changing the cut-off to 1.60 m finds only 46% of all women. Changing to 1.80 m finds 89.7% of all women but only 58.2% of those identified were correct. The accuracy of the test drops from 77.7% to 64.1% and 62.4%, respectively. This

illustrates how our choices allow us either to be confident in the result (specific) or not to miss out on positive things (sensitive). Translated to drug screening, this means either quickly reducing the number of possible substances or being careful not to lose the good ones.

The prevalence problem in drug discovery

When developing new drugs, researchers face a tricky problem: many of the effects they seek, both positive and negative, are quite rare. For example, out of thousands of drug candidates tested, only a small percentage end up being sufficiently safe and effective to bring to market as approved treatments. Regarding safety, dangerous side effects may also only occur in a tiny fraction of patients, still prohibiting their use. This means that, even when using very good laboratory tests and clinical trials, it can be hard to reliably detect these rare events—a drug could fail late in development over toxicity seen in 1 in 10,000 people, for instance. So scientists must test large numbers of drug candidates and use very large patient groups, which takes extra time and money. Careful testing design and statistics are key to properly estimate the likely benefits and risks of dealing with such low probabilities. Just as diagnostic tests in medicine work best for common diseases, the drug development process works far better for more prevalent drug effects. Clever ways to accurately find "needles in the haystack" during development is a permanent challenge; the intrinsic challenge of identifying rare yet significant events hampers the discovery of a truly effective drug or the detection of uncommon toxic effects of drug candidates. The vast majority of compounds investigated do not make it to market, either due to lack of efficacy or to adverse effects that may only be evident in a small fraction of the population or under specific conditions. This "needle in a haystack" problem is compounded by the fact that preclinical models, such as animal studies, do not always accurately predict human responses. Consequently, a drug that appears promising in preclinical trials may fail in clinical phases due to unforeseen toxicities or lack of therapeutic effect. On the other hand, a potentially useful drug might be erroneously discarded if its benefits are not readily apparent in the early stages of testing or if its side effects are overrepresented in preclinical models. Therefore, the efficiency of drug development is often hindered by the difficulty of extrapolating data from a limited set of preclinical results to the diverse human population, where genetic, environmental, and lifestyle factors can greatly influence drug responses. The prevalence problem underscores the need for more predictive models and testing methods that can better capture the complexity of human biology and disease. As in diagnostics, the predictive value of clinical trials decreases dramatically the less prevalent an outcome is. Companies must account for this limitation with very large and lengthy studies, at substantial cost. Clever trial designs to accurately detect these "needles in the haystack" remain an ongoing necessity in drug development.

Rare phenomena of high impact are sometimes called "black swan events". Nicolas Taleb in his book *The Black Swan* (2007) used this metaphor to especially describe events on the stock market. He defines black swan events by the "*. . .triplet: rarity, extreme impact and retrospective (though not prospective) predictability.*" This is exactly what drug discovery is: real hits are rare, they are a goldmine, and arguably, we can explain why they work so well only in retrospect. The identification of a new marketable drug requires much searching and luck. The same can be said inversely of the rare toxic effects of drugs coming to the market. Side effects which only occur in one in 1000 or 10,000 patients cannot be predicted: they are black swans. Taleb notes, "*What is surprising is not the magnitude of our forecast errors, but our absence of awareness of it.*" This is when the black swan hurts. "*True, our knowledge does grow, but it is threatened by greater increases in confidence, which makes our increase in knowledge at the same time an increase in confusion, ignorance, and conceit.*" This notion can easily be translated to adverse drug effects, where late discoveries of highly problematic side effects are rare but game-changing events.

A prime example is the case of the painkiller Vioxx (rofecoxib). Vioxx was initially hailed as a breakthrough for its effectiveness in relieving pain with fewer gastrointestinal side effects than other painkillers. This was a significant development, given that gastrointestinal complications are a common and serious side effect of the long-term use of such drugs. It was approved and marketed for 5 years before being withdrawn due to increased risk of heart attack and stroke. During clinical trials, it was observed that 2.4% of the 1,287 participants taking Vioxx suffered serious cardiac events, such as heart attacks, chest pain, or sudden death. This rate was notably higher than the less than 1% of patients who received a placebo. This significant increase in risk, although relatively small in percentage terms, led to the drug's withdrawal from the market due to safety concerns. The problem was that these cardiovascular risks occurred in only a small proportion of patients—about 1 in 200 over a year of treatment based on later analyses. So even with thorough testing, this rare side effect was initially missed. The company had to spend over $100 million on one study alone to properly detect these risks, requiring a huge sample of over 24,000 arthritis patients. Since heart disease progresses at a background rate regardless, only by analyzing such large numbers could Vioxx's small but real added risk be identified. The Vioxx case illustrates why finding rare adverse events or benefits is so difficult during development; even the most rigorous testing can miss effects that occur at rates of less than 1 in 1000. In preclinical trials and early clinical studies, Vioxx did not show significant adverse effects and was therefore approved by the FDA. However, after it was widely marketed and prescribed, it became apparent that there was an increased risk of heart attack and stroke associated with its use, which was not evident in the smaller, controlled clinical trials. The economic consequences of the Vioxx withdrawal were profound and multifaceted. Vioxx, which had been on the market since 1999, was generating over $2.5 billion annually for Merck, accounting for approximately 10% of its worldwide sales. When the drug was withdrawn in September 2004, Merck's sales plummeted, and the company's stock value took a significant hit. Moreover, the withdrawal triggered numerous high-profile product-liability lawsuits, leading to years of litigation that cost Merck billions of dollars. The Vioxx case remains a cautionary tale in the pharmaceutical industry, illustrating the staggering financial risks when safety concerns emerge post-market. The industry continued to feel the repercussions of the Vioxx withdrawal up to a decade later as it highlighted the vulnerabilities in drug safety surveillance and the potential for significant economic loss when

widely prescribed medications are retracted. The Vioxx case demonstrates the prevalence problem where rare but critical adverse events may not be detected until after a drug is approved and taken by a large and diverse patient population. It also shows the limitations of preclinical models in predicting human real-life outcomes, given that the cardiovascular risks associated with Vioxx use were not captured in earlier studies. This highlights the need for more comprehensive and sensitive methods for detecting rare events in drug safety and efficacy evaluations. We will later discuss the opportunities of human-relevant bioengineered models (microphysiological systems), mechanistic understanding, and big-data-driven analyses and modeling.

Another prime example is the cholesterol-lowering drug Lipobay (cerivastatin), which was withdrawn in 2001 after reports of serious muscle toxicity (rhabdomyolysis). This side effect occurred in approximately 1 out of every 1000 patients per year who took the approved dose. While quite rare, the results could be fatal. Even though cerivastatin had undergone extensive laboratory testing and clinical trials with thousands of patients prior to approval, this low probability meant that the risk was initially missed. The analysis of over a million patient years of post-approval prescription data was required to finally detect and quantify the risk. The Lipobay case, like that of Vioxx, demonstrates how developing or approved drugs can fail to identify rare but dangerous risks that only show up when tested in extremely large populations. As in medical diagnostics, even rigorous testing can easily miss outcomes that occur at rates less than around 1 in 1000. Companies must account for this limitation by conducting very large and lengthy studies to properly estimate safety and efficacy; however, even these might not be large enough to conclusively rule out some risks. Careful trial analysis for faint signals in the data is crucial.

Some number games and the difficulty of finding rare things

As seen above, drug discovery means ultimately finding one marketable drug out of more than 10,000 chemicals. The problem is that our tools are far from perfect. This holds for both animal tests and their alternatives. This is like solving a riddle with glasses not tailored to our eyesight. Let us assume that an animal tests deliver 90% correct results—a relatively high bar, with no more than 80% accuracy much more likely; however, for illustration, assume a 90% accurate animal test to try to discover one approvable drug out of 10,000 chemical candidates. Testing 10,000 chemicals, the 90% accurate test would correctly identify the one truly effective compound that will ultimately make it to market. However, with 10% false positives, it would also flag around 999 other chemicals as "hits" that will actually fail later. So, while not missing the promising needle in the haystack, initial results are unable to distinguish it from almost a thousand false positives. A large fraction of those 1000 extras would drop out in further rounds due to other limitations of course—but companies might still fruitlessly pursue 100 through later stages as if they were promising, based on the inaccurate early read. This thought exercise illustrates why, despite relatively good animal tests, failure rates in human trials remain high—rare actual positives get lost amongst the noise of greater numbers of false signals when working in domains of very low prevalence. Clever multi-parameter testing is important, but statistics dictate inevitable disappointment much of the time.

Here is an illustration of what would happen if the 1000 chemicals flagged as positives from the first 90% accurate animal test were run through a second, independent 90% accurate animal test: putting those 1000 chemicals through a second 90% accurate animal test independent of the first (an unlikely assumption, but useful here) might help refine the list, but major issues remain due to the low prevalence. The true promising drug would be confirmed, while around 900 of the original 999 false positives would now test negative and could be set aside. However, around 100 (10%) of those false leads would be incorrectly flagged positive again. Therefore, out of 110 total positives between the two tests, only one is the real winner, over 100 remain misleading false leads, and optimization between tests still cannot avoid this. Even added testing helps far less than intuition would suggest when fundamental probabilities are so low. Statistics dictate that reliability decreases exponentially the less prevalent the needles sought in research haystacks become. At huge scale, noise drowns signal without escape. While these are simple examples for illustration, these dynamics genuinely occur in real drug development pipelines, contributing to late failure and showing how proper expectations are vital when hunting for rare events like a 1-in-10,000 for a future drug (Figure 11).

If we assume a series of independent 90% accurate animal tests to narrow down the initial 10,000 compounds, we can analyze the number of tests to get to 10 remaining candidates, and the cumulative risk of losing the one truly promising compound:

* Round 1 test: ~1000 compounds flagged as positives (= candidates, ~1 true, ~999 false)
* Round 2 test: ~100 compounds flagged as positive (on average ~1 true, ~99 false)
* Round 3: ~10 flagged (~1 true, ~9 false)

It thus takes three sequential 90% accurate tests to narrow the 10,000 down to 10 compounds (one likely true positive, nine remaining false positives). However, there is in each round a 10% chance of the truly promising compound testing negative in one of the rounds and being incorrectly discarded. Over three tests, this means that there is actually a 27% chance ($1–0.9^3$) that the best candidate is lost along the way. This demonstrates how prevalence limitations mean that even an unlikely-to-achieve series of nearly perfect laboratory tests carries large risks of losing the rare "needle" when searching complex multidimensional haystacks like potential drug spaces. Confirming true signals remains improbable until late, so balancing information gain *versus* discarding promising niche opportunities remains an ongoing challenge throughout the drug discovery pipeline. What happens when we use less than ideal tests? Using 80% accurate tests, we need four rounds to get us to ~16 compounds, and the likelihood of still including the golden one is 41%. Using a series of 70% accurate tests, six test rounds will get us to ~7 compounds with only a 12% chance of still having the one we are looking for included.

These calculations assume that there is only one marketable substance in the 10,000 we start with. That is probably not the case. Assuming that were ten suitable compounds among the 10,000 at

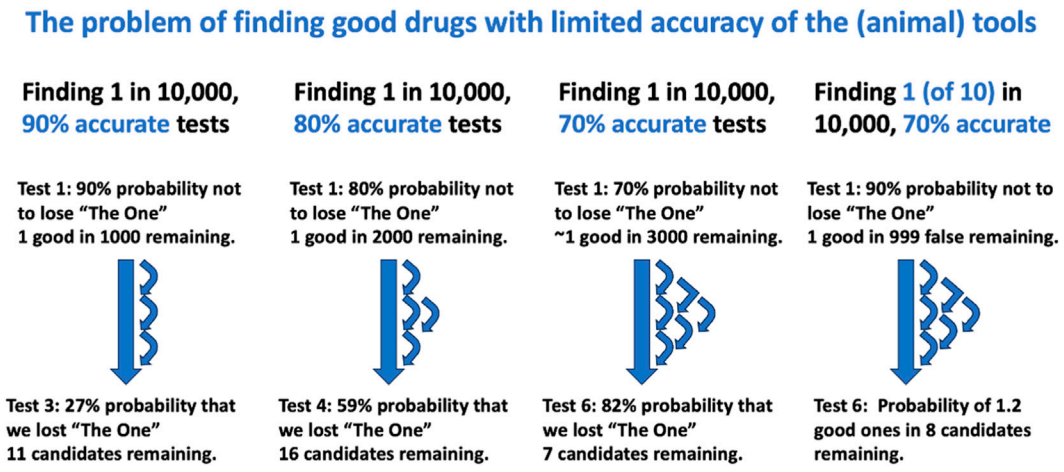


FIGURE 11
Illustration of the consequences of limited accuracy of finding 1 in 10,000 drug candidates The calculations show how less accurate tests require more testing rounds to bring the candidates down to about 10, which can be managed in clinical trials but also increase the risk of losing the winning one. If we assume that there are 10 equally good candidates, there is a much better chance that at least one will proceed to clinical trials.

start, a series of 90% accurate tests gets us, in three rounds, to ~17 compounds including seven possible winners, and 80% accurate tests in four rounds to ~20 compounds with four possible winners, as well as 70% accurate tests leading in six rounds to ~9 candidates including one promising one. The latter scenario seems to best agree with the experience that one in ten compounds later prove to work in clinical trials, with some of the numbers around 70% for the reproducibility of animal studies.

If we come back to the sensitivity vs. specificity discussion from above, we can illustrate the consequences here. We saw that 70% accurate tests (equal sensitivity and specificity) in six rounds brought us to nine compounds (one good). If we now use 80% sensitive/60% specific tests, we need seven rounds to get to 19 (two good ones), while the opposite 60% sensitive/80% specific gets us to 17 in four rounds (one good one). This again illustrates the compromise between sensitivity and specificity: higher specificity sorts the compounds faster at the risk of losing the winner; higher sensitivity means more effort (seven rounds) but no real gain in the probability of including the winner.

Do we know the accuracy (sensitivity and specificity of animal tests)? Often not, because that requires an assessment of the assay against some reference, such as chemicals, which are known to do what the drug discovery is seeking. We call this "target validation". The above calculations thus better serve the purpose of explaining why so much testing does not necessarily lead to substances which succeed in the clinic.

Safety testing—that is, toxicology —traditionally occurs just before human trials and in part concurrently. This means that the ~10 compounds entering the clinical phase of drug development need to be considered. Applying the estimated 70% safety accuracy of animal studies to a scenario with a 20% prevalence of a toxicity across 10 candidate compounds means that two compounds would be truly toxic to humans and eight would be truly safe. Using a hypothetical 70% sensitive/specific animal test on the two toxic compounds, it would correctly flag one or two as toxic; of the eight safe compounds, it would correctly identify six to seven. It would also incorrectly flag one to two of the safe compounds as toxic and misidentify one of the truly toxic compounds. Therefore, even with a relatively high prevalence toxicity of 20% and a good animal test with 70% accuracy, predictions can easily miss 25% of the unsafe human compounds while allowing unsafe candidates through at a 10%–20% rate. This demonstrates how testing limitations can quickly add up, even under idealized conditions—rare but dangerous outcomes get missed completely, and false safety signals erode confidence in labeling. Layered risk mitigation is key, but balancing information value against decision risk given the constraints around rare event prediction remains highly challenging throughout pharmaceutical pipelines.

How realistic are these number games?

In real life, not all steps will be run on all compounds. Such a brute-force approach is simply not realistic and affordable. Early rounds will likely be done with simple *in vitro* and *in chemico* tests with limited scope but better reproducibility. With additional information such as the intellectual properties for chemicals, ease of synthesis, estimated environmental stability, and chemicophysical properties, lead compound selection will proceed faster. Often, new chemical structure variants will be brought in on the lead-optimization phase. This does not necessarily improve the odds of ultimate success as this is somewhat a gamble based on experience and circumstantial information. It is quite possible that these considerations have a similar accuracy of about 70% and thus leave us within the calculations; in fact, 20 years ago, Romualdo Benigni and colleagues had scientists guess the outcome of cancer tests on chemicals and achieved 60%–65% accuracy. So, they were about as good as mathematical models or the reproducibility of cancer testing itself. The above scenario also assumes that the different tests per round are independent; this is very unlikely as they are all built around the same pharmacological target, which reduces the probability of success. So this represents a theoretical

exercise of testing all and everything in a sequence of test rounds, which serves mainly to illustrate how the tools stand up against the task.

What can be done to improve the probabilities of finding good drug candidates in preclinical research?

There are a few approaches that can help improve the odds when searching for extremely rare positive events, like 1-in-10,000 successful drug candidates:

1) Test more compounds: this helps detect more of the few true signals hiding amidst the noise. This is accomplished with robotized testing—so-called high-throughput testing with libraries of often millions of chemicals. Artificial intelligence can examine even more theoretical structures, but the contribution of this new approach is still to be shown. However, returns of actual testing more compounds diminish quickly and costs scale up, limiting feasibility. This is only possible with broader use of non-animal methods.
2) Using more replicates per test such as larger animal groups: while this increases the accuracy of tests when variability is the problem, it again increases costs, effort, and animal use.
3) Multi-parameter testing: assessing multiple aspects of each compound provides backup if the primary indicator is misleading. However, interpreting interactions quickly becomes complex, and such "multiple testing" can weaken statistical power.
4) Seek supplementary data: extra information like structure analyses or genetic associations can flag higher probability starting points tied to known biology. This aids in prioritizing what to screen first, especially when the quantity of substances is limited.
5) Refine models over time: statistical models predicting success can incrementally improve as more test data accumulates across pipelines.
6) Limit false positives: overly sensitive screenings should be avoided, even if they capture most true hits; generating excessive false leads that consume resources is counterproductive when positives are the priority.
7) Expect imperfection: appreciating prevalence constraints means properly setting expectations around reliability and uncertainty given the state of knowledge.
8) Use methods with higher reproducibility, fidelity, and accuracy. Most cell culture systems and, certainly, computational models are more reproducible than animal experiments. With respect to modeling human responses, at least microphysiological systems (MPS) promise fewer species differences. In general, models which are based on the same mechanisms as in humans promise better fidelity. With respect to accuracy, determining which model is more accurate must be shown case-by-case; however, AI models have already outperformed animal tests for a number of toxicological hazards.

In the end, no solution can avoid the direct implications of probability theory that extremely rare events intrinsically strain the predictive capacity of any analytical approach. However, combining strategic testing with an understanding of these inherent limitations helps maximize the likelihood of teasing out promising needles from the early biomedical haystacks faced during drug discovery.

The pharmaceutical industry differs from other industries in its use of animal testing and adoption of alternative methods

Pharmaceutical companies conduct a lot of animal testing during drug development to establish safety and efficacy. Animal studies play a critical role in preclinical testing and are used more extensively in pharmaceutical R&D than in most other industries, such as industrial chemicals, consumer products, or food. However, pharmaceutical companies have also pioneered many alternative methods and been early adopters of new technologies to reduce animal use. Some key differences to other industries are as follows:

- Market pressures are different, with enormous upfront investments rewarded by higher prices and profit margins. Pharmaceutical companies face strong incentives to bring drugs to market quickly, so they are motivated to use the most predictive methods, whether animal or alternative. Speed and human relevance are more valued than following traditional protocols.
- R&D spending is massive, estimated at more than $2.6 billion per successful drug development. Pharma devotes tens of billions annually to R&D, giving it resources to implement new technologies. The scale of animal use also makes reduction efforts very cost-relevant.
- There is extensive regulation, but also flexibility. Drug development is heavily regulated to ensure safety, but regulatory agencies allow some discretion in test methods. Pharma takes advantage of opportunities to waive animal tests when alternatives exist.
- The range of tests required is broader. Pharma must assess a wide range of endpoints, from pharmacodynamics to carcinogenicity, which requires a diverse arsenal of animal and non-animal methods.
- Product development cycles are long because development is much more sophisticated. Drug development takes about 12 years on average, so new alternative methods may take time to impact animal use. However, each marketed drug is tested for years, so replacements can eventually have great impacts.
- Focus on mechanism. Understanding drug molecular mechanisms, especially with omics technologies (i.e., simultaneously measuring as many active genes, or proteins and metabolite changes, as possible) informs human biology relevance and helps justify waivers of animal tests.

In summary, pharmaceutical companies are highly motivated to implement improvements in safety testing that can accelerate drug development, improve clinical predictivity, and reduce costs. This has made them forerunners in adopting alternative methods for efficacy testing, despite continuing extensive animal testing requirements for safety.

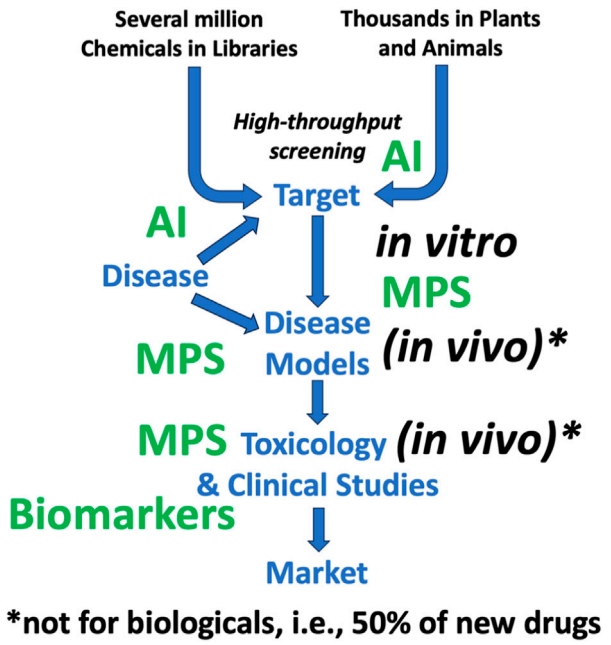


FIGURE 12
Increasing use of alternatives to animal testing in drug development The figure uses the same simplified work flow as Figure 4, where the new alternative approaches (green) discussed in this article are predominantly used. MPS = microphysiological systems, AI = artificial intelligence. It also makes the point that –50% of new drugs are biologicals, which usually cannot be adequately tested in animals anyway.

What are alternatives to animal testing in preclinical drug development?

The philosopher Peter Singer once said, "*I don't think there's much point in bemoaning the state of the world unless there's some way you can think of to improve it. Otherwise, don't bother writing a book; go and find a tropical island and lie in the sun.*" So, how can we improve? The main alternatives to animal testing are *in vitro* and *in silico* approaches (Figure 12). *In vitro* methods, while cheaper and faster, face issues like genetic instability and non-physiological culture conditions. However, advances in technology and practices, such as Good Cell Culture Practice (GCCP), are helping to overcome these limitations. *In silico* methods are now central to life sciences as they have evolved significantly, especially based on AI and also in regulatory contexts. Tools like Good Read-Across Practices and automated read-across, which leverage large toxicological databases, are increasingly used in drug discovery and other applications. Integrated testing strategies (ITS) are emerging which combine *in vitro*, *in silico*, and sometimes *in vivo* methods, recognizing that no single method can fulfill all information needs. This strategy, still in its early stages, is gaining traction in safety sciences with a more mechanistic design approach.

The shift towards non-animal methods aligns with a stronger focus on mechanistic research in biochemistry and molecular biology, offering a deeper understanding of physiology and disease. It is challenging to identify disease mechanisms in whole organisms or test specific mechanisms using complex animal models. Systems biology approaches are increasingly modeling this complexity.

Increasingly, mechanistic studies—that is, work elucidating the cellular and molecular aspects of disease and drug action—lead to surrogate measures ("biomarkers") of drug effects which can then be used in clinical trials to monitor efficacy more subtly and earlier than by clinical outcomes. This is also known as "translational medicine", which translates from preclinical to clinical work.

In summary, the reliance on animals to study human physiology and diseases is being questioned due to the emergence of alternative methods. These alternatives, although partial and simplistic, offer cheaper, faster, and potentially more robust means of data generation. Combining these methods in ITS or systems biology approaches is helping to overcome the limitations of each method, leading to a decreased reliance on animal testing in the scientific process.

Microphysiological systems such as microfluidic human organ chips for more predictive drug testing

A major challenge in developing new medications is that animal studies often fail to accurately predict whether a drug will be safe and effective in human patients. Animals differ from people in their biology and physiology, so drugs may behave differently in humans than in test animals. Conventional cell cultures also lack key features of real human organs.

To address this problem, scientists have developed innovative "organs on chips" that use microfluidic culture systems to mimic aspects of living human organs and tissues. Tiny channels allow cells to be cultured with flowing fluids that recreate blood flow and breathing motions. Multiple cell types can interact, like blood vessel cells linked to immune cells. Some systems even connect chips of different organs, like gut, liver, and brain.

These "microphysiological systems" (MPS) aim to model human biology more accurately than animal studies or regular cell cultures. Their ultimate goal is to better predict patient responses to drugs before human trials and thus reduce failures. Early studies suggest organ chips could help:

- Model complex human diseases involving multiple organs.
- Identify possible targets for drug action on cellular and molecular levels.
- Identify lead compounds out of a set of candidates.
- Optimize lead compounds by comparative testing of modifications.
- Identify biomarkers of clinical success to be measured later in clinical studies.
- Support an IND (initial drug development) review to move into clinical studies ("first in humans").
- Detect dangerous side effects missed in animal tests.
- Predict drug absorption, distribution, metabolism, and excretion.
- Test patient-derived cells for personalized medicine.

Challenges remain in validating organ chips and gaining regulatory acceptance. Nevertheless, combined with computer models and small, careful human studies, they could transform drug development to efficiently deliver effective, safe medicines matched to individual patients. While still experimental, the organ chip approach shows promise in providing more reliable human data on drug effects than animal models.

Adverse outcome pathways to improve drug safety testing

A major focus in the development of new drugs is detecting potential safety issues early, before patients are harmed in clinical trials. However, current safety testing methods often fail to predict all the adverse effects that emerge later. This leads to expensive late-stage drug failures and withdrawals of approved drugs.

To address this problem, the concept of "adverse outcome pathways" (AOPs) is gaining interest. AOPs map the chain of events from initial chemical–cell interactions to subsequent organ responses that ultimately lead to adverse health effects. They organize existing mechanistic knowledge into a sequence of:

- Molecular initiating events—how a chemical first interacts with a biomolecule.
- Key events—cellular, tissue, and organ responses.
- Adverse outcome—the adverse health effect.

AOPs aim to represent established pathways that lead to toxicity. Their development was driven by chemical safety regulations but they are relevant for drug toxicity as well. AOPs could improve drug toxicity prediction by the following:

- Elucidating species differences in toxicity pathways.
- Justifying when animal toxicity findings may not apply to humans.
- Allowing more mechanism-based safety testing methods.

AOPs are strengthened by broader "pathway of toxicity" (PoT) approaches that experimentally map early molecular perturbations using advanced omics technologies. PoTs provide detailed, dynamic networks while AOPs summarize established knowledge. Used together, AOPs and PoTs can enhance mechanistic understanding and modeling of drug safety.

Further efforts are still needed to expand and validate AOPs and PoTs. Nevertheless, mapping adverse outcome pathways promises to ultimately provide a compendium of toxicity mechanism knowledge. This could transform chemical and drug safety assessment to rely less on animal studies and more on human-relevant pathway-based approaches. Understanding toxicity pathways will enable the earlier and more reliable detection of key human hazards, vastly improving the drug development process.

The promise of AI in transforming drug development and toxicology

Pharmacology and toxicology have experienced a data revolution, transitioning from a historically small-scale discipline to one generating vast and heterogeneous evidence from high-throughput assays, omics technologies, electronic health records, and more. This exponential growth, coupled with increasing computational power, has created major opportunities for integrating artificial intelligence (AI) techniques to enhance chemical selection and hazard assessment. Early rule-based expert systems have given way to modern machine learning and, especially, deep learning models that find patterns in large datasets to predict toxicity. Notably, these methods are agnostic with respect to what effects are predicted, and similar approaches are available to predict pharmacological effects. Key developments include the following:

- Quantitative structure–activity relationships (QSARs) relating chemical descriptors to bioactivity.
- Public toxicity data repositories like Tox21, enabling AI model development.
- Deep neural networks that integrate chemical and bioassay data to predict diverse hazards.
- Natural language processing, exemplified by the current boom in large language models and mining legacy animal studies and literature.
- Explaining model behavior through explainable AI (xAI) techniques.

AI promises to transform areas like predictive toxicology, drug design, mechanistic understanding, risk assessment, and evidence integration. It can handle multifaceted data and capture uncertainties for robust probabilistic risk modeling. AI-derived knowledge graphs could link to adverse outcome pathways. However, biases, reproducibility, and interpretability remain challenges. AI models require extensive curated training data. Multidisciplinary collaboration is essential for human-centered, trustworthy systems tailored to enhance chemical safety decisions. AI is not a panacea but rather an enabling tool that must be thoughtfully designed and utilized alongside ongoing efforts to improve primary evidence generation and appraisal. It increasingly qualifies as a copilot but is not yet ready to take the pilot's seat. Overall, the symbiotic integration of AI and modern data-rich toxicology has immense potential to transition the field into a more predictive, mechanistic, and evidence-based scientific discipline to effectively promote human and environmental health.

Moving beyond animal testing with integrated approaches

Toxicity testing has traditionally relied heavily on animal models. However, differences between species mean that animal data do not always accurately predict human responses. There is a growing focus on new approach methodologies (NAMs) to replace or reduce animal use for ethical and scientific reasons, but individual alternative tests are often limited, requiring combination into integrated strategies.

Integrated approaches to testing and assessment (IATA) strategically combine results from multiple NAMs. Sources can include computer models, cell cultures, organ chips, and lower animal species. IATAs also incorporate existing data via weight-of-evidence assessment.

IATAs follow three key steps: 1) compile existing data on a chemical, 2) evaluate data to determine if they are sufficient for decision-making or if new data are needed, and 3) generate new data through targeted testing to fill gaps. IATAs have been developed for skin and eye hazard testing, incorporating animal-free methods like reconstructed human tissue models. Work is ongoing for complex endpoints like cancer and developmental toxicity.

Challenges for IATAs include 1) determining the best test combinations, 2) assessing predictivity, 3) validating integrated approaches, and 4) obtaining regulatory acceptance. IATAs do not necessarily avoid animal tests completely, but they do strategically combine new methods to significantly reduce and refine animal use. An intelligent combination of advanced models with computational approaches offers a path to enhanced predictivity of human outcomes. IATAs represent a pragmatic approach to transition from an animal-centered paradigm to more human-relevant 21st century toxicology. Animal models provide useful but limited data on drug effects in humans. Their flaws lead to many trial failures and to unsafe drugs reaching patients. However, animals remain necessary where better techniques are lacking. Ongoing advances in human cell studies, tissue chips, computer models, and innovative small human trials can make drug development more accurate, ethical, and effective for the diverse spectrum of patients needing safe and beneficial new treatments. The concept of integrated testing strategies originated mainly from the testing challenges for industrial chemicals. However, the concept applies well to the drug discovery process. This includes the combined use of tests with complementary characteristics, with results analyzed together and not sequentially.

Conclusion: the paradox of preclinical animal research in drug development

The journey of drug development is marked by a paradoxical reliance on preclinical animal research, despite its inherent flaws and limitations. This review has considered the complexities of this relationship, highlighting both the indispensable role and significant challenges posed by animal studies in the realm of therapeutic development.

Flaws and limitations of animal studies, while being a cornerstone of drug development, include physiological differences and misleading outcomes. Even systematic approaches, while improving methodological robustness, cannot fully overcome the inherent limitations of animal models, such as species-specific biological differences. The low translation to clinical use of approximately of 8% within two decades underscores the pitfalls of animal studies. The reproducibility of animal testing, a cornerstone for the reliability of any test, is under significant scrutiny. Biological variation and extreme standardization in trial designs contribute to this crisis, leading to poor reproducibility and questionable accuracy in predicting human efficacy and safety—the *reproducibility crisis*.

Drug discovery confronts a unique challenge: the rarity of both positive and negative effects sought in drug candidates. The "needle in a haystack" problem is exacerbated by the fact that preclinical models, such as animal studies, do not always accurately predict human responses, making the identification of effective drugs or the detection of toxic effects exceptionally difficult. Drug discovery has been likened to searching for black swan events—rare, impactful, and mostly unpredictable occurrences. The identification of new marketable drugs and the late discovery of significant side effects in widely used drugs are quintessential examples of such events in the pharmaceutical industry.

Several strategies to enhance the probability of identifying successful drug candidates have been discussed here. In summary, while animal research remains a crucial element in therapeutic development, its limitations and flaws necessitate continuous improvement in study design, methodology, and complementary human-relevant systems. The complexity of translating animal data to human applications underlines the need for more predictive models and testing methods that can better capture human biology and disease nuances, including but not limited to microphysiological systems (MPS) and artificial intelligence (AI). As the field evolves, a balanced approach that acknowledges both the value and the limitations of animal studies will be essential in advancing drug development. The new approach methods are already performing, sometimes astonishingly well. Franz Kafka once said, "*There is a goal but no way; what we call the way is our hesitation!*" This sentiment resonates in the context of animal experiments in drug safety testing, and perhaps we should stop hesitating and just embrace the new opportunities.

Author contributions

TH: writing–original draft and writing–review and editing.

Funding

The author declares that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Akhtar, A. (2015). The flaws and human harms of animal experimentation. *Camb. Q. Healthc. Ethics* 24 (4), 407–419. doi:10.1017/S0963180115000079

Allen, A. (2006). Of mice or men: the problems with animal testing. Slate. Available at: https://slate.com/technology/2006/06/does-animal-testing-work.html (Accessed December 11, 2023).

Arrowsmith, J. (2012). A decade of change. *Nat. Rev. Drug Discov.* 11, 17–18. doi:10.1038/nrd3630

Bailey, J., and Balls, M. (2019). Recent efforts to elucidate the scientific validity of animal-based drug tests by the pharmaceutical industry, pro-testing lobby groups, and animal welfare organisations. *BMC Med. Ethics* 20, 16. doi:10.1186/s12910-019-0352-3

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature* 533, 452–454. doi:10.1038/533452a

Check Hayden, E., and Hayden, T. (2014). Strike threat over jailed primatologist. *Nature* 448, 634. doi:10.1038/448634a

Daneshian, M., Busquet, F., Hartung, T., and Leist, M. (2015). Animal use for science in Europe. *ALTEX* 32, 261–274. doi:10.14573/altex.1509081

DiMasi, J. A., Grabowski, H. G., and Hansen, R. W. (2016). Innovation in the pharmaceutical industry: new estimates of R&D costs. *J. Health Econ.* 47, 20–33. doi:10.1016/j.jhealeco.2016.01.012

Everitt, J. I. (2015). The future of preclinical animal models in pharmaceutical discovery and development: a need to bring in cerebro to the *in vivo* discussions. *Toxicol. Pathol.* 43, 70–77. doi:10.1177/0192623314555162

Fratta, I. D., Sigg, E. B., and Maiorana, K. (1965). Teratogenic effects of thalidomide in rabbits, rats, hamsters, and mice. *Toxicol. Appl. Pharmacol.* 7, 268–286. doi:10.1016/0041-008X(65)90095-5

Frommlet, F. (2020). Improving reproducibility in animal research. *Sci. Rep.* 10, 19239. doi:10.1038/s41598-020-76398-3

Frommlet, F., and Heinze, G. (2021). Experimental replications in animal trials. *Lab. Anim.* 55, 65–75. doi:10.1177/0023677220907617

Harris, R. (2017). *Rigor mortis: how sloppy science creates worthless cures, crushes hope, and wastes billions*. New York: Basic Books.

Hartung, T. (2007). Food for thought . . . on cell culture. *ALTEX* 24, 143–152. doi:10.14573/altex.2007.3.143

Hartung, T. (2008). Food for thought. on animal tests. *ALTEX* 25, 3–16. doi:10.14573/altex.2008.1.3

Hartung, T. (2009). Per aspirin *ad astra*. *ATLA - Altern. Lab. Anim.* 37 (Suppl. 2), 45–47. doi:10.1177/026119290903702S10

Hartung, T. (2013). Look Back in anger – what clinical studies tell us about preclinical work. *ALTEX* 30, 275–291. doi:10.14573/altex.2013.3.275

Hartung, T. (2016). Making big sense from big data in toxicology by read-across. *ALTEX* 33, 83–93. doi:10.14573/altex.1603091

Hartung, T. (2023). ToxAIcology - the evolving role of artificial intelligence in advancing toxicology and modernizing regulatory science. *ALTEX* 40, 559–570. doi:10.14573/altex.2309191

Hooijmans, C. R., de Vries, R. B. M., Ritskes-Hoitinga, M., Rovers, M. M., Leeflang, M. M., IntHout, J., et al. (2018). Facilitating healthcare decisions by assessing the certainty in the evidence from preclinical animal studies. *PLoS ONE* 13, e0187271. doi:10.1371/journal.pone.0187271

Hooijmans, C. R., Rovers, M. M., de Vries, R. B., Leenaars, M., Ritskes-Hoitinga, M., and Langendam, M. W. (2014). SYRCLE's risk of bias tool for animal studies. *BMC Med. Res. Methodol.* 14, 43. doi:10.1186/1471-2288-14-43

Leist, M., and Hartung, T. (2013). Inflammatory findings on species extrapolations: humans are definitely no 70-kg mice. *Arch. Toxicol.* 87, 563–567. doi:10.1007/s00204-013-1038-0

Meigs, L., Smirnova, L., Rovida, C., Leist, M., and Hartung, T. (2018). Animal testing and its alternatives – the most important omics is economics. *ALTEX* 35, 275–305. doi:10.14573/altex.1807041

Nuffield Council on Bioethics (2005). The ethics of research involving animals. Available at: https://www.nuffieldbioethics.org/wp-content/uploads/The-ethics-of-research-involving-animals-full-report.pdf (last accessed December 12, 2023).

Pound, P. (2023). *Rat trap – breaking free from the illusion of progress in animal research*, 304. Matador.

Pound, P., and Bracken, M. B. (2014). Is animal research sufficiently evidence based to be a cornerstone of biomedical research? *Brit. Med. J.* 348, g3387. doi:10.1136/bmj.g3387

Pound, P., and Ritskes-Hoitinga, M. (2018). Is it possible to overcome issues of external validity in preclinical animal research? Why most animal models are bound to fail. *J. Transl. Med.* 16, 304. doi:10.1186/s12967-018-1678-1

Richter, S. H., Garner, J. P., and Würbel, H. (2009). Environmental standardization: cure or cause of poor reproducibility in animal experiments? *Nat. Methods.* 6, 257–261. doi:10.1038/nmeth.1312

Ritskes-Hoitinga, M., and Pound, P. (2022). The role of systematic reviews in identifying the limitations of preclinical animal research, 2000 – 2022. JLL Bulletin: commentaries on the history of treatment evaluation. Available at: https://www.jameslindlibrary.org/articles/the-role-of-systematic-reviews-in-identifying-the-limitations-of-preclinical-animal-research-2000-2022/ (Accessed December 12, 2023).

Skrabanek, P., and McCormick, J. (1989). *Follies and fallacies in medicine*. Glasgow: Tarragon Press. Available at: https://chagall.med.cornell.edu/Skrabanek/Follies-and-Fallacies-in-Medicine.pdf.

Spanagel, R. (2022). Ten points to improve reproducibility and translation of animal research. *Front. Behav. Neurosci.* 16, 869511. doi:10.3389/fnbeh.2022.869511

Van Norman, G. A. (2019). Limitations of animal studies for predicting toxicity in clinical trials: is it time to rethink our current approach? *JACC Basic Transl. Sci.* 4, 845–854. doi:10.1016/j.jacbts.2019.10.008

Van Norman, G. A. (2020). Limitations of animal studies for predicting toxicity in clinical trials: Part 2: potential alternatives to the use of animals in preclinical trials. *JACC Basic Transl. Sci.* 5, 387–397. doi:10.1016/j.jacbts.2020.03.010

Vargesson, N. (2015). Thalidomide-induced teratogenesis: history and mechanisms. *Birth Defects Res. C Embryo Today* 105, 140–156. doi:10.1002/bdrc.21096

Voelkl, B., Altman, N. S., Forsman, A., Forstmeier, W., Gurevitch, J., Jaric, I., et al. (2020). Reproducibility of animal research in light of biological variation. *Nat. Rev. Neurosci.* 21, 384–393. doi:10.1038/s41583-020-0313-3

von Aulock, S., Busquet, F., Locke, P., Herrmann, K., and Hartung, T. (2023). Engagement of scientists with the public and policymakers to promote alternative methods. *ALTEX* 39, 543–559. doi:10.14573/altex.2209261

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 11 April 2024
DOI 10.3389/fpubh.2024.1276536

OPEN ACCESS

EDITED BY
Katya Tsaioun,
Johns Hopkins University, United States

REVIEWED BY
Jean-Luc Poyet,
INSERM U976 Immunologie, Dermatologie, Oncologie, France
Minjun Chen,
National Center for Toxicological Research (FDA), United States
Alecia M. Fair,
Vanderbilt University, United States
Manuel Martín-Díaz,
Andusian Health Service, Spain

*CORRESPONDENCE
Yurong Zhang
✉ zyr876281137@126.com

†These authors have contributed equally to this work

RECEIVED 19 September 2023
ACCEPTED 25 March 2024
PUBLISHED 11 April 2024

CITATION
Wang L, Zhang R, Liu J, Xu R, Zhao L, Li E and Zhang Y (2024) An online survey on public awareness of drug clinical trials in inland cities of northern China.
Front. Public Health 12:1276536.
doi: 10.3389/fpubh.2024.1276536

An online survey on public awareness of drug clinical trials in inland cities of northern China

Li Wang[1,2†], Ruiguang Zhang[3†], Jingxuan Liu[2†], Rui Xu[1], Lili Zhao[1], Enze Li[1] and Yurong Zhang[1*]

[1]Department of Scientific Research, Faculty of General Family Medicine, the First Affiliated Hospital of Xi'an Medical University, Xi'an, Shaanxi, China, [2]Department of Health Statistics, Faculty of Public Health, Shaanxi University of Chinese Medicine, Xi'an, Shannxi, China, [3]Xi'an Evidence Based Pharmaceutical Technology Co., Ltd, Xi'an, China

Objective: The aims of this survey were to investigate the public awareness of drug clinical trials (DCTs) and willingness to participate the DCTs, and provide references for propaganda and science popularization of DCTs.

Methods: A self-designed questionnaire named "an online survey questionnaire on public awareness of DCTs" was used to conduct an online survey from January to March 2022. The demographic characteristics and the response of participants to the awareness and willingness to participate the DCTs were collected. The factors affecting the public awareness of DCTs were analyzed by single factor and binary logistic regression analysis.

Results: One thousand three hundred eighty valid questionnaires were collected, and the respondents' awareness rate of DCTs was 61.1%. Thirteen demographic characteristics including age, gender, education, occupation, work fields, household type, marital status, city type, income, medical insurance, medical expenditure, pressure to seek medical care, financial pressure, both significantly affected the qualified rate of participants' awareness of DCTs ($p<0.001$) by single factor analysis. Binary logistic regression analysis indicated that education level, work fields, city type, medical insurance, and medical expenditure affected independently the participants' awareness rate of DCTs ($p<0.001$). 52.9% of the participants were willing to take part in DCTs. "to promote medical progress" (54.4%) or "believe doctors" (31.1%) were the most frequent reasons for subjects participating in DCTs.

Conclusion: The public awareness rate of DCTs and the willingness to participate in drug clinical were significantly affected by the demographic characteristics of subjects. Thus, targeting the needs of the public, propaganda, and science popularization of DCTs should be carried out and served public health.

KEYWORDS

drug clinical trials, public, awareness of drug clinical trials, willingness to participate the drug clinical trials, in northern China

1 Introduction

With the advancement of the economy, living standards, and national cultural in China, there has been a discernible shift towards a more scientifically informed public awareness regarding clinical trials. Consequently, an increasing number of pharmaceutical enterprises are directing significant attention towards harnessing clinical research resources within China.

Presently, China has emerged as a pivotal hub for clinical trials, owing to its vast population scale, abundant medical infrastructure, and comparatively lower research expenditure (1). Encumbered by the high costs associated with new drug research and limited clinical resources elsewhere, foreign pharmaceutical enterprises are progressively relocating their research centers to China.

The 14th Five-Year Plan has introduced a range of policies, notably including "Opinions on Reforming the Review and Approval System of Drugs and Medical Devices",[1] aimed at overhauling the regulatory framework governing drugs and medical devices while fostering biomedical innovation. Consequently, the research and development capabilities pertaining to innovative drugs, vaccines, and high-end medical devices have experienced consistent augmentation. This has led to a notable increase in the number of new drugs and medical devices advancing into clinical stages. Consequently, the demand for clinical trial resources, such as subjects, has witnessed a substantial annual upsurge. The quantity of registered clinical trials in China approached nearly 10,000 (2), yielding an impressive output value of 700 billion US dollars from 2016 to 2020. According to data from Informa Pharma Intelligence, the average annual growth rate for Phase I to Phase IV clinical trials stood at 20% (3). These data underscore the heightened investment in funds for new drug research and development and signify the escalating health consciousness within China.

According to statistics, approximately 500,000 individuals participate to the DCTs in China annually, this figure that remains relatively small compared to the nation's vast population base. However, the emergence of the COVID-19 pandemic (4) has reignited public interest in keywords such as "drug/vaccine clinical trials," "clinical trials," and "subjects." Clinical trials serve as crucial avenues for offering additional treatment options, particularly for patients lacking access to effective remedies (5, 6), thereby presenting an opportunity for disease management and potential cure. Nonetheless, obstacles to subject participation in clinical trials consistently result in recruitment failure (7). Hence, it is crucial to explore public awareness of DCTs for subject recruitment. Unfortunately, a thorough investigation into the public awareness of DCTs in China has yet to be undertaken. Thus, the present study aims to fill this gap by investigating public awareness of DCTs and identifying key factors influencing this awareness through the design and implementation of a questionnaire. The findings seek to furnish a theoretical foundation for enhancing public engagement in clinical trials.

2 Methods

2.1 Study population

This investigation constitutes a cross-sectional survey which employed non-random convenience and snowball sampling methodologies. From January to March 2022, healthy adults aged 18 and above, possessing autonomy, adept in independent completion of electronic questionnaires, and exhibiting high cooperative tendencies, were recruited for participation. It is important to note that all participants voluntarily enrolled in this study. Excluding individuals with mental disorders, cognitive impairment, and inability to independently complete all items in the survey questionnaire.

2.2 Recruitment and enrollment

Adhering to the principles delineated by Professor Yan Yan in "Medical Statistics" (8), the sample size was determined to be 5–10 times the count of independent variables, hence yielding a calculated sample volume of 420 individuals. To address potential errors and bolster questionnaire effectiveness, the sample size was augmented by 10%, culminating in a confirmed cohort of 470 individuals.

The survey questionnaire was formulated and exported in the form of a two-dimensional code from the "Jinshuju" online data platform. Before the formal survey commenced, 20 preliminary questionnaires were disseminated, revealing incomplete responses. Measures were implemented to constrain each IP address to a single response and to mandate the completion of all items, reinforcing the scientificity and preciseness of the survey. Healthy individuals over 18 years of age from five provinces in the northern region of China were chosen as the study cohort. A liaison officer was designated to each province to disseminate information online within the community, including the objective and significance of survey, instructions for completing the questionnaire and notes on filling out the questionnaires. WeChat groups were established in each community for the distribution and retrieval of questionnaires. These questionnaires were scanned and completed using the two-dimensional code, enabling online data automatic collection.

Participants were tasked with anonymously addressing 42 survey inquiries, with response validity hinging on completeness and the absence of missing elements. Subsequent to an elucidation of the research objectives, questionnaire completion protocols, and principles of confidentiality, the electronic questionnaire was shared with respondents following the procurement of their informed consent. Respondents were encouraged to independently progress through the questionnaire, receiving standardized instructions and annotations for item completion. The formulation of the questionnaire deliberately avoided suggestive prompts. Data in Excel format was exported from "Jinshuju" data platform and underwent reconciliation and cleaning by the researcher, including logical error checking to enhance the accuracy of source material, ultimately facilitating the identification of valid surveys while excluding those with incomplete or insincere responses. A total of 1,462 surveys were collected in, with 82 deemed invalid, resulting in an effective response rate of 94.4%.

2.3 Development and validation of the survey questionnaire

This research initiative assembled a team of 5 investigators dedicated to questionnaire development, comprising 2 individuals with senior professional designations and 3 with intermediate professional affiliations. Their primary duties encompassed literature review and analysis, determination of questionnaire items, data collection and statistical analysis, questionnaire refinement,

1 https://www.nmpa.gov.cn/ylqx/ylqxjgdt/20150818200801163.html

assessment of questionnaire reliability and validity, and ultimately, questionnaire finalization.

Employing English search terms such as "drug/vaccine clinical trials," "clinical trials," "subjects," "cognitively," "willingness & drug clinical trials," "attitude & drug clinical trials" and "awareness," comprehensive literature searches were conducted across prominent databases including PubMed, Embase, The Cochrane Library, Biomedical Literature Databases, CNKI, and Wanfang Data Knowledge Service Platform. The exploration encompassed subject heading terms from the inception of the databases until December 2021. Following the removal of duplicate literature, the remaining titles, abstracts, and full texts underwent sequential review. Encompassed study types included clinical practice guidelines, systematic reviews, clinical randomized controlled trials, cohort studies, observational studies, expert opinions, and other research. Exclusion criteria comprised duplicated publications or translated literature, incomplete data, and inaccessible full texts. Drawing from the outcomes of the literature review and analysis, the metrics of the questionnaire were defined, and the questionnaire items were initially formulated.

Following collaborative discussion, the initial iteration of the Online Survey Questionnaire on Public Awareness of DCTs was formulated, encompassing two sections. The first section covers general demographic information, consisting of 13 entries such as age, gender, education, occupation, work fields, household type, marital status, city type, income, medical insurance, medical expenditure, pressure to seek medical care, and financial pressure. The second section entails the public awareness survey of DCTs, incorporating two dimensions: public awareness of DCTs and willingness to participate the DCTs. Within the Public Awareness Survey of DCTs, 20 items were related to cognitive assessment of DCTs, encompassing aspects such as new drug clinical trial procedures, allocation methods, compensation for damages to subjects, free trial drugs and related physical examinations, ethics and more. Each item was evaluated on a 3-point scale: "yes," "no," "do not know" correlating to respective scores of 2, 1, and 0. A total of 20 items need to score according to the participants' options of clinical trial-related questions in the questionnaires. Participants who achieved a score exceeding 24 points were deemed to exhibit a proficient understanding of DCTs. Each item on the scale was positively evaluated, with higher aggregate scores indicative of an elevated comprehension of DCTs. Moreover, the Public Willingness Survey to Participate the DCTs comprised 9 items addressing the inclination to partake in clinical trials, factors motivating participation, and related considerations. The outcome variables focused on public awareness of DCTs.

2.4 Assessment of the Questionnaire's reliability and validity

Reliability was assessed through the utilization of Cronbach's alpha coefficient, while validity underwent evaluation via Bartlett's sphericity test. A significance level of $p < 0.05$ was stipulated for statistical significance. Generally, a Cronbach's α coefficient exceeding 0.700 is deemed acceptable (9), with a value falling within the range of 0.800 to 0.900 indicating an outstanding level of internal consistency reliability (10), thus meeting the benchmark for a foundational research instrument. The questionnaire was subjected to reliability and validity testing using SPSS 26.0, resulting in a total Cronbach's α coefficient of 0.973 and a KMO coefficient of 0.975, both approximating 1. Additionally, the Bartlett's sphericity test produced an approximate χ^2 value of 28933.950, with a $p < 0.001$, aligning with the intended objectives. These outcomes signify robust reliability and validity within the questionnaire, rendering it suitable for implementation.

2.5 Quality control

The survey questionnaire is distributed online, and all respondents are willing to fill it out voluntarily. Measures are taken to ensure that all items in the questionnaire are completed before submission. Therefore, the collected questionnaires are those with all items completed. Through repeated distribution and snowball like online forwarding, extensive and multiple online education sessions have been conducted, expanding the scope of dissemination and increasing the number of participants. Through these methods, participant compliance is ensured, participant acceptance is improved, and sample size requirements are also met. Additionally, it is essential to select appropriate measurement tools and assessment methodologies, establish uniform standards and guidelines for researchers, and conduct thorough data review and verification to enhance data quality.

2.6 Statistical analysis

Utilizing SPSS 26.0 for data analysis, categorical data description entails the use of frequency and constituent ratio, Measures obeying normal distribution were analyzed by t-test and F-test for hypothesis testing of their overall means, whereas ($\overline{x} \pm s$) is utilized for describing continuous data. While measures not obeying normal distribution were analyzed by non-parametric tests. The χ^2 test is applied to compare discrepancies in the awareness of DCTs across demographics. Logistic regression analysis is employed to gauge the influence of demographic variables on public awareness of DCTs. Noteworthy demographic variables are singled out through univariate analysis and integrated into the ultimate logistic regression model. For the variables encompassed in the regression model, their odds ratios (OR) and 95% confidence intervals are computed. A significance threshold of $p < 0.05$ denotes statistical significance.

3 Results

3.1 Demographic characteristics statistics

A total of 1,462 questionnaires were collected in the survey, among them, 82 questionnaires with incomplete information or obvious filling errors were removed and 1,380 valid questionnaires (94.4%) were used to analyze the study.

This study comprised 1,380 valid questionnaires, with 1,380 distributed across 5 provinces and municipalities in northern China. Specifically, there were 328 from Shaanxi, 345 from Shanxi, 259 from

Gansu, 247 from Ningxia Hui Autonomous Region, and 201 from Henan Province. The survey respondents from Shaanxi and Shanxi were almost occupied 50%, all with a relatively even distribution of the population in each province and a good representation of the sample.

The study included 562 male participants (40.7%) and 818 female participants (59.3%). Individuals under the age of 40 constituted 81.23% of the total respondents. Additionally, 47.9% of the participants had attained a bachelor's degree or higher level of education. Among the participants, the most prevalent occupations were students (43.5%), followed by individuals employed in public institutions (19.6%), and those working in enterprises (12.7%). Regarding household registration type, 56.6% of participants hailed from urban areas. Approximately 43.8% of the participants were married, and 47.3% reported a monthly income level ranging from 0 to 2000 CNY. In terms of medical insurance, the majority of participants were enrolled in medical insurance for urban and rural residents (36.8%), followed by urban employee medical insurance (31.5%), and self-funded medical insurance (7.7%). Furthermore, 38.2% of respondents stated that their medical expenditure accounted for less than half of their monthly income. Notably, over half of the participants experienced varying degrees of medical (56.4%) and financial pressure (67.9%). The demographic characteristics are summarized in Table 1.

3.2 Awareness rate of DCTs related knowledge for participants

To gauge public awareness of DCTs, we computed the qualified awareness rate among participants (refer to Table 2). Out of 1,380 participants, the qualified awareness rate regarding DCTs stood at 61.1%. Notably, 70.6% of participants affirmed that they had heard of DCTs, while 18.6% stated they were aware but not familiar with them. It is evident that individuals employed in medical-related fields exhibited a notably higher awareness rate compared to those in non-medical related occupations. Awareness rates exhibited variability across different demographic characteristics, spanning from 47.4 to 70.6%. Notably, awareness concerning compensation for injury events arising during clinical trials was notably low, registering at 47.4 and 48.9%, respectively. Additionally, only 59.1% of participants demonstrated an awareness of the basic definition of DCTs.

3.3 The influence of demographic characteristics on the qualified rate of awareness about DCTs by a single factor analysis

To explore the factors influencing the qualified awareness rate of DCTs, we conducted a single-factor analysis to compare differences across various demographic categories. The findings revealed that demographic characteristics such as gender, age, education, occupation, Work fields, household type, marital status, city type, income, medical insurance, medical expenditure, pressure to seek medical care, and financial pressure all significantly impacted the qualified awareness rate of participants, regarding DCTs (Table 2, $p < 0.001$).

3.4 Binary logistic regression analysis of influencing the participants' awareness of DCTs

To mitigate the influence of interactions between factors on participants' awareness rate of DCTs, binary logistic regression analysis was employed to evaluate the independent effects of each demographic characteristic. In this analysis, demographic characteristics served as independent variables, while qualification status (qualified = 1, unqualified = 0) was the dependent variable. As presented in Table 3, education, field of work, city type, medical insurance, and medical expenditure were identified as independent factors significantly impacting participants' awareness rate of DCTs ($p < 0.05$).

Further analysis revealed significant associations between specific demographic characteristics and the awareness rate of DCTs. Specifically, higher education levels were linked to elevated awareness rate. For instance, participants with postgraduate degree exhibited a 4.905-folds increase in awareness rate compared to those with under junior college education (95% CI: 2.405 ~ 10.002). Moreover, individuals engaged in medical-related occupations demonstrated a substantially higher awareness rate, with a 3.217-folds increase compared to those in non-medical related fields (95% CI: 2.287–4.527). City type also exerted a notable influence on participants' awareness of DCTs. Notably, individuals residing in second-tier cities exhibited a mere 0.612-fold awareness rate compared to those in first-tier cities (95% CI: 0.404–0.928). Additionally, participants covered by medical insurance for urban and rural residents demonstrated a 1.8268-folds increase in awareness rate compared to the general population covered by public medical insurance (95%CI:1.077 ~ 3.093). Furthermore, heightened medical expenditures were associated with enhanced awareness rate of DCTs. Specifically, participants allocating more than 50% of their monthly income to medical expenditure exhibited a 2.008-folds increase in awareness rate compared to those with no medical expenditures (95% CI: 1.019–3.956).

3.5 Analysis of willingness to participate in DCTs

In the survey, facilitators and barriers for participants considering participation in DCTs were analyzed. Among the participants, 52.9% expressed willingness to partake in such trials. The primary motivations cited included a desire to "promote medical progress" (54.4%) or a belief in the guidance of doctors (31.1%). Conversely, 47.1% of participants declined participation in DCTs. The foremost reasons for refusal included concerns regarding the risks associated with the trial (62.1%), lack of awareness regarding DCTs (54.6%), reluctance to be treated akin to laboratory mice (32.8%), and unwillingness to allocate additional time and effort (24.3%).

4 Discussion

Clinical trials play a crucial role in evaluating the efficacy and safety of new drugs prior to their introduction to the market. The success of clinical trials hinges largely on the recruitment of an adequate number of subjects (11). However, several obstacles hinder

10.3389/fpubh.2024.1276536

TABLE 1 Demographic characteristics statistics.

Item	Class	*n*	%
Age	18–20 years old	484	35.07
	21–30 years old	336	24.35
	31–40 years old	301	21.81
	41-60 years old	225	16.30
	> 60 years old	34	2.46
Gender	Male	562	40.72
	Female	818	59.28
Education	Under junior college	481	34.86
	Junior college	239	17.32
	Undergraduate	470	34.06
	Postgraduate	190	13.77
Occupation	Administration organization	38	2.75
	Public institution	271	19.64
	Enterprise	175	12.68
	Student	600	43.48
	Farmer	99	7.17
	Retired	31	2.25
	Other	166	12.03
Work fields	Non-medical related	737	53.41
	Medical-related	643	46.59
Household type	Urban	599	43.41
	Rural	781	56.59
Marital status	Unmarried	775	56.16
	Married	605	43.84
City type	First-tier city	300	21.74
	Second-tier city	323	23.41
	Third-tier city	246	17.83
	Fourth-tier city and below	511	37.03
Income	0 ~ 2000 CNY	653	47.32
	2001 ~ 3,000 CNY	111	8.04
	3,001 ~ 5,000 CNY	232	16.81
	5,001 ~ 8,000 CNY	227	16.45
	> 8,000 CNY	157	11.38
Medical insurance	Public medical care	124	8.99
	Urban employee medical insurance	435	31.52
	Urban and rural residents medical insurance	508	36.81
	Out-of-pocket medical expenses	106	7.68
	Private/commercial insurance	20	1.45
	Other	187	13.55
Medical expenditure	Negligible	793	57.46
	< 50% of monthly income	527	38.19
	≥50% of monthly income	60	4.35
Pressure to seek medical care	None	602	43.62
	Mild	475	34.42
	Moderate	220	15.94
	Severe	83	6.01
Financial pressure	None	443	32.10
	Mild	459	33.26
	Moderate	317	22.97
	Severe	161	11.67

the subject recruitment process, including patient preferences, concerns stemming from uncertainty, and apprehensions regarding information and consent procedures, among others (12). Hence, it is imperative to assess the public's awareness of clinical trials and their willingness to participate in such endeavors. This survey scrutinized the public's awareness of DCTs, their propensity for participation, and the factors influencing awareness in China.

A comprehensive questionnaire comprising 42 inquiries was developed to elicit respondents' demographic characteristics, awareness of DCTs, and willingness to participate in such trials in China. Thirteen demographic characteristics were evaluated, including age, gender, education, occupation, field of work, household type, marital status, city type, income, medical insurance, medical expenditure, pressure to seek medical care, and financial pressure. This extensive range of demographic factors surpasses those reported in previous studies (13, 14). This approach allowed for the meticulous capture of a broad range of factors influencing public awareness of DCTs, providing deeper insight into participant attitudes and perceptions.

The 1,380 respondents were widely distributed across several cities in five northern Chinese provinces, with a significant concentration in Shaanxi and Shanxi, comprising 328 and 345 individuals, respectively. The population exhibited a more even distribution across provinces, thus offering a representative sample. In terms of age distribution, participants were evenly spread across four age brackets spanning 18 to 60 years, including 484 individuals aged 18–20, 336 aged 21–30, 301 aged 31–40, and 225 aged 41–60. There were fewer individuals above the age of 60, possibly attributed to the online nature of the survey and the extensive 42-item questionnaire, which may have deterred older adult participation. Challenges in engaging with new media platforms (e.g., WeChat and TikTok) and smartphone technology may have contributed to decreased participation among the elderly due to lower usage frequency. Additionally, students, being more interested in novel experiences such as DCTs, exhibiting relatively high compliance and ease in completing questionnaires, represented a significant proportion of the survey at 43.5%.

The findings revealed that 61.1% of subjects exhibited a qualified awareness rate of drug clinical trials-related knowledge in China. This percentage closely aligns with a survey conducted in Saudi Arabia, where 58% of cancer patients demonstrated awareness of clinical

10.3389/fpubh.2024.1276536

TABLE 2 Factors affecting the qualified rate of awareness of DCTs.

Item	Class	Unqualified		Qualified		χ^2 test	
		n	%	*n*	%	χ^2	*p*
Age	18–20 years old	267	55.2%	217	44.8%	112.338***	0.000
	21–30 years old	69	20.5%	267	79.5%		
	31–40 years old	99	32.9%	202	67.1%		
	41-60 years old	73	32.4%	152	67.6%		
	> 60 years old	17	50.0%	17	50.0%		
Gender	Male	250	44.5%	312	55.5%	16.685***	0.000
	Female	275	33.6%	543	66.4%		
Education	Under junior college	325	67.6%	156	32.4%	305.686***	0.000
	Junior college	85	35.6%	154	64.4%		
	Undergraduate	98	20.9%	372	79.1%		
	Postgraduate	17	8.9%	173	91.1%		
Occupation	Administration organization	8	21.1%	30	78.9%	201.261***	0.000
	Public institution	41	15.1%	230	84.9%		
	Enterprise	37	21.1%	138	78.9%		
	Student	289	48.2%	311	51.8%		
	Farmer	68	68.7%	31	31.3%		
	Retired	21	67.7%	10	32.3%		
	Other	61	36.7%	105	63.3%		
Work fields	Non-medical related	420	57.0%	317	43.0%	240.839***	0.000
	Medical-related	105	16.3%	538	83.7%		
Household type	Urban	148	24.7%	451	75.3%	79.857***	0.000
	Rural	377	48.3%	404	51.7%		
Marital status	Unmarried	328	42.3%	447	57.7%	13.733***	0.000
	Married	197	32.6%	408	67.4%		
City type	First-tier city	94	31.3%	206	68.7%	103.177***	0.000
	Second-tier city	61	18.9%	262	81.1%		
	Third-tier city	101	41.1%	145	58.9%		
	Fourth-tier city and below	269	52.6%	242	47.4%		
Income	0 ~ 2000 CNY	323	49.5%	330	50.5%	130.555***	0.000
	2001 ~ 3,000 CNY	59	53.2%	52	46.8%		
	3,001 ~ 5,000 CNY	82	35.3%	150	64.7%		
	5,001 ~ 8,000 CNY	38	16.7%	189	83.3%		
	> 8,000 CNY	23	14.6%	134	85.4%		
Medical insurance	Public medical care	42	36.9%	82	63.1%	116.329***	0.000
	Urban employee medical insurance	82	18.9%	353	81.1%		
	Urban and rural residents medical insurance	239	47.0%	269	53.0%		
	Out-of-pocket medical expenses	48	45.3%	58	54.7%		
	Private/commercial insurance	8	40.0%	12	60.0%		
	Other	106	56.7%	81	43.3%		

(Continued)

TABLE 2 (Continued)

Item	Class	Unqualified		Qualified		χ^2 test	
		n	%	*n*	%	χ^2	*p*
Medical expenditure	Negligible	308	38.8%	485	61.2%	22.782***	0.000
	< 50% of monthly income	178	33.8%	349	66.2%		
	≥50% of monthly income	39	65.0%	21	35.0%		
Pressure to seek medical care	None	243	40.4%	359	59.6%	23.537***	0.000
	Mild	146	30.7%	329	69.3%		
	Moderate	90	40.9%	130	59.1%		
	Severe	46	55.4%	37	44.6%		
Financial pressure	None	195	44.0%	248	56.0%	25.782**	0.000
	Mild	141	30.7%	318	69.3%		
	Moderate	111	35.0%	206	65.0%		
	Severe	78	48.4%	83	51.6%		

trials, indicating a similar level of awareness between the two studies (15). Moreover, only 3.37% of subjects were familiar with the rights and interests afforded to participants in DCTs, while 58% were knowledgeable about ethical committees and their functions in such trials. Although 70.6% of participants had heard of DCTs, a mere 16.2% had actually participated in them to a limited extent. Additionally, 69.2% of participants were aware of the importance of protecting the privacy and personal information during DCTs, underscoring the significance attached to privacy by the public. However, awareness regarding subjects' rights, such as the entitlement to receive clinical trial drugs free of charge and the ability to withdraw from the trial at any time for any reason, was relatively low at 57.5 and 50.5%, respectively. This suggests a limited understanding among the public regarding their rights to participate in DCTs and the evolving trend toward patient-centered DCTs.

In recent years, there has been a notable increase in the approval and implementation of DCTs in China (16), leading to a gradual shift in focus toward China in international multi-center trials. The examination of factors influencing awareness rates regarding DCTs in this study encompassed both univariate and binary logistic regression analyses. Univariate analysis revealed that all demographic characteristics collected significantly impacted participants' awareness rates of DCTs. However, binary logistic regression analysis highlighted that only education level, field of work, city type, medical insurance, and medical expenditure exerted significant effects on participants' awareness rate of DCTs. These findings suggest that education level, field of work, city type, medical insurance, and medical expenditure are the primary factors influencing public awareness of DCTs in China. This aligns with findings by Primo et al., who identified income and education level as factors related to awareness of cancer clinical trials (17). Similarly, a survey conducted in Korea identified age, religion, financial level, and education level as significant determinants of public awareness of cancer clinical trials (18). These collective results underscore the crucial role of education level as an influential factor in shaping public awareness of DCTs on a global scale. Individuals with higher education levels are likely to have greater exposure to media, internet resources, and other forms of learning, facilitating a deeper understanding of DCTs and related knowledge.

In our survey, 52.9% of participants expressed a willingness to partake in clinical trials, while 50.6% stated that they would recommend their relatives and friends to participate as well. However, this relatively low motivation to engage in clinical trials among the public may be indicative of insufficient awareness and a general distrust of current clinical trial practices. A national survey study reported that only 25% of respondents were willing to participate in clinical trials (19), a markedly lower percentage compared to our findings. This discrepancy could potentially be attributed to variations in the demographic characteristics of the participants involved in the respective surveys.

The pressure of undergoing medical treatment emerged as a significant factor influencing respondents' willingness to participate the DCTs. Participants tended to exercise caution in selecting treatment methods when they felt they had some control over their medical situation. However, when the pressure of medical treatment exceeded a certain threshold, individuals became more inclined to explore new treatment options. Consequently, it is essential to tailor communication approaches to individuals based on their level of pressure to seek medical care. For those experiencing mild pressure to seek medical care, recruiters should employ strategies aimed at alleviating their concerns through psychological counseling, thereby encouraging voluntary participation in clinical trials. Conversely, individuals facing high levels of pressure to seek medical care require a more nuanced approach. Before deciding whether to participate in clinical trials, it is crucial to emphasize the potential advantages and disadvantages of such involvement. This approach ensures that individuals fully understand the implications of their decision within the context of their medical circumstances.

Moreover, medical expenditure demonstrated a significant correlation with both participants' awareness rates and their willingness to participate in DCTs. Individuals with high medical expenditures often undergo frequent medical treatments for various illnesses. Consequently, they tend to possess a better understanding of hospital procedures and have more opportunities to encounter ongoing DCTs, thereby increasing their awareness of drug clinical trial-related knowledge and their willingness to participate. It's worth noting that DCTs typically offer drugs and relevant examinations free

10.3389/fpubh.2024.1276536

TABLE 3 Factors influenced the participants' awareness rate of DCTs by logistic regression analysis.

Variables	Exp (95% CI)	*p*
Age		
18–20 years old	1.000	
21 ~ 30 years old	0.587 (0.208 ~ 1.660)	0.315
31–40 years old	0.933 (0.339 ~ 2.565)	0.893
41–60 years old	0.590 (0.230 ~ 1.517)	0.274
> 60 years old	0.994 (0.396 ~ 2.497)	0.991
Gender		
Male	1.000	
Female	1.100 (0.835 ~ 1.451)	0.497
Education		
Under junior college	1.000	
Junior college	1.719 (1.130 ~ 2.616)	0.011
Undergraduate	2.759 (1.792 ~ 4.247)	0.0011
Postgraduate	4.905 (2.405 ~ 10.002)	0.001
Occupation		
Administration organization	1.000	
Public institution	1.323 (0.314 ~ 1.057)	0.577
Enterprise	1.119 (0.314 ~ 1.057)	0.701
Student	1.552 (0.314 ~ 1.057)	0.127
Farmer	0.542 (0.314 ~ 1.057)	0.067
Retired	0.550 (0.314 ~ 1.057)	0.083
Others	0.400 (0.314 ~ 1.057)	0.080
Work fields		
Non-medical related	1.000	
Medical-related	3.217 (2.287 ~ 4.527)	0.000
Household type		
Urban	1.000	
Rural	1.002 (0.711 ~ 1.410)	0.993
Marital status		
Unmarried	1.000	
Married	0.801 (0.477 ~ 1.343)	0.400
City type		
First-tier city	1.000	
Second-tier city	0.612 (0.404 ~ 0.928)	0.021
Third-tier city	0.1.140 (0.745 ~ 1.746)	0.545
Fourth-tier city and below	1.015 (0.704 ~ 1.465)	0.936
Income		
0 ~ 2000 CNY	1.000	
2001 ~ 3,000 CNY	0.829 (0.394 ~ 1.741)	0.620
3,001 ~ 5,000 CNY	0.613 (0.288 ~ 1.306)	0.205
5,001 ~ 8,000 CNY	0.528 (0.278 ~ 1.003)	0.051
> 8,000 CNY	1.061 (0.546 ~ 2.062)	0.862
Medical insurance		
Public medical care	1.000	
Urban employee medical insurance	1.687 (0.948 ~ 3.003)	0.076
Urban and rural residents medical insurance	1.826 (1.077 ~ 3.093)	0.025
Out-of-pocket medical expenses	1.382 (0.927 ~ 2.060)	1.112
Private/commercial insurance	1.505 (0.859 ~ 2.636)	1.153
Other	0.971 (0.329 ~ 2.867)	0.958
Medical expenditure		
None	1.000	
< 50% of monthly income	1.758 (0.890 ~ 3.472)	1.104
≥50% of monthly income	2.008 (1.019 ~ 3.956)	0.044
Pressure to seek medical care		
None	1.000	
Mild	1.358 (0.684 ~ 2.805)	0.366
Moderate	1.414 (0.583 ~ 2.235)	0.700
Severe	0.815 (0.413 ~ 1.608)	0.555
Financial pressure		
None	1.000	
Mild	0.980 (0.555 ~ 1.732)	0.946
Moderate	1.113 (0.658 ~ 1.881)	0.690
Severe	1.185 (0.708 ~ 1.983)	0.518
Constants	2.033	0.364

of charge, thereby reducing the overall treatment costs for patients. Consequently, individuals with lower incomes may exhibit a higher willingness to participate in DCTs compared to those with higher incomes. This is because participation in such trials presents an opportunity to access treatment at no cost, potentially alleviating the financial burden associated with medical care.

Some limitations existed in the survey. This study employed convenience sampling and snowball sampling techniques, resulting in a restricted representativeness and broad diversity in terms of occupation and age within the sampled population, inducing a measure of bias in the research outcomes. WeChat groups were established in each community for questionnaire distribution and retrieval, facilitated by the use of two-dimensional code for online data collection. For this reason, there are significant limitations regarding external validity, as the sample population may not be representative of the entire Chinese population. Moreover, calculating the survey response rate is unfeasible, and assessing internal validity becomes challenging.

The variables "pressure to seek medical care" and "financial pressure" are moderately subjective and liable to bias. These variables were incorporated to evaluate the impact of "pressure to seek medical care" and "financial pressure" on clinical trial participation. Patients encountering heightened "pressure to seek

medical care" and "financial pressure" manifest a higher inclination to partake in clinical trials and are also more prone to reap benefits from their participation.

The limited sample size of this study may not fully represent the broader situation in China. Therefore, future research should consider employing hierarchical randomization methods for demographic factors, expanding the sample size, and incorporating more representative questions related to DCTs.

In summary, our findings provide valuable insights into the public's awareness of DCTs. Strengthening effective publicity efforts regarding drug clinical trials-related knowledge in the future is crucial. It is evident that the public's willingness to participate in clinical trials in China is relatively low and significantly influenced by their professional backgrounds and relevant experiences. Therefore, it is imperative to conduct scientific outreach initiatives tailored to the public's needs. This approach will not only enhance public awareness but also foster a greater willingness to participate in clinical trials. Ultimately, such efforts will contribute to the smooth progression of DCTs, promote advancements in medical care, and serve to improve public health overall.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study of human participants in accordance with the local legislation and institutional requirements. Written informed consent from the patients/participants was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

LW: Writing – review & editing, Supervision. RZ: Writing – original draft. JL: Data curation, Formal analysis, Writing – review & editing. RX: Writing – original draft. LZ: Writing – original draft. EL: Writing – original draft. YZ: Writing – original draft, Data curation.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Education Department Service Local Special Program Project of Shaanxi Province of China (20JC031), the Supporting fund project of the First Affiliated Hospital of Xi'an Medical University of China (XYFYPT-2021-02), the Science and Technology Innovation Team Project of Xi'an Medical University (2021TD14) and the Education and Teaching Reform Research Project of Xi'an Medical University (2022JG-67), Key Research & Development Program Projects of Shaanxi Province, China (2023-YBSF-443).

Acknowledgments

The authors thank all subjects who participate in the survey.

Conflict of interest

RZ was employed by Xi'an Evidence Based Pharmaceutical Technology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2024.1276536/full#supplementary-material

References

1. Cao Y, Liao L, Liu X, Zheng Q, Xu Z, Niu H. Trend of DCTs in mainland China from 2009 to 2020. *Curr Med Res Opin*. (2022) 38:1499–07. doi: 10.1080/03007995.2022.2103960

2. Zhang L, Huang XX, Chen HF. An exploration of the protective effects of Investigators' ethical awareness upon subjects of DCTs in China. *J Bioeth Inq*. (2018) 15:89–100. doi: 10.1007/s11673-017-9826-5

3. Annie Siu. China clinical trials landscape. Available at: https://pharmaintelligence.informa.com/resources/product-content/clinical-trials-landscape, Pharma Intelligence-Informa

4. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA*. (2020) 323:1061–9. doi: 10.1001/jama.2020.1585

5. Limkakeng A, Phadtare A, Shah J, Vaghasia M, Wei DY, Shah A, et al. Willingness to participate in clinical trials among patients of Chinese heritage: a meta-synthesis. *PLoS One*. (2013) 8:e51328. doi: 10.1371/journal.pone.0051328

6. Bernstein SL, Feldman J. Incentives to participate in clinical trials: practical and ethical considerations. *Am J Emerg Med*. (2015) 33:1197–00. doi: 10.1016/j.ajem.2015.05.020

7. Rodríguez-Torres E, González-Pérez MM, Díaz-Pérez C. Barriers and facilitators to the participation of subjects in clinical trials: an overview of reviews. *Contemp Clin Trials Commun*. (2021) 23:100829–46. doi: 10.1016/j.conctc.2021.100829

8. Yan Y, Tong W. *Medical statistics*. Beijing: People's Health Publishing House (2020).

9. Zheng RC. *Psychological measurement and testing*. Beijing: People's University of China Press (2008).

10. Su H, Li S, Cheng Y. Evaluation of internal consistency of scale assessment and application of Cronbach's alpha coefficient. *J Clin Psychos Dis*. (2009) 15:85–6. doi: 10.3969/j.issn.1672-187X.2009.01.047.0085-02

11. Ross MW, Jeffords K, Gold J. Reasons for entry into and understanding of HIV/AIDS clinical trials: a preliminary study. *AIDS Care*. (1994) 6:77–82. doi: 10.1080/09540129408258027

12. Farrell B, Kenyon S, Shakur H. Managing clinical trials. *Trials*. (2010) 11:78–83. doi: 10.1186/1745-6215-11-78

13. Ha TV, Nguyen AMT, Nguyen HST. Public awareness about antibiotic use and resistance among residents in highland areas of Vietnam. *Biomed Res Int*. (2019) 2019:1–8. doi: 10.1155/2019/9398536

14. Moorcraft SY, Marriott C, Peckitt C, Cunningham D, Chau I, Starling N, et al. Patients' willingness to participate in clinical trials and their views on aspects of cancer research: results of a prospective patient survey. *Trials*. (2016) 17:17–28. doi: 10.1186/s13063-015-1105-3

15. Bazarbashi S, Hassan A, Eldin AM, Soudy H, Hussain F. Awareness and perceptions of clinical trials in Cancer patients and their families in Saudi Arabia. *J Cancer Educ*. (2015) 30:655–9. doi: 10.1007/s13187-015-0797-0

16. Xu Y, Dong M, Liu X. Characteristics and trends of clinical studies primarily sponsored by China in WHO primary registries between 2009 and 2018: a cross-sectional survey. *BMJ Open*. (2020) 10:e037262. doi: 10.1136/bmjopen-2020-037262

17. Lara PN, Paterniti DA, Chiechi C, Turrell C, Morain C, Horan N, et al. Evaluation of factors affecting awareness of and willingness to participate in cancer clinical trials. *J Clin Oncol*. (2005) 23:9282–9. doi: 10.1200/JCO.2005.02.6245

18. Kim JW, Kim SJ, Chung YH, Kwon JH, Lee HJ, Chung YJ, et al. Cancer patients' awareness of clinical trials, perceptions on the benefit and willingness to participate: Korean perspectives. *Br J Cancer*. (2008) 99:1593–9. doi: 10.1038/sj.bjc.6604750

19. Chu SH, Kim EJ, Jeong SH, Park GL. Factors associated with willingness to participate in clinical trials: a nationwide survey study. *BMC Public Health*. (2015) 15:10–7. doi: 10.1186/s12889-014-1339-0

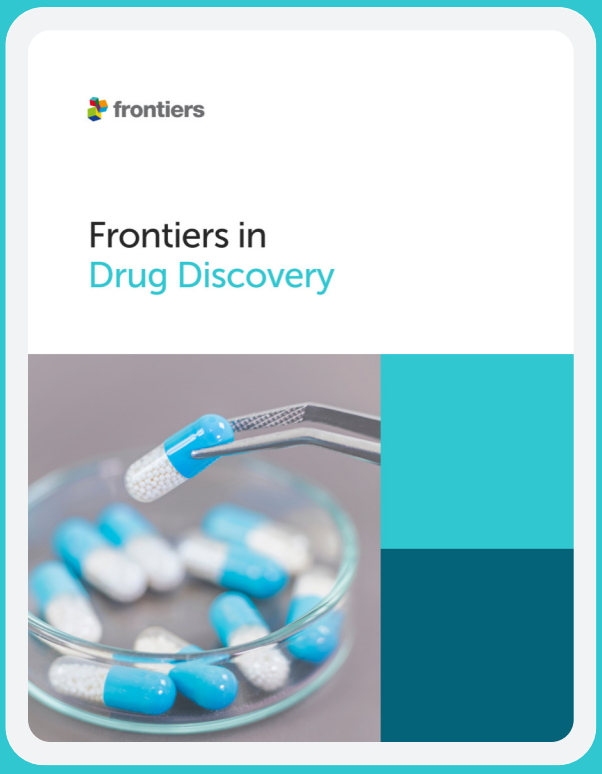
Frontiers in
Drug Discovery

Explores the discovery and development of new drugs and therapies

A multidisciplinary journal that focuses on drug discovery and the development lifecycle, from the identification of drug entities to the design and development of new candidate medications and therapies.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact