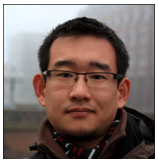
ELSEVIER

Teaser It is high time to integrate computational models at multiple levels to understand drug mechanism and safety.

Multiscale modelling of drug mechanism and safety

Jitao David Zhang[1,2], Lisa Sach-Peltason[1], Christian Kramer[1], Ken Wang[1] and Martin Ebeling[1]

[1] Pharma Early Research and Development, Roche Innovation Center Basel, F. Hoffmann-La Roche, Grenzacherstrasse 124, 4070 Basel, Switzerland
[2] University of Basel, Department of Mathematics and Computer Science, Spiegelgasse 1, 4051 Basel, Switzerland



Jitao David Zhang was a Marie-Curie Fellow at the European Institute of Bioinformatics, undertaking PhD research in computational biology at the German Cancer Research Center, and then joining Roche in 2011. His research, often embedded in collaborations with industrial and academic partners, focuses on *in silico* aspects of integrated approaches to understanding drug mechanism and safety. David co-develops new techniques supporting discovery projects, such as molecular phenotyping, which deciphers how drug candidates modulate biological pathways. David enjoys publishing open-source software and open-access publications, working with students and postdoctoral researchers, and teaching applied mathematics and informatics in drug discovery at the University of Basel.

Here, we introduce models at three levels—molecular level, cellular and omics level, and organ and system level—that study drug mechanism and safety in preclinical drug discovery. The models differ in both their scope of study and technical details, but are all rooted in mathematical descriptions of complex biological systems, and all require informatics tools that handle large-volume, heterogeneous, and noisy data. We present principles and recent developments with examples at each level and highlight the synergy by a case study. We proffer a multiscale modelling view of drug discovery, call for a seamless flow of information in the form of models, and examine potential impacts.

``Prédire n'est pas expliquer' (To predict is not to explain) – René Thom"

Introduction

A drug discovery program ideally maximises the knowledge about a molecule before progressing to test it in humans. Understanding the mechanism, namely knowing how and why a molecule works in *in vitro* and *in vivo* disease models (or not), is desired in both target-based and phenotypic discovery programs. It not only satisfies curiosity, but also enriches our knowledge in biology, guides optimisation of the molecule, enables targeted therapies and drug combinations, and reveals unexpected disease targets [1].

Understanding drug mechanism and safety is more than target identification and describing its mechanism of action. Target identification, reviewed in [2,3], is an important step towards mechanistic understanding. Identification of the mechanism of action (MoA), the specific biochemical interaction between a drug and its target through which the substance exerts its pharmacological effect, is another step forward. However, target and MoA are not all there is to know about the mechanism of a drug for several reasons. First, proteins, RNAs, and metabolites form complex intracellular and intercellular biological networks. The drug–target interaction is propagated and amplified through the networks. Therefore, a drug can regulate genes and

Corresponding author. Zhang, J.D. (jitao_david.zhang@roche.com)

https://doi.org/10.1016/j.drudis.2019.12.009

pathways in a cell identity- and context-specific manner [4]. Some effects are only secondary to MoA, for instance, negative feedback regulation of the target. Nevertheless, they are part of the mechanism for better or worse. In addition, many modalities, including small molecules and nucleotide-based agents, display polypharmacology, interacting with multiple targets, and with varying affinities [5–7]. Therefore, MoA is a simplified discrete model of a continuous spectrum of binding affinities at best. Moreover, drugs of distinct MoAs can have similar pharmacological effects when regulation of distinct upstream targets converges on the same downstream pathway [8]. Last but not least, even when the target and molecular MoA are unknown, drug mechanism can still be elucidated experimentally and quantitatively. Therefore, we call for a systematic, multiscale study of drug mechanism, expanding the scope of target and MoA identification.

A favourable safety profile of a module is at least as important as a favourable efficacy profile. Given that both mechanism and safety share the same physicochemical and biological foundations (drug-target interaction, gene/pathway perturbation, etc.), they are treated similarly, and sometimes simultaneously, from a modelling perspective. Mechanistic understanding of safety findings informs about the safety margin of a molecule, dissects target-related from molecule-specific or class-specific toxicity, and enables the identification of molecules with better safety profiles. By serendipity, mechanistic understanding of safety findings can repurpose a molecule and reveal unexpected insight into human biology, probably best highlighted by the history of sildenafil (Viagra®). More generally, experts consider that both refinement of *in silico* tools and greater mechanistic understanding might provide future opportunities to better identify drug safety liabilities [9].

Drug mechanism and safety are studied traditionally in animal models *in vivo* and in cellular systems *in vitro*. Since the 1970s, computational approaches have emerged, engaging *in vitro* and *in vivo* assays on the one hand, and mathematical and computational modelling *in silico* on the other hand. By the scale of the study, mathematical and computational models in preclinical drug discovery are classified into three levels: (i) *molecular-level modelling*, which models molecular structure and interactions between molecules, using physicochemical principles as well as molecular modelling and simulation techniques; (ii) *cellular- and omics-level modelling*, which probes all molecules of a particular kind in cells, their spatial organisation, as well as cellular morphology, using high-throughput techniques as well as bioinformatics and statistical analysis; and (iii) *organ- and system-level modelling*, which examines how drug and body interact and affect each other over time.

The three modelling approaches correspond to microscopic, mesoscopic, and macroscopic descriptions of complex systems, respectively [10]. On the molecular level, a large number of interacting entities (RNA species, proteins, metabolites, etc.) exist and only a small subset of relevant interactions can be modelled. On the cellular and omics level, well-designed experiments can reveal gene-, pathway-, and network-level changes following perturbation. The information is particularly valuable if the cell system in use is relevant for the disease or for the safety-relevant question [11]. On the organ and system level, pharmacokinetic (PK) and pharmacodynamic (PD) properties of selected drug candidates are modelled and scrutinised to ensure appropriate exposure and target engagement [12].

Population modelling, which examines distributions of PK and PD parameters in a human population, represents the fourth level of modelling, the highest in the hierarchy. Although it is beyond the scope of this review, population modelling is an important instrument that characterises individualised dose–exposure–response relationships [13,14].

The models form a hierarchy in which models at lower levels are the basis of models at higher levels. Models at each level have their own virtues and limitations, of which scientists and decision-makers must be aware to avoid being misinformed or misguided. Furthermore, to understand the mechanism and safety profile of a molecule, no individual model at any single level would suffice. Instead, it calls for a multiscale modelling approach.

We can implement such an approach in two steps. First, we establish mathematical, ideally mechanistic, models of drug–body interactions on the molecular level, the cellular and omics level, and the organ and system level. Next, we integrate modelling results at different levels, in the hope of gaining a holistic view of how and why the drug exerts both pharmacological and toxicological effects.

Results, both on individual levels and integrated across different levels, of multiscale modelling are impactful. On one side, they can bring deeper and sometimes unexpected insight into biology, pharmacology, and toxicology. On the other side, they allow scientists and decision-makers to quantitatively assess the probability of success of discovery projects. Such assessments can be crucial, given the high attrition rate of the industry [15], to prioritise and accelerate projects with a higher probability of success. Early but encouraging results were reported by colleagues from AstraZeneca [12] and from Merck & Co. Inc. [16], who both adopted assessment systems based on the integration of evidence from multiple levels and saw productivity increasing.

Multiscale modelling needs both knowledge and data to thrive. On the one hand, it requires human expertise that translates biological knowledge into mathematical descriptions and models that can be handled by computational tools, and the human mind to interpret the results. On the other hand, it requires high-quality and reproducible data that fulfil the Findable, Accessible, Interoperable, and Reusable (FAIR) standard [17]. To enable multiscale modelling of drug mechanism and safety, experimentalists and modelling scientists should work together to create a model-driven process that starts with hypotheses, which are rejected, accepted or refined, and retested iteratively, to gradually improve the models (Fig. 1).

Modelling preclinical drug mechanism and safety at three levels

Molecular-level modelling

At the most fundamental level, drugs exert their effects by interacting with molecular targets, including DNA, RNA, proteins, and so on. We illustrate the principles of molecular-level modelling by choosing drug–protein interactions as examples because proteins constitute most efficacy targets [18]. Although the therapeutic potential of many proteins remains to be explored [19], DNA and RNA (including epigenetic modifications), the other two components of the central-dogma model of molecular biology, are increasingly pursued as targets [20–23]. Molecular-level

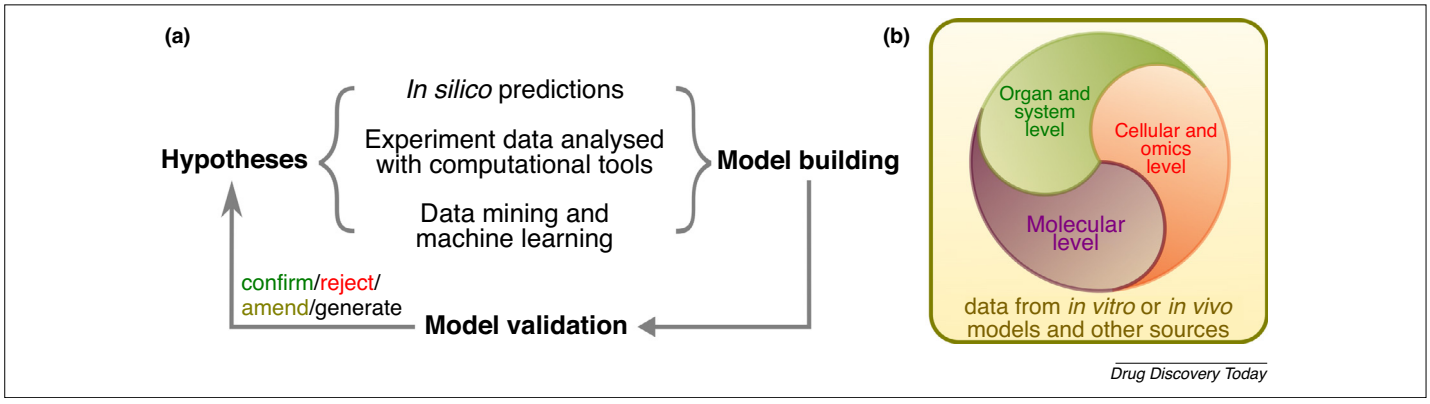


FIGURE 1

Multiscale modelling empowers preclinical drug discovery. **(a)** A drug discovery program can be modelled as an iterative process of hypothesis generation, model building, model verification (not shown for simplicity), model validation, and actions that confirm, reject, or amend the working hypotheses, and last but not least, generate new hypotheses for further testing. Computational approaches can make predictions alone, analyse experiment data, and mine existing data and identify patterns to build models. **(b)** Models are built at molecular, cellular and omics, and organ and system levels. Models of different scales inform and benefit from each other. Data generation, management, interpretation, and integration are equally important to characterise drug mechanism and safety by multiscale modelling.

modelling is also applied in these cases, with sometimes different techniques being used.

When we design novel drugs or improve existing ones, we need to tailor the molecules to have strong interactions with the desired target protein(s) and maximally weak interactions with any undesired off-target proteins. The strength of the interactions largely depends on the shape and electrostatic complementarity of the drugs with their protein binding site, famously coined the lock-and-key principle by E. Fischer [24]. Figure 2 provides an overview of molecular modelling techniques.

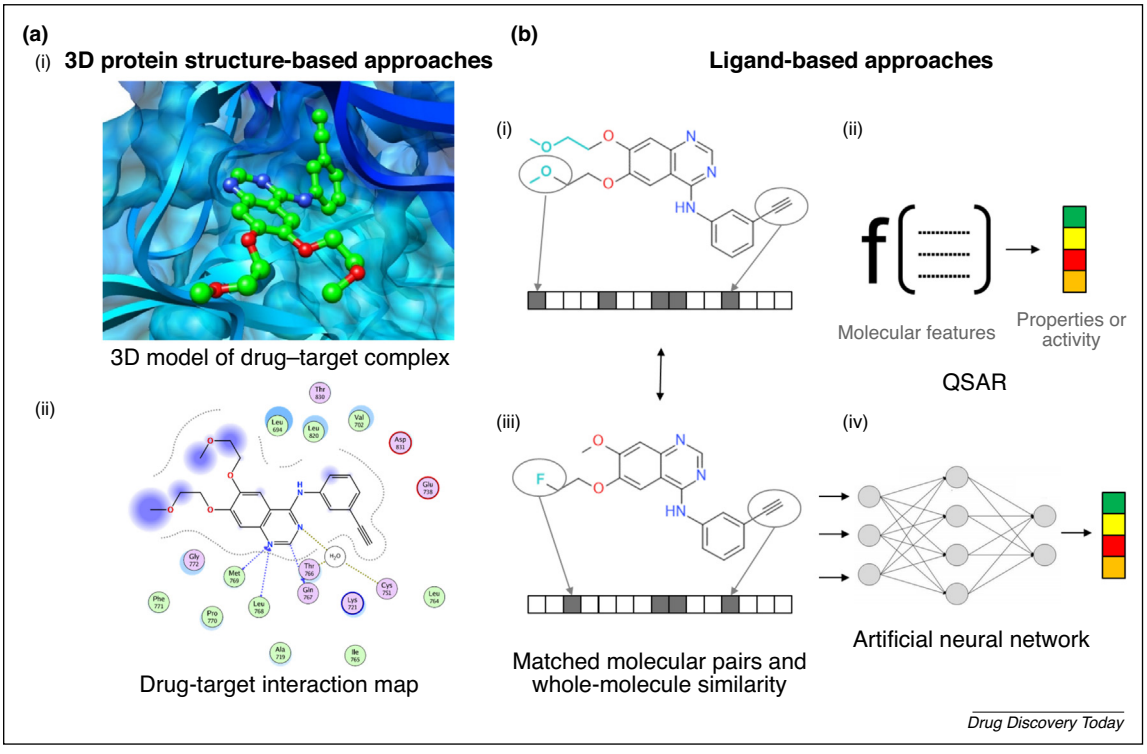


FIGURE 2

Overview of molecular-level modelling techniques. **(a)** 3D protein structure-based approaches. Molecular interactions between a drug and its target protein are key to the biological activity of the drug. Modelling these interactions requires a 3D model of the protein structure and the binding of the drug [(i) 3D model of erlotinib in complex with its target Epidermal Growth Factor Receptor, EGFR]. Different types of interaction between the drug and its target can be modelled at the atomistic or molecular level [(ii) interaction map of erlotinib with its target EGFR and solvent; created with Molecular Operating Environment (MOE) 2019.01, Chemical Computing Group]. **(b)** Ligand-based approaches. Comparison of chemical structures can be used to infer physicochemical and absorption, distribution, metabolism, and excretion (ADME) properties or biological activities. Matched molecular pair analysis focuses on local structural modifications and their effect on properties or activity, whereas molecular similarity analysis takes into account features of entire molecule structures, often encoded in a binary fingerprint [2D structure of (i) erlotinib and (iii) a closely related analogue CHEMBL2068878 [201]; modified parts are highlighted in cyan). Quantitative predictions can be derived from molecular features, using either classical quantitative structure–activity relationship (QSAR) regression methods (ii) or neural networks (iv).

Structure-based molecular modelling

To build models for molecule–target interactions, some structural information is required as a starting point. Currently, X-ray crystallography is the major experimental source for 3D protein-structure data, with currently more than 150 000 protein structures publicly available in the Protein Data Bank (PDB) [25]. Over the past few years, cryo-electron microscopy has emerged as an alternative technology that can deliver atomic-resolution structures of proteins, especially larger protein aggregates and membrane proteins [26]. Many structures of G-protein-coupled receptors (GPCRs) and ion channels have been solved, which now offers the possibility to apply structure-based drug design even for these transmembrane protein targets [27].

Although there are still several drug targets for which no 3D protein structure is known, that gap is becoming smaller [28]. In cases where no 3D structure is available, 3D homology models can be developed if structures of related proteins are known [29], for instance with tools such as SWISS-MODEL [30]. If no 3D structure of any related amino acid sequence is known, *de novo* protein structure prediction methods can be used. They are not yet highly reliable, but the field is developing steadily, as measured by the regular Critical Assessment of protein Structure Prediction (CASP) assessments, and has made promising advances, for instance with neural-network based methods [31,32].

With a 3D model available for the protein target, models of the drug–protein complex can be generated. In the case of small-molecule drugs, this can be done using either automated docking programs or manual modelling, or a combination thereof. The quality of the model can be assessed by docking scores that are based on atomistic force fields, interaction frequency statistics, or other empirical formulae [33]. Recently, machine-learning-based scoring functions have been introduced, which might one day outperform classic scoring functions [34].

Model building is usually a process where the protein target is kept rigid, and only the ligand is relaxed and fit to the binding site to save computational resources. In a further refinement step, candidate models can be energy minimised by molecular dynamics or other methods to allow for a better fit [35]. In laboratory experiments, protein–ligand binding is typically measured at room temperature, where both protein and ligand are flexible. This can be simulated in molecular dynamics simulations. Molecular dynamics simulations can also be used to test whether the modelled protein–ligand complex is stable and, therefore, likely to be true [36,37]. Simulations have greatly improved over the past few years thanks to sophisticated atomic force fields that make fewer simplifications of the underlying physics. Further improvements in modelling the underlying physics, including longer simulated time frames, could lead to even broader application in drug design [38].

3D protein–ligand interaction models are used to understand the atomistic interactions that stabilise the complex and to estimate whether a compound will show measurable binding energy. They can further be used to develop strategies for how molecules have to be modified to optimise the interaction or get rid of it in the case of off-targets [39]. 3D models can be even more impactful if they are validated with crystal structures, especially because compound binding can induce conformational changes that are not seen in the model. Important criteria for validation are: (i) whether a 3D model can explain previously observed experimental structure–activity relationships (SARs); and (ii) whether they are at least able to enrich compounds known to bind to the target among a set of decoy compounds.

Ligand-based molecular modelling

Another important class of model used in drug design are purely statistical ligand-based models. These can be rather simple matched-molecular pair models, where molecular transformations are associated with changes in molecular properties [40]. On a whole-molecule level, entire chemical structures can be represented by sets of features or descriptors in two or three dimensions [41], and the resulting representations are analysed to infer favourable molecular properties based on the 'similarity-property' principle [42]. The traditional way of modelling relationships between chemical compounds and correlating their properties to chemical structures is now being complemented, and partly even superseded, by recent advances in machine-learning algorithms [43–45]. Complex quantitative SAR (QSAR) or pharmacophore models have been generated, where a large set of computed molecular properties is used to train machine-learning models, such as random forests, support vector machines, deep neural networks, or an ensemble of models, that correlate descriptors with experimental measurements. In contrast to 3D protein–ligand models, extrapolation with statistical models needs caution because molecules of interest and their properties might not be well represented in the training set [46]. However, they come with several advantages compared with 3D models, because they are fast to calculate, can be systematically improved with more experimental training data, and typically come with a statistical estimate of the reliability of their predictions [47].

Applications of molecular-level models in drug design and discovery

In a practical drug design setting, both types of model are used for different applications. 3D models are used to optimise compound series for on- and off-target binding where a 3D structure of the target is available. They are mainly applied to generate hypotheses for beneficial chemical modifications of known structural classes in lead identification and lead optimisation campaigns [39]. Statistical models are typically used to predict physicochemical properties and absorption, distribution, metabolism, and excretion (ADME) properties of (virtual) candidate compounds. The quality of many physicochemical property predictions is often sufficient such that experiments for these properties are only occasionally necessary to spot check and improve the model for specific compound classes [48]. ADME predictions are usually followed up with experiments, but the models can also be used stand-alone to prioritise among otherwise promising compounds for synthesis [49]. During the early phases of drug discovery, when it comes to purchasing external compounds and selecting compound classes for further development, models are even more important because experimental data are either not available or usually too sparse to rank compound classes.

A special case is off-target prediction. Finding off-targets that lead to specific toxicities and ultimately adverse effects can be a daunting task. Several cases have been reported where statistical QSAR-type models have helped to identify off-targets of a given compound class [50,51]. Although the prediction of target activity is only coarse, those predictions might be sufficient to prioritise a

few out of many potential off-targets that can then be experimentally followed up. The knowledge of predicted off-targets can also be exploited for drug repurposing, where currently unknown interactions between ligands and predicted targets are exploited for novel therapeutic uses of known compounds [52]. Purely ligand similarity-based methods, as well as combinations of ligand and 3D protein modelling, have been applied successfully to predict unforeseen targets and to identify repurposing opportunities for existing drugs [53,54]. In general, however, it remains challenging to predict on- and off-targets *in silico* with high precision and recall [55].

An emerging application of known and predicted drug–target interactions is the selection and application of chemical probes to understand the molecular mechanisms in phenotypic disease models [56]. Compounds that are well characterised in terms of their molecular MoA, off-target profile, and safety can be used to perturb disease models. Molecules that evoke desired changes of the observed phenotype are analysed regarding their annotated targets and profiled with omics experiments. Systematic computational network analysis techniques are applied to infer molecular targets that are likely causal for the observed phenotype [57]. These target hypotheses can be used to focus laborious biochemical target validation experiments and ultimately lead to a better understanding of the molecular mechanisms underlying the phenotypic model. This approach can help turn a phenotypic drug discovery project into a target-based one, with the benefit of making molecular-level modelling approaches applicable and, thus, accelerating the drug discovery program [58]. Further consideration from the cellular and omics modelling perspective is provided in the subsection '*Chemogenomic and functional genomics tools*'.

Although earlier discussions focus on small molecules, molecular modelling is also indispensable to other modalities. The success of antibodies would not be imaginable without a detailed understanding of the SAR or without creative designs that lead to new MoAs [59]. Antisense oligonucleotides (ASOs) are another beneficiary of molecular modelling. Given that ASOs exert their functions by DNA/RNA interaction mediated by complementary sequence matching [60], it is feasible to model target-binding affinity and off-target potential by integrating biological sequence analysis, RNA structure, and physicochemical models [6].

In summary, molecular modelling has an important role, particularly during lead identification and optimisation. This process can be time-consuming and cumbersome, with success taking many years until a good candidate molecule is found. Computational models can be used to reduce the number of compounds that have to be synthesised, making the entire process more time- and cost-effective.

Cellular- and omics-level modelling

A cell is the smallest unit of life where the mechanism of a drug manifests. Omics technologies quantify and characterise all molecules of one kind (e.g., DNA, RNA, proteins, or metabolites) in one or more cells. Imaging technologies, such as high-content microscopy, capture morphological changes of cells upon perturbation, as well as expression changes of selected genes with either tagged molecules or surrogate markers, such as substrates or products of biochemical reactions [61]. Complementary to these technologies, functional genomics tools, among others chemical mutagenesis, RNA interference (RNAi), overexpression, and CRISPR-Cas9 loss-of-function and gain-of-function screens, offer ways to manipulate thr activity and function of individual genes [62]. Novel cellular models are used for disease and safety modelling, including induced pluripotent stem cells [63] and microphysiological systems, such as organ-on-a-chip [64] and organoids [65]. Taken together, technological and biological tools deliver huge volumes of data that allow computational tools to characterise drug mechanism and safety at the cellular and omics levels.

Properly designed, conducted, and analysed comparative experiments are key to building cellular- and omics-level models. The experiments are comparative in the sense that an absolute value, such as the speed of light, is not of interest; rather, a molecule is compared with controls or with other molecules of known mechanism [66]. Given the high dimensionality, large volume, and noisy nature of omics and imaging data, it is necessary to use appropriate statistical models to infer differences and to incorporate prior knowledge of biological systems to infer the mechanism and safety profiles of the molecule.

Modelling with omics data

Omics approaches are indispensable to modern drug discovery. They combine high-throughput biophysical and biochemical assays with statistical modelling, data mining, network analysis, and machine learning to offer a comprehensive view of gene-, pathway-, and network-level regulation by drug candidates [67–69]. Models trained with data from nontoxic and toxic compounds identify signatures applicable in preclinical safety screening [70,71].

Remarkably, omics data and models build an essential link between molecular-level modelling and organ-system level and population modelling. They can characterise molecular variations in a human population, such as single-nucleotide polymorphisms (SNPs) and other structural variants, expression levels, and splicing isoforms, which inform the status of the drug target and other genes that potentially affect PK/PD profiles in healthy or diseased individuals. For instance, a pharmacogenomics study of GPCRs demonstrated the power of omics linking models at other levels, where integrative analysis of the genetic architecture of a large population recovered individual differences in response to pharmacological modulation [72].

RNA-sequencing-based gene expression profiling is currently the most widely used omics technique to understand drug mechanisms, as judged by statistics from public data repositories. Complementary to standard full-transcriptome sequencing, several platforms exist that lower cost and reduce information redundancy. Examples include L1000, where ~1000 genes were selected to maximise the retained variance of gene expression data [67], DRUG-Seq, which is empowered by miniaturisation and shallow sequencing [73], and molecular phenotyping, where ~1200 pathway reporter genes were selected to infer pathway activity [68]. Independent of the variant, a typical data analysis workflow includes mapping reads to a reference genome or transcriptome, technical and biological quality control (QC) [74], dimension reduction and clustering, and differential gene expression analysis [75]. The differential expression profile is then compared with those of compounds with known MoA to infer the mechanism, following the 'guilt-by-association' principle [67]. In parallel, the profile can be interpreted with prior knowledge of gene regulatory

networks to infer latent genes and pathways that are likely regulated by the compound, for instance using gene-set enrichment analysis [76,77].

Our knowledge of gene regulatory networks, particularly how they vary by cell identities and by context, is scarce. One of the few things that we are certain about is that these networks are intrinsically complex. It is challenging to recover direct binding targets unambiguously from gene-expression data. Nevertheless, gene expression can offer valuable insight into the mechanism and safety profile of compounds, and comparative studies with more than one compound can reveal common and distinct mechanisms underlying the same phenotype. In a previous study, for instance, we identified two compounds in a phenotypic screening that convert stem cell-derived white adipocytes to display the brown-adipocyte phenotype. RNA-sequencing data analysis revealed that both molecules inhibit the Janus kinase/signal transducers and activators of transcription (JAK/STAT) signalling pathway. Whereas one molecule, tofacitinib, is a specific JAK3 inhibitor, the other, R406, shows polypharmacology, inhibiting multiple pathways simultaneously [78]. In another study, specific inhibition of epidermal growth factor (EGF) target genes was identified with molecular phenotyping. Together with other results, the finding helps to establish inhibition of EGF uptake as a trait of nephrotoxic antisense drugs [71].

RNA sequencing (RNA-seq) provides more information than gene expression. Splicing isoform quantification, for instance, led to the finding that the *CD44* isoform status predicts response to treatment with an anti-CD44 antibody both in cancer cell lines and in patients [79]. Another example, differential splicing of the *SMN2* gene, is discussed in detail later.

Over the past few years, RNA-seq technology has evolved from profiling many cells ('bulk-mode') to being able to sequence single cells ('single-cell mode') [80]. This progress, together with other single-cell technologies such as flow cytometry [81], mass cytometry time-of-flight (CyTOF) [82], and nuclear magnetic resonance (NMR) [83], motivated computational approaches to study biology at the resolution of single cells [84,85]. Although still in their early days, single-cell techniques have greatly expanded our ability to understand how drug perturbation affects individual cells and the communication between them.

Beyond the transcriptional level, proteomics characterises either expression or post-translational modifications of proteins *in vitro* or *in vivo* [86]. For small molecules permissive to chemical modifications, chemoproteomics identifies binding targets of modified drug candidates using the 'bait-and-prey' principle [87]. Modification-free methods, such as antibody-based cellular thermal shift assay (CETSA) and mass-spectrometry-based variants, are also available [88,89]. Compared with RNA-seq, where both devices and software solutions are more standardised, the data generation and analysis landscape of proteomics used to be fragmented. Recently, several community initiatives have benchmarked statistical analysis strategies and established best practices [90–92].

Imaging

Imaging is a sister technique of omics to study the cellular effects of drug candidates. Traditionally, it uses optical systems and computer-vision techniques to analyse cell morphology [93,94]. Emerging optics-free molecular imaging systems, such as NMR, positron emission tomography (PET) [95], and DNA microscopy [96], hold the promise of revealing the spatial organisation of biological molecules of interest.

Computational approaches are indispensable to imaging, because the sheer amount of data prohibits manual analysis, and the set-up of image assays and the analysis pipeline engage and influence each other [97]. Relevant computational concepts and non-commercial software are reviewed in [98].

Whole-slide imaging in pathology with automated image analysis is an important application that links cellular and organ-level information [99]. Cell imaging can also be used to cluster drug candidates based on their MoA [93,94]. This is of particular relevance for antimicrobial discovery because morphological changes can be concordant with the MoA of bacteria killing [100,101]. When profiling compound libraries that are annotated with target information, high-content imaging was reported to be able to predict biological activity, namely back-translating from phenotype to target [102], which, if generalisable, can be useful to identify target profiles of poorly characterised compounds.

Chemogenomics and functional genomics tools

Complementary to quantification methods, perturbations of different kinds can be used in combination with a drug candidate to interrogate its target and MoA. This orthogonal perturbation can be induced by chemical probes [33] or functional genomics tools [42]. The underlying assumption is that the MoA of the drug candidate is mediated by one type of biological molecule, and knockdown, knockout, or inhibition of that type of molecule abolishes the pharmacological effect of the drug. Similarly, overexpression or activation augments the effect. A combinatorial screening can reveal the biological molecule of interest.

One pitfall of these approaches is that, given the complex structure of gene networks, modulation of many genes can partially or fully antagonise the effect of the molecule, leading to a long list of candidates for validation. When the chemogenomics or functional genomics tool displays off-target effects, target deconvolution is even more challenging. Nevertheless, with the rapid development of *in vitro* and *in vivo* use of gene-editing systems, functional genomics approaches will likely have a huge impact on target identification and validation [103]. In such campaigns, bioinformatics tools are used to design specific agents to interfere with the expression and biological function of the putative target, and statistical models, equipped with prior knowledge, are engaged to identify the most likely targets that deserve further experimental validation.

A call for model-driven omics data interpretation and integration

We argue that different cellular and omics disciplines need to be combined and to be supplemented by models of various kinds, including molecular-level and organ- and system-level models, to enable scientific understanding of a biological process. Figure 3, a simplified diagram inspired by Thomas *et al.* [104], illustrates our point.

The centre diagram of Fig. 3 describes schematically the homeostasis of arginine biosynthesis in *Escherichia coli*. Arginine, the final biosynthesis step of which is catalysed by ornithine transcarbamylase (OTCase), binds to a transcriptional repressor and allosterically triggers its binding to promoter elements in both the OTCase and the repressor genes. When arginine is lacking,

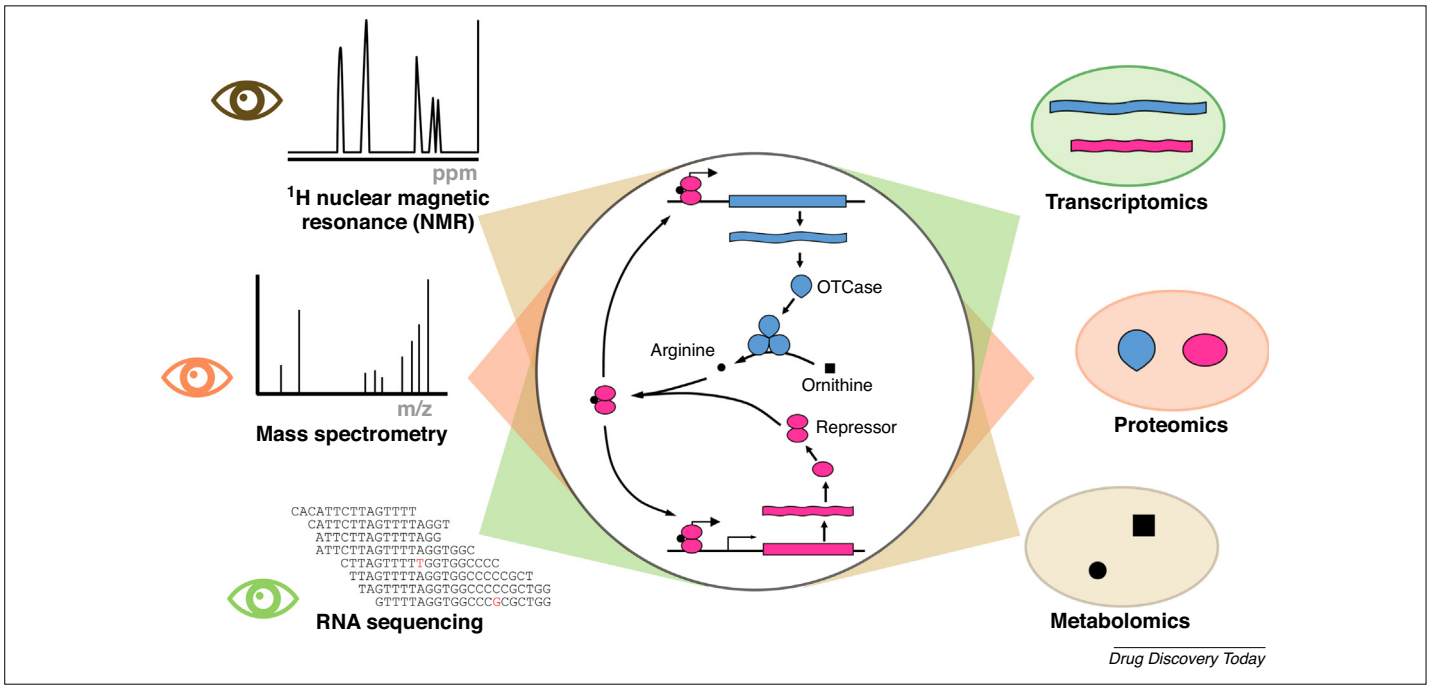


FIGURE 3

Omics data are projections of high-dimensional biological space. The centre diagram illustrates the homeostasis of arginine biosynthesis in *Escherichia coli*. Detailed descriptions can be found in the main text. Eyes indicate different perspectives of omics studies. High-throughput technologies, with ^{1}H nuclear magnetic resonance (NMR), mass spectrometry, and RNA sequencing as examples illustrated on the left side, examine all molecules of a particular kind in the cell system. They generate data of the metabolome (all metabolites), proteome (all proteins), and transcriptome (all RNAs), respectively, shown on the right side. These data, despite their high dimensionality, are low-dimensional projections of high-dimensional biological space. The same principle also applies to cellular imaging data.

repression is relieved, allowing new arginine to be produced by OTCase. Low-level, constitutive expression of the repressor gene ensures that the repressor will be available to bind arginine again once its levels have risen above a certain threshold.

The three projections surrounding the central diagram in Fig. 3 represent metabonomic, proteomic, and transcriptomic views of the system, respectively. They are generated by high-throughput technologies, namely NMR, mass spectrometry, and RNA-seq, as depicted. Based on repeated measurements of the system under various conditions, it will be possible to establish positive (ornithine and arginine) and negative (repressor and OTCase) correlations in the amounts of the analytes for any one of the three omics disciplines. However, a comprehensive understanding of even this simple network would not be possible based on any one of them alone. Thus, based on this example, albeit arguably theoretical, and other observations, we call for caution in the use of model-free approaches to drug mechanism and safety (Box 1).

A classical approach towards studying such a relatively small network would be to derive a schematic network diagram with (abstract) nodes and edges, either from literature data, dedicated molecular interaction databases, or based on correlation analyses of omics data generated under a range of conditions. Such a schema, coloured using quantitative data from an omics experiment, is often used in publications to cover the network aspect of a biological system of interest. With more complex data available, collected under different conditions, time points, or perturbations, statistical analyses enable the construction of correlation networks or Bayesian networks, suggesting functional or regulatory links between the measured entities. From an algorithmic point of view, correlation networks are liable to high false-positive and false-negative rates, even using single-cell expression data [105]. One structural drawback of Bayesian networks is that they produce directed acyclic graphs (DAGs), which cannot model feedback loops, a feature considered to be central to biological processes. Despite these limitations, it is possible to improve the results of network inference by using single-cell time-series data [106], by following the consensus of diverse algorithms [107], and by combining information from several omics disciplines, such as genetics, proteomics, and genomics [108,109].

It is important to distinguish between modelling the data and modelling the underlying biological processes. The latter types of model are mechanistic in nature. Ideally, inferring correlation or a Bayesian network model from an omics experiment will ultimately inspire a mechanistic model of the underlying biological process, but they are not guaranteed to be mechanistic *per se*. As the simple example in Fig. 3 demonstrates, there are several key aspects for understanding the mechanism and, ultimately, the dynamics of a biological network, that are nonobvious even from a combined omics point of view. These include: (i) the nature of the functional protein complexes, including the repressor dimer and the OTCase trimer; (ii) the concrete physical interactions, including those of the small molecule with repressor proteins, those of the repressor protein with DNA-binding sites, and those of the enzyme catalysis that leads to the production of arginine; and (iii) the presence and functional relevance of regulatory elements in the gene promoters.

BOX 1

Beware of model-free omics'Omics' approaches, by definition of the neologism, cover the entirety of detectable (and, typically, quantifiable) species of a certain kind, such as all mRNAs, all proteins, all miRNAs, and so on. Thus, their appeal stems from a combination of comprehensiveness and lack of bias: a properly done transcriptomics study will not miss any present transcript, and it does not make any prior assumptions about the relevance or role of any one of the measured entities. All the possibly attainable information about the biological system under study, it appears, must be contained.

Based on these premises, it is clear that omics data lend themselves readily for modern data analysis approaches, and can even be considered examples of 'big data'. Given a sufficiently large collection of, say, genetics or proteomics data sets, most of the biologically relevant patterns (of heritability, or of protein interactions) should be detectable in an almost 'model-free' approach, based on either statistical measures of correlation and significance, or causal models where the assumptions of causal inference are met. The only aspects of such studies that depend on scientific model assumptions are the identification of the relevant '-ome', and the knowledge of how to detect and measure its instances.

However, these aspects turn out to be very important and often underestimated in their consequences.

First, omics approaches never imply the measurement of all relevant players in a biological system. For example, transcriptomics, proteomics, and metabonomics represent three projections onto partially correlated subspaces of biological entities, none of which will provide more than the most superficial understanding by itself. Merging the three in 'systems biology' or 'integrated omics' efforts requires much stronger use of state-of-the-art biological models. Foregoing such integrated approaches leads in essence to a 'targeted approach' based at least on the (far-reaching) assumption that certain species of biological entities will turn out to be more relevant for a given question. From there, it is just one more step to truly targeted approaches (e.g., to monitor only the 'kinome' of protein kinases, or the 'pathway reporter genes' of the molecular phenotyping platform), methods that depend heavily on assumptions of scientific models.

Second, the theory and practice of generating omics data has a substantial influence on our interpretation of data and derived results. A well-known example is the era of hybridisation-based 'array' methods in transcriptomics. While opening the door for feasible, meaningful, and affordable omics data generation, they were, at least in their early years, mostly agnostic of splice variants. This must clearly have had an influence on the generation of hypotheses and interpretation of data, because what we cannot assess experimentally is the subject of speculation only and, if no improvement of methodology is provided, might even tend to be forgotten over time: it essentially 'ceases to exist'. Modern omics approaches studying the entirety of genetic enhancer elements, CCCTC-binding factor (CTCF) binding sites, histone modifications, or 5-methylcytidine provide evidence that the discovery of new, biologically relevant 'omes' has certainly not yet come to an end.

Finally, scientific models become vital when it comes to establishing crosslinks with other domains discussed in this review, the molecular and the organism level, and to transfer knowledge between them.

These concepts ultimately need to be incorporated into a mathematical model of this biological process, especially when our aim is to build mechanistic models predicting and explaining biological phenomena.

As discussed by Thomas *et al*. [104], the next step towards a truly integrated understanding of the network is therefore to build a mechanistic model, such as the simple one in Fig. 3, and to analyse its feedback loops, steady states, and parameter space using 'logical network analysis'. Such an analysis not only allows the identification of stable states or homeostatic loops, but also predicts effects of mutants, such as a repressor mutant with reduced affinity to arginine, or a promoter mutant that lacks the repressor binding site, at least in a qualitative way.

The ultimate, and often the admittedly distant, goal is to learn the kinetic parameters of the biological processes being studied, to describe the system using differential equations, and to simulate fully integrated dynamical models of all the involved components [110]. The availability of such models would then open the possibility to run truly *in silico* experiments and to computationally predict effects of perturbations that were never observed in an experiment.

The increased complexity in human and eukaryotes compared with prokaryotes makes network inference with omics data an intimidating task [107]. Nevertheless, examples abound where insight into human disease biology is gained by data integration [109] and modelling [111]. Readers interested in the mathematical and methodological aspects of these approaches are welcome to consult two well-written reviews [112,113]. We believe that predictive models of disease biology and drug safety integrating molecular, cellular and omics data will substantially increase our ability to prioritise targets and to understand drug mechanism and safety. Such models can be integrated into larger organ- and system-level models, as described later.

Organ- and system-level modelling

Pharmacokinetics (PK), the fate of drug substances administered to a living system, and pharmacodynamics (PD), the biochemical and physiological effects of drug substances, jointly determine the efficacy and safety profile of a drug. Mechanistic or semimechanistic mathematical models are important tools to quantify both the PK and PD profiles of drug candidates. Such models integrate information at the organ and whole-body level to establish the relationship between dose, exposure (measured in plasma, at the target site, and at off-target sites), and response, including efficacy and adverse effects. The applications of PK and PK/PD modelling in drug discovery include: (i) quantitative evaluation of the therapeutic window and, hence, informing go-no-go decisions [114,115]; (ii) integration and translation of preclinical data to inform the first-in-human dose [116–118]; (iii) informing clinical study design and drug labelling [119–121]; (iv) characterisation of patient-to-patient variability and informing optimal dose for subpopulations [122,123]; and (iv) extrapolation to special subpopulations, such as the paediatric population [119]. The adoption of PK/PD models has been strongly encouraged by health authorities, for instance through the Model-Informed Drug-Development Pilot Program (MIDD) from the US Food and Drug Administration (FDA).

Although empirical models have been widely and frequently used in PK and PK/PD modelling, here we focus on physiologically

based or mechanistically detailed models, where human physiology and pathophysiology is described mathematically as an underlying system with which the compound interacts. These models are intrinsically multiscale. On the molecular level, they can describe subcellular processes, such as enzyme reactions [124], ion channel kinetics [125,126], transporter activities [127], and subcellular signalling processes [128]. On the cellular level, they can model processes such as proliferation and apoptosis [129] as well as cell–cell interactions [130]. One level above, they can model organ- and system-level processes, for instance, whole-heart electrophysiological activities [131] and processes involved in ADME [132]. Last but not least, physiologically based or mechanistically detailed models can characterise individual variability in a population [133].

The modelling of physiological and/or pathophysiological systems is based on our understanding of the underlying biology. The models are parameterised by mechanistic data, for instance, cell proliferation rate, receptor expression level, or ion channel electrophysiology. In areas where biology is not well understood and data are scarce, an empirical or semimechanistic approach is required. Otherwise, if the biology is well studied and data are available from both human and preclinical species, these models can quantitatively capture interspecies differences and serve as a powerful tool to translate between preclinical discovery and clinical development [134–136]. If observational data confirm model predictions, we gain confidence in our understanding of the biological system and accept our hypothesis of the drug. Otherwise, if the data contradict model predictions, and if we are convinced that the data are free from bias and error, we shall revise the model and, thus, gain new insight.

To model the drug effect on the system, the models integrate compound-specific mechanistic data, such as ADME properties and target or off-target site interaction data, to predict PK, safety endpoint, or efficacy biomarkers under different dosing regimens. Identical or similar model structures can be parameterised to study the same processes or mechanisms in different physiological systems (e.g., preclinical species, healthy adults, and paediatric or patient subpopulations). Mathematically, these models are often systems of ordinary differential equations (ODEs), describing variables changing over time, for example, drug concentration in time course or the number of tumour cells at each time point. Partial differential equations (PDEs) [137], stochastic differential equations [138,139], and agent-based models [140,141] are also occasionally used. A schematic illustration of the physiological based systems model is shown in Fig. 4.

To highlight how these models assist drug mechanism and safety understanding, we focus here on two aspects: physiological-based PK (PBPK) models, and systems models for safety, using cardiac safety models as an example.

PBPK models

PK profiles of drugs are classically modelled by compartment models with first-order kinetics (see [142] for a concise introduction and [143] for a comprehensive treatment). Compartments in such models do not necessarily have a physiological meaning. By contrast, a PBPK model comprises compartments that represent organs or tissues (e.g., liver, heart, brain, etc.), and these compartments are connected by blood or lymph flow. The model is typically a system of coupled ODEs and contains physiological parameters such as intestinal fluid volume, intestinal pH, blood flow rate, organ/tissue volumes, and transporter and metabolic enzyme expression levels. These system parameters can be adjusted to reflect different physiology conditions. For instance, the demographic parameter and the metabolic enzyme abundance parameter can be adjusted to build a virtual patient cohort with a particular ethnic and genetic background [144]. The model uses drug-specific parameters, such as solubility, permeability, plasma protein binding, tissue-to-plasma partition coefficient, enzyme kinetics and intrinsic clearance, most of which are determined

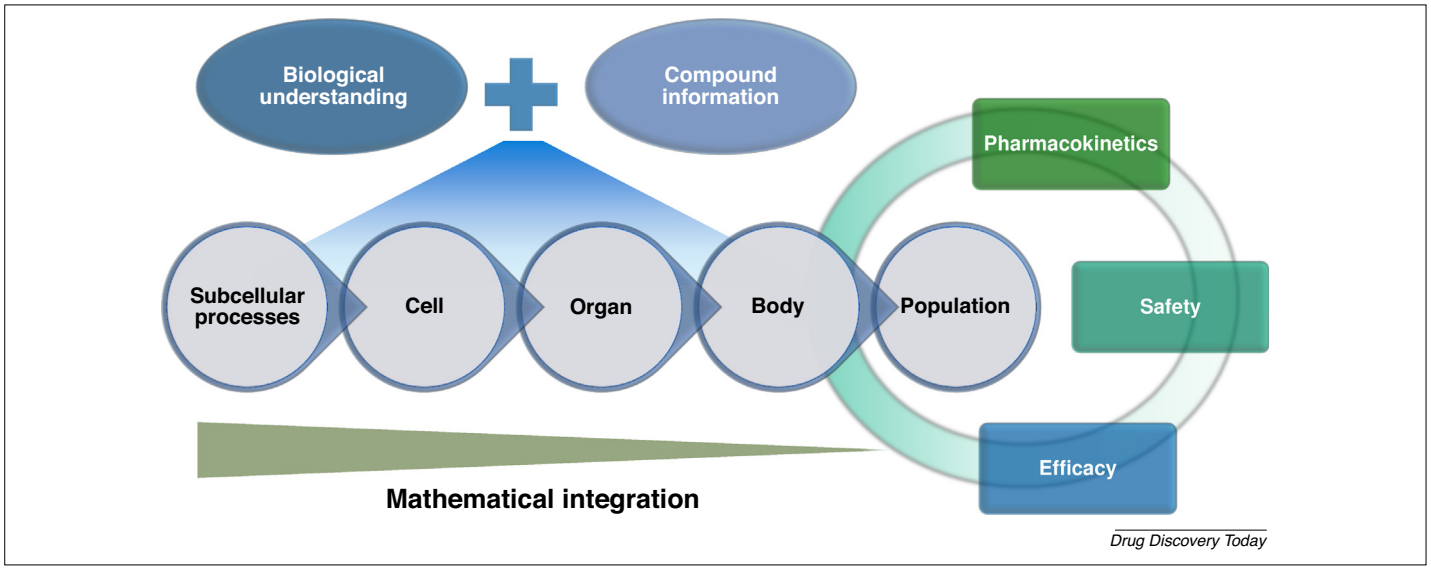


FIGURE 4

Schematic illustration of system- and organ-level models. Such models often take the form of a system of coupled ordinary differential equations (ODEs), with biological understanding and compound information encoded as parameters. Information of different scales is integrated to inform pharmacokinetics, safety, and efficacy profiles of drugs over time.

by preclinical mechanistic studies. The compound distribution into each organ can be typically modelled as either perfusion limited, in which the distribution is limited by the blood flow rate, or permeability limited, in which the distribution is limited by the ability of the compound crossing the membrane. The basic concept of the PBPK model, in particular for small molecules, was nicely reviewed by Jones and Rowland-Yeo [134]. Reviews and examples of PBPK models for biologics can be found elsewhere [145–147].

PBPK, especially small-molecule PBPK, is one of the most established systems models and has had a significant regulatory-included impact in the pharma industry. Between 2008 and 2017, the FDA's Office of Clinical Pharmacology received 130 investigational new drug (IND) applications and 94 new drug applications (NDA) that contain PBPK modelling, excluding *de novo* PBPK analysis that informed the regulatory decision for 30 submissions [119]. By the end of 2015, the European Medicines Agency (EMA) had received a total of 67 submissions of PBPK models, 20 of which were suggested by the regulator [148]. Between 2014 and 2016, the Japanese Pharmaceuticals and Medical Devices Agency (PMDA) received 17 submissions of PBPK analysis in NDAs [149]. The main areas of PBPK regulatory applications include, but are not limited, to prediction of drug–drug interactions (DDI) mediated by P450 enzymes [150,151] or transporters [127,152], prediction of paediatric exposure to inform the dosing regimen [153], prediction of exposure for patients with organ impairment [154–156], and prediction of food effects [156].

Systems models for safety

Systems models offer an integrative approach to combine mechanistic preclinical safety data with PK and patient-specific characteristics. They can assess safety liabilities under different dosing regimens in different populations quantitatively and identify mechanisms underlying liabilities, which may offer insight for compound optimisation. Similar to PBPK models, these safety models serve as a platform and are not restricted to particular disease areas.

Cardiac safety models are some of the most advanced systems safety models and have gained regulatory attention and encouragement [157]. These models describe cardiac electrophysiology at multiple levels mathematically. At the cellular level, the opening, closing, and inactivation of cardiac ion channels, the activities of ion pumps and exchangers, the intracellular calcium ion handling, and the cyclic transmembrane potential change are modelled. At the tissue and/or organ level, both the electrical coupling between cardiac myocytes and the electrical signal propagation through the heart are modelled. At the whole-body level, the electrical signal propagation to the torso surface and, hence, the electrocardiogram signal, are modelled [158]. The cellular electrophysiology is modelled as coupled ODEs and the organ and/or torso models are governed by the bidomain model, a PDE model, coupled with the cellular ODE models.

Similar to PBPK models, cardiac safety models also require a set of systems parameters, such as ion channel kinetics (e.g., transition rates between different states), ion channel conductance, which is associated with their abundance, as well as heart and torso anatomy. The models integrate information on drug interaction with different ion channels, which for instance can be inferred from *in vitro* ion channel inhibition data from the patch-clamp experiment, to predict the impact on cellular action potential, whole-heart electrical signals, or electrocardiogram (ECG). One typical aim of such predictions is to assess the proarrhythmic potential of drug candidates. The development and use of cardiac models for safety assessment have been reviewed elsewhere [159]. Recently, the FDA-led Comprehensive *in vitro* Proarrhythmia (CiPA) initiative published a series of papers on training [160,161] and validation [162] of a model to predict the risk of torsade de pointes, a rare but potentially fatal arrhythmia, which demonstrated the feasibility of establishing a predictive cardiac safety model.

Apart from guiding compound and dose selection to minimise proarrhythmic risk, these models also inform the intensity of clinical ECG monitoring and drug labelling, including prohibited concomitant medications [121,163]. In November 2018, a concept paper from the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) was released [163]. It indicated that the cardiac safety model, following the path of the PBPK model, will soon become another systems model in the regulatory context. Furthermore, other systems safety models are emerging, such as those that predict drug-induced liver injury [164].

Although we have focussed on applying systems models for safety so far, systems modelling has also been widely used to quantitatively describe disease biology. For example, a model constructed by Clausznitzer *et al.* described lipid dysregulation in the brain of patients with Alzheimer's disease (AD). The model predicted that activation of sphingosine-1-phosphate receptor 5 (S1PR5) can reverse the lipid dysregulation in AD and, therefore, that S1PR5 may serve as a promising target [165]. In addition to building confidence in a target and revealing potential new targets, these models are also used to optimise dosing regimens [166,167] and to stratify patients into responders and nonresponders [168].

Towards multiscale modelling: a case study

We illustrated earlier the principles of the three levels of modelling. The different modelling approaches share mathematics as the common language to describe the complex system of human biology and how drugs assert their effects. They also share informatics as the common tool to handle data, perform computations, make inferences, and transform models and data into human interpretable results. Moreover, they complement each other to shed light on human disease biology and the mechanism and safety profile of drugs in a holistic view.

In the parable 'Blind men and an elephant', each blind man perceives an elephant, depending on where he touches it, as a snake (the trunk), a fan (the ear), a tree trunk (the leg), a wall (the side), a rope (the tail), or a spear (the tusk) [169]. Each model only reveals part of the truth at best. To understand the elephant, it is necessary to combine different views, synthesise the evidence and models into a unifying model, and put that model to test. In our case, drug mechanism and safety is the elephant, and clinical trials are the test. We foresee that a multiscale modelling approach will improve our understanding of drug mechanism and safety profiles over isolated modelling approaches.

To illustrate the potential, let us examine a case elephant. Small-molecule splicing modifiers of the human survival of motor neuron 2, centromeric gene (*SMN2*) in the context of spinal muscular

atrophy (SMA) provide a recent, concrete example of how multiscale modelling can help elucidate mechanism and safety profiles of a drug candidate. Discussions below are based on published data [170–173].

The transcriptomes of patient cells in the presence or absence of test compounds were profiled with RNA sequencing. Although differential expression profiles were readily available using standard algorithms, these effects were deemed secondary to the primary effect of the drug, namely corrected splicing of the *SMN2* gene and potentially other, off-target, genes. In support of this notion, a chemoproteomics analysis with protein extracts from patient cells provided evidence for an interaction of a derivative of the candidate compound with the cellular splicing machinery.

Therefore, a second analysis was based on only those sequencing reads that uniquely represented one of the two possible alternatives in any of the about ~300 000 alternative, local splicing events described in the RefSeq human transcriptome database [174]. Based on this, every local splicing event (as opposed to a full-length splice variant, which can depend on the combination of any number of local alternative splicing events) was monitored and characterised by changes in a 'percent spliced in' score (ΔPSI) [171]. Although all of the tested compounds exerted comparable effects on the target splicing event in *SMN2*, they showed different, although largely overlapping, profiles in terms of additional, off-target events. A comparable analysis by Palacino *et al.* [170] had already revealed a preferred splice site consensus pattern of GA|GUAAGU in targeted splice sites, as opposed to the canonical donor site AG|GURAGU, where the vertical bars indicate the exon–intron boundary. Direct binding of the drug candidates to the complex formed from such 5′ donor splice sites and U1 snRNA was then demonstrated by solution-state ^{1}H NMR, and a structural model was derived (Fig. 2 in [171]). Through this integrated approach, patterns in an omics data set were ultimately explained in terms of ever more-refined mechanistic models, which finally elucidated the MoA at an atomic resolution [173]. The findings strengthened confidence in the original models and the compounds under investigation.

To explain the differences in off-target potential between different candidate compounds, an additional search for shared sequence patterns was performed that identified a 12-base pair-long stretch of purine residues ~25 base pairs upstream of the crucial donor splice sites in both the primary target and the most persistent off-target splice site across all the experiments performed. This sequence stretch had already been described as an exonic splicing enhancer (ESE) in the case of the target *SMN2* gene [175]. Subsequent analysis using surface plasmon resonance confirmed that the most selective compounds gained their increased selectivity through the joint interaction with the splice site and this purine-rich ESE [171,173].

As summarised in [172], substantial additional efforts in the areas of pharmacology, drug metabolism and PK (DMPK), and nonclinical safety were required to further understand the *in vivo* effects of the studied compounds and to ultimately allow a few of them to progress to clinical trials. A particularly interesting example is the characterisation of 'compound 2' (Fig. 1 in [172]). After oral administration in rodents and cynomolgus monkeys, it was found to be dealkylated *in vivo* (Fig. 5 in [172]), yielding a related 'compound 6', which reached plasma concentrations of up to 9% of those of the originally administered 'compound 2', but being tenfold more potent than the parent compound. This metabolite was also peripherally restricted (i.e., did not reach the brain) and, thus, showed a different overall profile compared with the parent 'compound 2'. These findings led to the design and characterisation of more advanced and advantageous compounds as detailed in [172].

In summary, transcriptomics data and their context-specific analysis have inspired chemoproteomics, structural and bioinformatics pattern searches, and DMPK analyses that, in combination, provided deep insight into the mechanism and safety profile of risdiplam, a drug candidate for SMA, enabling its progression to clinical trials. Thus, multiscale modelling integrates learnings across disciplines to accelerate discovery and development.

Concluding remarks and outlook

Here, we introduced molecular-level, cellular- and omics-level, and organ- and system-level models, examined their applications, and illustrated the potential impact of a multiscale modelling approach in drug discovery with the case study of risdiplam.

The concept of multiscale modelling has been proposed in many different areas. In physics and material science, it summarises mathematical concepts and tools that exploit the disparity of different scales and provides solutions to multiphysics problems [176,177]. In biochemistry, multiscale modelling that integrates classical molecular modelling and quantum models had a fundamental role in elucidating complex biological and chemical systems, for which Martin Karplus, Michael Levitt, and Arieh Warshel were awarded the Nobel Prize in Chemistry 2013. In biology, it has been realised that a multiscale modelling approach is indispensable to understand the hierarchical nature and the complexity of diverse biological systems [178–180]. In drug discovery, we propose multiscale modelling as a process of building mathematical, preferably mechanistic, models at individual levels and integrating them. We believe that multiscale modelling will not only lift productivity and reduce the attrition rate, but also lead to deeper and novel insights into disease biology and, ultimately, to new therapeutics.

We have limited our discussions to preclinical research. We note that other aspects of drug discovery and development, including biomarker and clinical development, have benefited from model-driven approaches, such as pharmacometrics and quantitative systems pharmacology, as comprehensively reviewed in [181] and [182], respectively. The trend of using models to inform decisions is also observed beyond the pharma industry. For instance, European Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) legislation already accepts predictions made by QSAR models in special cases [183].

Although the short-term goal of multiscale modelling is to enrich the knowledge of a particular molecule, either to gain confidence or to stop and fail early, one key long-term goal is to accumulate knowledge in disease biology, pharmacology, and toxicology in the form of testable, predictive, and ideally personalised, models. Reverse translation, the practice of extracting patterns and building models from clinical trials as well as other real-world evidence to inform preclinical discovery projects, is a viable option to feed population models and patient-relevant data into the multiscale modelling schema [184]. Multiscale modelling

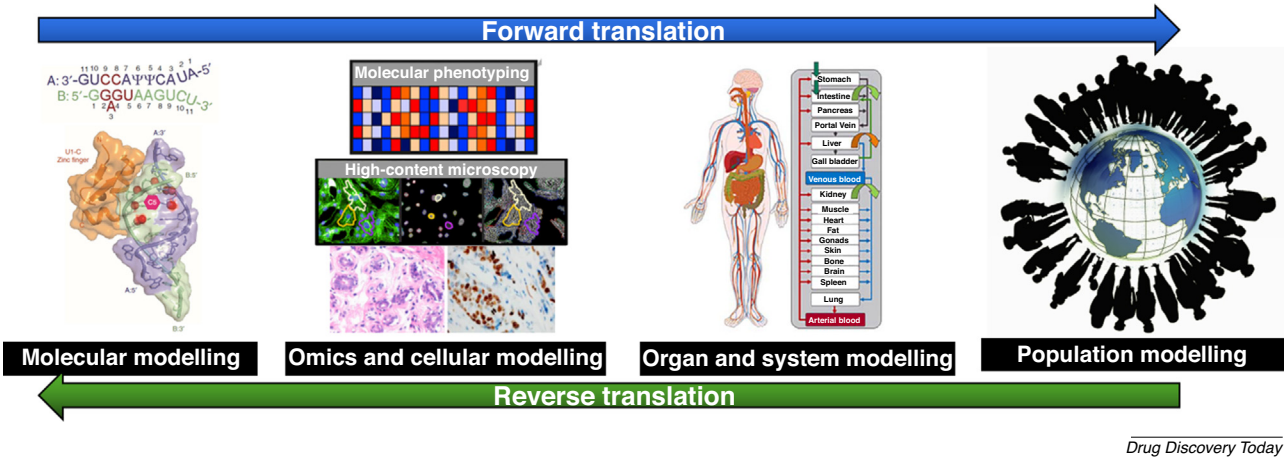


FIGURE 5

Multiscale modelling as a computational engine serves forward and reverse translation in drug discovery. Modern drug discovery depends on both forward translation, the process of turning molecular findings into drugs, as well as reverse translation, the process of turning real-world evidence and data into knowledge that catalyses future therapeutics. Multiscale modelling supports both processes by modelling biological systems at different levels and integrating models of different scales. Adapted from [171] (molecular modelling), [68] (cellular and omics modelling), [132] (organ and system modelling), and obtained from pixabay.com (user: geralt) under the Pixabay License (population modelling).

as a computational engine has the potential to serve both forward and reverse translation, as illustrated in Fig. 5.

Understanding drug mechanism and safety takes hard, creative work and a long time. Take acetylsalicylic acid (Aspirin®) as an example. It was first synthesised and manufactured between 1897 and 1899. Not until 2019 was a new mechanism reported, in which acetylsalicylic acid directly acetylates cyclic GMP–AMP synthase (cGAS) to prevent DNA-induced autoimmunity [185]. Although it might appear anecdotal, historical reviews suggest that the time span between drug discovery and mechanism and safety understanding can take years to decades, well beyond the usual timescale of a preclinical drug discovery program [186,187].

The discrepancy between the time and effort needed for a mechanistic understanding of drug mechanism and safety at multiple levels and business life-cycles of the pharma industry implies that new ways of research are needed. Collaborations between industry, academia, and research institutes to build multiscale models jointly might be beneficial. Indeed, technical solutions that allow private and practical pharmacological collaboration already exist [188]. Community-wide efforts, such as CASP assessments and DREAM challenges, participated in by both academic and industrial research groups, have demonstrated their value to identify pros and cons of individual models and to generate consensus models that outperform individual models [107,189]. We foresee that more interdisciplinary research across boundaries will transform the process of understanding drug mechanism and safety.

In contrast to the disciplines mentioned earlier, where multiscale modelling has been established and applied, multiscale modelling in drug discovery and development is still in its infancy. Its scope needs to be refined, its methodology to be established, and its impact to be scrutinised. Key open questions that are beyond the scope of this review beg answers. For instance, how can data and models, in line with the FAIR principles, be shared and managed so that they can be reused for other purposes than the original quest? Are there mathematical methods and models available, for example, Bayesian optimisation, as some experts have suggested, that have the promise to take the human out of the loop and to jointly optimise parameters of heterogeneous models [190]? If the answer is 'yes', what prevents us from building and using them? If the answer is 'no', at least for a given period of time, how can we make sure that the quantitative outputs of multiscale models are meaningfully processed by humans and translated into appropriate decisions?

Neither answers to these questions nor any success recipe are available to us. Nevertheless, we are inspired by the success of multiscale modelling in many scientific fields and are convinced of its potential in drug discovery as well. We invite all interested researchers to join forces with us in implementing and practising it. In Box 2, we share ten open questions, in the hope of initiating a dialogue with the community and learning from best practices.

Despite the many open questions and uncertainties, we envision that the implementation of multiscale modelling in drug discovery, in which modelling activities at different scales synergise and inform each other, will lead to insight in disease biology, which in turn will translate into effective and safe medicines. We are witnessing more powerful and accurate tools for molecular modelling and dynamics [191], accumulating single cell-level knowledge of human biology at temporal and spatial resolutions [192–194], physiology-emulating biological modelling systems [64], development of precise genome-editing tools [22,195], higher degree of automation [196], and last but not least, increased computational power and scalability [197]. It is hoped that these progresses will catalyse future models that operate in the 'what-if' mode. By combining *in silico* modelling and simulation with (semi-)automatic *in vitro* biological assays intelligently, the 'what-if' models shall assist drug-discovery scientists in maximising the knowledge of their drug candidate before human clinical testing.

BOX 2

Ten open questions on multiscale modelling of drug mechanism and safety We identified a set of open questions on how our three levels of modelling can benefit from each other. We appreciate open discussions and experience sharing within the community.

1 How can we assess off-target relevance in a cell-identity-, dose-, and MoA-specific manner, integrating evidence from different levels?
2 How good do compound property predictions need be so that they are useful in PK/PD modelling?
3 How can we integrate molecular modelling results as prior information for omics studies?
4 How can we effectively translate omics readouts, which are often relative because of the comparative nature of experiments, into parameters of organ- and system-level models?
5 How can we use the output of basic models (e.g., QSAR predicted ADME properties or ion channel IC_{50} values) as input for physiologically based PK or PK/PD models, particularly for early-stage compounds when experimental data are not available?
6 How can we use cellular and omics information to identify relevant cellular and/or subcellular processes to be included in the PK/PD model and, hence, determine the model structure?
7 How can we use molecular, cellular, and omics information, such as the abundance of particular proteins and pathway activity, to inform parameter values of system models?
8 How can we connect models in preclinical research with clinical, population models, for both forward and reverse translation?
9 What methodologies and tools can we use to facilitate and assist multiscale modelling?
10 How can we scientists working in drug discovery implement and achieve multiscale modelling as a community?

Acknowledgements

The authors thank the many colleagues who shared insight and provided inputs during years of collaboration. The names, if all listed, shall consume a significant proportion of the expected length of the article. Nevertheless, we would like to particularly thank Manfred Kansy and Holger Fischer for their knowledge of, and insight into, quantitative aspects of drug discovery, and for their suggestions to improve the manuscript. Corinne Solier and Michael Prummer offered valuable feedback. We thank Fabian Birzele, Jérôme Hert, Barbara Endler-Jobst, Juergen Hammer, Benjamin Ribba, and Thomas Singer for their support.

We are in debt to Roald Hoffmann and Jean-Paul Malrieu for their captivating tripartite essay on 'Simulation versus Understanding' [198–200], which we discovered among others from the blog *In The Pipeline* by Derek Lowe during the revision process. We thank four anonymous reviewers for their invaluable criticisms and insightful suggestions. J.D.Z. thanks students of the 'Applied Mathematics and Informatics in Drug Discovery' course and colleagues supporting the lecture series at the Department of Mathematics and Informatics, University of Basel, for inspiration and motivation.

J.D.Z. and M.E. wish to dedicate their work to Clemens Broger (2017†), a pioneer of bioinformatics in drug discovery, and a man true to himself.

This work was funded by F. Hoffmann-La Roche Ltd.

References

1 Schenone, M. *et al.* (2013) Target identification and mechanism of action in chemical biology and drug discovery. *Nat. Chem. Biol.* 9, 232–240
2 Ziegler, S. *et al.* (2013) Target identification for small bioactive molecules: finding the needle in the haystack. *Angew. Chem. Int. Ed.* 52, 2744–2792
3 Comess, K.M. *et al.* (2018) Emerging approaches for the identification of protein targets of small molecules - a practitioners' perspective. *J. Med. Chem.* 61, 8504–8535
4 Miller-Jensen, K. *et al.* (2007) Common effector processing mediates cell-specific responses to stimuli. *Nature* 448, 604–608
5 Wang, Y. *et al.* (2016) Evidence-based and quantitative prioritization of tool compounds in phenotypic drug discovery. *Cell Chem. Biol.* 23, 862–874
6 Hagedorn, P.H. *et al.* (2017) Managing the sequence-specificity of antisense oligonucleotides in drug discovery. *Nucleic Acids Res.* 45, 2262–2282
7 Gao, J. *et al.* (2018) Small molecule interactome mapping by photoaffinity labeling reveals binding site hotspots for the NSAIDs. *J. Am. Chem. Soc.* 140, 4259–4268
8 Malone, C.F. *et al.* (2017) mTOR and HDAC inhibitors converge on the TXNIP/thioredoxin pathway to Cause catastrophic oxidative stress and regression of RAS-driven tumors. *Cancer Discov.* 7, 1450–1463
9 Weaver, R.J. and Valentin, J.-P. (2019) Today's challenges to de-risk and predict drug safety in human 'mind-the-gap'. *Toxicol. Sci.* 167, 307–321
10 Lachowicz, M. (2011) Microscopic, mesoscopic and macroscopic descriptions of complex systems. *Probab. Eng. Mech.* 26, 54–60
11 Horvath, P. *et al.* (2016) Screening out irrelevant cell-based models of disease. *Nat. Rev. Drug Discov.* 15, 751–769
12 Morgan, P. *et al.* (2018) Impact of a five-dimensional framework on R&D productivity at AstraZeneca. *Nat. Rev. Drug Discov.* 17, 167–181
13 Sheiner, L. and Wakefield, J. (1999) Population modelling in drug development. *Stat. Methods Med. Res.* 8, 183–193
14 Mould, D.R. and Upton, R.N. (2012) Basic concepts in population modeling, simulation, and model-based drug development. *CPT Pharmacomet. Syst. Pharmacol.* 1, 1–14
15 Waring, M.J. *et al.* (2015) An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat. Rev. Drug Discov.* 14, 475–486
16 Brown, F.K. *et al.* (2017) Data to decisions: creating a culture of model-driven drug discovery. *AAPS J.* 19, 1255–1263
17 Wilkinson, M.D. *et al.* (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 1–9
18 Santos, R. *et al.* (2017) A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* 16, 19–34
19 Oprea, T.I. *et al.* (2018) Unexplored therapeutic opportunities in the human genome. *Nat. Rev. Drug Discov.* 17, 317–332
20 Matsui, M. and Corey, D.R. (2017) Non-coding RNAs as drug targets. *Nat. Rev. Drug Discov.* 16, 167–179
21 Setten, R.L. *et al.* (2019) The current state and future directions of RNAi-based therapeutics. *Nat. Rev. Drug Discov.* 18, 421–446
22 Fellmann, C. *et al.* (2017) Cornerstones of CRISPR-Cas in drug discovery and therapy. *Nat. Rev. Drug Discov.* 16, 89–100
23 Jones, P.A. *et al.* (2016) Targeting the cancer epigenome for therapy. *Nat. Rev. Genet.* 17, 630–641
24 Fischer, E. (1894) Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte Dtsch Chem. Ges.* 27, 2985–2993
25 Berman, H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242
26 Fernandez-Leiro, R. and Scheres, S.H.W. (2016) Unravelling biological macromolecules with cryo-electron microscopy. *Nature* 537, 339–346
27 Hilger, D. *et al.* (2018) Structure and dynamics of GPCR signaling complexes. *Nat. Struct. Mol. Biol.* 25, 4–12

28 Hauser, A.S. *et al.* (2017) Trends in GPCR drug discovery: new agents, targets and indications. *Nat. Rev. Drug Discov.* 16, 829–842
29 Hillisch, A. *et al.* (2004) Utility of homology models in the drug discovery process. *Drug Discov. Today* 9, 659–669
30 Waterhouse, A. *et al.* (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303
31 Evans, R. *et al.* (2018) De novo structure prediction with deep-learning based scoring. *Annu. Rev. Biochem.* 77, 363–382
32 AlQuraishi, M. (2019) End-to-end differentiable learning of protein structure. *Cell Syst.* 8, 292–301
33 Li, J. *et al.* (2019) An overview of scoring functions used for protein–ligand interactions in molecular docking. *Interdiscip. Sci. Comput. Life Sci.* 11, 320–328
34 Wallach, I. *et al.* (2015) AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *arXiv* 2015, 1510.02855v1
35 Lee, G.R. and Seok, C. (2016) Galaxy7TM: flexible GPCR–ligand docking by structure refinement. *Nucleic Acids Res.* 44, W502–W506
36 Chen, Y.-C. (2015) Beware of docking! *Trends Pharmacol. Sci.* 36, 78–95
37 Yuan, H. and Merck D3R Team, (2017) *MD Simulation in Pose Refinement and Scoring Using AMBER Workflows*. Merck
38 Klepeis, J.L. *et al.* (2009) Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* 19, 120–127
39 Bissantz, C. *et al.* (2010) Medicinal chemist's guide to molecular interactions. *J. Med. Chem.* 53, 5061–5084
40 Kramer, C. *et al.* (2018) Learning medicinal chemistry absorption, distribution, metabolism, excretion, and toxicity (ADMET) rules from cross-company matched molecular pairs analysis (MMPA). *J. Med. Chem.* 61, 3277–3292
41 Nicholls, A. *et al.* (2010) Molecular shape and medicinal chemistry: a perspective. *J. Med. Chem.* 53, 3862–3886
42 Maggiora, G. *et al.* (2014) Molecular similarity in medicinal chemistry. *J. Med. Chem.* 57, 3186–31204
43 Sellwood, M.A. *et al.* (2018) Artificial intelligence in drug discovery. *Future Med. Chem.* 10, 2025–2028
44 Chen, H. *et al.* (2018) The rise of deep learning in drug discovery. *Drug Discov. Today* 23, 1241–1250
45 Lo, Y.-C. *et al.* (2018) Machine learning in chemoinformatics and drug discovery. *Drug Discov. Today* 23, 1538–1546
46 Sahigara, F. *et al.* (2012) Comparison of different approaches to define the applicability domain of QSAR models. *Molecules* 17, 4791–4810
47 Sheridan, R.P. (2013) Using random forest to model the domain applicability of another random forest model. *J. Chem. Inf. Model.* 53, 2837–2850
48 Meier, R.J. (2019) A way towards reliable predictive methods for the prediction of physicochemical properties of chemicals using the group contribution and other methods. *Appl. Sci.* 9, 1700
49 Lobell, M. *et al.* (2006) In silico ADMET traffic lights as a tool for the prioritization of HTS hits. *ChemMedChem* 1, 1229–1236
50 Keiser, M.J. *et al.* (2007) Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.* 25, 197
51 Awale, M. and Reymond, J.-L. (2017) The polypharmacology browser: a web-based multi-fingerprint target prediction tool using ChEMBL bioactivity data. *J. Cheminform.* 9, 11
52 Cha, Y. *et al.* (2018) Drug repurposing from the perspective of pharmaceutical companies. *Br. J. Pharmacol.* 175, 168–180
53 Keiser, M.J. *et al.* (2009) Predicting new molecular targets for known drugs. *Nature* 462, 175–181
54 Zhou, H. *et al.* (2015) Comprehensive prediction of drug-protein interactions and side effects for the human proteome. *Sci. Rep.* 5, 11090
55 Mathai, N. *et al.* (2019) Validation strategies for target prediction methods. *Brief Bioinform.* . http://dx.doi.org/10.1093/bib/bbz026
56 Bunnage, M.E. *et al.* (2013) Target validation using chemical probes. *Nat. Chem. Biol.* 9, 195–199
57 Hill, S.M. *et al.* (2016) Inferring causal molecular networks: empirical assessment through a community-based effort. *Nat. Methods* 13, 310–318
58 Jones, L.H. and Bunnage, M.E. (2017) Applications of chemogenomic library screening in drug discovery. *Nat. Rev. Drug Discov.* 16, 285–296
59 Beck, A. *et al.* (2010) Strategies and challenges for the next generation of therapeutic antibodies. *Nat. Rev. Immunol.* 10, 345–352
60 Bennett, C.F. and Swayse, E.E. (2010) RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annu. Rev. Pharmacol. Toxicol.* 50, 259–293
61 Perlman, Z.E. *et al.* (2004) Multidimensional drug profiling by automated microscopy. *Science* 306, 1194–1198
62 Nijman, S.M.B. (2015) Functional genomics to uncover drug mechanism of action. *Nat. Chem. Biol.* 11, 942–948
63 Lei, C.L. *et al.* (2017) Tailoring mathematical models to stem-cell derived cardiomyocyte lines can improve predictions of drug-induced changes to their electrophysiology. *Front. Physiol.* 8, 986
64 McAleer, C.W. *et al.* (2019) Multi-organ system for the evaluation of efficacy and off-target toxicity of anticancer therapeutics. *Sci. Transl. Med.* 11, eaav1386
65 Takahashi, T. (2019) Organoids for drug discovery and personalized medicine. *Annu. Rev. Pharmacol. Toxicol.* 59, 447–462
66 Bailey, R. (2008) *Design of Comparative Experiments*. Cambridge University Press
67 Subramanian, A. *et al.* (2017) A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 171, 1437–1452
68 Drawnel, F.M. *et al.* (2017) Molecular phenotyping combines molecular information, biological relevance, and patient data to improve productivity of early drug discovery. *Cell Chem. Biol.* 18, 624–634
69 Moffat, J.G. *et al.* (2017) Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nat. Rev. Drug Discov.* 16, 531–543
70 Zhang, J.D. *et al.* (2014) Data mining reveals a network of early-response genes as a consensus signature of drug-induced *in vitro* and *in vivo* toxicity. *Pharmacogenomics J.* 14, 208–216
71 Moisan, A. *et al.* (2017) Inhibition of EGF uptake by nephrotoxic antisense drugs *in vitro* and implications for preclinical safety profiling. *Mol. Ther. Nucleic Acids* 6, 89–105
72 Hauser, A.S. *et al.* (2018) Pharmacogenomics of GPCR drug targets. *Cell* 172, 41–54
73 Ye, C. *et al.* (2018) DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery. *Nat. Commun.* 9, 4307
74 Zhang, J.D. *et al.* (2017) Detect tissue heterogeneity in gene expression data with BioQC. *BMC Genomics* 18, 277
75 Love, M.I. *et al.* (2015) RNA-Seq workflow: gene-level exploratory analysis and differential expression. *F1000Research* 4, 1070
76 Goeman, J.J. and Bühlmann, P. (2007) Analysing gene expression data in terms of gene sets: methodological issues. *Bioinformatics* 23, 980–987
77 Geistlinger, L. *et al.* (2019) Towards a gold standard for benchmarking gene set enrichment analysis. *BioRxiv* 2019, 674267
78 Moisan, A. *et al.* (2015) White-to-brown metabolic conversion of human adipocytes by JAK inhibition. *Nat. Cell Biol.* 17, 57–67
79 Birzele, F. *et al.* (2015) CD44 isoform status predicts response to treatment with anti-CD44 antibody in cancer patients. *Clin. Cancer Res.* 21, 2753–2762
80 Luecken, M.D. and Theis, F.J. (2019) Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* 15, e8746
81 Edwards, B.S. and Sklar, L.A. (2015) Flow cytometry: impact on early drug discovery. *J. Biomol. Screen.* 20, 689–707
82 Anchang, B. *et al.* (2018) DRUG-NEM: Optimizing drug combinations using single-cell perturbation response to account for intratumoral heterogeneity. *Proc. Natl. Acad. Sci. U. S. A.* 115, E4294–E4303
83 Nikolaev, Y. *et al.* (2019) Systems NMR: single-sample quantification of RNA, proteins and metabolites for biomolecular network analysis. *Nat. Methods* 16, 743
84 Tanay, A. and Regev, A. (2017) Scaling single-cell genomics from phenomenology to mechanism. *Nature* 541, 331–338
85 Lotfollahi, M. *et al.* (2019) scGen predicts single-cell perturbation responses. *Nat. Methods* 16, 715–721
86 Liu, J.J. *et al.* (2018) *In vivo* brain GPCR signaling elucidated by phosphoproteomics. *Science* 360, eaao4927
87 Moellering, R.E. and Cravatt, B.F. (2012) How chemoproteomics can enable drug discovery and development. *Chem. Biol.* 19, 11–22
88 Molina, D.M. *et al.* (2013) Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* 341, 84–87
89 Savitski, M.M. *et al.* (2014) Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* 346, 1255784
90 Choi, M. *et al.* (2017) ABRF Proteome Informatics Research Group (iPRG) 2015 Study: detection of differentially abundant proteins in label-free quantitative LC–MS/MS experiments. *J. Proteome Res.* 16, 945–957
91 Hogrebe, A. *et al.* (2018) Benchmarking common quantification strategies for large-scale phosphoproteomics. *Nat. Commun.* 9, 1–13
92 Donnelly, D.P. *et al.* (2019) Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat. Methods* 16, 587–594
93 Fetz, V. *et al.* (2016) Target identification by image analysis. *Nat. Prod. Rep.* 33, 655–667
94 Scheeder, C. *et al.* (2018) Machine learning and image-based profiling in drug discovery. *Curr. Opin. Syst. Biol.* 10, 43–52
95 Rudin, M. and Weissleder, R. (2003) Molecular imaging in drug discovery and development. *Nat. Rev. Drug Discov.* 2, 123–131

96 Weinstein, J.A. *et al.* (2019) DNA microscopy: optics-free spatio-genetic imaging by a stand-alone chemical reaction. *Cell* 178 (1), 229–241.e16
97 Boutros, M. *et al.* (2015) Microscopy-based high-content screening. *Cell* 163, 1314–1325
98 Smith, K. *et al.* (2018) Phenotypic image analysis software tools for exploring and understanding big image data from cell-based assays. *Cell Syst.* 6, 636–653
99 Webster, J.D. and Dunstan, R.W. (2014) Whole-slide imaging and automated image analysis: considerations and opportunities in the practice of pathology. *Vet. Pathol.* 51, 211–223
100 Zoffmann, S. *et al.* (2019) Machine learning-powered antibiotics phenotypic drug discovery. *Sci. Rep.* 9, 1–14
101 Bray, M.-A. *et al.* (2016) Cell painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* 11, 1757–1774
102 Simm, J. *et al.* (2018) Repurposing high-throughput image assays enables biological activity prediction for drug discovery. *Cell Chem. Biol.* 25, 611–618
103 Ahmad, G. and Amiji, M. (2018) Use of CRISPR/Cas9 gene-editing tools for developing models in drug discovery. *Drug Discov. Today* 23, 519–533
104 Thomas, R. *et al.* (1995) Dynamical behaviour of biological regulatory networks-I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bull. Math. Biol.* 57, 247–276
105 Chen, S. and Mar, J.C. (2018) Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinformatics* 19, 232
106 Munsky, B. *et al.* (2012) Using gene expression noise to understand gene regulation. *Science* 336, 183–187
107 Marbach, D. *et al.* (2012) Wisdom of crowds for robust gene network inference. *Nat. Methods* 9, 796–804
108 Zhu, J. *et al.* (2012) Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation. *PLoS Biol.* 10, e1001301
109 Argelaguet, R. *et al.* (2018) Multi-omics factor analysis—a framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* 14, e8124
110 Karr, J.R. *et al.* (2012) A whole-cell computational model predicts phenotype from genotype. *Cell* 150, 389–401
111 Jensen, K.J. *et al.* (2016) Network architecture predisposes an enzyme to either pharmacologic or genetic targeting. *Cell Syst.* 2, 112–121
112 Bersanelli, M. *et al.* (2016) Methods for the integration of multi-omics data: mathematical aspects. *BMC Bioinformatics* 17, S15
113 Le Novère, N. (2015) Quantitative and logic modelling of molecular and gene networks. *Nat. Rev. Genet.* 16, 146–158
114 Lavé, T. *et al.* (2007) Challenges and opportunities with modelling and simulation in drug discovery and drug development. *Xenobiotica* 37, 1295–1310
115 Dockendorf, M.F. *et al.* (2018) Leveraging model-informed approaches for drug discovery and development in the cardiovascular space. *J. Pharmacokinet. Pharmacodyn.* 45, 355–364
116 Agoram, B.M. (2009) Use of pharmacokinetic/pharmacodynamic modelling for starting dose selection in first-in-human trials of high-risk biologics. *Br. J. Clin. Pharmacol.* 67, 153–160
117 Zou, P. *et al.* (2012) Applications of human pharmacokinetic prediction in first-in-human dose estimation. *AAPS J.* 14, 262–281
118 Graaf, vander P.H. and Benson, N. (2018) The role of quantitative systems pharmacology in the design of first-in-human trials. *Clin. Pharmacol. Ther.* 104 797–797
119 Grimstein, M. *et al.* (2019) Physiologically based pharmacokinetic modeling in regulatory science: an update from the U.S. Food and Drug Administration's Office of Clinical Pharmacology. *J. Pharm. Sci.* 108, 21–25
120 Reddy, V.P. *et al.* (2018) Development, verification, and prediction of osimertinib drug–drug interactions using PBPK modeling approach to inform drug label. *CPT Pharmacomet. Syst. Pharmacol.* 7, 321–330
121 Vicente, J. *et al.* (2018) Mechanistic model-informed proarrhythmic risk assessment of drugs: review of the 'CiPA' initiative and design of a prospective clinical validation study. *Clin. Pharmacol. Ther.* 103, 54–66
122 Knibbe, C.A.J. and Danhof, M. (2011) Individualized dosing regimens in children based on population PKPD modelling: are we ready for it? *Int. J. Pharm.* 415, 9–14
123 Zhou, Q.-T. *et al.* (2017) Meropenem dosing based on a population pharmacokinetic–pharmacodynamic model in elderly patients with infection of the lower respiratory tract. *Drugs Aging* 34, 115–121
124 Hu, Z.-Y. *et al.* (2014) Physiologically based pharmacokinetic modeling of impaired carboxylesterase-1 activity: effects on oseltamivir disposition. *Clin. Pharmacokinet.* 53, 825–836
125 Hodgkin, A.L. and Huxley, A.F. (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544
126 O'Hara, T. *et al.* (2011) Simulation of the undiseased human cardiac ventricular action potential: model formulation and experimental validation. *PLoS Comput. Biol.* 7 e1002061
127 Zhao, Y. and Hu, Z.-Y. (2014) Physiologically based pharmacokinetic modelling and *in vivo* I/Ki accurately predict P-glycoprotein-mediated drug-drug interactions with dabigatran etexilate. *Br. J. Pharmacol.* 171, 1043–1053
128 Wangorsch, G. *et al.* (2011) Time-resolved in silico modeling of fine-tuned cAMP signaling in platelets: feedback loops, titrated phosphorylations and pharmacological modulation. *BMC Syst. Biol.* 5, 178
129 Deveaux, W. *et al.* (2019) Defining rules for cancer cell proliferation in TRAIL stimulation. *NPJ Syst. Biol. Appl.* 5, 1–8
130 de Pillis, L.G. *et al.* (2005) A validated mathematical model of cell-mediated immune response to tumor growth. *Cancer Res.* 65, 7950–7958
131 Trayanova, N.A. and Winslow, R. (2011) Whole-heart modeling. *Circ. Res.* 108, 113–128
132 Kuepfer, L. *et al.* (2016) Concepts in PBPK modeling: how to build a PBPK/PD model. *CPT Pharmacomet. Syst. Pharmacol.* 5, 516–531
133 Emoto, C. *et al.* (2019) A theoretical physiologically-based pharmacokinetic approach to ascertain covariates explaining the large interpatient variability in tacrolimus disposition. *CPT Pharmacomet. Syst. Pharmacol.* 8, 273–284
134 Jones, H.M. and Rowland-Yeo, K. (2013) Basic concepts in physiologically based pharmacokinetic modeling in drug discovery and development. *CPT Pharmacomet. Syst. Pharmacol.* 2, 63
135 Howell, B.A. *et al.* (2012) *In vitro* to *in vivo* extrapolation and species response comparisons for drug-induced liver injury (DILI) using DILIsym™: a mechanistic, mathematical model of DILI. *J. Pharmacokinet. Pharmacodyn.* 39, 527–541
136 Delporte, M.-L. *et al.* (2017) *Dose Selection in Entry into Human (EIH) Studies: Learnings from a Retrospective Survey on EiH Studies with Small Molecules Conducted between 2004 and 2016 at Hoffman-La Roche.* Roche
137 Okada, J. *et al.* (2015) Screening system for drug-induced arrhythmogenic risk combining a patch clamp and heart simulator. *Sci. Adv.* 1, e1400142
138 Jha, S.K. and Langmead, C.J. (2012) Exploring behaviors of stochastic differential equation models of biological systems using change of measures. *BMC Bioinformatics* 13, S8
139 Donnet, S. and Samson, A. (2013) A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamic models. *Adv. Drug Deliv. Rev.* 65, 929–939
140 Wang, Z. *et al.* (2015) Integrated PK-PD and agent-based modeling in oncology. *J. Pharmacokinet. Pharmacodyn.* 42, 179–189
141 Kather, J.N. *et al.* (2017) Silico modeling of immunotherapy and stroma-targeting therapies in human colorectal cancer. *Cancer Res.* 77, 6442–6452
142 Mortensen, S.B. *et al.* (2008) *Introduction to PK/PD Modelling - With Focus on PK and Stochastic Differential Equations.* Technical University of Denmark
143 Gabrielsson, J. (2016) *Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications* (5th edn), Apotekarsocieteten
144 Wang, H. *et al.* (2016) Evaluating a physiologically based pharmacokinetic model for predicting the pharmacokinetics of midazolam in Chinese after oral administration. *Acta Pharmacol. Sin.* 37, 276–284
145 Wong, H. and Chow, T.W. (2017) Physiologically based pharmacokinetic modeling of therapeutic proteins. *J. Pharm. Sci.* 106, 2270–2275
146 Niederalt, C. *et al.* (2018) A generic whole body physiologically based pharmacokinetic model for therapeutic proteins in PK-Sim. *J. Pharmacokinet. Pharmacodyn.* 45, 235–257
147 Glassman, P.M. and Balthasar, J.P. (2019) Physiologically-based modeling of monoclonal antibody pharmacokinetics in drug discovery and development. *Drug Metab. Pharmacokinet.* 34, 3–13
148 Luzon, E. *et al.* (2017) Physiologically based pharmacokinetic modeling in regulatory decision-making at the European Medicines Agency. *Clin. Pharmacol. Ther.* 102, 98–105
149 Sato, M. *et al.* (2017) Quantitative modeling and simulation in PMDA: a Japanese regulatory perspective. *CPT Pharmacomet. Syst. Pharmacol.* 6, 413–415
150 Baneyx, G. *et al.* (2014) Physiologically based pharmacokinetic modeling of CYP3A4 induction by rifampicin in human: Influence of time between substrate and inducer administration. *Eur. J. Pharm. Sci.* 56, 1–15
151 Wagner, C. *et al.* (2016) Predicting the effect of CYP3A inducers on the pharmacokinetics of substrate drugs using physiologically based pharmacokinetic (PBPK) modeling: an analysis of PBPK submissions to the US FDA. *Clin. Pharmacokinet.* 55, 475–483

152 Hanke, N. *et al.* (2018) PBPK models for CYP3A4 and P-gp DDI prediction: a modeling network of rifampicin, itraconazole, clarithromycin, midazolam, alfentanil, and digoxin. *CPT Pharmacomet. Syst. Pharmacol.* 7, 647–659
153 Johnson, T.N. *et al.* (2019) Development of a physiologically based pharmacokinetic model for mefloquine and its application alongside a clinical effectiveness model to select an optimal dose for prevention of malaria in young Caucasian children. *Br. J. Clin. Pharmacol.* 85, 100–113
154 Morcos, P.N. *et al.* (2018) Effect of hepatic impairment on the pharmacokinetics of alectinib. *J. Clin. Pharmacol.* 58, 1618–1628
155 Rhee, S. *et al.* (2017) Physiologically based pharmacokinetic modelling and prediction of metformin pharmacokinetics in renal/hepatic-impaired young adults and elderly populations. *Eur. J. Drug Metab. Pharmacokinet.* 42, 973–980
156 Tistaert, C. *et al.* (2019) Food effect projections via physiologically based pharmacokinetic modeling: predictive case studies. *J. Pharm. Sci.* 108, 592–602
157 Strauss, D.G. *et al.* (2019) Comprehensive *In vitro* Proarrhythmia Assay (CiPA) Update from a Cardiac Safety Research Consortium/Health and Environmental Sciences Institute/FDA Meeting. *Ther. Innov. Regul. Sci.* 53, 519–525
158 Pathmanathan, P. and Gray, R.A. (2018) Validation and trustworthiness of multiscale models of cardiac electrophysiology. *Front. Physiol.* 9, 106
159 Davies, M.R. *et al.* (2016) Recent developments in using mechanistic cardiac modelling for drug safety evaluation. *Drug. Discov. Today* 21, 924–938
160 Dutta, S. *et al.* (2017) Optimization of an in silico cardiac cell model for proarrhythmia risk assessment. *Front. Physiol.* 8, 106
161 Li, Z. *et al.* (2017) Improving the in silico assessment of proarrhythmia risk by combining hERG (Human Ether-à-go-go-Related Gene) channel–drug binding kinetics and multichannel pharmacology. *Circ. Arrhythm. Electrophysiol.* 10 (2), 1–12
162 Li, Z. *et al.* (2019) Assessment of an in silico mechanistic model for proarrhythmia risk prediction under the CiPA initiative. *Clin. Pharmacol. Ther.* 105, 466–475
163 ICH (2018) *Final Concept Paper ICH S7B and E14 Q&A, Endorsed by the MC with Support of the Assembly on 15.* ICH
164 Watkins, P.B. (2019) The DILI-sim initiative: insights into hepatotoxicity mechanisms and biomarker interpretation. *Clin. Transl. Sci.* 12, 122–129
165 Clausznitzer, D. *et al.* (2018) Quantitative systems pharmacology model for Alzheimer disease indicates targeting sphingolipid dysregulation as potential treatment option. *CPT Pharmacomet. Syst. Pharmacol.* 7, 759–770
166 Moore, H. (2018) How to mathematically optimize drug regimens using optimal control. *J. Pharmacokinet. Pharmacodyn.* 45, 127–137
167 Ribba, B. *et al.* (2018) Prediction of the optimal dosing regimen using a mathematical model of tumor uptake for immunocytokine-based cancer immunotherapy. *Clin. Cancer Res.* 24, 3325–3333
168 Milberg, O. *et al.* (2019) A QSP model for predicting clinical responses to monotherapy, combination and sequential therapy following CTLA-4, PD-1, and PD-L1 checkpoint blockade. *Sci. Rep.* 9, 1–17
169 Haefner, J.W. (2005) *Modeling Biological Systems: Principles and Applications.* Springer
170 Palacino, J. *et al.* (2015) SMN2 splice modulators enhance U1–pre-mRNA association and rescue SMA mice. *Nat. Chem. Biol.* 11, 511–517
171 Sivaramakrishnan, M. *et al.* (2017) Binding to SMN2 pre-mRNA-protein complex elicits specificity for small molecule splicing modifiers. *Nat. Commun.* 8, 1476
172 Ratni, H. *et al.* (2018) Discovery of risdiplam, a selective survival of Motor Neuron-2 (SMN2) gene splicing modifier for the treatment of spinal muscular atrophy (SMA). *J. Med. Chem.* 61, 6501–6517
173 Campagne, S. *et al.* (2019) Structural basis of a small molecule targeting RNA for a specific splicing correction. *Nat. Chem. Biol.* 15, 1191–1198
174 O'Leary, N.A. *et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745
175 Cléry, A. *et al.* (2011) Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-β1. *Nat. Struct. Mol. Biol.* 18, 443–450
176 Weinan, E. and Lu, J. (2011) Multiscale modeling. *Scholarpedia* 6, 11527
177 Weinan, E. (2011) *Principles of Multiscale Modeling.* Cambridge University Press
178 Marée, A.F.M. *et al.* (2006) Polarization and movement of keratocytes: a multiscale modelling approach. *Bull. Math. Biol.* 68, 1169–1211
179 Coveney, P.V. and Fowler, P.W. (2005) Modelling biological complexity: a physical scientist's perspective. *J. R. Soc. Interface* 2, 267–280
180 Walpole, J. *et al.* (2013) Multiscale computational models of complex biological systems. *Annu. Rev. Biomed. Eng.* 15, 137–154
181 Visser, S.A.G. *et al.* (2013) Model-based drug discovery: implementation and impact. *Drug Discov. Today* 18, 764–775
182 Milligan, P.A. *et al.* (2013) Model-based drug development: a rational approach to efficiently accelerate drug development. *Clin. Pharmacol. Ther.* 93, 502–514
183 Benfenati, E. *et al.* (2011) The acceptance of in silico models for REACH: requirements, barriers, and perspectives. *Chem. Cent. J.* 5, 58
184 Mariathasan, S. *et al.* (2018) TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* 554, 544–548
185 Dai, J. *et al.* (2019) Acetylation blocks cGAS activity and inhibits Self-DNA-induced autoimmunity. *Cell* 176 (6), 1447–1460.E14
186 Drews, J. (2000) Drug discovery: a historical perspective. *Science* 287, 1960–1964
187 Gerald, M.C. (2013) *The Drug Book: From Arsenic to Xanax, 250 Milestones in the History of Drugs.* Sterling
188 Hie, B. *et al.* (2018) Realizing private and practical pharmacological collaboration. *Science* 362, 347–350
189 Choobdar, S. *et al.* (2019) Assessment of network module identification across complex diseases. *Nat. Methods* 16, 843–852
190 Shahriari, B. *et al.* (2016) Taking the human out of the loop: a review of Bayesian optimization. *Proc. IEEE* 104, 148–175
191 Chmiela, S. *et al.* (2018) Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* 9, 1–10
192 Regev, A. *et al.* (2017) Science forum: the human cell atlas. *eLife* 6, e27041
193 Burgess, D.J. (2019) Spatial transcriptomics coming of age. *Nat. Rev. Genet.* 20, 317
194 Satpathy, A.T. *et al.* (2019) Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* 37, 925–936
195 Smits, A.H. *et al.* (2019) Biological plasticity rescues target activity in CRISPR knock outs. *Nat. Methods* 16, 1087–1093
196 Schneider, G. (2018) Automating drug discovery. *Nat. Rev. Drug Discov.* 17, 97–113
197 Lyu, J. *et al.* (2019) Ultra-large library docking for discovering new chemotypes. *Nature* 566, 224
198 Hoffmann, R. and Malrieu, J-P. (XXXX) Simulation vs understanding a tension, in quantum chemistry and beyond. Part A: stage setting. *Angew. Chem. Int. Ed.* https://doi.org/10.1002/anie.201902527.
199 Hoffmann, R. and Malrieu, J-P. (XXXX) Simulation vs understanding a tension, in quantum chemistry and beyond. Part B: the march of simulation, for better or worse. *Angew. Chem. Int. Ed.* https://doi.org/10.1002/anie.201910283.
200 Hoffmann, R. and Malrieu, J-P. (XXXX) Simulation vs understanding a tension, in quantum chemistry and beyond. Part C: toward consilience. *Angew. Chem. Int. Ed.* https://doi.org/10.1002/anie.201910285.
201 Gaulton, A. *et al.* (2017) The ChEMBL database in 2017. *Nucleic Acids Res.* 45, D945–D954