Biochemical Society Transactions (2022) **50** 747–758
https://doi.org/10.1042/BST20200967

PORTLAND PRESS

Review Article

Drug repurposing *in silico* screening platforms

Jonathan G.L. Mullins[1,2]

[1]Genome and Structural Bioinformatics, Institute of Life Science, Swansea University Medical School, Singleton Park, Swansea SA2 8PP, U.K.; [2]Moleculomics Ltd, Institute of Life Science, Swansea University Medical School, Singleton Park, Swansea SA2 8PP, U.K.

Correspondence: Jonathan G.L. Mullins (j.g.l.mullins@swansea.ac.uk)

Over the last decade, for the first time, substantial efforts have been directed at the development of dedicated *in silico* platforms for drug repurposing, including initiatives targeting cancers and conditions as diverse as cryptosporidiosis, dengue, dental caries, diabetes, herpes, lupus, malaria, tuberculosis and Covid-19 related respiratory disease. This review outlines some of the exciting advances in the specific applications of *in silico* approaches to the challenge of drug repurposing and focuses particularly on where these efforts have resulted in the development of generic platform technologies of broad value to researchers involved in programmatic drug repurposing work. Recent advances in molecular docking methodologies and validation approaches, and their combination with machine learning or deep learning approaches are continually enhancing the precision of repurposing efforts. The meaningful integration of better understanding of molecular mechanisms with molecular pathway data and knowledge of disease networks is widening the scope for discovery of repurposing opportunities. The power of Artificial Intelligence is being gainfully exploited to advance progress in an integrated science that extends from the sub-atomic to the whole system level. There are many promising emerging developments but there are remaining challenges to be overcome in the successful integration of the new advances in useful platforms. In conclusion, the essential component requirements for development of powerful and well optimised drug repurposing screening platforms are discussed.

Introduction

Drug repurposing possesses great advantages in terms of establishing new therapeutic options for specific disorders by exploiting drugs used for other conditions that are already known to be largely safe. The specific advantages include leveraging pre-existing safety and clinical data to avoid lengthy Phase I safety studies and therefore entering clinical trials at Phase IIa. This not only reduces development times, and thus costs, but also removes some of the risks [1]

The notion of repurposing of existing drugs brings associated reservations for many too. There is the concern that advances rely too much on serendipity and convenience. This is countered by the acknowledgement that many of the recent repurposing successes, not least for the treatment of Covid-19 [2], have come about due to trials based on extensive prior clinical observation of patients receiving different drug treatments concurrently for other disorders. Indeed, trials often proceed on the basis of straightforward clinical rationale or hunch.

Phenotypic drug discovery has in broad terms been highly successful. The process closely simulates the normal physiological situation. It allows conception and testing of therapeutic relevance early in the drug discovery process and does not require knowledge of the molecular mechanism of action. However, modelling of more complex diseases is difficult. The chance of serendipitous discovery is enhanced but there is a low chance of developing 'best in class' molecules and it is not possible to utilise high technology platforms and so screening capacity is low [3].

Received: 9 December 2021
Revised: 8 February 2022
Accepted: 21 February 2022

Version of Record published:
14 March 2022

Biochemical Society Transactions (2022) **50** 747–758
https://doi.org/10.1042/BST20200967

Target-based drug discovery incorporates *a priori* knowledge or elucidation of the mechanism of action, so though the findings may not be as relevant to the physiological situation, there is a very high chance of developing 'best in class' molecules. Screening of vast chemical libraries and utilisation of high technology platforms is feasible [3].

With repurposing, the reservations of molecular scientists are centred on the narrow coverage of chemical space offered by the body of approved drug compounds; and the nagging feeling, and perhaps ultimately certain logic, that repurposed drugs, by definition, cannot represent the best possible compound for the secondary therapeutic purpose, unless the pharmacological target receptors are the same or at least similar to those engaged in the primary therapeutic purpose.

The field of the development and use of *in silico* repurposing drug screening platforms thereby lies at the meeting point of these two sets of drivers — acknowledgement of the growing pressure for convenient new therapies, being balanced by the need for maximal exploration and scrutiny of the potential alignment of the predicted molecular interactions with those known to be needed for therapeutic efficacy. The challenge for platform development is to firstly encompass and properly define the full and appropriate chemical and biological space for a given set of drug–receptor interactions, and then secondly to target these interactions with the most suitable computational algorithms that secure a practically meaningful indicator of drug efficacy. The recent drug repurposing successes are in no doubt very encouraging developments, but lead us back to the challenge of ensuring that we are not overly being driven by convenient selection of drug candidates for the purpose of expediency at a particular time, rather than the proper objective of identification of the single best drug repurposing option at the molecular level.

This review will reflect on the development of *in silico* drug repurposing tools of recent years, focusing on molecular docking challenges, network-based methodologies, systems that integrate complex information networks, and the development of generic platform technologies.

Development of *in silico* drug repurposing tools

A number of *in silico* tools have been developed to down-screen a field of potential candidate target molecules and to discover novel targets [4]. Semantic mining has been applied in the development of the DReSMin (Drug Repositioning Semantic Mining) system [5]. A small number of repurposing efforts have focused on physiological responses, these methods normally utilising the side effect data from SIDER [6] as a primary source of the known side effects. However, the main thrust of work in the realm of *in silico* drug repurposing is directed at exploiting knowledge of molecular interactions and building robust frameworks for appropriate and validated approaches while further expanding the underpinning knowledgebase.

High quality 3D protein structures can be obtained directly from the experimentally determined structures of the Protein Databank [7,8] where applicable, and otherwise can be modelled by a range of established methodologies that have advanced substantially in recent years, excellently reviewed by Kuhlman and Bradley [9]. The powerful neural network-driven protein structure predictions of AlphaFold and RoseTTAFold are exciting advances [10,11]. The compound sets have to be appropriately extensive and the algorithms and computational power sufficiently powerful. The primary chemical set, representing the 'low-hanging fruit' is the FDA-approved drug set, just over 4200 compounds as of November, 2021, just over 2700 of which are approved small molecule drugs, as listed in DrugBank [12].

Databases

Somewhat wider compound sets are sometimes used, such as the ReFRAME collection of 12 000 compounds [13], a 'best-in-class' drug repurposing library containing nearly all small molecules that have reached clinical development or undergone significant preclinical profiling. ReFRAME (Repurposing, Focused Rescue, and Accelerated Medchem) that was assembled by combining three widely used commercial drug competitive intelligence databases (Clarivate Integrity, GVK Excelra GoStar, and Citeline Pharmaprojects), together with extensive patent mining of small molecules that have been dosed in humans. The practical utility of this collection was illustrated by its screening against Cryptosporidium spp., a major cause of childhood diarrhoea in the developing world, and the resulting identification of two active compounds previously tested in humans for other therapeutic indications, and subsequently shown to be efficacious at clinically relevant doses in animal models of Cryptosporidium infection. An open-access data portal has been developed to share assayed screen hits (https://reframedb.org).

Pharmaceutical companies and start-ups have advanced drug discovery by extracting hitherto intractable patterns from biomedical data [14]. IBM's Watson Health platform searches vast volumes of textual data, including laboratory data, clinical reports, and scientific publications [15].

Algorithms for identifying drug–target interactions

Drug discovery and repurposing is primarily based on identification of new drug–target interactions. Many drugs are non-specific and show reactivity to additional targets besides their primary targets. Repurposing is made simpler if accurate structural prediction of the drug targets can be achieved.

Experimental confirmation of interactions is expensive and lengthy, and the resources are best focussed where there is the highest probability of success. This has underpinned the attraction of developing computational methods to predict potential drug–target interactions.

Wang et al. [16], developed ACID, a tool for drug repurposing using a consensus inverse docking strategy. The computational protocol combines several significantly different types of free docking methodologies into a consensus inverse docking (CID) scheme (i.e. different conformational search algorithm, different global and local optimisers, and different scoring functions), namely AutoDock [17], AutoDock Vina [18], DOCK [19], LEDOCK, PLANTS [20], and PSOVina [21] for binding pose search. The CID was optimised by use of collections of crystal complexes (from PDB-bind [22]) and binding data. Molecular Mechanics/Poisson–Boltzmann Surface Area (MM/PBSA) and X-SCORE were used for the final binding energy calculation to avoid bias towards the intrinsic scoring functions of the individual docking methodologies. The workflow demonstrated a ~10% enhancement in posing accuracy and prediction of binding modes compared with previous best method, which shows the benefit of well-constructed consensus methods. A web server — the auto *in silico* consensus inverse docking (ACID) (http://chemyang.ccnu.edu.cn/ccb/server/ACID/) was designed based on this workflow, incorporating the CID workflow program, the compound database of 2086 approved drugs with therapeutic information and a known target database containing 831 protein structures covering 30 therapeutic areas.

The Computational Analysis of Novel Drug Repurposing Opportunities (CANDO) platform was developed for shotgun multitarget drug discovery, repurposing, and design [23,24]. The platform screens and ranks the approved drugs for every applicable disease/indication through large scale modelling and analysis of interactions between comprehensive libraries of compounds and protein structures using hierarchical fragment-based docking with dynamics, CANDOCK [25]. The comparison of drug-proteome signatures and ranking approach yields benchmarking accuracies of 20–40% for ~1500 indications relative to random control accuracies of

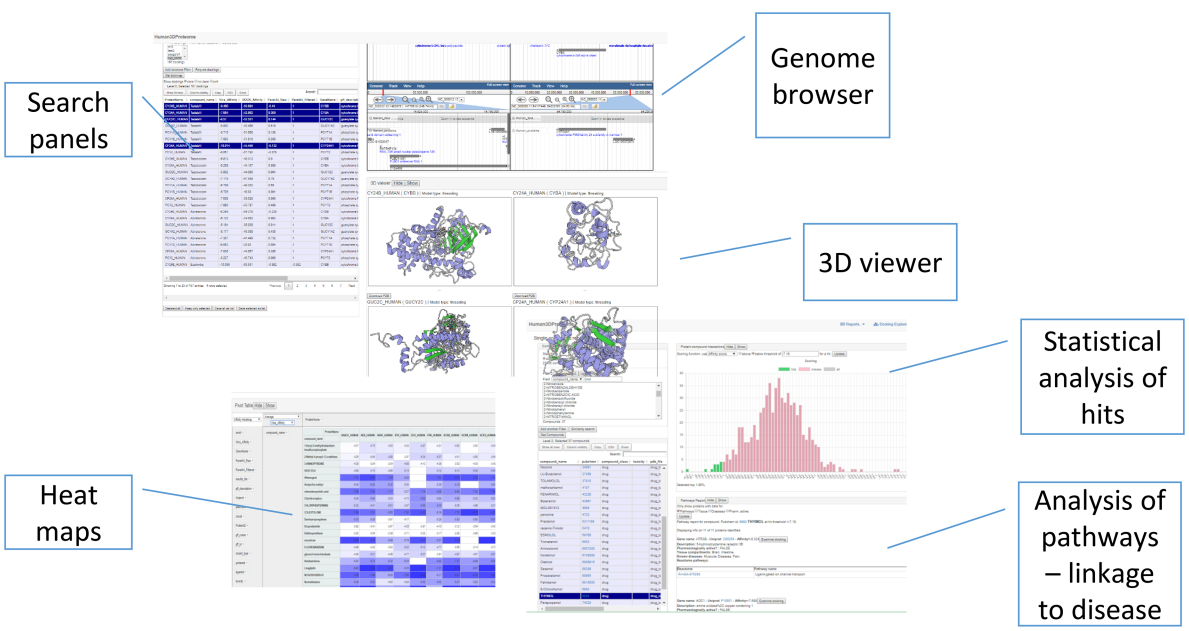

Figure 1. The Re-Drug screening platform user interface, showing the main views and functionalities and their synchronous use.

Biochemical Society Transactions (2022) 50 747–758
https://doi.org/10.1042/BST20200967

2–15%, when combined with *in vitro* validation studies. The 35% top ranking predictions had comparable or better activity than existing drugs across ten indications, identifying novel repurposed therapies for conditions such as dengue, dental caries, diabetes, herpes, lupus, malaria, and tuberculosis.

Combination of databases and drug–target interaction algorithms

The team of the author at Swansea have developed Re-Drug (Drug Repurposing Screening Platform), presented to the Biochemical Society Drug Repurposing meeting in Birmingham, U.K. in November, 2019 [26]. Re-Drug is a comprehensive Artificial Intelligence-based screening technology for the assessment of functionally feasible on-target and off-target interactions of a candidate therapeutic drawn from the pool of ∼1400 FDA approved drugs and screened for repositioning opportunities across more than 1400 known drug targets. This whole system framework provides a systematic assessment of the candidate drug's predicted behaviour based upon comparison with the established mechanistic data (molecular binding and metabolic and signalling pathway analysis via Reactome [27] and KEGG [28] for a broad range of currently used drugs, encompassing both drug efficacy and potential toxicity. Based on extraction from docking data across the 1400 drug receptor targets, interchangeable compounds with highly similar docking and pathway profiles are identified (Figure 1).

Multiple parameters of binding derived from docking simulations (including binding affinity, Van der Waals, solvent accessible surface area, rotatable bond entropy, atom distances, and H bonds) are used as training data for machine learning classifiers. A combination of docking methods is used (AutoDock Vina [29] and DOCK 6 [19] (http://dock.compbio.ucsf.edu/) in order to capture the advantages of both empirical (affinity) and force field (energy) approaches [18,19], and these consensus methods generate metascores with improved prediction accuracy.

For a drug of interest, the system reports known receptor targets, detailing the relevant pathways implicated, and identifying by the machine learning applied to *in silico* docking data, the most functionally similar alternative compounds in terms of binding parameters, based on specific comparison with the interactions of the drug of interest at those particular targets. Analysis of the indications and associated pathways of the implicated alternative drugs reveals new potential therapeutic areas for the drug of interest.

A measure is also provided of the similarity of molecular docking profiles of the drug of interest across the wider field of *in silico* docking data involving 1000+ receptor targets, allowing analysis of top-scoring new targets and their established ligands. Protein–ligand interactions with highly similar binding parameters and pathway profiles are predicted to be more highly functionally interchangeable and to present promising repurposing opportunities.

Molecular docking challenges

The challenge is that known drug–target interactions are relatively rare, and when the performance of prediction is being assessed for drugs without relevant target interaction information, validation by comparison with controls in the form of negative results samples is impossible because of the scarcity of experimentally verified negative examples [30].

Much work has involved using convolutional neural network (CNN) models trained on 3D structural information of protein–ligand complexes to distinguish binding from non-binding ligands for virtual screening. The paucity of reliable protein–ligand X-ray structures and binding affinity data has driven the generation of constructed datasets for the training and evaluation of neural net molecular recognition models, established sets of actives and decoys for benchmarking and training.

Decoy libraries may be used, such as the Directory of Useful Decoys Enchanced (DUD-E) that has often been used as a primary training set [31]. DUD-E is a benchmark comprised of 102 protein targets with active molecules and decoy data. The active-decoys ratio per target on DUD-E varies and mimics a real screening scenario in drug discovery; though for some targets only a few active molecules are available (as low as 40) while the number of decoys may be as high as 30 000. Therefore, though testing learning models on DUD-E may reveal the usefulness of the models in real virtual screening experiments, there is inherent bias, meaning that methods often reflect these biases in separation of actives and decoys, and do not necessarily learn to perform molecular recognition.

Chen et al. [32] investigated sources of bias in DUD-E to assess whether CNN models developed using DUD-E are properly learning the underlying physics that drives molecular recognition, or are instead reflecting biases inherent in the dataset itself, and found that superior enrichment efficiency in CNN models can be attributed to the analogue and decoy bias hidden in the DUD-E dataset rather than successful generalisation of the pattern of protein–ligand interactions. Comparing deep learning models trained on PDBbind datasets, they

found that enrichment performance using DUD-E was similar to that obtained just using AutoDock Vina, suggesting caution in applying constructed datasets to machine learning based methodology development. This fundamental issue hinders virtual screening method development.

A potential way forward is to filter DUD-E decoy sets and reduce reliance on a single source of decoys. This can be done by cross-referencing with the binding data (i.e. K_i, K_d, IC_{50}, and ΔG) of ligands of each protein target from chemical databases such as ChEMBL [33]. Compounds with affinities higher than 1.0 μM are considered active while any compounds with no measurable affinity up to 30 μM can be classified as experimental decoys. Decoy libraries can be generated from the ZINC database [34]. For each ligand the major protonation states were calculated and the molecular weight (MW), lipophilicity (AlogP), the number of hydrogen bond donors (HBD) and acceptors (HBA), number of rotatable bonds (nRotb) and net charges can be calculated. For each state, a set of property-matched decoys can then be generated. Similarity analysis can be conducted using Extended Connectivity Fingerprints 6 (ECFP6) fingerprints and the most dissimilar decoys retained and duplicates removed. ECFP6 and FCFP6 fingerprints can be generated using the RDKit cheminformatics platform (http://www.rdkit.org) module.

A deep learning method (DeepCoy) has been developed that generates decoys to a user's preferred specification in order to remove such biases or construct sets with a defined bias [35]. DeepCoy was validated using two established benchmarks, DUD-E and DEKOIS 2.0. The generated decoy molecules more closely matched the active molecules' physicochemical properties while introducing no additional risk of false negatives.

Network-based methodologies

Many methods for predicting potential drug–target interactions use a network representation. Network-based models are particularly suited to addressing this type of problem as they focus on interactions between separate data types, in this case, an interaction network can be constructed where nodes represent drugs and targets, and edges denote interactions. The interactions are derived from associations that may be pharmacological, clinical or molecular.

Network-based methodologies have been directed at screening of potential new indications of those FDA-approved drugs with well-characterised pharmacokinetics and pharmacodynamics and known safety and tolerability profiles by exploiting the known relationships between drug targets and diseases [36–38]. MFM (method of functional modules) approaches targeting breast, prostate, and leukaemia cancers [39] have been developed using resources such as the Functional Linkage Network, a drug response expression data set (The Library of Integrated Network-Based Cellular Signatures (LINCS) profiles) [40], CMap [41], DrugBank [12], OMIM [42], GEO [43], and The Cancer Genome Atlas (TCGA) portal [44].

The network of drug–target interactions can be extended by overlaying knowledge of protein–protein similarity and drug–drug similarity, resulting in powerful clustering methodologies that outperform the earlier models. If a given drug has known targets, other candidate targets can be ranked by measuring the structural similarity between them and the known targets. Similarly, drug analogy can be considered, the potential targets of the given drug are selected based on the target information of similar drugs. In this regard, the work of Yamanishi et al. [45] has demonstrated that if two drugs have similar structure, the chances of interaction with similar target proteins will be higher; and that two target proteins with high sequence similarity are more likely to interact with similar drugs. This approach can also widen the pool of negative controls.

For many in the field, chemical structures are compared in high throughput to find exact and similar matches and similar matches using packages such as Open Babel [46]. Similarity analysis consists of comparing the number of structural features shared between a reference molecule and test molecules in a database [47,48]. The degree of similarity between structures is calculated using a metric that considers how close or far a test molecule is to the reference set [48,49].

The SIMCOMP method is the main similarity measurement used in the Yamanishi studies, however, it can only be used for compounds which contain a KEGG database reference. Fingerprinting is therefore used more widely. For similarity searches on compound fingerprints, such as the PubChem and MACCS fingerprint formats, or the ECFP6 fingerprints, the most commonly used similarity metric is still the Tanimoto coefficient [47], despite its simplicity and limitations, such as the dependence on the size of the molecules and not considering the frequency of the features in the molecules under analysis [48–50].

Protein similarity is determined by comparison of amino acid sequences using methods such as the Smith and Waterman algorithm, developed in 1981 to perform local sequence alignment, which give high scores for

Biochemical Society Transactions (2022) 50 747–758
https://doi.org/10.1042/BST20200967

matching amino acid strings while penalising mismatches in sequence and length. The motivation behind these algorithms is to provide an alternative to comparing the entire protein sequence, instead focusing on the comparison of specific regions which can be of varying length to detect regions of the protein which are similar. These similar regions are more synonymous with similar functional domains. The Basic Local Alignment Search Tool (BLAST) [51] is the most widely used similarity tool. It is an open source heuristic algorithm developed by NCBI which generates a partially filled similarity matrix of high scoring matches. BLAST is a very powerful tool, and considered to be more efficient in terms of computation time and resource usage, however at a cost of thoroughness as BLAST will not score or report patterns which are more difficult to detect, that is those at more remote homology, which can often be the case when comparing human proteins that may be very different in sequence, even in global fold, but bind the same ligand.

Deep Learning Methods for Drug–Target Interaction Prediction encompass studies that predict compound properties [52,53]; target prediction for existing drugs, such as reverse docking simulation [54]; text mining approaches [55], literature-based medical knowledge graph methods [56], and drug binding using molecular dynamics simulation [57–59]. Deep learning has shown better performance than classic machine learning methods in repurposing of antibiotics [60] and more widely [61].

Systems that integrate complex information in COVID-19 drug repurposing

Researchers are increasingly realising the scope of integrating the different tiers of knowledge of the complex networks connecting drugs, targets, and diseases within *in silico* pipelines. The Covid-19 pandemic has served as a great driver of innovation, due to the pressing need for new routes to development of prevention and treatment strategies, by prioritisation of existing drugs for expedition to therapeutic development for COVID-19. Zeng et al. [62] built on their work on the FDA-approved drug set and developed a highly integrative network-based deep-learning framework, CoV-KGE, that employed a comprehensive knowledge graph including 15 million edges encompassing 39 relationship types linking drugs, diseases, proteins and genes, pathways, and expression from 24 million PubMed publications to identify repurposable drugs for COVID-19 (termed CoV-KGE). The framework identified 41 repurposable drugs (including dexamethasone, indomethacin, niclosamide, and toremifene) whose therapeutic associations with COVID-19 were validated by data from ongoing clinical trials, transcriptomic and proteomics data in SARS-CoV-2-infected human cells [62]. This demonstrates highly effective use of a powerful deep-learning methodology.

Dexamethasone, an approved glucocorticoid, prescribed for a diverse range of inflammatory and autoimmune conditions, was identified as a prime candidate for repurposing by CoV-KGE. The subsequent randomised COVID-19 therapy trial indicated that dexamethasone reduced death by a fifth in individuals requiring oxygen but not invasive mechanical ventilation, and by a third for more seriously ill patients who required invasive mechanical ventilation [63].

Sosa et al. [56] implemented a medical knowledge graph containing information from the biomedical literature relating to drugs, diseases, genes, and proteins and predicted links between drugs and diseases using graph embedding techniques. A knowledge graph developed by BenevolentAI forms a large repository of structured medical information, with machine learning used to extract connections from the scientific literature. In this way, a drug routinely used to treat rheumatoid arthritis, Baricitinib, that works through inhibition of AP2-associated protein kinase 1, was identified as a potential treatment for COVID-19 [64].

A network-based methodology that focuses on the virus–host interactome and related human drug targets was applied to drugs that could be repurposed and implicated 16 candidates for potential treatment of COVID-19 [65], including as a top candidate, toremifene, a first-generation non-steroidal selective oestrogenic receptor modulator approved for the treatment of breast cancer. *In vitro* work demonstrated that micromolar concentrations of toremifene blocked a range of viral infections, including Middle East Respiratory Syndrome coronavirus (MERS), severe acute respiratory syndrome coronavirus (SARS-CoV), and SARS-CoV-2 [66]. Beck et al. [67] developed 'Molecule Transformer-Drug Target Interaction' to ascertain whether any already available antiviral drugs could be potentially effective against SARS-CoV-2 infection and identified five, namely atazanavir, dolutegravir, efavirenz, remdesivir and ritonavir . Machine learning and statistical analysis approaches were used as part of the discovery that a poly-ADP-ribose polymerase 1 inhibitor, mefuparib (CVL218), blocked SARS-CoV-2 replication [68]. A network-based approach indicated that a combination of melatonin and toremifene held potential for use in the treatment of COVID-19 [69].

Network-based prediction of drug–target interaction has been shown to improve existing association prediction methods for measuring the topological similarities of bipartite (drug and target networks), extended to tripartite linked networks (drug, target, and disease networks) [59,70–72]. Zong et al. [55] employed tripartite networks through the DeepWalk method [73] to identify topology-based similarities, and revealed the potential of this method as a drug repurposing solution.

The advantage of these whole system approaches is that they can be used to address questions from the starting point of the drug of interest, the protein target of interest, the mechanism of disease or the disease itself,

Table 1 Computational tools used in drug repurposing

Tool	Database	Molecular docking/drug–target interaction	Platform
ksRepo [4]			✓
DReSMin [5]	✓		
SIDER [6]	✓		
DrugBank [12]	✓		
ReFRAME [13]	✓		
ACID [16]		✓	
AutoDock4/AutoDockTools4 [17]		✓	
AutoDock Vina [18]		✓	
DOCK 6 [19]		✓	
PLANTS [20]		✓	
PSOVina [21]		✓	
CANDO [23,24]			✓
CANDOCK [25]		✓	
Re-Drug [26]	✓	✓	✓
Reactome [27]	✓		
KEGG [28]	✓		
AutoDock Vina 1.2.0 [29]		✓	
DUD-E [31]	✓		
ChEMBL [33]	✓		
ZINC [34]	✓		
RDKit			✓
DeepCoy [35]			✓
MFM [39]	✓		
LINCS [40]	✓		
CMap [41]	✓		
OMIM [42]	✓		
GEO [43]	✓		
TCGA [44]	✓		
OpenBabel [46]		✓	
CoV-KGE [62]	✓		✓
BenevolentAI [64]	✓		✓
Molecule Transformer-Drug Target Interaction [67]		✓	
DeepWalk [73]		✓	
L1000FWD [74]	✓		✓
L1000CDS^2	✓		✓

Biochemical Society Transactions (2022) **50** 747–758
https://doi.org/10.1042/BST20200967

such as; (i) Repositioning approved drugs for new targets of interest; (ii) Identification of combinations of downstream pathway events and therapeutic actions for repositioning of drugs targeting multiple sites and effects; (iii) Identification of other marketed drugs with similar pharmacological profiles to that of a drug currently prescribed for a specific indication; (iv) Opportunities for 'therapeutic switching' by prediction of clinical effects of repurposed drug candidates based on the well characterised *in vivo* effects of drugs with similar profiles.

Polypharmacology

Instances of polypharmacology, where one drug is known to have multiple protein targets, i.e. many therapeutic small molecules that by their nature have a broader protein binding site specificity, serve as a natural foundation for further drug repurposing. Large-scale databases have been accumulated in recent years, such as the next-generation L1000-based Connectivity Map (CMap), a transcriptome database collecting gene-expression profiles of drug-treated human cancer cells (https://portals.broadinstitute.org/cmap/), mostly focussed on the FDA-approved drugs, and subsequently integrated in an analytic Web platform, the L1000FWD, for systematic analyses of polypharmacology and drug repurposing [74]; and L1000CDS^2 (https://maayanlab.cloud/L1000CDS2). For L1000FWD, two different classes of anti-cancer drugs provided proof-of-concept examples, namely histone deacetylase (HDAC) inhibitors and topoisomerase inhibitors. The study identified KM-00927 and BRD-K75081836 as novel HDAC inhibitors and mitomycin C as a topoisomerase IIB inhibitor. This is an excellent example of researchers integrating freely available public resources in systematic polypharmacology analysis and drug repurposing platform development.

The wealth of computational tools reviewed in this paper are detailed in Table 1, categorised according to whether they are primarily database tools, tools focused on molecular docking and/or drug–target interaction, platforms or a combination.

Platform requirements

In conclusion, the component requirements for development of powerful and well optimised drug repurposing screening platforms, that are the best possible representation of what can be achieved in the 2020's and on into

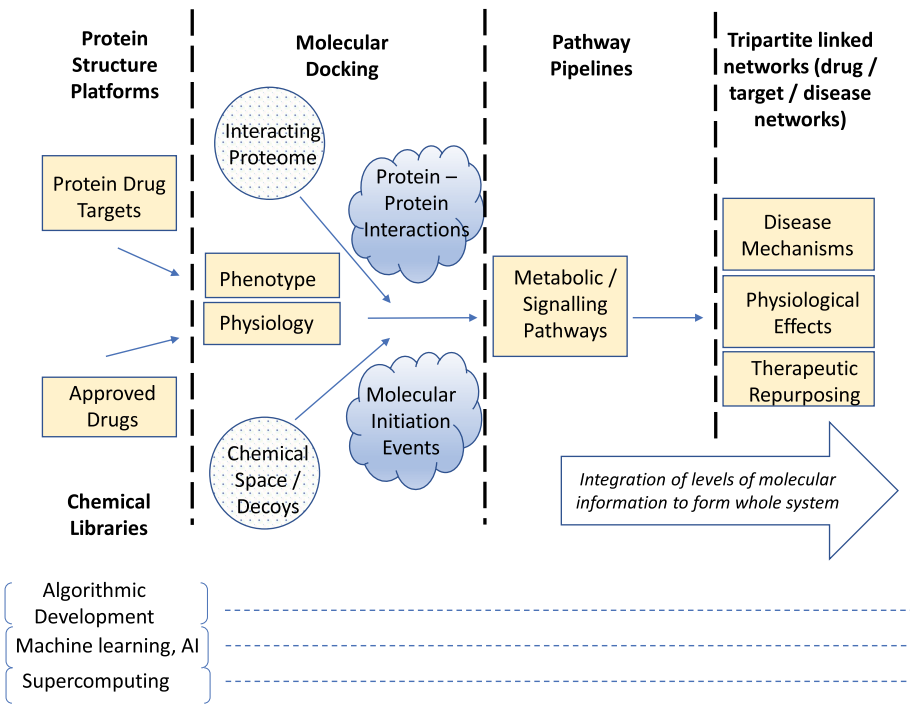

Figure 2. Systems approach drug repurposing screening platform, showing the main components and processes and their integration, underpinned by continuing algorithmic development, AI and ever increasing processing power.

the second quarter of the century, will include: (i) Relevant and 'clean' compound and protein structure data sets providing the most generic applicability possible; (ii) Selection of the most suitable and advanced molecular docking methodologies combined with machine learning or deep learning approaches that are continually validated and their performance meaningfully benchmarked against clinical and experimental data; (iii) Integration of metabolic and signalling pathway data and disease networks, such as into tripartite linked networks (drug/target/disease networks); (iv) Selection of the best AI algorithms; all incorporated within a dedicated systems model (Figure 2).

Looking forward, the burgeoning rate of current advance in applicable network-based and deep learning algorithm development [75–80] bodes well for better identification of repurposing opportunities. The challenge is to combine sufficiently technically robust approaches that are availed of the substantial advances in AI and network medicine, with the most appropriate selection of protocols and input training parameters.

Another big challenge to rise to is that we ensure that these components are fully and seamlessly integrated within functional, usable web-based platforms that are accessible to a wide population of researchers with little prior understanding or expertise in algorithm or platform development but who are seeking to acquire rapid knowledge relating to a specific repurposing possibility or even just seeking to prospectively view the field of potential opportunity. Technically simple, but rarely done properly, this is most vital if we are to leverage maximal benefit from the approved drugs that are readily available to better treat human disease.

Perspectives

- The field of development and use of *in silico* drug repurposing screening platforms lies at the meeting point of two sets of drivers — (i) acknowledgement of the growing pressure for convenient new therapies, being balanced by (ii) the need for maximal exploration and scrutiny of the potential alignment of the predicted molecular interactions with those known to be needed for therapeutic efficacy.
- Recent advances in algorithm development and artificial intelligence have shown great promise in identifying real-world repurposing opportunities and researchers are successfully integrating different types of data, such as that pertaining to metabolic and signalling pathway data and disease mechanisms, into linked networks.
- There will be a movement towards systems approach-based platform development that will yield powerful and generically applicable technology solutions dedicated to the unique challenge of drug repurposing.

Competing Interests

The author declares that there are no competing interests associated with this manuscript.

Abbreviations

ACID, auto *in silico* consensus inverse docking; BLAST, Basic Local Alignment Search Tool; CANDO, Computational Analysis of Novel Drug Repurposing Opportunities; CID, consensus inverse docking; CMap, Connectivity Map; CNN, convolutional neural network; DUD-E, Directory of Useful Decoys Enchanced; ECFP6, Extended Connectivity Fingerprints 6; HDAC, histone deacetylase; LINCS, Library of Integrated Network-Based Cellular Signatures; SARS-CoV, severe acute respiratory syndrome coronavirus; TCGA, The Cancer Genome Atlas.

References

1 Kepplinger, E.E. (2015) FDA's expedited approval mechanisms for new drug products. *Biotechnol. Law Rep.* **34**, 15–37 https://doi.org/10.1089/blr.2015.9999

2 Farne, H., Kumar, K., Ritchie, A.I., Finney, L.J., Johnston, S.L. and Singanayagam, A. (2020) Repurposing existing drugs for the treatment of COVID-19. *Ann. Am. Thorac. Soc.* **17**, 1186–1194 https://doi.org/10.1513/AnnalsATS.202005-566FR

Biochemical Society Transactions (2022) **50** 747–758
https://doi.org/10.1042/BST20200967

3 Kiriiri, G.K., Njogu, P.M. and Mwangi, A.N. (2020) Exploring different approaches to improve the success of drug discovery and development projects: a review. *Future J. Pharm. Sci.* **6**, 27 https://doi.org/10.1186/s43094-020-00047-9
4 Brown, A.S., Kong, S.W., Kohane, I.S. and Patel, C.J. (2016) Ksrepo: a generalized platform for computational drug repositioning. *BMC Bioinformatics* **17**, 78 https://doi.org/10.1186/s12859-016-0931-y
5 Mullen, J., Cockell, S.J., Tipney, H., Woollard, P.M. and Wipat, A. (2016) Mining integrated semantic networks for drug repositioning opportunities. *PeerJ* **4**, e1558 https://doi.org/10.7717/peerj.1558
6 Kuhn, M., Letunic, I., Jensen, L.J. and Bork, P. (2016) The SIDER database of drugs and side effects. *Nucleic Acids Res.* **44**, D1075–D1079 https://doi.org/10.1093/nar/gkv1075
7 Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H. et al. (2000) The protein data bank. *Nucleic Acids Res.* **28**, 235–242 https://doi.org/10.1093/nar/28.1.235
8 Burley, S.K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G.V. et al. (2021) RCSB protein data bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.* **49**, D437–D451 https://doi.org/10.1093/nar/gkaa1038
9 Kuhlman, B. and Bradley, P. (2019) Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Biol.* **20**, 681–697 https://doi.org/10.1038/s41580-019-0163-x
10 Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O. et al. (2021) Highly accurate protein structure prediction with alphaFold. *Nature (London)* **596**, 583–589 https://doi.org/10.1038/s41586-021-03819-2
11 Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R. et al. (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 https://doi.org/10.1126/science.abj8754
12 Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R. et al. (2018) Drugbank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 https://doi.org/10.1093/nar/gkx1037
13 Janes, J., Young, M.E., Chen, E., Rogers, N.H., Burgstaller-Muehlbacher, S., Hughes, L.D. et al. (2018) The ReFRAME library as a comprehensive drug repurposing library and its application to the treatment of cryptosporidiosis. *Proc. Natl Acad. Sci. U.S.A.* **115**, 10750–10755 https://doi.org/10.1073/pnas.1810137115
14 Fleming, N. (2018) How artificial intelligence is changing drug discovery. *Nature (London)* **557**, S55–S57 https://doi.org/10.1038/d41586-018-05267-x
15 Smalley, E. (2017) AI-powered drug discovery captures pharma interest. *Nat. Biotechnol.* **35**, 604–605 https://doi.org/10.1038/nbt0717-604
16 Wang, F., Wu, F., Li, C., Jia, C., Su, S., Hao, G. et al. (2019) ACID: a free tool for drug repurposing using consensus inverse docking strategy. *J. Cheminform.* **11**, 73 https://doi.org/10.1186/s13321-019-0394-z
17 Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S. et al. (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 https://doi.org/10.1002/jcc.21256
18 Trott, O. and Olson, A.J. (2010) Autodock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 https://doi.org/10.1002/jcc.21334
19 Allen, W.J., Balius, T.E., Mukherjee, S., Brozell, S.R., Moustakas, D.T., Lang, P.T. et al. (2015) DOCK 6: impact of new features and current docking performance. *J. Comput. Chem.* **36**, 1132–1156 https://doi.org/10.1002/jcc.23905
20 Korb, O., Stützle, T. and Exner, T.E. (2009) Empirical scoring functions for advanced protein–ligand docking with plants. *J. Chem. Inf. Model.* **49**, 84–96 https://doi.org/10.1021/ci800298z
21 Ng, M.C.K., Fong, S. and Siu, S.W.I. (2015) PSOVina: the hybrid particle swarm optimization algorithm for protein–ligand docking. *J. Bioinform. Comput. Biol.* **13**, 1541007 https://doi.org/10.1142/S0219720015410073
22 Liu, Z., Li, Y., Han, L., Li, J., Liu, J., Zhao, Z. et al. (2014) PDB-wide collection of binding data: current status of the PDBbind database. *Bioinformatics* **31**, 405–412 https://doi.org/10.1093/bioinformatics/btu626
23 Minie, M., Chopra, G., Sethi, G., Horst, J., White, G., Roy, A. et al. (2014) CANDO and the infinite drug discovery frontier. *Drug Discov. Today* **19**, 1353–1363 https://doi.org/10.1016/j.drudis.2014.06.018
24 Chopra, G., Kaushik, S., Elkin, P.L. and Samudrala, R. (2016) Combating ebola with repurposed therapeutics using the CANDO platform. *Molecules (Basel, Switzerland)* **21**, 1537 https://doi.org/10.3390/molecules21121537
25 Fine, J., Konc, J., Samudrala, R. and Chopra, G. (2020) CANDOCK: chemical atomic network-based hierarchical flexible docking algorithm using generalized statistical potentials. *J. Chem. Inf. Model.* **60**, 1509–1527 https://doi.org/10.1021/acs.jcim.9b00686
26 Walshe, J., Edwards, W., Austin-Muttitt, K., Witts, J. and Mullins, J.G.L. (2019) Re-Drug: Drug Repurposing Screening Platform. Proceedings of the Biochemical Society meeting on Drug Repurposing
27 Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A. et al. (2020) The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 https://doi.org/10.1093/nar/gkz1031
28 Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. and Tanabe, M. (2021) KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 https://doi.org/10.1093/nar/gkaa970
29 Eberhardt, J., Santos-Martins, D., Tillack, A.F. and Forli, S. (2021) Autodock Vina 1.2.0: new docking methods, expanded force field, and python bindings. *J. Chem. Inf. Model.* **61**, 3891–3898 https://doi.org/10.1021/acs.jcim.1c00203
30 Chen, H. and Zhang, Z. (2013) A semi-supervised method for drug-Target interaction prediction with consistency in networks. *PLoS ONE* **8**, e62975 https://doi.org/10.1371/journal.pone.0062975
31 Mysinger, M.M., Carchia, M., Irwin, J.J. and Shoichet, B.K. (2012) Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* **55**, 6582–6594 https://doi.org/10.1021/jm300687e
32 Chen, L., Cruz, A., Ramsey, S., Dickson, C.J., Duca, J.S., Hornak, V. et al. (2019) Hidden bias in the DUD-E dataset leads to misleading performance of deep learning in structure-based virtual screening. *PLoS ONE* **14**, e0220113 https://doi.org/10.1371/journal.pone.0220113
33 Bento, A.P., Gaulton, A., Hersey, A., Bellis, L.J., Chambers, J., Davies, M. et al. (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Res.* **42**, D1083–D1090 https://doi.org/10.1093/nar/gkt1031
34 Irwin, J.J., Tang, K.G., Young, J., Dandarchuluun, C., Wong, B.R., Khurelbaatar, M. et al. (2020) ZINC$_{20}$-a free ultralarge-Scale chemical database for ligand discovery. *J. Chem. Inf. Model.* **60**, 6065–6073 https://doi.org/10.1021/acs.jcim.0c00675

35 Imrie, F., Bradley, A.R. and Deane, C.M. (2021) Generating property-Matched decoy molecules using deep learning. *Bioinformatics (Oxford, England)* **37**, 2134–2141 https://doi.org/10.1093/bioinformatics/btab080
36 Cheng, F., Desai, R.J., Handy, D.E., Wang, R., Schneeweiss, S., Barabási, A. et al. (2018) Network-based approach to prediction and population-based validation of *in silico* drug repurposing. *Nat. Commun.* **9**, 2691–2612 https://doi.org/10.1038/s41467-018-05116-5
37 Cheng, F., Kovács, I.A. and Barabási, A. (2019) Network-based prediction of drug combinations. *Nat. Commun.* **10**, 1197 https://doi.org/10.1038/s41467-019-09186-x
38 Cheng, F., Lu, W., Liu, C., Fang, J., Hou, Y., Handy, D.E. et al. (2019) A genome-wide positioning systems network algorithm for *in silico* drug repurposing. *Nat. Commun.* **10**, 3476–3414 https://doi.org/10.1038/s41467-019-10744-6
39 Chen, H., Sherr, D.H., Hu, Z. and DeLisi, C. (2016) A network based approach to drug repositioning identifies plausible candidates for breast cancer and prostate cancer. *BMC Med. Genomics* **9**, 51 https://doi.org/10.1186/s12920-016-0212-7
40 Keenan, A.B., Jenkins, S.L., Jagodnik, K.M., Koplev, S., He, E., Torre, D. et al. (2018) The library of integrated network-based cellular signatures NIH program: system-level cataloging of human cells response to perturbations. *Cell Syst.* **6**, 13–24 https://doi.org/10.1016/j.cels.2017.11.001
41 Lamb, J. (2007) The connectivity Map: a new tool for biomedical research. *Nat. Rev. Cancer* **7**, 54–60 https://doi.org/10.1038/nrc2044
42 Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F. and Hamosh, A. (2015) OMIM.org: online Mendelian inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **43**, D789–D798 https://doi.org/10.1093/nar/gku1205
43 Chen, G., Ramírez, J.C., Deng, N., Qiu, X., Wu, C., Zheng, W.J. et al. (2019) Restructured GEO: restructuring gene expression omnibus metadata for genome dynamics analysis. *Database* **2019**, bay145 https://doi.org/10.1093/database/bay145
44 Lee, H., Palm, J., Grimes, S.M. and Ji, H.P. (2015) The cancer genome atlas clinical explorer: a web and mobile interface for identifying clinical-genomic driver associations. *Genome Med.* **7**, 112 https://doi.org/10.1186/s13073-015-0226-3
45 Yamanishi, Y. (2013) Chemogenomic approaches to infer drug-target interaction networks. *Methods Mol. Biol.* **939**, 97–113 https://doi.org/10.1007/978-1-62703-107-3_9
46 O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T. and Hutchison, G.R. (2011) Open babel: an open chemical toolbox. *J. Cheminform.* **3**, 1–14 https://doi.org/10.1186/1758-2946-3-1
47 Bajusz, D., Rácz, A. and Héberger, K. (2015) Why is tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* **7**, 20–23 https://doi.org/10.1186/s13321-015-0069-3
48 Fradera, X. and Babaoglu, K. (2017) Overview of methods and strategies for conducting virtual small molecule screening. *Curr. Protoc. Chem. Biol.* **9**, 196–212 https://doi.org/10.1002/cpch.27
49 Cereto-Massagué, A., Ojeda, M.J., Valls, C., Mulero, M., Garcia-Vallvé, S. and Pujadas, G. (2015) Molecular fingerprint similarity search in virtual screening. *Methods (San Diego, Calif.)* **71**, 58–63 https://doi.org/10.1016/j.ymeth.2014.08.005
50 Gimeno, A., Ojeda-Montes, M.J., Tomás-Hernández, S., Cereto-Massagué, A., Beltrán-Debón, R., Mulero, M. et al. (2019) The light and dark sides of virtual screening: what Is there to know? *Int. J. Mol. Sci.* **20**, 1375 https://doi.org/10.3390/ijms20061375
51 Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 https://doi.org/10.1016/S0022-2836(05)80360-2
52 Winter, R., Montanari, F., Noé, F. and Clevert, D. (2019) Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. *Chem. Sci. (Cambridge)* **10**, 1692–1701 https://doi.org/10.1039/C8SC04175J
53 Shang, C., Liu, Q., Chen, K., Sun, J., Lu, J., Yi, J. et al. Edge attention-based multi-relational graph convolutional networks. arXiv preprint arXiv:1802.04944 2018
54 Zeng, X., Zhu, S., Lu, W., Liu, Z., Huang, J., Zhou, Y. et al. (2020) Target identification among known drugs by deep learning from heterogeneous networks. *Chem. Sci. (Cambridge)* **11**, 1775–1797 https://doi.org/10.1039/C9SC04336E
55 Zong, N., Kim, H., Ngo, V. and Harismendy, O. (2017) Deep mining heterogeneous networks of biomedical linked data to predict novel drug-target associations. *Bioinformatics (Oxford, England)* **33**, 2337–2344 https://doi.org/10.1093/bioinformatics/btx160
56 Sosa, D.N., Derry, A., Guo, M., Wei, E., Brinton, C. and Altman, R.B. (2020) A literature-Based knowledge graph embedding method for identifying drug repurposing opportunities in rare diseases. *Pac. Symp. Biocomput.* **25**, 463–474 PMID:31797619
57 Gentile, F., Agrawal, V., Hsing, M., Ton, A., Ban, F., Norinder, U. et al. (2020) Deep docking: a deep learning platform for augmentation of structure based drug discovery. *ACS Cent. Sci.* **6**, 939–949 https://doi.org/10.1021/acscentsci.0c00229
58 Husic, B.E., Charron, N.E., Lemm, D., Wang, J., Pérez, A., Majewski, M. et al. (2020) Coarse graining molecular dynamics with graph neural networks. *J. Chem. Phys.* **153**, 194101 https://doi.org/10.1063/5.0026133
59 Hassan-Harrirou, H., Zhang, C. and Lemmin, T. (2020) RosENet: improving binding affinity prediction by leveraging molecular mechanics energies with an ensemble of 3D convolutional neural networks. *J. Chem. Inf. Model.* **60**, 2791–2802 https://doi.org/10.1021/acs.jcim.0c00075
60 Stokes, J.M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N.M. et al. (2020) A deep learning approach to antibiotic discovery. *Cell* **180**, 688–702.e13 https://doi.org/10.1016/j.cell.2020.01.021
61 Zeng, X., Zhu, S., Liu, X., Zhou, Y., Nussinov, R. and Cheng, F. (2019) deepDR: a network-based deep learning approach to *in silico* drug repositioning. *Bioinformatics (Oxford, England)* **35**, 5191–5198 https://doi.org/10.1093/bioinformatics/btz418
62 Zeng, X., Song, X., Ma, T., Pan, X., Zhou, Y., Hou, Y. et al. (2020) Repurpose open data to discover therapeutics for COVID-19 using deep learning. *J. Proteome Res.* **19**, 4624–4636 https://doi.org/10.1021/acs.jproteome.0c00316
63 Dhasmana, D.J. (2021) Dexamethasone in hospitalized patients with COVID-19. *N. Engl. J. Med.* **384**, 693–704 https://doi.org/10.1056/NEJMoa2021436
64 Richardson, P., Griffin, I., Tucker, C., Smith, D., Oechsle, O., Phelan, A. et al. (2020) Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *Lancet* **395**, e30–e31 https://doi.org/10.1016/S0140-6736(20)30304-4
65 Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W. and Cheng, F. (2020) Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.* **6**, 14 https://doi.org/10.1038/s41421-020-0153-3
66 Jeon, S., Ko, M., Lee, J., Choi, I., Byun, S.Y., Park, S. et al. (2020) Identification of antiviral drug candidates against SARS-CoV-2 from FDA-Approved drugs. *Antimicrob. Agents Chemother.* **64**, e00819-20 https://doi.org/10.1128/AAC.00819-20

Biochemical Society Transactions (2022) **50** 747–758
https://doi.org/10.1042/BST20200967

67 Beck, B.R., Shin, B., Choi, Y., Park, S. and Kang, K. (2020) Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. *Comput. Struct. Biotechnol. J.* **18**, 784–790 https://doi.org/10.1016/j.csbj.2020.03.025

68 Ge, Y., Tian, T., Huang, S., Wan, F., Li, J., Li, S. et al. (2021) An integrative drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. *Signal Transduct. Target. Ther.* **6**, 165 https://doi.org/10.1038/s41392-021-00568-6

69 Cheng, F., Rao, S. and Mehra, R. (2020) COVID-19 treatment: combining anti-inflammatory and antiviral therapeutics using a network-based approach. *Cleve. Clin. J. Med.* https://doi.org/10.3949/ccjm.87a.ccc037

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

71 Townshend, R.J., Vögele, M., Suriana, P., Derry, A., Powers, A., Laloudakis, Y. et al. Atom3d: tasks on molecules in three dimensions. arXiv 2012.04035 2020

72 Vázquez, J., López, M., Gibert, E., Herrero, E. and Luque, F.J. (2020) Merging ligand-based and structure-based methods in drug discovery: an overview of combined virtual screening approaches. *Molecules (Basel, Switzerland)* **25**, 4723 https://doi.org/10.3390/molecules25204723

73 DeepWalk: ACM; Aug 24, 2014

74 Liu, T., Hsieh, Y., Chou, C. and Yang, P. (2018) Systematic polypharmacology and drug repurposing via an integrated L1000-based connectivity Map database mining. *R. Soc. Open Sci.* **5**, 181321 https://doi.org/10.1098/rsos.181321

75 Zhang, Y., Jiang, Z., Chen, C., Wei, Q., Gu, H. and Yu, B. (2021) DeepStack-DTIs: predicting drug–target interactions using LightGBM feature selection and deep-stacked ensemble classifier. *Interdiscip. Sci.* https://doi.org/10.1007/s12539-021-00488-7

76 Wang, X., Cao, T., Jia, C.M., Tian, X. and Wang, Y. (2021) Quantitative prediction model for affinity of drug–target interactions based on molecular vibrations and overall system of ligand-receptor. *BMC Bioinformatics* **22**, 1–497 https://doi.org/10.1186/s12859-020-03881-z

77 Kim, J., Park, S., Min, D. and Kim, W. (2021) Comprehensive survey of recent drug discovery using deep learning. *Int. J. Mol. Sci.* **22**, 9983 https://doi.org/10.3390/ijms22189983

78 Zheng, S., Li, Y., Chen, S., Xu, J. and Yang, Y. (2020) Publisher correction: predicting drug–protein interaction using quasi-visual question answering system. *Nat. Mach. Intell.* **2**, 551 https://doi.org/10.1038/s42256-020-0224-z

79 Zheng, J., Xiao, X. and Qiu, W. (2021) iCDI-W2vCom: identifying the Ion channel–Drug interaction in cellular networking based on word2vec and node2vec. *Front. Genet.* **12**, 738274 https://doi.org/10.3389/fgene.2021.738274

80 Thafar, M.A., Olayan, R.S., Albaradei, S., Bajic, V.B., Gojobori, T., Essack, M. et al. (2021) DTi2vec: drug–target interaction prediction using network embedding and ensemble learning. *J. Cheminform.* **13**, 71 https://doi.org/10.1186/s13321-021-00552-w