



Cite this: *RSC Adv.*, 2021, **11**, 36181

## Computational methods directed towards drug repurposing for COVID-19: advantages and limitations†

Prem Prakash Sharma,‡<sup>a</sup> Meenakshi Bansal, ‡<sup>a</sup> Aaftaab Sethi,‡<sup>b</sup> Poonam, Lindomar Pena, <sup>d</sup> Vijay Kumar Goel, <sup>e</sup> Maria Grishina, <sup>f</sup> Shubhra Chaturvedi, <sup>g</sup> Dhruv Kumar\*<sup>h</sup> and Brijesh Rathi \*<sup>a</sup>

Novel coronavirus disease 2019 (COVID-19) has significantly altered the socio-economic status of countries. Although vaccines are now available against the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a causative agent for COVID-19, it continues to transmit and newer variants of concern have been consistently emerging world-wide. Computational strategies involving drug repurposing offer a viable opportunity to choose a medication from a rundown of affirmed drugs against distinct diseases including COVID-19. While pandemics impede the healthcare systems, drug repurposing or repositioning represents a hopeful approach in which existing drugs can be remodeled and employed to treat newer diseases. In this review, we summarize the diverse computational approaches attempted for developing drugs through drug repurposing or repositioning against COVID-19 and discuss their advantages and limitations. To this end, we have outlined studies that utilized computational techniques such as molecular docking, molecular dynamic simulation, disease–disease association, drug–drug interaction, integrated biological network, artificial intelligence, machine learning and network medicine to accelerate creation of smart and safe drugs against COVID-19.

Received 10th July 2021  
 Accepted 7th October 2021

DOI: 10.1039/d1ra05320e  
[rsc.li/rsc-advances](http://rsc.li/rsc-advances)

### 1. Introduction

The mean investment required to bring a new therapeutic agent to the market has been estimated to be ~\$1.3 billion, accounting for the costs of failed clinical trials.<sup>1</sup> Apart from the cost, the time involved in a conventional drug discovery process needs to be substantially reduced especially when response to a novel disease must be rapid, as is the case with the COVID-19 pandemic. It is an essential prerequisite that any new chemical

entity being explored as a drug, must be safe and non-toxic.<sup>2</sup> However, investigating the safety profile of any new chemical agent is a long and tedious procedure, therefore numerous attempts have made to discover anti-SARS-CoV-2 inhibitors from existing drugs.<sup>3,4</sup>

Drug repurposing or repositioning refers to the usage of pre-existing drugs to tackle a disease other than the one that the drug was originally designed to target. It has several advantages such as reduction of time, risk and cost associated with the development of a new drug molecule.<sup>2,5,6</sup> The benefit of drug repurposing is also highlighted by the fact that nearly 30% of new market entrants are derived from pre-existing drugs.<sup>7</sup> There are several reasons for which a drug repurposing strategy can be used, some of them include, (a) reduction in time and cost of clinical trials, as evaluation in phase I and II might not be required; (b) already established pharmaceutical supply chains for distribution; (c) the additive or symbiotic effects of combinatorial therapy with other drugs in treatments in contrast to monotherapy; and (d) may explore new mechanisms of action for old drugs.<sup>8</sup> The drug repurposing strategy has some limitations such as patent barriers. However, other factors such as the complexity of regulatory pathways, absence of funding opportunities, greater access to data from other industry-sponsored clinical trials, and the heterogeneity of the population for new clinical studies, have been instrumental in evolving drug repurposing as one of the most promising approaches for the

<sup>a</sup>Laboratory For Translational Chemistry and Drug Discovery, Department of Chemistry, Hansraj College, University of Delhi, Delhi, 110007, India. E-mail: [brijeshrathi@hrc.du.ac.in](mailto:brijeshrathi@hrc.du.ac.in)

<sup>b</sup>Department of Medicinal Chemistry, National Institute of Pharmaceutical Education and Research (NIPER), Hyderabad, India

<sup>c</sup>Department of Chemistry, Miranda House, University of Delhi, Delhi, 110007 India

<sup>d</sup>Department of Virology, Aggeu Magalhaes, Institute (IAM), Oswaldo Cruz Foundation (Fiocruz), Recife 50670-420, Pernambuco, Brazil

<sup>e</sup>School of Physical Sciences, Jawaharlal Nehru University, New Delhi, 110067, India

<sup>f</sup>South Ural State University, Laboratory of Computational Modelling of Drugs, Pr. Lenina, 76 454080, Russia

<sup>g</sup>Division of Cyclotron and Radiopharmaceutical Sciences, Institute of Nuclear Medicine and Allied Sciences, New Delhi 110054, India

<sup>h</sup>Amity Institute of Molecular Medicine & Stem Cell Research (AIMMSCR), Amity University Uttar Pradesh, Noida 201313, India

† Electronic supplementary information (ESI) available. See DOI: [10.1039/d1ra05320e](https://doi.org/10.1039/d1ra05320e)

‡ Contributed equally.



drug development with unprecedented swiftness. Drug repurposing is also sometimes referred as drug redirecting, drug repositioning, drug reprofiling, drug retasking, and therapeutic switching.<sup>9</sup>

Drug repurposing has resulted in the successful development of various medicines. Several examples of drugs repositioned to treat diseases or disorders have come to the market, which include minoxidil, originally developed for hypertension and later used for hair loss; sildenafl, identified to treat erectile dysfunction as a result of a clinical trial for angina; and thalidomide, originally for morning sickness and now used to treat symptoms of leprosy.<sup>10</sup> Several other medicines were repurposed and are being used to treat a multitude of ailments.<sup>11</sup> A report published in July 2020 has shown that 8565 existing drugs are being repositioned for COVID-19.<sup>12</sup>

The conventional drug development process involves *de novo* detection and identification of new molecular entities, including the five phases, (a) discovery and preclinical; (b) safety review; (c) clinical research; (d) FDA review; and (e) FDA post-market safety monitoring. It is a time-consuming and expensive procedure with a high risk of failure as highlighted briefly before, while on the other hand, the drug repositioning has only four steps, including (a) compound identification; (b) compound acquisition; (c) development; and (d) FDA post-market safety monitoring.<sup>13</sup> Several reports state that drug repurposing requires clearance from the FDA and/or the European Medicines Agency (EMA) for approval, but it approximately takes anywhere between 3–12 years and at a decreased cost of 50–60 percent.<sup>13</sup> Drug repurposing has great significance in the current COVID-19 pandemic scenario where no standard and safe therapy is available yet against COVID-19.

Bioinformatics tool and other recently evolved computational methods based on Artificial Intelligence (AI), Machine Learning (ML) and network medicine are promising and on many occasions have been found to be robust in identifying new indications of existing drugs.<sup>14,15</sup> The present article aims to provide highlights on computational approaches in drug development through drug repurposing for COVID-19, where each method is categorized into its core methodology, advantages and limitations. There are many studies published on a day-to-day basis describing a computational approach to identify drugs to treat SARS-CoV-2. This article is not a comprehensive overview of all those approaches as discussing the complete list of published studies would sound like a mere regurgitation and not be of any significant impact. Herein, we have tried to make an effort to provide relevant snippets of information from a huge volume of studies. There are several reviews published on this topic, but these articles have either focused extensively on drug repurposing approaches in general but not with respect to COVID-19 or a discussion on computational approaches has been limited to only a few popular techniques.<sup>6,16–18</sup> We believe that the collective information on a wide variety of computational drug repurposing methods discussed in this article alongside with their critical analysis and scrutiny would be of great interest to a wide range of researchers and scientists.

## 2. Drug repurposing by virtual screening and molecular dynamics (MD) simulations

A large number of *in silico* studies have been carried out during the period of global lockdown. Structure-based drug design (SBDD) is a popular method in computational drug design in which X-ray crystallographic or NMR based 3D structures of proteins/targets are utilized, and small molecule are generally studied to evaluate the nature of interaction between protein-ligand.<sup>16,19,20</sup> It also explains the conformational fitting of ligand within the binding pocket of protein. Therefore, the whole point of exploiting this technique is to find “binders” not just “fitters”.<sup>21</sup> Ligands are small molecules that can act as activators or inhibitors of a protein by binding to it.<sup>22,23</sup> Scoring functions are very useful and practical to score and rank all the screened ligands in the binding pocket based on the conformations generated, which leads to estimation of “binding affinity”, where the potential binders are referred to as “Hits”.<sup>21</sup> There are several softwares available that can be utilized for molecular docking, a few are open-access, such as AutoDock Vina while others are commercial suites like Glide, GOLD, *etc.*

Docking results must be validated with MD simulation to understand the interactions between receptor-ligand and their stability. It helps in estimation of the accurate thermodynamics and kinetics associated with drug-target recognition and binding.<sup>24</sup> MD simulation are generally also helpful in separating true positives from false positives. There are several programs, *e.g.*, Desmond and GROMACS among others that can be implemented for MD simulation on GUI interface or command line.

### 2.1. Methods

Approaches implemented for drug repurposing virtual screening and MD simulations have been described in Fig. 1, and include the following key steps.

- Retrieving experimentally determined crystal or cryo-EM protein structures from public databases. If unavailable, utilizing homology modelled protein structures or retrieving predicted structures from sources like AlphaFold (<https://alphafold.ebi.ac.uk/>);
- Protein and ligand preparation prior to docking;
- Screening of molecules that can bind to the target proteins;
- Selection of hit compounds based on respective scoring function algorithm; and
- Validation of most promising molecules by MD simulations.

The structures of drug molecules described in this article are provided in Table 1. Recently, Gimeno *et al.*<sup>25</sup> integrated the predictions of three molecular docking softwares (Glide, FRED, and AutoDock Vina) to repurpose drugs against the main protease ( $M^{Pro}$ ) (PDB: 6LU7) of SARS-CoV-2. The H41 and C145 are catalytic residues responsible for  $M^{Pro}$  functionality. Given its pivotal role in mediating the replication and transcription of SARS-CoV-2, antiviral therapies targeting  $M^{Pro}$  have the



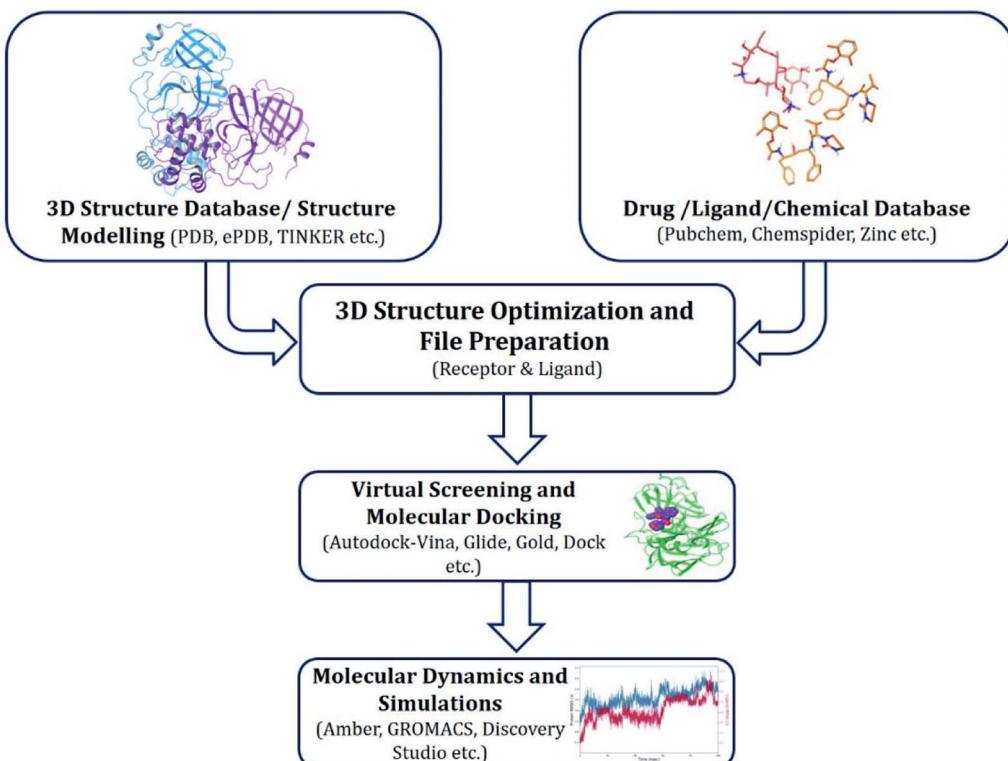


Fig. 1 Workflow for structure-based drug design.

potential to disrupt critical steps in the viral life cycle. The carprofen interacted with both catalytic residues while celecoxib did not. *In vitro* study of carprofen and celecoxib showed inhibition of  $M^{pro}$  by 3.97 and 11.90% at 50  $\mu$ M, respectively.<sup>25</sup> Wang *et al.* proposed the potential of lopinavir along with other four drugs, namely, carfilzomib, eravacycline, valrubicin, and elbasvir, as promising molecules to inhibit  $M^{pro}$  (PDB: 6LU7) activity which were also validated by 100 ns MD simulation.<sup>26</sup> These molecules interacted to hot spot residues H41, M49, N142, H164, M165, E166, and Q189. The combination of lopinavir and ritonavir showed a potential to better inhibit  $M^{pro}$  than lopinavir and ritonavir alone.<sup>27</sup> For decades, caffeine has been utilized clinically to treat apnea in preterm infants and to enhance respiratory function because of its potential role in reducing inflammation, immune modulation, antiviral and improving respiratory symptoms.<sup>28</sup> A computational study proposed a promising inhibiting nature of caffeine and linagliptin against  $M^{pro}$  (PDB: 6Y2E).<sup>29</sup> Both caffeine and linagliptin interacted with the catalytic residues but caffeine interacted *via* H-bond and van der Waals interactions while linagliptin interacted through van der Waals interaction alone. Koulgi *et al.*, proposed ensemble docking for *in silico* drug repurposing against  $M^{pro}$  (PDB: 6LU7) by screening two databases (DB), FDA approved drug-DB and SWEETLEAD-DB.<sup>30</sup> Authors found that indinavir, ceftin, ivermectin, neomycin, vasopressin-tennate, and amikacin as top ranked molecules against  $M^{pro}$ . Indinavir, ceftin and ivermectin were observed to show interactions with T24, L27, V42, T45, R60, K61, N142, E166 and Q189. Neomycin, vasopressin-tennate, and amikacin interacted to

either H41 or C145. In another study, Khan *et al.*, reported that remdesivir, saquinavir, darunavir and natural products such as flavones and coumarin derivatives as promising hits for  $M^{pro}$  (PDB: 6LU7) inhibition.<sup>31</sup> Both molecule interacted to C145, Q189, E166, and T190. Cherrak *et al.*, proposed glycosylated flavonoid quercetin-3-O-rhamnoside as a promising  $M^{pro}$  (PDB: 6LU7) inhibitor.<sup>32</sup> The quercetin-3-O-rhamnoside interacted to residues T26, F140, and L141 *via* H-bond while interacted to residues H163, G143, C145, M49, and R188 *via* other non-covalent interactions. Upadhyay *et al.*, proposed a few plant extracts including green tea, black tea and haritaki as alternative therapeutic candidates for SARS-CoV-2 infection. These plant products have the polymeric polyphenols thearubigin, quercetin-3-O-rutinoside, and hesperidin which can potentially inhibit  $M^{pro}$  (PDB: 6LU7).<sup>33</sup> The thearubigin interacted to catalytic residue, C145 while quercetin-3-O-rutinoside interacted to L141, N142, G143, and E166. Hesperidin hydrophobically interacted to M165 and Q189 while H-bond interacted to N142, and G146 residues.

Multi-targeted drugs such as ergotamine, dihydroergotamine, bromocriptine, dutasteride, conivaptan, paliperidone, and tipranavir were reported as potential promising inhibitors of both  $M^{pro}$  (PDB: 6Y84) and RNA-dependent RNA polymerase (RdRp) (PDB: 6NUR).<sup>34</sup> The RdRp is essential for SARS-CoV-2 replication and has been considered as a major drug target.<sup>14</sup> Remdesivir is an FDA-approved adenosine analogue, which has an EC<sub>90</sub> value of 1.76  $\mu$ M against SARS-CoV-2 in Vero E6 cell and even displayed activity against the virus in human liver cancer Huh-7 cells.<sup>35</sup> Yin *et al.*, developed remdesivir derivative, which



Table 1 2-Dimensional structure of drugs proposed as therapeutic agents to treat COVID-19 during several drug repurposing studies. The SDF file for these molecules can be found in ESI

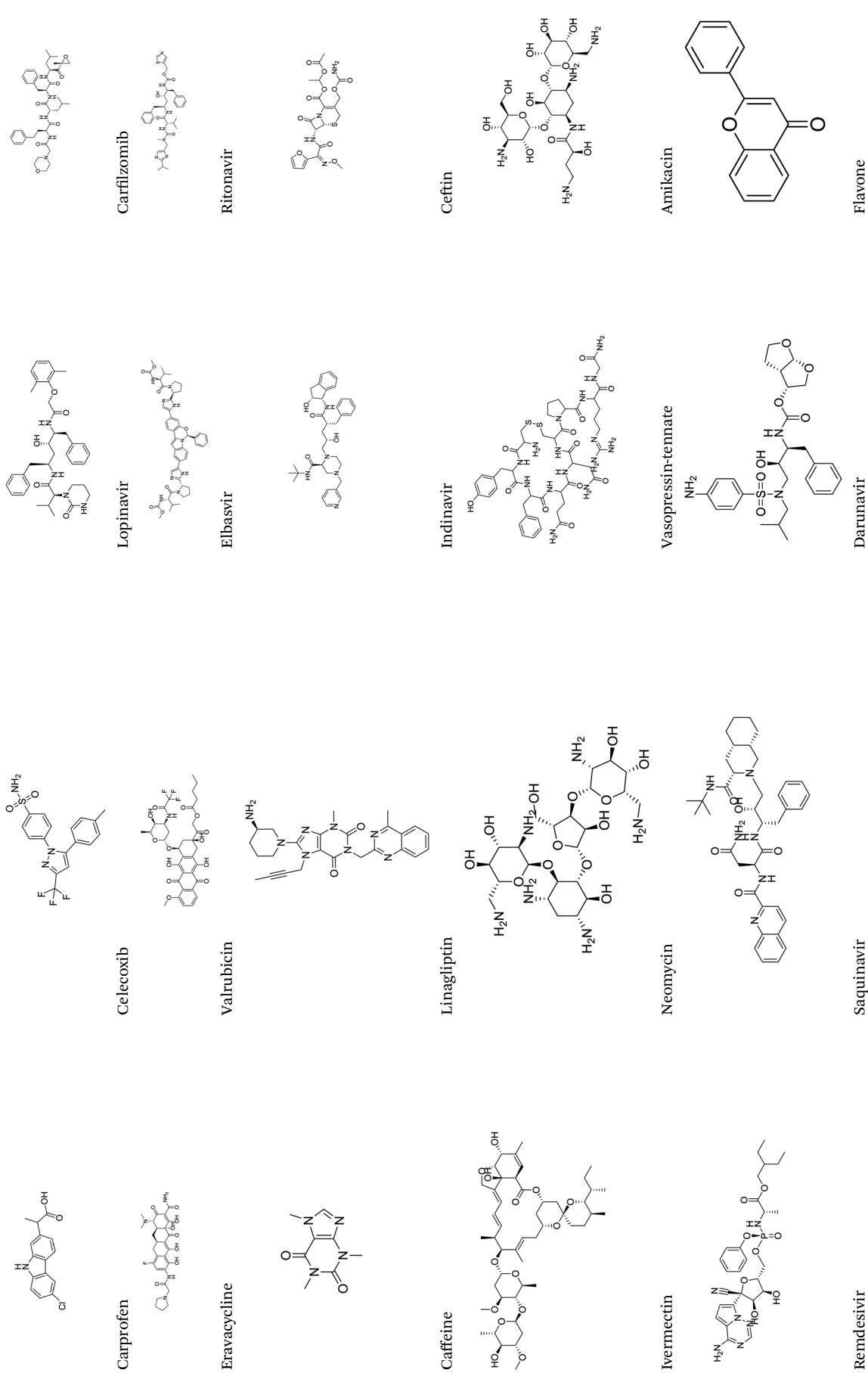


Table 1 (Contd.)

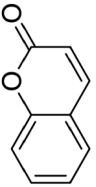
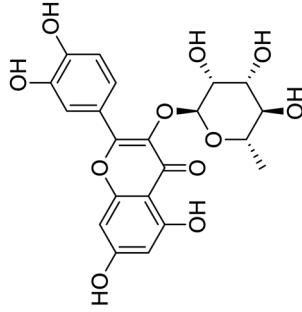
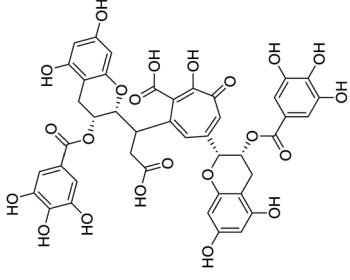
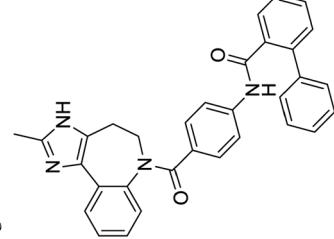
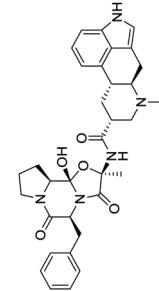
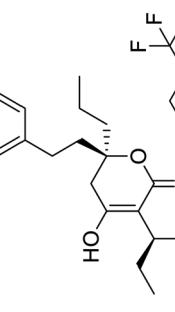
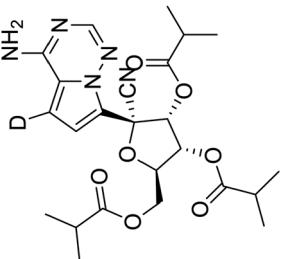
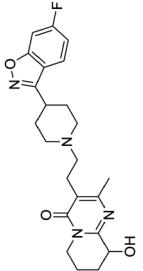
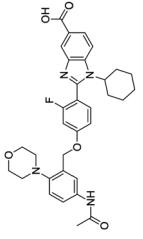
	Coumarin		Quercetin-3-O-rhamnoside		Thearubigin		Ergotamine		Dihydroergotamine		Bromocriptine		Dutasteride		Paliperidone		Tipranavir		Remdesivir derivative IN-6		Conivaptan
---	----------	---	--------------------------	---	-------------	---	------------	---	-------------------	--	---------------	--	-------------	--	--------------	---	------------	---	-------------------------------	--	------------

Table 1 (Cont'd.)

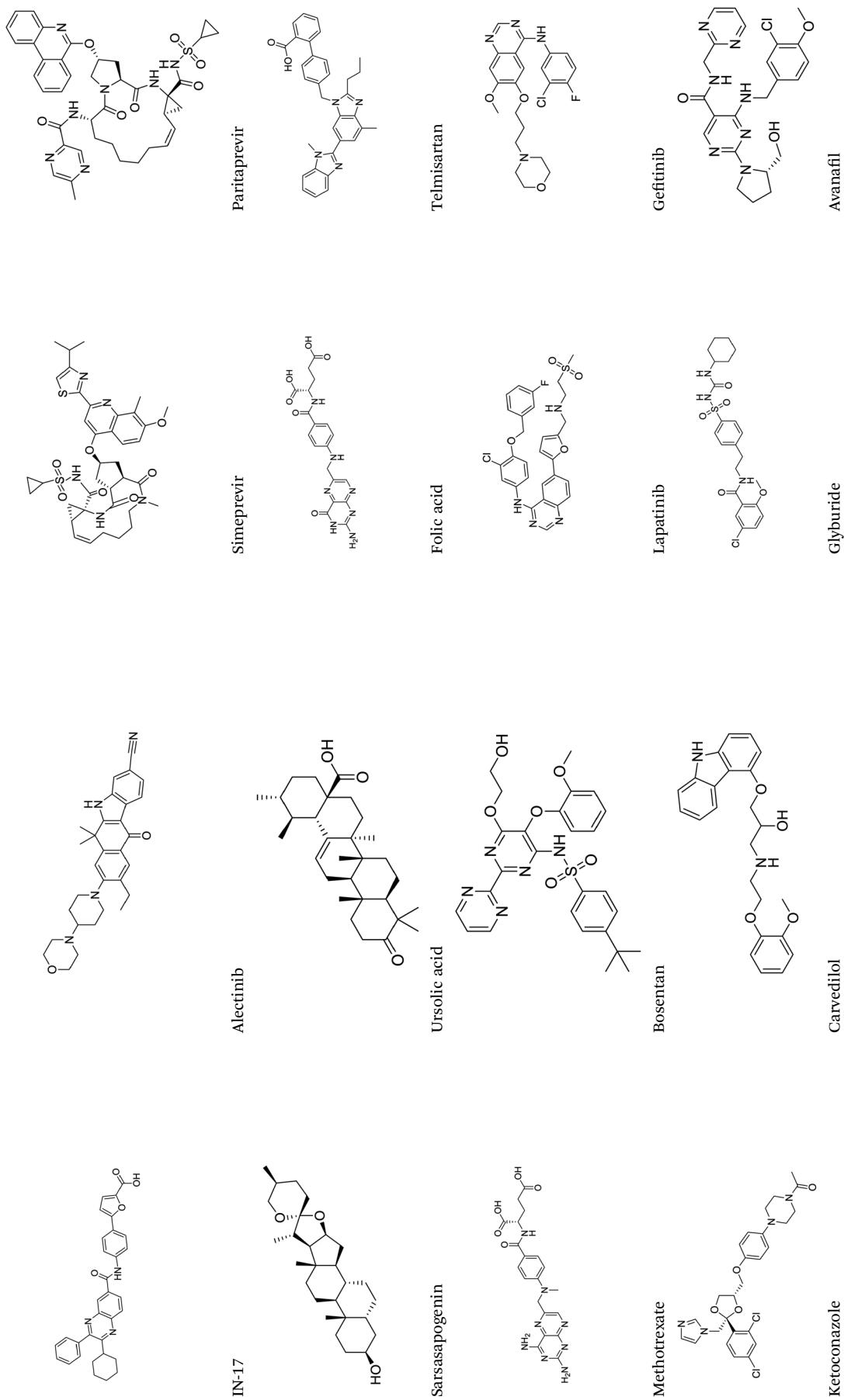


Table 1 (Cont'd.)

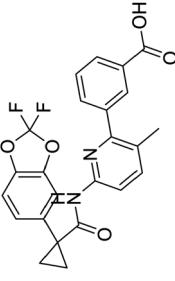
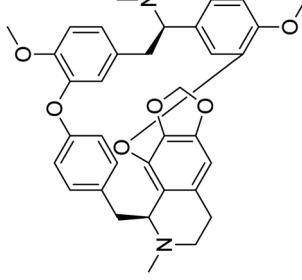
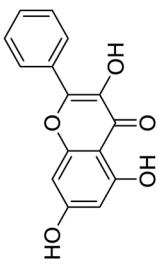
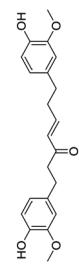
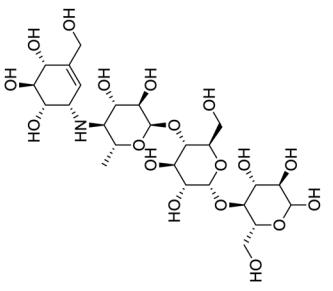
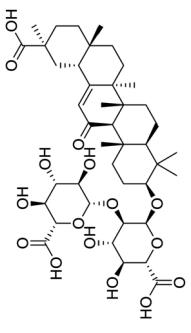
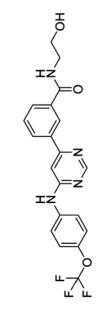
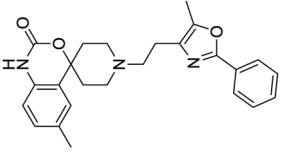
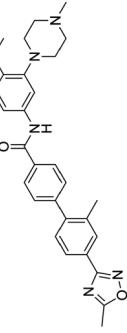
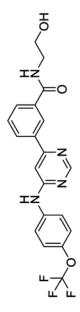
	Lumacaftor		Cepharanthine		Edodoxidine		Esculin		Acarbose		Glycyrrhetic acid		Galangin		Gingerenone A		Shogaol		GR 127935		GNF5		RS504393
---	------------	---	---------------	---	-------------	---	---------	---	----------	---	-------------------	--	----------	--	---------------	--	---------	---	-----------	--	------	---	----------

Table 1 (Cont'd.)

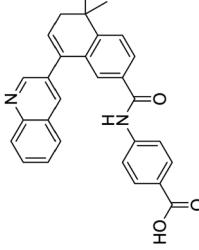
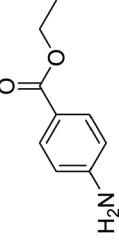
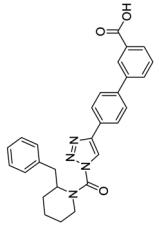
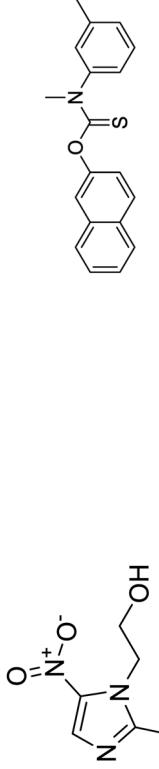
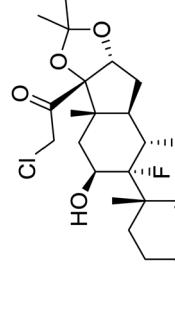
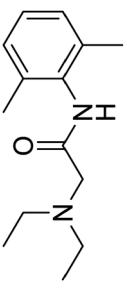
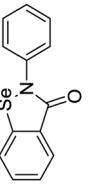
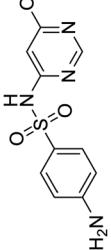
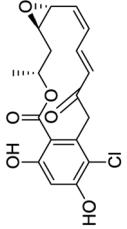
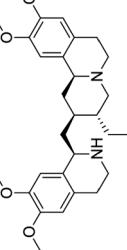
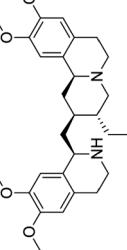
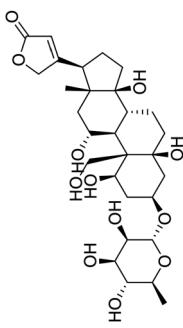
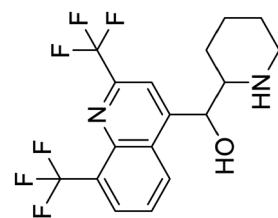
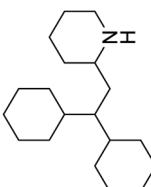
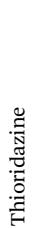
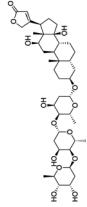
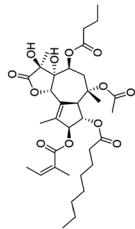
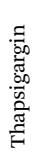
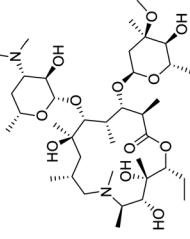
	BMS195614	
	KT203	
	Epitifibatide acetate	
	GSK1838705A	
	Halimonide	
	Lidocaine	
	Tolnaftate	
	Benzocaine	
	Lansoprazole	
	Ebselen	
	Sulfamonomethoxine	
	HSP 90 inhibitor (radicicol)	
	Emetine	

Table 1 (Cont'd.)

	Ouabain		Mefloquine		Perhexiline		Niclosamide		Digoxin		Thioridazine		Homoharringtonine		Thapsigargin		Azithromycin
---	---------	---	------------	---	-------------	---	-------------	---	---------	---	--------------	---	-------------------	---	--------------	---	--------------



**Gene Expression Omnibus (GEO):** Database repository of high throughput gene expression data and hybridization arrays, chips and microarrays

## Datasets Normalization

**RPKM:** Reads Per Kilobase of exon model per Million mapped reads,  
**TMM:** Trimmed Mean of M values, **RLE:** Relative Log Expression,  
**Quantiles** etc.



**Bioconductor:** Open software development for computational biology and bioinformatics based on R statistical programming language (DESeq, DESeq2, GRanges, Biostrings etc.)

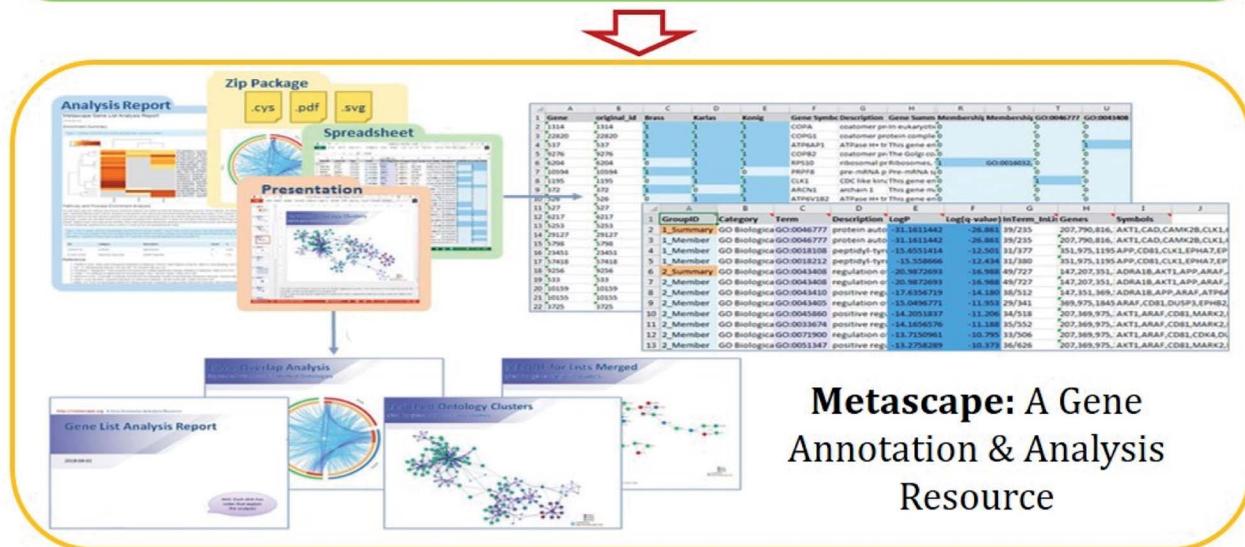


Fig. 2 Drug repurposing by disease–disease association method.

was five-fold potent than remdesivir in inhibition of viral replication.<sup>14</sup> Singh *et al.*, screened >350 potential RdRp (PDB: 7BV2) inhibitors from the BRENDA library and proposed JN-6-

and IN-17 as two promising hits compounds with the potential to inhibit RdRp with enhanced potency than remdesivir (Table 1).<sup>36</sup> These two molecules maintained interaction with

key residues K551, R553, R555, and T556. Alectinib is also proposed as a promising RdRp inhibitor, which was first used for non-small cell lung cancer (NSCLC) treatment.<sup>37</sup>

Nonstructural protein (NSP) 15, also referred as uridylate-specific endoribonuclease, primarily cleaves uridine at 3' end and plays an important role in host cells for viral replication. Studies have shown that viral replication can be delayed after inhibition of NSP15.<sup>38</sup> Khan *et al.*, proposed simeprevir and paritaprevir, antivirals originally developed against hepatitis C infection as potential inhibitors of SARS-CoV-2 NSP15 (PDB: 6W01).<sup>39</sup> The simeprevir and paritaprevir interacted with residues *via* H-bond (H235, K290), pi-interaction (A232, H235), and van der Waals interaction (E234). Kumar *et al.*, found computationally two phytochemicals sarsasapogenin and ursolic acid that may be potent inhibitors of NSP15 (PDB: 6VWW).<sup>38</sup> The ursolic acid interacted to residues K290, G248, T341, H250, and H235 while sarsasapogenin interacted to K290. Jung *et al.*, proposed FDA approved compounds such as folic acid, telmisartan, methotrexate, bosentan, lapatinib, gefitinib, ketoconazole, carvedilol, glyburide and avanafil as potential NSP3 (PDB: 6W0J) inhibitor, which also hold the promise of potential activity against SARS-CoV-2 with mutations in the spike protein.<sup>40</sup>

NSP13 helicase is a crucial component of viral replication among the 16 known coronavirus NSP proteins and shares the highest sequence conservation within the coronavirus family, highlighting its significance for viral viability.<sup>41</sup> White *et al.*, proposed lumacaftor and ceftaravine inhibiting NSP-13 ATPase activity after screening ~970 000 compounds from two libraries (the enamine libraries (AC and PC) and ZINC-in-trials

library) with estimated IC<sub>50</sub> values of 0.3 and 0.4 mM, respectively.<sup>41</sup> Based on sequence similarity of SARS-CoV-2 NSP 1 protein to the NSP 1 protein of SARS-CoV (PDB: 2HSX), Sharma *et al.*, hypothesized that NSP 1 inhibits host cell expression and inhibition of NSP 1 may result into SARS-CoV-2 attenuated form.<sup>42</sup> Authors proposed that a few FDA approved compounds like remdesivir, edoxudine, esculin, acarbose, glycyrrhizic acid, galangin, gingerenone, and shogaol may interact with NSP-1.<sup>42</sup> The esculin interacted with residues (R62, S63, A68, F72, and M74) *via* H-bond. The cidofovir, edoxudine, brivudine, remdesivir, SN00003849, and SN00003832 interacted to R62, R66, and M74 *via* H-bond. There were also H-bond interactions involved in stabilization of compounds iopromide (H2, V75, R113, N115), a cabose (R32, E76, L77, N115, G116, K114) and SN00216190 (R66, H72, M74, A65). It is suspected that the binding of SARS-CoV-2 spikes with human angiotensin-converting enzyme 2 (ACE2) protein promotes the entry of SARS-CoV-2 into human cells, making this mechanism a successful target.<sup>43</sup> Choudhary *et al.*, discovered that FDA-approved molecules (GR 127935 hydrochloride hydrate, GNF-5, RS504393, TNP, and eptifibatide acetate) bind effectively to ACE2 receptor virus binding motifs (PDB: 2AJF), whereas molecules KT203, BMS195614, KT185, RS504393, and GSK1838705A were identified to bind with the viral S-protein (homology model, template PDB: 6VSB) virus binding location.<sup>43</sup> The ACE2 receptor has K31 and K353 residues, which are reported as two main sites of virus binding to SARS-CoV. It was observed that GR 127935 hydrochloride hydrate, and GNF-5 interacted to K353 *via* hydrophobic interactions. The receptor-binding domain of S -protein interacted

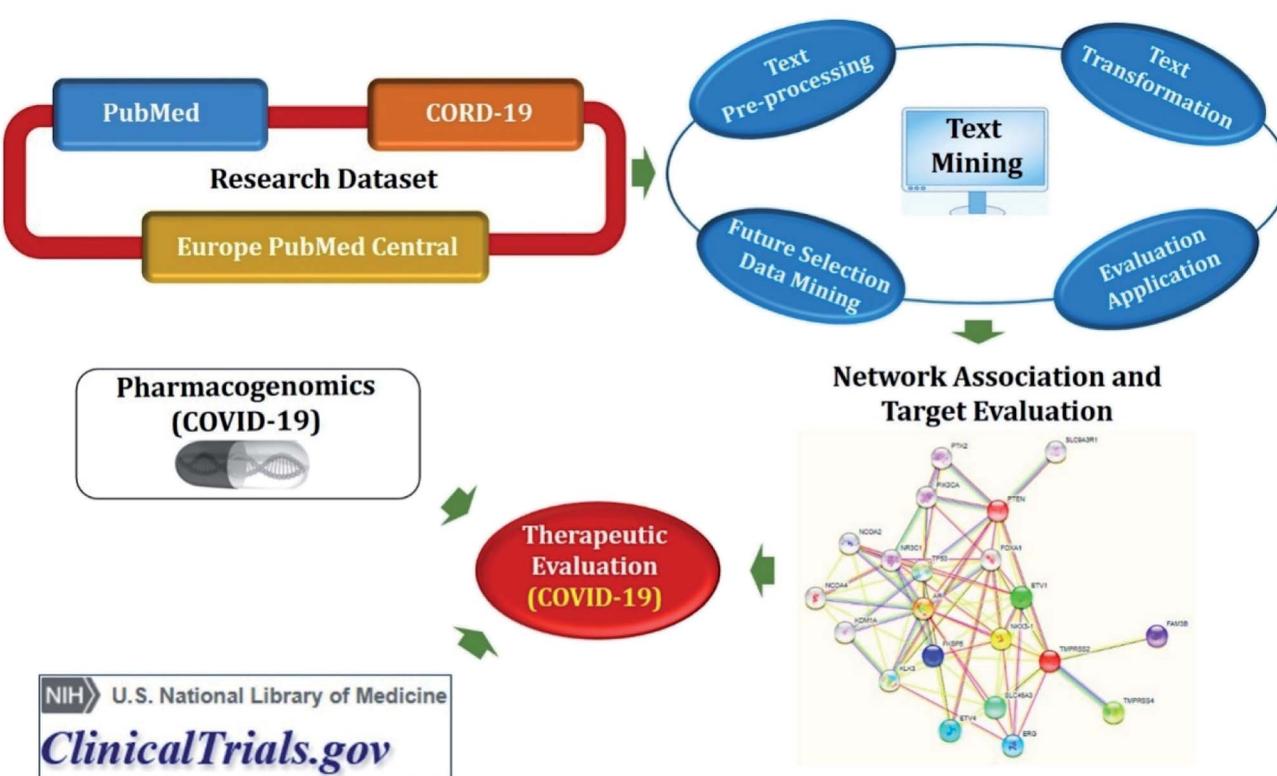


Fig. 3 Drug repurposing by integrated biological network approach.



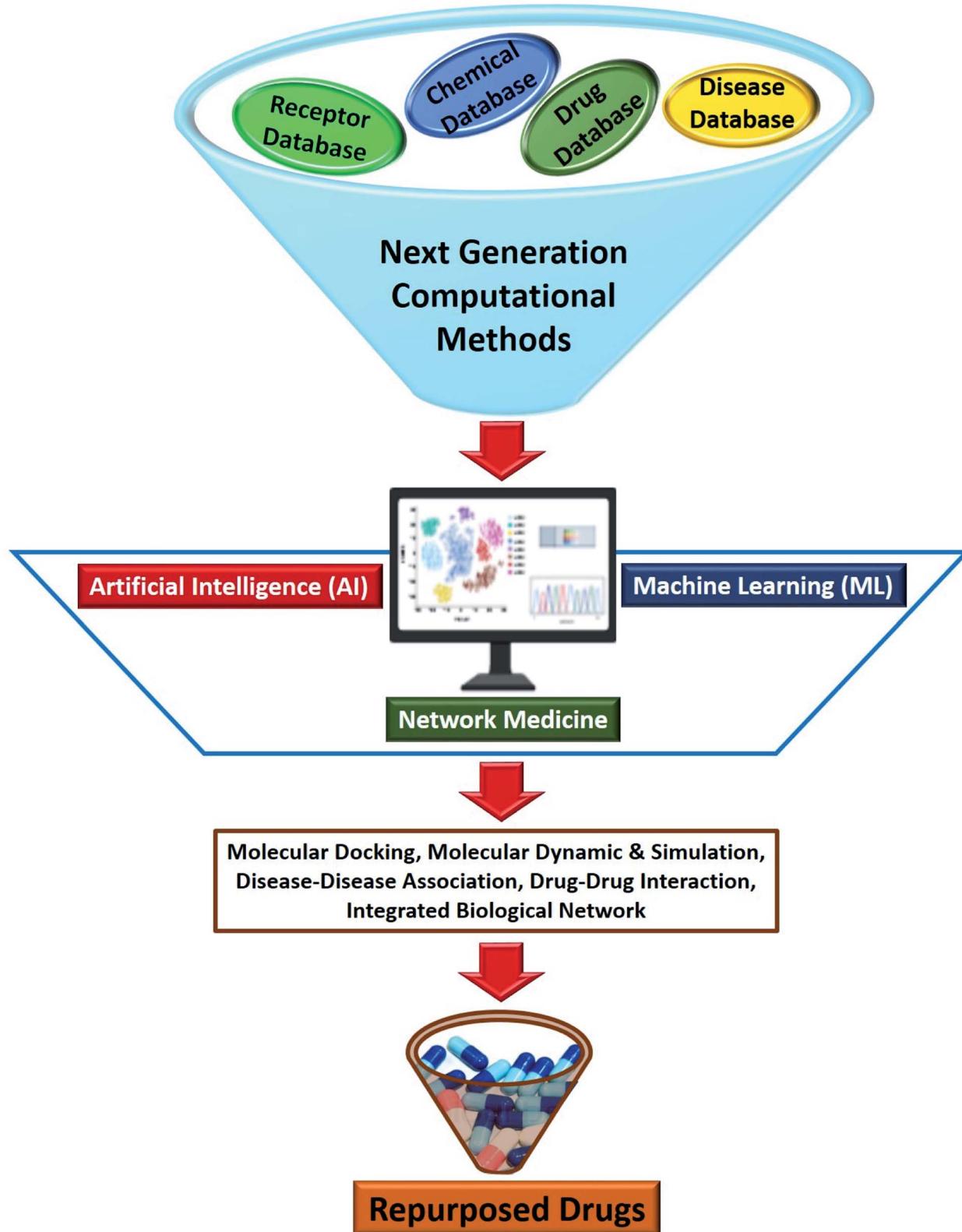


Fig. 4 Next generation computational methods based on Artificial Intelligence (AI), Machine Learning (ML) and network medicine for drug repurposing.

with ACE2 protein through residues L455, F486, Q493, N487, and S494. The residues L455, F486, Q493 interacted to K31 residues while N487, and S494 interacted to K353 residues of

ACE2. The molecules KT203, BMS195614, KT185, RS504393, and GSK1838705A interacted to residues of receptor-binding domain majorly by hydrophobic interaction except



BMS195614 (H-bond N487, and S494), and RS504393 (H-bond N487). Sethi *et al.*, proposed TD-139, a galectin inhibitor currently in phase IIb clinical trials to treat idiopathic pulmonary fibrosis, to be able to disrupt the SARS-CoV-2 spike protein and ACE2 interaction (PDB: 6M0J).<sup>44</sup> It was shown in several prior studies that the coronaviruses have a galectin fold within their spike protein and additionally that glycosylated ligands have the potential to bind to spike protein. This led the authors to investigate galectin inhibitors with the potential to disrupt SARS-CoV-2:ACE2 interaction. The TD-139 interacted with residues E471, I472, and Q474 through H-bond. TD-139 is now being assessed in phase III clinical trials as a rapid experimental medicine to treat COVID-19.<sup>45</sup>

These pre-developed and pre-evaluated molecules hold the potential to assist in the control of COVID-19 pandemic by not only inhibiting the entry or replication of virus but also serving as anti-inflammatory agents that may offer relief from the pulmonary inflammation.

**Limitations.** Screening of large ligand libraries require high-end system or workstation, which are not accessible commonly, similar problem plagues small research groups carrying out MD simulations. In addition, this methodology relies on structural files, which are not immediately available on identification of a new or orphan target.<sup>46</sup> Moreover, majority of the studies carried out are not backed by any experimental validation that does not build confidence in this methodology. Certain experimental studies in collaboration or otherwise on these *in silico* approaches and their hit enrichment would be welcomed and considered valuable.

### 3. Diseases–disease association

Different diseases may have overlapping molecular mechanisms underlying their pathobiology, which is known as disease–disease association. Molecular relationships between distinct disease pathologies need to be explored for drug repurposing, even if there is no apparent similarity at the phenotypic or clinical level.<sup>9</sup> There are a few methods of drug repurposing based on only disease–disease relationships. In these computational methods, the transcriptional signatures of the drugs and targets, the pathways perturbed as well as the diseases and side effects serve as multifarious modes for carrying out the drug repurposing. Transcriptional signatures of molecules could be used to explore therapeutic relationships between known drugs and new indications.<sup>47</sup> It is based on a similarity between the transcriptomic signature of a particular disease and the specific signature of a drug, resulting in a connectivity map (CMap) being developed.<sup>48</sup>

#### 3.1 Methodology

Disease–disease associations-based drug repurposing strategies use the DESeq2 algorithms,<sup>49</sup> which show the strength of associations in *p*-values as shown in Fig. 2.

The key steps involved in disease–disease association are illustrated below.

(a) Download microarray dataset from Gene Expression Omnibus (GEO) datasets of National Center for Biotechnology Information (NCBI);

(b) Normalize datasets *i.e.*, the value distributions are median-centered across samples;<sup>50</sup>

(c) Determine differential gene expression using DESeq2;

(d) Analyse gene expression based on *p*-value results.

There are some other platforms like DisGeNET that may also be explored for drug repurposing against COVID-19.<sup>47</sup> In order to find alternative therapies for COVID-19, Mousavi and his team members conducted drug repurposing analysis focused on the CMap concept.<sup>48</sup> They studied differentially expressed genes (DEGs) in NHBE, a primary human lung epithelial cell line and A549, a transformed lung alveolar cell line infected with SARS-CoV-2. Then, the identified DEGs were analyzed by connectivity map (CMap) with the goal of identifying similarly acting drug candidates.

Several drugs demonstrated positive effects in reversing cell gene expression profiles in SARS-CoV-2 infected cells, including lansoprazole, folic acid, sulfamonomethoxine, tolnaftate, diclofenamide, halcinonide, saquinavir, metronidazole, ebselen, lidocaine and benzocaine, histone deacetylase (HDAC) inhibitors, and heat shock protein 90 (HSP 90) inhibitors (Table 1).<sup>48</sup> Killick *et al.* used the transcriptional response of lung epithelial cells following infection by the original Severe Acute Respiratory Syndrome coronavirus (SARS) to identify repurposable drugs against SARS-CoV-2.<sup>51</sup> The authors combined CMap and LINCS repositories of the transcriptomic effects of approved drugs, which provide signature data enrichment. They found positive correlation of the SARS signature and the striking result that half of the high scoring drugs have reported anti-viral activities (23/45 in CMap and 22/45 in LINCS). Among the drugs are emetine, ouabain, digoxin, niclosamide, homoharringtonine, mefloquine, perhexiline, thioridazine, and thapsigargin shown to have the potential to inhibit SARS-CoV-2.

**Limitation.** The effectiveness of these methods would be constrained by the capacity to quantify the molecular pathology underlying the disease and represent it. The process of modeling the molecular state of diseases by computational methods is complicated because of the abundance of different incorporated molecular entities and organ systems in a disease pathology. There are some issues related with this method of drug repurposing such as the compound set of CMap was not large enough for performing integrative statistical analysis. Additionally, to the best of our knowledge, there are no published success stories for such repurposing yet.

### 4. Drug–drug interaction (DDI)

This section has been included even though drug–drug interaction does not directly aid in drug repurposing, because it is a crucial component to keep in mind when repurposing drugs. The adverse events due to DDI can be avoided with a prior knowledge. Many commercial DDI knowledge bases are widely used in hospitals and pharmacies such as Micromedex, Multum, Medispan, First Databank (FDB) *etc.*<sup>52</sup> We discuss IBM Micromedex in detail.

#### 4.1. Methods

Drug–drug interaction method includes the following steps<sup>53</sup>



- (a) Search for potential DDIs between diseases on IBM Micromedex;
- (b) Select medication for diseases from the search field and add to the 'drugs to check' in the interaction tool;
- (c) Collect information on DDIs manually using Micromedex. Micromedex grades the severity of DDIs as contraindicated, significant, mild, uncertain; and
- (d) Additionally, micromedex also offers an overview of all DDIs and clinical management data on some cases.

With IBM Micromedex's drug interaction checker, Baburaj and his team members performed a test for possible DDIs between lung cancer treatments and repurposed COVID-19 products.<sup>53</sup> A total of 61 potential DDIs between COVID-19 repurposed drugs and lung cancer medications along with their severity were identified from Micromedex. Mechanism of DDI of repurposed COVID-19 drugs with lung cancer pharmacotherapies includes several factors such as pharmacodynamic interactions, QT prolongation, pharmacokinetic interactions, drug absorption, drug metabolism, and drug transportation. Authors found a high chance for the occurrence of major DDIs associated with the concomitant use of COVID-19 repurposed treatments with lung cancer medications, with QT prolongation being the most commonly identified DDI.<sup>53</sup> The QT prolongation, measured on an electrocardiogram (ECG), represents the duration of the ventricular electrical systole results onto cardiac arrhythmias.<sup>54</sup> In COVID-19 patients, the addition of azithromycin to hydroxychloroquine resulted in superior viral clearance compared with hydroxychloroquine alone.<sup>55</sup> But chloroquine and hydroxychloroquine are associated with QT prolongation while azithromycin exposure has been reported to cause ventricular arrhythmia, and sudden cardiac arrest. Hence, concomitant use of azithromycin with hydroxychloroquine can cause a greater change in the QT prolongation than hydroxychloroquine alone.<sup>53</sup> This raises alert for clinicians and pharmacists for developing novel strategies for screening, identifying, reporting, and managing potential DDIs in lung cancer patients co-infected with COVID-19 infection. Administration of ritonavir along with oral prednisolone may increase systemic corticosteroid exposure and result in the development of Cushing syndrome.<sup>56</sup> Therefore, concomitant use of these drugs with powerful inhibitors of CYP3A, such as lopinavir and ritonavir, can raise the risk of adverse reactions and toxicity. Thus, possible medication interactions between cancer supportive care medications and repurposed COVID-19 drugs are necessary to consider.<sup>53</sup>

**Limitations.** Variability in different knowledge databases is a significant limitation of the technique. A standard set of DDI database needs to be developed and maintained to account for this variability.

## 5. Integrating biological networks in drug repositioning

Protein–protein interaction (PPI) data represent the proteins, which can physically interact, but it does not suggest any information about how an organism will react to the stimuli.

Gene expression data, however may display how an organism reacts to stimuli regarding the amount of RNA produced, but does not provide information on the physical mechanisms by which changes in the actions of the organism are manifested. Every type of link in a network represents specific information about an organism. Therefore, the integration of distinct types and ESI† would seem necessary.<sup>9</sup> Drug repositioning semantic mining (DReSMin)<sup>57</sup> as well as bipartite network projection and prioritization algorithms<sup>58</sup> suggest integration of different biological networks to reposition drugs. In these approaches, each prediction is made separately based on distinct metrics, and then these metrics are aggregated to arrive at the final result.<sup>9</sup> Zhang *et al.*, proposed an approach that combines literature-based discovery (lbd) and knowledge graph completion for COVID-19 drug repurposing where the ABC model was considered for lbd in their study.<sup>59</sup>

### 5.1. Method

Semantic predications (SemMedDB and CORD-19 relations) having semantic relations between subject and object, automatically extracted from biomedical literature (PubMed and CORD-19) utilizing SemRep natural language processing tool (Fig. 3). SemMedDB has supported a wide range of computational applications, ranging from gene regulatory network inference<sup>60</sup> to *in silico* screening for drug repurposing,<sup>61</sup> medical reasoning,<sup>62</sup> and has also found widespread applicability for the literature-based knowledge discovery and hypothesis generation.<sup>63</sup> LitCovid, NCBI's bibliographic database for more than 83k COVID-19 literature (accessed on 27-12-2020).<sup>64</sup> COVID-19 Open Research Dataset (CORD-19), which contains articles more than 200k articles<sup>65</sup> but not all of them are included in PubMed.<sup>59</sup> Both, PubMed and CORD-19 were used for adequate coverage of COVID-19 knowledge. Data were preprocessed before developing biomedical knowledge graph to remove some biases such as (i) eliminated relations involving generic biomedical concepts and relations with identical subject and object arguments; (ii) excluded a subset of predicate types that were not expected to be useful for drug repurposing; (iii) excluded the relations in which the subject or the object belongs to one of the following semantic groups: activities & behaviors, concepts & ideas, objects, occupations, organizations, and phenomena; and (iv) eliminated uninformative semantic relations. Normalization was also done to the range [0; 1] and summed them up into a final score.

Knowledge graph completion methods include multiple algorithms such as TransE<sup>66</sup> and RotatE<sup>67</sup> for translational models, DistMult<sup>68</sup> and ComplEx<sup>69</sup> for semantic matching models, and STELP<sup>70</sup> for context-based encoding. The TransE has simplicity and good prediction performance but is able to model only one-to-one relations and fails to embed one-to-many, many-to-one, and many-to-many relations. Therefore, RotatE is considered to counter TransE limitations. DistMult is also the simplest approach among semantic matching models but is limited only to symmetric relations. Hence, ComplEx extends DistMult to the complex domain. Semantic Triple Encoder for Link Prediction (STELP) is a context-based



encoding approach to knowledge graph completion. At its core is a Siamese BERT model that leverages sharing one set of weights across two models to produce encoded, contextual representations of the predication that are then fed to either multi-layer perceptron (MLP) for classification or a similarity function for contrasting. By using these models, open discovery patterns can be explored to identify drugs that can be repurposed for COVID-19 as well as can propose plausible mechanisms for the drug identification.

**Limitations.** As this approach relies on the accuracy of the predication extracted by SemRep. SemRep precision is about 0.70. Using models like vanilla BERT, BioBERT, Bio-ClinicalBERT, BlueBERT, and PubMedBERT were not enough to rectify all errors impacting the knowledge graph completion task. In addition, forming large graphs create obstruction to apply computationally intensive models like STELP. Therefore, a sub-graph is examined but there is a loss of information available to knowledge graph completion techniques. Our methods were limited to knowledge from the literature. Other types of biological data (*e.g.*, protein–protein interactions, drug–target interactions, gene/protein sequences, pharmacogenomic and pharmacokinetic data) are likely to benefit identification of drug candidates, as shown to some extent by other studies. However, the computational resources needed for training models based on such massive data can be prohibitive.<sup>9</sup> Semantic MEDLINE Database (SemMedDB) is usually not as accurate due to the limitations of the extraction algorithms used. For example, the extraction precision of SemRep, the tool used to generate SemMedDB, is only 74.5%. However, because it contains many noisy information, directly applying SemMedDB sometimes may lead to suboptimal performance.<sup>71</sup>

## 6. Artificial intelligence (AI), machine learning (ML) and network medicine approaches

In recent years, the development of next generation computational methods using Artificial Intelligence (AI), Machine Learning (ML) and network medicine approaches have positively impacted the different stages of a drug development process (Fig. 4).<sup>72,73</sup> AI and ML-based tools have been the main focus of the pharmaceutical industry which have significantly helped in property prediction and target proteins and small molecule search. These tools can support drug repurposing approaches by rapidly accelerating the pace of all the approaches discussed in previous sections. The first AI invented drug to enter clinical trials took 12 months as opposed to 5 years through a conventional method.<sup>74</sup> Additionally, using the AI and ML approaches, researchers have found many similarities between the regulatory proteins of SARS-CoV-2 and SARS.<sup>75</sup> Huge amount of COVID-19 data including clinical as well as basic research data is now available in public domain, which can be further used to train the AI and ML based model to repurpose smart drugs against SARS-CoV-2 infection.<sup>76</sup> Several reviews have elegantly summarized the principles and approaches to drug discovery using ML/AI.<sup>77,78</sup>

### 6.1. Methods

As summarized in the reviews,<sup>71,72</sup> AI has found applications in different sectors of COVID management, (1) tracking through contact tracing; (2) diagnosis; (3) screening of potential drug candidates; (4) proteomics and genomics; and lastly (5) automation of procedures to reduce health worker exposure. The general approach has been to fuse the structure-based and ligand-based screening methods with AI algorithms to build prediction models. The AI and ML algorithms like deep learning, support vector machine (SVM), random forest (RF), Naive Bayesian, and neural networks have been used extensively for high throughput screening with thousand plus dataset of molecules. Specialized statistical parameters, namely root-mean-square error (RMSE), Z-score, area under the curve (AUC), mean absolute error (MAE), Cohen's *K* value, correlation coefficient ( $R^2$ ), *etc.* are used to evaluate the performance (precision + accuracy) of the models.

**Limitations.** This approach is very promising and has proved its effectiveness in other domains. For *ex* – natural language processing has proved its utility in translation, spell check *etc.* However, AI/ML based techniques require large data to train the models. The lack of accessibility of structured and standardized data associated with assays and clinical trials could compromise predictive ability of the tools. Moreover, the majority of models developed are local models, meaning they're specific to one problem and there does not exist any global model or suite, which can be supportive in resolving or querying a wide variety of problems that a drug discovery team might encounter frequently.

## 7. Conclusion

Drug repurposing or repositioning is a technique in which existing drugs are remodeled to treat other diseases. In last few months, COVID-19 has greatly affected mankind. Though several vaccines have been developed recently, yet, no therapeutic drugs have been found to be effective. However, there are number of available drugs approved by the FDA for the treatment of other diseases, and these can be potentially repurposed for COVID-19. In recent years, the development of next generation computational methods including molecular docking, MD simulation, disease–disease association, drug–drug interaction, integrated biological network, artificial intelligence (AI), machine learning (ML) and network medicine have paved the way to repurpose drugs rapidly. Although drug repurposing has huge potential to decrease the time, resources and money required for a drug to reach the market compared to the conventional drug discovery process, several challenges still remain that need to be overcome. These challenges are twofold.

One inherently associated with drug repurposing; the commercial and legal barriers. Repurposed drugs are still a high-risk investment. One may argue that a pre-approved drug has a suitable safety profile and thereby phase I and II studies could be avoided for repurposed drug. However, as regulatory science progresses so do its standards, and therefore, the previous data might no longer be applicable for the new



indication. Even if the safety profile is found to meet present standards, the variation in benefit for the new indication might alter the overall benefit-risk profile of the drug. Even from the perspective of patentability, it faces certain hurdles. If there are studies in scientific literature/public domain that highlight the potential role of the preexisting drug for the new indication, this affects the novelty of the repurposed drug and thereby patentability. Thus, there is an urgent need to incentivize the repurposing of drugs as many believe that existing drugs would be beneficial in treating other ailments.

The second challenge is associated with the current pitfalls of the use of computational tools. The predictability of computational models has been time and again questioned. Specifically, the binding mode of interaction in molecular docking. Different software packages use different algorithms and scoring functions which leads to variability and inconsistency in data associated with docking. AI//ML based tools rely on large datasets, however, the lack of dissemination of clinical trial data by pharmaceutical companies could hamper predictability of AI//ML algorithms due to insufficient size of training data. Additionally, the data which is available is heterogenous, which makes integration of this data an extremely pain staking endeavor. Obtaining openly accessible standardized and structured data is crucial for progress of computational tools in drug repurposing. Every method has its own set of pitfalls, even *in vitro* data generated in anti-viral assays failed to translate *in vivo* due to phospholipidosis.<sup>79</sup> Thus, results generated by every tool at our disposal need to be critically analyzed and scrutinized, to improve it further.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

This work was supported by Department of Science & Department of Science & Technology (DST) under BRICS multilateral project (DST/INT/BRICS/COVID-19/2020). This study was funded by RFBR, DST, CNPq, SAMRC, project number 20-53-80002.

## References

- 1 O. J. Wouters, M. McKee and J. Luyten, *JAMA, J. Am. Med. Assoc.*, 2020, **323**, 844.
- 2 M. Placchi and R. J. Phillips, *The Benefits and Pitfalls of Repurposing Drugs*, <https://oliver-dev.s3.amazonaws.com/2018/11/07/21/59/08/240/Benefits%20and%20Pitfalls%20of%20Repurposing%20Drugs.pdf>, accessed February 26, 2021.
- 3 L. Riva, S. Yuan, X. Yin, L. Martin-Sancho, N. Matsunaga, L. Pache, S. Burgstaller-Muehlbacher, P. D. De Jesus, P. Teriete, M. V. Hull, M. W. Chang, J. F.-W. Chan, J. Cao, V. K.-M. Poon, K. M. Herbert, K. Cheng, T.-T. H. Nguyen, A. Rubanov, Y. Pu, C. Nguyen, A. Choi, R. Rathnasinghe, M. Schotsaert, L. Miorin, M. Dejosez, T. P. Zwaka, K.-Y. Sit, L. Martinez-Sobrido, W.-C. Liu, K. M. White, M. E. Chapman, E. K. Lendy, R. J. Glynne, R. Albrecht, E. Ruppin, A. D. Mesecar, J. R. Johnson, C. Benner, R. Sun, P. G. Schultz, A. I. Su, A. García-Sastre, A. K. Chatterjee, K.-Y. Yuen and S. K. Chanda, *Nature*, 2020, **586**, 113.
- 4 L. Yang, R.-j. Pei, H. Li, X.-n. Ma, Y. Zhou, F.-h. Zhu, P.-l. He, W. Tang, Y.-c. Zhang, J. Xiong, S.-q. Xiao, X.-k. Tong, B. Zhang and J.-p. Zuo, *Acta Pharmacol. Sin.*, 2020, **42**, 1347.
- 5 S. Naylor and J. M. Schonfeld, *Therapeutic Drug Repurposing, Repositioning and Rescue Part I: Overview*, <https://www.ddw-online.com/therapeutic-drug-repurposing-repositioning-and-rescue-part-i-overview-1463-201412/>, accessed March15, 2021.
- 6 G. Galindez, J. Matschinske, T. D. Rose, S. Sadegh, M. Salgado-Albarrán, J. Späth, J. Baumbach and J. K. Pauling, *Nat. Comput. Sci.*, 2021, **1**, 33.
- 7 A. Persidis, *The Benefits of Drug Repositioning*, <https://www.ddw-online.com/the-benefits-of-drug-repositioning-1779-201104/>, accessed January 26, 2021.
- 8 S. G. V. Rosa and W. C. Santos, *Revista Panamericana de Salud Pública*, 2020, **44**, e40.
- 9 M. Lotfi Shahreza, N. Ghadiri, S. R. Mousavi, J. Varshosaz and J. R. Green, *Briefings Bioinf.*, 2017, **19**, 878.
- 10 R. A. Hodos, B. A. Kidd, K. Shameer, B. P. Readhead and J. T. Dudley, *Wiley Interdiscip. Rev.: Syst. Biol. Med.*, 2016, **8**, 186.
- 11 P. Polamreddy and N. Gattu, *Drug Discovery Today*, 2019, **24**, 789.
- 12 K. Gao, D. D. Nguyen, J. Chen, R. Wang and G.-W. Wei, *J. Phys. Chem. Lett.*, 2020, **11**, 5373.
- 13 M. Rudrapal, S. Khairnar and A. Jadhav, in *Drug Repurposing - Hypothesis, Molecular Aspects and Therapeutic Applications*, ed. F. A. Badria, IntechOpen, 2020, DOI: 10.5772/intechopen.93193.
- 14 W. Yin, X. Luan, Z. Li, Y. Xie, Z. Zhou, J. Liu, M. Gao, X. Wang, F. Zhou, Q. Wang, Q. Wang, D. Shen, Y. Zhang, G. Tian, H. A. Aisa, T. Hu, D. Wei, Y. Jiang, G. Xiao, H. Jiang, L. Zhang, X. Yu, J. Shen, S. Zhang and H. E. Xu, *bioRxiv*, 2020, DOI: 10.1101/2020.11.01.363812.
- 15 A. Tuerkova and B. Zdravil, *J. Cheminf.*, 2020, **12**, 71.
- 16 A. B. Gurung, M. A. Ali, J. Lee, M. A. Farah and K. M. Al-Anazi, *BioMed Res. Int.*, 2021, **2021**, 8853056.
- 17 T. N. Jarada, J. G. Rokne and R. Alhajj, *J. Cheminf.*, 2020, **12**, 46.
- 18 H. Tayara, I. Abdelbaky and K. To Chong, *Brief. Bioinform.*, 2021, DOI: 10.1093/bib/bbab339.
- 19 E. Hernández-Domínguez, L. Castillo-Ortega, Y. García-Esquivel, V. Mandujano-González, G. Diaz-Godínez and J. Alvarez Cervantes, in *Computational Biology and Chemistry*, ed. P. Behzadi, IntechOpen, 2019. DOI: DOI: 10.5772/intechopen.89594.
- 20 R. Angles, M. Arenas-Salinas, R. García, J. A. Reyes-Suarez and E. Pohl, *BMC Bioinf.*, 2020, **21**, 85.
- 21 F. Sohraby, M. Bagheri and H. Aryapour, *Methods Mol. Biol.*, 2019, **1903**, 23.
- 22 G. Vettoretti, E. Moroni, S. Sattin, J. Tao, D. A. Agard, A. Bernardi and G. Colombo, *Sci. Rep.*, 2016, **6**, 23830.



23 Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L. W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao and H. Yang, *Nature*, 2020, **582**, 289.

24 M. De Vivo, M. Masetti, G. Bottegoni and A. Cavalli, *J. Med. Chem.*, 2016, **59**, 4035.

25 A. Gimeno, J. Mestres-Truyol, M. J. Ojeda-Montes, G. Macip, B. Saldivar-Espinoza, A. Cereto-Massagué, G. Pujadas and S. Garcia-Vallvé, *Int. J. Mol. Sci.*, 2020, **21**, 3793.

26 J. Wang, *J. Chem. Inf. Model.*, 2020, **60**, 3277.

27 M. Mahdi, J. A. Mótyán, Z. I. Szojka, M. Golda, M. Miczi and J. Tőzsér, *Virol. J.*, 2020, **17**, 190.

28 F. Monji, A. Al-Mahmood Siddiquee and F. Hashemian, *Eur. J. Pharmacol.*, 2020, **887**, 173561.

29 A. O. Elzupir, *J. Biomol. Struct. Dyn.*, 2020, DOI: 10.1080/07391102.2020.1835732.

30 S. Koulgi, V. Jani, M. Uppuladinne, U. Sonavane, A. K. Nath, H. Darbari and R. Joshi, *J. Biomol. Struct. Dyn.*, 2020, **39**, 5735.

31 S. A. Khan, K. Zia, S. Ashraf, R. Uddin, Z. Ul-Haq and Z. Ul-Haq, *J. Biomol. Struct. Dyn.*, 2020, **39**, 2607.

32 S. A. Cherrak, H. Merzouk and N. Mokhtari-Soulimane, *PLoS One*, 2020, **15**, e0240653.

33 S. Upadhyay, P. K. Tripathi, S. Raghavendar, M. Bhardwaj, A. K. Patel and M. Singh, *Phytother. Res.*, 2020, **34**, 3411.

34 S. Gul, I. H. Kavakli, O. Ozcan, I. Baris, I. H. Kavakli, S. Asar and A. Okyar, *J. Biomol. Struct. Dyn.*, 2021, **39**, 6772.

35 M. Wang, R. Cao, L. Zhang, X. Yang, J. Liu, M. Xu, Z. Shi and Z. Hu, *Cell Res.*, 2020, **30**, 269.

36 P. K. Singh, S. Pathania and R. K. Rawal, *SAR QSAR Environ. Res.*, 2020, **31**, 857.

37 S. Barage, A. Karthic, R. Bavi, N. Desai, R. Kumar, V. Kumar and K. W. Lee, *J. Biomol. Struct. Dyn.*, 2020, DOI: 10.1080/07391102.2020.1841026.

38 S. Kumar, P. Kashyap, S. Chowdhury, S. Kumar, A. Panwar and A. Kumar, *Phytomedicine*, 2020, **85**, 153317.

39 R. J. Khan, R. K. Jha, E. Singh, M. Jain, G. M. Amera, R. P. Singh, J. Muthukumaran and A. K. Singh, *J. Biomol. Struct. Dyn.*, 2020, DOI: 10.1080/07391102.2020.1814870.

40 L. S. Jung, T. M. Gund and M. Narayan, *Protein J.*, 2020, **39**, 619.

41 M. A. White, W. Lin and X. Cheng, *J. Phys. Chem. Lett.*, 2020, **11**, 9144.

42 A. Sharma, V. Tiwari and R. Sowdhamini, *J. Biosci.*, 2020, **45**, 100.

43 S. Choudhary, Y. S. Malik and S. Tomar, *Front. Immunol.*, 2020, **11**, 14.

44 A. Sethi, S. Sanam, S. Munagalasetty, S. Jayanthi and M. Alvala, *RSC Adv.*, 2020, **10**, 29873.

45 *Rapid Experimental Medicine for COVID-19 (DEFINE)*, <https://clinicaltrials.gov/ct2/show/NCT04473053>, accessed September 26, 2021.

46 Q. Vanhaelen, P. Mamoshina, A. M. Aliper, A. Artemov, K. Lezhnina, I. Ozerov, I. Labat and A. Zhavoronkov, *Drug Discovery Today*, 2017, **22**, 210.

47 P. Balasundaram, R. Kanagavelu, N. James, S. Maiti, S. Veerappapillai and R. Karuppaswamy, *Methods Mol. Biol.*, 2019, **1903**, 129.

48 S. Z. Mousavi, M. Rahamanian and A. Sami, *Infect. Genet. Evol.*, 2020, **86**, 104610.

49 M. I. Love, W. Huber and S. Anders, *Genome Biol.*, 2014, **15**, 550.

50 T. Barrett, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C. L. Robertson, N. Serova, S. Davis and A. Soboleva, *Nucleic Acids Res.*, 2012, **41**, D991.

51 R. Killick, C. Ballard, P. Doherty and G. Williams, *Virus Res.*, 2020, **290**, 198176.

52 K. W. Fung, J. Kapusnik-Uner, J. Cunningham, S. Higby-Baker and O. Bodenreider, *J. Am. Med. Inform. Assoc.*, 2017, **24**, 806.

53 G. Baburaj, L. Thomas and M. Rao, *Arch. Med. Res.*, 2020, **52**, 261.

54 L. Guglielmetti, S. Tiberi, M. Burman, H. Kunst, C. Wejse, T. Togonidze, G. Bothamley and C. Lange, *Eur. Respir. J.*, 2018, **52**, 1800537.

55 P. Gautret, J.-C. Lagier, P. Parola, V. T. Hoang, L. Meddeb, M. Mailhe, B. Doudier, J. Courjon, V. Giordanengo, V. E. Vieira, H. Tissot Dupont, S. Honoré, P. Colson, E. Chabrière, B. La Scola, J.-M. Rolain, P. Brouqui and D. Raoult, *Int. J. Antimicrob. Agents*, 2020, **56**, 105949.

56 N. E. Albert, S. Kazi, J. Santoro and R. Dougherty, *Am. J. Med. Sci.*, 2012, **344**, 72.

57 J. Mullen, S. J. Cockell, H. Tipney, P. M. Woppard and A. Wipat, *PeerJ*, 2016, **4**, e1558.

58 A. J. Berenstein, M. P. Magariños, A. Chernomoretz and F. Agüero, *PLoS Neglected Trop. Dis.*, 2016, **10**, e0004300.

59 R. Zhang, D. Hristovski, D. Schutte, A. Kastrin, M. Fiszman and H. Kilicoglu, *J. Biomed. Inf.*, 2021, **115**, 103696.

60 G. Chen, M. J. Cairelli, H. Kilicoglu, D. Shin and T. C. Rindflesch, *PLoS Comput. Biol.*, 2014, **10**, e1003666.

61 T. Cohen, D. Widdows, C. Stephan, R. Zinner, J. Kim, T. Rindflesch and P. Davies, *CPT: Pharmacomet. Syst. Pharmacol.*, 2014, **3**, e140.

62 S. R. Sukumar, *A Reasoning And Hypothesis-Generation Framework Based On Scalable Graph Analytics*, Oak Ridge National Lab (ORNL), Oak Ridge, TN (United States), 2016.

63 G. Crichton, S. Baker, Y. Guo and A. Korhonen, *PLoS One*, 2020, **15**, e0232891.

64 Q. Chen, A. Allot and Z. Lu, *Nature*, 2020, **579**, 193.

65 L. L. Wang, K. Lo, Y. Chandrasekhar, R. Reas, J. Yang, D. Eide, K. Funk, R. M. Kinney, Z. Liu, W. Merrill, P. Mooney, D. Murdick, D. Rishi, J. Sheehan, Z. Shen, B. Stilson, A. D. Wade, K. Wang, C. Wilhelm, B. Xie, D. Raymond, D. S. Weld, O. Etzioni and S. Kohlmeier, 2020, ArXiv, <https://arxiv.org/abs/2004.10706>.

66 A. Bordes, N. Usunier, A. García-Durán, J. Weston and O. Yakhnenko, *NeurIPS*, 2013, vol. 2, p. 2787.

67 Z. Sun, Z.-H. Deng, J.-Y. Nie and J. Tang, 2019, arXiv e-prints, arXiv:1902.10197.



68 B. Yang, W.-t. Yih, X. He, J. Gao and L. Deng, 2014, arXiv e-prints, arXiv:1412.6575.

69 T. Trouillon, J. Welbl, S. Riedel, É. Gaussier and G. Bouchard, 2016, arXiv e-prints, arXiv:1606.06357.

70 B. Wang, T. Shen, G. Long, T. Zhou and Y. Chang, ArXiv, 2020, abs/2004.14781.

71 D. Zhang, D. He, N. Zou, X. Zhou and F. Pei, 2018 *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2018. DOI: 10.1109/BIBM.2018.8621316.

72 Y. Zhou, F. Wang, J. Tang, R. Nussinov and F. Cheng, *The Lancet Digital Health*, 2020, 2, e667.

73 W. P. Walters and R. Barzilay, *Expert Opin. Drug Discovery*, 2021, **16**, 1–11.

74 J. Wakefield, *Artificial intelligence-created medicine to be used on humans for first time*, <https://www.bbc.com/news/technology-51315462>, accessed July 26, 2021.

75 H. S. Gns, S. Gr, M. Murahari and M. Krishnamurthy, *Biomed. Pharmacother.*, 2019, **110**, 700.

76 K.-K. Mak and M. R. Pichika, *Drug Discovery Today*, 2019, **24**, 773.

77 A. Keshavarzi Arshadi, J. Webb, M. Salem, E. Cruz, S. Calad-Thomson, N. Ghadirian, J. Collins, E. Diez-Cecilia, B. Kelly, H. Goodarzi and J. S. Yuan, *Front. Artif. Intell.*, 2020, **3**, 1–13.

78 Z. Tanoli, M. Vähä-Koskela and T. Aittokallio, *Expert Opin. Drug Discovery*, 2021, **16**, 977.

79 T. A. Tummino, V. V. Rezelj, B. Fischer, A. Fischer, M. J. O'Meara, B. Monel, T. Vallet, K. M. White, Z. Zhang, A. Alon, H. Schadt, H. R. O'Donnell, J. Lyu, R. Rosales, B. L. McGovern, R. Rathnasinghe, S. Jangra, M. Schotsaert, J.-R. Galarneau, N. J. Krogan, L. Urban, K. M. Shokat, A. C. Kruse, A. García-Sastre, O. Schwartz, F. Moretti, M. Vignuzzi, F. Pognan and B. K. Shoichet, *Science*, 2021, 373, eabi4708.

