Bioorg. Med. Chem. Lett. 83 (2023) 129171

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

Target identification by structure-based computational approaches: Recent advances and perspectives

Simona De Vita [a], Maria Giovanna Chini [b], Giuseppe Bifulco [a,*], Gianluigi Lauro [a,*]

[a] Department of Pharmacy, University of Salerno, Via Giovanni Paolo II 132, 84084 Fisciano (SA), Italy
[b] Department of Biosciences and Territory, University of Molise, Contrada Fonte Lappone, 86090 Pesche (IS), Italy

ARTICLE INFO

Keywords:
Target identification
Drug repurposing
Computational methods
Small molecules
Polypharmacology

ABSTRACT

The use of computational techniques in the early stages of drug discovery has recently experienced a boost, especially in the target identification step. Finding the biological partner(s) for new or existing synthetic and/or natural compounds by "wet" approaches may be challenging; therefore, preliminary *in silico* screening is even more recommended. After a brief overview of some of the most known target identification techniques, recent advances in structure-based computational approaches for target identification are reported in this digest, focusing on Inverse Virtual Screening and its recent applications. Moreover, future perspectives concerning the use of such methodologies, coupled or not with other approaches, are analyzed.

Introduction

The research for new drugs is a long, risky, and expensive process that frequently ends up in a project failure. Considering that a single drug requires approximately 2.5 billion dollars and 15–17 years to be finally marketed, and that only 13 % of drugs receive the required authorizations, each inconclusive study represents a massive loss for the pharmaceutical industry[1–4] (Figure 1). In this scenario, the target identification phase is of particular interest since it represents the core of the whole pre-clinical pipeline and dictates the route of the next synthetic efforts. Related to this, observed side effects, caused by the interaction between the drug and targets that fall outside the mechanism of action, represent another important aspect and a key reason why many drugs fail clinical trials.[5,6]

Therefore, to balance the risk necessary to commercialize a new product, speed up the process, and limit the investment/loss ratio, more and more companies are resorting to drug repositioning, also known as drug repurposing, representing a subgroup of target identification methods. With this approach, which can also aid polypharmacology profiling (i.e., a single molecule can hit multiple targets),[6–9] it is possible to use already approved molecules for new therapeutic applications, avoiding the preclinical tests and safety profile assessment (Phase II),[2–3,8–11] On the other hand, the use of a single drug instead of a drug combination in the treatment of a disease can help fight the emerging phenomenon of drug resistance, derived from mutations in the target or variation in its expression, thus reducing the risk of cross-interactions between different medications.[3,9] Also, in the latter case, developing fast and efficient *in silico* methods aimed at accelerating the target identification step are even more required to shed light on the selectivity profile of pharmacologically relevant compounds.

Indeed, while in the past century drug repurposing was frequently achievable only by serendipity (e.g., when a particular side effect was observed, which was inexplicable through the presumed mechanism of action),[10,12] great support to the target identification process is nowadays given by computational chemistry, which can highlight new targets among thousands of proteins in a short time, disclosing new therapeutic applications and elucidating the occurrence of the observed side effects.[4] The improvement in computer processing power and the almost endless availability of data have boosted the use of *in silico* screenings in the pharmaceutical industry for both synthetic and natural products (NPs).[2,13,14] Concerning the NPs, in particular, the amount of the metabolites extracted from the natural source may be very small, and preliminary tests that do not consume the samples are preferable to better direct the next experiments onto the most promising pathways.

A clear and recent sign of the importance of target identification techniques is represented by the high number of papers published during the SARS-CoV-2 pandemic involving *in silico* screening and, specifically, computational drug repurposing[15–20] (37 % and 30 % of the total papers published in 2020 and 2021 respectively) (Figure 2).

In this respect, many known drugs were screened using *in silico*

* Corresponding authors.
E-mail addresses: bifulco@unisa.it (G. Bifulco), glauro@unisa.it (G. Lauro).

https://doi.org/10.1016/j.bmcl.2023.129171
Received 5 August 2022; Received in revised form 15 December 2022; Accepted 1 February 2023
Available online 3 February 2023

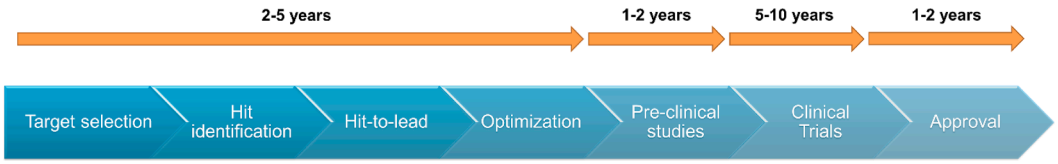

Figure 1. Drug development process with corresponding estimated time.

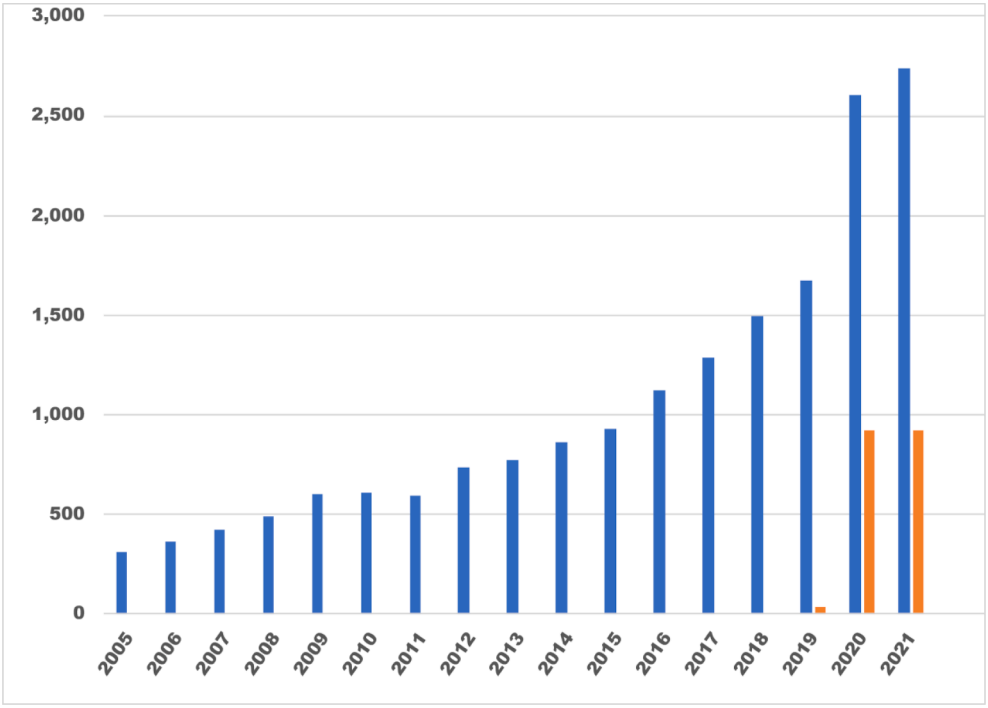

Figure 2. Total drug repurposing papers published per year (blue bars) and SARS-CoV-2-related ones (orange bars). The textual search was carried out on SciFinder selecting "computational drug repurposing" as query.

methods to obtain rapid and reliable responses to the overwhelming pandemic, and some examples[21] are reported in Figure 3. Interestingly, it can be noticed that both NPs and synthetic compounds were repurposed for this disease.

Computational methods for target identification are typically divided into data-driven (DD), ligand-based (LB), or structure-based (SB) according to whether they analyze the available data, known compounds, or the biological macromolecule involved (Figure 4).[3,7,13,22] The scope of this work is to analyze the latest papers published (2015–2022) concerning the techniques used in structure-based target identification, among which Inverse Virtual Screening (IVS) has gained great importance. This methodology, which will be discussed extensively in the following sections, is a docking-based technique that comprises both the exploration of the potential binding of a small molecule towards a panel of protein targets and the use of a set of "decoy" molecules, the latter needed for normalizing the obtained predicted ligand/macromolecule binding affinities. Therefore, we will briefly discuss LB and DD approaches, before moving to SB ones. For a complete discussion of these methods, please refer to other interesting recent reviews.[23–26]

Data-driven methods (DD) are straddling ligand and structure-based approaches as they can use either compound or target-related information to derive new possible applications. They represent the most innovative type of target identification approach as they emerged with the development of artificial intelligence (AI) and genomics. Conventionally they are divided into machine learning (ML), network analysis (NA), and text mining (TM) (Figure 5).

The considerable amount of health-related data available online has enabled researchers to implement ML and its subdisciplines like deep learning (DL) in the target identification process. In this respect, gene expression analysis is particularly useful to guide the selection of the most suitable molecular counterpart for a given molecule or to point out the protein responsible for a particular biological effect.[27] The capability to analyze intricate data sets and take decisions by continuously "learning" from the input data that could be structural or omic-related,[28,29] represents the main advantage of this approach. The reasons for the rapid popularity that DL methods are gaining in all the drug discovery fields and, in particular, in target identification are their scalability and reliability,[30–31] but their main limitation is related to the data availability because having a small dataset can lead to overfitting problems that will mislead data interpretation.[29] Despite that, numerous free algorithms have been developed and are available to the scientific community.[27] The two main categories of DL methods are Supervised Learning and Semi-Supervised; the first one requires both positive and negative data in the input, whereas the second one relies only on a small quantity of positive-labeled input data.[28] Handling all this information requires a not neglectable effort in developing algorithms capable of integrating and interpreting large-scale datasets and, in this scenario, network-based approaches have been extensively applied.[12,32–34]

Network analysis represents the most used among the data-driven approaches because it is capable of putting together heterogeneous data types and connecting them.[29] In this way, the pathology is fragmented into multiple subnetworks that can be addressed by pharmacological therapies, allowing the identification of common actors in

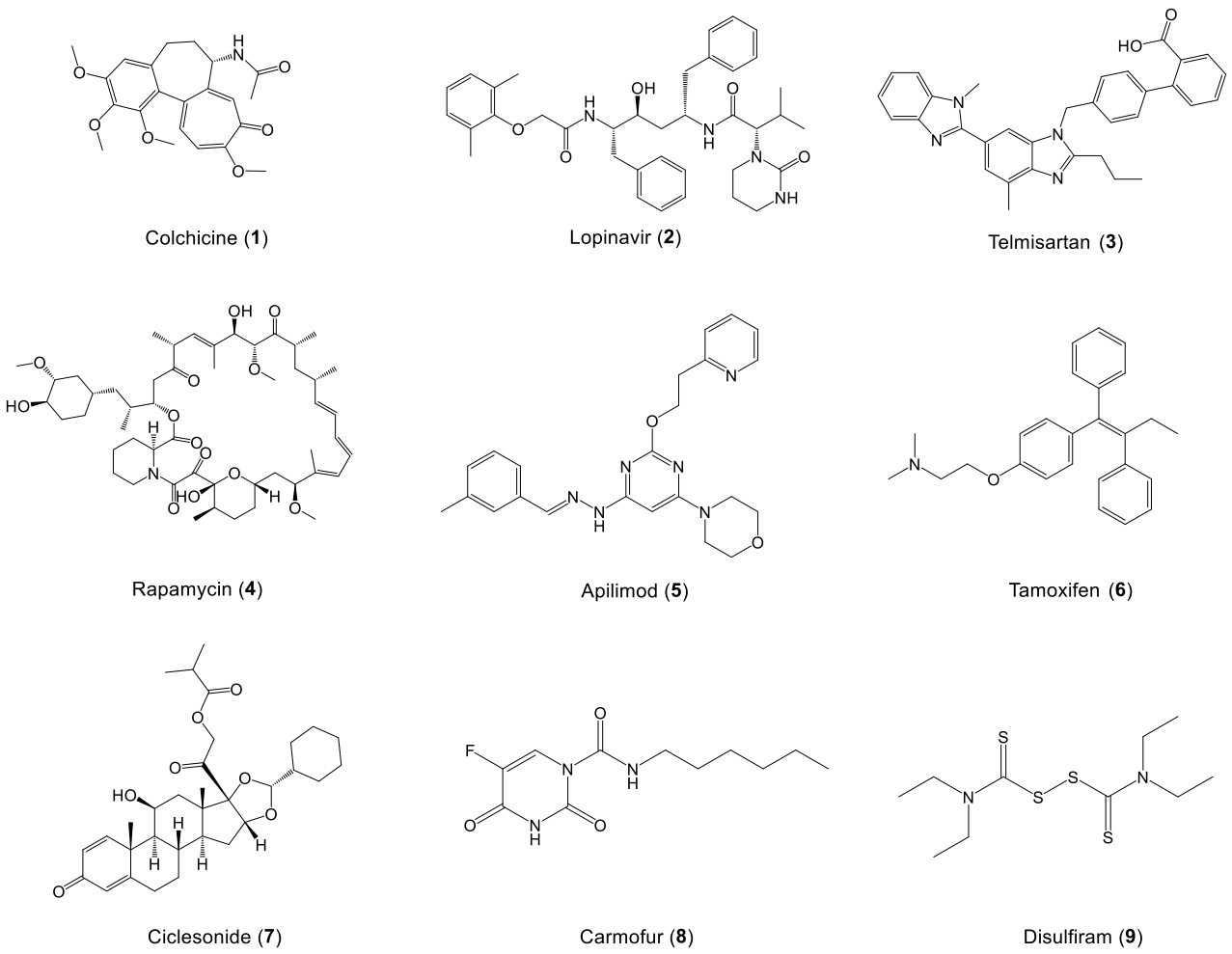

Figure 3. Examples of known drugs repurposed for the treatment of COVID-19.

different pathological states that can be targeted by known compounds. The construction of network-based analysis methods and employing omics technologies are the only way to look into otherwise very complex diseases.[4] The corresponding algorithms and their applications are fully described elsewhere.[33,35] Finally, also text-mining and semantic searches can be carried out using ML on medical and disease databases to extract the required information.

Ligand-based (LB) methods sink their roots into the knowledge and exploitation of already existing ligands, trying to derive from them the features required for the pharmacological activity (Figure 6);[26] this type of methodology has the strong limitation of being knowledge-based and, therefore, can hardly lead to the discovery of brand-new chemical scaffolds.[4,7]

The similarity search is the most common ligand-based method.[36] It uses known molecules as comparison terms to highlight, as the name suggests, similar compounds available in online databases like DrugBank,[37] PubChem[38], or ChEMBL[39] and, based on the principle that analogous structures could exert the same biological effect, predict new interacting partners.[7,36,40]

To perform the search, the ligand structure must be broken down into its physical–chemical properties, called "descriptors", that can help in grouping molecules that meet particular requirements, regardless of the structure.[4,7,32,40] These molecular descriptors are computer-readable multidimensional vectors that carry a particular property, and the correspondence between the query molecule and the database can be estimated by various parameters like the Tanimoto coefficient.[36] In particular, the 3D descriptors could be assembled to form pharmacophores, namely the chemical features that a compound must contain to match the characteristics of a protein binding site and generate a biological response.[41] In this way, if a high degree of similarity is retrieved by comparing the pharmacophores derived from multiple ligands, this can suggest that the two molecules could bind to the same targets. It is the case of a study reported by Steindl *et al.*[42] in which they used a "parallel screening" to assign the correct target to a group of HIV antiviral drugs. Analogously, the quantitative structure-activity relationship (QSAR) uses the structural characteristics of the ligand to create a correlation with its biological activity using mathematical manipulations. Unlike classical SAR, which tries to predict compounds' activity through the analysis of their structure, QSAR assumes that it is possible to quantitatively compare the biological activities through the characteristics of their structural components.[43]

Concerning structure-based (SB) methods, with the increasing number of macromolecule structures deposited online in dedicated databases, e.g., ~201,000 crystals available in the Protein Data Bank[44] (accessed on February 2023), the data regarding the three-dimensional characteristics of a particular protein are gaining a predominant role in computational methodologies and will be the focus of the remaining part of this digest (Figure 7).

As discussed in the previous section, the molecular descriptors, together with the structural information inferable from the analysis of protein/ligand co-complexes, could also be combined to generate a 3D structure-based pharmacophore model, as was recently reported for BRD9 by Pierri *et al.* [45]. In this respect, clear evidence of the improvement provided by the use of pharmacophores in drug screening is reported by Meslamani *et al.* [46]. In addition, greater attention is being paid to the development of online pharmacophore databases to speed up

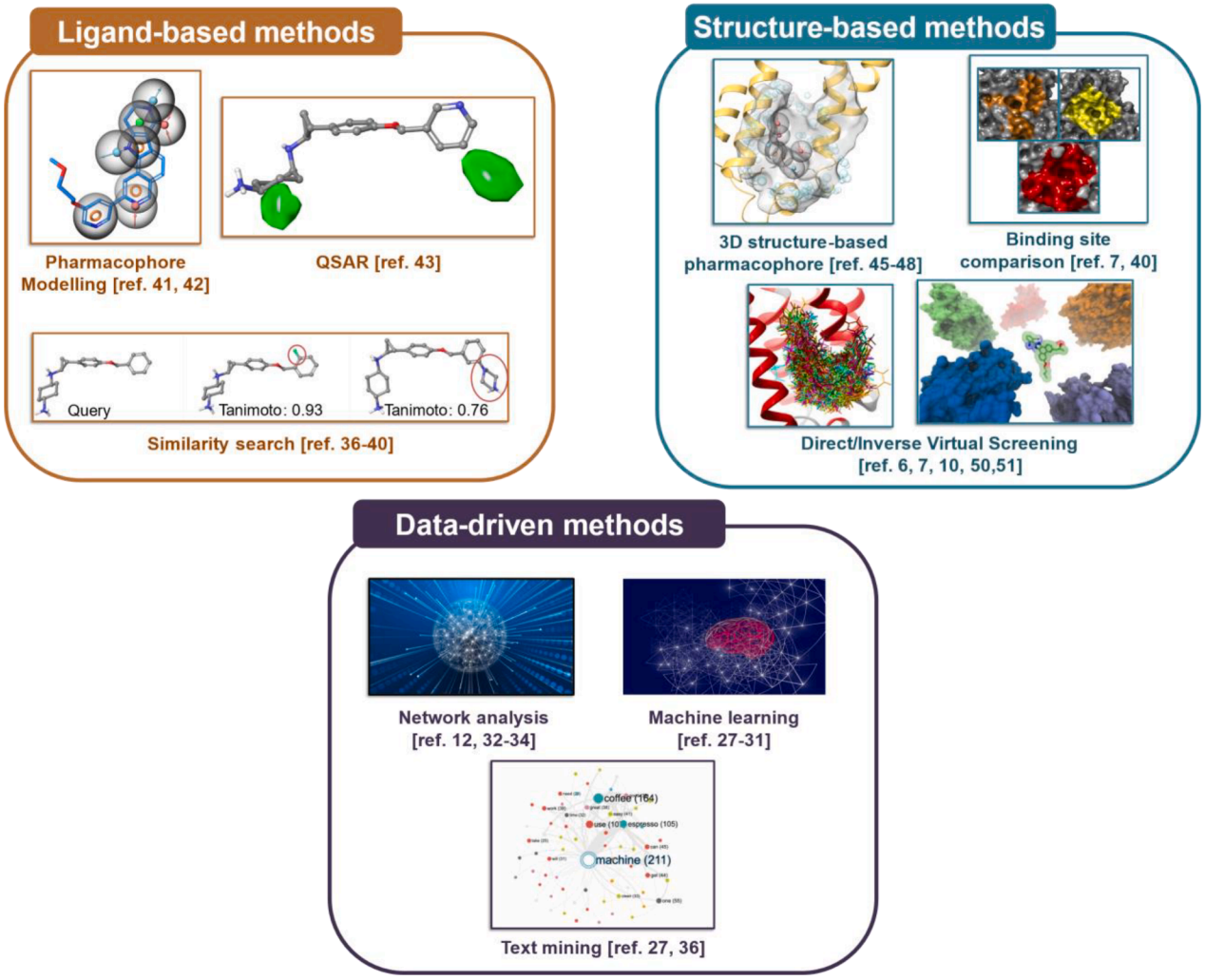

Figure 4. Methods in computational target identification discussed in this digest.

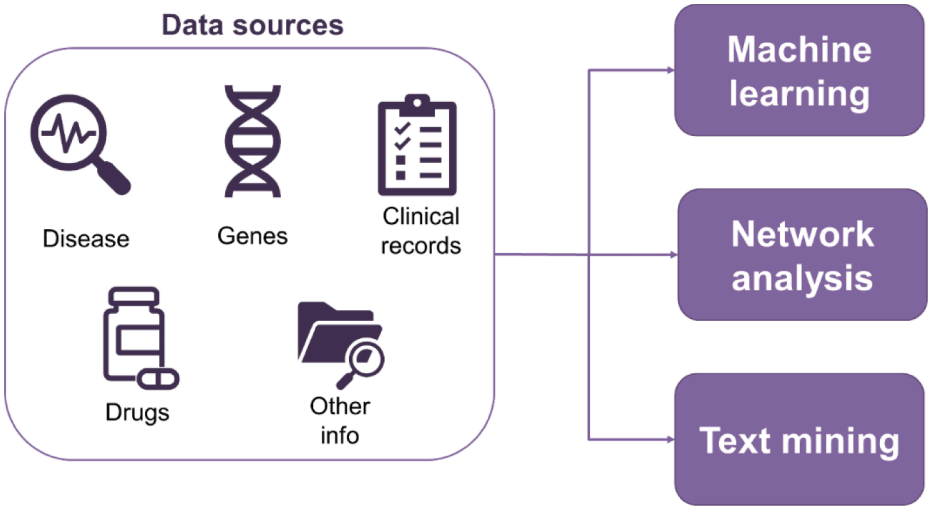

Figure 5. Overview of the data-driven (DD) methods for target identification discussed in this digest and the corresponding sources of information.

screening campaigns.[47,48] A pharmacophore can also be modeled considering the properties of the binding cavity by analyzing the surfaces and creating corresponding features required for the interaction. Intriguingly, comparing pharmacophoric models or testing a single drug against multiple pharmacophores (Reverse Pharmacophore Modeling) can highlight macromolecules that can be targeted by the same compounds.[7] Moreover, a simple comparison of the cavity can be made, assuming that similar binding sites will accommodate the same

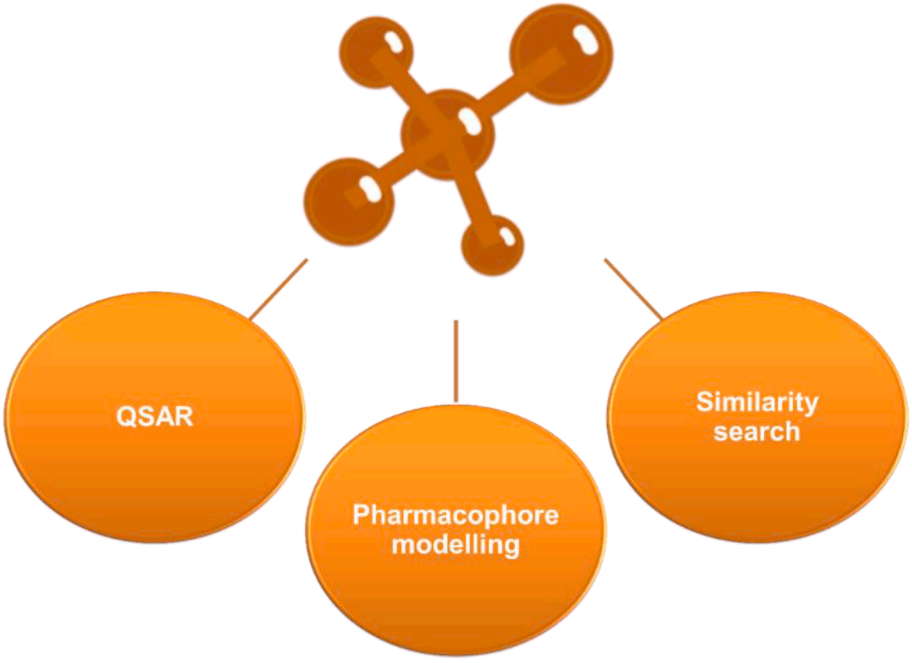

Figure 6. Overview of the ligand-based methods reported in this digest.

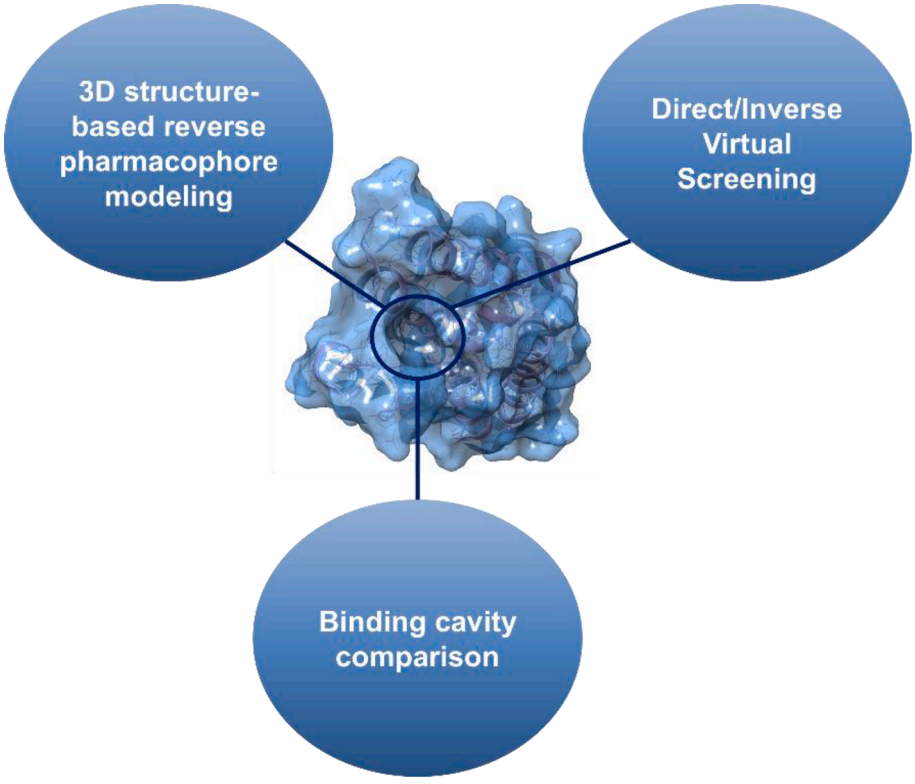

Figure 7. Structure-based approaches in target identification described in this digest.

ligands.[7,40]

In this scenario, the target identification method par excellence is, undoubtedly, molecular docking[6,7,10] that can estimate the direct binding between a ligand and the presumed macromolecule counterpart and, accordingly, belongs to the "structure-based" approaches. Many examples are reported to confirm the importance of these techniques in target identification.[6] In target identification, though, due to the need to test the same compounds on many targets, a methodology called Inverse Virtual Screening (IVS) was designed and will be discussed in-depth in the next section.

Inverse Virtual Screening. Despite the first appearance of the so-called "Inverse Docking" was reported for the first time by Chen *et al.* in 2001,[49] the complete procedure for Inverse Virtual Screening (IVS), comprising both structure-based calculations and the subsequent normalization-based selection step (*vide infra*), was developed in the last decade to obtain rapid and reliable target identification.[50–51] The term

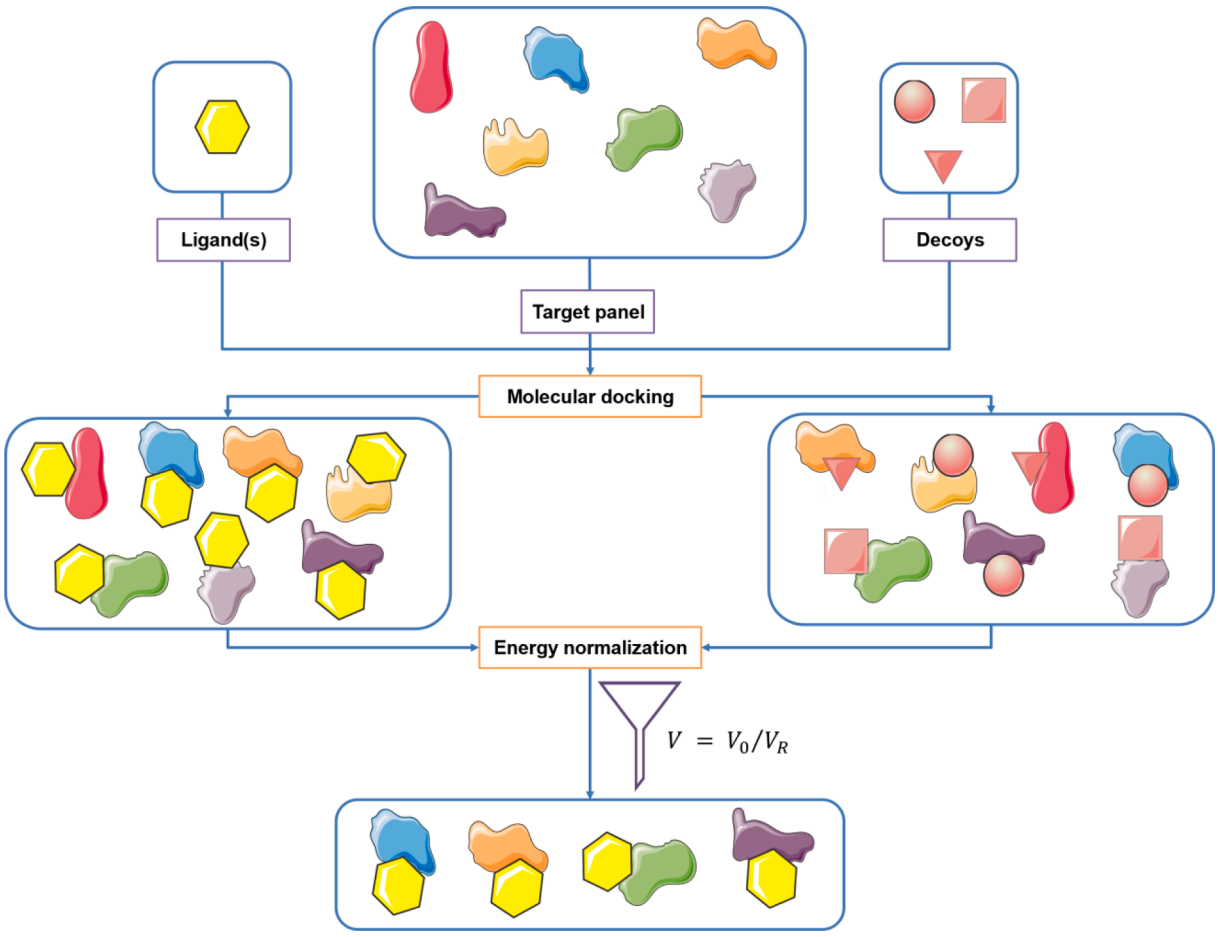

Figure 8. Schematic representation of the Inverse Virtual Screening approach.[14,51,55]

IVS, in fact, designates not only the "Inverse" or "Reverse" docking method (i.e., the docking of a single molecule against multiple targets) but also the following phases for filtering the most promising ligand/target complexes through the definition of a reliable ranking parameter. Such an aspect is not trivial due to false positive/false negative data; accordingly, starting from the pilot study that introduced IVS, a normalization of the docking-related predicted binding affinities was introduced in the IVS pipeline. In more detail, the final purpose of the IVS protocol is to select one or more protein partners for a small set of molecules by carrying out a series of molecular docking experiments, where each member of the library is docked against a wide panel of targets to point out the most probable interacting partners (Figure 8). Many reviews stress the importance of three aspects when performing *in silico*-based target identification experiments: a) the target panel, b) the software used, and c) the ranking method.[4,26,36,52] The type of panel used for the calculations represents the first key aspect of IVS. Many online databases are available that can help in gathering the necessary protein structures: Protein Data Bank,[44] Therapeutic Target Database (TTD),[53] and sc-PDB[54] are the most used. In the early stages of IVS development,[50,51] the screening was carried out on a relatively small number of proteins (~120 items) because of the necessity of manually collecting and preparing the required structures, which resulted in a long and tedious procedure. However, with the increase in the number of crystal structures deposited in the Protein Data Bank, it became clear that considering a single three-dimensional structure for each target would not be sufficient to correctly represent all the possible variations that occur in a protein (e.g., bound/unbound state, multiple isoforms, different resolution, etc.).

This is why an *in silico* procedure that automatically generates the panels was recently developed for the subsequent IVS experiments[55] starting from a customizable list of PDB codes. In detail, the corresponding structures are downloaded, deprived of the unnecessary elements (like solvents or ions), and parametrized for the calculations (https://www.computorgchem.unisa.it/cloe). Once ready, the required grids are generated considering the possible presence of co-crystallized ligand(s). This approach is useful to create pathology- or family-specific sets of targets that can help in elucidating the mechanism of action of the selected molecules. Molecular docking calculations are the core of the whole procedure and, accordingly, the use of robust software is crucial for obtaining consistent results. Many molecular docking software are available, either free or licensed, to sample the ligand inside the binding pocket. Among the most known there are Glide,[56–58] AutoDock Vina,[59] and DOCK,[60] based on algorithms capable of performing an exhaustive conformational search and returning the most reliable poses.[61] The obtained binding energies cannot be taken *per se*, because molecular docking programs still have poor accuracy issues in the sampling and scoring steps and may mislead the selection process. In this respect, docking score values are then normalized by performing the same procedure using a set of "decoy" molecules that share physicochemical properties with the compounds under study but have different chemical structures. After collecting the data, a parameter (called *V*) is calculated to rank the predicted ligand/protein complexes; this mathematical manipulation is shown below:

$$V = V_0/V_R$$

where V_0 is the predicted binding energy of the compound for the target and V_R is the average binding energy computed for the decoys for the same protein. Specifically, *V* values above 1 indicate that the energy

Table 1
Papers selected as examples of the application of structure-based computational methods in target identification and drug discovery. For the IVS experiments the number of targets used is also indicated.

	Method	N° of targets in the panel	Confirming biological assays
Crocetti L., *et al.*, *Struct. Chem.*, **2022**, 33, 769–793[63]	IVS	31	No
Fatima I., *et al.*, *Sci. Rep.*, **2022**, 12, 9260[64]	VS/MD		No
De Vita S., *et al.*, *Plants*, **2022**, 11, 1671[65]	IVS	3,789	No
Zhao C., *et al.*, *Chin. J. Nat. Med.*, **2021**, 19, 454-463[66]	RS	127	Yes
De Vita S., *et al.*, *Biomolecules*, **2021**, 11, 1490[67]	IVS	3,060	Yes
Potenza M., *et al.*, *Bioorg. Chem.*, **2021**, 111, 104897[68]	IVS	312	Yes
Wang F., *et al.*, *Int. J. Mol. Sci.*, **2019**, 20, 115[69]	RS	1,044	No
Ostacolo C., *et al.*, *Eur. J. Med. Chem.*, **2019**, 167, 61-75[70]	IVS	312	Yes
Cilibrizzi A., *et al.*, *J. Enzyme Inhib. Med. Chem.*, **2019**, 34, 44-50[71]	IVS	32	No
Di Micco S., *et al.*, *Bioorg. Med. Chem.*, **2018**, 26, 3935-3957[72]	IVS	308	Yes
Gazzillo E., *et al.*, *Molecules*, **2022**, 27, 3866[73]	VS		Yes
Bharti H., *et al.*, *Sci. Rep.*, **2022**, 12, 918[74]	VS		Yes
Giatti S., *et al.*, *J. Mol. Struct.*, **2022**, 1268, 133690[75]	VS/MD		Yes
Aziz M., *et al.*, *Sci. Rep.*, **2022**, 12, 6404[76]	VS/MD		Yes
Jiang H., *et al.*, *Biochem. Biophys. Res. Commun.*, **2022**, 606, 87-93[77]	RS	17,000	Yes
De Vita S., *et al.*, *RSC Adv.*, **2020**, 10, 40867-40875[62]	VS/IVS	628	No
De Vita S., *et al.*, *J. Chem. Inf. Model.*, **2019**, 59, 4678-4690[55]	IVS	2789	Yes
Bhardwaj P., *et al.*, *Chemosphere*, **2019**, 235, 976-984[52]	RS	226	No
Giordano A., *et al.*, *Eur. J. Med. Chem.*, **2018**, 152, 253-263[78]	IVS	302	Yes

IVS = Inverse Virtual Screening; VS = Virtual Screening; RS = Reverse Screening; MD = Molecular Dynamics.

computed for the considered molecule is higher than the average energy of the decoys and, for this reason, it is possible to argue that the protein–ligand interaction may correspond to a promising binding. After this normalization, the *V* value obtained can be used to highlight the best interactors among the ones contained in the panel.

A new interesting application of IVS regards the introduction of a new parameter called "IVS_{ratio}" to corroborate the results deriving from a starting Virtual Screening campaign. In more detail, the most promising compounds that emerged in the Virtual Screening step, thus focusing on a specific target, are further processed with IVS experiments using a large panel of proteins in which the original target was inserted. Then, the binding affinities calculated for the original ligand/target complexes are divided by the best binding affinity obtained by each compound considering the entire panel:

$$IVS_{ratio} = BA_{ligand}/BA_{max}$$

In this way, with values spanning from 0 to 1, it is possible to estimate the reliability of the ligand/protein complexes arising from the original Virtual Screening. This approach was applied in a drug repurposing campaign (*vide infra*).[62]

The benefits of this type of screening comprise the possibility of testing thousands of proteins in a short time, highlighting unprecedented targets that may also explain side effects or suggest new mechanisms of action. In summary, the possible applications of IVS are: a) target identification, b) drug repurposing, c) side effects prediction, and d) elucidation of the mechanism of action.

Applications of computational methods in target identification. As said before, computational chemistry and structure-based approaches, in particular, have boosted the target prediction capability, reducing the time required for the screening considerably. In this section, we report some examples (Table 1) of the use of these techniques in target identification, with particular emphasis on the application of the IVS pipeline (Figure 9).

The IVS procedure is particularly useful for the analysis of natural products (NPs), which are frequently obtained in small amounts that prevent the implementation of an extensive *in vitro* protocol similar to IVS. More recently, De Vita *et al.*[65,67] used IVS in two different studies. The first aimed to clarify the biological properties of gentiopicroside and loganic acid (**10** and **11**), two secondary metabolites of *Gentiana lutea* L., known for their anti-inflammatory properties. After the extraction and the structural characterization, **10** and **11** were screened against a wide panel of targets involved in cancer and inflammation. Among the top-ranked results, COX-2 was finally confirmed as a binding partner by *in vitro* assays.[67] The second work was based on the elucidation of the anti-inflammatory action of a newly discovered secondary metabolite (2α-hydroxy-$\Delta^{3,7}$-cannabitriol, **12**), cannabidiol (**13**), and cannabidiolic acid (**14**), contained in the *Cannabis sativa* L. vr. *Futura 75*. The methanolic extract components were characterized before carrying out an extensive IVS campaign to highlight the most probable targets that could help explain the biological effects.[65]

NPs are also used often as the starting point to generate derivatives with better pharmacokinetic/pharmacodynamic features; it is the case of the studies reported by Di Micco *et al.*[72] and Potenza *et al.*[68]. In the first paper, magnolol (**15**) and four derivatives were screened against 308 cancer and inflammation targets to point out tankyrase-2 as the principal interactor for compound **16**; the *in silico* predictions were confirmed by surface plasmon resonance experiments and anti-proliferative assays. In the second work, the farnesoid X receptor (FXR) was reported to be the biological partner for steviol (**17**) after a combined IVS/MD evaluation. At first, the secondary metabolite was submitted to an IVS evaluation on a panel of 312 cancer and inflammation targets and, after confirming *in vitro* the FXR/steviol agonist activity, five derivatives of the case-study metabolite were designed. These compounds were first docked directly in the binding site of FXR to verify their interaction capabilities and, then, screened by IVS on the same panel, confirming what was already found with steviol. Moreover, all the docking poses generated were compared to the co-crystallized ligand to highlight similarity. Surprisingly, biological data indicated that these derivatives (**18** and **19**), unlike their parent compound, have antagonist properties.

Additionally, Ostacolo *et al.*[70] and Cilibrizzi *et al.*[71] in 2019, and Crocetti *et al.*[63] in 2022, used IVS to find biological applications for different synthetic compounds. Specifically, Ostacolo *et al.*, during the search for *Varicella zoster* virus inhibitors, came across 1,3,5-trisubstituted indole derivatives with cytotoxic activity on HeLa cells. Prompted by these findings, those compounds were tested on 312 protein targets, involved in cancer and inflammation, and 14 kinases were selected for *in vitro* evaluations. The main outcome that emerged from these studies was the effect of compound **20** on targets belonging to the ERK pathway, which could explain its anticancer activity. Cilibrizzi *et al.*[71] first, and Crocetti *et al.*[63] then, tested the anticancer profile of 43 and 32 heterocyclic hits, respectively by docking them against 31 and 32 key proteins and evaluating their pharmacokinetic properties. In both cases, several proteins, especially kinases, were highlighted as binders for compounds **21**–**25** (Figure 9).

It was already discussed before that drug repurposing is a particular form of target identification. One of the first applications of IVS for drug

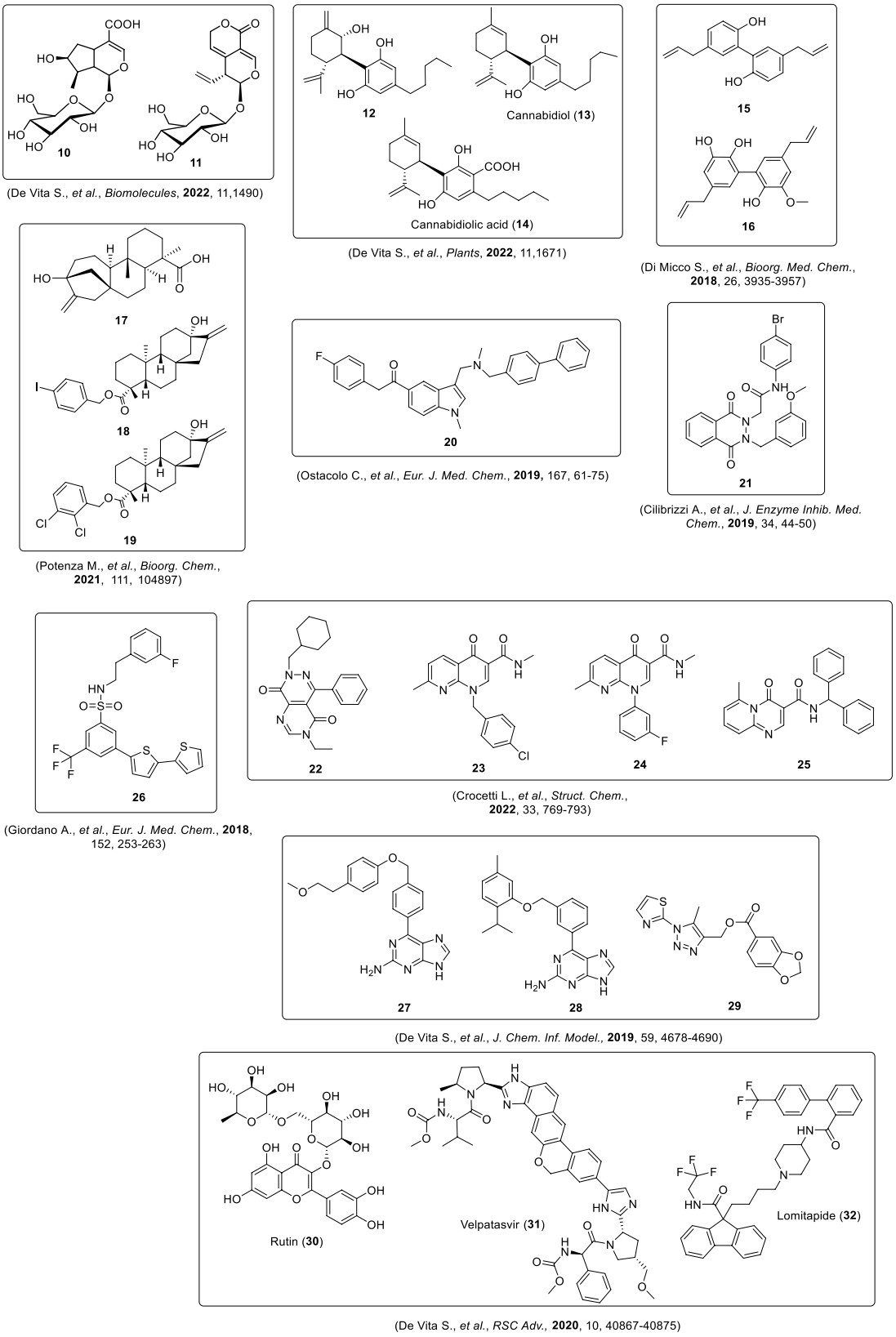

Figure 9. Chemical structures of the repurposed compounds reported as IVS-based examples and listed in Table 1.

repurposing was reported by Giordano *et al.*[78] in 2018. In their research, they tried to find new scopes for molecules originally meant to be Jumonji domain-containing protein D3 (JMJD3) inhibitors, but for which low/no activity was detected *in vitro*. The eighteen compounds were tested on a panel of 302 proteins belonging to cancer-related classes and normalized with 162 decoy molecules. Among the whole panel, Erb-B2 Receptor Tyrosine Kinase 4 (erbB4) emerged as the most promising partner, also due to the high degree of similarity calculated

between the case-study molecules and the co-crystallized inhibitors available in the structures. Compound **26** (Figure 9), in particular, showed a low micromolar IC_{50} in cell proliferation studies. In the same year, De Vita *et al.*[55] introduced a new automatic protocol for the preparation of the required protein panel. To test the robustness of this new approach, an IVS procedure was employed to reposition three molecules from an in-house library (**27**–**29**, Figure 9) using a panel of more than 2000 targets. After the normalization step, tankyrase 2 emerged as promising off-target and the binding was later confirmed by SPR assays. Prompted by the pandemic caused by SARS-CoV-2, De Vita *et al.*[62] tested the entire FDA-approved drug database on three key viral proteins (main protease, papain-like protease, and spike protein). Out of 2748 molecules composing the original library, the best protein/ligand complexes for each target (27 complexes in total) were selected and blindly re-checked through a new parameter called "IVS_{ratio}" (*vide supra*). The MM-GBSA energy was the final filter to select the three most promising target/ligand complexes (**30**–**32**, Figure 9), which were simulated with molecular dynamics, and the interaction suggested was supported by literature data.[79–80]

Concerning the non-IVS approaches, Fatima *et al.*[64] used a combined cross-docking/MD approach to highlight new Rift Valley fever virus (RVFV) natural inhibitors. For this purpose, over 6000 NPs were screened against the main viral protein targets and the five top-ranked (calyxin C, calyxin D, calyxin J, gericudranins A, and blepharocalyxin C) were submitted to MD simulations with MM-GBSA free energy evaluation. A particular case of Reverse Screening studies involving molecules derived from natural sources is represented by Wang *et al.* [69]. Dityrosine is generated from oxidative processes and is detected in high levels in several pathological conditions. Here, *cis* and *trans* dityrosine (CDT and TDT respectively), the dimeric form of tyrosine, were screened against a panel of 1044 targets involved in different pathologies to highlight the putative interactions with biological macromolecules. Three complexes (CDT/Tubulin, TDT/Tubulin, and TDT/thyroid hormone receptor beta-1) were refined with molecular dynamics (MD) to confirm the interaction.

Recently, Zhao *et al.*[66] used a Reverse Docking approach to find putative partners for bufotenine, a compound isolated from *Bufo bufo gargarizans* Cantor and *Bufo melanostictus* Schneider, and its ammonium salt. Out of 127 targets, acetylcholinesterase (AChE) and α4β2 nicotinic acetylcholine receptor (nAChR) were confirmed by *in vitro* and *in vivo* tests.

Another example is provided by Bhardwaj *et al.*[52] in which they repositioned the antimicrobial compound triclosan. Starting from the effects of this compound on the endocrine system, they performed a Reverse Screening analysis on 226 targets (collected by text-mining) that could explain the macroscopic biological outcome. Jiang *et al.*[77] have recently applied an evaluation to confirm the inhibitory activity of montelukast, a leukotriene receptor inhibitor, against flaviviruses proteases NS2B-NS3. The action of this drug on the proteases of Zika and Dengue viruses was already known, but the mechanism of action was unclear. With a combined RS/MD protocol, they were able to define NS2B-NS3 as the main interactor for this compound.

An interesting application of a non-IVS-based drug repurposing campaign combining cross-docking and pharmacophore evaluation was recently published by Gazzillo *et al.* [73]. Starting from 190 molecules with no activity on the target they were designed for, a repositioning was carried out leading to the identification of 2 new soluble epoxide hydrolase inhibitors. Initially, the library was filtered using a 3D pharmacophoric hypothesis to eliminate the compounds that did not contain the required features. The 89 molecules that survived the filter were docked against the enzyme crystal structure, and the resulting poses were further compared to the pharmacophore without the conformational search. In this way, the spatial orientation is kept, and the superposition with the 3D pharmacophore can indicate whether the compound not only possesses the structural features required for the interaction but also if those features are correctly oriented. In the end, two compounds were selected and submitted to *in vitro* tests that corroborated the above predictions.

Final considerations and future perspectives

In silico techniques provide a fast and efficient screening method that helps to direct the following stages of drug research, avoiding part of the downstream inconveniences. Conventionally, target identification can be performed by exploiting target information (structure-based methods), ligand structures (ligand-based methods), or available data (data-driven methods). In this digest, we discussed the applications of these computational chemistry methods, with particular emphasis on structure-based approaches and IVS, to identify and select new targets. In this scenario, the development of new methodologies dramatically improves the quality of the results and reduces the number of failures that can occur in the later stages. In this respect, machine learning and artificial intelligence represent the direct evolution of drug repurposing and, in a more general meaning, target identification. In general, to ensure the optimal outcomes of a target identification and/or drug repurposing study, some precautions should be taken. With particular regard to IVS, the panel of targets should be wide enough to include multiple states of the same protein to provide adequate variability in the protein pool. This is important to consider all the conformation that a protein can take and find the one that accommodates the ligand in the best way possible. Then, a reliable molecular docking software should be used to perform the necessary calculations to correctly sample and score the binding modes and, therefore, return consistent results and, eventually, select a sufficient number ($\geq$10) of decoys to normalize the results, preferably with scaffolds that are different from each other, to highlight the selectivity towards the chemical structure of the ligand.

In this respect, the intermingling between *in silico* and "wet" procedures is becoming the new forefront technology in target identification and drug repurposing. In particular, techniques like the Drug Affinity Responsive Target Stability (DARTS)[81] (https://computorgchemunisa.org/target-identification/) may represent promising partners to be paired with computational methods like IVS in order to quickly find new treatments, either from natural or synthetic sources, for urging pathologies like cancer. Another key aspect is the breakthrough provided by artificial intelligence and the machine learning techniques that, due to the impressive amount of data available nowadays, are capable of providing reliable models for target prediction.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

The research leading to these results has received funding from AIRC under IG 2018 - ID. 21397 project – P.I. Bifulco Giuseppe and MFAG 2017 - ID. 20160 project – P.I. Lauro Gianluigi.

References

1 Wouters OJ, McKee M, Luyten J. Estimated research and development investment needed to bring a new medicine to market, 2009–2018. *J Am Med Assoc*. 2020;323:844–853.
2 Hua Y, Dai X, Xu Y, et al. Drug repositioning: Progress and challenges in drug discovery for various diseases. *Eur J Med Chem*. 2022;234, 114239.
3 Sonaye HV, Sheikh RY, Doifode CA. Drug repurposing: Iron in the fire for older drugs. *Biomed Pharmacother*. 2021;141, 111638.

4 Agamah FE, Mazandu GK, Hassan R, et al. Computational/in silico methods in drug target and lead prediction. *Brief Bioinform*. 2020;21:1663–1675.
5 Jenkinson S, Schmidt F, Rosenbrier RL, Delaunois A, Valentin J-P. A practical guide to secondary pharmacology in drug discovery. *J Pharmacol Toxicol Methods*. 2020; 105, 106869.
6 Pinzi L, Rastelli G. Molecular docking: Shifting paradigms in drug discovery. *Int J Mol Sci*. 2019;20:4331.
7 Sydow D, Burggraaff L, Szengel A, et al. Advances and challenges in computational target prediction. *J Chem Inf Model*. 2019;59:1728–1742.
8 Kabir A, Muth A. Polypharmacology: The science of multi-targeting molecules. *Pharmacol Res*. 2022;176, 106055.
9 Zhang W, Pei J, Lai L. Computational multitarget drug design. *J Chem Inf Model*. 2017;57:403–412.
10 Pushpakom S, Iorio F, Eyers PA, et al. Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov*. 2019;18:41–58.
11 Jarada TN, Rokne JG, Alhajj R. A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions. *J Cheminform*. 2020;12:46.
12 Freeman TB, Lum P. Computational approaches to drug target identification. In: Loging WT, ed. *Bioinformatics and Computational Biology in Drug Discovery and Development*. Cambridge: Cambridge University Press; 2016:17–46.
13 Sabe VT, Ntombela T, Jhamba LA, et al. Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review. *Eur J Med Chem*. 2021;224, 113705.
14 Chini MG, Lauro G, Bifulco G. Addressing the target identification and accelerating the repositioning of anti-inflammatory/anti-cancer organic compounds by computational approaches. *Eur J Org Chem*. 2021;2021:2966–2981.
15 Kato Y, Nishiyama K, Nishimura A, et al. Drug repurposing for the treatment of COVID-19. *J Pharmacol Sci*. 2022;149:108–114.
16 Sharma PP, Bansal M, Sethi A, et al. Computational methods directed towards drug repurposing for COVID-19: advantages and limitations. *RSC Adv*. 2021;11: 36181–36198.
17 White MA, Lin W, Cheng X. Discovery of COVID-19 Inhibitors Targeting the SARS-CoV-2 Nsp13 Helicase. *J Phys Chem Lett*. 2020;11:9144–9151.
18 Romeo I, Ambrosio FA, Costa G, Corona A, Alkhatib M, Salpini R, Lemme S, Vergni D, Svicher V, Santoro MM, Tramontano E, Ceccherini-Silberstein F, Artese A, Alcaro S. Targeting SARS-CoV-2 nsp13 Helicase and Assessment of Druggability Pockets: Identification of Two Potent Inhibitors by a Multi-Site In Silico Drug Repurposing Approach. *Molecules*. 2022;27.
19 Ribaudo G, Yun X, Ongaro A, et al. Combining Computational and Experimental Evidence on the Activity of Antimalarial Drugs on Papain-Like Protease of SARS-CoV-2: A Repurposing Study. *Chem Biol Drug Des*. 2022. in press.
20 Fuzo CA, Martins RB, Fraga-Silva TFC, et al. Celastrol: A lead compound that inhibits SARS-CoV-2 replication, the activity of viral and human cysteine proteases, and virus-induced IL-6 secretion. *Drug Dev Res*. 2022;83:1623–1640.
21 Aronskyy I, Masoudi-Sobhanzadeh Y, Cappuccio A, Zaslavsky E. Advances in the computational landscape for repurposed drugs against COVID-19. *Drug Discov Today*. 2021;26:2800–2815.
22 Talevi A. Computer-aided drug design: An overview. In: Gore M, Jagtap UB, eds. *Computational Drug Discovery and Design*. Springer, New York: New York, NY; 2018: 1–19.
23 Shaker B, Ahmad S, Lee J, Jung C, Na D. *In silico* methods and tools for drug discovery. *Comput Biol Med*. 2021;137, 104851.
24 Galati S, Di Stefano M, Martinelli E, Poli G, Tuccinardi T. Recent advances in *in silico* target fishing. *Molecules*. 2021;26:5124.
25 Frye L, Bhat S, Akinsanya K, Abel R. From computer-aided drug discovery to computer-driven drug discovery. *Drug Discov Today Technol*. 2021;39:111–117.
26 Xu X, Huang M, Zou X. Docking-based inverse virtual screening: methods, applications, and challenges. *Biophys Rep*. 2018;4:1–16.
27 Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol Divers*. 2021;25:1315–1360.
28 Tanoli Z, Vähä-Koskela M, Aittokallio T. Artificial intelligence, machine learning, and drug repurposing in cancer. *Expert Opin Drug Discov*. 2021;16:977–989.
29 Kim H, Kim E, Lee I, Bae B, Park M, Nam H. Artificial intelligence in drug discovery: A comprehensive review of data-driven and machine learning approaches. *Biotechnol Bioprocess Eng*. 2020;25:895–930.
30 Song T, Wang G, Ding M, Rodriguez-Paton A, Wang X, Wang S. Network-Based approaches for drug repositioning. *Mol Inform*. 2022;41:2100200.
31 Lo Y-C, Rensi SE, Torng W, Altman RB. Machine learning in chemoinformatics and drug discovery. *Drug Discov Today*. 2018;23:1538–1546.
32 Katsila T, Spyroulias GA, Patrinos GP, Matsoukas M-T. Computational approaches in target identification and drug discovery. *Comput Struct Biotechnol J*. 2016;14: 177–184.
33 Wu Z, Li W, Liu G, Tang Y. Network-based methods for prediction of drug-target interactions. *Front Pharmacol*. 2018;9:1134.
34 Pan X, Lin X, Cao D, et al. Deep learning for drug repurposing: Methods, databases, and applications. *Wiley Interdiscip Rev: Comput Mol Sci*. 2022;12:e1597.
35 Chen R, Liu X, Jin S, Lin J, Liu J. Machine learning for drug-target interaction prediction. *Molecules*. 2018;23:2208.
36 Yang S-Q, Ye Q, Ding J-J, et al. Current advances in ligand-based target prediction. *Wiley Interdiscip Rev: Comput Mol Sci*. 2021;11:e1504.
37 Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46:D1074–D1082.
38 Kim S, Thiessen PA, Bolton EE, et al. PubChem substance and compound databases. *Nucleic Acids Res*. 2016;44:D1202–D1213.
39 Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res*. 2012;40:D1100–D1107.
40 Byrne R, Schneider G. In silico target prediction for small molecules. In: Ziegler S, Waldmann H, eds. *Systems Chemical Biology: Methods and Protocols*. Springer, New York: New York, NY; 2019:273–309.
41 Wang X, Pan C, Gong J, Liu X, Li H. Enhancing the enrichment of pharmacophore-based target prediction for the polypharmacological profiles of drugs. *J Chem Inf Model*. 2016;56:1175–1183.
42 Steindl TM, Schuster D, Laggner C, Langer T. Parallel screening: A novel concept in pharmacophore modeling and virtual screening. *J Chem Inf Model*. 2006;46: 2146–2157.
43 Rudrapal M, Chetia D. Virtual screening, molecular docking and QSAR studies in drug discovery and development programme. *J Drug Deliv Ther*. 2020;10:225–233.
44 Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28:235–242.
45 Pierri M, Gazzillo E, Chini MG, et al. Introducing structure-based three-dimensional pharmacophore models for accelerating the discovery of selective BRD9 binders. *Bioorg Chem*. 2022;118, 105480.
46 Meslamani J, Li J, Sutter J, Stevens A, Bertrand H-O, Rognan D. Protein–ligand-based pharmacophores: generation and utility assessment in computational ligand profiling. *J Chem Inf Model*. 2012;52:943–955.
47 Moumbock AFA, Li J, Tran HTT, et al. ePharmaLib: A versatile library of e-pharmacophores to sddress small-molecule (poly-)pharmacology. *J Chem Inf Model*. 2021;61:3659–3666.
48 Wang X, Shen Y, Wang S, et al. PharmMapper 2017 update: a web server for potential drug target identification with a comprehensive target pharmacophore database. *Nucleic Acids Res*. 2017;45:W356–W360.
49 Chen YZ, Zhi DG. Ligand–protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins: Struct Funct Bioinform*. 2001;43:217–226.
50 Lauro G, Masullo M, Piacente S, Riccio R, Bifulco G. Inverse Virtual Screening allows the discovery of the biological activity of natural compounds. *Biorg Med Chem*. 2012; 20:3596–3602.
51 Lauro G, Romano A, Riccio R, Bifulco G. Inverse Virtual Screening of antitumor targets: pilot study on a small database of natural bioactive compounds. *J Nat Prod*. 2011;74:1401–1407.
52 Bhardwaj P, Biswas GP, Bhunia B. Docking-based inverse virtual screening strategy for identification of novel protein targets for triclosan. *Chemosphere*. 2019;235: 976–984.
53 Chen X, Ji ZL, Chen YZ. TTD: Therapeutic Target Database. *Nucleic Acids Res*. 2002; 30:412–415.
54 Kellenberger E, Muller P, Schalon C, Bret G, Foata N, Rognan D. sc-PDB: An annotated database of druggable binding sites from the Protein Data Bank. *J Chem Inf Model*. 2006;46:717–727.
55 De Vita S, Lauro G, Ruggiero D, Terracciano S, Riccio R, Bifulco G. Protein preparation automatic protocol for High-Throughput Inverse Virtual Screening: Accelerating the target identification by computational methods. *J Chem Inf Model*. 2019;59:4678–4690.
56 Friesner RA, Murphy RB, Repasky MP, et al. Extra Precision Glide: Docking and scoring injcorporating a model of hydrophobic enclosure for protein−ligand complexes. *J Med Chem*. 2006;49:6177–6196.
57 Halgren TA, Murphy RB, Friesner RA, et al. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem*. 2004;47:1750–1759.
58 Friesner RA, Banks JL, Murphy RB, et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of focking accuracy. *J Med Chem*. 2004;47:1739–1749.
59 Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010;31:455–461.
60 Allen WJ, Balius TE, Mukherjee S, et al. DOCK 6: Impact of new features and current docking performance. *J Comput Chem*. 2015;36:1132–1156.
61 Su M, Yang Q, Du Y, et al. Comparative assessment of scoring functions: The CASF-2016 update. *J Chem Inf Model*. 2019;59:895–913.
62 De Vita S, Chini MG, Lauro G, Bifulco G. Accelerating the repurposing of FDA-approved drugs against coronavirus disease-19 (COVID-19). *RSC Adv*. 2020;10: 40867–40875.
63 Crocetti L, Floresta G, Nazir S, et al. Synthesis and inverse virtual screening of new bi-cyclic structures towards cancer-relevant cellular targets. *Struct Chem*. 2022;33: 769–793.
64 Fatima I, Ahmad S, Alamri MA, Mirza MU, Tahir ul Qamar M, Rehman A, Shahid F, Alatawi EA, Alkhayl FFA, Al-Megrin WA, Almatroudi A. Discovery of Rift Valley fever virus natural pan-inhibitors by targeting its multiple key proteins through computational approaches. *Sci Rep*. 2022;12:9260.
65 De Vita S, Finamore C, Chini MG, et al. Phytochemical analysis of the methanolic extract and essential oil from leaves of industrial hemp Futura 75 Cultivar: Isolation of a new cannabinoid derivative and biological profile using computational approaches. *Plants*. 2022;11:1671.
66 Zhao C, Chen M, Sun S-L, et al. Bufotenine and its derivatives: synthesis, analgesic effects identification and computational target prediction. *Chin J Nat Med*. 2021;19: 454–463.
67 De Vita S, Chini MG, Saviano G, et al. Biological profile of two *Gentiana lutea* L. metabolites using computational approaches and *in vitro* tests. *Biomolecules*. 2021;11: 1490.

68 Potenza M, Cavalluzzi MM, Milani G, et al. Inverse Virtual Screening for the rapid re-evaluation of the presumed biological safe profile of natural products. The case of steviol from Stevia rebaudiana glycosides on farnesoid X receptor (FXR). *Bioorg Chem*. 2021;111:104897.

69 Wang F, Yang W, Hu X. Discovery of high affinity receptors for dityrosine through Inverse Virtual Screening and docking and molecular dynamics. *Int J Mol Sci*. 2019; 20:115.

70 Ostacolo C, Di Sarno V, Lauro G, et al. Identification of an indol-based multi-target kinase inhibitor through phenotype screening and target fishing using inverse virtual screening approach. *Eur J Med Chem*. 2019;167:61–75.

71 Cilibrizzi A, Floresta G, Abbate V, Giovannoni MP. iVS analysis to evaluate the impact of scaffold diversity in the binding to cellular targets relevant in cancer. *J Enzyme Inhib Med Chem*. 2019;34:44–50.

72 Di Micco S, Pulvirenti L, Bruno I, et al. Identification by Inverse Virtual Screening of magnolol-based scaffold as new tankyrase-2 inhibitors. *Biorg Med Chem*. 2018;26: 3953–3957.

73 Gazzillo E, Terracciano S, Ruggiero D, et al. Repositioning of quinazolinedione-based compounds on soluble epoxide hydrolase (sEH) through 3D structure-based pharmacophore model-driven investigation. *Molecules*. 2022;27:3866.

74 Bharti H, Singal A, Saini M, et al. Repurposing the Pathogen Box compounds for identification of potent anti-malarials against blood stages of *Plasmodium falciparum* with PfUCHL3 inhibitory activity. *Sci Rep*. 2022;12:918.

75 Giatti S, Domizio AD, Diviccaro S, et al. Identification of a novel off-target of paroxetine: Possible role in sexual dysfunction induced by this SSRI antidepressant drug. *J Mol Struct*. 2022;1268, 133690.

76 Aziz M, Ejaz SA, Tamam N, et al. Identification of potent inhibitors of NEK7 protein using a comprehensive computational approach. *Sci Rep*. 2022;12:6404.

77 Jiang H, Zhang Y, Wu Y, et al. Identification of Montelukast as flavivirus NS2B-NS3 protease inhibitor by inverse virtual screening and experimental validation. *Biochem Biophys Res Commun*. 2022;606:87–93.

78 Giordano A, Forte G, Massimo L, Riccio R, Bifulco G, Di Micco S. Discovery of new erbB4 inhibitors: Repositioning an orphan chemical library by inverse virtual screening. *Eur J Med Chem*. 2018;152:253–263.

79 Abd El-Mordy FM, El-Hamouly MM, Ibrahim MT, et al. Inhibition of SARS-CoV-2 main protease by phenolic compounds from Manilkara hexandra (Roxb.) Dubard assisted by metabolite profiling and in silico virtual screening. *RSC Adv*. 2020;10: 32148–32155.

80 Agrawal PK, Agrawal C, Blunden G. Rutin: A Potential Antiviral for Repurposing as a SARS-CoV-2 Main Protease (Mpro) Inhibitor. *Nat Prod Commun.*. 2021:16, 1934578X21991723.

81 Pai MY, Lomenick B, Hwang H, et al. Drug Affinity Responsive Target Stability (DARTS) for small-molecule target identification. In: Hempel JE, Williams CH, Hong CC, eds. *Chemical Biology: Methods and Protocols*. Springer, New York: New York, NY; 2015:287–298.