Current Topics in Medicinal Chemistry, 2016, *16*, 2088-2106

BENTHAM SCIENCE

The Repurposing of Old Drugs or Unsuccessful Lead Compounds by *in Silico* Approaches: New Advances and Perspectives

Annamaria Martorana, Ugo Perricone and Antonino Lauria*

Dipartimento di Scienze e Tecnologie Biologiche Chimiche e Farmaceutiche "STEBICEF" – Università di Palermo, Viale Delle Scienze-ED. 17-90128 Palermo, Italy

Antonino Lauria

Abstract: Have you a compound in your lab, which was not successful against the designed target, or a drug that is no more attractive? The drug repurposing represents the right way to reconsider them. It can be defined as the modern and rationale approach of the traditional methods adopted in drug discovery, based on the knowledge, insight and luck, alias known as serendipity. This repurposing approach can be applied both *in silico* and *in wet*. In this review we report the molecular modeling facilities that can be of huge support in the repurposing of drugs and/or unsuccessful lead compounds. In the last decades, different methods were proposed to help the scientists in drug design and in drug repurposing. The steps strongly depend on the approach applied. It could be a ligand or a structure based method, correlated to the use of specific means. These processes, starting from a compound with potential therapeutic properties and a sizeable number of toxicity passed tests, can successfully speed up the very slow development of a molecule from bench to market. Herein, we discuss the facilities available to date, classifying them by methods and types. We have reported a series of databases, ligand and structure stand-alone software, and of web-based tools, which are free accessible to scientific community. This review does not claim to be exhaustive, but can be of interest to help in drug repurposing through *in silico* methods, as a valuable tool for the medicinal chemistry community.

Keywords: Drug design, Drug repositioning, Drug repurposing, *In silico* approaches, Lead compound, Ligand based, Structure based

1. INTRODUCTION

A crucial point in Medicinal Chemistry is the discovery of promising lead compounds and their optimization, in order to obtain new hits with a good bioavailability and a high affinity for the selected biological target.

In lead compound optimization, the medicinal chemists have at their disposal several approaches more or less intuitive, such as the synthesis of drug analogues, isosteres and bioisosteres, the modification of ring systems or the rational pharmacophore identification by computer-assisted design or by quantitative structure-activity relationships (QSARs).

The need to be in step with the times justifies the frenzied research of more effective molecules based on the critical equilibrium between healthy and economic aspects of pharmaceutical corporate governances.

Currently, the average time, from the lead compound discovery to the approval as new drug, is about 14 years, with a failure rate exceeding 90 percent [1]. This high failure rate involves a huge set of many partially developed clinical molecules.

In this context, the drug repurposing (also called "drug reprofiling", "drug repositioning", "therapeutic switching" or "drug re-tasking") is a emerging strategy to assign new clinical uses to known drugs or to promising molecules halted during the clinical trial processes.

The therapeutic drug repositioning has the potential to provide new drugs more quickly than starting from scratch, by speeding the pace in the finding of new pharmacological therapies.

1.1. Successful Examples of Drug Repurposing

Growing in importance, the drug repurposing allowed pharmaceutical companies to reach a wide number of successes. Four examples of well-known examples of therapeutic switching are reported in Fig. (**1**).

Thalidomide was marketed in West Germany (1957), and used as sedative, sleep-inducing agent. Afterwards, it was employed in the treatment of morning sickness in pregnant women. The drug was expected to be safe, based on the *in vivo* good results in rodents [2]. Tragically, for humans thalidomide caused phocomelia in children born from women taking the drug. For this reason, the molecule was quickly withdrawn from the market. In 1964, thalidomide was repurposed by the physician Jacob Sheskin [3] as therapeutic treatment for *erythema nodosum leprosum*, a deeply painful inflammatory condition with typical red nodules under the skin. In 1993, thalidomide was also tested for the treatment of multiple myeloma, due to its anti-angiogenic effect [4]. In 2006, after seven years from the first report issued in oncol-

*Address correspondence to this author at the Department STEBICEF, University of Palermo, Palermo, Italy; Tel: +39-091-23896818; E-mail: antonino.lauria@unipa.it

1873-5294/16 $58.00+.00

ogy literature, the U.S. Food and Drug Administration (FDA) approved the use of thalidomide in combination with dexamethasone for the clinical therapy of newly diagnosed multiple myeloma patients [5,6]. To date, thalidomide sales amount to about 200 million dollars per year, this molecule, harmful for pregnant women population and highly helpful for a big part of patients population, is a suitable example of successful drug reprofiling. Another example of fruitful repositioned drug is raloxifene (Fig. **1**). This molecule, marketed as Evista, is a selective estrogen receptor modulator (SERM), like tamoxifen. Raloxifene started its life as anticancer agent against breast tumors, after it was repurposed in the treatment of osteoporosis, and finally was approved in the protocol for breast cancer prevention [7-9]. The re-tasking of raloxifene is based on preclinical experimental evidences and is not due to a surprising side effect emerged during the clinical trials.

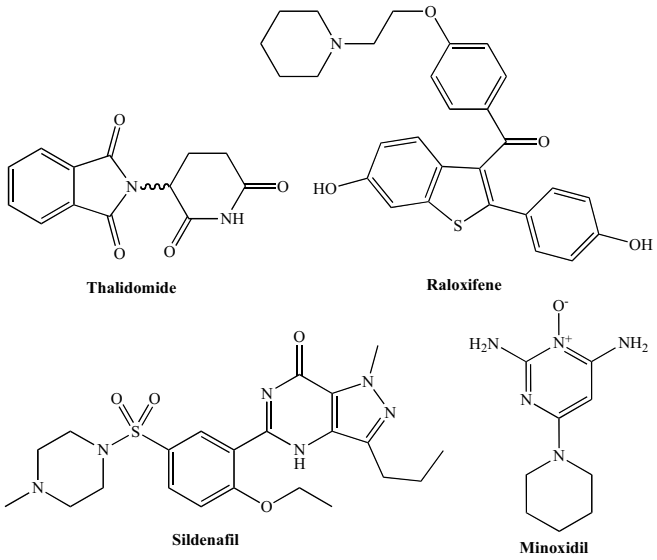

Fig. (1). Known drugs as examples of drug repurposing.

Sildenafil (Viagra) (Fig. **1**), was originally developed for the treatment of hypertension and angina pectoris by Pfizer in 1980s. This molecule is an inhibitor of cyclic guanosine monophosphate (cGMP)-specific phosphodiesterase of type 5 (PDE5), an enzyme involved the angina pectoris processes [10]. In the Phase I clinical trials the patients treated with sildenafil for hypertension and angina pectoris manifested a marketed induction of penile erections. When sildenafil failed in Phase II clinical trials, this peculiar side effect inspired the drug reprofiling of this molecule. In 1998 FDA approved sildenafil for the treatment of erectile dysfunctions, becoming rapidly a blockbuster. The last well-known example, here reported, in drug repurposing is minoxidil (Fig. **1**). Similar to the history of sildenafil, minoxidil was early studied, as peripheral vasodilator, in the treatment of hypertension. During the clinical trials, sodium retention and hirsutism were observed as predominant side effects and, in the late 1980s, minoxidil (Rogaine) was, therefore, cleared by the FDA as remedy for pattern baldness and androgenic alopecia [11,12]. In addition to these examples, several old and idle drugs were proposed in therapeutic switch projects, such as doxycycline early used as antibacterial and then employed in the treatment of periodontitis, bupropion firstly used against depression and now re-tasked on the smoking cessation, and ropinirole, generally active in the hypertension therapy, repositioned as Parkinson's disease remedy.

1.2. Why Drug Repurposing is Intriguing in Medicinal Chemistry?

In the discovery of new therapeutic molecules, the drug re-tasking is still an untapped approach of huge potential. Compared to the traditional drug development, drug repositioning provides few advantages, which fulfil both biological and trade aspects. A therapeutic switch project ensures novel clinical molecules, with pharmacokinetic and pharmacodynamic properties tested, reduced time and cost, and assures new patents.

The molecules, selected as promising candidates for re-tasking developments, should have some common characteristics: a) they could be well-known drugs, often available as generics, or lead compounds suspected of intriguing clinical applications; b) the toxicology profile has to be good, even with chronic dosing. The most of potential candidates come from failed Phase I or II clinical trials. These molecules have already passed a sizeable number of toxicity tests and their safety might be deeply explored; c) a mechanism of action should be supposed with secondary or off-target predictions. New tasking can be suggested by structural analogy with other drugs or by evident side effects arisen from clinical trials tests.

These aspects are crucial for a successful repurposing project in order to save time and cost, necessary to drive a drug from bench to market.

The repositioning of a molecule could be conducted following different approaches. The early repurposed pathways were generally based on serendipitous and systematic observations of the clinical performances of tested molecules. A single drug may interact with more than one target. Therefore, a reprofiling approach could be based on the evidence of relevant side effects and consequently on the identification of new protein targets correlated with new therapeutic indications. This type of reprofiling project presents some advantages, for example the application target- and cell-based of high throughput screening experimentations, to quickly identify or validate a re-tasked drug. On the contrary, the requirement of physical collections of existing drugs creates no negligible limits, entailing time and labor consuming as well as higher research costs.

In recent times, the predictive computational approaches have caught on in the therapeutic switch protocols. Using bioinformatics and public databases tools, *in silico* studies, in repurposing field, allow to predict new correlations between idle drugs and a wide range of biological targets. The advantages of these approaches lie in the possibility to realize a selection of best focused biological targets, with increased speed and reduced research cost.

2. DISCUSSION

The application of computational techniques in drug discovery and development is rapidly gaining in popularity, in implementation and, above all, in reliability. Moreover, as

mentioned before, computational methods are recently playing a crucial role in drug repurposing as well, thanks to the capability to offer means of increased efficacy in the biomedical research field, by improving the drug repositioning tasks. The fast growth in the *in silico* drug discovery has been possible not only thanks to the advances in software and more powerful hardware development, but also to the extensive amount of available biological data. In fact, in the past few decades, there was an huge increasing in the availability of databases containing pharmacological properties of compounds (drugs and not), high-resolution crystals structures of biological targets and small molecules involved in their modulation. Nowadays, software tools have reached a very huge reliability and quality in prediction and simulations, as well as a great versatility for various operative systems, making them of common use. The development of new Central Processing Units (CPUs), capable to scale works in a multithreaded way, has pushed up the speed calculations and therefore the performance in application work on large biomolecular systems (proteins, enzymes, DNAs, etc).

Recently, the new tools take advantage of the speed performance offered by graphics processing units (GPUs) as well. Modern GPUs contain hundreds of arithmetic units that are harnessed to provide tremendous acceleration for numerically intensive scientific applications such as molecular modeling. These increased capabilities and flexibility of GPUs hardware combined with high-level GPU programming languages, such as CUDA and OpenCL, let to scientists an easy access to computational approaches, with notable implications in medicinal chemistry progresses. Many molecular modeling applications are well suited for GPUs, thanks to their extensive computational requirements, and to data-parallel implementations [13,14]. These recent advances in computer-aided drug design can also be applied with huge advantage on drug repurposing.

The challenge in the application of molecular modeling on this field is the search of new indications to test for an old drug or a unsuccessful lead compound [15]. Generally, computational drug repositioning is performed by the design and the validation of workflows that generate hypotheses of new therapeutic applications for candidate drugs [16]. Computational methods have been demonstrated to help researchers to simultaneously generate and evaluate many molecules against several disease pathways, increasing the productivity of traditional drug repositioning projects [16]. These efforts are constantly enhanced by the easy availability of databases in literature [17]. Many computational drug repositioning strategies have been developed over the years, and this review is focused on few of the most used methods. The strategies can broadly be classified into two categories: drug-based and disease-based strategies.

The first starts from the chemical or pharmaceutical perspective trying to find repositioning opportunities for compounds (candidate drugs) considering the traditional ligand based and structure based approaches. This method is preferred when large databases (structures, proteins, enzymes, etc.) are available. In the second strategy, it is necessary to start by examining the symptoms, mechanisms, or pathology of the disease, and this approach is usually adopted when the interest is mainly focused on a specific disease or therapeutic category, or when information about the drug is lacking [15]. In the last years these two approaches have been integrated thanks to new computational strategies based on innovative machine learning processes, as kernel methods, and increasingly articulated artificial networks [18,19]. This synergy between different areas such as chemistry, biology, and pharmacology would be difficult to achieve if not through the aid of well-developed hardware and software. In this context, the capability of drug repurposing prediction becomes an interesting, albeit highly challenging, task leading to continuous improvement efforts in the characterization of possible ligand-receptor interactions [20-22]. Many researchers have adopted the application of these strategies, using chemical, genotypic, and phenotypic similarities among different molecules [23]. In recent years Shoichet and his research group used a statistic-based chemoinformatic approach, in order to obtain a network map able to predict new off-target effects [24].

2.1. Approaching Drug Repurposing Through *In Silico* Methods

How to drive a compound (old drug or unsuccessful lead compound) to drug repurposing? There are two ways that can be approached (Fig. **2**).

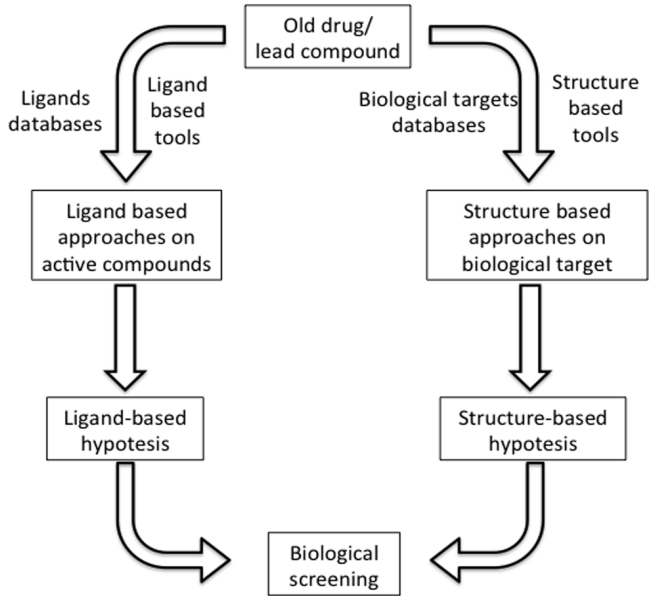

Fig. (2). *In silico* approaches in drug repurposing.

The first is to find a molecule with a specific biological activity by the matching of fingerprint sequences of the repurposed compound with extensive databases. This approach can be defined ligand-based. In particular, the targeted compound is converted in a numeric string that is matched with databases of strings of compounds endowed with specific biological activity. In this case the facilities to use might be ligand databases and ligand-based software (stand-alone or online). The second approach can be defined structure-based. In this case the compound under investigation is tested towards a set of biological targets. Commonly, the compound is submitted, for example, to the so-called reverse docking against a set of biological targets to find the higher affinity. To perform these methods, there are a lot of facilities helpful

for the researchers (databases and software tools). In the next section several of these tools are reported (Table **1**), analyzing their characteristics, uses, and when possible cases of success.

2.2. Ligand and Structure Based Approaches in Drug Repurposing

In the last years there was a trend inversion in computational medicinal chemistry research, because of the developing of drug repurposing strategies. For a long time, the molecular modeling challenge was the prediction of affinity of molecules towards biological targets or mechanisms of action, in order to provide important information useful for lead compound optimization. Recently, computational medicinal chemists are requested to propose ligands for a collection of suitable biological targets. This path can be put in place, only using new approaches like target fishing, in which the ligand is used as bait for different targets. Another important mean for repositioning evaluation is ligand profile approach, where a pharmacological template is pointed out for a specific compound. Nevertheless, classical approaches based on chemical similarity have been demonstrated to be quite good in assessing whether two compounds with analogue structure can be also considered similar in terms of pharmacological effect [25]. In this context, ligand and structure based methods could be applied both in synergy or alone to search for the maximum similarity between compounds or to assess the affinity of biological targets for a ligand. Likewise to traditional molecular modeling approach, structure based methods are used when the 3D structure of the protein is known. The application of the structure based methods is the most useful, because allows the study of the target-ligand interactions. As reported in a recent review [26], the most useful approaches in drug repurposing, could be represented by molecular docking, 3D-pharmacophore modeling and similarity methods on both ligands and binding sites. However, in the lacking of structural information about the biological target, the ligand-based approach can be put in place with good results in the profiling study. Some examples of the mentioned methods will be briefly discussed, underlining the advantages of use for each of them.

2.2.1. Ligand-Based Approaches

In the last decades lots of progresses have been made in the resolution of protein structures and nowadays large

Table 1. ***In silico*** **drug design/repurposing facilities.**

Ligand based means		**Structure based means**	
database	software	***database***	***software***
Binding DB	CANVAS	PDB	AutoDock
ChEMBL	CODESSA	Sc-PDB	#DPS
ChemSpider	CORE HOPING		DOCK
DrugBank	CORINA Symphony		#DvD
G-quadruplex DB	DRAGON		#GES GLIDE
GRAC	#DRUGSERV		GOLD
IUPHAR/BPS	INSTANT JCHEM		MOE
NCI DTP DB	ISIDA Fragmentor		Rosetta
PDBind	JAGUAR		SiteMap
PDSP Ki	LIGAND SCOUT		
PubChem	#LIGSIFT		
SuperTarget	MOE		
TherapeuticTargets DB	PaDEL Descriptor		
ZINC	PHASE		
Data mining tools			
KNIME			
RapidMiner			
Web-based tools			
#BalestraWeb		#K-Map	
CELMINER		#Mantra	
COMPARE (NCI)		#NFFinder	
#CPDB		#PharmDB	
#CRCMeta		ProBis	
DINIES		STICH	
#DT-Web		SuperPred	
		TarFisDock	

#*Drug repurposing dedicated (see paragraph 2.8).*

amount of three dimensional X ray and NMR proteins are available. Nevertheless, ligand-based methods are still of considerable relevance. In fact, ligand based approaches are yet the most developed and used in the repositioning of bioactive compounds. In particular, one of the most used strategy exploits the application of molecular descriptors in sinergy to classical kernel methods, in order to build and validate models that can be applied to predict biological activities for new molecules [27-29]. In these works, there is a perfect cooperation in the use of classical molecular descriptors, such as topological ones or molecular fingerprints and machine learning techniques usually based on statistics. Another recent method is based on the use of the protein-ligand fingerprints [30]. In this case the method allows to translate 3D structural binding information, from protein-ligand complexes, into one-dimensional binary strings, where each byte is referred to a specific interaction. Each fingerprint represents the "structural interaction profile" of the complex that can be used to organize, analyze, and visualize the rich amount of information encoded in ligand-receptor complexes, and also to assist database mining. The most interesting and attractive feature of the ligand-based approaches is represented by the possibility of connecting them to the genomic information about different pathologies [31,32]. All of these methods can be used to perform ligand-based drug repositioning. The common idea of these approaches is the search of molecules structurally related to the ones of a database of interest, containing a set of known active ligands. In the field of drug repositioning, the weakness of ligand-based approaches is the limitation of searching structural similarity between new molecules and only approved compounds.

2.2.2. Structure-Based Approaches

The most common structure based method, used in drug repositioning, is without any doubt the molecular docking (or protein-ligand docking). The aim of this approach is to predict the three-dimensional binding orientation (also called "pose") of a ligand into the protein binding pocket, analyzing energies involved in the binding interactions. Differently to the classical virtual screening approach in which many ligands are tested towards one biological target, this technique, in drug repurposing, has been recently adopted in a reverse manner to find novel targets for a pharmacologically interesting ligand. In this approach a single molecule is screened towards many biological target structures in order to establish the most suitable binding sites. Several studies have been published based on this approach. For example, recently Lee and Kim reported a large-scale reverse docking profile study in which several proteins from yeast and human were tested against 35 drugs used for different pathologies [33]. Another interesting application of reverse docking method is described by Grinter *et al.*, where this approach has been put in place to find new anticancer targets for a quinuclidinone derivative [34]. The reverse docking approach is considered one of the best techniques for target fishing, even though the major problem with this method is represented by the need of high quality of crystal resolution that sometimes is not always available [35].

Another relevant approach to study target-ligand interaction is represented by structure-based pharmacophore methodology. The use of pharmacophore profiling for similarity searching has been for long time used as a ligand based approach based on the 2D pharmacophore sites of molecules. The structure-based pharmacophore application has always been overshadowed by the docking method. Nevertheless, screening using 3D structure-based pharmacophores has been demonstrated to be much faster than docking, which still remains of great importance especially when used in inverse screening applications. In addition, as demonstrated in one of the latest works published by Langer, 3D pharmacophores give the researcher a transparent view of the crucial part characterizing ligand receptor interactions [36]. Several recent studies have demonstrated the efficacy of the structure-based pharmacophore approach [37,38]. Besides this method, protein binding cavity similarity can be considered a possible route in order to find new candidate targets for existing drugs or for new molecules. In this context, computational methods have been revealed of fundamental importance. In fact, it is very difficult to detect binding site similarities just from amino acid sequences, thus 3D computational methods represent essential means for quantifying global or local similarities between protein cavities [39,40]. In this approach key residues for interacting with ligand are compared in the three dimensional space throughout a superposition, then RMSD calculation is performed in order to establish the most similar cavities, and rank them with an algorithm to find the most reliable target issued for a ligand profile.

Clearly, in the age of the "*in silico* big bang", computational drug repurposing is becoming more and more a promising strategy for discovering new uses from existing drugs or unsuccessful lead compounds. The massive growth of available databases containing molecular structures, genomic information, and biological data represents the key of the drug repurposing development, especially for neglected or rare/orphan diseases. In this context it is however necessary to remark the importance of databases in terms of quantity and quality of data contained. It is quite obvious, in fact that the quality of prediction strongly depends on the quality of the data used to building models. For these reasons in the next section we propose a list of available databases, useful in a drug repositioning process, with a short description for everyone.

2.3. Public Domain Databases

In recent years, a number of accessible databases have been created in order to facilitate computational chemists' studies focused on medicinal chemistry. This has increased the rapidity and at the same time the reliability of the molecular modeling tools. One of the key points about availability of data is the improvement of information technology and the development of the new opportunities offered by World Wide Web for dissemination of data. For example, medicinal chemists take advantage of these opportunities through data resources such as the Protein Data Bank (PDB) [41] for 3D biological structures, or BindingDB [42], the first public protein-ligand database aimed at serving the drug discovery community. These resources have substantially grown up in the last years and their use has been also integrated with other sub-databases focused on specific scopes and goals. For example, Pathguide is a Web resource for online databases containing 547 biological pathways and molecular

interaction related resources [43]. Such means are of increasing interest not only for their common use, like finding and downloading structure–activity relationship (SAR) data, they can be adopted for new scopes like the drugs repurposing.

In this section, we report the most common freely accessible databases that are useful in the drugs repurposing processes. We first describe the resources focused on small molecule structures, their binding affinity and biological activity, after the protein structure databases will be described.

2.3.1. Ligand Databases

In this section small molecules databases are reported, providing an overview of the content of each library and how the information is structured, as well as the description about how each resource works.

BindingDB is the most important ligand-protein affinity database, created in the 90s at the University of Maryland [42]. All the data of this repository have been collected with a huge quality of matching information concerning assay conditions such as pH, temperature, and buffer composition. Particularly, BindingDB focuses on quantitative data, such as Ki, Kd, IC50 and EC50 measurements relatives to a well-defined protein target. Currently, these data consist of 2,291 protein-ligand crystal structures, with BindingDB affinity measurements for proteins with 100% sequence identity, and of 5,816 crystal structures matching proteins to 85% sequence identity. More than 60,000 stored data have been manually extracted from literature, including some sets directly submitted by authors. Large amount of these data come from other open databases such as ChEMBL [44] and PubChem [45,46], as well as PDSP Ki [47], a special database containing psychoactive drugs. Very recently (March 2015) Antibodypedia, an open-access database of publicly available antibodies against human protein targets [48-50] and Unichem an EMBL EBI tool that efficiently produces cross-references between chemical structure identifiers from different databases [51,52] has been also integrated. In each case, BindingDB carries out additional processing to ensure that all imported data meet current BindingDB criteria [53].

ChEMBL is a database containing bioactive molecules with drug-like properties [54]. It is sustained by the European Bioinformatics Institute (EBI), of the European Molecular Biology Laboratory (EMBL), based at the Wellcome Trust Genome Campus, Hinxton, UK. All of the pharmaceutical data stored in ChEMBL have been retrieved from the scientific literature. All compounds in the database are matched with biological activities, including protein–ligand affinities and cell-based assay data. Differently to BindingDB, ChEMBL does not include details like buffer composition and experimental conditions for biological assays. About 40% of ChEMBL data are imported from PubChem, and fused with other several large screening data sets. Since January 2015, the ChEMBL database contains about 1,715,135 compound records for 1,463,270 compounds (of which for 1,456,020 hits mol files are available), 13,520,737 activities, 1,148,942 assays, and 10,774 targets [54].

ChemSpider owned by Royal Society of Chemistry [55], represents one of the most important free accessible chemical databases. As reported on the website, the actual version contains 34 million molecules available in about 488 data sources (as compound catalogs or databases). For every molecule, different links, containing information about properties are available. One of the most attractive features of the Web interface is the use of a crowdsourcing approach to expand and improve the data set, by allowing users to enter or correct entries. Compound name, structure, database identifier, and molecular properties generate queries. For every compound searched, the web site gets all the available information such as names, properties, spectra, vendors, data sources, and patents. One of the most interesting aspects inside this database is that it can be updated with user contributions including chemical structure deposition and spectra deposition. This crowdsourcing approach has permitted to develop a very extended online chemistry database [56].

DrugBank [57] is characterized by a consistent public database of approved and experimental drugs even while it contains a very small amount of data if compared with the previous ones. DrugBank combines drug data (i.e. chemical, pharmacological, and pharmaceutical) with drug target information (i.e. sequence, structure, and biochemical pathway). The current database houses a total of 7,759 drug entries including 1,600 FDA-approved small molecule drugs, 160 FDA-approved biotech (protein/peptide) drugs, 89 nutraceuticals, and over 6,000 experimental drugs. All of the drugs data entries are linked to 4,282 non-redundant protein sequences (i.e. drug target/enzyme/transporter/carrier). The data set includes pharmacological and pharmacokinetic data, dosage forms, solubility, drug–drug interactions, metabolism information, target, and pathway data [58].

G-quadruplex binders DB is actually the only and well-organized database of DNA G-quadruplexes ligands. Users have different query options for retrieve and analyze known G-quadruplex binders, as well as use information for the design of new G-quadruplex ligands. It contains molecule structures in different formats for their download and some basic information about molecules such as LogP, formal charges, and donor and acceptor groups of hydrogen bonds. It also provides a DNA G-quadruplex structure database to use for online molecular docking [59,60].

IUPHAR/BPS is the new version of the previous IUPHAR and GRAC database. It collects small molecule activities and affinities, for GPCRs, ion channels, and nuclear receptors. These data are fully integrated with the pharmacological profile of receptors associated with the proposed molecules [61-63]. Moreover they are linked to related information reported in other online resources. To date these databases contain data for about 7,586 small molecules divided into categories such as drugs, natural product, metabolites or antibodies, and 2,726 different proteins, classified on the base of their biochemical role, covering the targets of about half of all current licensed drugs [64].

NCI DTP DB (National Cancer Institute Developmental Therapeutics Program) is a huge repository server containing data available for chemical structure and related grow inhibition and lethal dose data on different cancer cells. It also contains protein expression profile for more than 60 human cancer cell lines collected as RNA microarray data. The main aim of this database is to help researchers to develop molecules and let them to be attractive hits for medicinal

applications in the cancer and AIDS therapy. The repository at first contained about 600,000 molecules tested, but annually has been continuously grown by the insertion of new compounds provided by the users community. The first NCI DTP imprint was compound-oriented, but during these years it has rather become a disease-panel oriented mean acquiring a great reference for different biological assays to be conducted on cells from solid tumors. In the repository biological data are represented as a mean graph that displays growth inhibition in a standard bar graph representation. This graph is a projection of the inhibition values (as bars) deviating on the left or right of the mean depending on whether cells results more or less sensitive compared to the mean of the entire panel studied. Thus, the entire mean graph provide a real fingerprint of a compound activity on the cell lines tested, getting users to be able to compare different compounds efficacy. This is supported by the fact that many of the molecules included in the repository present a known mechanism of action [65].

PDBbind [66] is a database collecting measured affinities for many complexes reported in the Protein Data Bank (PDB). These are protein–small molecule, protein–protein, and nucleic acid–small molecule systems. The PDBbind database is annually updated in order to keep up with the growth of the PDB. The current release contains data from the PDB version officially released on Jan 1st, 2014. This database, classifiable, so far, as the most representative of this type, provides binding affinity data and structural information for a total of 12,995 biomolecular complexes, including protein-ligand (10,656), nucleic acid-ligand (87), protein-nucleic acid (660), and protein-protein complexes (1,592). Binding data, included in the latest version, have increased by 20.6% as compared to previous one, this is also given by a huge increase in the PDB repository. All of these data, extracted from about 30,000 scientific references, have been submitted to double-check in order to ensure the correct matching with the relative complex structures in the Protein Data Bank [67,68]. Moreover, this repository allow researchers to work on a "refined set" and a "core set" information, representing a high-quality data sets of protein-ligand complexes to be used for docking/scoring studies. It is possible to use it for free for academic and commercial use, on a previous acceptance of a license agreement [69].

PDSP Ki, (database of the Psychoactive Drug Screening Program) is provided by North Carolina University. It actually contains more than 50,000 binding measurements, involving more than 7,500 drugs acting on 750 receptors, clustered into neurotransmitter transporters, ion channels, and enzymes [47]. The web interface contains different fields to be used for the query. Users can submit their research based on Ki range values, target or ligands. It is free for academics engaged in mental health research. The web application also provides a huge variety of assays for experimental compound screening, these include for example bioavailability predictions [70].

PubChem database [45,46,71] is provided by the National Institutes of Health within the National Center for Biotechnology Information (NCBI) and mainly contains biological activities of small molecules. At first it was only based on data collected from the high-throughput compound screening programs supported by NIH's Molecular Libraries Roadmap Initiative, but during the years, it has been increased with other sources also allowing users to enrich database with external data. It is organized as linked database within the NCBI's Entrez information retrieval system. The three main branches of the database are PubChem Substance, PubChem Compound, and PubChem BioAssay. PubChem also provides a fast chemical structure similarity search tool. It currently contains more than 30 million entries activity data collected from NIH Molecular Libraries assays and completed with data from about 50,000 journal articles, and other sources, such as pharmaceutical companies and academy research groups [72].

SuperTarget is a repository containing information for about 200,000 compounds interacting with more than 6,200 targets, for a comprehensive data amount of about 330,000 interactions patterns. All the data are divided into Drug-target related pathways (282) and Drug-target ontologies (6,532). In the latest version there are also data for 63 P450 cytochrome metabolic pathways related to all the structure stored into the repository. The structures search could be made for target categories such as function or cellular location [73-75].

Therapeutic Targets Database (TTD) [76] provides a huge amount of information as links related to disease ontologies specifying all the pharmacological pathways involving different protein targets and DNA. All of these information are focused on proven and prospective drug targets and their associated drugs and candidate drugs. It allows users to download data extracted from a query based on both text and chemical similarity searches. The current version (2014) covers more than 2,300 targets (of which 388 successful and 461 clinical trial targets), 20,600 drugs, of them 2,003 approved 3,147 clinical trial drugs, 20,000 multitarget agents against almost 400 target-pairs and the activity data of 1,400 agents against 300 cell lines. Indeed, information of almost 1,800 biomarkers for 300 disease conditions and 200 drug scaffolds for 700 drugs has been added in the latter version [77].

ZINC, is a free database containing over 35 million molecules created at the University of California, San Francisco. All the compounds have been stored in ready-to-dock 3D format in order to facilitate users in virtual structure-based screening [78]. Compounds are organized into various subsets according to different issues, such as, structure, properties, target etc. Users are also allowed to combine different molecules, assembling small subsets according to their own needs. ZINC was firstly created to be used for virtual screening (molecular docking), and this scope still remains its main focus. However, today it is also useful for many other uses including the search of purchasable compounds, the building of small molecules libraries to use for ligand based approaches, find compounds by similarity to a starting one (SAR-by-catalog), for a particular target (via ChEMBL), predict analogues for a particular target (via SEA /ChEMBL or docking) [79].

2.3.2. Biological Target Databases

A good drug repositioning strategy should be always supported by the structural knowledge of protein–ligand

complexes. This aspect has gained considerable value in the recent years, and thanks to the joint efforts of researchers from different biomedical sciences areas, nowadays a huge amount of well-defined biological target data are available for scientific community. Here we report two of the most important biological targets resources available for computer-based drug repositioning.

The Protein Data Bank (PDB) is a repository of three-dimensional (3D) structures of biological macromolecules created in the 70s. Since then, thanks to the efforts of many worldwide scientists, the PDB has always represented a valid connecting platform resource for different databases such as the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB, USA), the Protein Data Bank Europe (PDBe), and the Protein Data Bank Japan (PDBj). All of these resources have been recently merged in a unique one, named worldwide PDB (wwPDB) [80]. To date, it contains more than 108,000 entries, however the number of structures stored in PDB is in constant increase.

In the early 2000s a huge rise in data amount has been observed due to structural genomics initiatives, this trend continued to grow thanks to the effort made by researchers from different scientific areas [81,82]. In the effort made to crystallize and solve proteins, particular attention was paid to the quality of the structures, especially for particular folds and protein families [83]. The format of structures contained in the repository is PDB type, a format similar to the connection table of MOL files [84], but with an incomplete description of the molecular structure. In the CONECT section of every file, in fact, no information is provided about atomic bonds within biological target residues. Indeed, checking for ligands information, there is not any information about bond orders and the connectivity data may be sometimes missing or wrong. Fortunately, recently all the database has been reviewed with a consistent remediation of files errors and validation methods [85-87].

Directly connected to the PDB Database, the **sc-PDB** (screening Protein Data Bank) has increasingly becoming a noteworthy resource. Created in 2002, this database, considerable as a PDB branch, was designed to set up a collection of protein–ligand binding sites, with the purpose of reverse docking applications [88], therefore as an essential resource for the drug repositioning. The first publicly available version of the database was released in 2004 [88]. Since its first release, it contained the atomic coordinates of proteins referred to the "druggable" binding pocket. This has been possible thanks to a special algorithm for detecting proteins in order to find whether binding sites can be classified as druggable or not. Indeed, all atoms in the crystal structure were represented, including hydrogen atoms generally not described in the native PDB crystal structures. In the succeeding versions sc-PDB files also contained atomic coordinates of ligands, covering some lack of information typical of the PDB database. The sc-PDB is annually updated and regularly enriched with new ligand and interactions site information [89], and new functionalities [90]. The latest release, updated to 2013 contains 9,283 entries, 3,678 proteins, and 5,608 ligands. It has also been enriched with a fragment library very useful for bioisosteric substitution assays [91].

2.4. Ligand Based Tools

Ligand based tools could be considered as the basic approach for a drug repurposing workflows. They let the computational chemists to perform some of the principal tasks, in order to start any molecular modeling process, such as molecules drawing, 3D structure optimization and properties calculation. Modern applications go beyond these simple uses, opening to several types of analysis and giving to researchers the possibility of evaluating pharmaceutical issues from different point of view. Herein we report some of the most widespread tools according to scientific community references available.

CANVAS has been designed to be an innovative versatile, performant, and powerful suite of chemoinformatic tools based on a modern technology. It offers solutions such as ultra-fast substructure searching, scaffold decomposition, maximum common substructure search combined with univariate and bivariate statistical analysis applications [92,93]. One of the main tool strengths is represented by the ease of use for every expertise level in molecular modeling.

CODESSA PRO (Comprehensive Descriptors for Structural and Statistical Analysis) is stand-alone software, where mathematical, statistical and computational engines are perfectly integrated to calculate a large variety of molecular descriptors on the basis of the 2D/3D geometrical structure and/or quantum-chemical wave function of chemical compounds. The combination of these approaches allows users the development of quantitative structure-activity/property relationships (QSAR/QSPR). Moreover, (multi) linear and non-linear QSPR models on the chemical and physical properties or biological activity of chemical compounds can be set up. It also provides a set of statistical evaluations such as cluster analysis of the experimental data and molecular descriptors, interpretation of the previous developed models and finally prediction of any property values for chemical compounds with known molecular structure [94].

CORE HOPING, a powerful tool for lead optimization based on ligand-based scaffold exploration. It offers the possibility of search for attachment-based core hopping and shape-based core hopping, which exploits another huge tool, Phase (see below).

CORINA Symphony is an application designed to manage, manipulate and profile molecular data, in order to apply them for drug discovery and lead optimization experiments. It exploits structure representation and descriptors calculation. These descriptors provide a systematic representation of molecules comprising different scaffolds and can directly be used for data analysis methods such as for profiling of data sets, grouping and categorization techniques or predictive model building. Furthermore, a robust database providing chemical data sets along with biological and physicochemical properties can be stored, managed and manipulated. In addition, it uses a chemogenomic approach based on structural features of molecules that may exhibit similar biologically-related pathways. Both representations are presented to the user for selection and are organized in a chemically intuitive hierarchy [95].

DRAGON is a tool primarily designed for the molecular descriptors calculation [96]. The whole descriptors pattern

can be used to evaluate molecular structure-activity or structure-property relationships (QSAR, QSPR), as well as for similarity analysis and high-throughput screening of molecule databases. Actually DRAGON is widely used in scientific studies based on chemoinformatic researches, especially when the focus is based on QSAR approach [97]. It contains a molecular visualization tool and some statistical applications that allow users to perform PCA or other calculations. The latest add-ins features have made DRAGON a very powerful mean for chemometrics and chemoinformatics applications.

INSTANT JCHEM, by Chemaxon, is a tool especially designed for chemists. With this facility users can create, explore, and share chemical and non-chemical data in local and remote databases. The latest version presents a wide and continuous growing amount of functionality, including integration of databases with external data in order to customize models to use for predictions. The software allows users to calculate physicochemical properties, molecular descriptors and topology analysis data. Recently some of the most important properties such as Lipinski drug-likeness or LogP used in drug discovery have been also implemented in the application [98].

ISIDA Fragmentor 2014, designed by the Laboratoire de Chémoinformatique, Chimie de la Matière Complexe (SMS UMR 7140), Université de Strasbourg (France), is a part of the ISIDA project, which stands for "*In Silico* Design and data Analysis". The main aim of the program is quantitative structure-activity modeling (QSAR) and virtual screening and it is composed by tools for the calculation of descriptors, the navigation in chemical space useful for building models and for predictions [99].

JAGUAR is a comprehensive high-performance package for *ab initio* calculations. It allows evaluation for different chemical properties such as NMR, IR, UV-vis, VCD, pKa, partial charges, multipole moments, polarizabilities, molecular orbitals, electron density, electrostatic potential, Fukui functions, Mulliken population, and NBO analysis. Users can exploit all calculations applying them to both gas and solution phases [100].

LIGAND SCOUT is a tool for pharmacophore modeling characterized by an extreme ease of use. Once created, the pharmacophore can be exported in different file formats, so that users are allowed to use it importing in any other software for modeling [101]. Pharmacophores can be created from a set of ligands in case of protein structure absence (models can be clustered in terms of common features), or from a ligand–protein complex, but obviously excluded volumes are only considered in the latter manner. The software allows alignment for either a set of ligand molecules, or a set of pharmacophores. All the models created can be merged in order to build an all-encompassing consensus pharmacophores, containing all features of input models. Alternatively, the previous built models can be fused in a shared pharmacophore, which only extracts the features (including volumes) that are common to all the input models, this will create a common map of ligand binding to the protein pocket. Indeed, pharmacophores can be manually edited if required setting some features as optional or modifying tolerance features [102].

MOE^{TM} (Molecular Operating Environment) could be configured as a suite of applications for the manipulation and the analysis of large collections of chemical and pharmaceutical data. In the ligand based section, users can find molecular descriptors calculation tasks, chemometrics and QSAR tools, fragment based applications, as well as an entire pharmacophore discovery and alignment section. MOE also contains a big database containing molecular data (small molecules, proteins, antibodies, alignments, docking results, etc.) as well as mining tools for compound library design and analysis. One of the most interesting aspects of MOE is that users can import their data into a database and use them for clustering or in the similarity selection and QSAR modeling [103-105].

PaDEL Descriptor is a software able to calculate molecular descriptors and fingerprints. PaDEL currently provides 1,875 descriptors subdivided into 1,444 1D/2D descriptors and 431 3D descriptors. The software also offers the possibility of analyzing molecules with 12 types of fingerprints for a total of 16,092 bits. The calculation engine is based on the Chemistry Development Kit, furthermore it contains additional descriptors and fingerprints such as atom type electrotopological state descriptors, Crippen's logP and MR, extended topochemical atom (ETA) descriptors, McGowan volume, molecular linear free energy relation descriptors, ring counts, count of chemical substructures identified by Laggner, binary fingerprints and count of chemical substructures identified by Klekota and Roth [106]. Another characteristic of PaDEL Descriptor is its possible use as node in the KNIME or RapidMiner environment (see below), enhancing the integration of this tool with others frequently used.

PHASE is a powerful tool for hit generation and lead hopping. It contains a complete package of pharmacophore modeling tools that offers scientists a good level of control at any step of the model creation. The users can use without difficulty the highly configurable process workflow [107,108].

2.5. Structure Based Tools

AutoDock is a tool designed for molecular docking of small molecules, such as substrates or drug candidates, which can interact with the binding pocket of biological targets. AutoDock (4.2), the latest distributions consist of two different software: AutoDock and AutoDock Vina [109]. AutoDock is structured as a merge of two different engines working together. The first part (autogrid) operates a target grid calculation (this allowing a mapping of the active pocket in the protein) and the second (autodock) performs the ligand docking on a set of previously calculated grids describing the target. Some of the frequent uses for Autodock are virtual screening (HTS), protein-protein docking, lead optimization and combinatorial library design. AutoDock Vina, differently to Autodock, does not require a precalculation of the grid maps for the binding pockets. Instead, it calculates instantly the grids for each atom types that are present [110]. The latest version of Autodock and Autodock Vina are completed by a graphical user interface called AutoDockTools (ADT), where the user can easily set up

every feature for the docking, especially establishing rotamers of the ligands to be considered for the docking process.

DOCK is one of the first software for docking small molecules into receptor binding pocket. Differently from other analogues software, the receptor's binding site is characterized analyzing spheres that are subsequently matched with molecules properties. This is performed thanks to geometrical algorithms. The molecule being docked is characterized by ligand centers, which may be its non-hydrogen atoms or volume-filling spheres calculated in Sphgen. The ligand pose is defined according to calculated distances between ligand center and receptor center of the previously described spheres. Sets of ligand centers match sets of receptor centers, if all the internal distances match within a deviation value of tolerance. Once a good orientation has been found, it is evaluated by any of several scoring functions. This software can be used, as the previous ones, to perform protein-ligand docking virtual screenings for both a huge library and/or a small set of molecules to be tested. In the latest version some new functions have been introduced such as receptor flexibility, the full AMBER molecular mechanics scoring function with implicit solvent, conjugate gradient minimization, and molecular dynamics simulation capabilities [111,112].

GLIDE is a complete solution for ligand-receptor docking, offering to users the possibility of choosing between speed based applications or maximal accuracy in calculation as for the extra precision mode. All of these methods are characterized by a high accuracy in binding site mode prediction, reliably finding the correct binding modes for a large set of test cases. This has been also confirmed from some works, where docking pose RMS of native crystalized ligand was tested comparing different docking programs. Glide always obtained the smallest value in a variety of biological targets. The only weakness of this program is given by the force field adopted (OPLS 2005), that is not suitable for studies on nucleic acids receptors [113-115].

GOLD resulted one of the most reliable software for molecular docking from several comparative studies. It revealed a great capability in finding a correct pose for different type of targets [116,117]. One of the most attractive features is the high customizability that gives users the full advantage for the maximization of the docking performance [118]. As for GLIDE, GOLD also offers a wide range of constraints that can be employed to ensure the fulfilling of some key interactions, or bias some docking results towards a known binding motif, allowing to eliminate some erroneous ligand pose conformations [119]. All of the tool features are characterized by an extreme ease of use and a very intuitive graphical interface.

MOE (Molecular Operating Environment), as previously reported is a suite of different molecular modeling tools very useful for medicinal chemistry researchers. It provides different applications useful for the evaluation of binding site features and for better understanding ligand-receptor interactions. The tool is based on a specific α-shapes algorithm that provides a good detection of the possible ligand-protein interaction site [120]. This software is designed also for a multi fragment approach. In fact, it gives the possibility of testing binding site properties by populating the cavity with dummy atoms of chemical fragments, which are subjected to an energy minimization protocol, in order to establish the mapping and preferred locations of specific chemical features in the receptor structure [121]. The resulting group locations are clustered, scored (including solvation effects) and written to a database for subsequent visualization and analysis. Analyzing docking application, one of the most interesting issues is that it is very reliable in terms of correct ligand poses and it contains different force fields available to be set according to the studied target [122].

Rosetta software is a configurable tool suite containing different applications useful for protein analysis and ligand-target inspection. It is free for academic usage, but not for commercial one. In the recent version it has been improved with bioinformatics features, including *de novo* protein design, enzyme design, and structure prediction of biological macromolecules and macromolecular complexes. The first and main use of Rosetta still remains the molecular docking for ligand-protein interactions study. The basic tasks and operations used by Rosetta are represented by the combination of the inner libraries with some algorithm in order to provide flexible molecular modeling protocols. All of these applications can either be used as self-contained units, or they can be chained together to accomplish more complex tasks in a workflow. The software is able to predict, design, and analysis on a diverse set of bio-molecular systems, including proteins, RNA, DNA, peptides, small molecules, and non-canonical or derivatized amino acids [123-125].

SiteMap combines new discovered techniques for a reliable, rapid and easy recognition of the protein binding site. According to users settings, the program identifies the binding pocket of an entire protein creating the correct size of it and moreover all the functionality hot spots of the cavity, considering also the solvent exposure area. The scoring function used by SiteMap to assess ligand ability to correctly bind protein site, concedes to rank ligands in order to take into account just the best ones, dismissing those pharmaceutically irrelevant. The integration with GLIDE is one of the most relevant features. In fact users can build up docking grids according to all of the important outcomes derived from the SiteMap analysis. This application could be fundamental in case of binding site analogies approach for drug repurposing, giving to users the possibility of analyze binding active maps with all of the residues necessary for the correct binding of ligands [126,127].

2.6. Data Mining Tools

The above-described methods allow the matching of chemical compounds towards a whole range of biological targets or specific pathologies. In drug design and repurposing projects, the definition of similarity between molecules is highly dependent on the data analysis system adopted for the combination of some molecular descriptors and similarity measurement, in order to give the best results for a specific target. There is a great amount of data available to researchers useful in drug repositioning. These data include chemical structure similarity, pharmacophore similarity, basic biochemical measurements, cellular assay results and *in vivo* activities. As already discussed, several databases provide an excellent support for the *in silico* applications. But data available have to be treated with suitable tools to make them

reliable. To do so, today there are lots of software able to build data networks and maps. High computational speed and an intuitive representation of the results are, hence, crucial criteria for the development of new methods. Therefore, it is not surprising that data mining and machine-learning techniques widely applied in computer science are convenient for this task. Herein, we report two of the most used tools for network analysis.

KNIME, is a modern data mining integrated platform where users can input different data to analyze, project, and connect them. The analytic platform is an open source tool where different users and software-houses can contribute in a crowdsourcing way, adding their own nodes for the community. Every analysis is conceived in KNIME as a workflow to build with the different pattern useful to manage a process. Indeed, user with friendly graphical tools can perform all of the data analysis. A workbench characterizes the central part of the graphical user interface where combining nodes represent the various step of a workflow. Mostly important features can be clustered in data access, data transformation, initial investigation, powerful predictive analytics, and visualization nodes. KNIME also provides the ability to develop reports. In the node repository hundreds of processing nodes for data I/O, preprocessing and cleaning, modeling, analysis and data mining as well as various interactive views, such as scatter plots, parallel coordinates and others can be easily found thanks to the logical sorting used to collect them. Amongst the various analysis nodes users can find some modules directly taken from the famous data mining tool WEKA, and some plugins based on R script. KNIME contains several network building tools such as Multilayer Perceptron or PNN (Probabilistic Neural Network). Thanks to the integration with different molecular modeling software as Schrodinger suite, Ligand scout, PaDEL descriptors, and many others KNIME results a perfect mean for data mining in pharmaceutical issues [128,129].

RapidMiner is a machine learning software platform providing a very useful environment for data analysis and text mining. Its application in life science as well as training and teaching is very recent and maybe this is the reason why RapidMiner is less known than KNIME. The logical structure is very similar to the KNIME one and the GUI is designed on the idea of the entire data mining process conceived as a highly customizable workflow to build including results visualization, validation, and optimization [130]. Differently to the previous reported tool, here, the single pathways of the workflow are called operators and each of them is performed as single task within the process. The output of each operator forms the input of the next one. RapidMiner contains several R scripts to use for statistics as well as different WEKA learning models. Recently some pharmaceutical applications as PaDEL descriptors have developed operators to be included in the software workflow [131].

2.7. Web-Based Tools

In a context such as drug repurposing the online tools can surely play a key role. The majority of online tools used to asses a repositioning hypothesis base their prediction on 2D and 3D structural similarity of a tested molecule towards those belonging to a database and with known pharmacological and biological properties. Other applications, more complex and more reliable for prediction quality, work instead on complex network algorithms. Herein, we report a few available online tools, where the user can carry out studies of drug repurposing with a good result. Some of these tools provide users an outcome of possible target for a query molecule, others propose putative biochemical pathways or target on the basis of different matched information. A powerful characteristic of these applications is definitely the ability to integrate data from different databases to track as many information as possible about molecular similarity not only of structural but, in a wider view, of ontological type.

NCI-DTP server provides also the NCI COMPARE, an online tool able to match between one-dose and five-dose data of cytotoxicity for compounds showing similar activity profiles. In this sub repository there are molecules previously tested by the DTP program on 60 different cell lines (see paragraph 2.3.1). The aim of the program is to find analogies between compounds showing similar cytotoxicity patterns. To do so a compound with its biological data is used as a seed to find out all the most similar compounds showing similar activity fingerprints. A correlation coefficient is also expressed relating the closeness of the seed to those agents listed. In case of high correlation between two compounds, pharmacological considerations can be extrapolated in terms of possible common biochemical pathways for the compounds, despite eventual chemical dissimilarities. The importance of NCI DTP program has been largely demonstrated through the great ability it has got in providing reliable information about cytotoxicity of drugs candidate, before the *in vivo* testing, revealing very often results in accordance with the latter and avoiding huge expenses [132].

Directly related to NCI DTP there is another important tool, **CELLMINER**. It is a web interface to be used for advanced querying about molecules stored in the NCI database. It is a freely tool for data mining applicable to different pharmacological information data, including molecular activity related to proteomic and genomic cell profiles. Once data has been queried, the analysis outcomes can be downloaded along with comprehensive information on experimental and analytic methods for each data set. Users can also adopt a specific tool capable to extrapolate information about the list of common genes and proteins in two or more different analyzed pattern data sets. In addition to its role as an integrative resource for the NCI-60, the CELLMINER package can be used for the incorporation of pharmacological profiles of new compounds on different cells than those used as pattern of reference by the NCI repository [133,134].

DINIES (Drug-target Interaction Network Inference Engine based on Supervised Analysis) exploits drugs and proteomic-connected data to predict potential interactions between input molecules and putative targets. The approach used by the algorithm engine of this tool is pharmacogenomic, where data such as molecular structure, target interactions, and correlated pathologies are linked. Input data uploaded by users are automatically converted into matrices to be used for similarity search. The prediction of biological targets for an input molecule is performed by network analysis, where chemical structure similarities are related to different biological data such as amino acid sequences and pro-

tein domains analogies. The originality of DINIES lies into the well-integrated machine learning engine able to work on external heterogeneous biological data and its compatibility with the KEGG (Kyoto Encyclopedia of Genes and Genomes) database, a resource that integrates genomic, chemical, and systemic functional information. The DINIES server also allows users to load matrices containing previous calculated similarity matches, providing that the file format is a standard file table format. For the predictive model, user can also load interactions information data, or choosing those present in the KEGG repository [135]. The algorithm implemented in DINIES web tool has been adopted by Takarabe to evaluate interesting predicted interactions between the anti-obesity agent D07627 (cathine) and mu-opioid receptor (hsa:4988) [136].

ProBis is a web server based on the search of structural similarity between various binding sites through a local alignment of proteins stored in the PDB repository. The latest version engine contains an algorithm recently developed by Konc and Janezic that reduce the time of similarity calculation from hours to minutes. The web application can also be used just for consulting the database containing already done similarity pairings for about 30,000 non-redundant proteins. The ProBiS web server has also a complementary tool, Probis ligand, based on the similarity of crystallized ligands within the PDB structures [137].

STITCH is an interactive mean for studying detailed molecular-target interactions. The dataset included contains information taken from experiments, databases, and literature, relatively to molecules-molecules or molecules-proteins relations. It provides information about interactions for more than 2.5 million proteins derived from different organism and a set of 300,000 molecules. All of these data integrate different interaction information derived from metabolic pathways, crystal structures, drug-target relationship, and binding experiments. The outcomes of the queries are presented as network for single drug-target interaction or for big data screening analysis. All of the proposed possible interactions maps can be traced back to the original source in order to deepen some information or expand knowledge about the binding pocket of the target [138]. The STITCH dataset has been employed in 2014 by Liu group to assess the capability of a DIPR algorithm to reposition drugs for hypertension, human immunodeficiency virus, and malaria. Results from the cross-validation protocol have demonstrated that STITCH dataset contains very useful data to be adopted in the repurposing approach [139].

SuperPred is another web server based on the molecule-target interactions information. Here drug-like compound are assigned to possible target, based on chemical and properties similarities with known approved drugs. At first, the server did not contain much drug-target information, but during the years it has been enriched with lots of data and the latest version consists of more than 660,000 interaction data. This big amount of data collected gives a realistic prediction power and reliability of outcomes. Indeed, new features as the 3D similarity or the occurrence of fragments and the concordance of physico-chemical properties are taken into consideration in the similarity search engine for the latest version. All of the previously mentioned data are used by the server to build a statistical model for prediction to be applied to new input molecules. Data and information collected allow building a molecular fingerprint to be compared with the others in the database, in order to give some reasonable biological target predictions. The application also grants the possibility of evaluating the method and lets on the customization of descriptors used for the prediction, through a retrospective analysis of a drug with known mechanism of actions or targets. If the similarity found out from the analysis is sufficiently high and moreover if the model is correctly built, the web engine will propose indication about the therapeutic area of the novel compound, helping researchers in the optimization of new lead compounds [140].

TarFisDock is a web-based tool that uses the target fishing approach. It automates the research for possible protein interactions of a query molecule over a large amount of protein structures. It covers about 15 therapeutic known pathways with a database containing more than 700 protein structures clustered according to these 15 pathways. These proteins can be used for reverse docking assays in order to find new biological targets or promiscuous ones for a given molecule. In contrast to conventional ligand-protein docking, reverse ligand-protein docking used by TarFisDock tests a molecule over different proteins. Once uploaded the query molecule, in mol2 file format, the server starts a docking procedure on all the proteins collected in the repository. The docking engine used by TarFisDock is the DOCK, and its binding energies parameters are used in order to rank the best possible target for new molecule. Therefore, TarFisDock represents a very useful tool for target identification applied to old drugs, promiscuous molecules, and natural product to be discovered for new medical applications. Validation of TarFisDock has been proven by the application on Vitamin E and 4H tamoxifen, with results for possible off targets confirmed by experimental data [141,142].

2.8. Drug Repositioning Dedicated Tools

The above-mentioned tools can be used both in new drugs design and old drugs repositioning. Due to the interest in the repurposing strategies, recently various drug-repurposing dedicated tools were proposed. The advantage in the use of these tools is the availability of data ready-to-use for repurposing, the weakness is the limited amount of data available when compared to generic tools of molecular modeling. Below the most recent tools and databases, focused on drug repurposing, and a few example of their applications are reported.

BalestraWeb

It is a web server designed to let users assessing predictions based on the potential occurrence of interactions between a specific drug–target pair. Users can compare the most likely interaction partners for a given molecule with drugs or targets listed in the DrugBank repository [143]. This mean allows to identify most similar drugs or most similar targets analysing interaction patterns, comparing them, and showing possible common features. The comparison has done by computing the distance between the latent variable (LV) vectors of a set of molecules. The LV similarity between two drugs could be compared to similar target profiles. Therefore drugs with similar target profiles will have high LV similarity and vice versa.

The Outcomes are useful data helping the development of hypotheses about drug repurposing as well as potential side effects for new molecules. Based on this algorithm, Balestra can be used to find out drugs that are most similar to a query or targets showing high interaction patterns similarity.

The tool, strongly based on machine learning system, exploits a probabilistic matrix factorization method where all the variables are trained using the GraphLab [144] collaborative filtering toolkit, in order to validate the model created. Even though the engine works well, the only limitation of this tool could be the database repository, only DrugBank is used as reference.

CPDB

Consensus Path Data Base is an online tool created by Max Plank Institute for Molecular Genetics integrating interaction networks referred to Homo sapiens, yeast, and mouse. The information included, cover different biochemical pathways as well as protein-protein, genetic, metabolic, gene regulatory, and drug-target interactions. Data originate from more than 30 public resources for interactions such as PDB [80], CheMBL [54], DrugBank [58], and many others. Interactions available to be used in CPDB are completely taken from the literature references. One of the very interesting aspects of this tool is that the interaction data are very well integrated in a complementary manner, avoiding any kind of information redundancies. The result is a perfect interaction network containing different types of interactions to be used as ontological data supporting drug repurposing issues based on network approaches [145].

CRCMeta

It is an integrative approach that combines ontology reasoning with network-assisted gene ranking to predict new drug targets. Colorectal cancer (CRC) as a proof-of-concept use case was used to illustrate the approach. Starting from FDA-approved CRC drugs and the relationships among disease, drug, gene, pathway, and SNP in an ontology representing PharmGKB data, 113 potential CRC drug targets were identified. These genes, based on their relationships with CRC disease genes in the context of human protein-protein interaction networks, were further filtered. Thus, among the 113 potential drug targets, 15 were selected as the promising drug targets, including some genes that are supported by previous studies [146].

DPS

Drug Pair Seeker (DPS) is a tool to predict and prioritize pairs of drugs using the Connectivity Map dataset. Users can enter lists of up and down differentially expressed genes from their own experiments to receive a ranked list of drug combinations that would either reverse or aggravate the condition of their cells or tissue. Zhong *et al.* successfully applied this tool to the search of angiotensin-converting enzymes and Histone deacetylase inhibitors as renoprotective agents [147].

DRUGSURV

It is configured as a repository containing data about experimental and approved drugs in oncology extracted from clinical dataset. Based on this architecture, DRUGSURV allows researchers to study potential anticancer effect of new compounds by similarity searching with known drugs. This relation is made by a chemogenomic manner, exploiting patients' survival data and connecting them to the chemical structure of drugs used in therapy for those particular cases. The expression analysis of related genes shows high correlation with the specific pathology condition analyzed. The latest version contains data for about 50 different clinical cancer gene expression patterns, each referred to more than 100 patients. The pharmaceutical dataset stores about 1,700 approved drugs by the FDA and 5,000 experimental drugs presenting dose information, drug delivery system, and ADMET information. The chemogenomic approach adopted by DRUGSURV places it as one the most important tools used in anticancer drug repurposing processes as demonstrated by the application of this protocol by Amelio *et al.* for the repositioning of thioridazine, a well known antipsycothic drug, repurposed in a successful way for anticancer therapy [148].

DT-Web

It is a web-based interface that uses an algorithm to assess network analysis based on Drug-Target interactions. This technique, combined with domain-specific knowledge expressing drugs and targets similarity, can be used in order to find applications for a given drug. The main scope of this web tool is that users, once uploaded their data to submit to the algorithm research, can inspect all the data network produced by the query, this surely facilitates the early stages of drug combinations, repositioning, substitution, or resistance studies. The use of network approach, in fact, is the most helpful to find out drugs that can act simultaneously on multiple targets in a multi-pathway environment [149]. DT algorithm has been used by Jin and Wong, in connection with others, to repurpose simvastatin and ketoconazole drugs for breast cancer treatment [150].

DvD

It is a powerful tool to be integrated with R or Cytoscape. The main scope is the comparison of drug and disease gene expression profiles from public microarray repositories. Uncorrelated profiles can be used to generate hypotheses of drug-repurposing, whereas the correlated profiles may be used to infer drugs side effects [151]. In an application of DvD in drug repurposing, Pacini *et al.* analyzed several disease datasets comparing them with the 1309 compounds in the CMap and considered significant those connections with q-value < 0.05. From these analysis interesting results has been obtained for Ranitidine, a Histamine receptor type-2 (H2) antagonist, to be used on the breast cancer as also demonstrated by *Bolton et al*. [152].

GeneExpressionSignature (GES)

It is an R package developed for the large-scale analysis of gene expression signatures. The package implements two rank-merging algorithms and two similarity-scoring algorithms. The functions of this tool provide a flexible solution for gene expression signature-based studies and hold great potential in biomedical research applications focused also on drug repurposing [153].

K-Map

A user-friendly web-based program that systematically connects a set of query kinases to kinase inhibitors, based on quantitative profiles of the kinase inhibitor activities. Kim *et al.* applied this tool to demonstrate the goodness of prediction in the repurposing of crizotinib to be used on p53 wild-type cancer cells. It led to inhibition of proliferation and enhancement of cell killing in *in vitro* experiments. This supports the finding that the K-Map could reveal new applications for kinase inhibitors [154].

LIGSIFT

This facility is a algorithm tool exploiting Shape-based alignment of small molecules. It reveals its efficacy when applied to ligand based drug repurposing approach focused on bioisosteric replacement. This tool could be employed when potential off-target interactions and cross-reactivity of known drugs are explored. The main scope of this algorithm is to propose a valid alternative to the common Tanimoto Coefficient (TC) used in most existing software for aligning small molecules. The main aspects of this *in silico* tool can be summed up into the size-independent scoring function for evaluating molecular similarity between small molecules and a huge statistical assessment based on alignment significance (authors state that millions of random comparisons has been made for the validation protocol). Moreover, the algorithm created for large-scale applications is well described and has been benchmarked on a standard database of active and decoy molecules for 40 pharmaceutically relevant protein targets, listed in the Directory of Useful Decoys (DUD) in order to evaluate its capability. This computer method seems to be a very precise and useful mean for drug repurposing, especially when a structural comparison is made on new and old molecules or for analysing possible structural similarities between different drug classes [155].

MANTRA

Mode of Action by NeTwoRk Analysis (MANTRA) is a computational tool based on the analysis of the mechanism of action of new drugs throughout the comparison with known and approved candidates in order to use information for drug repositioning. The central engine of this tool is based on the network theory and non-parametric statistics exploiting gene expression data. In order to evaluate possible off-target uses for a drug, users have to submit a query containing a genome list with genes sorted according to their differential expression under treatment with the studied molecule. Starting from the query, the drug analysed is automatically integrated into a huge network of compounds showing all the possible similarities and differences related to mechanism of action. Exploring the drug network created, users are allowed to make novel hypothesis on known and FDA approved drugs, hence to find repositionable drugs [156]. The MANTRA approach has been applied to the discovery of Fasudil as potential inducing autophagy agent by Iorio et al [157].

NFFinder

It is a bioinformatics tool for identifying potential useful drugs to be adopted for orphan diseases. NFFinder engine is based on the research of possible relationship between drugs, disease, and phenotype of interest. The information matching has created exploiting transcriptomic data relating gene expression as typical pattern of a specific disease. Users can also find all the references published on the research domain. Thanks to its architecture, this bioinformatics web-based tool appears very useful for create hypotheses in drug repositioning especially in the context of orphan diseases like Neurofibromatosis (NF) as initially developed but recently it has been improved to be applied to any other disorder. NFFinder tool requires an input of two lists of up- and down-regulated genes or a list of microRNAs related to the disease condition. Then, a selective query against the inner database, which contains curated DataSets from GEO, CMap and DrugMatrix, is performed to compare the input with the available data in the tool repository.

The output result consists of a series of graphics and tables designed to allow users to formulate repositioning hypotheses and identify potential biological relationships between drugs and diseases. Depending on the input genes and the options selected, users will be able to explore a wide experiments data where phenotypes similar or opposite to the input one are showed, this helping to discover multiple drugs with similar effect or with an opposed profile to the one's particular disease. The interface of this tool is very easy to use, nonetheless the results seems to be very accurate and well built. Its usage becomes fundamental especially for orphan disease, where genomic expression comparison with known pathologies results the best way to find out possible drugs considering the lack of experimental data. Its weakness is that user has to possess all the genomic data available for a disease in order to let the analysis reliable. The lack of some data could surely influence the quality of the information. Setoran *et al.* have demonstrated the successful application of this tool in different cases [158].

PharmDB

It is a database that integrates data associated with disease indications, drug development, and associated proteins, and known interactions extracted from various established databases. To explore linkages of known drugs to diseases of interest from within PharmDB, the Shared Neighborhood Scoring (SNS) algorithm was defined, and to facilitate exploration of tripartite (Drug-Protein-Disease) network, a graphical data visualization software program called phExplorer was developed, It allows to browse PharmDB data in an interactive and dynamic manner. This tool has been applied in a successful way on the repositioning of Benzthiazide (TBZT), known as diuretic drug, as potential agent for lung cancer treatment [159].

3. CONCLUSION

Modern medicinal chemistry, in the last decades, has exploited the use of *in silico* tools. These new means were well established thanks to the huge development of software and hardware facilities focused on drug discovery. At the same time, new drug design approaches have emerged, and the drug repurposing, also called “drug reprofiling”, “drug repositioning”, “therapeutic switching” or “drug retasking”, is one of the more interesting.

Table 2. Selected application of web tools to drug repurposing.

Tool Name	Application	Ref.
DINIES	Cathine on mu-opioid receptor	[136]
DVD	Ranitidine as possible drug for breast cancer	[152]
DPS	Discovery of new possible uses for Histone decetylase inhibitors as renoprotective agents	[147]
DRUGSURV	Thioridazine, a well known antipsycothic drug, repurposed as agent for anticancer therapy	[148]
DTWEB	Simvastatin and Ketoconazole as possible agents acting on breast cancer treatment	[150]
KMAP	Crizotinib and Nutlin-3 as antiproliferative agents on p53 wild type cells	[154]
MANTRA	Fasudil, Rho-kinase inhibitor, as authophagy inducing agent	[157]
NFF FINDER	Several possible drugs for orphan diseases like Neurofibromatosis (NF)	[158]
PHARMDB	Benzthiazide, a diuretic drug, as potential anticancer drug	[159]
STITCH	Possible repositioning for different agents on hypertension, human immunodeficiency virus, and malaria disease	[139]
TARFISDOCK	Vitamine E and Tamoxifen on different targets confirmed by experimental data	[141,142]

To date, the medicinal chemists can be supported in the drug repurposing process through various tools that, if well integrated, can give an excellent push in drug discovery. A few examples of integration between biological data and chemometric tools were developed by us in the VLAK [160] and BIOTA protocols [161]. These approaches, focused on the use of molecular descriptors allowed to improve biological activities of unsuccessful molecules with good results [162-164]. These methods have the great advantage that begin, in the search of new drugs, from lead compounds with known biological profile, shorting the long way that lead to drugs.

The description of available means in drug design/repurposing, reported in this review, although not exhaustive, can help the choosing of the right facilities to repurpose compounds.

This opportunity can make idle compounds interesting. All medicinal chemists have in their labs a lot of compounds that through *in silico* (and low-cost) methods could become a new "serendipitous" drug.

The tools reported in this review can give an idea on how the repurposing is a hot topic. Many cases of success are present in literature, although there are large possibilities of improvement. In Table **2** is reported a summary of examples applications of the most reliable tools reported in this review.

The latest advances and probably the future ones in drug repurposing research consist in the use of genomic information, integrated with target pathways involved, target interaction pattern, and chemical structure of drugs. Very recently lots of works has been realized exploiting this approach thanks to the powerful machine learning methods such as network analysis [165].

One interesting aspect is that this kind of approach finds its best application for orphan disease, a very important topic considering that pharmaceutical firms are not interested in investing so much money to find any possible solution.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work was in part supported by an FFR2012 – University of Palermo grant (CUP:B78C12000380001).

REFERENCES

[1] Pammolli, F.; Magazzini, L.; Riccaboni, M. The productivity crisis in pharmaceutical R&D. *Nat. Rev. Drug Discov.,* **2011**, *10*, 428-438.

[2] Stephens, T.D.; Brynner, R. Dark remedy: The impact of thalidomide and its revival as a vital medicine, 1st ed.; Basic Books: New York, **2001**.

[3] Sheskin, J. Thalidomide in lepra reaction. *Int. J. .Dermatol.,* **1975**, *14*(8), 575-576.

[4] D'Amato, R.J.; Loughnan, M.S.; Flynn, E.; Folkman, J. Thalidomide is an inhibitor of angiogenesis. *Proc. Natl. Acad. Sci. USA,* **1994**, *91*, 4082-4085.

[5] Singhal, S.; Mehta, J.; Desikan, R.; Ayers, D.; Roberson, P.; Eddlemon, P.; Munshi, N.; Anaissie, E.; Wilson, C.; Dhodapkar, M.; Zeddis, J.; Barlogie, B. Antitumor activity of thalidomide in refractory multiple myeloma. *N. Engl. J. Med.,* **1999**, *341*, 1565-1571.

[6] Ning, Y.M.; Gulley, J.L.; Arlen, P.M.; Woo, S.; Steinberg, S.M.; Wright, J.J.; Parnes, H.L.; Trepel, J.B.; Lee, M.J.; Kim, Y.S.; Sun, H, Madan RA, Latham L, Jones E, Chen CC, Figg WD, Dahut W.L. Phase II trial of bevacizumab, thalidomide, docetaxel, and prednisone in patients with metastatic castration-resistant prostate cancer. *J. Clin. Oncol.,* **2010**, *28*, 2070-2076.

[7] Jordan, V.C.; Phelps, E.; Lindgren, J.U. Effects of anti-estrogens on bone in castrated and intact female rats. *Breast cancer Res. Treat.,* **1987**, *10*(1), 31-35.

[8] Ashburn, T.T.; Thor, K.B. Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.,* **2004**, *3*(8), 673-683.

[9] FDA Raloxifene hydrochloride. http://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/ucm129243.htm. (Accessed September, **2007**).

[10] Boolell, M.; Allen, M.J.; Ballard, S.A.; Gepi-Attee, S.; Muirhead, G.J.; Naylor, A.M.; Naylor, A.M.; Osterloh, I.H.; Gingell, C. Sildenafil: an orally active type 5 cyclic GMP-specific phosphodiesterase inhibitor for the treatment of penile erectile dysfunction. *Int. J. Impot. Res.,* **1996**, *8*, 47-52.

[11] Joekes, A.M.; Thompson, F.D.; O'Regan, P.F. Clinical use of minoxidil (Loniten). *J. R. Soc. Med.,* **1981**, *74*(4), 278-282.

[12] Olsen, E.A.; Whiting, D.; Bergfeld, W.; Miller, J.; Hordinsky, M.; Wanser, R.; Zhang, P.; Kohut. B. A multicenter, randomized, placebo-controlled, double-blind clinical trial of a novel formulation of 5% minoxidil topical foam versus placebo in the treatment of androgenetic alopecia in men. *J. Am. Acad. Dermatol.,* **2007**, *57*(5), 767-774.

[13] Stone, J.E.; Phillips, J.C.; Freddolino, P.L.; Hardy D.J.; Trabuco, L.G.; Schulten, K. Accelerating molecular modeling applications with graphics processors, *J. Comput. Chem.,* **2007**, *28* (16), 2618–2640.

[14] Stone, J. E.; Hardy, D. J.; Ufimtsev, I. S.; Schulten, K. GPU-accelerated molecular modeling coming of age, *J. Mol. Graph. Model.,* **2010**, *29*(2), 116-125.

[15] Dudley, J.T.; Desphande, T.; Butte, A.J. Exploiting drug–disease relationships for computational drug repositioning, *Brief Bioinform.,* **2011**, *12*(4), 303-311.

[16] Waldman, S.A.; Terzic, A. Systems-Based Discovery Advances Drug Development, *Clin. Pharmacol. Ther*., **2013**, *93*(4), 285–287.

[17] Gaulton, A.; Overington, J.P. Role of open chemical data in aiding drug discovery and design, *Fut. Med. Chem.,* **2010**, *2*, 903–907.

[18] Azencott, C. A.; Ksikes, A.; Swamidass, S.J.; Chen, J. H.; Ralaivola, L.; Baldi, P. One- to four-dimensional kernels for virtual screening and the prediction of physical, chemical, and biological properties, *J. Chem. Inf. Model.,* **2007**, *47*, 965–974.

[19] Cheng, F.; Sutariya, V. Applications of Artificial Neural Network Modeling in Drug Discovery, *Clin. Exp. Pharmacol.,* **2012**, *2*(3), 113-116.

[20] Siegel, M.G.; Vieth, M. Drugs in other drugs: a new look at drugs as fragments, *Drug Discov. Today,* **2007**, *2*, 71–79.

[21] Millera, J.R.; Dunhama, S.; Mochalkina, I.; Banotaia, C.; Bowmana, M; Buista, S; Dunklea, B; Hannaa, D; Harwoodb, H.J.; Hubanda, M.D; Karnovskya, A; Kuhna, M.; Limberakisa, C; Liua, J.Y; Mehrensa, S.; Muellera, T.S; Narasimhana, L.; Ogdena, A.; Ohrena, J.; Vara Prasada, J.V.N.; Shellya, J.A.; Skerlosa, L.; Sulavika, M.; Thomasa, V.H.; VanderRoesta, S.; Wangc, L.; Wanga,Z.; Whittona, A; Zhua, T; Stovera, C.K. A class of selective antibacterials derived from a protein kinase inhibitor pharmacophore. *P. Natl. Acad. Sci. USA,* **2009**, *106*, 1737–1742.

[22] Walsh, C.T.; Fischbach, M.A. Repurposing libraries of eukaryotic protein kinase inhibitors for antibiotic discovery, *P. Natl. Acad. Sci. USA,* **2009**, *106*, 1689–1690.

[23] Young, D.W; Bender, A.; Hoyt, J.; McWhinnie, E.; Chirn, G.W.; Tao C.Y.; Tallarico J.A.; Labow M.; Jenkins J.L.; Mitchison T.J.; Feng Y. Integrating high-content screening and ligand-target prediction to identify mechanism of action, *Nat. Chem. Biol.,* **2008**, *4*(1), 59-68.

[24] Keiser, M.J.; Setola, V.; Irwin, J.J.; Laggner C.; Abbas, A.I.; Hufeisen, S.J.; Jensen, N.H.; Kuijer, M.B.; Matos, R.C.; Tran, T.B.; Whaley, R.; Glennon, R.A.; Hert J.; Thomas, K.L.H.; Edwards, D.D.; Shoichet, B.K.; Roth, B.L. Predicting new molecular targets for known drugs, *Nature,* **2009**, *462*, 175–181.

[25] Bender, A.; Jenkins, J.L.; Glick, M.; Deng, Z.; Nettles, J.H.; Davies, J.W. Bayes affinity fingerprints improve retrieval rates in virtual screening and define orthogonal bioactivity space: when are multitarget drugs a feasible concept?, *J. Chem. Inf. Model.,* **2006**, *46*, 2445–2456.

[26] Rognan, D. Structure-based approaches to target fishing and ligand profiling, *Mol. Inf*., **2010**, *29*(*3*), 176–187.

[27] Teo, C.Y.; Abdul Rahman, M.B.; Thean Chor A.L.; Salleh A.B.; Ballester P.J.; Tejo B.A. Ligand-Based Virtual Screening for the Discovery of Inhibitors for Protein Arginine Deiminase Type 4 (PAD4), *Metabolomics,* **2013**, *3*, 118-122.

[28] Swamidass, S.J. Mining small-molecule screens to repurpose drugs, *Brief Bioinform.,* **2011**, *12*(*4*), 327–35.

[29] Wang, Y.; Chen, S.; Deng, N. Drug repositioning by kernel- based integration of molecular structure, molecular activity, and phenotype data, *PLoS One,* **2013**, *8*(11), e78518.

[30] Desaphy, J.; Raimbaud, E.; Ducrot, P.; Rognan D. Encoding protein-ligand interaction patterns in fingerprints and graphs *J. Chem. Inf. Model.,* **2013,** *53*(3), 623–637.

[31] Tan, F.J.; Yang, R.Z.; Xu X.X.; Chen, X.; Wang, Y.; Ma, H.; Liu, X.; Wu, X.; Chen, Y.; Liu, L.; Jia, X. Drug repositioning by applying 'expression profiles' generated by integrating chemical structure similarity and gene semantic similarity, *Mol. Biosyst.*, **2014**, *10*(5), 1126–38.

[32] Ng, C.; Hauptman, R.; Zhang, Y.; Bourne, P.E.; Xie, L. Anti-infectious drug repurposing using an integrated chemical genomics and structural systems biology approach, *Pac. Symp. Biocomput.,* **2014**, 136–47.

[33] Lee, M.; Kim, D. Large-scale reverse docking profiles and their applications, *BMC Bioinfo.,* **2012**, *13*(17), S6, 1-10.

[34] Grinter, S.Z. Liang, Y.; Huang, S.Y.; Hyder, S.M.; Zou, X. An inverse docking approach for identifying new potential anti-cancer targets, *J. Mol. Graph. Model.,* **2011**, *29*, 795–799.

[35] Henrik, K.; Feng, Z; Bluhm, W.F.; Dimitropoulos, D.; Doreleijers, J.F.; Dutta, S.; Flippen-Anderson, J.L.; Ionides, J.; Kamada, C.; Krissinel, E.; Lawson, C.L.; Markley, J.L.; Nakamura, H.; Newman, R.; Shimizu, Y.; Swaminathan, J; Velankar, S.; Ory, J.; Ulrich E.L.; Vranken, W.; Westbrook, J.; Yamashita, R.; Yang, H.; Young, J.; Yousufuddin, M.; Berman, H.M.; Remediation of the Protein Data Bank archive, *Nuc. Acids Res.,* **2008**, *36*, D426–D433.

[36] Langer, T. Pharmacophores in drug research, *Mol. Info.,* **2010**, *29*(6-7), 470–475.

[37] Duwensee, K.; Schwaiger, S.; Tancevski, I.; Eller, K.; Van Eck, M.; Markt, P.; Linder, T.; Stanzl, U.; Ritsch, A.; Patsch, J.R.; Schuster, D.; Stuppner, H.; Bernhard, D.; Eller, P. Leoligin, the major lignan from Edelweiss, activates cholesteryl ester transfer protein, *Atherosclerosis,* **2011**, *219*, 109–115.

[38] Nashev, L.G.; Vuorinen, A.; Praxmarer, L.; Chantong, B.; Cereghetti, D.; Winiger, R.; Schuster, D.; Odermatt, A. Virtual screening as a strategy for the identification of xenobiotics disrupting corticosteroid action, *Plos One,* **2012**, *7*(10), e46958.

[39] Haupt, V.J.; Daminelli, S.; Schroeder, M. Drug Promiscuity in PDB: Protein Binding Site Similarity Is Key, *Plos One,* **2013**, *8*(6), e65894.

[40] Weill, N.; Rognan D. Alignment-Free Ultra-High-Throughput Comparison of Druggable Protein−Ligand Binding Sites, *J. Chem. Inf. Model., **2010**, 50*(1), 123–135.

[41] Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank, *Nucleic Acids Res.,* **2000**, *28*, 235–242.

[42] Liu, T.; Lin, Y.; Wen, X.; Jorrisen, R.N.; Gilson, M.K. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities, *Nucleic Acids Res.,* **2007**, *35*, D198-D201

[43] Bader, G.D.; Cary, M.P.; Sander, C. Pathguide: a pathway resource list, *Nucleic Acids Res.,* **2006**, *34*, D504-D506.

[44] Gaulton, A.; Bellis, L.J.; Bento, A.P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J.P. ChEMBL: a large-scale bioactivity database for drug discovery, *Nucleic Acids Res.,* **2012**, *40*, D1100–D1107.

[45] Bryant, S. PubChem: an information resource linking chemistry and biology. *Abstr. Pap. Am. Chem. S.,* **2007**, 231.

[46] Bolton, E.E.; Wang, Y.; Thiessen, P.A.; Bryant, S.H. PubChem: integrated platform of small molecules and biological activities, *Annu. Rep. Com. Chem.,* **2008**, *4*(12), 217–241.

[47] Evans, J.M.; Lopez, E.; Roth, B.L. The NIMH Psychoactive Drug Screening Program's Ki database: an on- line, searchable database of receptor- ligand affinity values, *Soc. Neurosci. Abs.,* **2001**, *27*, 2076

[48] Björling, E.; Uhlén, M. Antibodypedia - a portal for sharing antibody and antigen validation data, *Mol. Cell. Proteomics,* **2008**, *7*, 2028-2037.

[49] Cottingham, K. Antibodypedia seeks to answer the question: How good is your antibody? *J. Proteome Res.,* **2008**; *7*(10), 4213–4213.

[50] Kiermer V. Antibodypedia - A web portal to share antibody validation data, *Nat. Methods,* **2008**, *5*, 860-870.

[51] Chambers, J.; Davies, M.; Gaulton, A.; Hersey, A.; Velankar, S.; Petryszak, R.; Hastings, J.; Bellis, L.; McGlinchey, S.; Overington,

J.P. UniChem: a unified chemical structure cross-referencing and identifier tracking system, *J. Cheminform.,* **2013**, *5*(1), 3-11.
[52] Chambers, J.; Davies, M.; Gaulton, A.; Papadatos, G.; Hersey, A.; Overington, J.P. UniChem: extension of InChI-based compound mapping to salt, connectivity and stereochemistry layers, *J. Cheminform.,* **2014**, *6*(1), 43-53.
[53] The Binding Database. www.bindingDB.org (Accessed June 03 **2015**).
[54] ChEMBL. www.ebi.ac.uk/chembl/ (Accessed June 03 **2015**).
[55] Pence, H. E.; Williams, A. ChemSpider: An Online Chemical Information Resource, *J. Chem. Educ*., **2010**, *87*(11), 1123-1124.
[56] ChemSpider. www.chemspider.com (Accessed June 03 **2015**).
[57] Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V.; Djoumbou, Y.; Eisner, R.; Guo, A.C.; Wishart, D.S. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs, *Nucleic Acids Res*., **2011**, *39*, D1035–D1041.
[58] DrugBank. www.drugbank.ca (Accessed June 03 **2015**).
[59] Li, Q.; Xiang, J.F.; Yang, Q.F.; Sun, H.X.; Guan, A.J.; Tang, Y.L. G4LDB: a database for discovering and studying G-quadruplex ligands, *Nucleic Acids Res*., **2013**, *41*(1), D1115-D1123.
[60] G-quadruplex ligands database. http://www.g4ldb.org/ci2/index.php (Accessed June 03 **2015**).
[61] Alexander, S.P.; Mathie, A.; Peters, J. A. Guide to Receptors and Channels (GRAC), 5th edition. *British J. Pharmacol*., **2011**, 164(1), S1–S324.
[62] Sharman, J.L.; Mpamhanga, C.P.; Spedding, M.; Germain, P.; Staels, B.; Dacquet, C.; Laudet, V.; Harmar, A.J.; Nc, I. IUPHAR-DB: new receptors and tools for easy searching and visualization of pharmacological data, *Nucleic Acids Res*., **2011**, *39*, D534–D538.
[63] Harmar, A.J.; Hills, R.A.; Rosser, E.M.; Jones, M.; Buneman, O.P.; Dunbar, D.R.; Greenhill, S.D.; Hale, V.A.; Sharman, J.L.; Bonner, T.I.; Catterall, W.A.; Davenport, A.P.; Delagrange, P.; Dollery, C.T.; Foord, S.M.; Gutman, G.A.; Laudet, V.; Neubig, R.R.; Ohlstein, E.H.; Olsen, R.W.; Peters, J.; Pin, J.P.; Ruffolo, R.R.; Searls, D.B.; Wright, M.W.; Spedding, M. IUPHAR-DB: the IUPHAR database of G protein-coupled receptors and ion channels, *Nucleic Acids Res.,* **2009**, *37*, D680–D685.
[64] IUPHAR BPS. www.guidetopharmacology.org (Accessed June 03 **2015**).
[65] DTP Developmental Therapeutics Program NCI/NIH. http://dtp.nci.nih.gov (Accessed June 03 **2015**).
[66] Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind database: collection of binding affinities for protein–ligand complexes with known three-dimensional structures, *J. Med. Chem*., **2004**, *47*, 2977–2980.
[67] Li Y.; Liu Z.H.; Li J.; Han L.; Liu J.; Zhao Z.X.; Wang R.X. Comparative Assessment of Scoring Functions on an Updated Benchmark: I. Compilation of the Test Set, *J. Chem. Inf. Model.,* **2014**, *54*(6), 1700–1716.
[68] Li Y.; Han L.; Liu Z.H.; Wang R.X. Comparative Assessment of Scoring Functions on an Updated Benchmark: II. Evaluation Methods and General Results, *J. Chem. Inf. Model.,* **2014**, *54* (6), 1717–1736.
[69] PdbBind. www.pdbbind.org.cn (Accessed June 03 **2015**).
[70] PDSP-NIMH Psychoactive Drug Screening Program https://pdspdb.unc.edu/pdspWeb/ (Accessed June 03 **2015**).
[71] Wang, Y.; Xiao, J.; Suzek, T.O.; Zhang, J.; Wang, J.; Zhou, Z.; Han, L.; Karapetyan, K.; Dracheva, S.; Shoemaker, B.A.; Bolton, E.; Gindulyte, A.; Bryant, S.H. PubChem's BioAssay Database. *Nucleic Acids Res.,* **2012**, *40*, D400–D412.
[72] PubChem. http://pubchem.ncbi.nlm.nih.gov/ (Accessed June 03 **2015**).
[73] Gunther, S.; Kuhn, M.; Dunkel, M.; Campillos, M.; Senger, C.; Petsalaki, E.; Ahmed, J.; Urdiales, E.G.; Gewiess, A.; Jensen, L.J.; Schneider, R.; Skoblo, R.; Russell, R.B.; Bourne, P.E.; Bork, P.; Preissner, R. SuperTarget and Matador: resources for exploring drug– target relationships, *Nucleic Acids Res*., **2008**, *36*, D919–D922.
[74] Hecker, N.; Ahmed, J.; von Eichborn, J.; Dunkel, M.; Macha, K.; Eckert, A.; Gilson, M.K.; Bourne, P.E.; Preissner, R. SuperTarget goes quantitative: update on drug–target interactions, *Nucleic Acids Res*., **2012**, *40*, D1113–D1117.
[75] SuperTarget. http://bioinf-apache.charite.de/supertarget_v2/ (Accessed June 03 **2015**).
[76] Qin, C.; Zhang, C.; Zhu, F.; Xu, F.; Chen, S.Y.; Zhang, P.; Li, Y.H.; Yang, S.Y.; Wei, Y.Q.; Tao, L.; Chen, Y.Z. Therapeutic target database update 2014: a resource for targeted therapeutics, *Nucleic Acids Res*., **2014**, *42*, D1118-D1123.
[77] Therapeutic Target Database (TTD). http://bidd.nus.edu.sg/group/cjttd/ (Accessed June 03 **2015**).
[78] Irwin, J.J.; Sterling, T.; Mysinger, M.M.; Bolstad, E.S.; Coleman, R.G. ZINC: A Free Tool to Discover Chemistry for Biology, *J. Chem. Inf. Model.,* **2012**, *52*(7),1757-1768.
[79] Zinc Database. http://zinc.docking.org/ (Accessed June 03 **2015**).
[80] Berman, H.; Henrick, K.; Nakamura, H.; Markley, J.L. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data, *Nucleic Acids Res*., **2007**, *35*, D301–D303.
[81] Nair, R.; Liu, J.; Soong, T.T.; Acton, T.B.; Everett, J.K.; Kouranov, A.; Fiser, A.; Godzik, A.; Jaroszewski, L.; Orengo, C.; Montelione, G.T.; Rost, B. Structural genomics is the largest contributor of novel structural leverage, *J. Struct. Funct. Genomics,* **2009**, *10*, 181–191.
[82] Terwilliger, T. C.; Bricogne, G. Continuous mutual improvement of macromolecular structure models in the PDB and of X-ray crystallographic software: the dual role of deposited experimental data, *Acta Cryst.,* **2014**, *70*, 2533-2543
[83] Brown, E.N.; Ramaswamy, S. Quality of protein crystal structures, *Acta Cryst*., **2007**, *63*, 941–950.
[84] Dalby, A.; Nourse, J.G.; Hounshell, W.D.; Gushurst, A.K.I.; Grier, D.L.; Leland, B.A.; Laufer, J. Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited, *J. Chem. Inf. Comp. Sci.,* **1992**, *32*, 244–255.
[85] Read, R. J.; Adams, P. D.; Arendall, W. B. A new generation of crystallographic validation tools for the Protein Data Bank, *Structure,* **2011,** 19, 1395-1412.
[86] Montelione, G. T.; Nilges, M.; Bax, A.; Güntert, P.; Herrmann, T.; Richardson, J.S.; Schwieters, C.D.; Vranken, W.F.; Vuister, G.W.; Wishart, D.S.; Berman, H.M.; Kleywegt, G.J.; Markley, J.L.; Recommendations of the wwPDB NMR Structure Validation Task Force, *Structure,* **2013**, *21*, 1563-1570.
[87] The Protein Data Bank. http://www.rcsb.org/pdb/home/home.do (Accessed June 03 **2015**).
[88] Paul, N.; Kellenberger, E.; Bret, G.; Muller, P.; Rognan, D. Recovering the true targets of specific ligands by virtual screening of the Protein Data Bank, *Proteins,* **2004**, *54*, 671–680
[89] Marcou, G.; Rognan, D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints, *J. Chem. Inf. Model.,* **2007**, *47*, 195–207.
[90] Meslamani, J.; Rognan, D.; Kellenberger, E. sc-PDB: a database for identifying variations and multiplicity of "druggable" binding sites in proteins, *Bioinformatics,* **2011**, *27*, 1324–1326.
[91] Desaphy, J.; Bret, G.; Rognan, D.; Kellenberger, E. sc-PDB: a 3D-database of ligandable binding sites-10 years on, *Nucleic Acids Res*., **2015**, *43*, D399-D404.
[92] Duan, J.; Dixon, S.L.; Lowrie, J.F.; Sherman, W. Analysis and Comparison of 2D Fingerprints: Insights into Database Screening Performance Using Eight Fingerprint Methods, *J. Molec. Graph. Model.,* **2010**, *29*, 157-170.
[93] Sastry, M.; Lowrie, J.F.; Dixon, S.L.; Sherman, W. Large-Scale Systematic Analysis of 2D Fingerprint Methods and Parameters to Improve Virtual Screening Enrichments, *J. Chem. Inf. Model.,* **2010**, *50*, 771-784.
[94] Katritzky, A. R.; Karelson, M.; Maran, U.; Wang, Y. QSPR and QSAR Models Derived Using Large Molecular Descriptor Spaces. A Review of CODESSA Applications, *Collect. Czech. Chem. Commun.,* **1999**, *64*, 1551-1560.
[95] Yang, C.; Tarkhov, A.; Marusczyk, J.; Bienfait, B.; Gasteiger, J.; Kleinoeder, T.; Magdziarz, T.; Sacher, O.; Schwab, C.; Schwö bel, J.; Terfloth, L.; Arvidson, K.; Richard, A.M.; Worth, A.P.; Rathman, J.F. A New Publicly Available Chemical Query Language, CSRML, to support Chemotype Representations for Application to Data Mining and Modeling, *J. Chem. Inf. Model.,* **2015**, *55*(3), 510–528.
[96] Mauri A.; Consonni V.; Pavan M.; Todeschini R.; Dragon Software: An easy approach to molecular descriptors calculations, MATCH *Commun. Math. Comput. Chem*., **2006**, *56*, 237-248.
[97] Sawada, R.; Kotera, M.; Yamanishi, Y. Benchmarking a wide range of chemical descriptors for drug-target interaction prediction

using a chemogenomic approach, *Mol. Inform*., **2014**, *33* (11-12), 719-731.

[98] ChemAxon Instant. *J. Chem.,* https://www.chemaxon.com/ products/instant-jchem-suite/instant-jchem/ (Accessed June 03 **2015**).

[99] Varnek, A.; Fourches, D.; Horvath, D.; Klimchuk, O.; Gaudin, C.; Vayer, P.; Solov'ev, C.; Hoonakker, F.; Tetko, I.; Marcou, G. ISIDA - platform for virtual screening based on fragment and pharmacophoric descriptors, *Curr. Comput. Aided. Drug. Des.,* **2008**, *4*, 191–198.

[100] Bochevarov, A.D.; Harder, E.; Hughes, T.F.; Greenwood, J.R.; Braden, D.A.; Philipp, D.M.; Rinaldo, D.; Halls, M.D.; Zhang, J.; Friesner, R.A. Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences, *Int. J. Quantum Chem.*, **2013**, *113*(18), 2110-2142.

[101] Wolber, G.; Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters, *J. Chem. Inf. Model.,* **2005**, *45*(1), 160–169.

[102] Wolber, G.; Dornhofer, A.A.; Langer, T. Efficient overlay of small organic molecules using 3D pharmacophores, *J. Comput. Aided. Mol. Des.,* **2007**, *20*(12), 773–788.

[103] Clark, A.M.; Labute, P.; Santavy, M. 2D Structure Depiction, *J. Chem. Inf. Model.,* **2006**, *46*, 1107-1123.

[104] Clark, A.M.; Labute, P. Detection and Assignment of Common Scaffolds in Project Databases of Lead Molecules, *J. Med. Chem.,* **2008**, *52*, 469-483.

[105] Clark, A.M.; Labute, P. SD File Processing with MOE Pipeline Tools, *JCCG,* **2008**.

[106] Yap, C.W. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints, *J. Comput. Chem*., **2011**, *32*(7), 1466-1474.

[107] Dixon, S.L.; Smondyrev, A.M.; Knoll, E.H.; Rao, S.N.; Shaw, D.E.; Friesner, R.A. PHASE: A New Engine for Pharmacophore Perception, 3D QSAR Model Development, and 3D Database Screening. Methodology and Preliminary Results, *J. Comput. Aided Mol. Des*., **2006**, *20*, 647-671.

[108] Dixon, S.L.; Smondyrev, A.M.; Rao, S.N. PHASE: A Novel Approach to Pharmacophore Modeling and 3D Database Searching, *Chem. Biol. Drug Des*., **2006**, *67*, 370-372.

[109] Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S.; Olson, A.J. Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility, *J. Comp. Chem*., **2009**, *16*, 2785-2791.

[110] Huey, R.; Morris, G.M.; Olson, A.J.; Goodsell, D. S. A Semiempirical Free Energy Force Field with Charge-Based Desolvation, *J. Comp.Chem.,* **2007**, *28*, 1145-1152.

[111] Lang, P.T.; Brozell, S.R.; Mukherjee, S.; Pettersen, E.T.; Meng, E.C.; Thomas, V.; Rizzo, R.C.; Case, D.A.; James, T.L.; Kuntz, I.D. DOCK 6: Combining Techniques to Model RNA-Small Molecule Complexes, *RNA,* **2009**, *15*,1219-1230.

[112] Graves, A.P.; Shivakumar, D.E.; Boyce, S.E.; Jacobson, M.P.; Case D.A.; Shoichet B.K. Rescoring Docking Hit Lists for Model Cavity Sites: Predictions and Experimental Testing, *J. Mol. Biol*., **2008**, *377*, 914-934.

[113] Friesner, R.A.; Murphy, R.B.; Repasky, M.P.; Frye, L.L.; Greenwood, J.R.; Halgren, T.A.; Sanschagrin, P.C.; Mainz, D.T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes, *J. Med. Chem*., **2006**, *49*, 6177–6196.

[114] Halgren, T.A.; Murphy, R.B.; Friesner, R.A.; Beard, H.S.; Frye, L.L.; Pollard, W.T.; Banks, J.L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening*, J. Med. Chem.,* **2004**, *47*, 1750–1759.

[115] Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J. J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shaw, D.E.; Shelley, M.; Perry, J.K.; Francis, P.; Shenkin, P.S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy, *J. Med. Chem*., **2004**, *47*, 1739–1749.

[116] Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C.W.; Taylor, R. D. Improved protein-ligand docking using GOLD, *Proteins,* **2003**, *52*, 609-623.

[117] Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T. M.; Mortenson, P. N.; Murray, C.W. Diverse, high-quality test set for the validation of protein-ligand docking performance, *J. Med. Chem*., **2007**, *50*, 726-741.

[118] Kirton, S. B.; Murray, C.W.; Verdonk, M. L.; Taylor, R. D. Prediction of binding modes for ligands in the cytochromes P450 and other heme-containing proteins, *Proteins,* **2005**, *58*, 836-844.

[119] Brewerton, S. C. The use of protein-ligand interaction fingerprints in docking, *Curr. Opin. Drug Discov. Devel*., **2008**, *11*(3), 356-64.

[120] Soga, S.; Shirai, H.; Kobori, M.; Hirayama, N. Use of Amino Acid Composition to Predict Ligand-Binding Sites, *J. Chem. Inf. Model.,* **2007**, *47*, 400–406.

[121] Das, S.; Kokardekar, A; Breneman C.M. Rapid comparison of protein binding site surfaces with Property Encoded Shape Distributions (PESD), *J. Chem. Inf. Model.,* **2009**, *49*(12), 2863–2872.

[122] Corbeil, C.R.; Williams, C.I.; Labute, P. Variability in Docking Success Rates Due to Dataset Preparation, *J. Comp.-Aided Mol. Des*., **2012**, *26*, 775–786.

[123] O'Meara, M.J.; Leaver-Fay, A; Tyka, M.D.; Stein, A.; Houlihan, K.; DiMaio, F.; Bradley, P.; Kortemme, T.; Baker, D.; Snoeyink, J.; Kuhlman B. Combined Covalent-Electrostatic Model of Hydrogen Bonding Improves Structure Prediction with Rosetta, *J. Chem. Theory Comput*., **2015**, *11*(2), 609–622.

[124] Cheng C.Y.; Chou F.C.; Das R. Modeling complex RNA tertiary folds with Rosetta, *Methods. Enzymol*., **2015**, *553*, 35-64.

[125] Johnson D.K.; Karanicolas J. Selectivity by small-molecule inhibitors of protein interactions can be driven by protein surface fluctuations, *PLoS Comput. Biol.,* **2015**, *11*(2), e1004081.

[126] Halgren, T. Identifying and characterizing binding sites and assessing druggability, *J. Chem. Inf. Model.,* **2009**, *49*, 377–389.

[127] Halgren, T. New method for fast and accurate binding-site identification and analysis, *Chem. Biol. Drug Des.,* **2007**, *69*, 146–148.

[128] Beisken, S.; Meinl, T.; Wiswedel, B.; De Figueiredo, L. F.; Berthold, M.; Steinbeck, C. KNIME-CDK: Workflow-driven Cheminformatics, *BMC Bioinformatics,* **2013**, *14*, 257-260.

[129] Tiwari, A.; Sekhar, A.K.T. Workflow based framework for life science informatics, *Comput. Biol. Chem*., **2007**, *31*(5-6), 305–319.

[130] Forman, G.; Scholz, M. Apples-to-Apples in Cross-Validation Studies: Pitfalls in Classifier Performance Measurement, *ACM SIGKDD,* **2009**, *12*(1), 49-57.

[131] Kornhuber, J.; Terfloth, L.; Bleich S.; Wiltfang J.; Rupprecht R. Molecular properties of psychopharmacological drugs determining non-competitive inhibition of 5-HT3A receptors, *Eur. J. Med. Chem*., **2009**, *44*(6), 2667-2672.

[132] Monga, M.; Sausville, E.A. Developmental therapeutics program at the NCI: molecular target and drug discovery process, *Leukemia,* **2002**, *16*(4), 520-526.

[133] Shankavaram, U.T.; Varma, S.; Kane, D.; Sunshine, M.; Chary, K.K.; Reinhold, W.C.; Pommier, Y.; Weinstein, J.N. CellMiner: a relational database and query tool for the NCI-60 cancer cell lines, *BMC Genomics,* **2009**, *10*, 277-286.

[134] Reinhold, W.C.; Sunshine, M.; Liu, H.; Varma, S.; Kohn, K.W.; Morris, J.; Doroshow, J.; Pommier, Y. CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set, *Cancer Res.,* **2012**, *72*(14), 3499–3511.

[135] Yamanishi, Y.; Kotera, M.; Moriya, Y.; Sawada, R.; Kanehisa, M.; Goto, S. DINIES: drug–target interaction network inference engine based on supervised analysis, *Nucl. Acids Res.,* **2014**, *42*(1), 39-45.

[136] Takarabe M.; Kotera M.; Nishimura Y.; Goto S.; Yamanishi Y. Drug target prediction using adversed event report systems: a pharmacogenomic approach. *Bioinformatics,* **2012,** *28,* i611-i618

[137] Konc, J.; Janezic, D.; ProBiS–2012: web server and web services for detection of structurally similar binding sites in proteins, *Nucleic Acids Res*., **2012**, *40*, 214-221.

[138] Kuhn, M.; Szklarczyk, D.; Pletscher-Frankild, S.; Blicher, T.H.; von Mering, C.; Jensen, L.J.; Bork, P. Stitch 4: integration of protein–chemical interactions with user data*, Nucl. Acids Res*., **2014**, *42* (D1), D401-D407.

[139] Liu,R.; Singh, N.; Tawa, G.J.; Wallqvist A.; Reifman J. Exploiting large-scale drug-protein interaction information for computational drug repurposing, *BMC Bioinformatics,* **2014**, *15,* 210-226.

[140] Nickel, J.; Gohlke, B.; Erehman, J.; Banerjee, P.; Rong, W.W.; Goede, A.; Dunkel, M.; Preissner, R. SuperPred: update on drug classification and target prediction, *Nucl. Acids Res*., **2014**, *42*(1), 26-31.

[141] Gao, Z.; Li, H.; Zhang, H.; Liu, X.; Kang, L.; Luo, X.; Zhu, W.; Chen, K.; Wang X.; Jiang, H. PDTD: a web-accessible protein database for drug target identification, *BMC Bioinformatics,* **2008**, *9*,104-110.

[142] Li, H.; Gao, Z.; Kang, L.; Zhang, H.; Yang, K.; Yu, K.; Luo, X.; Zhu, W.; Chen, K.; Shen, J.; Wang X.; Jiang, H. TarFisDock: a web server for identifying drug targets with docking approach, *Nucl. Acids Res*., **2006**, *34*, 219-224.

[143] Cobanoglu, M.C; Oltvai, Z.N.; Taylor, D.L.; Bahar, I. BalestraWeb: efficient online evaluation of drug–target interactions, *Bioinformatics*, **2015**, *31* (1), 131-133.

[144] Low, Y.; Gonzalez, J.; Kyrola, A.; Bickson, D.; Guestrin, C.; Hellerstein, J. M. GraphLab: A New Framework for Parallel Machine Learning, the 26th Conference on Uncertainty in Artificial Intelligence, Catalina Island, California, July 8-11, **2010**.

[145] Kamburo, A.; Stelzl, U.; Lehrach, H.; Herwig, R. The ConsensusPathDB interaction database: 2013 update, *Nucl. Acids Res.,* **2013**, *41* (D1): D793-D800

[146] Tao, C.; Sun, J; Zheng, W.J.; Chen, J.; Xu, H; Colorectal cancer drug target prediction using ontology-based inference and network analysis, *Database* (*Oxford*) *2015,* **2015**: bav015.

[147] Zhong, Y.; Chen, E.Y.; Liu, R.; Chuang, P.Y.; Mallipattu, S.K.; Tan, C.M.; Clark, N.R.; Deng, Y.; Klotman, P.E.; Ma'ayan, A.; He, J.C. Renoprotective effect of combined inhibition of angiotensin-converting enzyme and histone deacetylase, *J. Am. Soc. Nephrol*., **2013**, *24*(5), 801-11.

[148] Amelio, I.; Gostev, M.; Knight, R.A.; Willis, A.E.; Melino, G.; Antonov A.V.; DRUGSURV: a resource for repositioning of approved and experimental drugs in oncology based on patient survival information, *Cell. Death Dis.,* **2014**, *6*(5), 1051-1056.

[149] Alaimo, S.; Bonnici, V.; Cancemi, D.; Ferro, A.; Giugno, R.; Pulvirenti, A. DT-Web: a web-based application for drug-target interaction and drug combination prediction through domain-tuned network-based inference, *BMC Syst Biol.,* **2015**, *9* (3:S4) 1-11.

[150] Wong, J.G. STC: Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines**.** *Drug Discov. Today,* **2014**, *19* (5) 637-644

[151] Pacini, C.; Iorio, F.; Gonçalves, E.; Iskar, M.; Klabunde, T.; Bork, P.; Saez-Rodriguez, J. DvD: An R/Cytoscape pipeline for drug repurposing using public repositories of gene expression data. *Bioinformatics,* **2013**, *29*(1), 132-134.

[152] Bolton, E.; King, J.; Morris, D.L. H2-antagonists in the treatment of colon and breast cancer, *Semin Cancer Biol*., **2000**, *10*(1), 3-10.

[153] Li, F.; Cao, Y.; Han, L.; Cui, X.; Xie, D; Wang, S.; Bo, X. GeneExpressionSignature: an R package for discovering functional connections using gene expression signatures, *OMICS,* **2013**, *17*(2),116-118.

[154] Kim, J.; Yoo, M.; Kang, J.; Tan, A.C. K-Map: connecting kinases with therapeutics for drug repurposing and development, *Hum. Genomics,* **2013**, 7-20.

[155] Roy, A.; Skolnick, J. LIGSIFT: an open-source tool for ligand structural alignment and virtual screening, *Bioinformatics,* **2015**, *31*(4), 539-544.

[156] Carrella, D.; Napolitano, F.; Rispoli, R.; Miglietta, M.; Carissimo, A.; Cutillo, L.; Sirci, F.; Gregoretti, F.; Di Bernardo, D. Mantra 2.0: an online collaborative resource for drug mode of action and repurposing by network analysis, *Bioinformatics,* **2014**, *130* (12), 1787-1788.

[157] Iorio, F.; Bosottic, R.; Scacheric, E.; Belcastro, V.; Mithbaokar, P.; Ferriero, R.; Murino, L.; Tagliaferri, R.; Brunetti-Pierria, N.; Isacchic, A.; di Bernard, D. Discovery of drug mode of action and drug repositioning from transcriptional responses, *PNAS,* **2010**. *107* (33), 14621–14626.

[158] Setoain, J.; Franch, M.; Martínez, M.; Tabas-Madrid, D.; Sorzano, C. O. S.; Bakker, A.; Gonzalez-Couto, E.; Elvira, J.; Pascual-Montano, A. NFFinder: an online bioinformatics tool for searching similar transcriptomics experiments in the context of drug repositioning, *Nucl. Acids Res.,* **2015**, *43*(W1), W193-99.

[159] Lee, H.S.; Bae, T.; Lee, J.H.; Kim, D.G.; Oh, Y.S.; Jang, Y.; Kim, J.T.; Lee, J.J.; Innocenti, A.; Supuran, C.T.; Chen, L.; Rho, K.; Kim, S. Rational drug repositioning guided by an integrated pharmacological network of protein, disease and drug, *BMC Syst. Biol*., **2012**, *6*, 80-90.

[160] Lauria, A.; Tutone, M.; Almerico, A. M. "Virtual lock-and-key approach: The *in silico* revival of Fischer model by means of molecular descriptors", *Eur. J. Med. Chem.,* **2011**, *46*(9), 4274-4280.

[161] Lauria, A.; Tutone, M.; Barone, G.; Almerico, A. M. Multivariate analysis in the identification of biological targets for designed molecular structures: The BIOTA protocol, *Eur. J. Med. Chem*., **2014**, *75*, 106-110.

[162] Lauria, A.; Patella, C.; Abbate, I.; Martorana, A.; Almerico, A.M. Lead optimization through VLAK protocol: New annelated pyrrolo-pyrimidine derivatives as antitumor agents, *Eur. J. Med. Chem.,* **2012**, *55*, 375-383.

[163] Lauria, A.; Abbate, I.; Patella, C.; Martorana, A.; Dattolo, G.; Almerico, A.M.; New Annelated Thieno[2,3-e][1,2,3]triazolo[1,5-a]pyrimidines, with Potent Anticancer Activity, Designed through VLAK Protocol, *Eur. J. Med. Chem*., **2013**, *62*, 416-424.

[164] Martorana, A.; Gentile, C.; Perricone, U.; Piccionello, A. P.; Bartolotta, R.; Terenzi, A.; Pace, A.; Mingoia, F.; Almerico, A. M.; Lauria, A.; Synthesis, antiproliferative activity, and *in silico* insights of new 3-benzoylamino-benzo[b]thiophene derivatives, *Eur. J. Med. Chem*., **2015**, *90*, 537-546.

[165] Craddock ,T.J.A.; Harvey, J. M.; Nathanson, L.; Barnes, Z.M.; Klimas, N.G.; Fletcher, M.A.; Broderick, G. Using gene expression signatures to identify novel treatment strategies in gulf war illness, *BMC Med. Genomics,* **2015**, *8*, 36-49.

Received: June 12, 2015 **Revised: August 09, 2015** **Accepted: November 01, 2015**