NPR

Chemo- and bioinformatics resources for in silico drug discovery from medicinal plants beyond their traditional use: a critical review

Journal:	*Natural Product Reports*
Manuscript ID:	NP-REV-05-2014-000068.R1
Article Type:	Review Article
Date Submitted by the Author:	11-Jun-2014

Complete List of Authors:	Lagunin, Alexey; Institute of Biomedical Chemistry RAMS, Bioinformatics; Russian National Research Medical University, Goel, Rajesh; Punjabi University, Pharmaceutical Sciences and Drug Research Gawande, Dinesh; Punjabi University, Pharmaceutical Sciences and Drug Research Pahwa, Priynka; Punjabi University, Pharmaceutical Sciences and Drug Research Gloriozova, Tatyana; Institute of Biomedical Chemistry RAMS, Bioinformatics Dmitriev, Alexander; Institute of Biomedical Chemistry RAMS, Bioinformatics Ivanov, Sergey; Institute of Biomedical Chemistry RAMS, Bioinformatics Konova, Anastassia; Institute of Biomedical Chemistry RAMS, Bioinformatics Pavel, Pogodin; Institute of Biomedical Chemistry RAMS, Bioinformatics; Russian National Research Medical University, Biochemistry Druzhilovsky, Dmitry; Institute of Biomedical Chemistry RAMS, Bioinformatics Poroikov, Vladimir ; Institute of Biomedical Chemistry RAMS, Bioinformatics; Russian National Research Medical University, Biochemistry

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

REVIEW

Chemo- and bioinformatics resources for *in silico* drug discovery from medicinal plants beyond their traditional use: a critical review

Alexey A. Lagunin,*[a,c] Rajesh K. Goel,*[b] Dinesh Y. Gawande,[b] Priynka Pahwa,[b] Tatyana A. Gloriozova,[a] Alexander V. Dmitriev,[a] Sergey M. Ivanov,[a] Anastassia V. Rudik,[a] Varvara I. Konova,[a] Pavel V. Pogodin,[a,c] Dmitry S. Druzhilovsky[a] and Vladimir V. Poroikov*[a,c]

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
DOI: 10.1039/b000000x

In silico approaches have been widely recognised to be useful for drug discovery. Here, we consider the significance of available databases of medicinal plants and chemo- and bioinformatics tools for *in silico* drug discovery beyond the traditional use of folk medicines. This review contains a practical example of the application of combined chemo- and bioinformatics methods to study pleiotropic therapeutic effects (known and novel) of 50 medicinal plants from Traditional Indian Medicine.

1 Introduction

Natural products have been used in folk medicine for thousands of years. One-third of the adult population in industrially developed countries and more than 80% of the population in developing countries uses herbal medicinal products to promote health and to treat common illnesses such as colds, inflammation, heart diseases, diabetes and central nervous system disorders. It is believed that plants interact with changing environmental stresses and adapt to these changes[1]. This adaptation is accompanied by unusual phytochemical diversity. More than 70% of new chemical substances (New Chemical Entities - NCEs), introduced into medical practice from 1981 to 2006 were derived from natural products.[2] These data confirm the assertion by Dhawan that the study of plants, based on their use in traditional systems of medicine, is a viable and cost-effective strategy for the development of new drugs.[3] Because there are several thousand pharmacological targets and because most natural compounds exhibit pleiotropic effects by interacting with different targets, computational methods are the methods of choice in drug discovery based on natural products.[4] The use of chemo- and bioinformatics methods for the exploration of their pleiotropic pharmacological potential beyond the traditional uses may be possible with the availability of medicinal plant databases including data on chemical structures and therapeutic uses of phytoconstituents identified over the years from medicinal plants.

Chemoinformatics and bioinformatics tools help in the identification of complementary leads and targets. Many of these approaches facilitate lead discovery against individual targets using molecular docking,[5,6] pharmacophores,[4] (Q)SAR, and machine learning methods.[7,8] Combinatorial approaches straightforwardly conduct parallel searches against each individual target to find virtual hits that simultaneously interact with multiple targets. Despite wide use of 3D target-based approaches they have limitations with respect to the number of targets with 3D structures. Therefore, the use of 2D structures appears more reasonable in facilitating leads for multiple targets. Plant phytoconstituents are explored either based on bioactivity-guided fractionation or through random screening of plant extracts. To date, only the bioactive principles for traditional activities have been used as templates for new drug discovery for known bioactivities using molecular docking. Thus, phytochemicals for which the biological activity is unknown are largely unexplored. Their potential can be efficiently investigated using multi-targeted *in silico* approaches.

Bioinformatics and systems biology approaches are becoming increasingly important along with the above-mentioned chemoinformatics methods to study the therapeutic potential of medicinal plants.[9] They are used to select targets for docking and to identify relationships between the revealed actions of phytochemicals on targets and the known therapeutic effects of medicinal plants. Thus, the aim of this review is a critical consideration of the various available databases of medicinal plants and *in silico* tools for their utilisation in new drug discovery based on expanding the use of folk medicinal plants through the exploration of phytochemical diversity.

2 Medicinal plant databases

The databases of medicinal plants are collections of particular information about plants used in folk medicine. Dozens databases and Internet sources partially containing such information became available during the last decade. We collected a list of currently available sources with English interface and data which were mentioned in peer-reviewed scientific journals and may be useful in studies of medicinal plants (Table 1). These databases were analysed for the following information:

1. Availability (freely accessible or commercial).
2. Plant Name.
3. Traditional uses.
4. Plant parts which are used for treatment.
5. Phytoconstituents.
6. Phytoconstituents with their 2D/3D structures.
7. Pharmacological and toxic activities of the phytoconstituents.
8. Possibility of download of phytoconstituent structures and properties.

The majority of databases contain different types of data on medicinal plants; therefore, in practical applications, scientists usually have to combine data from several databases. Most of the databases contain the botanical name, vernacular name and traditional uses of each of these plants; examples include databases on ethno-medicinal plants and the PFAF database. Other databases, such as the Dictionary of Natural Products, TradiMed, and SuperNatural, contain information about phytoconstituents discovered in these plants.

In this review we discuss application of virtual screening for identification of new leads for different biological targets beyond traditional use from these medicinal plants. Therefore the medicinal plant databases ideally requires to provide phytochemical and pharmacological information of medicinal plants, so we further screened 54 selected databases with respect to the phytochemistry and found 14 databases, discussing the phytochemistry of natural product (BoDD, Cardiovascular Disease Herbal Database (CVDHD), Chemical Abstract Services (CAS), Dictionary of Natural Product Database (DNP), Ethanobotany of the Peruvian amazon, Herbal think-TCM, Indo-Russian traditional medicine database, IBS natural product library, KNApSAcK core DB, Pubchem substance database, Super natural II database, Traditional Chinese Medicine Information Database (TCMID), TIPdb and TradiMed). Four from 14 databases have a limited access (paid access), which limits their use in new drug discovery: CAS, DNP, Herbal think-TCM and TradiMed. Plants are composed from a plurality of compounds belonging to different chemical classes and exhibiting diverse biological activities. They may enhance each other's actions or compensate for toxicity, or their interaction can lead to side effects. Traditional national medicines (TNM) rely on a wide range of plants and herbs. For example, approximately 1,250 plants are used in various Traditional Indian Medicine (TIM) preparations.[20] Both parts of plants and the whole plants are used in TNM. TNM compositions can also include various parts of different plants. The parts of plant can be extracted and prepared by various ways, and constituents of the same plants may vary in different geographical regions. We confirm the statement of Polur that, at the present time, there is no database reflecting all the current knowledge about the composition, method of preparation and use of medicinal plants, but existing databases can be used for *in silico* work associated with the drug discovery process.[21] The main database that may be used for drug discovery based on TNM is DNP, containing over 270,000 records. Unfortunately, it does not allow the extraction of data on sets of structures in SD files, but extracting single molecules is possible. CVDHD, TCMID, TIPdb and TradiMed are others useful resources of data containing information both about phytochemicals and therapeutic effects. Most databases include information about taxonomy, TNM usage and photos of plants. SuperNatural II database and IBS Natural products library containing structures of phytoconstituents may be the basis for virtual screening. Taxonomic databases (e.g. The Plant List, Tropicos) are helpful for validation of plant taxonomy.

The number of available databases increases every year. This opens up the possibility for the detailed study of plants and the utilisation of knowledge of their traditional use for the drug discovery process. In general, the data in a database should be as comprehensive as possible to allow its use in various biomedical studies.

3 Cheminformatics tools for exploring biological activity of medicinal plants

The amount of available data on the biological activity of the investigated compounds (including herbal medicines) and the number of target macromolecules related to their therapeutic effects increase every year (e.g., ChEMBLdb contains data on 1.5 million compounds acting on more than 9,000 targets). At the same time, the pool of data on compositions of medicinal plants has also increased (e.g., Natural Product Database contains information for more than 226,000 natural products with approximately 210,000 structures). Therefore, the need for use of *in silico* methods to determine the biological activity of medicinal plants is obvious.

The classic methods of (Q)SAR (quantitative or qualitative "structure-activity" relationships), molecular modelling and virtual screening widely applied for synthetic compounds in drug discovery may be used for exploring the biological activity of medicinal plants if information about the structures of phytochemicals is available. All these methods are based on assumption that activity of compounds depends on their structures. Three key components are necessary for creation of (Q)SAR models:

(1) Noncontradictory data on structures and biological activity of studied compounds;
(2) Descriptors for structures' presentation (structural fragments, fingerprints, constitutional, topological, electro-topological, quantum-chemical and physicochemical descriptors);
(3) Machine learning methods (multiple linear regressions, neural networks, support vector machine, random forest, similarity etc.) for identification of the relationship between descriptors, which are traditionally used as independent variables, and biological activity.

(Q)SAR models created on the basis of heterogeneous data are considered as global models with wide applicability domain and may be used for virtual screening, prediction of biological activity and target fishing. (Q)SAR models created on the basis of homogeneous data are called local models. They are traditionally used for optimisation of hit or lead compounds. Local 3D-QSAR models may be also used for pharmacophore generation. Pharmacophore describes a group of atoms in the molecule which is considered to be responsible for a pharmacological action.

Natural Product Reports

Dynamic Article Links ►

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

REVIEW

Table 1. Databases and Internet resources about plants, their therapeutic effects, names of phytoconstituents and their structures.

Source[Reference]	Description and URL	Freely Available	Number of species	Plant Name	Plant Image	Traditional Uses	Plant Part	Phytoconstituents	Phytoconstituents Structures	Stereochemistry	Pharmacological activities	Toxicity	Download structure(s)
A Guide to Medicinal and Aromatic Plants[10]	Information about medicinal, spice and aromatic plants: *http://www.hort.purdue.edu/newcrop/med-aro/default.html*	Y	510	Y	N	Y	Y	Y	N	N	N	N	N
AGRIS[10]	International Information System for Agricultural Sciences and Technology. Bibliographic data: *http://agris.fao.org/agris-search/index.do*	Y	ND	Y	Y	N	Y	Y	N	N	Y	N	N
Ayurvedic Medicinal Plants of Sri Lanka	Medicinal plants used in all of the traditional medicine systems in Sri Lanka and Ayurveda: *http://www.ayurvedicmedicinalplantssrilanka.org/*	Y	1635	Y	Y	Y	Y	ND	N	N	N	N	N
Botanical Dermatology Database (BoDD)[10]	Description of plants used in the treatment of dermatological diseases, medicinal use and adverse effects: *http://www.botanical-dermatology-database.info/*	Y	300	Y	N	Y	Y	ND	Y	Y	Y	Y	N
Botanical.com[10]	The electronic version of "A Modern Herbal" by Maud Grieve, published in 1931: *http://www.botanical.com/botanical/mgmh/comindx.html*	Y	800	Y	N	Y	Y	N	N	N	Y	Y	N
Chemical Abstracts Service (CAS)[10]	Collect and organize publicly disclosed chemical substance information including plant components: *http://www.cas.org*	N	ND	Y	N	N	N	Y	Y	Y	Y	Y	N
Chinese Herbal Medicine Dictionary[10]	Includes also examples of recipes and dosages of plants: *http://alternativehealing.org/chinese_herbs_dictionary.htm*	Y	~900	Y	N	Y	Y	Y	N	N	Y	Y	N
ClinicalTrials.gov[11]	Database of publicly and privately supported clinical studies of human participants including studies of plant extracts: *http://clinicaltrials.gov/*	Y	ND	Y	N	N	Y	ND	N	N	Y	ND	N
CMKb[10]	Medicinal plant used by Australian Aborigines: *http://biolinfo.org/cmkb*	Y	456	Y	Y	Y	Y	Y	N	N	N	N	N
Cardiovascular Disease Herbal Database (CVDHD)[12]	Provides docking results between phytocomponents and 2398 target proteins, cardiovascular-related diseases, pathways and clinical biomarkers: *http://pkuxxj.pku.edu.cn/CVDHD/index.php*	Y	3518	Y	N	N	N	35230	Y	Y	N	N	Y
Database on ethno-medicinal plants	Medicinal plants and their active components that can be used for the development of new drugs: *http://www.assamphytocure.org/scien.php*	Y	80	Y	Y	Y	Y	Y	N	N	Y	N	N
Dictionary of Natural Products (DNP)[10]	Major commercial source of chemical information on natural products: *http://dnp.chemnetbase.com*	N	ND	Y	N	N	Y	210000	Y	Y	Y	Y	Y
Dr Duke's Phytochemical and Ethnobotanical Databases[13]	Provides search tools for plant selection and information on ethnobotanical use, phytochemicals and activities: *http://www.ars-grin.gov/duke*	Y	1000	Y	N	Y	Y	Y	N	N	N	Y	N
eBDB[10]	International Ethnobotany Database provides multilingual data on plants	Y	ND	Y	Y	N	Y	N	N	N	N	N	N

	from Ecuador, Peru, Kenya and Hawai'i: *http://ebdb.org*											
ECOPORT[10]	Wiki like database including ethnobotanical data: *http://ecoport.org/ep*	Y	88291	Y	Y	Y	Y	N	N	N	N	Y	N
Ethnobotany of the Peruvian Amazon[10]	Medicinal and useful plants in the Amazonian region of Perú: *http://www.biopark.org/Plants-Amazon.html*	Y	16	Y	Y	Y	Y	Y	Y	N	Y	N	N
EXTRACT Database[10]	An expert-based knowledge system' on medicinal plants *http://www.plant-medicine.com/index.asp*	Y	24	Y	Y	Y	Y	N	N	N	Y	N	N
FDA Poisonous Plant Database[10]	References in the literature describing studies of the toxic effects of plants. *http://www.accessdata.fda.gov/scripts/plantox/index.cfm*	Y	ND	Y	N	N	Y	N	N	N	Y	Y	N
FRLHT Indian Medicinal Plants database[10]	Covers natural resources used in the Indian System of Medicine, geo-distribution data, propagation and trade information: *http://envis.frlht.org/*	Y	6198	Y	Y	N	Y	N	N	N	Y	N	N
GBIF[10]	Global Biodiversity Information Facility database includes also data on medicinal plants: *http://www.gbif.org/*	Y	1454695	Y	Y	Y	Y	N	N	N	Y	Y	N
GlobinMed[10]	Data on medicinal herbs and plants from different countries including dosage and interactions with drugs and herbs: *http://www.globinmed.com*	Y	ND	Y	Y	Y	Y	Y	N	N	Y	Y	N
HerbalThink-TCM[10]	Interactive software to learn aspects of Traditional Chinese Medicine: *http://www.rmhiherbal.org/herbalthink/index.html*	N	430	Y	Y	Y	Y	Y	Y	N	Y	Y	N
Herbalist[10]	Description of principles of medicinal plant usage at appropriate diseases and data on medicinal plants: *http://www.hoptechno.com/herbmm.htm*	N	161	Y	Y	Y	Y	Y	N	N	Y	N	N
HerbMed[10]	Categorised, evidence-based resource for herbal information, with hyperlinks to clinical and scientific publications: *http://herbmed.org/*	Y	242*	Y	Y	Y	N	N	N	N	Y	Y	N
MedlinePlus: Herbs and Supplements[10]	Dietary supplements and herbal remedies, their effectiveness, dosage, drug interactions: *http://www.nlm.nih.gov/medlineplus/druginfo/herb_All.html*	Y	80	Y	Y	N	Y	N	N	N	Y	Y	N
Herbs&Auyrveda	Ayurveda plants: *http://herbsandayurveda.wordpress.com*	Y	20	Y	Y	Y	N	Y	N	N	Y	N	N
Indian-Russian Traditional Indian Medicine database	Plants used in Traditional Indian Medicine, includes pharmacological activities of plants and their phytoconstituents (experimental and predicted by PASS software): *http://ayurveda.pharmaexpert.ru/*	Y	50	Y	N	Y	Y	2100	Y	N	Y	N	Y
IBS Natural products library	Information on natural compounds and their derivatives, with samples available for biological activity screening: *http://www.ibscreen.com/*	Y	ND	N	N	N	N	45895	Y	Y	N	N	Y
KNApSAcK Core DB[14]	Metabolites related to plants, medicinal/edible plants that are related to the geographic zones: *http://kanaya.naist.jp/KNApSAcK_Family/*	Y	1432	Y	N	N	N	Y	Y	Y	Y	N	Y
MAROWINA FACTS®[10]	Natural remedies, dietary supplements, medicinal plants and herbs of Surinam: *http://www.tropilab.com/medsupp.html*	Y	43	Y	Y	Y	Y	Y	N	N	Y	N	N
MMPD[10]	Myanmar Medicinal Plant Database: *http://www.tuninst.net/MMPD/MMPD-indx.htm*	Y	100	Y	Y	Y	Y	N	N	N	Y	N	N
MPBD[10]	Medicinal Plants of Bangladesh: *http://www.mpbd.info/*	Y	900	Y	Y	Y	Y	Y	N	N	Y	N	N
NAPRALERT[10]	Database of natural products, extracts of organisms, case reports, non-clinical and clinical studies: *http://napralert.org/*	N	ND	Y	N	Y	Y	Y	N	N	Y	N	N
Native American Ethnobotany Database[10]	Plants used as drugs, foods, dyes, and more, by native Peoples of North America with links to PLANTS Database: *http://herb.umd.umich.edu/*	Y	4029	Y	N	Y	Y	N	N	N	N	Y	N
Natural Standard[10]	Systematic reviews of foods, herbs & supplements including drug interactions, dosages and clinical trials: *http://www.naturalstandard.com*	N	ND	Y	Y	Y	Y	Y	N	N	Y	Y	N

NCCAM Herbs at a Glance[10]	A series of brief fact sheets that provides basic information about specific herbs or botanicals *http://nccam.nih.gov/health/herbsataglance.htm*	Y	48	Y	Y	Y	Y	N	N	N	Y	Y	N
PLANTS Database[10]	Standardised information about the vascular plants, mosses, liverworts, hornworts, and lichens of the U.S.: *http://plants.usda.gov/java/*	Y	1049	Y	Y	N	N	N	N	N	N	Y	N
Plants for A Future (PFAF)[10]	A resource and information centre for edible and otherwise useful plants: *http://www.pfaf.org/user/default.aspx*	Y	7000	Y	Y	N	N	N	N	N	Y	Y	N
PRELUDE Medicinal Plants Database[10]	The use of plants in different traditional veterinary and human medicines in Africa: *http://www.africamuseum.be/collections/external/prelude*	Y	2357	Y	Y	Y	Y	N	N	N	Y	ND	N
Prosea[10]	Plants of South-East Asia: *http://proseanet.org/prosea/eprosea.php*	Y	6697	Y	N	Y	Y	Y	N	N	Y	Y	N
PROTA[10]	Plant Resources of Tropical Africa: *http://www.prota.org*	Y	7400	Y	Y	Y	Y	Y	N	N	N	N	N
Provisional Global Plant Checklist	Taxonomic records from 6 major floristic datasets and 7 specialised plant family datasets: *http://bgbm3.bgbm.fu-berlin.de/IOPI/GPC/query.asp*	Y	201397	Y	N	N	N	N	N	N	N	N	N
PubChem Substance Database[15]	Samples from a variety of sources including medicinal plants, and links to biological screening results: *http://www.ncbi.nlm.nih.gov/pcsubstance*	Y	ND	N	N	N	N	Y	Y	Y	Y	Y	Y
Raintree[10]	Phytochemical information, taxonomic, ethnobotanical and clinical data for plants of the Amazon Rainforest: *http://www.rain-tree.com/*	Y	251	Y	Y	Y	N	Y	N	N	Y	N	N
Richters catalog	Description of plants and their parts, which are sold: *http://www.richters.com/Web_store/web_store.cgi*	Y	1062	Y	Y	Y	Y	N	N	N	Y	Y	N
RxList Supplements[16]	Descriptions of Herbs, and Dietary Supplements, their mode of action and drug-interactions: *http://www.rxlist.com/supplements/article.htm*	Y	ND	Y	N	Y	Y	Y	N	N	Y	Y	N
SuperNatural II database[17]	A database of purchasable natural products: *http://bioinf-applied.charite.de/supernatural_new/index.php*	Y	ND	N	N	N	N	355076	Y	Y	N	N	Y
TCMID[10]	Traditional Chinese Medicine Information Database: *http://tcm.cz3.nus.edu.sg/group/tcm-id/tcmid.asp*	Y	1098	Y	N	Y	Y	9852	3D	Y	Y	Y	Y
The Plant List[10]	The accepted Latin names with links to all synonyms by which that species has been known in other databases: *http://www.theplantlist.org/*	Y	1244871	Y	N	N	N	N	N	N	N	N	N
TIPdb[18]	Database of anti-cancer, anti-platelet, and anti-tuberculosis phytochemicals from indigenous plants in Taiwan: *http://cwtung.kmu.edu.tw/tipdb/*	Y	ND	Y	N	N	Y	8856	3D	Y	Y	Y	Y
TradiMed[10]	Commercial database of plants with symptom(s), efficacy, target organ(s), property, safety measures: *http://www.tradimed.com/*	N	502	Y	Y	Y	Y	20012	3D	Y	Y	Y	Y
TRAMEDIII[10]	South African Traditional Medicines Database *http://www.mrc.ac.za/Tramed3*	Y	ND	Y	N	Y	Y	N	N	N	N	N	N
TRAMIL[10]	Traditional Medicines in the Islands (Carrabean): *http://www.tramil.net/*	Y	365	Y	Y	Y	Y	Y	N	N	Y	Y	N
Tropicos[19]	The nomenclatural, bibliographic, and specimen data collected for the past 25 years. *http://www.tropicos.org/Home.aspx*	Y	1200000	Y	Y	N	N	N	N	N	N	N	N

Y - Yes; N - Not; ND - Not Defined; * - part of information is available for a fee.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

REVIEW

Docking is a method which predicts the preferred orientation of one molecule to another molecule when bound to each other to form a stable complex.[22] It is frequently used for prediction of the binding orientation of small drug-like molecules to their protein targets in order to estimate the affinity and predict activity of the molecule. 3D structures of targets are necessary for docking procedure. They may be extracted from PDB database (www.pdb.org) or calculated by molecular modelling methods.

Both ligand-based (when experimental data on the biological activity and structures of ligands are available) and target-based (if 3D structure of a target is available) strategies may be applied. Table 2 shows commercially and freely available software that may be used in these studies. They include completely ready tools for prediction of biological activities (e.g., INVDOCK, Selnergy, PASS, PredictFX, GUSAR); tools for docking (e.g., AutoDock, FlexX, Glide) and pharmacophore generation (e.g., Accelrys' Discovery-studio, Schrödinger's small-molecule drug discovery suite, SYBYL-X), calculation of molecular descriptors (e.g., DRAGON, ISIDA, Mol2D, PaDEL) and fingerprints (e.g., OpenBabel, PaDEL); (Q)SAR modelling (e.g., Accelrys' Discovery studio, SYBYL-X, ChemBench, KNIME, StarDrop, GUSAR).

If investigators would like to estimate the potential targets of a new phytochemical, the tools mentioned in Table 2 can be used for exploring its biological activity: (1) pair similarity with known compounds (e.g., Tanimoto coefficient calculation based on fingerprints – implemented in ChEMBLdb)[23]; (2) docking with the set of proteins (e.g., INVDOCK[24]); (3) pharmacophore-based virtual screening[23]; (4) classification prediction of biological activity spectra based on Bayesian statistics and substructural descriptors (e.g., PASS[26,27]) or fingerprints.[28] Despite the ease of use and fast calculation of pair similarity assessment by fingerprints[29], it has a limitation in fingerprint-level description of molecules (different types of fingerprints may better describe particular classes of targets[30]) and an "activity cliff" problem (when structurally similar compounds have different activities or values of activity[31]).

The advantages of docking are the use of data only for targets and that it does not require knowledge about active compounds. Docking has limitations in the number of available 3D structures of targets and in the estimation of docking results, which is a nontrivial task for selection of scoring function. Targeted scoring functions for virtual screening were reviewed by Seifert.[32] Pharmacophore generation is a quick method for virtual screening of a large database with 3D structures of chemicals, but it is also limited to known active compounds, requires conformational sampling (which does not guarantee the sampling of biologically relevant conformers) and has rigid frames of the search which is sensitive to the initial suppositions of conformers.[33,34] The use of docking and pharmacophore approaches for virtual screening of multi-targeted ligands was discussed by Ma and co-authors.[35] Classification methods have also limitations in the number of known active compounds for the appropriate targets and in the interpretation of probabilistic assessments of interaction with a target.

Although it is considered that natural products have more preferable ADME/T (absorption, distribution, metabolism, excretion and toxicity) properties in comparison with synthetic chemicals,[36] the study of ADME/T properties for phytochemicals is also very important in drug development. The tools listed in Table 2 for creation of QSAR models to predict ADME/T properties may be used for this purpose. At the same time, there are software packages with appropriate QSAR models for the prediction of ADME/T properties for chemicals (including phytochemicals) based on their structures. Adsorption, distribution and excretion depend on physical-chemical properties and on interactions with transporter proteins and blood proteins. QSAR models for estimation of such properties are provided by Discovery Studio (Accelrys), ACD/Percepta (ACD/Labs), ADME QSAR module for StarDrop (Optibrium), PASS (interaction with protein-transporters), PreADMET, QikProp (Schrödingerand), ADMET Predictor (Simulation Plus Inc.). Metabolism of phytochemicals depends on interactions with drug-metabolising enzymes (e.g., P450 cytochromes). Software for prediction of interactions with drug-metabolising enzymes and possible metabolites is provided by Discovery Studio (Accelrys), ACD/Percepta (ACD/Labs), Meteor Nexus (Lhasa Ltd.), METAPC (Multicase Inc.), ADME QSAR module for StarDrop (Optibrium), PASS (interaction with drug-metabolising enzymes) and ADMET Predictor (Simulation Plus Inc.). Apart from the above-mentioned commercial products, there are at least two freely available web services for the prediction of possible metabolic sites for different isoforms of cytochrome P450 – SMARTCyp and RS-Web Predictor. Kirchmair and co-authors recently published a review of software and *in silico* methods for the estimation of chemical interactions with drug-metabolising enzymes and the prediction of possible metabolites.[37] For prediction of different toxicity types, such as cardio-, hepato-, and renal toxicity, as well as teratogenicity and carcinogenicity, one can use DEREK (Lhasa Ltd.), TOPKAT (Accelrys), MCASE (Multicase) or PASS. In addition to the above-mentioned software, there are publications on the prediction of side effects of drugs through estimation of their interactions with antitargets (proteins related with manifestation of side effects) using the methods of molecular docking,[38,39] pair similarity assessment,[40] Bayesian-like statistics[41] and QSAR models.[42] Prediction of LD_{50} values for phytochemicals tested on rodents is also important for the estimation of their safety. Despite the absence of QSAR models that have been specially created for the prediction of LD_{50} values for phytochemicals, there are several tools with general QSAR models that may be used for this purpose (ACD/Labs, Accelrys, GUSAR). A web service for the prediction of LD_{50} values for rats via four routes of administration[43] and interaction with a set of antitargets[42] based on GUSAR software is freely available at http://www.way2drug.com/GUSAR.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

REVIEW

Table 2. Commercial and freely available software for prediction of biological activity, docking, generation of descriptors and QSAR modelling.

Name[Reference]	FA	Description and URL
Software package with descriptor and pharmacophore generation, (Q)SAR modelling tools, docking and (Q)SAR models		
ACD/Percepta[44]	N	Prediction of ADME/T and physico-chemical properties: *http://www.acdlabs.com/products/percepta*
ADMET Predictor[45]	N	Prediction of ADME/T and physico-chemical properties: *http://www.simulations-plus.com*
CDK[46]	Y	The Chemistry Development Kit (CDK) is a Java library for structural chemo- and bioinformatics applications. It includes generation of 260 types of descriptors: *http://cdk.sourceforge.net*
ChemBench[47]	Y	Chemoinformatics research support by integrating robust model builders, generators of descriptors, property and activity predictors, virtual libraries of available chemicals with predicted biological and drug-like properties, and special tools for chemical library design: *http://chembench.mml.unc.edu*
Discovery studio[48]	N	QSAR modelling and pharmacophore generation, for data analysis and structures optimisation: *http://accelrys.com/products/discovery-studio*
GUSAR[42,43]	N	QSAR modelling, antitarget interactions and LD_{50} values prediction based on atom-centric QNA and MNA descriptors: *http://www.way2drug.com/GUSAR*
KNIME[49]	Y	Graphical workbench for the entire analysis process, including plug-ins for descriptor generation, creation of QSAR models, and work with SD files: *http://www.knime.org*
MOE[50]	N	Calculates over 600 molecular descriptors including topological indices, structural keys, E-state indices, physical properties, topological polar surface area (TPSA) and CCG's VSA descriptors. MOE includes tools for creation of QSAR/QSPR models using probabilistic methods and decision trees, PCR and PLS methods: *http://www.chemcomp.com/software-chem.htm*
Molinspiration[51]	Y	Cheminformatics software with tools supporting molecule manipulation and processing, including SMILES and SDfile conversion, normalisation of molecules, generation of tautomers, molecule fragmentation, and calculation of various molecular properties needed in QSAR, molecular modelling and drug design: *http://www.molinspiration.com*
OpenTox[52]	Y	Interoperable, standards-based framework for the support of predictive toxicology including APIs and services for compounds, datasets, features, algorithms, models, ontologies, tasks, validation, and reporting which may be combined into multiple applications satisfying a variety of different user needs: *http://www.opentox.org*
PASS[26,27]	N	Prediction of Activity Spectra for Substances (PASS) – software for creation of SAR models based on Multilevel Neighbourhoods of Atoms (MNA)descriptors and modified Bayesian algorithm. It predicts several thousand types of biological activity, including pharmacological effects, mechanisms of action, toxic and adverse effects, interaction with metabolic enzymes and transporters, influence on gene expression: *http://www.way2drug.com*
PreADMET[53]	N	Calculates more than 2,000 2D and 3D descriptors, ADME/T and drug-likeness properties prediction: *http://preadmet.bmdrc.org*
PredictFX[54]	N	QSAR modelling and simulation suite that provides prediction of off-target pharmacology, associated side effect profile and affinity profiles on 4,790 targets for drug lead compounds: *http://www.certara.com/products/molmod/predictfx*
QSARpro[55]	N	QSAR modelling including calculation of over 1,000 molecular descriptors of various classes: *http://www.vlifesciences.com/products/QSARPro/Product_QSARpro.php*
Scigress Explorer, SCIGRESS[56]	N	Molecular and QSAR modelling including generation of physico-chemical descriptors for small organic molecules, inorganics, polymers, materials systems and whole proteins. *http://www.fqs.pl/chemistry_materials_life_science/products/scigress_explorer*
Selnergy[57]	N	Combination of docking software to predict interaction energies of a ligand with a protein, database of 7000 protein structures with annotated biological properties and Greenpharma Core Database: *http://www.greenpharma.com/services/selnergy-tm*
Small-molecule drug discovery suite[58]	N	2D/3D QSAR with a large selection of fingerprint options, Shape-based screening, with or without atom properties, Ligand-based pharmacophore modelling, docking, R-group analysis: *http://www.schrodinger.com/productsuite/1*
StarDrop[59]	N	QSAR modelling, data analysis and structures optimisation, R-group analysis and ADME/T

		prediction: *http://www.optibrium.com*
SYBYL-X[60]	N	QSAR modelling, pharmacophore hypothesis generation, molecular alignment, conformational searching, ADME prediction, docking and virtual screening: *http://www.tripos.com*
T.E.S.T.[61]	Y	Estimation of toxicity values and physical properties of organic chemicals based on the molecular structure of the organic chemical entered by the user: *http://www.epa.gov/nrmrl/std/qsar/qsar.html*
On-line services for prediction of biological activities, sites of biotransformation and docking to drug-targets		
1-CLICK DOCKING	Y*	Docking to 9,871 targets or user targets: *https://mcule.com/apps/1-click-docking*
DIGEP-Pred[62]	Y	Prediction of drug-induced changes in the gene expression profile based on structural formulae of drug-like compounds: *http://www.way2drug.com/GE*
GUSAR (web-service)[42,43]	Y	Prediction of acute rodent toxicity (LD_{50} values), interaction with antitargets and ecotoxicity endpoints: *http://www.way2drug.com/GUSAR*
INVDOCK[24]	Y	Automatically searches a protein and nucleic acid 3-D structure database (this database currently covers 9000 protein and nucleic acid entries) to identify the protein, RNA or DNA molecule that the small molecule can bind to: h*ttp://bidd.nus.edu.sg/group/softwares/invdock.htm*
Osiris[63]	Y	Guides performance of Risk Assessment and integrated testing strategies on skin sensitisation, repeated dose toxicity, mutagenicity, carcinogenicity, bioconcentration factor, and aquatic toxicity: *http://osiris.simpple.com/OSIRIS-ITS/itstool.do*
PASS Online[64-66]	Y	Prediction of several thousand types of biological activity, including pharmacological effects, mechanisms of action, toxic and adverse effects, interaction with metabolic enzymes and transporters, influence on gene expression based on structural formula of chemical *http://www.way2drug.com/PASSOnline*
RS-WebPredictor[67]	Y	Prediction of cytochrome P450-mediated sites of metabolism on drug-like molecules: *http://reccr.chem.rpi.edu/Software/RS-WebPredictor*
SMARTCyp[68]	Y	Prediction of the sites in molecules that are most liable to cytochrome P450 mediated metabolism: *http://www.farma.ku.dk/smartcyp*
TarFisDock[69]	Y	Identification of drug targets from Potential Drug Target Database with docking approach: *http://www.dddc.ac.cn/tarfisdock*
Docking and pharmacophore software		
AutoDock[70]	Y	Molecular modelling simulation software including protein-ligand docking: *http://autodock.scripps.edu/*
FlexX[71]	N	Prediction of protein-ligand interactions (docking): *http://www.biosolveit.de/FlexX*
Glide[72]	N	Schrodinger's ligand-protein docking software: *http://www.schrodinger.com/productpage/14/5/21*
GOLD[73]	N	Prediction of protein-ligand interactions (docking), virtual screening, lead optimisation, and identifying the correct binding mode of active molecules: *http://www.ccdc.cam.ac.uk/Solutions/GoldSuite/Pages/GOLD.aspx*
LigandScout[74]	N	Virtual screening based on 3D chemical feature pharmacophore models: *http://www.inteligand.com/*
Molegro Virtual Docker[75]	N	Prediction of protein-ligand interactions: *http://www.molegro.com/mvd-product.php*
OEDocking[76]	N	Molecular docking tools and their associated workflows: *http://www.eyesopen.com/oedocking*
Descriptor generators		
AFGen[77]	Y	A fragment-based descriptors generator with three different types of topologies: paths, acyclic subgraphs, and arbitrary topology sub-graphs: *http://glaros.dtc.umn.edu/gkhome/afgen/overview*
CODESSA[78]	N	Calculation over 500 types of constitutional, topological, geometrical, electrostatic, thermodynamic and quantum-chemical descriptors: *http://www.codessa-pro.com*
DRAGON[79]	N	Calculation almost all types of descriptors (4885 types of descriptors in total): *http://www.talete.mi.it*
E-DRAGON[80]	N	A remote version of DRAGON: *http://www.vcclab.org*
ISIDA[81]	Y	Calculation of Substructural Molecular Fragments (SMF) and ISIDA Property-Labelled Fragments (IPLF) descriptors: *http://infochim.u-strasbg.fr/spip.php?rubrique49*
MODEL[82]	Y	Web service for calculating approximately 4000 molecular descriptors based on 3D structure of a molecule: *http://jing.cz3.nus.edu.sg/cgi-bin/model/model.cgi*
Mol2D[83]	Y	Calculation more 700 descriptors: *http://www.fda.gov/ScienceResearch/BioinformaticsTools/Mold2*
MOLGEN[84]	Y	Web service for calculating 708 arithmetical, topological and geometrical descriptors: *http://molgen.de*
OpenBabel[85]	Y	Molecular fingerprint generation and similarity searching: *http://openbabel.org*
PaDEL[86]	Y	Calculates 729 1D, 2D descriptors and 134 3D descriptors, and 10 types of fingerprints: *http://padel.nus.edu.sg/software/padeldescriptor*

Statistics and Machine learning software

MATLAB[87]	N	Interactive environment, using their own language, and includes almost all of the most commonly used mathematical methods in QSAR: *http://www.mathworks.com*
R[88]	Y	Statistical calculation and graphics creation with own language of programming. R provides a wide range of tools used for QSAR modeling (linear and nonlinear modelling, classical statistical tests, consistent analysis, classification, clustering): *http://www.r-project.org*
PSPP	Y	Free available alternative of SPSS with similar possibilities: *https://www.gnu.org/software/pspp*
SAS/STAT	N	Different regression methods, Bayesian and multivariate analyses: *http://www.sas.com*
SIMCA[89]	N	Machine learning methods for QSAR: *http://www.umetrics.com/products/simca*
SPSS Statistics	N	Linear and non-linear methods for QSAR modeling: *http://www-01.ibm.com/software/analytics/spss*
Statistica[89]	N	Data processing environment includes almost all of the most frequently used machine learning methods in QSAR: *http://www.statsoft.com/*
WEKA[90]	Y	A collection of machine learning algorithms for data analysis. Contains tools for data pre-processing, classification, regression, clustering and visualisation: *http://www.cs.waikato.ac.nz/ml/weka*

* - part of services is available for a fee; FA – freely available.

4 Applications of QSAR modelling and docking in studies of phytochemicals

A review on QSAR modelling and docking applications in studies of Chinese herbal medicines was recently published by Barlow with co-authors.[91] Here, we pay more attention to *in silico* studies of phytochemicals from plants used in traditional Indian medicine. Several papers were published in which QSAR modelling was used for studying the properties of Indian herbal medicines. Most of these studies combine QSAR modelling of the appropriate therapeutic activity with docking for revealing possible targets or mechanisms of action of the studied phytochemicals (Table 3).

QSAR and molecular docking studies were performed to explore the immunomodulatory activity of derivatives of natural coumarinolignoids isolated from the seeds of *Cleome viscose*. Immunostimulatory activity was predicted using a QSAR model, developed by forward stepwise multiple linear regression and 52 physico-chemical descriptors from Scigress Explorer. The final QSAR model included dipole moment, steric energy, amide group count, lambda max (UV-visible) and molar refractivity as descriptors. Docking studies revealed the possible binding affinity of coumarinolignoids to different immunomodulatory receptors: TLR-4, iNOS, COX-2, CD14, IKK β, CD86 and COX-1.[92,93] Similar tools and approaches were used for prediction of the anticancer activity of glycyrrhetinic acid analogues against the human lung cancer cell line A-549[94] and of the immunomodulatory/anti-inflammatory activity of gallic acid derivatives.[95] Glycyrrhetinic acid is a pentacyclic triterpenoid derivative of beta-amyrin obtained by hydrolysis of glycyrrhizic acid, found mainly in the root of *Glycyrrhiza glabra* (liquorice). The docking studies showed high binding affinity of the predicted active compounds with the lung cancer target EGFR.[94] A molecular docking of gallic acid derivatives showed that the compounds had high binding affinities for INFα-2, IL-6, and IL-4 receptors.[95]

Table 3. QSAR and docking studies of phytocomponents.

No	Compounds/ Source Reference	Effects	Descriptors	Method	Docking software	N_{dt}	N_{sdt}
1	Ursolic acid analogues / *Eucalyptus hybrid* leaves[96]	Cytotoxic activity against human lung (A-549) and CNS (SF-295) cancer cell lines	50 physico-chemical descriptors (SYBYL-X 1.3)	Forward Stepwise Multiple Linear Regressions	-	-	-
2	Coumarinolignoids / *Cleome viscose* seeds[92,93]	Immunomodulatory and anti-inflammatory activity	52 physico-chemical descriptors (Scigress Explorer)		Scigress Explorer	22	7
3						24	7
4	Triterpenoids / *Eucalyptus tereticornis* and *Gentiana kurroo*[97]	Immunomodulatory and anti-inflammatory activity				11	5
5	Polyhalogenated and ester derivative of cleomiscosin A / *Cleome viscosa*[98]	Anti-inflammatory				3	3
6	Gallic acid derivatives[95]	Immunomodulatory and anti-inflammatory activity				3	3
7	Artemisinin Derivatives / *Artemisia annua*[99]	Antimalarial				1	1
8	Withanolides / *W. Somnifera*[100]	Cytotoxicity against human breast cancer cell line (MCF7)	404 descriptors (PaDEL)	ANN based QSAR-map-model	AutoDock 4.2	1	1

ANN - Artificial Neural Network; N_{dt} – number of studied drug targets; N_{sdt} – number of selected drug targets which were interacted with phytocomponents by docking studies.

QSAR modelling based on a forward stepwise multiple linear regression and 50 physical-chemical descriptors from SYBYL-X 1.3 were used for the creation of QSAR models for the prediction of cytotoxic activity of ursolic acid analogues against human lung (A-549) and CNS (SF-295) cancer cell lines.[96] The QSAR study indicated that the LUMO energy, ring count, and solvent-accessible surface area were strongly correlated with cytotoxic activity against human lung cancer cells (A-549). Similarly, the QSAR model for cytotoxic activity against the human CNS cancer cell line (SF-295) indicated that dipole vector and solvent-accessible surface area were strongly correlated with the activity.

QSAR modelling and docking studies were performed using Scigress Explorer for study of immunomodulatory and anti-inflammatory activity for the triterpenoids ursolic acid and lupeol isolated from *Eucalyptus tereticornis* and *Gentiana kurroo*.[97] Docking results suggested that the studied triterpenoids showed immunomodulatory and anti-inflammatory activity due to high binding affinity to human receptors and enzymes: NF-kappaB p52, tumour necrosis factor (TNF-alpha), nuclear factor NF-Kappa-B p50 and cyclooxygenase-2. Previously, hepatoprotective, antigestagenic and other biological activities were also predicted for the triterpenoids of the lupane group using PASS software.[101]

Five novel polyhalogenated derivatives and an ester derivative were synthesised from cleomiscosin A methyl ether and studied by QSAR modelling and docking with Scigress Explorer.[98] QSAR modelling results showed that two compounds possessed anti-inflammatory activity comparable to or even higher than diclofenac sodium. Docking results predicted that these compounds had high binding affinity to IL6, TNF-α and IL1β.

QSAR modelling of antimalarial activity and docking to Plasmepsins (Plm-II) using Scigress Explorer was performed in a study of artemisinin derivatives from *Artemisia annua*.[99] One of the predicted active compounds was chemically synthesised and tested *in vivo* in mice infected with a multidrug-resistant strain of *Plasmodium yoelii nigeriensis*. The experiment showed antimalarial activity of the selected compound.

Cytotoxic activity against a human breast cancer cell line (MCF7) was studied for withanolides from *W. somnifera* using an artificial neural network (ANN)-based QSAR model created from 37 previously tested compounds containing androstenedione-like skeletons.[100] PaDEL descriptors (404 in total) and MATLAB were used for QSAR modelling. AutoDock 4.2 was used for docking of withanolides to aromatase (PDBID: 3EQM). The study showed that four selected compounds had promising binding affinity values with aromatase in comparison to the reference, the co-crystallised control compound androstenedione.

Lipinski's rule of five and ADME properties of the studied phytocomponents were calculated with Schrödinger (QikProp from Small-molecule drug discovery suite) software or Molinspiration & T.E.S.T software[100] in almost all of the above-mentioned publications. In some publications, the calculation of toxicity risks parameters such as mutagenicity, carcinogenicity, irritation and reproductive risk of compounds was performed by Osiris software.[93, 95,98]

Most of the authors provided the estimation of the applicability domain of the created QSAR models that simplified the evaluation of the adequacy of QSAR model applications to the tested compounds. Nevertheless, it should be noted that despite the high value of calculated internal accuracy of the models (R^2 > 0.9 and R^2_{cv} >0.8), almost all of the authors did not provide the experimental values for the tested compounds, so external validation of the actual accuracy and benefit of the models is impossible. The experimental results for the part of tested compounds were given only in the paper by Kalani and co-authors.[96] This paper allowed us to calculate R^2 and RMSE for the prediction results of the experimentally tested compounds. It appeared that R^2 and RMSE were 0.05 and 0.809, respectively (although the internal validation of QSAR model showed a high accuracy of prediction: R^2 = 0.99, R^2_{cv} >0.96 and $RMSE_{cv}$ = 0.565). These results correlate with the opinion of Golbraikh and Tropsha about the imperfection of estimation of accuracy of QSAR models based on Q^2 (or R^2_{cv}) calculated for the training sets.[102] The requirement of external validation is also stated in the OECD principles for QSAR modelling (http://www.oecd.org/env/ehs/risk-assessment/37849783.pdf).

Another disadvantage of many of these publications is the practice of modelling the values in mg/kg, whereas it is known that the use of molar units (mol/kg, mmol/kg, nmol/kg, etc.) better reflects the relationships between structures of compounds and the values of their biological activity.[103]

In all of the above-mentioned publications, the authors used docking to identify the possible targets of the studied compounds. This approach is called inverse docking or target fishing. Unfortunately, none of the authors provided direct experimental evidence to confirm the interaction of the studied compounds with the predicted targets. Further, it should be remembered that, according to Leach, the best docking programs have approximately 70% accuracy.[104]

Several studies were published in which authors used docking methods to clarify the mechanisms of action for phytocomponents with known therapeutic activity (Table 4).

Let us consider in more detail two of the studies described in Table 4. Puppala and co-authors presented the bioassay-guided isolation and structure elucidation of 1-O-galloyl-b-D-glucose (b-glucogallin), a major component from the fruit of the gooseberry (*Emblica officinalis)* that displays selective and relatively potent inhibition (IC_{50}=17 μM) of human aldehyde reductase (AKR1A1) *in vitro*.[105] Molecular modelling demonstrated that b-glucogallin was able to bind favourably in the active site.

Anti-venom activity of a polyherbal formulation from aqueous extracts of leaves and roots of *Aristolochia bracteolata Lam*., *Tylophora indica* (Burm.f.) Merrill, and *Leucas aspera S.* was evaluated *in vivo* on mice against venoms from Russell's viper and Indian cobra. It was shown that these extracts provided protection against venoms in a dose of 200 mg/kg. Docking studies confirmed the interaction of leucasin (a component of *Leucas aspera S.)* and aristolochic acid (a component of *Aristolochia bracteolata Lam.)* with phospholipase A2 type I, which is considered a target in anti-venom activity.[106]

Table 4. Docking studies of mechanisms of action for phytoconstituents.

No	Compounds/Source [Reference]	Effects	Targets	Docking software
1	b-Glucogallin / *Emblica officinalis* [105]	Anticataract, antidiabetic	Aldose Reductase	Discovery Studio, Ligand Scout
2	Leucasin and aristolochic acid / *Aristolochia bracteolata* Lam., *Tylophora indica, Leucas aspera S* [106]	Antivenom against Russell's viper and cobra venom	Type I phospholipase A2	Discovery studio and Glide
3	Capsaicin / *hot chilli* [107]	Antibiotic	NorA efflux pump	SiteMap module of Schrodinger, Extra Precision scoring function of Glide
4	Catechol alkenyls / *Semecarpus anacardium* [108]	Alzheimer's disease treatment	Acetylcholinesterase	GOLD 3.1
5	Withanolides / *Withania somnifera* [109]	Cancer treatment	Mortalin, p53, p21, Nrf2	AutoDock 4.2
6	Phytoconstituents from aqueous root extracts / Gentiana lutea [110]	Inhibition of Vascular Smooth Muscle Cell Proliferation	MEK1	AutoDock4
7	Withanolide derivatives / *Withania somnifera* [111]	Antimycobacterial	Protein kinase G	Glide
8	4b-[(4-alkyl)-1,2,3-triazol-1-yl] podophyllotoxin derivatives / *Podophyllum peltatum, Podophyllum hexandrum* [112]	Antineoplastic	ATPase domain of Topoisomerase-II	Glide
9	Saponins / *Parthenium hysterophorus* [113]	Anti-inflammatory	TNF-a	Cerius2, LigandFit, Glide

In addition to the above-mentioned individual studies, a comparison of different inverse docking strategies including GOLD, FlexX, Tarfisdock, TarSearch-X and TarSearch-M with the appropriate scoring functions was made by Hui-fang and co-authors on the data from 1,594 known drug targets covering 18 biochemical functions and eight ligands as a test set.[114] Several publications have demonstrated how the pharmacophore models and docking can be used in target fishing for compounds extracted from medicinal plants (Table 5).

In these studies, the sets of phytoconstituents from appropriate plants were evaluated for interactions with possible targets by pharmacophore models or docking. The obtained results were compared with those for known reference compounds for which interaction with a target was experimentally confirmed. If the calculated values of interaction with a target for a studied phytoconstituent exceeded those for the reference compounds, that target was considered as a target for the studied phytoconstituent.

A virtual screening through a database with structures of natural products could be realised for some drug-targets if they have been a priori selected as a reason of a certain therapeutic activity. Suhitha and co-authors chose alpha glucosidase, aldose reductase and PTP1B enzymes as anti-diabetic targets and PLA2 as an anti-inflammatory target and performed virtual screening of natural ligands for these targets.[117]

5 Bioinformatics and systems biology tools for analysis of OMICs data for plant extracts

A large amount of accumulated biomedical knowledge has led to the use of bioinformatics approaches for genomics, proteomics and metabolomics data analysis. With the help of computer analysis techniques and special software, it has become possible to analyse the existing OMICs data. Existing databases allow data mining, modelling of biochemical pathways and protein-protein interactions, and they are used for research into Traditional Medicines. Genome-wide functional screening for promising pharmacological targets is the most advanced and successful approach in the post-genomic era. The combination of OMICs technologies with robust ethnobotanical and ethnomedical studies of traditional medicines leads to synergistic and reciprocal benefits for development of new inexpensive, accessible, safe and reliable medicines and treatments.[118,119]

The fast growth of network analysis methods for finding drug targets is based on the observation that over 80% of the new drugs tend to bind targets that are connected to the network of previous drug targets.[120] Network pharmacology as the next paradigm in drug discovery was proposed by Hopkins in 2007.[121,122] It uses network analysis methods to explore the pharmaceutical action of molecules in the context of biological networks to understand the mechanisms of action and to evaluate the drug efficacy.[123]

Table 5. Examples of studies on target fishing.

No	Compounds/Source [Reference]	Selected Targets	Targets	Software
1	16 secondary metabolites isolated from the aerial parts / *Ruta graveolens L.* [115]	Acetylcholinesterase, human rhinovirus coat protein, cannabinoid receptor type-2	2208 pharmacophore models	Discovery Studio
2	Epsilon viniferins / *Vitis vinifera* [57]	9 proteins including PDE4	700 proteins	Selnergy (Sybyl 6.9, FlexX)
3	Meranzin / *Limnocitrus littoralis* [116]	COX1, COX2, PPAR gamma	400 proteins	Selnergy

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

REVIEW

The work of Gu and co-authors[124] is an example of such study for natural compounds. They docked structures of 197,000 natural products to 332 target proteins of FDA-approved drugs and then explored the properties of the natural product-target networks. It has been found that polypharmacology was greatly enriched for those compounds with large numbers of network neighbours and that they play a critical role (e.g., bottleneck) in the network.[124]

The most important bioinformatics and systems biology resources and tools needed for such studies are represented in Table 6.

An exhaustive review of contemporary methods of network and pathway analysis was presented by Csermely and co-authors.[120] A review of these tools for the study of Traditional Chinese Medicine pharmacology was published by Zhao and co-authors.[137]

Deocaris and co-authors studied the functions of drug-responsive genes using systems biology approaches and compared gene regulatory circuits in response to a herbal preparation against the responses to some of its bioactive components.[133] This approach allows the explanation of the phenotypic response in the target cells, such as tumour cells, as well as how the activities of individual components influence each other. For the *Withania somnifera*, an anti-cancer therapeutic plant of Ayurveda, bioinformatics analysis was performed with Ingenuity Pathway Analysis to explore how the identified gene targets functionally interact with each other and to gain insights from the differences in the networks that may correlate with the bioactivities of natural products.[133] This study identified the critical signal transduction pathways involved in the biological response and also suggested that the minor changes in gene expression were sufficient to evoke major responses.

Table 6. Bioinformatics and systems biology resources useful for analysis of OMICs data.

Name[Reference]	FA	Description and URL
BiologicalNetworks[125]	Y	Software platform for the creation, visualisation and analysis of biological networks. Allows the use of microarray gene expression data for pathway analysis: *http://biologicalnetworks.org*
CellDesigner[126]	Y	Software for visualisation, editing and simulation of gene-regulatory and biochemical networks: *http://www.celldesigner.org*
Cell Illustrator[127]	Y*	Software for creation, visualisation and simulation of biological pathway models based on hybrid Petri-net with extensions theory. It permits executing simulations with discrete or/and continuous elements: *http://www.cellillustrator.com*
Cytoscape[128]	Y	Open-source software platform for visualisation of biological networks and their integration with various attribute data. It has many plugins for various kinds of bioinformatics analysis: *http://www.cytoscape.org*
ConsensusPathDB[129]	Y	Database integrating protein-protein, genetic, signalling, metabolic, gene regulatory, and drug-target interactions from various sources. Provides search of interactions and network visualisation based on user-defined genes and allows the performance of some types of gene/metabolite set analysis: *http://cpdb.molgen.mpg.de*
DAVID[130]	Y	A set of tools for functional interpretation of large lists of genes derived from genomic and proteomic studies: *http://david.abcc.ncifcrf.gov*
GSEA[131]	Y	Software based on Gene Set Enrichment analysis allowing the determination of whether a defined set of genes from a microarray experiment shows statistically significant differences between two biological states: *http://www.broadinstitute.org/gsea*
GeneXplain[132]	N	Platform supporting all types of gene expression analyses including normalisation and statistical analysis of microarray data, functional and pathway analysis, identification of master network regulators and possible drug targets, and analysis of regulatory gene regions: *http://genexplain.com*
Ingenuity Pathway Analysis[133]	N	Provides various kinds of pathway and network analysis of complex OMICs data: *http://www.ingenuity.com*
MetaCore[134]	N	Integrated software for functional analysis of OMICs data. Provides pathway analysis of OMICs data, knowledge mining of the database, target and biomarker assessment, model disease pathways and investigation of causal mechanisms: *http://thomsonreuters.com/metacore*
Pathway Studio[135]	N	A biological decision support tool helping to understand disease mechanisms and predict putative functionality and target-drug interactions by analysing and visualising them in a biological context: *http://www.elsevier.com/online-tools/pathway-studio*
VANESA[136]	Y	Software for the creation, modelling, visualisation, analysis and Petri net simulation of biological networks: *http://vanesa.sf.net*

* - freely available with some restrictions; FA – freely available.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

REVIEW

Klein and co-authors used transcriptome analysis in combination with pathway-focused bioassays as a helpful approach for gaining deeper insights into the complex mechanisms of the actions of multicomponent herbal preparations in living cells.[138] These authors also used Ingenuity Pathway Analysis to identify molecular targets and pathways. They suggested several mechanisms that underlie the biological activity of the preparation Padma 28.

Lamb and co-authors introduced Connectivity Map (CMap) as a phenotypic-based drug discovery approach based on the comparison of the disease gene signature and drug-induced changes in gene expression profiles in 2006.[139] The analysis of mapping disease-specific and drug-specific gene signatures can be used for drug repositioning. It is considered that if the expression of genes that are included in a specific signature of a particular disease is inversely correlated with the expression of the appropriate genes in drug gene signatures, this drug can be used for the treatment of the disease.[140-142] The use of human disease-drug networks created by the gene expression profile similarities between the pairs of drug-drug, drug-disease and disease-disease relationships provides the possibility to identify new drug indications, potential side effects, molecular targets and biological pathways that are affected by a drug.[142,143] Disease-drug connectivity mapping has been used to reveal common mechanisms underlying both diseases and drug side effects.[144]

A CMap approach was used in several studies of natural products.[145-147] For example, Wen and co-authors used it to discover the molecular mechanisms of the traditional Chinese medicinal formula Si-Wu-Tang (SWT).[148] Human breast cancer cells (MCF-7) treated with 0.1-μM estradiol or 2.56 mg/ml of SWT showed dramatic gene expression changes, but no significant change was detected for ferulic acid, a known bioactive compound of SWT. Pathway analysis using differentially expressed genes related to the treatment effect showed that expression of genes in the nuclear factor erythroid 2-related factor 2 (Nrf2) cytoprotective pathway was most significantly affected by SWT and was not affected by estradiol or ferulic acid. The gene expression profile of differentially expressed genes related to SWT treatment was compared with those of 1,309 compounds in Connectivity Map (CMAP) database. The CMAP profiles of estradiol-, withaferin A- and resveratrol-treated MCF-7 cells showed high similarity to the SWT profiles. This study identified SWT as an Nrf2 activator and phytoestrogen, suggesting its use as a nontoxic chemopreventive agent.

There are at least three freely available databases with microarray data related to drugs and diseases that may be used for the creation of disease gene signatures: NCBI Gene Expression Omnibus,[149] CMAP [139] and Comparative Toxicogenomics Database (CTD).[150] The CMap approach is limited due to its applicability only to drugs with experimentally determined drug-induced changes of gene expression, and it cannot be used for virtual screening of new drugs or new drug candidates. This limitation may be partly overcome by the prediction of possible drug-induced changes of gene expression for new natural compounds based on existing experimental microarray data. Such a possibility was recently realized on a freely available DIGEP-Pred web service[62] (http://www.way2drug.com/GE). This web service also provides the links between gene names in the predicted drug-induced changes of gene expression and the Comparative Toxicogenomics Database,[150] which simplifies the interpretation of predicted results through access to relationships of genes with diseases, side effects and biological pathways.[62]

The above-mentioned studies show that the therapeutic efficacy of medicinal plants is based on multi-target effects and that bioinformatics tools should therefore provide a system of representation to contribute to the development and enrichment of traditional herbal medicine.

6 Example of a study with combined chemo- and bioinformatics methods in drug discovery from plant natural products

The estimation of the biological action of medicinal plant extracts is a complex task requiring the involvement of chemoinformatics methods for the prediction of interactions between the phytoconstituents and drug targets, as well as knowledge about the effects associated with such interactions. These multiple interactions may lead to pleiotropic action as well as synergism for certain effects. Apart from the direct "cause-effect" relationships known from the literature (e.g., inhibition of angiotensin-converting enzyme leads to antihypertensive effect), it is possible to find indirect (hidden) relationships between the actions of drug-like compounds on drug targets and therapeutic or adverse effects through biological pathways.

Let us introduce some types of relationships between targets, pathways and effects which are used for computer formalisation and further discussion (Fig. 1):

- "Target-effect" relationships describe relations between action on a target and biological effect. For example, inhibition of cyclooxygenase 1 (COX-1) causes anti-inflammatory effect (Fig. 1).
- "Target-pathway" relationships describe relations between a target and a pathway when the target is included in the appropriate pathway. For example, COX-1 is included in "arachidonic acid metabolism" pathway of KEGG, so action on COX-1 leads to changing "arachidonic acid metabolism" (Fig. 1).
- "Pathway-effect" relationship means that there is a relation between action on a pathway and a biological effect (for example, action on "arachidonic acid metabolism" pathway related with anti-inflammatory effect (Fig. 1).
- "Target-pathway-effect" relationships describe relations between action on a target, a pathway and a biological effect (Fig. 1).

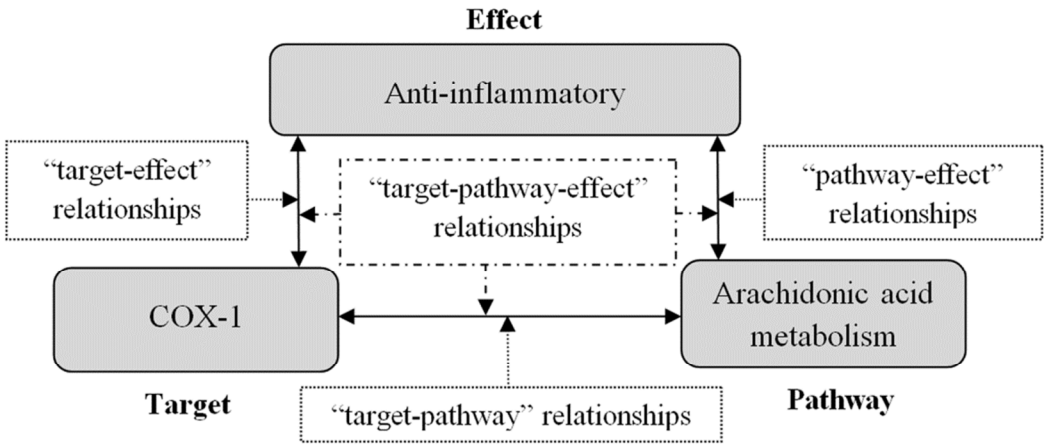

Fig. 1. Example of "target-effect", "target-pathway", "pathway-effect" and "target-pathway-effect" relationships.

The study of "cause-effect" relationships leads to better drug discovery. For example, if we know that an appropriate effect is caused by the action of a drug on an appropriate target (e.g., Target 1 at Fig. 2) and that this target belongs to a certain pathway, we may suggest that influence on this pathway is related to the observed effect. In this case, we may expect that if a new drug acts on any of the other targets (e.g., Target 2 at Fig. 2) of this pathway, it may also cause the same effect. This is quite possible when we use plant extracts because of the presence of many phytocomponents. For example, if we analyse "arachidonic acid metabolism" pathway of KEGG which contains 80 enzymes we can see that there are experimental data for compounds interacting with 42 enzymes of this pathway in ChEMBLdb 14. Some active compounds interacting with these proteins lead also to anti-inflammatory effect. So, we may expect that interaction with some other proteins may also lead to anti-inflammatory effect.

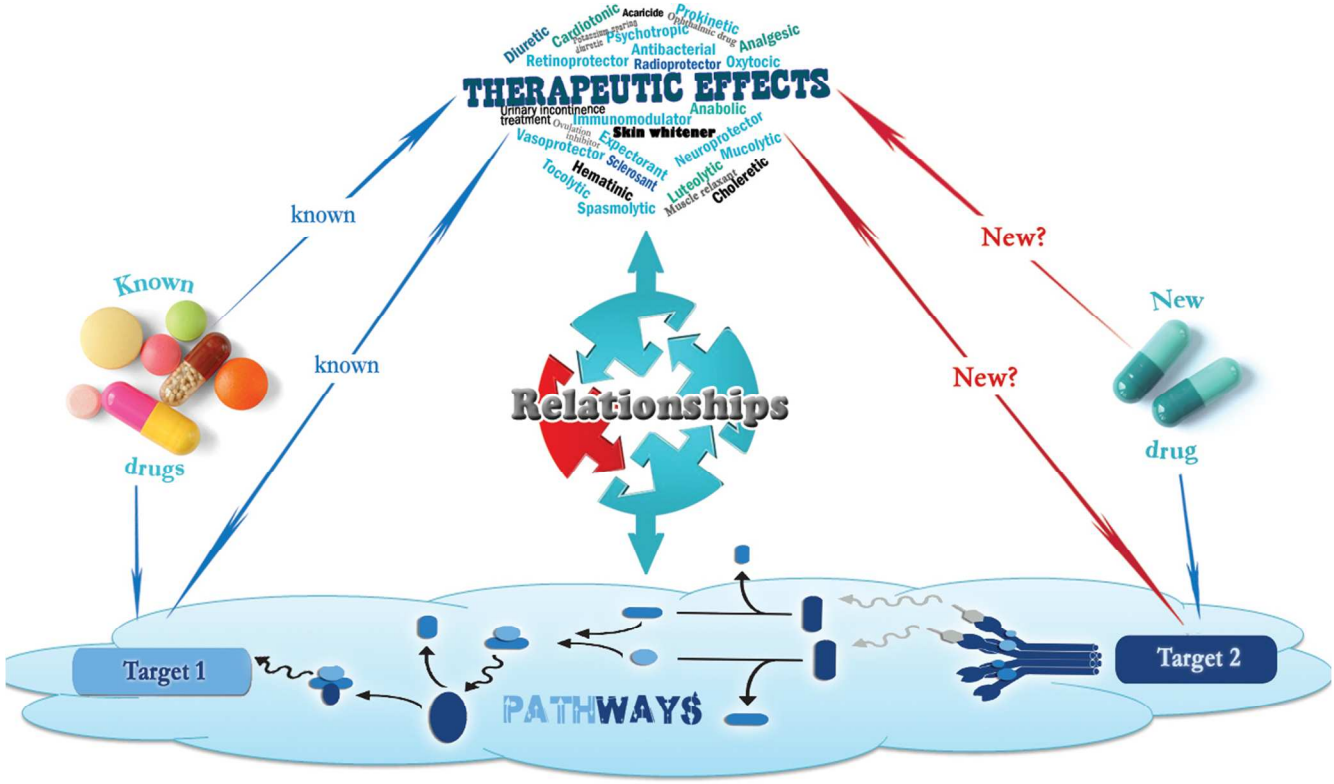

Fig. 2. General scheme of "target-pathway-effect" relationships analysis.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

REVIEW

Because there are several pathway databases, first we need to select those resources with the highest number of "target-pathway" relationships, keeping in mind the available data about the known compounds studied on interaction with the analysed targets. For this analysis, we calculated the number of "target-pathway" relationships for human proteins in twelve well-known, freely available pathway databases (Table 7). The human targets with known active compounds were extracted from the ChEMBLdb 14 database. Data from the studied pathway databases were downloaded at the beginning of 2013.

Table 7. The studied pathway databases.

Name of database	Description and URL	Number of			
		all pathways	pathways with human targets from ChEMBLdb	human targets from ChEMBLdb	"target – pathway" relationships
BioCarta	Signalling and metabolic pathways as dynamic graphic models: *http://www.biocarta.com/*	254	246	643	2915
EHMN	Human-specific metabolic pathways with data on locations of subcellular reactions: *http://www.ehmn.bioinformatics.ed.ac.uk*	70	60	429	932
HumanCyc	Human-specific metabolic pathways: *http://humancyc.org/*	297	224	402	808
INOH	Signal transduction and metabolic pathways. Uses hierarchical, event-centric data modelling with a compound graph and has its own set of literature-based ontologies for pathway annotation: *http://www.inoh.org/*	93	92	732	2848
KEGG	Signal transduction, regulatory, metabolic, disease-specific pathways and drug development pathways: *http://www.genome.jp/kegg/pathway.html*	**275**	**248**	**1877**	**8952**
NCI pathways	Human signalling and regulatory pathways: *http://pid.nci.nih.gov/*	**227**	**220**	**921**	**4536**
NetPath	Human immune and cancer signalling pathways: *http://www.netpath.org/*	32	26	461	1154
PharmGKB	Drug action pathways: *http://www.pharmgkb.org/*	103	92	544	1191
Reactome	Human signalling, regulatory and metabolic pathways. Human pathways are used to infer orthologous events in 20 non-human species: *http://www.reactome.org*	**1442**	**1202**	**1727**	**17288**
Signalink	Signalling pathways with analysis of the regulation of pathways members by scaffold and endocytosis-related proteins, miRNA, transcriptional factors and regulatory enzymes: *http://signalink.org/*	15	15	248	307
SMPDB	Small molecule pathways including human metabolic, metabolic disease, metabolite signalling and drug-action pathways with information about relevant organs, subcellular compartments, protein cofactors, protein and metabolite locations, chemical structures and protein quaternary structure: *http://www.smpdb.ca/*	411	340	470	1388
Wikipathways	Community resource of signalling, metabolic, and other pathways, in which any registered user can contribute additional pathways: *http://www.wikipathways.org*	443	335	1664	6680

Table 8. Publications with experimental confirmation of PASS prediction results for natural products.

No	Natural products[Reference]	Activity	Experimental confirmation
1	Natural products different sources[157]	Antimycobacterial	*in vitro*
2	Phytocomponents of *Nelumbo, Polygonum, Aristolochia*[158]	Disclosure of new activity	literature
3	Spirosolenol from roots of *S. anguvi*[159]	Antiinflammatory	*in vitro*
4	Violacein[160]	Antiprotozoal (Leishmania) Antiviral	literature
5	Phytocomponents of *Vitex negundo*[161]	Antioxidant, Antineoplastic	*in vitro*
6	Phytocomponents of *Ficus religiosa L.(Moraceae)*[162]	Anticonvulsant GABA aminotransferase inhibitor	*in vivo* *in vitro*
7	Quercetin[163]	Antiinflammatory, Antibacterial	*in vitro*
8	Taxol, vinblastine, vincristine, topotecan, irinotecan, etoposide, teniposide[164]	Antineoplastic	literature
9	Polyketides from marine-derived fungus *Ascochyta salicorniae*[165]	Protein phosphatase inhibitor	*in vitro*

Based on the data represented in Table 7, we selected KEGG, Reactome and NCI human pathways for further study as the most known databases with the highest number of "target – pathway" pairs for which active compounds interacting with targets are known.

Previously, we have shown the capabilities of computational methods in the prediction of therapeutic effects and mechanisms of action along with the evaluation of multi-targeted actions, possible additive/synergistic and/or antagonistic effects of natural products with PASS[151-156] and information about known direct "cause-effect" relationships generated by PharmaExpert software.[27] There are several publications where authors used freely available web service PASSOnline for prediction of biological activity spectrum for natural products with experimental confirmation of prediction results (Table 8).

In this study, we collected data on structural formulae and known biological activities for approximately 2,100 phytoconstituents along with therapeutic applications (122 therapeutic effects) of extracts of 50 plants used in Traditional Indian Medicine. All data are represented in a specially created and freely available web-resource: http://ayurveda.pharmaexpert.ru. For these phytoconstituents, we used PASS software[26,166,167] to predict probable molecular mechanisms of action (interaction with drug-targets) and biological effects based on their structural formulae. PASS (version 2012) predicted 6,400 types of biological activity including 380 therapeutic and 314 adverse effects, 3,634 mechanisms of action and 1,604 activities related to drug-induced changes in gene expression. The average accuracy calculated by a leave-one-out cross-validation procedure during the training was approximately 95%. The lists of 122 known therapeutic effects for 50 analyzed medicinal plants and 3,634 mechanisms of action that were predicted by PASS and were used in this study with data on number of active compounds and accuracy of prediction for each type of biological activity are represented in Supplement material (Table S1, S2). The PASS predicted activity spectrum is presented as a list of activities with probabilities "to be active" - Pa and "to be inactive" - Pi. Pa and Pi values vary from 0 to 1. The list of predicted activities is arranged in descending order of Pa–Pi values. Thus, the more probable types of activity are at the top of the list. If the user chooses a very high value of Pa as a cut-off for selection of probable activities, the chance to confirm the predicted activities by the experiment is also high, but many actual activities may be lost. The more detailed description of PASS approach is represented in the Supplement.

The activities from PASS related to action on drug targets were linked to the names of pathways from KEGG, Reactome and NCI human pathways databases in PharmaExpert software earlier developed for the analysis of PASS prediction results based on published knowledge on "activity-activity" relationships.[167-169] PharmaExpert contained a database with information for more than 10,000 "mechanism – effect" relationships for 5,823 proteins and 1,704 effects. Based on these data, the names of pathways were related to the PASS predicted biological effects likewise as described in Fig. 1, 2. The parameters of initial and selected data with number of "target-pathway", "pathway-effect" and "target-pathway-effect" relationships for targets from different organisms (including human, mammalians, viruses, bacteria, parasites and others) given in PharmaExpert for the used pathway databases are shown in Table 9.

Table 9. "Target-pathway", "pathway-effect" and "target-pathway-effect" relationships for KEGG, Reactome and NCI human pathways databases in PharmaExpert.

Database	Number of						
	All pathways	All targets	Targets with known ligands from ChEMBL	Selected pathways	"target-pathway" relationships	"pathway-effect" relationships	"target-pathway-effect" relationships
KEGG	259	14,022	3,137	250	15,460	9,563	8,410,688
NCI pathways	227	2,541	921	220	4,536	5,210	1,100,999
Reactome	3,660	12,479	2,982	1,202	30,030	12,720	20,455,249

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

REVIEW

By comparing the columns "All targets" and "Targets with known ligands from ChEMBL" in Table 9, it is clear that only a small part of targets (on average approximately 27%) taking part in the pathways has known ligands. Nevertheless, these proteins belong to the majority of pathways KEGG, NCI pathways and Reactome databases. We generated "target-pathway", "pathway-effect" and "target-pathway-effect" relationships based on the idea shown on Fig. 1 and 2. One can see that with pathway data, the number of relationships between action on targets (protein) and possible effects is considerably increased from 10000 "mechanism-effect" relationships to millions of "target-pathway-effect" relationships. These relationships allow us, based on PASS prediction results, to estimate possible interactions of phytoconstituents with known regulatory pathways and hence to analyse possible causes for known and new pharmacological effects of medicinal plants.

To demonstrate this approach, we compared the known pharmacological effects of 50 TIM plants with PASS predicted results for their phytoconstituents. As an example, the results of the prediction of therapeutic effects of *Aloe vera* based on analysis of its phytocomponents are shown in Table 10. The analysis was made without (column Effects in Table 10) and with consideration of the information on "mechanism-effect" (column MOA in Table 10) and "target-pathway-effect" (columns KEGG, NCI pathways and Reactome in Tables 10) relationships by PharmaExpert.

Table 10. Comparison of direct PASS prediction results of known effects for phytoconstituents of *Aloe vera* with prediction of known effects through "mechanism-effect" (column MOA) and "target-pathway-effect" (columns KEGG, NCI pathways and Reactome) relationships from PharmaExpert.

		PASS Prediction at cut-off Pa>0.5					
N	**Known effects**	**Effects**	**MOA**	**KEGG**	**NCI pathways**	**Reactome**	**Any mode**
1	Antibacterial	+	+	+	+	+	+
2	Antifungal	+	+	+	+	+	+
3	Anti-inflammatory	+	+	+	+	+	+
4	Antimutagenic	+	-	-	-	-	+
5	Antioxidant	+	+	-	-	+	+
6	Antiprotozoal (Leishmania)	+	-	-	-	-	+
7	Antiulcerative	+	+	+	-	+	+
8	Cardioprotectant	+	-	+	+	+	+
9	Cytostatic	+	-	-	-	+	+
10	Cytotoxic	+	-	+	+	+	+
11	Hepatoprotectant	+	+	+	+	+	+
12	Hypoglycemic	-	+	+	+	+	+
13	Hypolipemic	-	+	+	-	+	+
14	Immunostimulant	-	-	+	+	+	+
15	Neurodegenerative diseases treatment	-	-	+	+	+	+
16	Wound-healing agent	-	+	+	+	+	+
True Positives (TP)		**11**	**9**	**12**	**10**	**14**	**16**
True Negatives (TN)		67	63	54	67	49	28
False positives (FP)		39	43	52	39	57	78
False negatives (FN)		5	7	4	6	2	0
Sensitivity, TP/(TP+FN)		0.69	0.56	0.75	0.63	0.88	1.00
Specificity, TN/(TN+FP)		0.63	0.59	0.51	0.63	0.46	0.26
Precision, TP/(TP+FP)		0.22	0.17	0.19	0.20	0.20	0.17

Effects – known therapeutic effects predicted by PASS with Pa>0.5; **MOA** – known therapeutic effects for which some mechanism of their action was predicted with Pa>0.5; **Any mode** - known therapeutic effects, which were revealed by any modes.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

REVIEW

We collected information about the structures of 31 phytocomponents of *Aloe vera*. PASS prediction was made for these structures with a threshold of Pa>0.5. The direct prediction of the known effect was considered as corresponding to known data if the effect was predicted for at least one structure. Thus, 11 of 16 known therapeutic effects of *Aloe vera* (or 69%) were correctly revealed by direct PASS prediction of these effects. The prediction of the known effect through their known mechanisms of action (MOA), i.e., with the use of "mechanism-effect" relationships, was considered as corresponding to known data if any MOA related to the effect was predicted for at least one structure. Thus, 9 of 16 known therapeutic effects of *Aloe vera* (or 56%) were estimated correctly by PASS prediction of MOAs related with these effects. Despite the lower number of predicted known effects by prediction of MOA, some of the MOA-predicted effects are different from the directly predicted effects (Hypoglycemic, Hypolipemic and Wound-healing agent).

Prediction of known effects by "target-pathway-effect" relationships analysis for the pathways databases along with the increased number of predicted known effects (e.g., 12 from 16 (75%) for KEGG and 14 from 16 (87.5%) for Reactome) provided additional explanations for unpredicted known effects by direct PASS prediction or through PASS prediction of their MOAs (Immunostimulant and Neurodegenerative diseases treatment). At the same time, there were effects that were predicted only by direct PASS prediction (Antimutagenic and Antiprotozoal (Leishmania)). This may be explained by insufficient data on their MOAs and related pathways in PharmaExpert. Summarising the results presented in Table 10 shows that all known effects were predicted by at least one method of analysis.

Similar analyses were performed for all 50 plants (Table S3 in Supplement material). The summary results of this analysis are represented in Fig. 3. They demonstrated that the number of revealed effects were higher in all modes using information about known "mechanism-effect" (MOA) or "target-pathway-effect" (KEGG, NCI pathways and Reactome) relationships. At that, precision values were also higher for these modes than for direct prediction of therapeutic effects.

Fifty TIM plants have different numbers of known phytocomponents that reflects the degree of their study. It should be noted that the predicted biological activity of a single molecule from plant extracts is not always correlate with biological activity of plant extract. The same situation is observed and for correlation between experimental determined biological activity of a single molecule and plant extract because of interactions between phytocomponents and insufficient concentration to reveal these effects. Nevertheless, despite the small number of known phytocomponents for some plants, in most cases, they are the main components that are considered to relate to the plants' therapeutic effects. Therefore, we considered that the well-known therapeutic effects of plants might be revealed by the use of the proposed approach. This is confirmed by the data presented in Fig. 3, which shows the average percentage of predicted known therapeutic effects by their direct prediction of PASS or through predicted MOA or "target-pathway-effect" relationships based on data from the pathways databases.

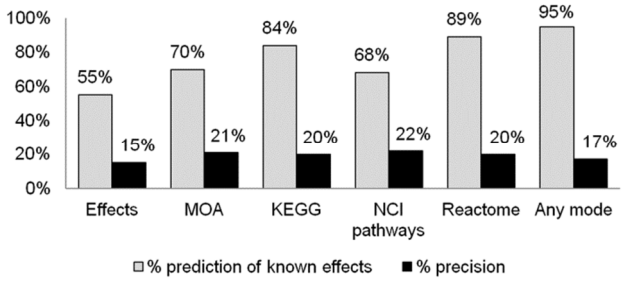

Fig. 3. Average percentages of known plants' effects predicted by different modes for phytoconstituents from 50 TIM plants.

Thus, our study shows that the number of correctly predicted known therapeutic effects of medicinal plants is considerably increased by the use of ligand-based prediction of biological activity spectra of their phytoconstituents along with the data on related mechanisms of action and "target-pathway-effect" relationships. This study may be a starting point for further *in silico* network and pathway analysis by the methods discussed in the previous part of the review for improving the identification of the true reasons for the observed therapeutic effects and reducing the false positives in the search for new therapeutic applications of plant extracts.

7 Conclusions

The use of *in silico* methods for drug discovery in natural products has increased during the previous decade. The appearance of new chemo- and bioinformatics methods, along with a growing range of OMICS data and data on phytochemical structures opens vast perspectives in the study of pharmacological activity of plant preparations. Nevertheless, scientists should consider the quality of the data and computational models used. For example, Kalliokoski and co-authors showed that standard deviations for the fitted Gaussian distributions of ChEBLdb data were $\sigma pIC_{50}=0.87$ and $\sigma pK_i=0.69$.[170] (Q)SAR models created from these data cannot be more accurate than the experimental data. The quality of publicly available databases was also discussed by Williams and Ekins.[171] The crystal structures of proteins used for docking and molecular modelling are also sometimes far removed from the native protein conformations.[172] Thus, the results of predictions given by both ligand- and target-based drug design approaches must be experimentally tested. Simultaneous use of different computational methods (consensus models) for estimation of ligand-target interactions allows decreasing variance of separate methods.[173] All individual models contain varying proportions of predictions with uncertainty, and

averaging them leads to more reliable predictions.[174,175]

The quality and reproducibility of microarray-based microRNA profiling,[176] protein-protein interactions,[177] proteomics[178] and metabolomics[179] data are sometimes poor and should be further analysed and curated.[180,181] Moreover, the number of known phytoconstituents and their structures is only a portion of the total diversity of plant phytocomponents, and many new phytoconstituents will be discovered in the future. Therefore, despite the increasing number of known phytoconstituents, not all pharmacological effects of medicinal plants may be modelled by their action on drug targets. However, even the application of currently available chemo- and bioinformatics resources and approaches provides valuable information for discovery of novel applications of medicinal plants beyond their traditional use.

8 Acknowledgments

This work was supported by RFBR/DST grant No. 11-04-92713-IND_a/RUSP-1176.

Notes

[a] *Orekhovich Institute of Biomedical Chemistry of Rus. Acad. Med. Sci., 119121, Pogodinskaya Str. 10/8, Moscow, Russia. E-mail: alexey.lagunin@ibmc.msk.ru*
[b] *Department of Pharmaceutical Sciences and Drug Research, Punjabi University, Patiala-147002, India E-mail: goelrkpup@gmail.com*
[c] *Russian National Research Medical University, Medico-biologic faculty, 117997, Ostrovitianov str. 1, Moscow, Russia. E-mail: vladimir.poroikov@ibmc.msk.ru*
† Electronic Supplementary Information (ESI) available: Description of PASS approach, Tables S1-S3. See DOI: 10.1039/b000000x/

9 References

1 H. J. Li, Y. Jiang, P. Li, *Nat. Prod. Rep.*, 2006, **23**, 735.
2 D. J. Newman and G.M. Cragg, *J. Nat. Prod.*, 2007, **70**, 461.
3 B. N. Dhawan, in *Decade of the Brain: India/USA Research in Mental Health and Neurosciences*, ed. S.H. Koslovo, M.R. Srinivasa and G.V. Coelho, National Institute of Mental Health, Rockville, MD, 1995, pp. 197-202.
4 J. M. Rollinger, D. Schuster, B. Danzl, S. Schwaiger, P. Markt, M. Schmidtke, J. Gertsch, S. Raduner, G. Wolber, T. Langer, H. Stuppner, *Planta Medica*, 2009, **75**, 195.
5 D. K. Yadav, A. Meena, A. Srivastava, D. Chanda, F. Khan, S. K. Chattopadhyay, *Drug Des. Devel. Ther.* 2010, **4**, 173.
6 G. Sakthivel, A. Dey, Kh. Nongalleima, M. Chavali, R.S. Rimal Isaac, N.S. Singh, L. Deb, *Evid Based Complement Alternat Med.*, 2013, 2013:781216. DOI: 10.1155/2013/781216.
7 D. K. Yadav, F. Khan, A. S. Negi. *J. Mol. Model.*, 2012, **18**, 2513.
8 O. Prakash, F. Khan, R. S. Sangwan, L. Misra, *Comb. Chem. High Throughput Screen.,* 2013, **16**, 57.
9 D. J. Barlow, A. Buriani, T. Ehrman, E. Bosisio, I. Eberini, P. J. Hylands, *J. Ethnopharmacol.*, 2012, **140**, 526.
10 S.S. Ningthoujam, A.D. Talukdar, K.S. Potsangbam, M.D. Choudhury, *J. Ethnopharmacol.* 2012, **141**, 9.
11 V. Sharma, I.N. Sarkar, *J. Am. Med. Inform. Assoc.* 2013, **20**, 668.
12 J. Gu, Y. Gui, L. Chen, G. Yuan, X. Xu, *J. Cheminform.* 2013, **5**, 51.
13 J. A. Duke, In *Handbook of Phytochemical Constituents of GRAS Herbs and Other Economic Plants*. CRC Press, Boca Raton. 1992, 654 pp.
14 F. M. Afendi, T. Okada, M. Yamazaki, A. Hirai-Morita, Y. Nakamura, K. Nakamura, S. Ikeda, H. Takahashi, M. Altaf-Ul-Amin, L. K. Darusman, K. Saito, S. Kanaya, *Plant Cell Physiol.*, 2012, **53**, doi: 10.1093/pcp/pcr165.
15 H. Ming, C. Tiejun, W. Yanli, B.H. Stephen, *Sci China Chem.* 2013, **56**, doi: 10.1007/s11426-013-4910-0
16 K.A. Clauson, W.A. Marsh, H.H. Polen, M.J. Seamon, B.I. Ortiz, *BMC Med. Inform. Decis. Mak.,* 2007, **7**, 7.
17 M. Dunkel, M. Fullbeck, S. Neumann, R. Preissner, *Nucleic Acids Res.*, 2006, **34**, D678.
18 Y. C. Lin, C. C. Wang, I. S. Chen, J. L. Jheng, J. H. Li, C. W. Tung, *Scientific World Journal*, 2013, 736386.
19 R. Popescu, B. Kopp, *J. Ethnopharmacol.,* 2013, **147**, 42.
20 S. Dev, *Environ. Health Perspect.,* 1999, **107**, 783.
21 H. Polur, T. Joshi, C.T. Workman, C. Lavekar, I. Kouskoumvekaki, *Mol. Inform.*, 2011, **30**, 181.
22 T. Lengauer, M. Rarey, *Curr. Opin. Struct. Biol.* 1996, **6**, 402.
23 A. Koutsoukas, R. Lowe, Y. Kalantarmotamedi, H.Y. Mussa, W. Klaffke, J.B. Mitchell, R.C. Glen, A. Bender, *J. Chem. Inf. Model.*, 2013, **53**, 1957.
24 X. Chen, C. Y. Ung, Y. Chen, *Nat. Prod. Rep.*, 2003, **20**, 432.
25 J. M. Rollinger, H. Stuppner, T. Langer, In: Progress in Drug Research. Natural Compounds as Drugs. ed. P.L. Herrling and A. Matter, Birkhauser Verlag AG, Basel, Switzerland, 2008, Vol. 65, pp. 211-250.
26 D. Filimonov, V. Poroikov In: Chemoinformatics Approaches to Virtual Screening, ed. A. Varnek and A. Tropsha, RSC Publishing, Cambridge, UK, 2008, pp. 182-216.
27 A. Lagunin, D. Filimonov, V. Poroikov, *Curr. Pharm. Des.*, 2010, **16**, 1703.
28 F. Mohd Fauzi, A. Koutsoukas, R. Lowe, K. Joshi, T. P. Fan, R. C. Glen, A. Bender, *J. Chem. Inf. Model.*, 2013, **53**, 661.
29 P. Willett In. *Chemoinformatics and Computational Chemical Biology* (Methods in Molecular Biology, vol. 672). ed. J. Bajorath, Springer Science, New York, 2011, pp. 133-158.
30 T. Kogej, O. Engkvist, N. Blomberg, S. Muresan, *J. Chem. Inf. Model.,* 2006, **46**, 1201.
31 Y. C. Martin, J. L. Kofron, L. M. Traphagen, *J. Med. Chem.*, 2002, **45**, 4350.
32 M. H. Seifert, *Drug Discov. Today*, 2009, **14**, 562.
33 D. Horvath, In. *Chemoinformatics and Computational Chemical Biology* (Methods in Molecular Biology, vol. 672). ed. J. Bajorath, Springer Science, New York, 2011, pp. 261-298.
34 G. Wolber, T. Seidel, F. Bendix, T. Langer, *Drug Discov. Today*, 2008, 13, 23.
35 X. H. Ma, Z. Shi, C. Tan, Y. Jiang, M. L. Go, B. C. Low, Y. Z. Chen, *Pharm. Res.*, 2010, **27**, 739.
36 P. Ertl, S. Roggo, A. Schuffenhauer, *J. Chem. Inf. Model.*, 2008, **48**, 68.
37 J. Kirchmair, M. J. Williamson, J. D. Tyzack, L. Tan, P. J. Bond, A. Bender, R. C. Glen, *J. Chem. Inf. Model.,* 2012, **52**, 617.
38 Z. L. Ji, Y. Wang, L. Yu, L. Y. Han, C. J. Zheng, Y. Z. Chen, *Toxicol. Lett.*, 2006, **164**, 104.
39 L. Yang, K. Wang, J. Chen, A.G. Jegga, H. Luo, L. Shi, C. Wan, X. Guo, S. Qin, G. He, G. Feng, L. He, *Plos Comput. Biol.*, 2011, **7**, e1002016.
40 E. Lounkine, M.J. Keiser, S. Whitebread, D. Mikhailov, J. Hamon, J. L. Jenkins, P. Lavan, E. Weber, A. K. Doak, S. Côté, B. K. Shoichet, L. Urban, *Nature*, 2012, **486**, 361.
41 S. M. Ivanov, A. A. Lagunin, A. V. Zakharov, D. A. Filimonov, V. V. Poroikov, *Biochem. Mosc. Suppl. Ser. B Biomed. Chem.* 2013, **7**, 40.
42 A. V. Zakharov, A. A. Lagunin, D. A. Filimonov, V. V. Poroikov, *Chem. Res. Toxicol.*, 2012, **25**, 2378.
43 A. Lagunin, A. Zakharov, D. Filimonov, V. Poroikov, *Mol. Inform.*, 2011, **30**, 241.
44 M. Janicka, M. Sztanke, K. Sztanke, *J Chromatog. A*. 2013, 1318, 92.
45 D.B. Singh, M.K. Gupta, D.V. Singh, S.K. Singh, K. Misra, *Interdiscip Sci*. 2013, **5**, 1.
46 C. Steinbeck, C. Hoppe, S. Kuhn, M. Floris, R. Guha, E. Willighagen, *Curr. Pharm. Des.*, 2006, **12**, 2111.
47 T. Walker, C.M. Grulke, D. Pozefsky, A. Tropsha, *Bioinformatics*, 2010, **26**, 3000.
48 S.H. Basha, D. Talluri, N.P. Raminni, *BMC Complement. Altern. Med.*, 2013, **13**, 85.
49 M. P. Mazanetz, R. J. Marmon, C. B. Reisser, I. Morao, *Curr. Top Med. Chem.,* 2012, **12**, 1965.

50 S. Vilar, G. Cozza, S. Moro, *Curr. Top Med. Chem.*, 2008, **8**, 1555.
51 A. Jarrahpour, J. Fathi, M. Mimouni, T.B. Hadda, J. Sheikh, Z. Chohan, A. Parvez, *Med. Chem. Research*, 2012, **21**, 1984.
52 B. Hardy, N. Douglas, C. Helma, M. Rautenberg, N. Jeliazkova, V. Jeliazkov, I. Nikolova, R. Benigni, O. Tcheremenskaia, S. Kramer, T. Girschick, F. Buchwald, J. Wicker, A. Karwath, M. Gutlein, A. Maunz, H. Sarimveis, G. Melagraki, A. Afantitis, P. Sopasakis, D. Gallagher, V. Poroikov, D. Filimonov, A. Zakharov, A. Lagunin, T. Gloriozova, S. Novikov, N. Skvortsova, D. Druzhilovsky, S. Chawla, I. Ghosh, S. Ray, H. Patel, S. Escher, *J. Cheminform*., 2010, **2**, 7.
53 S. Singh, S.K. Gupta, A. Nischal, S. Khattri, R. Nath, K.K. Pant, P.K. Seth, *Hepat Mon.* 2011, **11**, 803.
54 PredictFX description on Certara web site: http://www.certara.com/images/uploads/files/Prediction_of_affinity_target_profiles_with_PredictFX.pdf.
55 VLifeMDS: Molecular Design Suite, VLife Sciences Technologies Pvt. Ltd., Pune, India, 2010 (www.vlifesciences.com).
56 K. L. Summers, A. K. Mahrok, M. D. Dryden, M. J. Stillman, *Biochem. Biophys. Res. Commun.*, 2012, **425**, 485.
57 Q. T. Do, I. Renimel, P. Andre, C. Lugnier, C. D. Muller, P. Bernard, *Curr. Drug Discov. Technol.*, 2005, **2**, 161.
58 G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, *J. Comput. Aid. Mol. Des.,* 2013, **27**, 221.
59 I. Yusof, M. D. Segall, *Drug Discov. Today*, 2013, **18**, 659.
60 M. Mudit, M. Khanfar, G. V. Shah, K. A. Sayed, *Methods Mol. Biol.*, 2011, **716**, 55.
61 H. Zhu, T. M. Martin, D. M. Young, A. Tropsha, *Chem. Res. Toxicol.*, 2009, **22**, 1913.
62 A. Lagunin, S. Ivanov, A. Rudik, D. Filimonov, V. Poroikov, *Bioinformatics*, 2013, **29**, 2062.
63 T. Vermeire, T. Aldenberg, H. Buist, S. Escher, I. Mangelsdorf, E. Pauné, E. Rorije, D. Kroese, *Regul. Toxicol. Pharm.*, 2013, **67**, 136.
64 A. Lagunin, A. Stepanchikova, D. Filimonov, V. Poroikov, *Bioinformatics*, 2000, **16**, 747.
65 A. Sadym, A. Lagunin, D. Filimonov, V. Poroikov, *SAR QSAR Environ Res.*, 2003, **14**, 339.
66 D.A. Filimonov, A.A. Lagunin, T.A. Gloriozova, A.V. Rudik, D.S. Druzhilovskii, P.V. Pogodin, V.V. Poroikov, *Chem. Heterocycl. Compd.*, 2014, **50**, 444.
67 . J. Zaretzki, C. Bergeron, P. Rydberg, T.-W. Huang, K. P. Bennett, C. M. Breneman, *J. Chem. Inf. Model.*, 2011, **51**, 1667.
68 R. Liu, J. Liu, G. Tawa, A. Wallqvist. *J. Chem. Inf. Model.*, 2012, **52**, 1698.
69 Z. Gao, H. Li, H. Zhang, X. Liu, L. Kang, X. Luo, W. Zhu, K. Chen, X. Wang, H. Jiang, *BMC Bioinformatics*, 2008, **9**, 104.
70 G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, *J. Comput. Chem.*, 2009, **16**, 2785.
71 G. L. Warren, C. W. Andrews, A. M. Capelli, B. Clarke, J. LaLonde, M. H. Lambert, M. Lindvall, N. Nevins, S. F. Semus, S. Senger, G. Tedesco, I. D. Wall, J. M. Woolven, C. E. Peishoff, M. S. Head, *J. Med. Chem.*, 2006, 49, 5912.
72 R. Palakurti, D. Sriram, P. Yogeeswari, R. Vadrevu, *Mol. Inform.*, 2013, **32**, 385.
73 S. Thangapandian, S. John, S. Sakkiah, K. W. Lee, *J. Chem. Inf. Model.,* 2011, **51**, 33.
74 G. Wolber and T. Langer, *J. Chem. Inf. Model.*, 2005, **45**, 160.
75 R. Thomsen and M. H. Christensen, *J. Med. Chem.*, 2006, **49**, 3315.
76 M. McGann, *J. Chem. Inf. Model.* 2011, 51, 578.
77 N. Wale, I. A. Watson, G. Karypis, Proceedings LSS Comput Syst Bioinform Conference, 2007. Vol. 6, 403-416.
78 M. Randić and M. Pompe, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 631.
79 R. Todeschini and V. Consonni, *Molecular Descriptors for Chemoinformatics.* WILEY - VCH, 2009, pp. 1257.
80 I. V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V. A. Palyulin, E. V. Radchenko, N. S. Zefirov, A. S. Makarenko, V. Y. Tanchuk, V. V. Prokopenko*, J. Comput. Aid. Mol. Des.,* 2005, **19**, 453.
81 A. Varnek, D. Fourches, D. Horvath, O. Klimchuk, C. Gaudin, P. Vayer, V. Solov'ev, F. Hoonakker, I. Tetko, G. Marcou, *Curr. Comput. Aided. Drug Des.*, 2008, **4**, 191.
82 Z. R. Li, L. Y. Han, Y. Xue, C. W. Yap, H. Li, L. Jiang, Y. Z. Chen, *Biotechnol Bioeng.*, 2007, **97**, 389.
83 H. Hong, Q. Xie, W. Ge, F. Qian, F. Fang, L. Shi, Z. Su, R. Perkins, W. Tong, *J. Chem. Inf. Model.,* 2008, **48**, 1337.
84 A. Kerber, R. Laue, M. Meringer, C. Rücker, *J. Chem. Inf. Model.*, 2007, **47**, 805.
85 N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, *J. Cheminform.*, 2011, **3**, 33.
86 C. W. Yap, *J. Comput. Chem.*, 2011, **32**, 1466.
87 J. Caballero, M. Fernández, *Curr. Top. Med. Chem.*, 2008, **8**, 1580.
88 S. Stowell, *Instant R: An Introduction to R for Statistical Analysis.* Jotunheim Publishing, 2012, pp 203.
89 L. Scotti, F.J. Bezerra Mendonça Junior, D.R. Magalhaes Moreira, M.S. da Silva, I.R. Pitta, M.T. Scotti, *Curr. Top. Med. Chem.* 2012, **12**, 2785.
90 M. Hall, I. Witten, E. Frank, *Data Mining: practical Machine Learning Tools and Techniques.* Morgan Kaufmann Publishers, San Francisco, 2011, pp. 629.
91 D. J. Barlow, A. Buriani, T. Ehrman, E. Bosisio, I. Eberini, P. J. Hylands, *J. Ethnopharmacol.*, 2012, **140**, 526.
92 D. K. Yadav, A. Meena, A. Srivastava, D. Chanda, F. Khan, S. K. Chattopadhyay, *Drug Des. Devel. Ther.*, 2010, **4**, 173.
93 A. Meena, D. K. Yadav, A. Srivastava, F. Khan, D. Chanda, *Chem. Biol. Drug Des.*, 2011, **78**, 567.
94 D. K. Yadav, K. Kalani, F. Khan, S. K. Srivastava, *Med. Chem.,* 2013, **9**, 1073.
95 D. K. Yadav, F. Khan, A. S. Negi, *J. Mol. Model.,* 2012, **18**, 2513.
96 K. Kalani, D. K. Yadav, F. Khan, S. K. Srivastava, N. Suri, *J. Mol. Model.*, 2012, **18**, 3389.
97 A. Maurya, F. Khan, D. U. Bawankule, D. K. Yadav, S. K. Srivastava, *Eur. J. Pharm. Sci.,* 2012, **47**, 152.
98 S. Sharma, S. K. Chattopadhyay, D. K. Yadav, F. Khan, S. Mohanty, A. Maurya, D. U. Bawankule, *Eur. J. Pharm. Sci*., 2012, **47**, 952.
99 T. Qidwai, D. K. Yadav, F. Khan, S. Dhawan, R.S. Bhakuni, *Curr. Pharm. Des.*, 2012, **18**, 6133.
100 O. Prakash, F. Khan, R. S. Sangwan, L. Misra, *Comb. Chem. High Throughput Screen.,* 2013, **16**, 57.
101 O. B. Flekhter, L. T. Karachurina, V. V. Poroikov, L. R. Nigmatullina, L. A. Baltina, F. C. Zarudii, V. A. Davidova, L. V. Spirikhin, I. P. Baikova, F. Z. Galin, G. A. Tolstikov, *Rus. J. Bioorg. Chem.*, 2000, **26**, 192.
102 A. Golbraikh and A. Tropsha, *J. Mol. Graph. Model.*, 2002, 20, 269.
103 J. C. Dearden, M. T. Cronin, K. L. Kaiser, *SAR QSAR Environ Res.*, 2009, **20**, 241.
104 A. Leach, *Molecular Modelling: Principles and Applications (2nd Edition).* Pearson Education Limited, Harlow, 2001, pp. 667.
105 M. Puppala, J. Ponder, P. Suryanarayana, G. B. Reddy, J. M. Petrash, D. V. LaBarbera, *PLoS One*, 2012, **7**, e31399.
106 G. Sakthivel, A. Dey, Kh. Nongalleima, M. Chavali, R. S. Rimal Isaac, N. S. Singh, L. Deb, *Evid. Based Complement. Alternat. Med.,* 2013, 781216.
107 N. P. Kalia, P. Mahajan, R. Mehra, A. Nargotra, J.P. Sharma, S. Koul, I. A. Khan, *J. Antimicrob. Chemother.*, 2012, **67**, 2401.
108 H. R. Adhami, T. Linder, H. Kaehlig, D. Schuster, M. Zehl, L. Krenn, *J. Ethnopharmacol.,* 2012, **139**, 142.
109 K. Vaishnavi, N. Saxena, N. Shah, R. Singh, K. Manjunath, M. Uthayakumar, S. P. Kanaujia, S. C. Kaul, K. Sekar, R. Wadhwa, *PLoS One*, 2012, **7**, e44419.
110 R. Kesavan, U. R. Potunuru, B. Nastasijević, A. T, G. Joksić, M. Dixit, *PLoS One*, 2013, **8**, e61393.
111 N. Santhi and S. Aishwarya, *Bioinformation*, 2011, **7**, 1.
112 D. M. Reddy, J. Srinivas, G. Chashoo, A. K. Saxena, H. M. Sampath Kumar, *Eur. J. Med. Chem.,* 2011, **46**, 1983.
113 B. A. Shah, R. Chib, P. Gupta, V. K. Sethi, S. Koul, S. S. Andotra, A. Nargotra, S. Sharma, A. Pandey, S. Bani, B. Purnimad, S. C. Taneja, *Org. Biomol. Chem.*, 2009, **7**, 3230.
114 L. Hui-fang, S. Qing, Z. Jian, F. Wei, *J. Mol. Graph. Model.*, 2010, **29**, 326.
115 J. M. Rollinger, D. Schuster, B. Danzl, S. Schwaiger, P. Markt, M. Schmidtke, J. Gertsch, S. Raduner, G. Wolber, T. Langer, H. Stuppner, *Planta Medica*, 2009, **75**, 195.

116 Q.T. Do, C. Lamy, I. Renimel, N. Sauvan, P. Andre, L. Morin-allory, Ph. Bernard, *Planta Med.*, 2007, **73**, 1235.
117 S. Suhitha, K. Gunasekaran, D. Velmurugan, *Bioinformation*, 2012, **8**, 1125.
118 L.T. Ngo, J. I. Okogun, W. R. Folk, *Nat. Prod. Rep.*, 2013, **30**, 584.
119 O. Pelkonen, M. Pasanen, J. C. Lindon, K. Chan, L. Zhao, G. Deal, Q. Xu, T.P. Fan, *J. Ethnopharmacol.*, 2012, **140**, 587.
120 P. Csermely, T. Korcsmáros, H. J. Kiss, G. London, R. Nussinov, *Pharmacol. Ther*., 2013, **138**, 333.
121 A. L. Hopkins, *Nature Biotechnology*, 2007, **25**, 1110.
122 A. L.Hopkins, *Nature Chemical Biology*, 2008, **4**, 682.
123 J. Hoeng, R. Deehan, D. Pratt, F. Martin, A. Sewer, T. M. Thomson, D. A. Drubin, C. A. Waters, D. de Graaf, M. C. Peitsch, Drug Discov. Today, 2012, 17, 413.
124 J. Gu, Y. Gui, L. Chen, G. Yuan, H. Z. Lu, X. Xu, *PLoS One*, 2013, **8**, e62839.
125 S. Kozhenkov, M. Sedova, Y. Dubinina, A. Gupta, A. Ray, J. Ponomarenko, M. Baitaluk, *BMC Syst. Biol.*, 2011, **5**, 7.
126 L. Grieco, L. Calzone, I. Bernard-Pierrot, F. Radvanyi, B. Kahn-Perlès, D. Thieffry. *PLoS Comput. Biol.* 2013, **9**, e1003286.
127 M. Nagasaki, A. Saito, E. Jeong, C. Li, K. Kojima, E. Ikeda, S. Miyano, *In Silico Biol.,* 2010, **10**, 5.
128 R. Saito, M. E. Smoot, K. Ono, J. Ruscheinski, P. L. Wang, S. Lotia, A.R. Pico, G. D. Bader, T. Ideker, *Nat. Methods*, 2012, **9**, 1069.
129 A. Kamburov, U. Stelzl, H. Lehrach, R. Herwig, *Nucleic Acids Res.*, 2013, **41(Database issue)**, D793.
130 W. Huang da, B. T. Sherman, R. A. Lempicki, *Nat. Protoc*., 2009, **4**, 44.
131 A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 15545.
132 Y. Shi, F. Nikulenkov, J. Zawacka-Pankau, H. Li, R. Gabdoulline, J. Xu, S. Eriksson, E. Hedström, N. Issaeva, A. Kel, E.S. Arnér, G.Selivanova, *Cell Death Differ.* 2014, **21**, 612.
133 C. C. Deocaris, N. Widodo, R. Wadhwa, S. C. Kaul, *J. Transl. Med.,* 2008, **6**, 14.
134 P.R. Subbarayan, M. Sarkar, L. Nathanson, N. Doshi, B.L. Lokeshwar, B. Ardalan, *Evid. Based Complement. Alternat. Med*. 2013, **2013**, 471739.
135 H.M. Chen, Y.W. Lin, J.L. Wang, X. Kong, J. Hong, J.Y. Fang, *Nutr. Cancer*. 2013, **65**, 1171.
136 S. Proß, S. J. Janowski, R. Hofestädt, B. Bachman, Online proceedings of the 2012 Winter simulation Conference, IEEE, 2012.
137 J. Zhao, P. Jiang, W. Zhang, *Brief Bioinform.*, 2010, **11**, 417.
138 A. Klein, O. A. Wrulich, M. Jenny, P. Gruber, K. Becker, D. Fuchs, J. M. Gostner, F. Uberall, *BMC Genomics*, 2013, **14**, 133.
139 J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M.J. Wrobel, J. Lerner, J.-P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, T. R. Golub, *Science*, 2006, **313**, 1929.
140 J.T. Dudley, M. Sirota, M. Shenoy, R. K. Pai, S. Roedder, A. P. Chiang, A. A. Morgan, M. M. Sarwal, P. J. Pasricha, A. J. Butte, *Sci. Transl. Med.*, 2011, **3**, 96ra76.
141 M. Sirota, J. T. Dudley, J. Kim, A. P. Chiang, A. A. Morgan, A. Sweet-Cordero, J. Sage, A. J. Butte, *Sci. Transl. Med.*, 2011, **3**, 96ra77.
142 A. Gottlieb, G. Y. Stein, E. Ruppin, R. Sharan, *Mol. Syst. Biol.*, 2011, **7**, 496.
143 G. Hu and P. Agarwal, PLoS One, 2009, 4, e6536.
144 H. Huang, C. C. Liu, X. J. Zhou, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 6823.
145 S. A. Khan, A. Faisal, J. P. Mpindi, J. A. Parkkinen, T. Kalliokoski, A. Poso, O. P. Kallioniemi, K. Wennerberg, S. Kaski, *BMC Bioinformatics*, 2012, **13**, 112.
146 L. R. Aramadhaka, A. Prorock, B. Dragulev, Y. Bao, J. W. Fox, Toxicon., 2013, 69, 160.
147 L. Mak, S. Liggi, L. Tan, K. Kusonmano, J. M. Rollinger, A. Koutsoukas, R. C. Glen, J. Kirchmair, *Curr. Pharm. Des.*, 2013, 19, 532.
148 Z. Wen, Z. Wang, S. Wang, R. Ravula, L. Yang, J. Xu, C. Wang, Z. Zuo, M. S. Chow, L. Shi, Y. Huang, *PLoS One*, 2011, **6**, e18278.
149 T. Barrett, D. B. Troup, S. E. Wilhite, P. Ledoux, D. Rudnev, C. Evangelista, I. F. Kim, A. Soboleva, M. Tomashevsky, R. Edgar, *Nucleic Acids Res.*, 2007, **35(Database issue)**, D760.
150 A.P. Davis, C.G. Murphy, R. Johnson, J. M. Lay, K. Lennon-Hopkins, C. Saraceni-Richards, D. Sciaky, B.L. King, M. C. Rosenstein, T. C. Wiegers, C.J. Mattingly, *Nucleic Acids Res.*, 2013, **41**, D1104.
151 V. M. Dembitsky, T. A. Gloriozova, V. V. Poroikov, *Mini Rev. Med. Chem.*, 2005, **5**, 319.
152 V. M. Dembitsky, D. O. Levitsky, T. A. Gloriozova, V. V. Poroikov, *Nat. Prod. Commun.*, 2006, **1**, 773.
153 V. M. Dembitsky, T. A. Gloriozova, V. V.Poroikov, *Mini Rev. Med. Chem.*, 2007, **7**, 571.
154 S. B. Zotchev, A. V. Stepanchikova, A. P. Sergeyko, B. N. Sobolev, D. A. Filimonov, V. V. Poroikov, *J. Med. Chem.*, 2006, 49, 2077.
155 J. Devillers, J. C. Doré, M. Guyot, V. Poroikov, T. Gloriozova, A. Lagunin, D. Filimonov, *SAR QSAR Environ Res.*, 2007, **18**, 629.
156 R. K. Goel, D. Singh, A. Lagunin, V. Poroikov, *Med. Chem. Res.*, 2011, **20**, 1509.
157 Ch. E. Salomon, L. E. Schmidt, *Curr. Top. Med. Chem.*, 2012, **12**, 735.
158 S.N. Hade *J. Chem. Pharm. Res.*, 2012, **4**, 1925.
159 Mathew B. *Chem. Sci. J.*, 2012, **2012**, CSJ-83.
160 F. Artiguenave, A. Lins, W.D. Maciel, A.C. Junior, C. Nacif-Coelho, M.M. de Souza Linhares, G.C. de Oliveira, L.H. Barbosa, J.C. Lopes, C.N. Junior, *OMICS*, 2005, **9**, 130.
161 F.A. Kadir, N.M. Kassim, M.A. Abdulla, W.A. Yehye, *BMC Complement. Altern. Med.*, 2013, **13**, 343.
162 D. Singh, D.Y. Gawande, T. Singh, V. Poroikov, R.K. Goel. Comput. Biol. Med. 2014, 47, 1.
163 P.G.R. Chandran, S. Balaji, *Ethnobotanical Leaflets*, 2008, **12**, 245.
164 A.J. De Britto, T.L.S. Raj, and D.A. Chelliah, *Ethnobotanical Leaflets*, 2008, **12**, 801.
165 S.F. Seibert, E. Eguereva, A. Krick, S. Kehraus, E. Voloshina, G. Raabe, J. Fleischhauer, E. Leistner, M. Wiese, H. Prinz, K. Alexandrov, P. Janning, H. Waldmann, G.M. König, *Org. Biomol. Chem.*, 2006, **4**, 2233.
166 V. Poroikov, D. Filimonov, Yu. Borodina, A. Lagunin, A. Kos, *J. Chem. Inf. Comput. Sci.*, 2000, **40**, 1349.
167 A. A. Lagunin, O. A. Gomazkov, D. A. Filimonov, T. A. Gureeva, E. A. Dilakyan, E. V. Kugaevskaya, Yu. E. Elisseeva, N. I. Solovyeva, V. V. Poroikov, *J. Med. Chem.*, 2003, **46**, 3326.
168 A. A. Geronikaki, A. A. Lagunin, D. I. Hadjipavlou-Litina, Ph. T. Eleftheriou, D. A. Filimonov, V. V. Poroikov, I. Alam, A. K. Saxena, *J. Med. Chem.*, 2008, **51**, 1601.
169 N. Benaamane, B. Nedjar-Kolli, Y. Bentarzi, L. Hammal, A. Geronikaki, P. Eleftheriou, A. Lagunin, *Bioorg. Med. Chem.*, 2008, **16**, 3059.
170 T. Kalliokoski, C. Kramer, A. Vulpetti, P. Gedeck. *PLoS One*, 2013, **8**, e61007.
171 A. J. Williams and S. Ekins, *Drug Discov. Today.*, 2011, **16**, 747.
172 P. Nollert, M. D. Feese, B.L. Staker, H. Kim, in *Drug Discovery Handbook.* ed. Gad S.C. Wiley. USA, 2005, pp. 373-456.
173 H. Zhu, A. Tropsha, D. Fourches, A. Varnek, E. Papa, P. Gramatica, T. Oberg, P. Dao, A. Cherkasov, I. V. Tetko, *J. Chem. Inf. Model.*, 2008, **48**, 766.
174 J. S. Geman, E. Bienenstock, R. Doursat, *Neural Comput.*, 1992, **4**, 1.
175 A. Lagunin, D. Filimonov, A. Zakharov, W. Xie, Y. Huang, F. Zhu, T. Shen, J. Yao, V. Poroikov, *QSAR Comb. Sci.*, 2009, **28**, 806.
176 K. W. Witwer, *Clin. Chem.,* 2013, **59**, 392.
177 P. Braun, *Proteomics*, 2012, **12**, 1499.
178 R. J. Goodwin, *J. Proteomics*, 2012, **75**, 4893.
179 A. Korman, A. Oh, A. Raskind, D. Banks, *Methods Mol. Biol.*, 2012, 856, 381.
180 Y. Yang, S. J. Adelstein, A. I. Kassis, *Drug Discov. Today.*, 2012, **17**, Suppl:S16.
181 J. Casado-Vela, A. Cebrián, M. T. Gómez del Pulgar, J. C. Lacal, Clin. Transl. Oncol., 2011, 13, 617.