Reverse Pharmacognosy: Another Way to Harness the Generosity of Nature

S. Blondeau[1,2,3], Q.T. Do[1,*], T. Scior[4], P. Bernard[1] and L. Morin-Allory[2,3]

[1]*Greenpharma SAS. 3 allée du titane 45100 Orléans, France,* [2]*Institut de Chimie Organique et Analytique, ICOA, Laboratoire de Chemoinformatique, Université d'Orléans, BP 6759, 45067 Orléans Cedex 2, France,* [3]*CNRS, UMR 6005, 45067 Orléans Cedex 2, France,* [4]*Department of Pharmacy, BUAP University of Puebla, Mexico*

Abstract: A huge amount of data has been generated by decades of pharmacognosy supported by the rapid evolution of chemical, biological and computational techniques. How can we cope with this overwhelming mass of information? Reverse pharmacognosy was introduced with this aim in view. It proceeds from natural molecules to organisms that contain them *via* biological assays in order to identify an activity. *In silico* techniques and particularly inverse screening are key technologies to achieve this goal efficiently. Reverse pharmacognosy allows us to identify which molecule(s) from an organism is (are) responsible for the biological activity and the biological pathway(s) involved. An exciting outcome of this approach is that it not only provides evidence of the therapeutic properties of plants used in traditional medicine for instance, but may also position other plants containing the same active compounds for the same usage, thus increasing the curative arsenal *e.g.* development of new botanicals. This is particularly interesting in countries where western medicines are still not affordable. At the molecular level, in organisms, families of metabolites are synthesized and seldom have a single structure. Hence, when a natural compound has an interesting activity, it may be desirable to check whether there are more active and/or less toxic derivatives in organisms containing the hit – this corresponds to a kind of "natural combinatorial" chemistry. At a time when the pharmaceutical industry is lacking drug candidates in clinical trials, drug repositioning – i.e. exploiting existing knowledge for innovation – has never been so critical. Reverse pharmacognosy can contribute to addressing certain issues in current drug discovery – such as the lack of clinical candidates, toxicity - by exploiting existing data from pharmacognosy. This review will focus on recent advances in computer science applied to natural substance research that consolidate the new concept of reverse pharmacognosy.

Keywords: Reverse pharmacognosy, inverse screening, Selnergy, natural compound database, ethnopharmacology, drug repositioning.

INTRODUCTION

After decades of drug design focused on synthetic compounds, Nature is still a source of inspiration for researchers [1-5]. Combinatorial chemistry has not lived up to its goal to find robust leads [6,7]. Clinical attrition rates are still high and there has not been any drastic reduction in the drug discovery time delay. Among the factors contributing to this situation, two stand out. The first is that while new health threats have arisen due to our modern lifestyle (*e.g.* obesity, diabetes II, cardiovascular problems, new infectious disease etc.), the number of new chemical entities per year has remained unchanged over the last 25 years [8] and has even decreased during the past decade [9]. Secondly, some old threats, that were expected to disappear with progress in medicine (*e.g.* cancer, malaria, tuberculosis, etc.), still remain. Does it make sense to "oppose" synthetic and natural molecules [10], or to "disdain" traditional remedies, given that some populations rely on them? [11,12]

The need for new drugs in certain therapeutic areas *e.g.* antibiotics and oncology are obvious. Natural substances are still good sources for finding new active molecules [13,14]. Pharmacognostic field studies have led to many successes in identifying molecules of interest from living organisms such as plants [15,16]. The identification process goes through several steps: (i) organism selection - ethnopharmacology is helpful in pinpointing particular organisms used in traditional medicine; (ii) iterative activity-guided fractionation and finally (iii) active molecule(s) purification, identification and characterisation. Thanks to the development of analytical chemistry, phytochemistry, computational chemistry and information systems, the wealth of data on organisms, their traditional uses, their biological properties and the compounds they contain, open new horizons to cross-link, dissect and exploit this knowledge on a large scale in order to generate new research hypotheses to unveil Nature's hidden treasures [17]. In particular, methods to exploit traditional know-ledge based on computational techniques have demonstrated the utility of such strategies [18-21].

Following this trend, new approaches inspired by pharmacognosy have been introduced such as "reverse pharmacology" (RPL) or "reverse pharmacognosy" (RPG) [22-24]. Herein, we will define the latter concept and detail the data and tools needed to implement it. Finally, we will report applications of RPG and how it helps in identifying new leads and harnessing Nature's hidden treasures. RPG bridges the gap between traditional and modern medicines, thus enriching our knowledge.

DEFINITIONS

1. Pharmacognosy

The term pharmacognosy comes from the Greek *pharmakon* which means drug or recipe and *gnosis* which means knowledge. A simple definition could be: "Pharmacognosy is the science which studies natural compounds with therapeutic applications." [25]

Pharmacognosy is a multi-disciplinary science that covers a wide variety of disciplines: botany, pharmacology, chemistry and ethnopharmacology. As traditional medicines are mostly based on plants, most of the known natural compounds are extracted from plants. For several decades, other sources have been extensively explored such as fungi, microorganisms, marine organisms, *etc*. Only a small part of this biodiversity has been studied so far and many new original compounds remain to be discovered, especially from organisms such as insects which represent an extremely large and diverse population.

The scope of pharmacognosy is not limited to medicinal uses of natural resources but also includes applications in cosmetics, the food industry, agriculture (*e.g.* pesticides) and dyes *etc*. Moreover, various activities are linked to pharmacognosy such as the characterization of natural compounds (source localization and its concentration), the determination of chemical families, structures and their physico-chemical characteristics and of course the search for potential biological activities and therapeutic applications in drug discovery.

*Address correspondence to this author at the Greenpharma SAS. 3 allée du titane 45100 Orléans; Tel: +33 2-38-25-99-80; Fax: +33 2-38-25-99-65; E-mail: quoctuan.do@greenpharma.com

1381-6128/10 $55.00+.00

In a drug discovery strategy, pharmacognosy is generally used as shown in Fig. (**1**), from living organism(s) to molecule(s) [3,4]. The purpose is to find new bioactive molecules from a natural source.

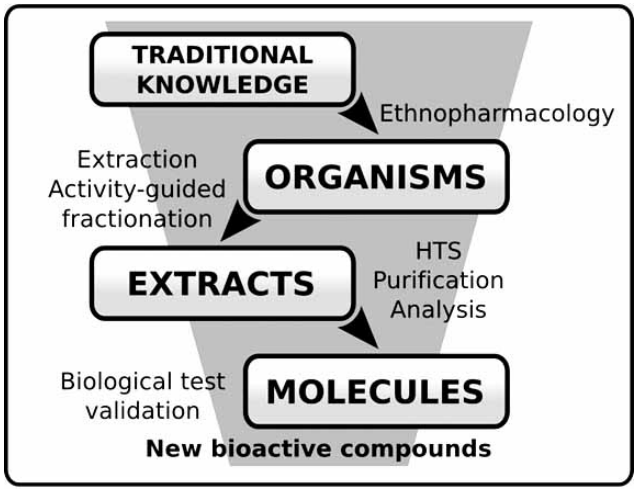


Fig. (1). Pharmacognosy process Classical pharmacognosy approach proceeds from organisms to bioactive compounds.

The first step consists in selecting organisms from traditional uses. This can be accomplished by studying traditional well documented medicines such as Traditional Chinese Medicine (TCM) or Ayurvedic medicine in India. Ethnopharmacology field studies are most helpful in oral traditions (*e.g.* African or South-American medicines) which have no written record or long standardization process. Ethnopharmacologists try to relate western medicinal concepts to traditional ones by questioning healers and shamans.

When organisms are selected from traditional uses, finding new active ingredients in the selected therapeutic areas under investigation is more likely than when using random trials [26].

In the next step, the extraction of ingredients is conducted on the selected organisms. Then activity-guided fractionation or High-Throughput Screening (HTS) retains only those extracts with the desired activity. Prior to determination of their structures by analytical chemistry *e.g.* HPLC (High-performance liquid chromatography), or NMR (Nuclear magnetic resonance), the compounds have to be identified and isolated. As a final result, active molecules are obtained from the selected organisms. Bioassays are crucial. *In vitro* assays are generally used in activity-guided fractionations followed by toxicity tests and *in vivo* validations to rule out any doubt about the alleged activity.

Pharmacognosy is a discipline that calls on several fields of expertise, as shown in Fig. (**2**). The main areas are :

- Ethnopharmacology to identify interesting organisms.
- Chemistry to identify interesting compounds by extracting, purifying and analyzing them; sometimes chemical synthesis can be a good solution to optimize a natural compound.
- Biology and biochemistry in order to evaluate biological activities, to search for targets and to understand the biological mechanisms which cause these activities.
- Pharmacology is very important to evaluate the effects of a substance, and to detect potential toxicity and adverse effects. ADME properties are also crucial for the drug development process.
- Information systems play an important role in collecting and formalizing heterogeneous data from other components of pharmacognosy into databases, in order to record knowledge, facilitate cross-analysis of data and build new hypotheses.

In order to exploit data accumulated by a century of pharmacognosy, new concepts have been introduced such as reverse pharmacognosy [22-24] and reverse pharmacology [27-30]. Before detailing reverse pharmacognosy, we will first define reverse pharmacology in order to avoid any confusion between the two concepts.

Reverse pharmacology (RPL) was established by Sir Ram Nath Chopra and Gananath Sen in the seventies and was subsequently developed by Indian scientists such as Patwardhan or Vaidya [29]. The main definition is: "*Reverse pharmacology is a rigorous scientific approach of integrating documented clinical experiences and experiential observations into leads by transdisciplinary exploratory studies and further developing these into drug candidates or formulations through robust preclinical and clinical research. In this process 'safety' remains the most important starting point and the efficacy becomes a matter of validation. The novelty of this approach is the combination of living traditional knowledge such as Ayurveda and the application of modern technology and processes to provide better and safer leads.*" [30]

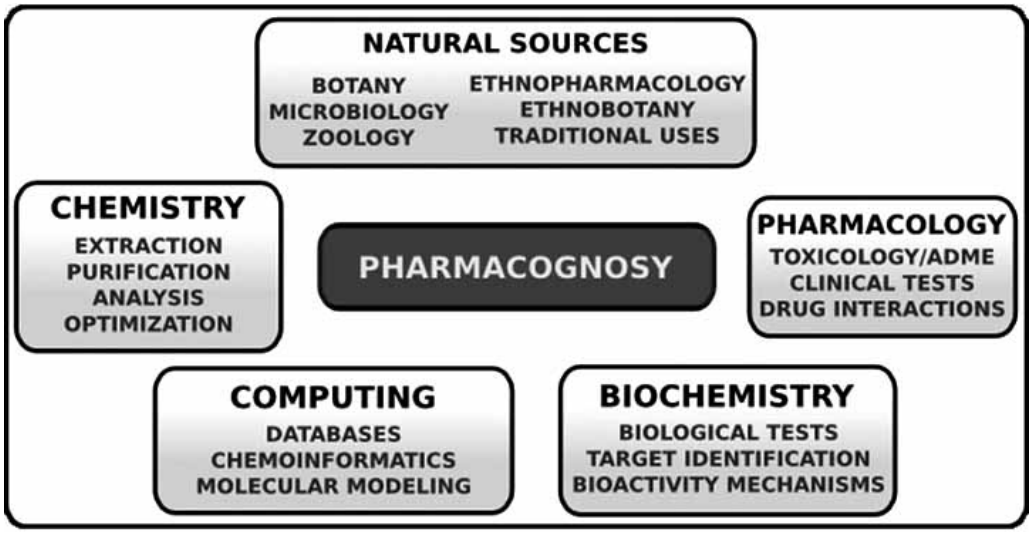


Fig. (2). Main scientific disciplines involved in pharmacognosy Classical pharmacognosy calls on several diverse fields of expertise such as ethnopharmacology (organism selection), chemistry (molecule identification) and pharmacology (biological tests).

RPL is an attractive new approach to promote the use of living organisms (mostly plants) and natural compounds by translating traditional medicine knowledge into western medicine standards. We now move on to the concept of reverse pharmacognosy.

2. Reverse Pharmacognosy

The aim of reverse pharmacognosy is to exploit the overwhelming amount of data generated by pharmacognosy. It was recently introduced and proposes to find new therapeutic activities among natural products and their sources by means of database mining and computational tools. RPG is a complementary approach to pharmacognosy which makes it possible to find applications for living organisms based on the bioactive compounds they contain and the biological properties of these compounds. Inverse screening and natural compound/source databases are essential components of RPG. Fig.(**3**) summarizes the key steps in the process.

1) Molecule Selection

RPG goes from molecules to organisms. Thus, the first step consists in selecting interesting natural compounds. Several criteria may be used. Compounds may be selected by structural criteria. For example, selections may include only molecules from the same chemical family (e.g. flavanols, triterpenoids...), and/or compounds with drug-like characteristics using derived Lipinski rules for natural compounds [31,32] and/or by selecting molecules by chemical diversity.

Another way to select these natural compounds is to consider their origins, using a natural compound/source database. For example, molecules from a particular plant could constitute a starting set of compounds. The organisms may be selected with regard to their cultivation conditions, biotopes, traditional uses or conservation status *i.e.* not on the IUCN (International Union for Conservation of Nature) Red List of endangered species [33]. Finally, economic and/or intellectual property parameters should also be considered.

2) Target Identification

The second step of RPG is the identification of targets which may bind selected compounds. Ligands can have multiple targets [34] and be involved in different metabolic pathways. Thus, all these interactions could have synergistic therapeutic effects or on the contrary, induce undesirable adverse effects.

A classical docking approach aims at finding ligand(s) for an interesting target, by screening large compound databases. In RPG, we wish to identify new biological properties for a set of pre-selected natural compounds by "inverse screening". This method tries to find protein(s) from a target database, which potentially bind(s) the interesting molecules. After the inverse screening process, each natural compound will have putative interacting protein partners and is consequently linked to related metabolic pathways. Thus, potential selectivity and/or synergy between all these ligands and/or targets may be estimated. This cannot be achieved with "classical" docking. In summary, classical docking tries to find a molecule which binds a protein target whereas inverse screening tries to find a potential target that may interact with a molecule.

3) Discovery of New Activities

Targets found in the previous step will allow molecule (re) positioning *i.e.* finding applications not yet known for the studied molecules. A target database with information on protein structures and protein biological properties is required for this purpose.

4) Biological Assays

Although improving the accuracy of virtual screening predictions is a matter of developing of new software releases, only experimental validations with *in vitro* binding tests can truly validate the predicted interaction models. As an alternative to biological assays as the first step in RPG, initial virtual screening makes it possible to focus on the most probable hits, thus decreasing the attrition rate and increasing efficiency in terms of time and cost.

5) Organisms Positioning

As biological activities can be ascribed to a certain molecule, organisms containing it – at a certain concentration – may possess the same biological property, provided that no toxicity or adverse effects are inherent to the organisms. Hence, instead of isolated molecules, we can limit our efforts to extracts to gain access to such organism-associated activities.

6) Correlating Biological Activities and Traditional Uses

Western medicine and traditional medicines use different concepts/systems to describe symptoms. For instance, if you ask a traditional medicine man to show you plants to treat inflammation, it will not make any sense to him. Instead, inquiring about plants to heal snake bites or insect stings will be meaningful for him. Bernard *et al*. [21] have shown that it is statistically more likely that a plant with anti-inflammatory properties will be found when those with a traditional use for snake bites or insect stings are selected.

Once biological properties have been attributed to an organism according to the molecules it includes, one can find "bridges" between modern medicine and folk medicines, and hence provide a scientific rationale for traditional cures.

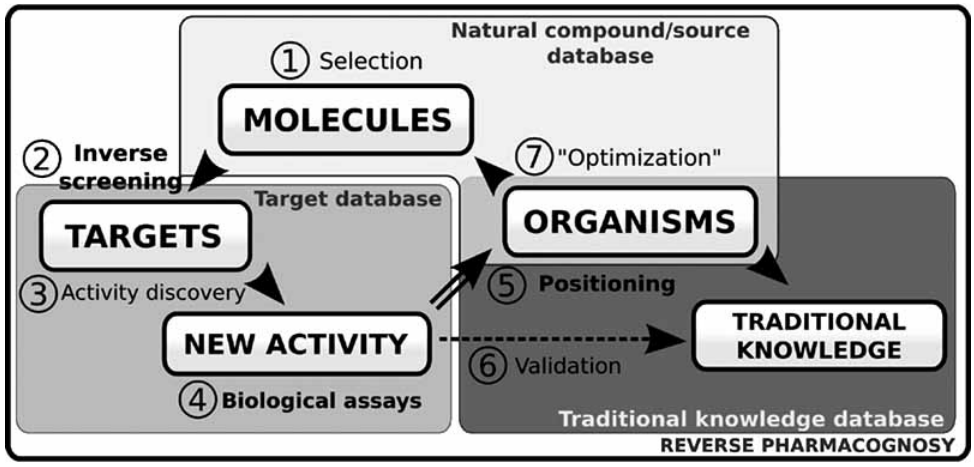


Fig. (3). Reverse pharmacognosy key steps Schematic display of the main steps in RPG from molecule selection (step 1), target identification (step 2) to activity prediction (step 3). Targets are experimentally validated (step 4), and organisms are (re)positioned (step 5) possibly with validation with traditional uses (step 6). An "activity optimization" is possible through derivative compounds search (step 7).

7) Activity Optimization

As a natural compound is a metabolite produced by living organisms, there are likely other derivatives with similar properties either in the same organism or in other organisms. RPG databases enable such derivatives to be retrieved. Among these derivatives, there may be some that are more potent, less toxic, more easily accessible or with better pharmacological profiles than the initial molecule of interest. Our novel feature of "natural combinatorial" optimization is depicted with its work-flow in Fig. (**4**).

RPG Scope

RPG may be applied for a variety of purposes:

- Active extracts: Especially in cosmetics, companies often use extracts of entire organisms (mostly plants) instead of isolated compounds. With RPG, new active extracts can easily be identified starting from an interesting compound. Active extracts are also relevant in therapeutic (*e.g.* botanicals) or nutraceutic applications.
- Traditional medicines: An estimated 65 to 80 percent of the world's population still largely uses traditional medicine [35] because of its accessibility, cost or simply its efficiency. RPG helps to find new properties for organisms, especially plants. Hence, RPG can contribute new plants or new applications of medicinal plants to extend the traditional healing arsenal. RPG and RPL have a similar approach by going back and forth between traditional and scientific knowledge and by promoting medicinal herbs coupled to scientific rationalization.
- Libraries for virtual screening in drug discovery: Natural compound or extract libraries can be generated by RPG, resulting in biofocused libraries.
- Intellectual property: An organism can be patented for a specific application. With RPG, multiple sources can be positioned for the same biological property based on the active molecules they contain, thus allowing "patent navigation". However, biopiracy problems of exploiting traditional knowledge should be carefully examined in order to equitably compensate all the actors of a valuable discovery in accordance with international treaties. RPG is less dependent on traditional knowledge than pharmacognosy because it starts from molecules and not from organisms selected *via* an ethnopharmacological process. Furthermore, once RPG has positioned an organism on a new activity, a partnership could be envisaged with the local population to develop the cultivation of the identified plant in a sustainable manner.

RPG is in need of multiple databases and a wide range of expertise. It involves new disciplines such as chemoinformatics, molecular modeling, information retrieval and management systems. Computer science is a key technology in RPG. Database and molecular modeling tools are compulsory in generating new research hypotheses from current knowledge, whether validated by experimental data or not. RPG does not replace pharmacognosy but proposes a new, original and complementary way to exploit traditional and scientific knowledge and ultimately bridges the gap between traditional knowledge and modern science.

REVERSE PHARMACOGNOSY: TOOLS AND APPLICATIONS

In the previous sections, the RPG process has been detailed. Each step needs specific tools and data types. The following section will lend insight into RPG components and their tools, in particular, Greenpharma´s reverse pharmacognosy platform.

1. General Cheminformatics Tools

They range from visualization, physico-chemical property calculation, and statistical tools, to chemical-oriented and similarity/diversity selection/searching algorithms. Some are available from open source communities, others from commercial companies. For example, OpenBabel [36], Chemistry Development Kit (CDK) [37] or ScreeningAssistant [38] are free software. They calculate physico-chemical descriptors and/or search by different criteria (substructure, similarity) *etc*. Others are proprietary software like Sybyl from Tripos Inc. [39] with the *Optisim* algorithm for building a set of diverse molecules or Chemaxon tools [40] with its popular MarvinSketcher.

2. Inverse Screening Tools and Target Database

In RPG, inverse screening is associated with a target database containing structural and activity data in such a way that potential target proteins become associated to a given natural compound of interest and its activity by what is called an inverse screening process, see Fig. (**5**).

a. Inverse Screening Process

"Classical" docking looks for molecules which may interact with an interesting protein. It has proven its efficiency through many different studies [41,42]. Three dimensional protein structures are required: they can be obtained by X-Ray crystallography, NMR or homology. Furthermore, a high resolution and the presence of a co-crystallized ligand may enhance the quality of predictions. Three dimensional structures of molecules to be docked should be computed, using specialized software such as *Concord* [43] or *Corina* [44]. Ligand - protein affinities are assessed by a scoring function that takes into account the different physico-chemical parameters, *e.g.* ionic interactions, steric hindrance, hydrogen bonding *etc*. Certain docking software packages have become more popular than others: *AutoDock* [45], *Dock* [46], *FlexX* [47], *Glide* [48], *Gold* [49], *Surflex* [50].

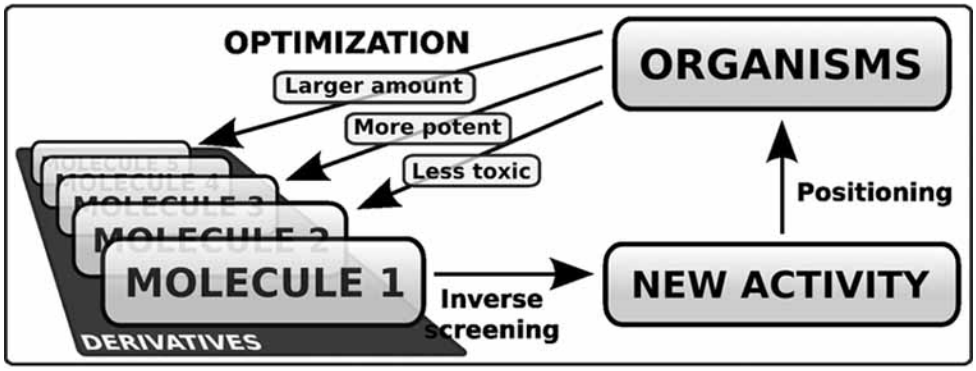


Fig. (4). The optimization step of reverse pharmacognosy Given a natural compound, its activities are predicted by inverse screening. Then in the following positioning step, organisms containing this compound or its molecular derivatives are located in association to the computed activities. The optimization step allows the search for less toxic and/or more potent representatives.

Fig. (5). Identification of new activities using inverse screening Given a compound of interest, inverse screening identifies proteins as potential target candidates. Additional bioassays are recommended to validate the predicted activities.

"Classical" screening consists of docking multiple compounds to a protein binding site. In inverse screening, the aim is to find which protein(s) can bind a particular compound. While there is no real inverse screening software available, docking tools adapted for this purpose exist, associated with a protein target database. Inverse screening software differ by their scoring function and the size of their target database. The crucial point is the scoring function which is computed for solutions containing many proteins as possible answers to docking studies with only a single compound. This is in stark contrast to "classical" docking questions where many new ligands are docked into just one receptor or enzyme protein. Ranking the solutions thereof becomes a tricky exercise. For instance, certain active site models do not discriminate between weak and strong binders, thus achieving high scores with any molecule. Several solutions have been envisaged to work around this problem such as consensus scoring *i.e.* using different types of scoring functions or their combinations to obtain a global robust one (*e.g.* Cscore [51]), MASC (Multiple active site corrections) procedure [52] or interaction fingerprints [53]. Another caveat is the size and the composition of the protein database in order to represent the underlying biological pathways satisfactorily. Recent developments in inverse screening follow in the next section.

Invdock [54]

INVDOCK, developed by the Bioinformatics and Drug Design group from the University of Singapore, is one of the first inverse screening software tools. It uses a protein structures database from the Protein DataBank (PDB) [55] and is associated to a database of target activities called TTD. INVDOCK was validated against compounds from organisms used in Traditional Chinese Medicine.

Tarfisdock [56]

Target Fishing Dock is an inverse screening platform freely accessible on the web at http://www.dddc.ac.cn/tarfisdock. Online users can directly forward a compound of interest to this interactive web site. The molecule is docked to proteins stemming from a remote database called Potential Drug Target Database. PDTD is constructed from PDB and other target databases. As a direct result, new target candidates are found (or fished) in the PDB pool, and then classified by their interaction energy score. TarFisDock is an extension to *DOCK* [46]

Selnergy [23]

Selnergy is Greenpharma's in-house inverse screening software. It is based on FlexX. It uses a validated target database of 7000 proteins of therapeutic interest. The features of Selnergy are detailed in part 4.

b. Target Databases

Such databases include two types of data: molecules and their annotations. Concerning the biomolecules, their 3D structures are required along with their names and synonyms, EC number for enzymes [57] or membership of a protein family, their therapeutic classes (*e.g.* CNS, dermatology...), the applications that may result in modulating the targets, the source organisms the proteins come from and bibliographic references. This combination of structural and non-structural data is pivotal for retrieving a subset of the proteome to be screened.

Protein 3D coordinates can be determined with different methods such as X-Ray crystallography, Nuclear Magnetic Resonance or homology modeling. RX and NMR are the best techniques to obtain suitable resolution whereas structures from homology modeling are less accurate. One major limit here is the lack of crystallized structures of some proteins and particularly of important families such as GPCR and ion channels. The annotations should contain hints on observed active sites when co-crystallized ligands are available. Such complexes serve as validation in related cases of blind docking comparing docked conformations to experimental geometries.

Certain proteins bind different ligands just as, conversely, certain ligands could bind into various active sites. Due to the flexibility of proteins [58], their spatial arrangements may change upon ligand binding. However, 3D coordinates in structure files are static and do not reflect dynamic phenomena. To deal with this important issue, it is interesting to have more than one structure for the same protein (*e.g.* co-crystallized with different types of ligands) in order to apprehend its flexibility [59] or to predict the induced fit. Some docking algorithms (*e.g.* FlexE [60]) can handle these multiple structures and compute interaction energies accordingly.

Nonstructural data provided with specific annotations about each target are found during reverse docking. When applied to an entire target database, however, this may often be time consuming and inefficient. For example, if compounds with cosmetic effects are sought, only targets that are expressed in the skin will be relevant, therefore it makes sense to focus only on them. Simulating the interaction of a ligand-protein pair can take several minutes so target filtering should not be neglected in order to gain time and above all to avoid analyzing a huge quantity of irrelevant results. Another useful filter is based on the source organism of the proteins. Thus, if the goal is to find antibacterial compounds, only bacterial targets will be selected, not plants' or mammals' ones. The protein family class (*e.g.* kinases, phosphodiesterases...) may be helpful in selectivity prediction for instance.

The biological properties linked to each target will allow the positioning of a predicted modulator for a given target to the related biological properties and consequently to the related applications. This RPG essential step of compound positioning should ultimately be validated with biological assays.

Different target databases exist and are freely available on the Internet. All of them have advantages and drawbacks considered in the light of a RPG strategy (Table **1**).

The Protein Data Bank [55, 61] at the Research Collaboratory for Structural Bioinformatics (RCSB-PDB) in the USA is the main reference database for three-dimensional structures in the field of proteins, nucleic acids and their complexes. The Web site offers tools for analysis and visualization. Structural data are stored in plain ASCII text file format with the extension dot-PDB, which is now widely used in the research community. A unique PDB code is attributed to each structure entry.

PDB structures are now only obtained with experimental methods such as X-ray crystallography, NMR or Electron Microscopy. The essential data for RPG are hence present: 3D structures can be used for inverse screening steps and non-structural data are helpful for RPG filtering and positioning processes. For historical reasons, there is a lack of information about biological activities, which is only accessible through personal examination of the literature whether referenced or not. This is the main limitation of the use of this database for RPG.

The Kyoto Encyclopedia of Genes and Genomes database [62] is another huge data repository, which contains useful information about proteins and their ligands from chemical structures to biological metabolic pathways, but also offers many search features, links to expert annotations and processed data focusing on specific issues such as genomes, species, mutations or associated diseases. KEGG has been maintained by the Bioinformatics Center of Kyoto University and the Human Genome Center at the University of Tokyo, Japan, since its creation in 1995. The main database contains several interconnected subdatabases such as KEGG PATHWAY, KEGG GENES and KEGG DISEASE or KEGG ENZYME [63]. The latter is particularly interesting for RPG with its 5000 enzymes. To each enzyme is associated a set of annotations dealing with coding genes in different species, its role in the metabolic pathways involved; products and subtracts or known ligands are some of the available data. The strength of this database is the amount of information and the dense relational links between data.

BRENDA is an acronym that stands for BRaunschweig ENzyme DAtabase [64]. Its strength is also its major drawback: an entirely enzyme-dedicated database omitting any other protein family. It was created in 1987 first in book version and then as a web database in 1998. Maintained by the Institute of Biochemistry and Bioinformatics at the Technical University of Braunschweig, Germany, this database embraces over 4800 enzyme entries as well as two data-mining tools called FRENDA and AMANDA [65].

Entries inform about nomenclature, enzyme classes, different enzymatic reactions, ligands, IC_{50} test values, host organisms and corresponding bibliographic references. Ligand - protein complexes retrieved under BRENDA can be compared to results of inverse screening for validation.

A disease section associates the enzymatic activities to their corresponding therapeutic applications and includes a bibliography for target filtering and positioning.

Like BRENDA, Therapeutic Target Database (TTD) [66] is a database focused on targets with therapeutic applications but unlike BRENDA, it includes drugs. It was created in 2001 by Dr Chen Yuzon from the Bioinformatics & Drug Design group (BIDD), a research group of the Department of Computational Science, at the National University of Singapore. In contrast to PDB-based data sets with added annotations, here general scientific literature, pharmacology books, reviews and original articles are the starting point of the collection, while links to PDB are added at the end. Focused on drugs, it delivers a vast number of ligands with references and details, *e.g*. 2D structures, CAS numbers, targets or biological properties.

BIDD has developed its own inverse screening software named INVOCK [54], which predicts protein-ligand interactions using data from the TTD and natural compounds contained in plants used in TCM.

The Potential Drug Target Database (PDTD) [67] was created in 2008 by the Drug Discovery and Design Center of Shanghai, China. PDTD belongs to the previously described inverse screening platform TarFisDock [56]. Targets included in PDTD are extracted from PDB and selected from entire and high-resolution structures, with a well-defined active site and the presence of co-crystallized ligands.

Though similar to TTD in its features, the main interest of PDTD is its integrated implementation with TarFisDock, offering some very useful tools for RPG.

The scope of this presentation was not to give an exhaustive description of existing databases but rather a selection of those that are compliant with and useful for RPG: PDB for 3D protein structures, KEGG and BRENDA for accessing biological pathways and relations, and finally, TDT and PDTD because of their similar approach with regard to ours.

Other databases containing rich and useful data on targets and ligands should also be mentioned here: sc-PDB [68], a processed PDB subset focused on druggable targets; DrugBank [69] for its contents of commercial drugs and their targets, mainly fed with FDA-approved (Food and Drug Administration) drugs and proteins; SuperTarget [70], a comprehensive drug-target relational database containing metabolic pathway data for different organisms. They are similar to those described above. For a detailed review of the open compound/activity databases, the reader is referred to the published literature [71]. The data mining tool STITCH [72] is described as "*a resource to explore known and predicted interactions of chemicals and proteins. Chemicals are linked to other chemicals and proteins by evidence derived from experiments, databases and the literature*" [73]. Though this is not a docking tool, it is a useful tool for drug positioning.

3. Natural Compounds and Traditional Knowledge Databases

(Re-) positioning an organism into therapeutic or cosmetic indications based on the molecules contained in these sources is a crucial task in RPG. It is therefore necessary to look for natural compounds and their organisms.

A traditional knowledge database includes hints on uses of organisms in different folk medicines and populations in remote areas, and can also be used to validate computed hits (cf. Fig(**3**)) by considering whether there are correlations between a traditional usage and a predicted activity.

a. Natural Compound/Source Databases

A desirable design for a natural compound/source database would enable the user to access organisms and their molecules with reference to organism names, families, genus, species and naturalist. For each hit combining molecules and organisms in a pair, information about the organ, the concentration and bibliographic references is strongly recommended. Other data could also be very useful, *e.g.* physico-chemical characteristics of molecules, known activities and targets, and 3D coordinates. For organisms, agricultural information about soil, climate, sowing, growing seasons and harvesting conditions (*e.g.* place, period, GPS – Global Positioning System – coordinates) are a priceless asset in tackling the hitherto unresolved issue of variable phyto-chemical compositions when it comes to the industrial standardization of a preparation of biological origin, irrespective of genetic variations between individuals.

Table 1. List of Target Databases Usable for a RPG Process

Database name	Accessibility	Data types	Advantages	Drawbacks
PDB [55]	Freely accessible http://www.rcsb.org/pdb	60 000 protein structures with unique PDB code	Reference database Standard PDB format	Lack of data about biological activities
KEGG ENZYME [62]	Freely accessible http://www.genome.jp/kegg	5 000 enzymes Ligands Biological pathways	Large amount and wide variety of data Dense relational links with other KEGG databases	No structural data No link between target and activities Only enzymes
BRENDA [64]	Freely accessible http://www.brenda-enzymes.org	4800 enzymes Ligands Organisms Biological activities	Very comprehensive database	Only enzymes
TTD [66]	Freely accessible http://bidd.nus.edu.sg/group/cjttd/TTD_HOME.asp	1900 targets 5000 ligands Biological pathways and activities Patents	Very useful for RPG Frequently updated	Relatively small amount of data
PDTD [67]	Freely accessible http://www.dddc.ac.cn/pdtd/index.php	1200 selected protein structures Biological activities Cross linked with other databases	Link to TarFisDock, an inverse screening platform	Relatively small amount of data
Sc-PDB [68]	Freely accessible http://bioinfo-pharma.u-strasbg.fr/scPDB	3D structures selected from PDB	Useful to enrich a target database for inverse screening	Only a subset of PDB
DrugBank [69]	Freely accessible http://www.drugbank.ca	2500 proteins 4800 drugs Pathways	Important part of FDA-approved drugs and proteins	Lack of data about biological activities
STITCH [73]	Freely accessible http://stitch.embl.de	Interaction between 74000 molecules and 2,5 million proteins Data mining	Interaction predictions Different approach	Not really a target database

b. Traditional Knowledge Database

This database should contain data from a variety of sources, ranging from knowledge collected by ethnopharmacologists in Amazonian or African tribes, to the detailed writings of TCM or Ayurvedic medicine, compiled in many books. Such information on traditional usage makes it possible for the scientist not only to select organisms but also to validate RPG organism positioning once a correlation between predicted biological activities and known uses has been established Fig. (**3**).

Due to the nonscientific nature of the information, we face several complex issues: the rationalization and standardization of contents, diverse cultures, medicinal concepts and language translations, semantics of words and local terms. Keywords have to be devised for efficient queries and data cross-over. Given the complexity, data types vary greatly. Certain data sets are specialized in ethnobotany [74] not only to exploit but also to preserve the ethnopharmacological heritage [75]. In this respect, the EbDB database [76] is an attempt to provide open access to a multi-language repository.

c. Existing Databases

Several databases are more or less suited to RPG. In this part, we will present generalist databases but there are many others, specialized in a specific medicine (*e.g.* Traditional Chinese Medicine Information Database [77]), a type of usage (*e.g.* plants used as food [78]), a biotope (*e.g.* databases about compounds from marine organisms [79,80]), geographic areas (*e.g.* Tropical plant database about plants of the Amazonian rainforest [81]), a type of chemical family [82], *etc*. Thus, all these databases should be

consulted for more specialized queries, because in these cases, they will certainly be more accurate than generalist databases. Some of these database characteristics are summarized in Table **2**. In 1971 *James Duke* created the platform hosted by the Agricultural Research Service of the U.S. Department of Agriculture, *Dr. Duke's phytochemical and ethnobotanical databases* [83]. Its contents deliver phytochemical data with a list of compounds, their plant sources, their concentrations, biological activities and bibliographic references. The usefulness of the database is unfortunately flawed by omitting molecular structures and naming without synonyms. The strength of Duke's database is the large amount of ethnobotanical data that it contains and the many different possible ways to query the database. However, access to the following fields is rather restricted: plants, compounds, acti-vities, traditional uses and references. Data are lacking about molecular properties and plant monographs describing biotopes, living conditions, crops, traditional administration and preparation modes concerning traditional medicine. Albeit time consuming, the linked references, however, enable the user to explore additional information.

The Dictionary of Natural Products (ChemNetBase) [84] (DnP) edited by Chapman & Hall, contains data on molecules, their names and synonyms, structures, characteristics and brief information on source organisms. DnP is a very comprehensive database of compounds, but unfortunately limited in organism information and molecular relations, which discourages its use in RPG. DnP is a commercial database with restricted access to the data (license or pay per view).

Napralert [85] is another relational database maintained by the University of Illinois at Chicago since 1975. The database documents organisms, natural compounds and biological activities. Access is free but only the number of hits appears on the screen. Once the fees have been paid, detailed results are sent in a summary with associated pharmacological and ethnopharmacological information and bibliographic references [86]. Napralert is a very exhaustive database and a reference in the field of natural products, but as it lacks molecular data it cannot be fully implemented in a RPG strategy.

The Plant for a future (Pfaf) database is maintained by the "*Plant for a future*" charity foundation. According to its founders: "20 species now provide 90% of our food" [87]. Starting from this fact, the Pfaf database proposes data about more than 7000 plants with reported traditional uses. Intriguingly, all the plants are quite original, seldom used and potentially edible. Furthermore, two

Table 2. List of Databases with Information on Natural Compounds, Organisms or Traditional uses, Usable for a RPG Process

Database Name	Accessibility	Data Types	Advantages	Drawbacks
Dr Duke's database [83]	Freely accessible http://www.ars-grin.gov/duke	2000 organisms 7500 molecules 2200 traditional uses Biological activities	Many ways to query the database Huge amount of data	Lack of molecule data (structures...) Not updated since 1998
ChemNetBase [84]	Searches are free, results browsing under license. http://dnp.chemnetbase.com	170 000 natural compounds	Very comprehensive Frequently updated	Lack of organism data Commercial database
Napralert [85]	Searches are free, but not results report http://www.napralert.org	200 000 publications annotated Organisms Molecules Biological activities Ethnopharmacological data	Very comprehensive Frequently updated	Lack of molecule data Commercial database Lack of flexibility in results presentation
Pfaf [87]	Freely accessible http://www.pfaf.org	7000 plants Traditional uses Medical and edible quality scores	Seldom used and original plants Highly suited to RPG	No molecule data
Supernatural [88]	Freely accessible http://bioinformatics.charite.de/supernatural	> 45 000 natural compounds Molecule characteristics Supplier data	Similarity searches	No organism data
GPDB	Greenpharma internal search	> 140 000 compounds (>80 000 naturals) > 160 000 organisms 4360 targets > 10 000 activities > 1000 traditional uses	Rich query system Structural searches Numerous links between data Suited to RPG	Lack of data but very frequently updated

scores between 1 and 5 are attributed to each plant, according to their edible and medical qualities. Pfaf is very comprehensive and the originality of the recorded plants makes this database particularly interesting for RPG.

The Greenpharma RPG platform integrates a heterogeneous database which contains target, natural compound, organism and traditional usage data. GPDB is specially designed and optimized for RPG. This database is presented in more detail in part 4.

In addition to specialized databases, many natural compound suppliers provide information on source organisms or even activities of the proposed molecules. In this respect the Supernatural database [88] can be mentioned. It is a natural compound database which collects molecules from several suppliers. Several databases such as Pubchem (http://pubchem.ncbi.nlm. nih.gov) or Chembank (http://chembank.broadinstitute.org) focus on molecules and contain a large number of natural compounds. However, there are often no related organism data, which limits their interest for RPG.

4. Greenpharma's Reverse Pharmacognosy Platform (GRPG Platform)

Reverse pharmacognosy requires several essential data types *i.e.* data on natural compounds, sources, traditional knowledge, protein targets and tools such as an inverse screening software and cheminformatics tools. These different types of data and tools are separately available online and may be used independently in a RPG approach. However in practice, it is inefficient and difficult to switch from one to another because of the lack of compatibility (*e.g.* input and output data have to be converted in multiple file formats) between tools.

The RPG concept was brought into practical existence by Greenpharma to support its research projects. A RPG-dedicated platform, see Fig. (**6**), was developed by integrating all the necessary tools. Data are standardized and can easily be used from one step to another. Data access is thus optimized and all the processes are time and cost effective. The core of this platform is a large database integrating very heterogeneous data with a web interface (intranet) which provides multiple search capabilities and numerous features for results analysis.

Thus, the GPDB contains data on natural compounds, organisms, traditional knowledge, biological targets accessible *via* a single web interface presented on Fig. (**7**). Cross-over of data is straightforward and complex queries can easily be built. For example, it is possible to search a compound by its name or by its structure and to restrain the search on some physico-chemical parameters *etc*., then display related biological properties and targets involved, identify source organisms and list their traditional uses. It is easy to hop from one molecule M1 to an organism O and display all the compounds in O in order to hop to a derivative of M1. Chemoinformatic tools, such as Openbabel [36] or Jmol [89], are included to handle molecular diversity selection, compute physico-chemical properties, calculate and display compound 3D structures or search molecules by substructures or similarity.

The other important part of the platform is Selnergy, an in-house inverse screening software. Selnergy is based on *FlexX* [47] and *Sybyl* [39] components and is linked to the target part of the GPDB. This represents about 7000 protein structures which have been mainly extracted from PDB (proteins of therapeutic interest and containing a co-crystallized ligand). A protein model is included in GPDB if Selnergy is able to simulate properly the binding mode of the co-crystallized ligand. A second validation step consists in mixing known active ligands with "decoy" molecules. Selnergy should be able to rank the active compounds on the top versus the inactive compounds according to the simulated interaction energy. Sometimes several structures are recorded for the same protein in order to take into account the flexibility of the protein. In GPDB, biological properties are linked to both targets and ligands. Upon inverse screening, activities can be inferred for docked compounds through annotated reports.

GPDB data come from 3 main sources: external databases (*i.e.* natural compound suppliers), input from the scientific literature, and internal experimental data. GPDB is constantly updated and currently contains almost 150,000 compounds, 160,000 organisms and more than 10,000 biological properties. Data mining and automatic data importing tools have been developed to facilitate GPDB updates.

Ultimately, the RPG platform will include a larger set of experimental evidence in order to enable the validation of new research ideas generated with GPDB *via* a network of partners for bioassays. Moreover, Greenpharma's analytical chemistry lab can extract, purify and characterize new natural compounds and generate physico-chemical experimental data, subsequently enriching the database.

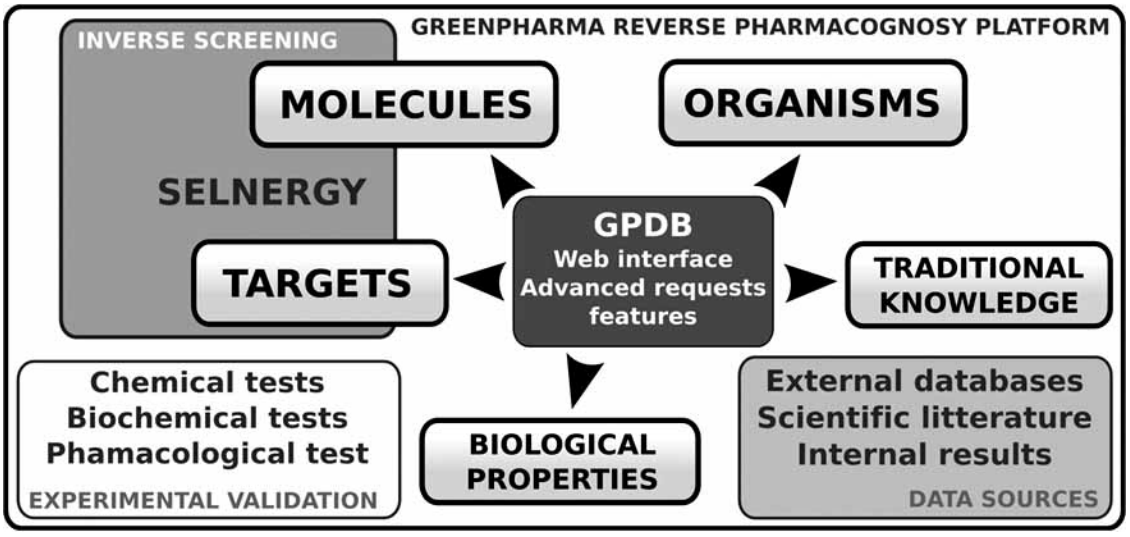


Fig. (6). The Greenpharma Reverse Pharmacognosy Platform The Greenpharma Database (GPDB) is the core of the platform. It links heterogeneous and multi-origin data required for RPG. A web interface allows advanced queries and proposes unique features to cross-reference data and efficiently exploit results. The inverse screening tool Selnergy predicts new applications for natural compounds and organisms. Activity positioning is confirmed by external validations through experiments.

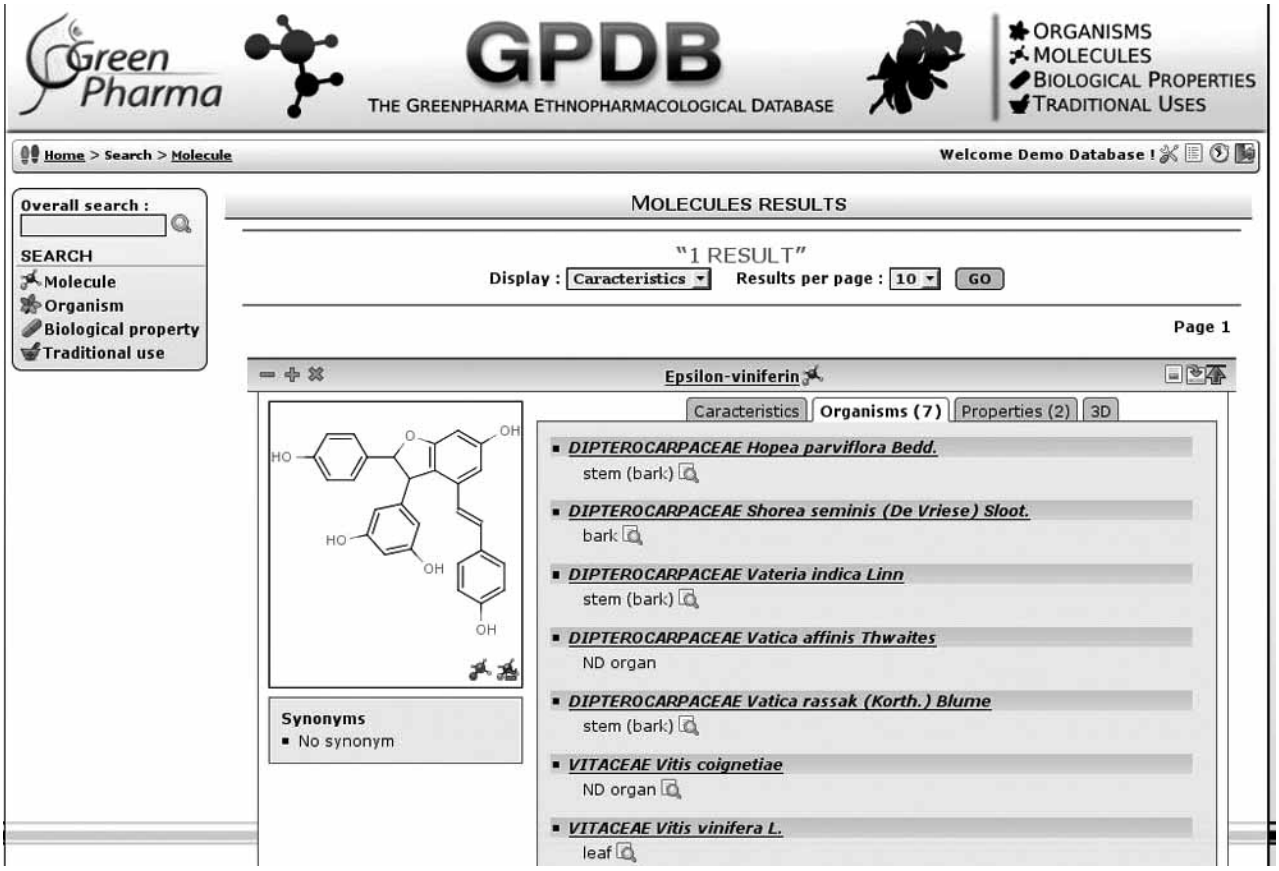


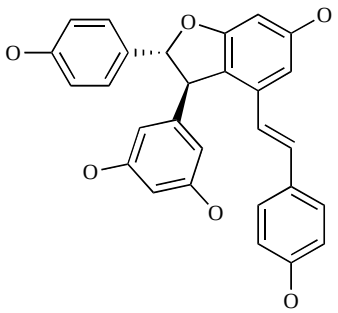
Fig. (7). GPDB interface, molecule results page for ε-viniferin

5. RPG Applications

RPG is not an abstract concept. In years of practical work, it has evolved and proven its performance through several publications. In the following we present some case studies:

a. ε-Viniferin [23]

The aim of this work was to find new cosmetic applications to ε-viniferin (**1**), a polyphenolic compound from *Vitis vinifera L.*, already known, for instance, for anti-tumor [90] or antioxidant activities [91]. First, targets with possible cosmetic activities (*i.e.* 400 proteins) were selected and an inverse screening with Selnergy was conducted. Phosphodiesterase 4 (PDE4) was selected as the best candidate with regard to interaction energies and therapeutic originality. Experimental tests were carried out. Indeed, it possesses anti-inflammatory effects, inhibiting the release of tumor necrosis factor alpha (TNF-α).



ε-viniferin (**1**)

Firstly, ε-viniferin was tested *in vitro* on different PDE bovine subtypes. Compound concentration in DMSO was 1%. At this concentration, DMSO did not affect the target activity. The IC50 was calculated from concentration-response curves with 6 different concentrations of ε-viniferin. The results in Table **3** represent the mean value of 3 assays with an experimental error around 15%. Results show an IC50 of 4.6µM and a selectivity for the PDE4 subtype (though the selectivity is not marked for PDE3). Indeed, only this PDE subtype was found by Selnergy, thus these tests confirm the prediction. Other assays on blood mononuclear cells showed a significant decrease in TNF-α release induced by ε-viniferin with an IC50 between 10 and 100µM, see Fig. (**8**). Thus, anti-inflammatory activity could be attributed to the molecule and was validated at the cell level.

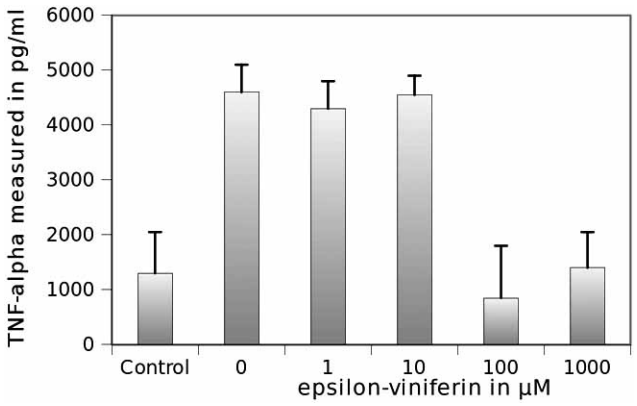


Fig. (8). Production of TNF- α measured in pg/mL. There is no effect of ε-viniferin at 1 and 10 µM. The reduction of TNF-α is significant at100µM and is not better at 1000µM. So 100µM is certainly the optimum concentration.

Table 3. *In Vitro* Inhibition Tests of ε-viniferin on PDE Subtypes

Target	PDE1	PDE2	PDE3	PDE4	PDE5	PDE6
% Inhibition	27%	30,9%	47,7%	68,3% IC50 = 4,6 μM	23,3%	16,7%

According to Selnergy prediction, ε-viniferin inhibits Phosphodiesterase with a selectivity for subtype 4.

Table 4. Organisms Containing ε-Viniferin

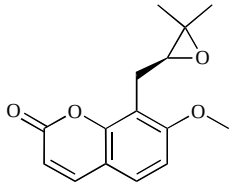
Family	Genus	Species	Naturalist	Organ	Reference
Dipterocarpaceae	*Hopea*	*parviflora*	Bedd	Stem bark	[92]
Dipterocarpaceae	*Shorea*	*seminis*	(De Vriese)	Bark	[93]
Dipterocarpaceae	*Vateria*	*indica*	L.	Stem bark	[90]
Dipterocarpaceae	*Vatica*	*rassak*	(Korth.) Blume	Stem bark	[94]
Fabaceae	*Sophora*	*davidii*	Skeels	Root	[95]
Gnetaceae	*Gnetum*	*hainanense*	Cheng	Liane	[96]
Gnetaceae	*Gnetum*	*klossii*	Markgraf	Stem	[97]
Gnetaceae	*Gnetum*	*latifolium*	Blume	Stem	[98]
Vitaceae	*Vitis*	*coignetiae*	Pulliat		[99]
Vitaceae	*Vitis*	*davidii*	Foex		[100]
Vitaceae	*Vitis*	*flexuosa*	Thunb	Stem	[101]
Vitaceae	*Vitis*	*thunbergii*	(Siebold Zucc.) Druce	Root	[102]
Vitaceae	*Vitis*	*vinifera*	L.	Leaf	[103]

By querying the GPDB database, organisms containing ε-viniferin may also be positioned with this new property (provided that there is no toxicity in the sources and that the molecule is in sufficient quantity). Cosmetics companies generally prefer plant extracts to molecules. Table **4** lists the different organisms containing ε-viniferin. All these plants could therefore be positioned for an anti-inflammatory application.

Since a given metabolite rarely appears as a single unrelated structure, a search for its derivatives in the initial source or in another source containing ε-viniferin can be interesting for finding better candidates in terms of potency, toxicity and accessibility.

b. Meranzin [24]

Meranzin (**2**) is a coumarin derivative and a major compound of *Limnocitrus littoralis* (Miq.) Swingle. This molecule was selected because it can be easily purified in huge quantities from available sources. Furthermore, the biological properties of meranzin were little known. Using Selnergy and the RPG platform, eight proteins were identified as potential interacting partners of meranzin. Upon post-processing of ligand-protein complexes, false positive hits were discarded from the protein hit list. As a direct result, three targets received highest scores when ranked by Selnergy: firstly COX1 and COX2 (anti-inflammatory activities [104]) then PPARγ (anti-obesity and diabetes activities [105]). Inhibition of COX1 and COX2 on binding assays was measured and confirmed inverse screening predictions (Table **5**) though the selectivity between both subtypes was not accessed by prediction. This is expected as the active sites of the enzymes only differ by a valine/leucine residue. The inhibition of COX2 is dose-dependent. For COX1, the results are less clear-cut.



Meranzin (**2**)

At the request of the journal referees, PPARγ activation was determined. A negative control was also required for elastase as this enzyme had been discarded as a false positive hit. Due to the promiscuity of the active sites of COX and 5-lipoxygenase and because this enzyme was not selected by Selnergy, an additional negative control was carried out.

Binding assays showed no interaction of meranzin with elastase at millimolar concentrations nor with 5-lipoxygenase at 100μM, demonstrating the validity of Selnergy predictions.

A cellular based assay was performed to prove PPARγ activation (Table **6**) by meranzin. The activation of the nuclear receptor is dose-dependent. The assay is performed with Rosiglitazone as the reference molecule. Meranzin possesses a similar activation rate at 100μM (tenfold the reference ligand concentration) as Rosiglitazone.

Meranzin constitutes a feasible starting point for finding new COX inhibitors or PPAR modulators either by finding natural derivatives in the same and/or different sources ("natural combinatorial chemistry"). We may also apply medicinal chemistry to optimize

Table 5. Inhibition Activity of Meranzin on COX1 and COX2

Meranzin (μM)	COX1 Inhibition (%)	COX2 Inhibition (%)
0.4	37.1	56.2
4.0	19.1	71.1
40	64.4	79.5

This table shows the inhibition level of meranzin on 2 predicted targets. COX1 inhibition is weaker than COX2. Meranzin inhibits COX2 with a dose-dependent effect, from the very low concentration of 0.4μM. Selnergy predictions are validated.

Table 6. Activation Rate of PPARγ

Meranzin (μM)	Activation Rate
1	1.0 ± 0.1
10	1.1 ± 0.0
100	1.4 ± 0.2
DMSO	1.0 ± 0.0
Rosiglitazone 10 μM	1.5 ± 0.0

Activation rates are relative to DMSO concentration. Meranzin is significantly active at 100μM on PPAR-gamma. The activity is dose-dependent. Rosiglitazone is the reference compound, meranzin is about 10 times less active.

the molecule. To the best of our knowledge, this is the first time that a coumarin derivative has been described for its activity towards PPAR. Thanks to RPG, we may have found a new scaffold for PPAR.

Since biological tests are positive, meranzin and all the known sources could be positioned on these 3 new activities. Three organisms containing meranzin are recorded in GPDB (Table **7**).

Obviously, plant tissues or organs in which compounds cumulate in a significant amount are preferred. Sometimes, inherent conflicts of ecological and economic interests can be avoided: thus, when the very same molecule is present in acceptable concentrations in leaves and roots, then using the leaves is the better choice because root sampling will kill the plants whereas leaf sampling will not. Besides, leaves can be cropped several times on the same plant and so have a better yield. Industrial interests must comply with the sustainable development of agricultural activity with an emphasis on local plant cultivation and biodiversity protection. Therefore, many different and complex criteria have to be considered: plant growing speed, raw material yield, intellectual property, biodiversity and positive returns for the local population.

CONCLUSION

RPG is a highly valuable asset in attempts to discover new activities among natural compounds and organisms containing them. This method is not intended to replace pharmacognosy research methods but rather provides a complementary strategy as well as a new way to study natural compounds and their potential applications. Indeed, pharmacognosy goes from organisms to molecules while RPG reverses this direction, going from compounds to organisms.

Different tools are required in a RPG process: chemoinformatic tools to handle molecules, an inverse screening software with a protein structure database to predict new therapeutic activities, and then a natural compound/source database to position organisms in new activities. Moreover, traditional medicine data help to correlate biological activities with traditional uses. These heterogeneous data are sparsely available on the Internet. We examined related databases, all of which are focused and found that - possessing their own purposes - they are unfortunately not amenable to RPG work. In practice, users would have to "manually" link tools and data, which all too often present incompatibilities with regard to input, processing and output requirements. This rather tedious and daunting task distracts the researchers who have to cope with heterogeneous data formats, not to mention the difficulties encountered during data analysis.

In order to solve such problems, the GRPG platform has been developed. It includes an appropriate database by integrating heterogeneous data with a unique user interface for an "all-in-one" study in search of molecules, organisms, biomolecular targets and biological activities coupled with collections about traditional uses. *Selnergy*, an inverse screening software, is included in the platform for target identification. Indeed, inverse screening associated to a target-focused database is a very powerful tool – not just an add-on for RPG. Many possible applications exist such as the creation of focused libraries containing modulators of targets with similar pharmacological effects or (re)positioning of natural products.

RPG users can not only save time and money, they can also concentrate more on the valorization of natural molecules and organisms. New applications identified for plant extracts by RPG may be of interest for the cosmetic and neutraceutic industries as demonstrated by our studies on meranzin and ε-viniferin. We may also end up with more pharmacologically interesting derivatives by searching through "natural sources of combinatorial chemistry".

Other applications are under consideration. In countries with poor access to medication, RPG could search for new properties of

Table 7. Organisms Containing Meranzin

Family	Genus	Species	Naturalist	Organ	References
Rutaceae	*Citrus*	*maxima*	Merr.	Fruit skin	[106]
Rutaceae	*Murraya*	*gleniei*	Thwaites ex Oliv	Leaf	[107]
Rutaceae	*Limnocitrus*	*littoralis*	(Micq.) Swingle	Leaf	Internal data

local plants and thus provide reliable, validated and cost-cutting therapeutic solutions. The local population would be encouraged to cultivate endemic species. Besides the commercial opportunities, efforts in botanical standardization could strictly control the cultivation and the extraction conditions. The use of fertilizers, pesticides or insecticides has to be monitored and frequent quality controls performed to ensure that the quality of the product is unchanged [108].

Alongside the aforementioned assets, RPG suffers from certain limitations, of which missing data are the most sensitive ones, especially regarding biological activities linked to targets. Collecting the knowledge for future investigation helps to fill in these gaps. To this end, the GRPG platform provides integrated data mining tools not only to enhance automatic data retrieval from diverse repositories of scientific publications and/or patents, but also to assist information management and retrieval. In the meantime, RPG has gained considerable importance as many natural sources and new biological processes have emerged. The larger the amount of data becomes, the more reverse pharmacognosy will be accurately conducted with ever-growing efficiency.

Altogether, it can be safely said that RPG has become a mature concept. It has already arrived at the crossroads between traditional medicine, modern medicine and economic development, for it is now ready to harness the generosity of Nature and above all to gratefully restitute it.

ABBREVIATIONS

ADME	=	Absorption, Distribution, Metabolism and Excretion
BRENDA	=	Braunschweig Enzyme Database
CDK	=	Chemistry Development Kit
COX-1	=	Cyclooxygenase 1
COX-2	=	Cyclooxygenase 2
DMSO	=	Dimethylsulfoxide
DnP	=	Dictionary of Natural Products
FDA	=	Food and Drug Administration
GPCR	=	G protein-coupled receptor
GPDB	=	Greenpharma Database
GPS	=	Global Positioning System
GRPG platform	=	Greenpharma Reverse Pharmacognosy platform
HPLC	=	High-performance Liquid Chromatography
HTS	=	High-Throughput Screening
IUCN	=	International Union for Conservation of Nature
KEGG	=	Kyoto Encyclopedia of Genes and Genomes
MASC	=	Multiple Active Site Corrections
NMR	=	Nuclear Magnetic Resonance
PDB	=	Protein Data Bank
PDE4	=	Phosphodiesterase 4
PDTD	=	Potential Drug Target Database
PPARγ	=	Peroxisome Proliferator-Activated Receptor gamma
RPG	=	Reverse Pharmacognosy
RPL	=	Reverse Pharmacology
RCSB	=	Research Collaboratory for Structural Bioinformatics
TCM	=	Traditional Chinese Medicine
TNF-α	=	Tumor Necrosis Factor alpha
TTD	=	Therapeutic Target Database

REFERENCES

[1] Harvey AL. Natural products in drug discovery. Drug Discov Today 2008; 13: 894-901.

[2] Koehn FE, Carter GT. The evolving role of natural products in drug discovery. Nat Rev Drug Discov 2005; 4: 206-20.

[3] Balunas MJ, Kinghorn AD. Drug discovery from medicinal plants. Life Sci 2005; 78: 431-41.

[4] Rates SMK. Plants as source of drugs. Toxicon 2001; 39: 603-13.

[5] Butler MS. Natural products to drugs: natural product-derived compounds in clinical trials. Nat Prod Rep 2008; 25: 475-516.

[6] Newman DJ, Cragg GM, Snader KM. Natural products as sources of new drugs over the period 1981-2002. J Nat Prod 2003; 66: 1022-37.

[7] Ganesan A. Natural products as a hunting ground for combinatorial chemistry. Curr Opin Biotechnol 2004; 15: 584-90.

[8] Newman DJ, Cragg GM. Natural products as sources of new drugs over the last 25 years. J Nat Prod 2007; 70:461-78.

[9] Woodcock J, Woosley R. The FDA Critical Path Initiative and Its Influence on New Drug Development 2008; 59: 1-12.

[10] Grabowski K, Baringhaus KH, Schneider G. Scaffold diversity of natural products: inspiration for combinatorial library design. Nat Prod Rep 2008; 25: 892-904.

[11] Farnsworth NR, Akerele O, Bingel AS, Soejarto DD, Guo Z. Medicinal plants in therapy. Bull World Health Organ 1985; 63: 965-81.

[12] Gurib-Fakim A. Medicinal plants: traditions of yesterday and drugs of tomorrow. Mol Aspects Med 2006; 27: 1-93.

[13] Newman DJ, Cragg GM, Snader KM. The influence of natural products upon drug discovery. Nat Prod Rep 2000; 17: 215-34.

[14] Feher M, Schmidt JM. Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. J Chem Inf Comput Sci 2009; 43: 218-27.

[15] Tiew P, Ioset JR, Kokpol U, Chavasiri W, Hostettmann K. Antifungal, antioxidant and larvicidal activities of compounds isolated from the heartwood of *Mansonia gagei*. Phytother Res 2003; 17: 190-3.

[16] Sartelet H, Serghat S, Lobstein A, Ingenbleek Y, Anton R, Petitfrere E. Flavonoids extracted from fonio millet (*Digitaria exilis*) reveal potent antithyroid properties. Nutrition 1996; 12: 100-6.

[17] Larsson S, Backlund A, Bohlin L. Reappraising a decade old explanatory model for pharmacognosy. Phytochem Lett 2008; 1: 131-4.

[18] Rollinger JM, Langer T, Stuppner H. Integrated in silico tools for exploiting the natural products' bioactivity. Planta Med 2006; 72: 671-8.

[19] Rollinger JM, Langer T, Stuppner H. Strategies for efficient lead structure discovery from natural products. Curr Med Chem 2006; 13: 1491-507.

[20] Rollinger JM, Haupt S, Stuppner H, Langer T. Combining ethnopharmacology and virtual screening for lead structure discovery: COX-inhibitors as application example. J Chem Inf Comput Sci 2004; 44: 480-8.

[21] Bernard P, Scior T, Didier B, Hibert M, Berthon JY. Ethno-pharmacology and bioinformatic combination for leads discovery: application to phospholipase A2 inhibitors. Phytochemistry 2001; 58: 865-74.

[22] Do Q, Bernard P. Pharmacognosy and reverse pharmacognosy: a new concept for accelerating natural drug discovery. IDrugs 2004; 7: 1017-27.

[23] Do QT, Renimel I, Andre P, Lugnier C, Muller CD, Bernard P. Reverse pharmacognosy: application of selnergy, a new tool for lead discovery. The example of epsilon-viniferin. Curr Drug Discov Technol 2005; 2: 161-7.

[24] Do QT, Lamy C, Renimel I, Sauvan N, André P, Himbert F, *et al.* Reverse pharmacognosy: identifying biological properties for plants by means of their molecule constituents: application to meranzin. Planta Med 2007; 73: 1235-40.

[25] Bruneton J. Pharmacognosie : Phytochimie et plantes médicinales. 2nd ed. Paris: Lavoisier 1993.

[26] Fabricant DS, Farnsworth NR. The value of plants used in traditional medicine for drug discovery. Environ Health Perspect 2001; 109(Suppl 1): 69-75.

[27] Patwardhan B, Vaidya ADB, Chorghade M. Ayurveda and natural products drug discovery. Curr Sci 2004; 86: 789-99.
[28] Patwardhan B. Ethnopharmacology and drug discovery. J Ethnopharmacol 2005; 100: 50-2.
[29] Vaidya A. Reverse pharmacological correlates of ayurvedic drug actions. Indian J Pharmacol 2006; 38: 311-5.
[30] Patwardhan B, Mashelkar RA. Traditional medicine inspired approaches to drug discovery: can Ayurveda show the way forward? Drug Discov Today 2009; 14: 804-11.
[31] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 1997; 23: 3-25.
[32] Ortholand JY, Ganesan A. Natural products and combinatorial chemistry: back to the future. Curr Opin Chem Biol 2004; 8: 271-80.
[33] International Union for Conservation of Nature, IUCN red list of threatened species™. Available at: http://www.iucnredlist.org/
[34] Wermuth CG. Multitargeted drugs: the end of the `one-target-one-disease` philosophy? Drug Discov Today 2004; 9: 826-7.
[35] Aschwanden C. Herbs for health, but how safe are they? Bull World Health Organ 2001; 79: 691-2.
[36] The Open Babel Package, version 2.2.1. (http://openbabel. sourceforge.net) Oct 2009.
[37] Steinbeck C, Hoppe C, Kuhn S, Floris M, Guha R, Willighagen EL. Recent developments of the chemistry development kit(CDK)-: An open-source java library for chemo- and bioinformatics. Curr Pharm Des 2006; 12: 2111-20.
[38] Monge A, Arrault A, Marot C, Morin-Allory L. Managing, profiling and analyzing a library of 2.6 million compounds gathered from 32 chemical providers. Mol Divers 2006; 10: 389-403.
[39] Tripos, Sybyl, version 8.0 (http://www.tripos.com/sybyl) Oct 2009.
[40] ChemAxon (http://www.chemaxon.com) Oct 2009.
[41] Sperandio O, Miteva MA, Delfaud F, Villoutreix BO. Receptor-based computational screening of compound databases: the main docking-scoring engines. Curr Protein Pept Sci 2006; 7: 369-94.
[42] Tang Y, Zhu W, Chen K, Jiang H. New technologies in computer-aided drug design: toward target identification and new chemical entity discovery. Drug Discov Today: Technol 2006; 3: 307-13.
[43] Pearlman RS, Balducci R, Rusinko A, Skell JM, Smith KM. CONCORD. St. Louis, MO: Tripos Associates 1994.
[44] Sadowski J, Wagener M, Gasteiger J. CORINA: Automatic Generation of High-Quality 3D-Molecular Models for Application in QSAR. In: Sanz F, Giraldo J, Manaut F, Eds. QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications. Spain: Prous Science Publishers 1995; 646-51.
[45] Goodsell DS, Morris GM, Olson AJ. Automated docking of flexible ligands: applications of AutoDock. J Mol Recognit 1996; 9: 1-5.
[46] Ewing TJA, Makino S Skillman AG, Kuntz ID. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J Comput Aided Mol Des 2001; 15: 411-28.
[47] Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. J Mol Biol 1996; 261: 470-89.
[48] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 2004; 47: 1739-49.
[49] Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997; 267: 727-48.
[50] Jain AN. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. J Med Chem 2003; 46: 499-511.
[51] Clark RD, Strizhev A, Leonard JM, Blake JF, Matthew JB. Consensus scoring for ligand/protein interactions. J Mol Graph Model 2002; 20: 281-95.
[52] Vigers GPA, Rizzi JP. Multiple active site corrections for docking and virtual screening. J Med Chem 2004; 47: 80-9.
[53] Kellenberger E, Foata N, Rognan D. Ranking targets in structure-based virtual screening of three-dimensional protein libraries: methods and problems. J Chem Inf Model 2008; 48: 1014-25.
[54] Chen X, Ung CY, Chen Y. Can an in silico drug-target search method be used to probe potential mechanisms of medicinal plant ingredients? Nat Prod Rep 2003; 20: 432-44.
[55] Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, *et al.* The protein data bank. Acta Crystallogr. Sect D: Biol Crystallogr 2002; 58: 899-907.
[56] Li H, Gao Z, Kang L, Zhang H, Yang K, Yu K, *et al.* TarFisDock: a web server for identifying drug targets with docking approach. Nucleic Acids Res 2006; 34: 219-24.
[57] Webb EC. Enzyme nomenclature 1992: recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes, 1992.
[58] Teague SJ. Implications of protein flexibility for drug discovery. Nat Rev Drug Discov 2003; 2: 527-41.
[59] Carlson HA. Protein flexibility is an important component of structure-based drug discovery. Curr Pharm Des 2002; 8: 1571-8.
[60] Claußen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular dockingconsidering protein structure variations. J Mol Biol 2001; 308: 377-95.
[61] Berman HM. The protein data bank: a historical perspective. Acta Crystallogr Sect A: Found Crystallogr 2008; 64: 88-95.
[62] Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000; 28: 27-30.
[63] Goto S, Okuno Y, Hattori M, Nishioka T, Kanehisa M. LIGAND: database of chemicalcompounds and reactions in biological pathways. Nucleic Acids Res 2002; 30: 402-4.
[64] Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, *et al.* BRENDA, the enzyme database: updates and major new developments. Nucleic Acids Res 2004; 32: 431-3.
[65] Chang A, Scheer M, Grote A, Schomburg I, Schomburg D. Brenda, amenda and frenda the enzyme information system: new content and tools in 2009. Nucleic Acids Res 2009; 37: 588-92.
[66] Chen X, Ji ZL, Chen YZ. TTD: therapeutic target database. Nucleic Acids Res 2002; 30: 412-5.
[67] Gao Z, Li H, Zhang H, Liu W, Kang L, Luo X, *et al.* PDTD: a web-accessible protein database for drug target identification. BMC Bioinformatics 2008; 9: 104.
[68] Kellenberger E, Muller P, Schalon C, Bret G, Foata N, Rognan D. sc-PDB: an annotated database of druggable binding sites from the Protein Data Bank. J Chem Inf Model 2006; 46: 717-27.
[69] Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, *et al.* DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 2008; 36: 901-6.
[70] Günther S, Kuhn M, Dunkel M, Campillos M, Senger C, Petsalaki E, *et al.* Super target and matador: resources for exploring drug-target relationships. Nucleic Acids Res 2008; 36: 919-22.
[71] Scior T, Bernard P, Medina-Franco JL, Maggiora GM. Large compound databases for structure-activity relationships studies in drug discovery. Mini Rev Med Chem 2007; 7: 851-60.
[72] Kuhn M, von Mering C, Campillos M, Jensen LJ, Bork P. STITCH: interaction networks of chemicals and proteins. Nucleic Acids Res 2008; 36: 684-8.
[73] STITCH, (http://stitch.embl.de) Oct 2009.
[74] Thomas MB, Lin N, Beck HH. A database Model for integrating and facilitating collaborative ethnomedicinal research. Pharmaceutical Biol 2001; 39: 41-52.
[75] Verpoorte R. Repository for ethnopharmacological survey data? J Ethnopharmacol 2008; 120: 127-8.
[76] Skoczen S, Bussmann RW. ebDB-Filling the gap for an International Ethnobotany Database. Lyonia 2006; 11: 71-81.
[77] Ji ZL, Zhou H, Wang JF, Han LY, Zheng CJ, Chen YZ. Traditional Chinese medicine information database. J Ethnopharmacol 2006; 103: 501.
[78] Gry J, Black L, Eriksen FD, Pilegaard K, Plumb J, Rhodes M, *et al.* EuroFIR-BASIS - a combined composition and biological activity database for bioactive compounds in plant-based foods. Trends Food Sci Technol 2007; 18: 434-44.
[79] Babu PA, Puppala SS, Aswini SL, Vani MR, Kumar CN, Prasanna T. A database of natural products and chemical entities from marine habitat. Bioinformation 2008; 3: 142-3.
[80] Lei J, Zhou J. A marine natural product database. J Chem Inf Comput Sci 2002; 42: 742-8.
[81] Taylor L. The Tropical Plant Database (http://www.rain-tree.com/plants.htm) Oct 2009.

[82] Holden JM, Eldridge AL, Beecher GR, Marilyn Buzzard I, Bhagwat S, Davis CS, *et al.* Carotenoid content of US foods: an update of the database. J Food Compost Anal 1999; 12: 169-96

[83] Duke J. Dr. Duke's phytochemical and ethnobotanical databases. ARS, USDA (http://www.ars-grin.gov/duke) Oct 2009.

[84] Buckingham J. Dictionary of natural products. London: Chapman & Hall; 1994.

[85] Loub WD, Farnsworth NR, Soejarto DD, Quinn ML. NAPRALERT: computer handling of natural product research data. J Chem Inf Comput Sci 1985; 25: 99-103.

[86] Farnsworth N. The Napralert database as an information source for application to traditional medicine. OMS Geneva, Switzerland 1993.

[87] Plant for a future database (http://www.pfaf.org) Oct 2009.

[88] Dunkel M, Fullbeck M, Neumann S, Preissner R. SuperNatural: a searchable database of available natural compounds. Nucleic Acids Res 2006; 34: 678-83.

[89] Jmol, version 11.2.14 (http://www.jmol.org) Oct 2009.

[90] Mishima S, Matsumoto K, Futamara Y, Araki Y, Ito T, Tanaka T, *et al.* Antitumor effect of stilbenoids from *Vateria indica* against allografted sarcoma S-180 in animal model. J Exp Ther Oncol 2003; 3: 283-8.

[91] Privat C, Telo JP, Bernardes-Genisson V, Vieira A, Souchard JP, Nepveu F. Antioxidant properties of trans-[epsilon]-viniferin as compared to stilbene derivatives in aqueous and nonaqueous media. J Agric Food Chem 2002; 50: 1213-7.

[92] Tanaka T, Ito T, Ido Y, Son TK, Nakaya K, Iinuma M, *et al.* Stilbenoids in the stem bark of *Hopea parviflora*. Phytochemistry 2000; 53: 1015-19.

[93] Aminah NS, Achmad SA, Aimi N, Ghisalberti EL, Hakim EH, Kitajima M, *et al.* Diptoindonesin A, a new C-glucoside of epsilon-viniferin from *Shorea seminis (Dipterocarpaceae)*. Fitoterapia 2002; 73: 501-7.

[94] Tanaka T, Ito T, Nakaya K, Iinuma M, Riswan S. Oligostilbenoids in stem bark of *Vatica rassak*. Phytochemistry 2000; 54: 63-9.

[95] Tanaka T, Ito T, Iinuma M, Ohyama M, Ichise M, Tateishi Y. Stilbene oligomers in roots of *Sophora davidii.* Phytochemistry 2000; 53: 1009-14.

[96] Huang KS, Wang YH, Li RL, Lin M. Five new stilbene dimers from the lianas of *Gnetum hainanense*. J Nat Prod 2000; 63: 86-9.

[97] Ali Z, Tanaka T, Iliya I, Iinuma M, Furusawa M, Ito T, *et al.* Phenolic constituents of Gnetum klossii. J Nat Prod 2003; 66: 558-60.

[98] Iliya I, Ali Z, Tanaka T, Iinuma M, Furusawa M, Nakaya K, *et al.* Stilbenoids from the stem of *Gnetum latifolium (Gnetaceae)*. Phytochemistry 2002; 61: 959-61.

[99] Oshima Y, Namao K, Kamijou A, Matsuoka S, Nakano M, Terao K, *et al.* Powerful hepatoprotective and hepatotoxic plant oligostilbenes, isolated from the Oriental medicinal plant *Vitis coignetiae (Vitaceae)*. Cell Mol Life Sci 1995; 51: 63-6.

[100] Yang JZ, Zhou LX, Ding Y. Studies on chemical constituents of oligostilbenes from *Vitis davidii* Foex. Zhongguo Zhong yao za zhi 2001; 26: 553-5.

[101] Li W, Li B. Flexuosol A, a new tetrastilbene from *Vitis flexuosa*. J Nat Prod 1998; 61:646-7.

[102] Huang YL, Tsai WJ, Shen CC, Chen CC. Resveratrol derivatives from the roots of *Vitis thunbergii*. J Nat Prod 2005; 68: 217-20.

[103] Langcake P. Disease resistance of *Vitis spp.* and the production of the stress metabolites resveratrol, epsilon-viniferin, alpha-viniferin and pterostilbene. Physiol Plant Pathol 1981; 18: 213-26.

[104] Fiorucci S, Meli R, Bucci M, Cirino G. Dual inhibitors of cyclooxygenase and 5-lipoxygenase. A new avenue in anti-inflammatory therapy? Biochem Pharmacol 2001; 62: 1433-8.

[105] Park KS, Ciaraldi TP, Abrams-Carter L, Mudaliar S, Nikoulina SE, Henry RR. PPAR-gamma gene expression is elevated in skeletal muscle of obese and type II diabetic subjects. Diabetes 1997; 46: 1230-4.

[106] Teng WY, Chen CC, Chung RS. HPLC comparison of supercritical fluid extraction and solvent extraction of coumarins from the peel of *Citrus maxima* fruit. Phytochem Anal 2005; 16: 459-62.

[107] Wickramaratne DBM, Kumar V, Balasubramaniam S. Murra-gleinin, a Coumarin from*Murraya gleinei* Leaves. Phytochemistry 1984; 23: 2964-6.

[108] Street RA, Stirk WA, Van Staden J. South African traditional medicinal plant trade-Challenges in regulating quality, safety and efficacy. J Ethnopharmacol 2008; 119: 705-10.

Received: February 8, 2010 **Accepted: February 26, 2010**