Natural products: sources and databases

Melanie Füllbeck,† Elke Michalsky,† Mathias Dunkel and Robert Preissner*

Received (in Cambridge, UK) 6 March 2006
First published as an Advance Article on the web 10th April 2006
DOI: 10.1039/b513504b

Covering: up to 2006

This Highlight gives a general survey of natural product databases, suppliers and manufacturers. It describes opportunities for researchers to obtain information about natural compounds and makes a proposal for successful identification of pharmaceutically relevant substances.

1 Introduction

About 200,000 natural compounds are currently known. Many natural products occur in almost all parts of the world (e.g. salicylic acid) and have been investigated in traditional medicines as well as in modern drug discovery.

In contrast to this, a great number of natural compounds occur in defined regions of the world, they are rare, or the extraction
of these compounds is rather difficult. It is supposed that only 5–15% of approximately 250 000 species of higher plants have been investigated for the presence of bioactive compounds. Actinomycetes have been and are still a major source for natural compounds used in drug discovery. The increasing number of natural compounds demonstrates that the tremendous potential of nature's diversity has not been exhausted yet. Because of many untapped origins (extreme environments), unthought of possibilities can be found in nature to be used for innovations in a variety of research fields. This is just one compelling argument for expanding the exploration of nature as a source of novel active agents. Nevertheless, a progressive organisation for an effective development in the field of natural product research is necessary.

For historical reasons, the first widely-used natural compounds were extracted from plants and animals. Later, microbial compounds came into focus. As 90% of the earth's forms of life inhabit the oceans, researchers’ attention today is focused on marine natural products. One reason for this may be that the molecular physiology of eukaryotic cells evolved in early marine ancestors. Marine microorganisms could become a major source for the discovery of new drugs, because the biological diversity in marine ecosystems like coral reefs or deep sea floors is probably higher than in the rainforest, and, furthermore, marine natural products released to water dilute rapidly, so that they have to be highly potent. Many marine natural compounds have been marketed as drugs, e.g. cytarabine (Fig. 1), or are under clinical trials for many different diseases.

Some marine natural products are undergoing clinical trials for treatment of cancer, analgesia, allergy, and cognitive diseases. The chemical diversity of these compounds is tremendous and may offer inspiration for innovations in the fields of medicine, nutrition, agrochemical research and life sciences.

The rise of combinatorial chemistry reduced the interest in natural compounds among pharmaceutical companies. However, not all promises of high-throughput synthesis were kept, because the produced compounds that are drug-like in size and properties have not evolved to interact with biomolecules and may therefore induce unexpected and possibly severe side effects. About half of all compounds that were successful in clinical trials during the past 20 years have at least been derived from natural origin, so interest in natural compounds has been rekindled. Some examples, such as heparin, vitamin K antagonists, hirudin (Fig. 2), streptokinase (Fig. 3) and urokinase (Fig. 4) are still among the most potent antithrombotic drugs.

Fig. 1 Three-dimensional structure of cytarabine (bold structure) and its conformers, which have been calculated with the Accelrys software Catalyst to demonstrate the flexibility of the molecule.

Fig. 2 Antithrombotic hirudin (red structure) in complex with thrombin (blue surface). The catalytic triangle of thrombin is highlighted in green (PDB code: 1HRT).

Fig. 3 Protein structure of streptokinase (PDB code: 1BML).

There are large (public) databases containing hundreds of thousands of compounds, but the origin and availability of these
compounds are often neglected. On the other hand, a number of suppliers provide information about their compounds in a variety of formats. In this article, we summarise suppliers and manufacturers of natural compounds and extracts. Furthermore, existing public and commercial databases and an exhaustive new database committed to storing information on available natural compounds are introduced.13 In addition to the direct screening of natural compounds, the new database allows promising building blocks or scaffolds to be selected for the design of novel drugs.

2 Public databases

This section lists molecular databases publicly available on the Internet, which contain, amongst others, natural compounds (see Table 1). Typically, they represent collections from other different databases. They supply, for instance, structures, names, synonyms, CAS Registry Numbers, chemical properties, information about the use of a compound, patents and corresponding literature references. Information about the number of contained natural compounds is only available in some cases.

2.1 ChEBI

The Chemical Entities of Biological Interest (ChEBI) database is produced by the European Bioinformatics Institute and contains naturally occurring and synthetic small molecular entities.

A text search in ChEBI, which may, for instance, be performed within the synonyms, the formula, or SMILES records, *inter alia* supplies a structure, names, synonyms, SMILES, the formula, several registry numbers. Structures can be saved as 2D Mol files. Additionally, each ChEBI entry is linked to associated UniProt, proteins, in such a way that information about known biological functions can be accessed directly.

As an outstanding feature, the database comprises an ontological classification. Here, the relationships between molecular entities or classes of entities and their parents and/or children are specified. With the aid of this classification, details about the molecule’s chemistry and biological activity can be derived. ChEBI Ontology is subdivided into three separate sub-ontologies: Molecular Structure Ontology, in which molecular entities and elementary particles are classified according to their structure; Biological Function Ontology, which distinguishes between functions, such as antibiotics, antiviral agents, coenzymes, enzyme inhibitors; Application Ontology, which classifies entities, where appropriate, on the basis of their applications, *e.g.* as pesticides, detergents, healthcare products, fuel.

Currently, ChEBI contains 6806 curated entries and 6129 structures. The data can be downloaded in four different formats, *i.e.* a flat file format, as Oracle SQL table dumps, in OBO Ontology format, and as generic SQL dumps.

2.2 ChemBank

ChemBank is a freely available collection of data about small molecules focusing on their effects on biology. ChemBank is under active development by the Broad Institute Chemical Biology Program (BCB) and is sponsored by the National Cancer Institute’s Initiative for Chemical Genetics (ICG).

Currently, ChemBank contains more than 1.1 million chemical structures, including 150 000 commercially available compounds with vendor catalogue numbers and a reference set of over 700 000 other drug-like compounds. Files containing 2D structures or SMILES of all bioactives as well as additional information are downloadable in .sdf or .xml format.

ChemBank enables text searches, which can, for example, be limited to natural products, bioactives or commercially available compounds. Synonyms, SMILES, structure pictures, molecular weight, molecular formulae, solubility, as well as information about classification, biological activity and vendors can be retrieved. Similarity or substructure searches can be performed by providing SMILES.

The Small Molecule Bioactives Database represents a subset of the entire database and contains chemical structures and biological activity data for over 6300 known bioactive compounds. Furthermore, publicly available Biological Assay Data from high-throughput chemical screens can be searched, analysed and downloaded.

Table 1: Public databases containing natural compounds. Web addresses, total numbers of compounds as well as numbers of contained natural compounds are given. N. A.: information not available

<table>
<thead>
<tr>
<th>Name</th>
<th>Homepage</th>
<th>Number of compounds</th>
<th>Number of natural compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChEBI</td>
<td><a href="http://www.ebi.ac.uk/chebi">www.ebi.ac.uk/chebi</a></td>
<td>&gt;6800</td>
<td>~3500</td>
</tr>
<tr>
<td>ChemBank</td>
<td>chembank.broad.harvard.edu</td>
<td>&gt;1 100 000</td>
<td>N. A.</td>
</tr>
<tr>
<td>ChemID</td>
<td>chem.sis.nlm.nih.gov/chemidplus</td>
<td>&gt;379 000</td>
<td>~2000</td>
</tr>
<tr>
<td>NCI</td>
<td>cactus.nci.nih.gov/ncidb3/down load_ncidb3.html</td>
<td>&gt;260 000</td>
<td>N. A.</td>
</tr>
<tr>
<td>PubChem</td>
<td>pubchem.ncbi.nlm.nih.gov</td>
<td>&gt;5 000 000</td>
<td>N. A.</td>
</tr>
</tbody>
</table>

This journal is © The Royal Society of Chemistry 2006
2.3 ChemIDplus

The free web-based search system ChemIDplus provides access to structure and nomenclature information for the identification of chemical substances cited in NLM (National Library of Medicine) databases. ChemIDplus contains over 379,000 chemical records, of which over 257,000 include chemical structures. The database is searchable by name, synonym, CAS Registry Number, molecular formula, classification code, locator code and structure. Physicochemical properties, e.g. molecular weight, logP, water solubility, melting and boiling points, can be used as search criteria; 2D similarity or substructure searches can be performed with user-provided structures. ChemIDplus Lite is available for name and registry number searching without the need for plugins or applets.

2.4 NCI

The National Cancer Institute provides downloadable structure files of their Open NCI Database Compounds. As of September 2003, they represent a collection of 260,071 structures, combined from several http://dtp.nci.nih.gov/ (Developmental Therapeutics Program) releases. For the major part, i.e. currently 192,310 of the structures, calculated 3D structures are available. A web search form, the Enhanced NCI Database Browser, can be found at the URL http://cactus.nci.nih.gov/ncidb2/. Here, 250,251 open structures from the Open NCI Database can be screened using search criteria such as NSC and CAS numbers, molecular formula and weight, ClogP, atom and ring counts, H-bond donor and acceptor counts. Moreover, 2D similarity or substructure and 3D pharmacophore searches can be performed.

2.5 PubChem

PubChem is part of the US National Institutes of Health’s ‘Molecular Libraries’ initiative and is organised as three linked databases within the NCBI’s http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi information retrieval system. These are PubChem Substance, PubChem Compound and PubChem BioAssay. PubChem Compound currently contains more than five million unique 2D structures with computed properties. It can be searched for names, chemical properties or chemical elements contained in a compound. Structural similarity or substructure searches can be performed on the basis of SMILES or an uploaded structure. Each compound is linked to those substances from PubChem Substance in which it is contained. Data can be downloaded in various file formats.

3 Commercial databases

There are large compendia of natural compounds distributed commercially. Usually, they provide similar search options as public databases, but access is fee-based. Some of them are delivered on CD-ROM or DVD, others are accessible via Internet accounts for a fee. Partly, printed counterparts are available. In some cases, free reduced or test versions are available. Information about availability and supply of the substances is not necessarily contained in commercial databases. This section lists a number of commercial natural product databases (see Table 2).

<table>
<thead>
<tr>
<th>Name</th>
<th>Homepage</th>
<th>Number of compounds</th>
<th>Number of natural compounds</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>AntiBase</td>
<td>en.wiley.com/WileyCDA</td>
<td>&gt;31,000</td>
<td>&gt;31,000</td>
<td>$8250 (CD-ROM) €8755 £4714</td>
</tr>
<tr>
<td>Chapman &amp; Hall</td>
<td><a href="http://www.cas.org">www.cas.org</a></td>
<td>&gt;190,000 (CD-ROM)</td>
<td>&gt;185,000 (Web)</td>
<td>$31,400 (CD-ROM) On inquiry</td>
</tr>
<tr>
<td>Chemical Abstracts</td>
<td><a href="http://www.cas.org">www.cas.org</a></td>
<td>&gt;27,000,000</td>
<td>~6000</td>
<td>$4,550 (CD-ROM) On inquiry</td>
</tr>
<tr>
<td>RöMPP Online</td>
<td><a href="http://www.roempp.com/thieme-chemistry/rmp/info">www.roempp.com/thieme-chemistry/rmp/info</a></td>
<td>~6000</td>
<td>~6000</td>
<td>On subscription basis (ONLINE versions)</td>
</tr>
<tr>
<td>The Beilstein Database</td>
<td><a href="http://www.beilstein-institut.de/englisch/1024/chemie/index.php3">www.beilstein-institut.de/englisch/1024/chemie/index.php3</a></td>
<td>&gt;9,300,000</td>
<td>&gt;14,200</td>
<td>On subscription basis (ONLINE versions)</td>
</tr>
<tr>
<td>The Merck Index</td>
<td><a href="http://www.merckbooks.com">www.merckbooks.com</a></td>
<td>&gt;10,000</td>
<td>N. A.</td>
<td>$65 (printed version); prices on request (CD-ROM)</td>
</tr>
</tbody>
</table>

Table 2 Commercial natural compound databases. Web addresses, overall numbers of compounds, numbers of natural compounds and prices are given. N. A.: information not available.
3.1 AntiBase

AntiBase 2005, published by H. Laatsch (University of Göttingen, Germany), is the latest version of a database of 31,022 natural compounds from microorganisms and higher fungi. The data have been collected from the primary and secondary literature and then carefully checked and validated. AntiBase provides descriptive data (molecular formula and mass, elemental composition, CAS Registry Number; physico-chemical data (melting point, optical rotation); spectroscopic data (UV, 13C-NMR, IR and mass spectra); biological data (pharmacological activity, toxicity); information on origin and isolation; and a summary of literature sources. Structural and substructural searches can be performed, also taking user-provided structures. This data collection can be obtained on CD-ROM in the formats ISIS/Base and ChemFinder. Updates are intended once per year.

3.2 Chapman & Hall

In 1997, Chapman & Hall published the most recent supplement to their ‘Dictionary of Natural Products’ in book form. Since then, it has been available on CD-ROM as a single-user or a network version. It is updated every 6 months, the latest version having been launched in January 2005. Currently, a web version which dates from the year 2000 is available. The Dictionary of Natural Products contains chemical, physical and structural data on over 190,000 natural products, organised into more than 43,000 entries. These are classified into alkaloids, antibiotics, flavonoids, carbohydrates, lignans, amino acids, peptides and proteins, polypeptides, polypeptides, steroids, tannins, terpenoids and products of unknown structure. Various information about the compounds is contained in the Dictionary of Natural Products: accurate, systematic chemical names, trade names, and trivial names; searchable chemical structures; CAS Registry Numbers; extensive UV and biological source data; molecular formulae, accurate mass and per cent composition; physical data including melting/boiling points, optical rotation, dissociation constants and refractive index; applications including biological use; hazard and toxicity information including LD50 values; concise citliography; biogenetic classification scheme, grouping all related natural products together. The features provided on the CD-ROM are simultaneous text and structure searching, flexible display format, customisable hit list display, enhanced coverage of UV spectra and biological activity. Special academic prices are offered on request.

3.3 CAS

CAS (Chemical Abstracts Service) is located in Columbus, Ohio, and is a division of the American Chemical Society. Data produced by CAS are accessible through the printed Chemical Abstracts, Chemical Abstracts on CD, STN, the CAS files distributed through licensed vendors, the SciFinder and SciFinder Scholar desktop research tools, and the STN Easy or STN on the Web services.

The CAS REGISTRY, produced by CAS, contains information about more than 27 million organic and inorganic substances, including more than 8 million commercially available chemicals, and 57 million DNA sequences. This computer system automatically identifies and catalogues all substance information that CAS selects from journal articles, patents, conference proceedings and other published sources. These data are enriched with calculated properties and experimental property data. CAS REGISTRY covers literature from 1957 to the present, with some classes going back to the early 1900s, and is updated daily with approximately 4000 new substance records.

When a new chemical substance is processed by CAS, its molecular structure diagram, systematic chemical name, molecular formula, and other identifying information are added to the Registry, and the latter is assigned a unique numerical identifier, the CAS Registry Number (CASRN). CAS Registry Numbers are currently used by many organisations around the world to identify chemical substances unambiguously.

The CAS REGISTRY can, for example, be explored by chemical structure, substructure or structural similarity, substance identifier, molecular formula or reactions, with availability for the latter depending on the particular product. For the commercially available chemicals, contact information is available depending upon the supplier, e.g., company name and address, catalogue name, order number, pricing terms, packaging and shipping information, chemical and trade names.

An additional service provided by CAS is the Document Detective Service (CAS DDS), which delivers print copies of documents found in the Chemical Abstracts at individual charges. Chemical Abstracts on CD demo disks and an STN Easy demo are available free of charge.

3.4 RÖMPP Online

RÖMPP Online is a large web-accessible collection of knowledge in chemistry and the life sciences. As a part of it, ‘RÖMPP Natural Products’ contains information on approximately 6000 natural products, namely chemical structure, stereochemistry and absolute configuration, nomenclature, molecular formula and weight, biological and physical properties, occurrence and biological sources, biosynthesis pathways, medical usage and application, citliographic data with current literature. It provides structural formulae and diagrams. The database is searchable by keywords and CAS numbers that can be combined by logical operators. The RÖMPP encyclopaedias are also available on CD-ROM. RÖMPP Online is updated quarterly and currently is available only in German. The book form of ‘RÖMPP Natural Products’ is also distributed in English and is intended to be published also as an English online version. One-year licences for RÖMPP Online are available, whereas prices are given on inquiry.

3.5 The Beilstein Database

Based on Beilstein’s ‘Handbuch der organischen Chemie’, the Beilstein Database covers organic chemistry from 1771 to the present. The latest release, made available in September 2001, contains over 9.3 million structures, supplemented with numerous reactions and property records.

The Beilstein Database also contains the Beilstein CrossFire Abstracts Database, which contains over 600,000 original author abstracts from 1980 to the present, and the EcoPharm database, which focuses on areas such as pharmacology, toxicology and ecological chemistry.

The Beilstein database can be accessed online using the CrossFire information system interface (MDL Information Systems).
CrossFire Beilstein is arranged in three primary data domains: substances, reactions and literature. The substance domain stores structural information and literature references, including chemical, physical and bioactivity data. The reaction domain details the preparation of substances, and thus allows chemists to find successful ways to optimally synthesise known and novel compounds. The literature domain includes citations, titles, and abstracts, which are hyperlinked to the substance and reaction domain entries.

The Beilstein database is updated quarterly and can be hosted for use at an organisation’s site or accessed via different online services, e.g. MDL CrossFire Direct, DiscoveryGate. Free evaluation versions are available on request. Searching opportunities include structure, substructure, reaction, property and keyword searches. Beilstein includes own registry numbers, the Beilstein Registry Numbers (BRN), but also CAS Registry Numbers.

### 3.6 The Merck Index

‘The Merck Index’ is a one-volume encyclopaedia of chemicals, drugs and biologicals that contains more than 10 000 monographs. Amongst others, biologicals and natural products as well as plants and herbal medicines are included. The following information is provided: chemical, common and generic names, trademarks and associated companies, CAS Registry Numbers, chemical structures, molecular formulae, weights and percentage composition, derivatives, scientific and patent literature references, physical and toxicity data, therapeutic and commercial uses, caution and hazard information.

A collection of supplementary tables contains physical, chemical and biomedical data and listings of pharmaceutical company names, locations and experimental drug codes. The Organic Name Reactions section contains over 400 reactions, featuring descriptions, literature references and graphical depictions.

Covering knowledge from 1889, ‘The Merck Index’ is now available on CD-ROM and in two different online versions. The current CD-ROM version and the web-accessible version, Thirteenth Edition (2003), are co-published by Merck & Co., Inc. and CambridgeSoft. This edition features text and substructure searching tools for exploring the database. It contains the text and structures of the monographs, the supplementary tables section and the Organic Name Reactions section. ‘The Merck Index’ Online is a text-searchable database that contains the monograph section of ‘The Merck Index’, Thirteenth Edition. Sample content is available free of charge. A complete online version requires subscription.

### 4 Supplier databases

Many suppliers and manufacturers around the world specialise in the discovery and development of new natural products from a variety of biological habitats. For many natural products, the original supplier often remains unspecified, because, after isolation, the natural compounds or extracts are sold to chemical catalogue companies. This section gives an overview of a variety of manufacturers, original suppliers, resellers, and a manufacturer spanning database which enables the user to obtain information about the original suppliers of a great number of natural compounds.

#### 4.1 Suppliers and manufacturers

There are numerous manufacturers of natural compounds. In this article, a selection of suppliers with diverse natural product libraries will be introduced. Since many natural products appear in defined regions the manufacturers are spread throughout the world (see Tables 3–5).

A large number of companies have their headquarters in the USA and offer natural products from a wide variety of biological habitats. The Gaia Chemical Corporation, which has its headquarters on the east coast of the United States, provides about 500 pure natural products and derivatives for experimental testing. The biological diversity of its compounds is ensured spanning database which enables the user to obtain information about the original suppliers of a great number of natural compounds.

#### Table 3 Companies and suppliers of natural compounds in America. Web addresses of the companies, number of compounds and the availability of the compounds for experimental testing. Yes: compounds are available for testing. N. A.: information not available

<table>
<thead>
<tr>
<th>Name</th>
<th>Homepage http://...</th>
<th>Number of compounds or extracts</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaia Chemical Corporation</td>
<td><a href="http://www.gaiachem.com">www.gaiachem.com</a></td>
<td>∼500</td>
<td>Yes</td>
</tr>
<tr>
<td>Kosan Biosciences</td>
<td><a href="http://www.kosan.com">www.kosan.com</a></td>
<td>N. A.</td>
<td>Synthesis after contract</td>
</tr>
<tr>
<td>Life Pharm Inc.</td>
<td><a href="http://www.lifepharms.com">www.lifepharms.com</a></td>
<td>150 000</td>
<td>Yes (after finished preparation)</td>
</tr>
<tr>
<td>Magellan BioScience Group Inc.</td>
<td><a href="http://www.magellanbioscience.com">www.magellanbioscience.com</a></td>
<td>&gt;10 000</td>
<td>Yes</td>
</tr>
<tr>
<td>Martek Biosciences</td>
<td><a href="http://www.martekbio.com">www.martekbio.com</a></td>
<td>N. A.</td>
<td>Yes</td>
</tr>
<tr>
<td>Mera Pharmaceuticals, Inc.</td>
<td><a href="http://www.aquasearch.com">www.aquasearch.com</a></td>
<td>N. A.</td>
<td>N. A.</td>
</tr>
<tr>
<td>Micro-Source, Discovery Systems, Inc.</td>
<td><a href="http://www.msdiscovery.com">www.msdiscovery.com</a></td>
<td>N. A.</td>
<td>N. A.</td>
</tr>
<tr>
<td>Sequoia Sciences</td>
<td><a href="http://www.sequoiasciences.com">www.sequoiasciences.com</a></td>
<td>N. A.</td>
<td>Yes</td>
</tr>
<tr>
<td>TimTec</td>
<td><a href="http://www.timtec.net">www.timtec.net</a></td>
<td>13 000</td>
<td>Yes</td>
</tr>
</tbody>
</table>

#### Table 4 Companies and suppliers of natural compounds in Asia and Australia. Web addresses of the companies, number of compounds and the availability of the compounds for experimental testing. Yes: compounds are available for testing. N. A.: information not available

<table>
<thead>
<tr>
<th>Name</th>
<th>Homepage http://...</th>
<th>Number of compounds or extracts</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Albany Molecular Research, Inc.</td>
<td><a href="http://www.albmolecular.com">www.albmolecular.com</a></td>
<td>∼300 000</td>
<td>Yes</td>
</tr>
<tr>
<td>Australia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cerylid Biosciences</td>
<td><a href="http://www.cerylid.com.au">www.cerylid.com.au</a></td>
<td>600 000</td>
<td>Yes</td>
</tr>
<tr>
<td>Entocosm Pty Ltd.</td>
<td><a href="http://www.csiro.au">www.csiro.au</a></td>
<td>N. A.</td>
<td>Yes (for collaborators)</td>
</tr>
</tbody>
</table>
by its contacts with botanical exporters from Chile, China, India, Peru and Russia. In contrast to this, **Kosan Biosciences** specialises, on the one hand, in developing drug candidates from a single important class of natural product compounds known as polyketides, but, on the other hand, it offers contract development and manufacturing services for all types of small molecule APIs (Active Pharmaceutical Ingredients) produced by fermentation, depending upon the needs of its collaborators. Various extracts of the fruiting bodies of basidiomycetes and ascomycetes are offered by **LifePharms, Inc.** The collection currently contains about 11 000 organisms. Over 10 000 natural product extracts from **Magellan BioScience Group, Inc.** are available for experimental testing. The biotic and geographic diversity represented by Magellan’s library allows a great chemical diversity to be brought to high-throughput screening projects. About 15 000 samples have been obtained from a great variety of biological habitats. More than 5000 bacteria and fungi have been extracted from plants and soils in tropical and temperate zones, marine sediments and marine invertebrates, Antarctic lakes, soil and sea ice. **Martek Biosciences Corporation** develops and sells products from microalgae used in supplements and functional foods for older children and adults. The natural products originate from two strains of algae (**Crypthecodinium cohnii** and **Schizochytrium**), and a fungus (**Mortierella alpina**). **MicroSource Discovery Systems, Inc.** maintains a growing collection of pure natural products from its archives and also from its natural product division, MicroBotanica. An established collaborative program with Perubotanica slr and the Universidad Nacional de la Amazonía Peruana (UNAP) also allows access to the resources of the flora of the Amazon Basin. The emphasis of **Mera Pharmaceuticals, Inc.** is on aquatic microbial plants, *i.e.* small but highly evolved organisms that generate a vast array of complex and unique organic molecules that may offer benefits for humans. Mera Pharmaceuticals has developed a program for accelerating the discovery of new drugs from organic molecules produced by microalgae. **Sequioa Sciences** offers a library of structurally diverse compounds isolated from plants. The proprietary design of these libraries allows the screening of plant compounds at optimum high-throughput screening concentrations without non-drug-like interference and facilitates the identification of potential leads. The database from **TimTec** contains plants collected from various regions of the former USSR, with about 2600 plants available from stock for extraction and 9000 plants available for collection. Furthermore, over 4000 pure natural compounds and natural product-derived compounds are available for immediate testing.

**Asia.** A supplier with one of the largest natural product libraries is **Albany Molecular Research, Inc.** Over 300 000 purified or partially purified natural product-derived samples are included in three different libraries. A collection of 109 000 natural product extracts of microbial fermentation derived from microorganisms, including those from marine environments, are stored in a diversity library. Another library of 170 000 wells consists of pre-fractionated extracts from plants, marine invertebrates and fermented microorganisms derived from the broadest geographical and environmental diversity available. To give a starting point for the identification of active molecules, a collection of over 1000 pure natural products and derivatives thereof with known mechanisms of action is available. Collaborators may access an entire collection or a specified subset of the natural products.

**Australia.** **Cerylid Biosciences** is an Australian bioscience company with a natural product library of over 600 000 extracts from more than 60 000 biotic samples. Cerylid currently has a culture collection of about 25 000 fungi and bacteria from plants and soils in tropical and temperate zones, marine sediment and marine animals, Antarctic lakes, soil and sea ice, composts and animal dung. The environments from which some of the 35 000 plant samples are derived include many different biological habitats. As the amount of each extract that is available would not suffice for developing any hits found in screens so as to identify active compounds, Cerylid does not provide samples of extracts without also selling the corresponding biotic samples from which they were derived. The chemistries of insects and other terrestrial invertebrates are still practically untouched as drug discovery leads. The taxonomical library of **Entocosm Pty Ltd.** contains about 1300 species of insects and terrestrial invertebrates obtained over several years of field collection. Most of them are already available as screen-ready extracts. The company would welcome discussions with commercial partners such as biotechnology and pharmaceutical companies as well as research institutions interested in screening the existing library to identify novel molecules and chemical structures with applications in the field of drugs, and in increasing the existing library, expanding its scope and developing a parallel library based on the unique microbial flora of insects.

**Europe.** A new section in **Ambinter’s** database contains about 500 compounds of marine origin. All compounds are provided on the basis of a budget that is to be defined. Only a few compounds are currently synthesised, but more would be available.

### Table 5: Companies and suppliers of natural compounds in Europe.

<table>
<thead>
<tr>
<th>Name</th>
<th>Homepage</th>
<th>Number of compounds or extracts</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambinter</td>
<td><a href="http://www.ambinter.com/">www.ambinter.com/</a></td>
<td>500–1000</td>
<td>Yes</td>
</tr>
<tr>
<td>Amp-Lab GmbH</td>
<td><a href="http://www.amplab.de/">www.amplab.de/</a></td>
<td>N. A.</td>
<td>Yes</td>
</tr>
<tr>
<td>Analyticon Discovery</td>
<td><a href="http://www.ac-discovery.com/">www.ac-discovery.com/</a></td>
<td>~14 000</td>
<td>Yes</td>
</tr>
<tr>
<td>Instituto Biomar, S. A.</td>
<td><a href="http://www.institutobiomar.com/">www.institutobiomar.com/</a></td>
<td>~15 000</td>
<td>Yes</td>
</tr>
<tr>
<td>InterBioScreen, Ltd GmbH</td>
<td><a href="http://www.ibscreen.com/">www.ibscreen.com/</a></td>
<td>41 000</td>
<td>Yes</td>
</tr>
<tr>
<td>IRIS Biotech GmbH</td>
<td><a href="http://www.iris-biotech.de/">www.iris-biotech.de/</a></td>
<td>~1200</td>
<td>Yes</td>
</tr>
<tr>
<td>Moscow MedChemLabs, Ltd</td>
<td><a href="http://www.mosmedchemlabs.com">www.mosmedchemlabs.com</a></td>
<td>~200</td>
<td>Yes</td>
</tr>
<tr>
<td>Specs</td>
<td><a href="http://www.specs.net/">www.specs.net/</a></td>
<td>~500</td>
<td>Yes</td>
</tr>
<tr>
<td>WERNER BioAgents</td>
<td><a href="http://www.webioage.com/">www.webioage.com/</a></td>
<td>9</td>
<td>Yes</td>
</tr>
</tbody>
</table>
after a customer’s selection. **Amp-Lab GmbH.** Amp-Lab offers first samples of natural products from the intertidal zone of the North Sea, which is characterised by extreme changes in environmental conditions. Organisms that inhabit the intertidal zone are excellently adapted and offer an enormous potential as natural sources of novel biologically active molecules. **Analyticon Discovery GmbH** provides about 8000 natural product small molecules (MEGAbolite) and about 5000 natural product analogue small molecules (NATDiverse). MEGAbolite libraries consist of pure and non-redundant natural compounds from plants and microorganisms, produced either from the clients’ biomaterial collections for their use or from Analyticon’s own collection. All compounds can be purchased from Analyticon. **Instituto Biomar, S. A.** has developed a research program with the objective of exploiting the sea’s biodiversity. Their microorganisms are obtained from sea sediments and marine invertebrates and vertebrates. More than 15 000 extracts (marine unicellular bacteria, actinomycetes and fungi) are available for testing. The company is interested in forming productive collaborations. Over 41 000 unique diverse, rare natural and related compounds are included in the natural compound collection of **InterBioScreen, Ltd.** The sources of natural products range from the former Soviet Union, Asia and Latin America to Europe. In percentage terms, the collection can be said to comprise 30–35% of strictly natural compounds isolated from plants, microorganisms, marine species etc.; approximately 40% are derivatives of natural compounds, i.e. modified alkaloids, terpenoids, flavonoids etc., and the remaining 25–30% are mimetics of strictly natural compounds, e.g. azosteroids, azocoumarins and oxasteroids. Of the whole natural compound library, 60–65% are of plant origin; 5–10% were isolated from microorganisms, about 5% from marine species, and the rest from other natural sources. Compounds in this collection are available in amounts ranging from several grams up to several hundred grams.

**IRIS Biotech GmbH** provides about 1200 natural compounds which are available for experimental testing. Another company is **Moscow MedChemLabs, Ltd.**, which specialises in producing small molecule libraries for high-throughput screening (HTS). The current library contains about 200 natural compounds. **Specs** quotes a diverse set of about 500 isolated or synthesised natural products and derivatives thereof from natural sources such as plants, fungi, bacteria and sea organisms. The selection of natural products consists of pure compounds and no extracts: all are available for experimental testing. **WERNER BioAgents** pursues a mail-order business for a worldwide sale of selected natural products and biologically active compounds, some of which are not available on the market or are of special interest for structural modifications by classic or modern combinatorial methods with the aim of developing leads and potential drugs.

### 4.2 Resellers

The original suppliers of many natural products often remain unidentified, because, after isolation, the natural compounds or extracts are sold to chemical catalogue companies. **Sigma-Aldrich** ([http://www.sigmaaldrich.com](http://www.sigmaaldrich.com)) is a leading supplier of compound libraries and coated multiwell plates to the drug discovery research community. It offers a growing selection of compounds for assay development and HTS. Furthermore, Sigma-Aldrich provides over 40 000 products for the scientific community. Many natural products are available for experimental testing.

**Merck** ([http://www.merck.de](http://www.merck.de)) is a global pharmaceutical and chemical company. More than 15 000 products have been developed, manufactured and marketed in the fields of pharmaceutically active ingredients, drugs, chemicals and laboratory products. All natural products offered by Merck are available for experimental testing.

Another reseller of fine chemicals for medicinal, analytical or biochemical research is **Acros Organics.** They provide a library of more than 18 000 molecules available at different purities for experimental testing.

### 4.3 Manufacturer spanning database

There are several commercial and public databases for natural products,[16–18] but the **SuperNatural database** is the first public resource that contains 3D structures and conformers for about 50 000 natural compounds, derivatives and analogues purchasable from different suppliers ([http://bioinformatics.charite.de/supernatural](http://bioinformatics.charite.de/supernatural)) (Fig. 5).[19] Currently, data from eight suppliers and a comprehensive list of further suppliers for natural compounds and extracts are available. To enable efficient identification of the desired compounds, substructure searches with typical templates have been implemented. The starting points for in silico screenings are about 2500 well-known and classified natural compounds. Possible medical applications can be identified via automatic searches for similar drugs in a free conformational drug database containing World Health Organization (WHO) indications.[19] Similarity searches based on fingerprints and Tanimoto coefficients are implemented in the SuperNatural database.[20] Another way to perform a similarity search is the Marvin Applet, which allows the user to build or import a molecular structure and compare it with the compounds in the database. Furthermore, it is possible to perform a 3D-superposition of two compounds. The algorithm compares all conformers of two compounds to find the best structural alignment.[21] For each natural compound, information is given about various structural and chemical properties. Using embedded screening functions, the SuperNatural database makes it possible to find new active natural compounds with a great diversity of biological activities.

### 5 Conclusions

Although the outstanding diversity and the great potential of natural products are widely accepted, their rational application has been restricted due to the lack of searchable resources for well-characterised compounds. During the last decade, compound collections have turned from exclusive in-house repositories of pharmaceutical companies into publicly available resources affordable to academia.[22] Analogously, open-source software and web technology enable the set-up of chemical databases and query systems, as described here. This process has only just begun, and it can be foreseen that the availability of natural compounds will rise because of technological developments, globalisation, and the negative pressure resulting from the dried drug pipelines of the pharmaceutical industry.
Fig. 5  Screenshot of the SuperNatural Database Web interface. Query results after similarity screening with gingkolide. The compounds can be rotated (left mouse button), different display styles are available (right mouse button) and more detailed information concerning the properties of each structure can be obtained by use of the ‘FULL INFO’ button.

Fig. 6  Workflow proposal for successful identification of natural compounds for experimental testing. Starting from a lead compound with known biological activity, different databases can be screened to identify compounds that potentially have similar activity. Resellers or suppliers can be contacted to obtain substances of interest. YES: found suitable compounds for experimental testing. NO: no suitable compounds found—perform another search in a different database.
In Fig. 6, the different kinds of natural compound databases and suppliers, which are described in this article, are organised as a possible workflow for screenings. This allows for successful identification of pharmaceutically relevant substances and gives support for purchasing them.

6 References