Journal of Ethnopharmacology 140 (2012) 526–534

Contents lists available at SciVerse ScienceDirect

Journal of Ethnopharmacology

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

In-silico studies in Chinese herbal medicines' research: Evaluation of *in-silico* methodologies and phytochemical data sources, and a review of research to date

D.J. Barlow[a,*], A. Buriani[a,b,c], T. Ehrman[a], E. Bosisio[d], I. Eberini[d], P.J. Hylands[a]

[a] *Institute of Pharmaceutical Science, King's College London, Franklin Wilkins Building, 150 Stamford Street, London SE1 9NH, UK*
[b] *Department of Pharmacology and Anesthesiology, Università degli Studi di Padova, Padova, Italy*
[c] *Data Medica, Padova, Italy*
[d] *Department of Pharmacological Science, Università degli Studi di Milano, Milan, Italy*

ARTICLE INFO

Article history:
Received 29 November 2011
Received in revised form 24 January 2012
Accepted 24 January 2012
Available online 2 February 2012

Keywords:
GP-TCM
Traditional Chinese Medicine
In-silico studies
Chinese Herbal medicine

ABSTRACT

The available databases that catalogue information on traditional Chinese medicines are reviewed in terms of their content and utility for *in-silico* research on Chinese herbal medicines, as too are the various protein database resources, and the software available for use in such studies. The software available for bioinformatics and 'omics studies of Chinese herbal medicines are summarised, and a critical evaluation given of the various *in-silico* methods applied in screening Chinese herbal medicines, including classification trees, neural networks, support vector machines, docking and inverse docking algorithms. Recommendations are made regarding any future *in-silico* studies of Chinese herbal medicines.

1. Introduction

With the goal of establishing guidelines for good practice in *in-silico* research into traditional Chinese medicines (TCM), the WP4 working party of the FP7-funded European consortium for GP-TCM (*Good Practice in Traditional Chinese Medicine Research in the Post-genomic Era*) (Uzuner et al., 2010) deemed it appropriate first of all to draw up a catalogue of the various resources currently available, and then to review the use of such resources in the *in-silico* TCM research reported in the literature to date. The computing resources considered included the various databases currently available and the software that has been or might be used in analyses of these data. The various different kinds of databases identified as relevant included those holding ethnobotanical and/or chemical and/or pharmacological and/or toxicological data on the herbs used in Chinese medicine, as well as those that hold data on known or potential molecular targets for the herbal constituents. The software tools considered as relevant included programs that provide for (a) virtual screening of natural product libraries and chemical libraries, (b) pattern recognition, (c) bioinformatics studies relating to TCM and (d) the various data visualisation and statistical analysis packages for proteomics/genomics/metabolomics studies involving TCM.

2. Terminology and communication problems between specialist and non specialist

Preliminary discussions of the *in-silico* WP4 working party focussed on the problems of terminology, and the difficulties associated with performing scientifically sound and fruitful *in-silico* studies in Chinese herbal medicines (CHM) research. For the purposes of the review, *in-silico* studies were defined as those involving virtual screening and/or cheminformatics, but more widely would also include those involving bioinformatics and the various different types of 'omics' studies. The view was universally held that any such studies could not be fruitfully performed by non-specialists – even though the software tools were often easily accessed and easily used by those unfamiliar with computational chemistry. It was unanimously accepted that the performance of *in-silico* studies in CHM research necessitated close interaction between computational chemists and CHM experimentalists. Such interaction is common in the cross-disciplinary areas of molecular pharmacology and structural biology, and *in-silico* CHM research would thus present no exception.

3. Catalogue of existing databases

It is vitally important to any *in-silico* studies of CHM that the associated phytochemical data are available in a format suitable for virtual screening (VS) or other informatics applications. Over recent years the information available for Chinese herbs has increased

* Corresponding author. Tel.: +44 020 7848 4827; fax: +44 020 7848 4800.
E-mail address: dave.barlow@kcl.ac.uk (D.J. Barlow).

0378-8741/$ – see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.jep.2012.01.041

Table 1
Catalogue of databases holding botanical information on TCM herbs and/or the composition of TCM formulae and their usage and/or information on the phytochemical constituents of plants (including those used in CHM).

Database	Content	Source	URL
TCM and chemical structure data			
TCM Database@Taiwan	Chinese medicine database that contains 3-D structural information of TCM constituents – ready for molecular docking simulation (database currently holds 37,170 (32,364 non-duplicate) TCM compounds from 352 TCM)	Chen (2011)	http://tcm.cmu.edu.tw/review.php?menuid=3
China Natural Products Database (CNPD)	Information on Chinese natural products including >40,000 structures. Full set of structures	Shen et al. (2003)	http://www.neotrident.com
3D Structure Database of Components from Chinese Herbs	3D structures (>10,000) from Chinese herbs (>2000), with descriptors and data on clinical uses. Full set of structures	Qiao et al. (2002)	–
Comprehensive Herbal Medicine Information System for Cancer (CHMIS-C)	Integrated information on cancer molecular targets, Chinese herbal recipes and phytochemical constituents. Some structures	Fang et al. (2005)	http://sw16.im.med.umich.edu/chmis-c/
Chinese herbal constituents database (CHCD) and Bioactive plant compounds Database (BPCD)	Information and structures for >13,000 constituents of ≈300 commonly used herbs. >2500 compounds active against ≈80 targets. Full set of structures	Ehrman et al. (2007a,b)	http://www.chemtcm.com/
Ethnopharmacological Database (GPNDB™)	100,000 natural products (3D structures), biological activities, ethnopharmacological data. In-house database of Greenpharma S.A.	Do and Bernard (2006)	http://www.greenpharma.com
Traditional Chinese Medicine Information Database (TCM-ID)	Information on 1197 formulas, 1098 herbs and 9852 constituents in relation to TCM diagnosis and prescription. Some structures	Chen et al. (2006)	http://tcm.cz3.nus.edu.sg/group/tcm-id/tcmid.asp
TCM data only			
Dictionary of Chinese Herbs	TCM herbal formulas, specificity, toxicity and side effects. No structures	–	http://Alternativehealing.org/Chinese_herbs_dictionary.htm
Traditional Chinese Medicine Database System	Bibliographic database (TCMLARS), and Chinese herb database (TCDBASE) in addition to other data	–	http://www.cintcm.com
TCMGeneDIT	Information on relations between TCM and gene regulation, protein-protein interactions and biological pathways	Fang et al. (2008)	http://tcm.lifescience.ntu.edu.tw/
TCM Knowledge Based Grid	TCM herb database, literature database, traditional Tibetan herb database. No structures	–	http://www.cintcm.com
TCM Assistant	TCM herbs, herbal formulas, diseases and patent prescriptions. No structures	–	http://www.tcmassistant.com
Phytochemical data only			
PhytochemDB	Chemical composition of 1278 taxa (>19,000 constituents), including Chinese herbs. No structures	–	http://ukcrop.net/perl/ace/search/PhytochemDB
Dr. Duke's Phytochemical and Ethnobotanical Databases	Information on phytochemicals from >1000 plants, including Chinese herbs. No structures	Duke (1992)	http://www.ars-grin.gov/duke/
Dictionary of Natural Products (DNP)	Major source of chemical information on natural products, including some biological sources, and pharmacological and toxicological data. Full set of structures	–	http://dnp.chemnetbase.com

Table 2
A catalogue of databases holding information on the structures of known and potential protein targets of the phytochemical constituents of herbs used in TCM.

Database	Content	Source	URL
Therapeutic Target Database (TTD)	Information on 1894 targets, 5028 drugs, diseases, and pathways	Zhu et al. (2010)	http://xin.cz3.nus.edu.sg/group/ttd/ttd.asp
Potential Drug Target Database (PDTD)	Information on 830 targets, protein and active site structures, biological functions, diseases and pathways	Gao et al. (2008)	http://www.dddc.ac.cn/pdtd/
Protein Data Bank	Information of 70,000+ protein structures determined by single crystal X-ray diffraction or high filed ^{1}H NMR studies	Berman et al. (2000)	http://www.rcsb.org/pdb

quite dramatically, and the phytochemical classes found within individual plants are often characterised by a large variety of closely related constituents, many of which have been investigated in detail over the past few decades. A catalogue of the databases currently available and relevant to computational studies of CHM is presented in Tables 1 and 2. The databases listed include those holding (a) botanical information on TCM herbs, and the composition of TCM formulae and their usage, (b) information on the phytochemical constituents of plants (including those used in TCM and others), and (c) information on known and potential protein targets of the phytochemical constituents of herbs used in TCM. The databases containing category (a) and/or category (b) data are shown in Table 1; those containing data in category (c) are shown in Table 2.

Of the TCM and natural product databases that have been constructed that provide information relevant to *in-silico* studies of CHM (Table 1), the majority (ten out of the total of fifteen) appear to contain either limited or no chemical data on the phytochemical constituents of the herbs, and would clearly, therefore, have limited utility in any bioinformatics/cheminformatics research.

As regards the protein/target databases holding information relevant to *in-silico* CHM studies (Table 2), the principle resource – since it contains details of *all* publicly available protein 3D structures – is provided by the Protein Data Bank (PDB; Berman et al., 2000). The Therapeutic Target Database (TTD; Zhu et al., 2010) and the Potential Drug Target Database (PDTD; Gao et al., 2008) are more focused resources, providing details of proteins of specific interest in development of new drug therapies and/or studies of mechanisms of drug/CHM phytochemical action.

In assessing the utility of the various databases available, the following considerations were made:

- The availability of a database is important to consider, and although sources which are not publicly available are unlikely to be of interest to most people, such databases might nevertheless be of interest to researchers who are perhaps on the lookout for unusual information, and who might, therefore, be prepared to negotiate with the database authors directly.
- Databases that are 'live' and curated must be held as superior to those that are static and unsupported.
- Databases that are reported in a peer-reviewed scientific publication must likewise be held as superior to those that are undocumented.
- Databases should ideally embody a range of data, including not just chemical structure data but also information on TCM usage, toxicology data, known or predicted target(s), *etc*. It is also to be preferred that the database cater for complex Boolean searches, allowing, for example, for a search for all compounds based a given natural product scaffold found in plants with a specific TCM usage.
- From a practical point of view it is important to consider the 'platform' required by the database (that is, whether it can be mounted and used on computers running UNIX, or Windows, *etc*.). For intensive use, a CD/DVD version of a database is much preferable to a database accessed online (*via* the internet).
- It is important to distinguish between databases where users can output a list or batch file of chemical structures from a search, and those where this is not possible. It is often the case, for example, particularly with commercial databases, that structures have to be downloaded individually, and this must be regarded as a significant limitation/deficiency.
- The nature of the available chemical data held within the database is crucial to consider since this will impose limitations on the type(s) of database search and/or the type(s) of analysis that can be performed. It is important to discriminate between those databases that hold only 2D chemical structure data, those that hold 3D data on single conformers, and those that hold 3D data on multiple conformers. It is also important to discriminate those databases in which compound stereochemistry is unequivocally defined and those where stereochemistry is unspecified.

4. Evaluation criteria for CHM databases

With due consideration of all the points noted above, the WP4 *in-silico* work group formulated a set of quality evaluation criteria/content and functionality check list for CHM databases The check list is shown in Table 3, and feature criteria relating to the database content (C1–C11), the versatility of Boolean searches that are catered for (C12–C14), the nature of the allowed chemical structure output (C15–C16), and the availability of associated reports in peer-reviewed publications (C17). These various features might be used (a) in defining the content to be acquired in creation of a new CHM database, and (b) as a means for users to judge the utility of an existing CHM database. The database features deemed (by the authors) to be of greatest importance include:

- those relating to content – particularly in terms of the chemical structure data held (C1–C11);
- the versatility of Boolean searches that are catered for – the ideal being that searches involving multiple search fields are allowed, one of which is preferably a molecular structure based search field (C12–C14);
- the nature of the allowed chemical structure output – the ideal being that multiple chemical structures can be output, each in 3D format (C15–C16).

In view of the fact that several of the databases catalogued in Table 1 were not available for evaluation by the WP4 *in-silico* work group, it was deemed inappropriate to present any evaluation scores for the databases. Instead, the databases were simply grouped (without ranking) according to their content. Those grouped at the top of Table 1 are those considered most useful for *in-silico* studies of CHM – in that they contain botanical, TCM and phytochemical data, including chemical structures. The databases listed in the second group are those that contain only TCM data, while those grouped last include only non-specific botanical and phytochemical information.

The majority of the databases listed in Tables 1 and 2 do not contain information relating to the toxicology of TCM, principally because the commonly used endpoints such as LD_{50} (either for individual constituents or for herbal extracts) are of limited value in understanding the possible side effects of herbs, particularly those that arise through long term use. While Chinese medicine has always paid close attention to this sort of information, the only English-language toxicological database available is that reported by Bensoussan et al.

5. Types of software for use in computational studies of CHM

As for the review of databases (above), examples of software tools relevant to computational studies of CHM were catalogued using the knowledge and expertise of individual WP4 members together with information presented in the review by Ehrman et al. (2010b).

Several categories of software were deemed of relevance in CHM research. Those of use for virtual screening and/or the identification of potential mechanisms of action of CHM constituents were identified as:

Ligand based screening programs – *Pre-requisite(s) for use*: knowledge of compounds with known activity; *use*: to

Table 3
Quality evaluation criteria for CHM databases.

Evaluation criteria	Description of criteria
C1	Contains information on CHM usage?
C2	Contains botanical information of plants used in CHM?
C3	Contains toxicological data on CHM phytochemicals?
C4	Contains (known) biological activity data on CHM phytochemicals?
C5	Contains (predicted) biological activity data on CHM phytochemicals?
C6	Contains chemical structure data on CHM phytochemicals?
C7	Contains 2D structures of CHM phytochemicals?
C8	Contains 3D structures of CHM phytochemicals?
C9	Contains data on the molecular mechanism(s) of CHM phytochemicals?
C10	Data on >1000 unique CHM phytochemicals
C11	Database is live, curated and updated?
C12	Database can be queried by chemical structure?
C13	Searches with ≥2 search fields possible?
C14	Searches can be performed using combinations of search fields?
C15	Chemical structure(s) can be output in 1D (as SMILES strings)? Chemical structure(s) can be output in 2D? Chemical structure(s) can be output in 3D (as single conformers)? Chemical structure(s) can be output in 3D (as conformer libraries)?
C16	Facility for (batch) output of multiple chemical structures
C17	Database is reported in an article in a peer reviewed journal

identify putatively active compounds; *tools available*: classification/regression trees (including Random Forest), linear discriminant analysis, artificial neural networks, support vector machines.

Pharmacophore programs – Can be either ligand-based (LB), or target-based (TB) (the latter being superior/preferable); *pre-requisite(s) for use*: 3D structures of known ligands to chosen targets (LB), or known 3D structures of target protein(s), and ideally known 3D structure(s) of known complex(es)) (TB); *use*: to identify putative active compounds; *programs available*: LigandScout (Wolber and Langer, 2005), Schrödinger's Phase program (Dixon et al., 2006) and Accelrys's Discovery Studio® Catalyst.

Docking programs – *Pre-requisites for use*: known 3D structure(s) of target proteins; *use*: to 'dock' potential small molecule ligands into protein active sites, optimising their topographical and chemical complementarity, and scoring their interaction. *Programs available*: FlexX (Rarey et al., 1996), Gold (Jones et al., 1997), Dock (Shoichet and Kuntz, 1993), Glide (Halgren et al., 2004), MolDock (Thomsen and Christensen, 2006), AutoDock (Goodsell et al., 1996) and LigandFit (Venkatachalam et al., 2003).

Other relevant types of software tool were identified as:

Pattern recognition software – *Use*: post-screening analyses (involving dimensionality reduction); *algorithms employed*: principle components analysis, multi-dimensional scaling, self-organising maps, and various forms of cluster analysis.

Proteomics and/or genomics data visualisation and analysis tools – *Use:* application specific programs for statistical processing and visualisation of data output from DNA micro-array experiments, MS proteomics experiments, *etc.*

6. Factors for consideration in the choice of *in-silico* tools used in the study of CHM

In consideration of the *in-silico* tools that are readily available for use in CHM research, and the factors that might mitigate against their productive application in such research it should be noted that although the majority of tools are readily available and, for the most part, easy to operate, they are *not* suitable for those lacking the appropriate computing background. *In-silico* tools should ideally *not*, therefore, be used by non-specialists, since almost everyone can obtain an output from a computational program, but – if not carefully produced/checked – the output obtained could be meaningless. That said, manuals/handbooks should of course be available and should be exhaustive. *In silico* approaches and tools that facilitate and complement molecular biology studies constitute a fast moving research area. Direct interaction with experts from that area thus provide the best option for CHM experimentalists to maximise the value of *in-silico* predictions.

In assessing the utility of the various *in-silico* tools available for the study of CHM, the hardware platform is relevant, since high throughput screening approaches require a very efficient use of RAM and a stable operating system, and UNIX/MacOSX/Linux operating systems may be more suitable in this regard than Windows – *albeit* that these are perhaps not as intuitive to set up and use as those implemented under Windows. In addition, it should be noted that the utility of web-based resources – although in many ways very convenient – is critically dependent on their availability.

It is not necessary that the algorithms to be used have been validated specifically for use in CHM research; and users should count it sufficient that the methods they propose to use have been validated on other data sets that are analogous to those to be investigated in the CHM studies. That said, it must also be noted that *in-silico* studies of natural products may not necessarily be conducted in quite the same way as those focused on synthetic chemical libraries – given that the two classes of compound have rather different profiles of chemical properties (Feher and Schmidt, 2003).

7. Reports of target-oriented *in-silico* screening of bioactive material in CHM

The following sections present a précis of the research reported over the period 2000–2010 wherein *in-silico* methods have been used in studies of the bioactive materials in CHM.

7.1. Virtual screening of Chinese herbs for drug discovery

Several studies dealing with the virtual screening (VS) of CHM constituents, most of which have been published within the last five years, have sought to employ various computational methods to identify potential ligands for protein targets of pharmacological/therapeutic interest. The screening techniques employed include pharmacophore search, molecular docking, and screening by means of chemical descriptors and fingerprints.

There are several cases reported in which compounds found in Chinese herbs have been identified as likely inhibitors of specific enzymes or receptors, and these predicted activities have subsequently been confirmed experimentally. (Such studies are clearly of greater interest than those where the *in-silico* predictions are not subsequently tested by experiment.)

It is particularly of interest to note here that for several of the reported studies significant information leading to the selection of appropriate plant material was derived from a consideration of the plants' use(s) in TCM – thus testifying to the importance of ethnopharmacological data in VS studies.

Molecular docking studies have shown that sieboldigenin binds to the active site of soybean lipoxygenase (SLOX), and this compound was later found to inhibit SLOX with an IC_{50} of 38 μM, as well as reducing carageenan-induced paw oedema (Khan et al., 2009). Sieboldigenin is a spirostane sterol found in various species of *Smilax* including *Smilax glabra* (tufuling) which is used in TCM as a 'heat clearing' herb, employed largely for arthritic joint pain and skin disease.

In like manner, leucovorin was discovered as a potential anti-HIV agent by screening Chinese natural products using a molecular fingerprint based on the HIV protease inhibitor, saquinavir. Molecular dynamics studies subsequently confirmed the favourable binding of this compound to the protease active site (Gao et al., 2007).

Combined molecular fingerprint studies and docking were also responsible for the discovery of aurantiamide acetate, from *Artemisia annua* (qinghao), as an inhibitor of severe acute respiratory syndrome coronavirus main proteinase (SARS-$CoVM^{pro}$; Wang et al., 2007).

In a search for ways to prevent surface biofilm formation, Zeng et al. (2008) performed VS studies of quorum sensing inhibitors of *Pseudomonas aeruginosa*. Fifty-one TCM compounds with known antibacterial activity were docked into the active site of the transcription activation factor TraR. *In vitro* screening of eight of the high scoring compounds subsequently showed that *Pseudomonas aeruginosa* growth was effectively inhibited by baicalein and that this compound acted synergistically with ampicillin.

Rollinger et al. used protein-based pharmacophores to screen for natural product inhibitors of acetylcholinesterase (AChE) and cyclooxygenase (COX). By this means they identified scopoletin as a potential AChE inhibitor and sanggenons as a potential COX inhibitor, and these predictions were then proved correct through *in vitro* studies (Rollinger et al., 2004, 2005).

In other VS studies of CHM consitituents the screening was either carried out without reference to TCM usage or else it afforded predictions of potential ligands which were not then substantiated experimentally.

Paoletta et al. (2008) selected three Chinese herbal constituents which were identified as potential aromatase inhibitors by VS based on molecular descriptors, using known phytochemical inhibitors in training, and then docked each of these into the active site of human aromatase. The three compounds studied were the flavonoids myricetin, liquiritigenin and gossypetin. Liquiritigenin was shown to have a lowest binding energy. Subsequent *in vitro* experiments confirmed these *in-silico* predictions, with liquiritigenin showing the highest potency – exhibiting an IC_{50} of 0.34 μM and thus presenting a 10-fold increase in inhibitory activity over the first generation inhibitor, aminoglutethimide.

Guided by the TCM use of *Epimedium* spp. as tonics for 'yang invigoration', Chen et al. sought to identify constituents of these herbs which might mimic the phosphodiesterase 5 (PDE5) inhibitory effect of sildenafil (Chen et al., 2009). They used pharmacophores, docking and QSAR methods. Several potential inhibitors were identified and some of these – similar in structure to ES-03b – showed docking scores which were comparable to that shown by another known PDE5 inhibitor taladafil (Chen et al., 2009).

Three similar studies have been reported involving *in-silico* screening of Chinese herbs for compounds active against various kinases that present as targets for cancer therapy – these being the aurora-A kinase, polo-like kinases and KDR kinase (Yu et al., 2007; Deng et al., 2008; Wang et al., 2008). In all three studies, a similar methodology was employed, using pharmacophores derived from known inhibitors which were used to screen the CNPD (Shen et al., 2003) (in the case of the aurora-A and polo-like kinase studies) and the TCMD (Qiao et al., 2002) (for the KDR kinase study). The various hits obtained were then filtered using the Lipinski rules, regression analysis predictions of IC_{50}, and molecular docking to relevant protein targets. In the case of the KDR kinase, a pterocarpan glucoside was identified as a hit from the TCMD, and this was found by experiment to have a dissociation constant of 30 μM.

7.2. Searching for multi-target ligands in Chinese herbs

Virtual screening studies have also been performed in the search for single phytochemicals with the potential to inhibit a variety of functionally/pathologically related targets, and also in searches for 'cocktails' of ligands, each of which is predicted to inhibit one of the targets in a series of (pathologically related) targets of interest.

Huang et al. (2007) have shown that several constituents of plants used in the Chinese formula *Xuefu zhuyu tang* could present as inhibitors of targets relevant to the treatment of cardiovascular disease. 'Drug-like' compounds with acceptable ADME profiles were first identified by Lipinski filtering, and these were then docked to a number of targets with known significance in cardiovascular disease, including renin, angiotensin-converting enzyme (ACE), VEGFR, 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) and P-glycoprotein (Pgp). A total of 283 compounds were obtained as possible inhibitors of all the selected protein targets. Out of the 11 herbs studied, the majority were identified as possible inhibitors of more than one target. Moreover, 10 of the herbs were shown to possess constituents with the potential to inhibit two or more targets, and 50% of these showed the potential to inhibit five targets.

Ehrman et al. (2007b) have used multiple decision trees to fish for compounds in Chinese herbs on the basis of structural similarity to known phytochemical inhibitors of targets involved in several disease processes, including inflammation, HIV and diabetes; they used Random Forest models to discriminate phytochemical inhibitors from 8000+ compounds found in 240 herbs. Target compounds that could not be distinguished from the known inhibitors for a given target on the basis of three different RF models (each employing a different set of molecular descriptors), were flagged as potential inhibitors of that target. Despite stringent selection criteria, around 62% of the plant species were hit at least once, with 50% of these predicted to have constituents active against two or more targets. Literature searches performed to corroborate these predictions, subsequently showed that 29% of the predicted classes within these herbs were supported by recent data, with ~80% of publications providing new information distinct from that used in constructing the RF models.

The same research team later reported protein-based screening studies (Ehrman et al., 2010a), employing multiple PDB ligand–receptor complexes, to identify possible ligands four major anti-inflammatory targets – COX, p38 MAP kinase (p38 MAPK), c-Jun terminal NH_2 kinase (JNK), and phosphodiesterase 4 (PDE4). A conformational database of Chinese herbal constituents was first screened using LigandScout pharmacophores (derived through analyses of ~60 PDB entries). Resulting hits were subsequently docked into the relevant target proteins, and the docking scores compared against the median binding energy for the crystal ligands for that target. Analysis of the cumulative scores for compounds from distinct phytochemical classes within each herb led to the identification of the 100 plants most likely to yield mimics comparable in potency to the known PDB ligands. Further analysis revealed that the Chinese formulas in which these herbs were

commonly employed, are very often used in the treatment of inflammatory conditions.

7.3. Inverse docking and target fishing

Inverse docking and target fishing are *in-silico* screening techniques that are used to identify protein targets *ab initio*. These techniques can thus be viewed as particularly pertinent in studies of phytochemicals. Indeed, two of the computer programs available have been specifically developed Chinese medicine in mind, to aid in elucidating the targets of TCM constituents. The first of these, INVDOCK (Chen and Zhi, 2001), is similar as regards docking methodology to Kuntz's DOCK program (Shoichet and Kuntz, 1993); it makes use of a database of protein cavities derived from PDB entries, and query compounds are tested for fit against some or all of these cavities to identify potential targets. The algorithm uses a shape matching alignment of the query compound to each protein cavity, with torsion optimisation and energy minimisation using the AMBER force field then employed to refine the fit. Protein cavities are defined using overlapping spheres following the approach of Kuntz et al. (Shoichet and Kuntz, 1993), and the database now holds ~9000 entries. Validation of the methodology, using phytochemicals with known targets (including ginsenoside Rg1, baicalin, quercetin, emodin, catchin and allicin), identified ~50% of the targets for which experimental data exist (Chen and Ung, 2002; Chen et al., 2003). INVDOCK has also been used to identify possible targets of the cytotoxic compound ganoderic acid D from the Chinese fungus *Ganoderma lucidum* (lingzhi), following investigation of its effects on expression of 21 different proteins. The results obtained suggest that it might bind eight of these receptors directly (Yue et al., 2008).

TarFisDock (Li et al., 2006) is another inverse docking program, and makes use of a similar methodology to INVDOCK. This program uses a protein database constructed from the PDB (with each entry comprising residues within 6.5 Å of a known ligand), with the active site defined by spheres as in INVDOCK. The docking of query compounds against the database sites is again based on Kuntz's DOCK algorithm. TarFisDock (http://www.dddc.ac.cn/tarfisdock/) is designed for use with the PDTD (Gao et al., 2008) and currently holds over 1100 entries covering 830 known or potential drug targets, each target being linked to information on biological function and disease. Users can access the program *via* the internet, with query compounds uploaded, and the database then searched for candidate targets *via* a web interface available at http://www.dddc.ac.cn/pdtd/. Successful applications of the program reported to date include the identification of a *Helicobacter pylori* target for the natural product N-*trans*-caffeoyltyramine in the enzyme peptide deformylase – this compound, and one if its derivatives, subsequently being shown to be potent inhibitors of the enzyme with IC_{50} values of 10.8 and 1.25 μM, respectively (Cai et al., 2006).

Greenpharma's 'reverse pharmacognosy' represents another inverse docking approach (Do and Bernard, 2006), very similar to INVDOCK and TarFisDock, but differing in its use of an in-house database of ethnopharmacological data. Where pharmacognosy seeks to identify bioactive compounds from plants based on extraction and assay, *reverse* pharmacognosy seeks to screen targets and diseases from individual compounds. In part this is achieved through use of compound and target databases linked by the docking program *Selnergy* – the use of which successfully identified ε-viniferin as an inhibitor of phosphodiesterase 4 (ε-viniferin being a stilbene found in *Sophora* spp. that are commonly used in TCM).

An alternative, related strategy used in identifying targets of phytochemical involves the screening of query compounds against pharmacophore models of PDB ligands. Such screening is significantly faster than molecular docking, and thus provides a rapid means to filter out compounds that are not direct mimics of the ligands from which the pharmacophore model is generated. This approach does have the drawback, however, that compounds that might find alternative modes of binding to the selected protein targets will not be identified as hits. The first use of this type of PDB-based pharmacophore method for target fishing for plant constituents was reported by Rollinger et al. (2009). In this study, 16 constituents of *Ruta graveolens* were screened against a database containing 2208 pharmacophores. *In vitro* screening against three targets – AChE, HRV coat protein and cannabinoid receptor type-2 (CB_2) – showed a close degree of congruity between the best hits and their respective IC_{50} values.

7.4. In-silico tools for proteomics, genomics and metabonomics data visualisation and analysis

The advent of information rich "omic" techniques has provided powerful new research tools for use in the study of CHM. Using genomic, proteomic or metabonomic analysis, it has become possible to examine simultaneous molecular effects that result when using chemical mixtures as in CHM, affording new opportunities to study biological effects holistically, rather than by following the classical reductionistic approach – a new vision which goes beyond single molecule pharmacology and target specificity, embracing the entire equilibrium of a biological system undergoing simultaneous perturbations. Such an approach can be considered more feasible to investigate the effects of multichemical mixtures on a plurality of biological targets, as in CHM formulae (for a review see Buriani et al., 2012).

Using such a strategy a multitude of experimental data can be obtained that need to be handled both in terms of multichemical identification and pathophysiological correlations. With the help of bio-informatics, specific software and databases have been developed in order to exploit the large amount of raw data derived from omics techniques and allow correlation of the multiple components of a given phytocomplex with its biological effects (Wang et al., 2011, 2012). This systems biology approach thus integrates powerful information rich technologies, computational tools and knowledge bases, making it possible to establish links between molecular patterns, biological functions and a wide range of human diseases and pharmacological interventions (Auffray et al., 2009). In particular, high-throughput omics techniques require several bioinformatics and knowledge-assembly tools for data processing, analysis, integration and interpretation using a top-down system biology approach (Naylor and Chen, 2010).

When using omics techniques, raw data are produced mostly from microarrays and mass spectrometry or NMR, depending on the technique used, and subsequent to data processing – which often needs multivariate analysis – sets of qualitative and quantitative data are obtained indicating patterns of induced molecular perturbations. *In-silico* tools can be used to assist in molecular identification, and databases as well as specific search software are available to assist in this task. These include the METLIN and MassBank databases for metabolomics (Sana et al., 2008; Wolf et al., 2010; Horai et al., 2010), and human protein and peptide databases for proteomics (Choi et al., 2008; Klimek et al., 2008; Ding et al., 2008; Wang et al., 2011, 2012).

Following analysis, the affected molecular networks can be identified and the observed perturbations correlated to a given pharmacotoxicological effect. This last correlation step is of utmost importance in order to exploit the full potential of information rich data resulting from the application of omics techniques. Several databases have been made available that can be used for guided analysis, while others can be applied to untargeted and genome-wide or metabolome-wide association studies (Smith and Newton-Cheh, 2009; Chadeau-Hyam et al., 2010). These databases

allow data mining, integrated correlations, and modelling of biochemical pathways, thus providing a platform of translational medicine that can also be used for CHM research (Li et al., 2010). Molecular databases can be thus used to identify unknown metabolites or peptides, to identify structure from known elemental composition, or to determine the biological function of a specific molecule.

The catalogue of databases and software developed for genome-scale metabolic reconstructions is ever-increasing, and many of these utilities are freely available for academic use. The databases and software that are currently available have been extensively reviewed elsewhere (Kouskoumvekaki and Panagiotou, 2011; Fiehn et al., 2011; Kamath et al., 2011), and the following section provides just a few selected examples.

- The *Human Metabolome Database* (HMDB) is considered the most complete bioinformatics and chemoinformatics medical information database (Kouskoumvekaki and Panagiotou, 2011) with more than 6800 metabolite entries; it is fully searchable with tools for viewing, sorting metabolites and pathways, and provides customised, clickable metabolic maps, as well as disease information (Wishart et al., 2009).
- BiGG, *Biochemical Genetic and Genomic knowledge base*, is a metabolic reconstruction of human metabolism using both genetics and literature-based data to assess whether a reaction is present, providing confidence level data and Boolean relationships between genes, proteins and reactions (Schellenberger et al., 2010).
- The *Small Molecule Pathway Database* (SMPDB) is an interactive, visual database containing a limited number of small-molecule pathways found in humans, most of which cannot be found in any other pathway database. The SMPDB is used in metabolomics, transcriptomics and proteomics studies (Frolkis et al., 2010).
- *Genomics Portals* provides a database of human, mouse, and rat genomics data with basic analytical visualisation tools (Shinde et al., 2010).
- Examples of proteomic databases include GPDE, *Griss Proteomics Database Engine*, PRIDE, *Proteomics IDEntifications database*, and *PeptideAtlas* (Griss et al., 2011).
- Some databases are dedicated to specific fields of study. For example, T3TB, the Toxin and Toxin-Target Database, provides descriptions, mechanisms of action, and information on toxins and toxin-targets (Lim et al., 2010). CEBS, Chemical Effects in Biological Systems, is an integrated public repository focused on toxicogenomics, integrating data describing histopathological and biological measures with microarray and proteomics data (Waters et al., 2008).

Even though the introduction of omics techniques in CHM research is relatively recent and is not yet widely disseminated among scientists in the field, meaningful examples of the application of a systems biology approach can be found in the current literature, and the results are suggestive of a great potential for application of omics and related bioinformatic techniques in the study of both Chinese syndromes and phytocomplexes from traditional Chinese medicine.

The "Connectivity Map" (CMAP) has been successfully used to examine the biological effects of Chinese formulae. CMAP is a database used for the identification of functional connections between drugs, genes and diseases (Lamb, 2007). Using gene expression microarrays it was possible to compare the gene expression signatures of MCF-7 cells treated with the Chinese medicine formulae Si-Wu-Tang, with those of 1309 compounds present in the CMAP database, thus allowing the matching of the formula-induced pattern with that of estradiol-treated cells, suggesting a phytoestrogenic effect (Wen et al., 2011).

Using a similar approach, Cheng et al. tested the effects of 15 different Chinese herbal formulae on mouse kidney and liver, analysing DNA expression on microarrays. The molecular signatures of the formulae were compared with those from drugs, or disease states in three different databases. In addition to using CMAP to compare the formulae-regulated genes with those of known drugs, the *Genetic Association Database* was used for comparing with gene expression patterns from disease states (Becker et al., 2004) and the *Environment, Drugs and Gene Expression* (EDGE) database to compare with the effect patterns resulting from exposures to chemicals (Hayes et al., 2005). The results showed numerous matches allowing predictions on both the pharmacological effects and the toxicological potential of the formulae (Li et al., 2010).

Transcriptomics was also used to analyse the effects of San-Huang-Xie-Xin-Tang and its herbal components on the HepG2 human hepatoma cell line. Using the Molecular Signature Database website, TIGR Multiexperiment Viewer and BiblioSphere Pathway Edition software, it was found that the formula displayed an antiproliferation pattern *via* the p53 pathway, signalling an effect that was also shown to be attributable primarily to a single herbal component (Cheng et al., 2008).

Pro-apoptotic and pro-inflammatory pathways were shown to be activated when using the Chinese herbal formula PHY906 on tumour bearing mice. In this case, gene functions were congregated into pathways and networks using Ingenuity Pathways Analysis (IPA) (Wang et al., 2011, 2012).

IPA was also used to analyse microarrays for gene expression in mice fed with *Ginkgo biloba*. The results showed alterations in several pathways, suggesting implications of both pharmacotoxicological and pharmacokinetic interest when using *Ginkgo biloba* (Guo et al., 2010).

A metabonomic study was performed analysing urine collected from rats fed with Morning Glory (*Pharbitis purpurea*) seeds, to study the nephrotoxicity induced by this plant – commonly used in TCM. Urine was analysed by ultra-performance liquid chromatography/mass spectrometry (UPLC/MS) and mass spectra were interpreted with the aid of biochemical databases, such as KEGG, METLIN, and SciFinder and eight endogenous metabolites as biomarkers were found to be significantly changed in the urine (Ma et al., 2010).

The 'Kidney-Yang Deficiency syndrome' induced by high doses of hydrocortisone and the therapeutic effects of the dried rhizome of *Drynaria fortunei*, traditionally used for treating the syndrome, were studied using a metabonomic approach. Urine samples were analysed with UPLC/MS and identification of potential markers, was sought using the HMDB, METLIN, Massbank, PubChem and KEGG databases. The work provided indications on biochemical pathways which correlated with the syndrome-related dysfunctions in energy metabolism as well as the efficacy of the *Drynaria* treatment (Lu et al., 2011).

A proteomics analytical approach was used to study the effect of *Salvia miltiorrhiza* on vascular atherosclerotic lesions, using an *in vitro* smooth muscle cell model. Two-dimensional electrophoresis coupled with mass spectrometry was used to analyse proteins under oxidative stress and treatment with the herbal extract. The MetaCore software was then used to investigate the biological pathways involved and it was found that *Salvia miltiorrhiza* exerts its protective effect by acting on the scavenging of ROS and subsequent modulation of protein carbonylation to inhibit cell proliferation (Hung et al., 2010).

A large amount of experimental data generated by high-throughput techniques is available through various public repositories. Knowledge about transcriptional regulation, molecular interaction networks and metabolic pathways is rapidly expanding thanks to the innovative system biology approach,

information rich omics technology and computer assisted analysis. In this rapidly evolving context the experimental study of multiple target, multichemical pharmacologic effects with phytocomplexes as in CHM, can greatly benefit from this new holistic scientific vision.

8. Conclusions

In-silico tools are best applied in CHM research as a means: to seek out potential mechanisms of action of their constituents; to identify putative new leads for drugs; and to summarise and/or visualise the complex patterns embedded within the output generated through associated 'omics studies. Given the complexity of CHM, and the ensuing difficulties in performance of the relevant experimental work, *in-silico* studies clearly offer an economical and efficient way of exploring the problem landscape and they can thus aid in the formulation of appropriate hypotheses for subsequent testing in (*in vitro* and/or *in vivo*) laboratory studies.

The utility and choice of *in-silico* screening tools are dependent on the reason(s) for their use and/or the nature of the output to be generated. The knowledge of the computational specialist is essential in guiding decisions made here, and it is imperative that the tools should be applied in CHM research through close collaboration between computational specialists and CHM experimentalists.

References

Auffray, C., Chen, Z., Hood, L., 2009. Systems medicine: the future of medical genomics and healthcare. Genome Medicine 1, 2.

Becker, K.G., Barnes, K.C., Bright, T.J., Wang, S.A., 2004. The genetic association database. Nature Genetics 36, 431–432.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. Nucleic Acids Research 28, 235–242.

Buriani, A., Garcia-Bermejo, M.L., Bosisio, E., Xu, Q., Li, H., Dong, X., Simmonds, M.S.J., Carrara, M., Tejedor, N., Lucio-Cazana, J., Hylands, P.J., 2012. Omic techniques in systems biology approaches to traditional Chinese medicine research: present and future. Journal of Ethnopharmacology 140, 535–544.

Chadeau-Hyam, M., Ebbels, T.M.D., Brown, I.J., Chan, Q., Stamler, J., Huang, C.C., Daviglus, M.L., Ueshima, H., Zhao, L., Holmes, E., Nicholson, J.K., Elliott, P., De Iorio, M., 2010. Metabolic profiling and the metabolome-wide association study: significance level for biomarker identification. Journal of Proteome Research 9, 4620–4627.

Cai, J.H., Han, C., Hu, T.C., Zhang, J., Wu, D.L., Wang, F.D., Liu, Y.Q., Ding, J.P., Chen, K.X., Yue, J.M., Shen, X., Jiang, H.L., 2006. Peptide deformylase is a potential target for anti-*Helicobacter pylori* drugs: reverse docking, enzymatic assay, and X-ray crystallography validation. Protein Science 15, 2071–2081.

Chen, C.Y.-C., 2011. TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. PLoS One 6, e15939, doi:10.1371/journal.pone.0015939.

Chen, Y.Z., Zhi, D.G., 2001. Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. Proteins 43, 217–226.

Chen, X., Zhou, H., Liu, Y.B., Wang, J.F., Li, H., Ung, C.Y., Han, L.Y., Cao, Z.W., Chen, Y.Z., 2006. Database of traditional Chinese medicine and its application to studies of mechanism and to prescription validation. British Journal of Pharmacology 149, 1092–1103.

Chen, C.Y., Chang, Y.H., Bau, D.T., Huang, H.J., Tsai, F.J., Tsai, C.H., Chen, C.Y.C., 2009. Discovery of potent inhibitors for phosphodiesterase 5 by virtual screening and pharmacophore analysis. Acta Pharmacologica Sinica 30, 1186–1194.

Chen, Y.Z., Ung, C.Y., 2002. Computer automated prediction of potential therapeutic and toxicity protein targets of bioactive compounds from Chinese medicinal plants. American Journal of Chinese Medicine 30, 139–154.

Chen, X., Ung, C.Y., Chen, Y.Z., 2003. Can an *in silico* drug–target search method be used to probe potential mechanisms of medicinal plant ingredients? Natural Products Reports 20, 432–444.

Choi, H., Ghosh, D., Nesvizhskii, A.I., 2008. Statistical validation of peptide identifications in large-scale proteomics using the target-decoy database search strategy and flexible mixture modeling. Journal of Proteome Research 7, 286–292.

Cheng, W.Y., Wu, S.L., Hsiang, C.Y., Li, C.C., Lai, T.Y., Lo, H.Y., Shen, W.S., Lee, C.H., Chen, J.C., Wu, H.C., Ho, T.Y., 2008. Relationship between San-Huang-Xie-Xin-Tang and its herbal components on the gene expression profiles in HepG2 cells. American Journal of Chinese Medicine 36, 783–797.

Deng, X.Q., Wang, H.Y., Zhao, Y.L., Xiang, M.L., Jiang, P.D., Cao, Z.X., Zheng, Y.Z., Luo, S.D., Yu, L.T., Wei, Y.Q., Yang, S.Y., 2008. Pharmacophore modelling and virtual screening for identification of new Aurora-A kinase inhibitors. Chemical Biology & Drug Design 71, 533–539.

Ding, Y., Choi, H., Nesvizhskii, A.I., 2008. Adaptive discriminant function analysis and reranking of MS/MS database search results for improved peptide identification in shotgun proteomics. Journal of Proteome Research 7, 4878–4889.

Dixon, S.L., Smondyrev, A.M., Rao, S.N., 2006. PHASE: a novel approach to pharmacophore modeling and 3D database searching. Chemical Biology & Drug Design 67, 370–372.

Do, Q.T., Bernard, P., 2006. Reverse pharmacognosy: a new concept for accelerating natural drug discovery. In: Khan, M.T.H., Ather, A. (Eds.), Lead Molecules from Natural Products. Elsevier, pp. 1–20.

Duke, J.A., 1992. Handbook of Phytochemical Constituents of GRAS Herbs and Other Economic Plants. CRC Press, Boca Raton.

Ehrman, T.M., Barlow, D.J., Hylands, P.J., 2007a. Phytochemical databases of Chinese herbal constituents and bioactive plant compounds with known target specificities. Journal of Chemical Information & Modeling 47, 254–263.

Ehrman, T.M., Barlow, D.J., Hylands, P.J., 2007b. Virtual screening of Chinese herbs with Random Forest. Journal of Chemical Information & Modeling 47, 264–278.

Ehrman, T.M., Barlow, D.J., Hylands, P.J., 2010a. In-silico search for multi-target anti-inflammatories in Chinese herbs and formulas. Bio-organic & Medicinal Chemistry 18, 2204–2218.

Ehrman, T.M., Barlow, D.J., Hylands, P.J., 2010b. Phytochemical informatics & virtual screening of herbs used in Chinese medicine. Current Pharmaceutical Design 16, 1785–1798.

Fang, X., Shao, L., Zhang, H., Wang, S., 2005. CHMIS-C: a comprehensive herbal medicine information system for cancer. Journal of Medicinal Chemistry 48, 1481–1488.

Fang, Y.C., Huang, H.C., Chen, H.H., Juan, H.F., 2008. TCMGeneDIT: a database for associated traditional Chinese medicine gene and disease information using text mining. BMC Complementary & Alternative Medicine 14, 8–18.

Feher, M., Schmidt, J.M., 2003. Property distributions: differences between drugs natural products and molecules from combinatorial chemistry. Journal of Chemical Information & Computer Science 43, 218–227.

Fiehn, O., Barupal, D.K., Kind, T., 2011. Extending biochemical databases by metabolomic surveys. Journal of Biological Chemistry 286, 23637–23643.

Frolkis, A., Knox, C., Lim, E., Jewison, T., Law, V., Hau, D.D., Liu, P., Gautam, B., Ly, S., Guo, A.C., Xia, J., Liang, Y., Shrivastava, S., Wishart, D.S., 2010. SMPDB: the small molecule pathway database. Nucleic Acids Research 38, D480–D487.

Griss, J., Haudek-Prinz, V., Gerner, C., 2011. GPDE: a biological proteomic database for biomarker discovery and evaluation. Proteomics 11, 1000–1004.

Gao, W.N., Li, Y., Zhang, R., Gao, H., Xu, W.R., Li, A.X., Du, Q.S., Zhang, X., Wei, D.Q., 2007. Screening of new HIV inhibitors based on the database of traditional Chinese medicine. Yaoxue Xuebao 2006, 241–246.

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

Goodsell, D.S., Morris, G.M., Olson, A.J., 1996. Automated docking of flexible ligands: applications of AutoDock. Journal of Molecular Recognition 9, 1–5.

Guo, L., Mei, N., Liao, W., Chan, P.C., Fu, P.P., 2010. *Ginkgo biloba* extract induces gene expression changes in xenobiotics metabolism and the Myc-centered network. OMICS 14, 75–90.

Halgren, T.A., Murphy, R.B., Friesner, R.A., Beard, H.S., Frye, L.L., Pollard, W.T., Banks, J.L., 2004. Glide: a new approach for rapid, accurate docking and scoring: 1. Method assessment of docking accuracy. Journal of Medicinal Chemistry 47, 1739–1749.

Hayes, K.R., Vollrath, A.L., Zastrow, G.M., McMillan, B.J., Craven, M., Jovanovich, S., Rank, D.R., Penn, S., Walisser, J.A., Reddy, J.K., Thomas, R.S., Bradfield, C.A., 2005. EDGE: a centralized resource for the comparison, analysis, and distribution of toxicogenomic information. Molecular Pharmacology 67, 1360–1368.

Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K., Tanaka, S., Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda, F., Sawada, Y., Hirai, M., Nakanishi, Y., Ikeda, H., Akimoto, K., Maoka, N., Takahashi, H., Ara, T., Sakurai, N., Suzuki, H., Shibata, D., Neumann, S., Iida, T., Tanaka, K., Funatsu, K., Matsuura, F., Soga, T., Taguchi, R., Saito, K.i., Nishioka, T., 2010. MassBank: a public repository for sharing mass spectral data for life sciences. Journal of Mass Spectrometry 45, 703–714.

Huang, Q., Qiao, X.B., Xu, X.J., 2007. Potential synergism and inhibitors to multiple target enzymes of Xuefu Zhuyu decocotion in cardiac disease therapeutics: a computational approach. Bioorganic & Medicinal Chemistry Letters 17, 1779–1783.

Hung, Y.-C., Wang, P.-W., Pan, T.-L., 2010. Functional proteomics reveal the effect of *Salvia miltiorrhiza* aqueous extract against vascular atherosclerotic lesions. Biochimica et Biophysica Acta 1804, 1310–1321.

Jones, G., Willett, P., Glen, R.C., Leach, A.R., Taylor, R., 1997. Development and validation of a genetic algorithm for flexible docking. Journal of Molecular Biology 267, 727–748.

Kamath, K.S., Vasavada, M.S., Srivastava, S., 2011. Proteomic databases and tools to decipher post-translational modifications. Journal of Proteomics 75, 127–144.

Khan, I., Nisar, M., Ebad, F., Nadeem, S., Saeed, M., Khan, H., Ullah, S., Khuda, F., Karim, N., Ahmad, Z., 2009. Anti-inflammatory activities of sieboldigenin from *Smilax china* Linn.: experimental and computational studies. Journal of Ethnopharmacology 121, 175–177.

Klimek, J., Eddes, J.S., Hohmann, L., Jackson, J., Peterson, A., Letarte, S., Gafken, P.R., Katz, J.E., Mallick, P., Lee, H., Schmidt, A., Ossola, R., Eng, J.K., Aebersold, R., Martin, D.B., 2008. The standard protein mix database: a diverse dataset to assist in the production of improved peptide and protein identification software tools. Journal of Proteome Research 7, 96–103.

Kouskoumvekaki, I., Panagiotou, G., 2011. Navigating the human metabolome for biomarker identification and design of pharmaceutical molecules. Journal of Biomedicine and Biotechnology 2011, 19, Article ID 525497.

Lamb, J., 2007. The connectivity map: a new tool for biomedical research. Nature Reviews 7, 54–60.
Li, H.L., Gao, Z.T., Kang, L., Zhang, H.L., Yang, K., Yu, K.Q., Luo, X.M., Zhu, W.L., Chen, K.X., Shen, J.H., Wang, X.C., Jiang, H.L., 2006. TarFisDock: a web server for identifying drug targets with docking approach. Nucleic Acids Research 34, 219–224.
Li, C.Y., Hou, Y.C., Lee Chao, P.D., Shia, C.S., Hsu, I.C., Fang, S.H., 2010. Potential ex vivo immunomodulatory effects of San-Huang-Xie-Xin-Tang and its component herbs on mice and humans. Journal of Ethnopharmacology 127, 292–298.
Lim, E., Pon, A., Djoumbou, Y., Knox, C., Shrivastava, S., Guo, A.C., Neveu, V., Wishart, D.S., 2010. T3DB: a comprehensively annotated database of common toxins and their targets. Nucleic Acids Research 38, D781–D786.
Lu, X., Xiong, Z., Li, J., Zheng, S., Huo, T., Li, F., 2011. Metabonomic study on 'Kidney-Yang Deficiency syndrome' and intervention effects of *Rhizoma drynariae* extracts in rats using ultra performance liquid chromatography coupled with mass spectrometry. Talanta 83, 700–708.
Ma, I., Bi, K., Zhang, M., Su, D., Fan, X., Ji, W., Wang, C., Chen, X., 2010. Metabonomic study of biochemical changes in the urine of Morning Glory seed treated rat. Journal of Pharmaceutical and Biomedical Analysis 53, 559–566.
Naylor, S., Chen, J.Y., 2010. Unraveling human complexity and disease with systems biology and personalized medicine. Personalized Medicine 7, 275–289.
Paoletta, S., Steventon, G.B., Wildeboer, D., Ehrman, T.M., Hylands, P.J., Barlow, D.J., 2008. Screening of herbal constituents for aromatase inhibitory activity. Bioorganic & Medicinal Chemistry 16, 8466–8470.
Qiao, X., Hou, T.J., Zhang, W., Guo, S.L., Xu, X.J., 2002. A 3D structure database of components from Chinese traditional medicinal herbs. Journal of Chemical Information & Computer Science 42, 481–489.
Rarey, M., Wefing, S., Lengauer, T., 1996. Placement of medium-sized molecular fragments in active sites of proteins. Journal of Computer-aided Molecular Design 10, 41–54.
Rollinger, J.M., Hornick, A., Langer, T., Stuppner, H., Prast, H., 2004. Acetylcholinesterase inhibitory activity of scoploin and scopoletin discovered by virtual screening of natural products. Journal of Medicinal Chemistry 47, 6248–6254.
Rollinger, J.M., Bodensieck, A., Seger, C., Ellmerer, E.P., Bauer, R., Langer, T., Stuppner, H., 2005. Discovering COX-inhibiting constituents of *Morus* root bark: activity-guided versus computer-aided methods. Planta Medica 71, 399–405.
Rollinger, J.M., Schuster, D., Danzl, B., Schwaiger, S., Markt, P., Schmidtke, M., Gertsch, J., Raduner, S., Wolber, G., Langer, T., Stuppner, H., 2009. In silico target fishing for rationalized ligand discovery exemplified on constituents of *Ruta graveolens*. Planta Medica 75, 195–204.
Sana, T.R., Roark, J.C., Li, X., Waddell, K., Fischer, S.M., 2008. Molecular formula and METLIN personal metabolite database matching applied to the identification of compounds generated by LC/TOF-MS. Journal of Biomolecular Techniques 19, 258–266.
Schellenberger, J., Park, J.O., Conrad, T.M., Palsson, B.Ø., 2010. BiGG: a biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. BMC Bioinformatics 11, 213.
Shen, J.H., Xu, X.Y., Cheng, F., Liu, H., Luo, X.M., Shen, J.K., Chen, K.X., Zhao, W.M., Shen, X., Jiang, H.L., 2003. Virtual screening on natural products for discovering active compounds and target information. Current Medicinal Chemistry 10, 1241–1253.
Shinde, K., Phatak, M., Johannes, F.M., Chen, J., Li, Q., Vineet, J.K., Hu, Z., Ghosh, K., Meller, J., Medvedovic, M., 2010. Genomics portals: integrative web-platform for mining genomics data. BMC Genomics 11, 27.
Shoichet, B.K., Kuntz, I.D., 1993. Matching chemistry and shape in molecular docking. Protein Engineering 6, 723–732.
Smith, J.G., Newton-Cheh, C., 2009. Genome-wide association study in humans. Methods in Molecular Biology 573, 231–258.
Thomsen, R., Christensen, M.H., 2006. MolDock: a new technique for high-accuracy molecular docking. Journal of Medicinal Chemistry 49, 3315–3321.
Venkatachalam, C.M., Jiang, X., Oldfield, T., Waldman, M., 2003. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. Journal of Molecular Graphics & Modeling 21, 289–307.
Uzuner, H., Fan, T.P., Dias, A., Guo, D.A., El-Nezami, H.S., Xu, Q., 2010. Establishing an EU-China consortium on traditional Chinese medicine research. Chinese Medicine 5, 42.
Wang, S.Q., Du, Q.S., Zhao, K., Li, A.X., Wei, D.Q., Chou, K.C., 2007. Virtual screening for finding natural inhibitor against cathepsin-L for SARS therapy. Amino Acids 33, 129–137.
Wang, E., Bussom, S., Chen, J., Quinn, C., Bedognetti, D., Lam, W., Guan, F., Jiang, Z., Mark, Y., Zhao, Y., Stroncek, D.F., White, J., Marincola, F.M., Cheng, Y.-C., 2011. Interaction of a traditional Chinese medicine (PHY906) and CPT-11 on the inflammatory process in the tumor microenvironment. BMC Medical Genomics 4, 38.
Wang, H.Y., Cao, Z.X., Li, L.L., Jiang, P.D., Zhao, Y.L., Luo, S.D., Yang, L., Wei, Y.Q., Yang, S.Y., 2008. Pharmacophore modeling and virtual screening for designing potential PLK1 inhibitors. Bioorganic & Medicinal Chemistry Letters 18, 4972–4977.
Wang, R., Fabregat, A., Ríos, D., Ovelleiro, D., Foster, J.M., Côté, R.G., Griss, J., Csordas, A., Perez-Riverol, Y., Reisinger, F., Hermjakob, H., Martens, L., Vizcaíno, J.A., 2012. PRIDE Inspector: a tool to visualize and validate MS proteomics data. Nature Biotechnology 30 (2), 135–137.
Waters, M., Stasiewicz, S., Merrick, B.A., Tomer, K., Bushel, P., Paules, R., Stegman, N., Nehls, G., Yost, K.J., Johnson, C.H., Gustafson, S.F., Xirasagar, S., Xiao, N., Huang, C.-C., Boyer, P., Chan, D.D., Pan, Q., Gong, H., Taylor, J., Choi, D., Rashid, A., Ahmed, A., Howle, R., Selkirk, J., Tennant, R., Fostel, J., 2008. CEBS—Chemical effects in biological systems: a public data repository integrating study design and toxicity data with microarray and proteomics data. Nucleic Acids Research 36, D892–D900.
Wen, Z., Wang, Z., Wang, S., Ravula, R., Yang, L., Xu, J., Wang, C., Zuo, Z., Chow, M.S., Shi, L., Huang, Y., 2011. Discovery of molecular mechanisms of traditional Chinese medicinal formula Si-Wu-Tang using gene expression microarray and connectivity map. PLoS One 6, e18278.
Wishart, D.S., Knox, C., Guo, A.C., Eisner, R., Young, N., Gautam, B., Hau, D.D., Psychogios, N., Dong, E., Bouatra, S., Mandal, R., Sinelnikov, I., Xia, J., Jia, L., Cruz, J.A., Lim, E., Sobsey, C.A., Shrivastava, S., Huang, P., Liu, P., Fang, L., Peng, J., Fradette, R., Cheng, D., Tzur, D., Clements, M., Lewis, A., De Souza, A., Zuniga, A., Dawe, M., Xiong, Y., Clive, D., Greiner, R., Nazyrova, A., Shaykhutdinov, R., Li, L., Vogel, H.J., Forsythe, I., 2009. HMDB: a knowledgebase for the human metabolome. Nucleic Acids Research 37, D603–D610.
Wolber, G., Langer, T., 2005. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. Journal of Chemical Information & Modeling 45, 160–169.
Wolf, S., Schmidt, S., Müller-Hannemann, M., Neumann, S., 2010. In silico fragmentation for computer assisted identification of metabolite mass spectra. BMC Bioinformatics 11, 148.
Yu, H., Wang, Z.L., Zhang, L.R., Zhang, J.F., Huang, Q., 2007. Pharmacophore modeling and in silico screening for new KDR kinase inhibitors. Bioorganic & Medicinal Chemistry Letters 17, 2126–2133.
Yue, Q.X., Cao, Z.W., Guan, S.H., Liu, X.H., Tao, L., Wu, W.Y., Li, Y.X., Yang, P.Y., Liu, X., Guo, D.A., 2008. Proteomics characterization of the cytotoxicity of mechanism of ganoderic acid D and computer-automated estimation of the possible drug target network. Molecular & Cellular Proteomics 7, 949–961.
Zeng, Z.R., Qian, L., Cao, L.X., Tan, H.M., Huang, Y.L., Xue, X.L., Shen, Y., Zhou, S.N., 2008. Virtual screening for novel quorum sensing inhibitors to eradicate biofilm formation of *Pseudomonas aeruginosa*. Applied Microbiology & Biotechnology 79, 119–126.
Zhu, F., Han, B., Kumar, P., Liu, X., Ma, X., Wei, X., Huang, L., Guo, Y., Han, L., Zheng, C., Chen, Y., 2010. Update of TTD: therapeutic target database. Nucleic Acids Research 38, D787–D791.