HIM-herbal ingredients in-vivo metabolism database

Hong Kang[1], Kailin Tang[1,2], Qi Liu[1], Yi Sun[1], Qi Huang[1], Ruixin Zhu[1], Jun Gao[1,3], Duanfeng Zhang[1], Chenggang Huang[4*] and Zhiwei Cao[1,2*]

Abstract

Background: Herbal medicine has long been viewed as a valuable asset for potential new drug discovery and herbal ingredients' metabolites, especially the *in vivo* metabolites were often found to gain better pharmacological, pharmacokinetic and even better safety profiles compared to their parent compounds. However, these herbal metabolite information is still scattered and waiting to be collected.

Description: *HIM* database manually collected so far the most comprehensive available *in-vivo* metabolism information for herbal active ingredients, as well as their corresponding bioactivity, organs and/or tissues distribution, toxicity, ADME and the clinical research profile. Currently *HIM* contains 361 ingredients and 1104 corresponding *in-vivo* metabolites from 673 reputable herbs. Tools of structural similarity, substructure search and Lipinski's Rule of Five are also provided. Various links were made to PubChem, PubMed, TCM-ID (Traditional Chinese Medicine Information database) and HIT (Herbal ingredients' targets databases).

Conclusions: A curated database *HIM* is set up for the *in vivo* metabolites information of the active ingredients for Chinese herbs, together with their corresponding bioactivity, toxicity and ADME profile. *HIM* is freely accessible to academic researchers at http://www.bioinformatics.org.cn/.

Keywords: TCM, *In vivo*, Metabolism, Metabolite, Biotransformation, Structure search

Background

As one of the naturally originated medical systems, Chinese herbal medicine (CHM) has developed for several thousand years and accumulated plenty of clinical experiences and pharmacological information to form its own integrated theory system [1]. Being a multi-component and multi-target therapy methodology, the studies on its molecular mechanism have made a great progress in recent years although much more still remains unclear [2-4]. In order to get a deeper insight into the mechanism of CHM, various modern scientific technologies have been applied to separate and purify the active ingredients from herbs and elucidate their pharmacodynamic characteristics [5]. Over the past few years, many active compounds have been separated and their pharmacological effects were tested [6-8]. More interestingly, during many researches [9-11], active metabolites were sometimes found to gain better pharmacological, pharmacokinetic and safety profiles compared to their respective parent compounds. For example, morphine is a widely used analgesic which was extracted from Papaver somniferum L and its major therapeutic benefit is mediated by morphine-6-glucuronide, an active metabolite of morphine [12,13]. Another example, Ginsenoside-Rb1, a major active ingredient of Panax ginseng, is found to have the antiallergic activity through its main metabolite named compound K instead of itself [14]. Similarly, glycyrrhizic acid, an alicyclic compound which was extracted from Glycyrrhiz glabra L, has no effect of anti-lipid peroxidation in rat hepatocyte, while its metabolite, glycyrrhetinic acid, has the inhibitory effect on lipid peroxidatioin in dose-dependent manner [15]. Many other similar instances can be found which implies an important message that metabolites of herbal ingredients could be highly valuable for new drug discovery.

Currently, abundant metabolism information of herbal active ingredients has been produced with the progress

* Correspondence: cghuang@mail.shcnc.ac.cn; zwcao@tongji.edu.cn
[4]Shanghai Institute of Materia Medica, Shanghai Institute for Biological Science, Chinese Academy of Sciences, Shanghai 201203, China
[1]School of Life Sciences and Technology, Tongji University, Shanghai 200092, China
Full list of author information is available at the end of the article

of TCM modernization. However, although there exist several databases such as MDL Metabolite Database [16] and Accelrys Metabolism Database [17] for synthetic compounds whose pharmacokinetic and metabolism data have been carefully stored, there is still lack of specific database to collect and store the corresponding information for herbal active ingredients. It is noted that synthetic compounds metabolic databases have made great contributions to new drug discovery [18]. Constructing a database collected the CHM ingredients metabolism information could also have substantial positive impact on TCM development.

In our previous work, in order to collect available resources of protein targets for FDA-approved drugs and the promising precursors, we developed HIT [19] (http://lifecenter.sgst.cn/hit/). Served as a serial work after HIT, HIM is a database which aims to provide the systematical and accurate data storage, data access as well as data analysis (i.e., structural similarity search and substructure analysis) for the herbal active ingredients *in vivo* metabolism information. In this work, the *in vivo* metabolism data of those active ingredients extracted from herbs were collected from literature, unpublished in-house experimental data and the Chinese herbal medical monographs. The information from all these heterogeneous data sources was further processed and integrated into a well-designed database. The information of each ingredient was divided into three categories: identification label, metabolic scheme and bioactivity information. Additionally, properties like the number of hydrogen bond (H-bond) donors and acceptors, molecular weight or the octanol-water partition coefficient logP, which allow the evaluation of the Lipinski's Rule of Five, can be found within the database. The 2D structures of all the compounds are available and the structure similarity search function and substructure search function are also provided. In summary, up to now, there are 361 active ingredients from 673 Chinese herbs and 1104 corresponding metabolites stored in HIM. All the data were freely accessible at http://www.bioinformatics.org.cn/ for academic researches.

Construction and content
Data source
The data in HIM were compiled from both primary and secondary sources.

First, Chinese herbal ingredients *in vivo* metabolism data were extracted from PubMed [20] literature by searching with key words: metabolism, metabolite, biotransformation, metabolic, CHM, Chinese herbal medicine, *in vivo*. Then a preliminary screening was carried out by browsing all the abstracts manually. After that we checked the full text for all the qualified articles and extracted the information according the database

criterion. At last, the data is confirmed when they passed the quality control process which consists of rechecking and revising. In summary, about one-third of all entries come from literatures.

Second, metabolism data were extracted from the book entitled "Absorption, Distribution, Metabolism, Excretion, Toxicity and Activity of The Chemical Constituents in Traditional Chinese Medicines" [21]. This book is a well-known TCM monograph which is concerning ADME/T (Absorption, Distribution, Metabolism, Excretion and Toxicity) of CHM active ingredients in China. Approximately half of the entries in the database are derived from this book, which made such valuable information available online for the first time.

Third, some unpublished *in vivo* experimental data about CHM ingredients metabolism are also gathered in HIM, which accounts for the remaining minority of the entries.

Content and details
The database HIM comprises three data fields for each active ingredient: identification label, metabolic scheme and bioactivity information.

Identification label
In this field, following information is provided for each record:

Common, Alias and Systematic Names. Both the Chinese pinyin and the common English name are provided. The aliases of each active ingredient that are obtained from the database SciFinder [22] are also listed. The systematic name presents precise details of a chemical structure which is generated based on the IUPAC names of natural product skeletal types.
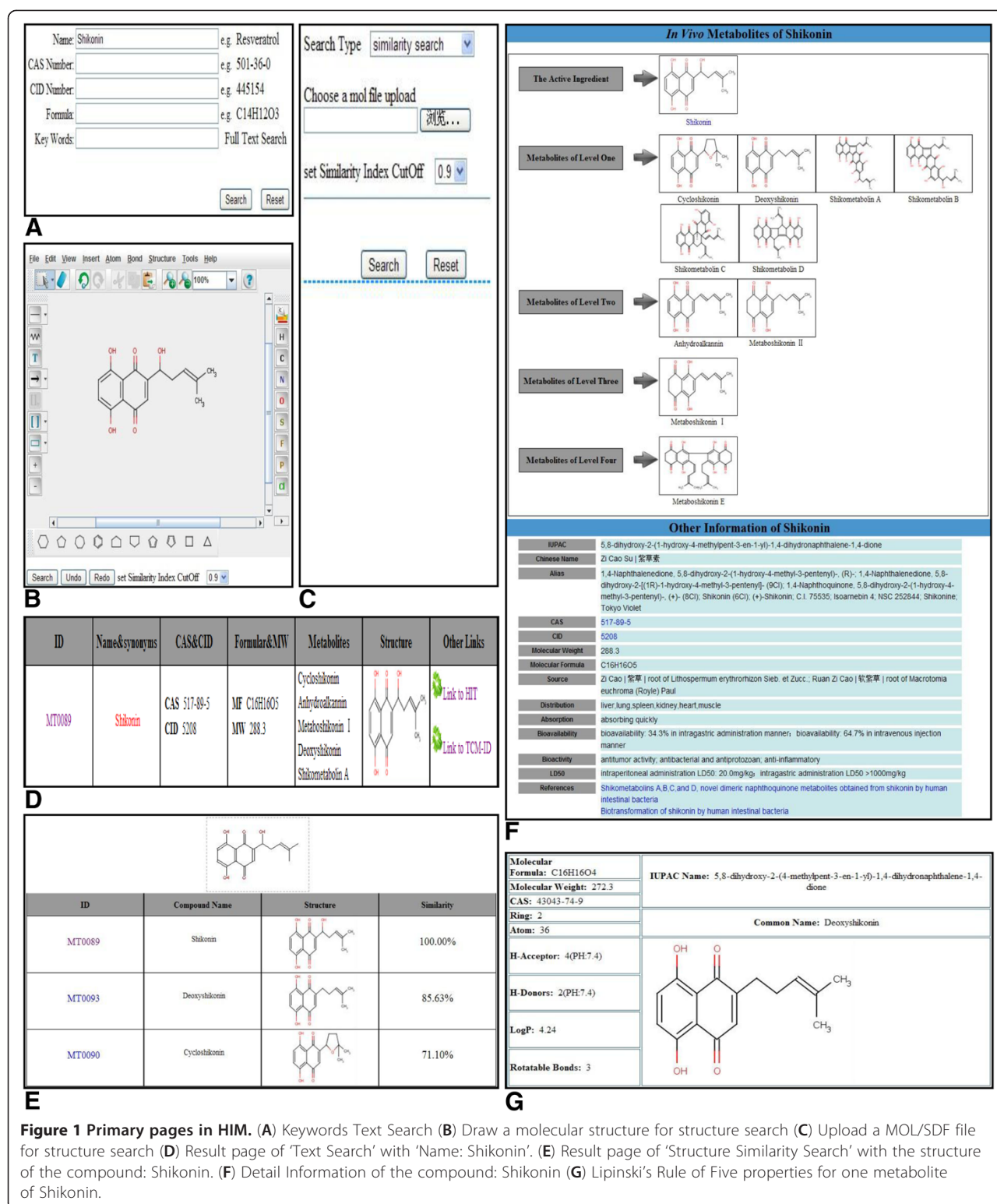
CAS Registry Number. CAS number provides a reliable link between different systems of nomenclature as well as an access to future information on every ingredient. In addition, the compounds are also annotated with the CID numbers provided by PubChem [23] with a hyperlink to it.

Botanical Species. Latin binomials of the herbs and the corresponding region of the plant in which the ingredient located are listed.

Metabolic scheme
In this field, detailed information about *in vivo* metabolism data of each CHM active ingredient is available.

Molecular Weight and Molecular Formula. The MW and MF of all compounds are given by using the functions provided by Marvin Bean [24], a JAVA package supplied by ChemAxon [25].

Figure 1 Primary pages in HIM. (**A**) Keywords Text Search (**B**) Draw a molecular structure for structure search (**C**) Upload a MOL/SDF file for structure search (**D**) Result page of 'Text Search' with 'Name: Shikonin'. (**E**) Result page of 'Structure Similarity Search' with the structure of the compound: Shikonin. (**F**) Detail Information of the compound: Shikonin (**G**) Lipinski's Rule of Five properties for one metabolite of Shikonin.

Structure. For each CHM active ingredient and its metabolites, a 2D structure which is stored in MDL mol format in HIM is shown in JPEG format on the web page.

Metabolite and Metabolic Scheme. All the metabolites of each active ingredient and the full view of *in vivo* metabolic process are provided in HIM.

Bioactivity information

In this field, extensive information about the pharmacokinetic (ADME) properties, bioactivity and toxicity of each active ingredient are listed.

Bioactivity

Some general concepts like anti-cancer, anti-inflammation, anti-bacteria, etc. rather than the diseases-related protein targets are used to describe the bioactivity of each active ingredient.

Toxicity

As being generally recognized, LD50 (median lethal dose) value is used to represent the toxicity with some concrete descriptions for each ingredient.

Other Information. Besides the bioactivity and toxicity, some other information such as absorption, distribution, clinical research and the main references are also available (see Figure 1).

Additional functions

Text search

The "Text Search" function in the homepage of the website provides five distinctive items to search the whole database: Compound name, CAS Number, CID Number, Molecular Formula and Keywords.

Structure similarity search

The "Structure Similarity Search" can be done by uploading a compound structure in MOL/SDF format or via drawing the structure as you want with an embedded molecule editor applet, Marvin Sketch [24]. The structure similarity search is performed by using the so-called structural fingerprint, a binary string with a length of 1024 bits which has encoded the structure characteristics of a given compound. Note that the fingerprint is generated by the Chemistry Development Kit (CDK) [26]. Then the Tanimoto coefficient is calculated by the background program. A molecule with a Tanimoto coefficient ≥ 0.85 to an active compound is often assumed to own similar biological activity [27].
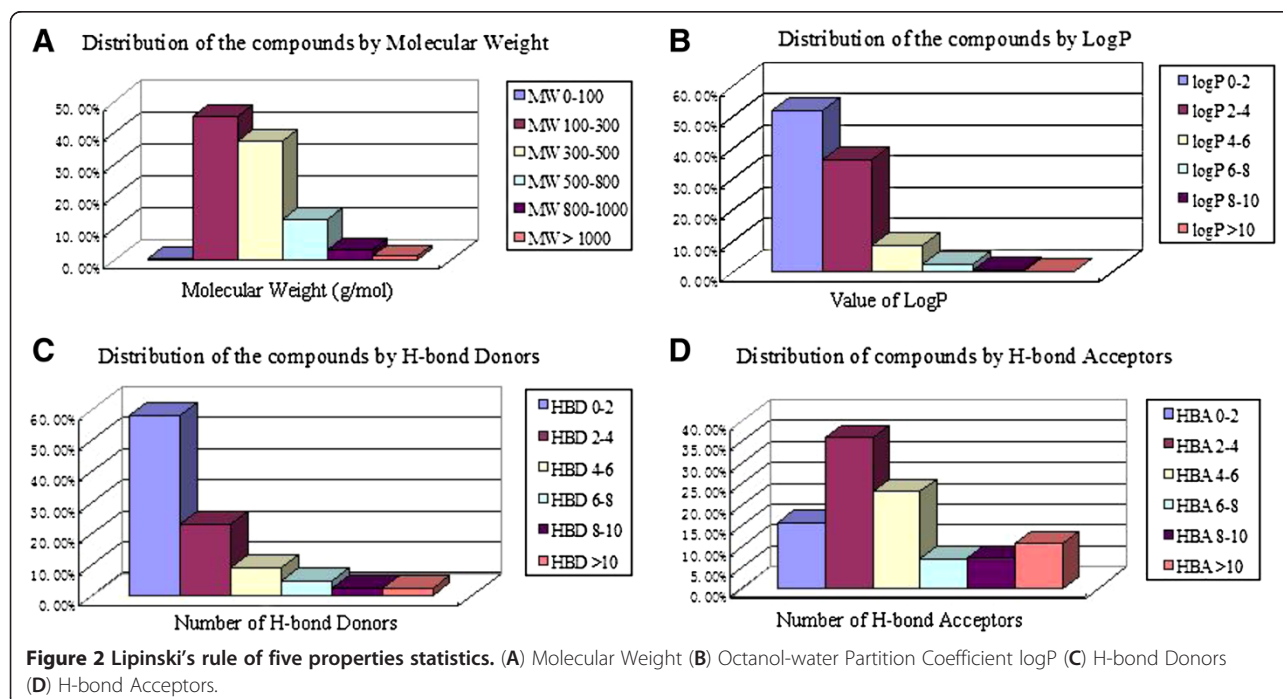
Substructure search

Chemical substructure-based *in silico* techniques have been wildly used as an effective and popular approach to reduce the cost in identifying molecules suitable for pharmaceutical development in early stage of drug discovery [28,29]. In our database *HIM*, substructure search is also available by *JChem* [30].

Website and server

HIM is available online at: http://www.bioinformatics.org.cn/. It is designed as a relational database and implemented in MySQL Server 5.0 with the Apache Tomcat 6.0 as the web server. For chemical calculation and structure drawing, *CDK package* and *Marvin Sketch applet* are embedded. The website is built in JSP, HTML and CSS.

Utility

HIM (http://www.bioinformatics.org.cn/), which is proposed in this work, is served as a serial work after HIT

Figure 2 Lipinski's rule of five properties statistics. (**A**) Molecular Weight (**B**) Octanol-water Partition Coefficient logP (**C**) H-bond Donors (**D**) H-bond Acceptors.

and concerns about the herbal active ingredients with explicit *in vivo* metabolism data, since the active metabolites of CHM were sometimes found to gain better pharmacological, pharmacokinetic and safety profiles compared to their respective parent compounds. With the help of HIM, researches could find out the mechanism of pharmacological action of CHM more comprehensively, which is expected to have a substantial positive impact on CHM development.

Discussion

Although having been used for thousands of years and own outstanding reputation, the mechanism of CHM is still largely unknown. One reason contributing to this is that it is unclear for the process of ADME/T *in vivo* such as *in vivo* metabolism of CHM. *HIM* is constructed as the first database to store almost all the CHM active ingredients *in vivo* metabolism data dated to January of 2012, as well as their corresponding bioactivity, toxicity, and ADME profile. The properties of Lipinski's rule of five for each compound are also given to the whole database. As one of the common rules, Lipinski's rule of five is widely used in drug screening and design. Blake's [31] study has shown that for the five stages from pre-clinical to approved, less and less compounds break the Lipinski's rules. It is indicated that compounds which are against the Lipinski's rule need to be modified too much and they are little probability to be a drug. In *HIM* about 90% of the compounds (herbal ingredients and their metabolites) meet the Lipinski's rule (Figure 2). We hope that *HIM* can be served as a valuable database to make the progress of CHM modernization and provide great assistance in the new drug discovery and developments.

Conclusion

HIM can be used to get the metabolites of the active ingredient which the researchers are interested in. The structure similarity search and substructure search can be applied to get compounds which potentially similar bioactivity to the query compound and can provide other chemical and biologic information of the query molecule. Moreover, the database is useful for the study of pharmacognosy. Although some active metabolic intermediates are unstable and hard to get, fermentation technology such as microbial transformation could be used to obtain the active compounds which are the *in vivo* metabolites of CHM. Crude herbal medicines could be fermented by certain microbial strains and get certain products [32,33]. *HIM* could provide valuable information for the researchers who are interested in TCM, drug design, pharmacognosy, drug metabolism, etc.

Author details
[1]School of Life Sciences and Technology, Tongji University, Shanghai 200092, China. [2]Shanghai Center for Bioinformation and Technology, 1278 Keyuan Road, Shanghai 201203, China. [3]College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China. [4]Shanghai Institute of Materia Medica, Shanghai Institute for Biological Science, Chinese Academy of Sciences, Shanghai 201203, China.

References
1. Hesketh T, Zhu WX: **Health in China. Traditional Chinese medicine: one country, two systems.** *BMJ* 1997, **315**(7100):115–117.
2. Cheng JT: **Review: drug therapy in Chinese traditional medicine.** *J Clin Pharmacol* 2000, **40**(5):445–450.
3. Yuan R, Lin Y: **Traditional Chinese medicine: an approach to scientific proof and clinical validation.** *Pharmacol Ther* 2000, **86**(2):191–198.
4. Nahin RL, Straus SE: **Research into complementary and alternative medicine: problems and potential.** *BMJ* 2001, **322**(7279):161–164.
5. Liu S, Yi LZ, Liang YZ: **Traditional Chinese medicine and separation science.** *J Sep Sci* 2008, **31**(11):2113–2137.
6. Li N, Lin G, Kwan YW, Min ZD: **Simultaneous quantification of five major biologically active ingredients of saffron by high-performance liquid chromatography.** *J Chromatogr A* 1999, **849**(2):349–355.
7. Chen X, Zhang J, Xue C, Hu Z: **Simultaneous determination of some active ingredients in anti-viral preparations of traditional Chinese medicine by micellar electrokinetic chromatography.** *Biomed Chromatogr* 2004, **18**(9):673–680.
8. Li HL, Zhang WD, Liu RH, Zhang C, Han T, Wang XW, Wang XL, Zhu JB, Chen CL: **Simultaneous determination of four active alkaloids from a traditional Chinese medicine Corydalis saxicola Bunting. (Yanhuanglian) in plasma and urine samples by LC-MS-MS.** *J Chromatogr B Analyt Technol Biomed Life Sci* 2006, **831**(1–2):140–146.
9. Jaiswal PK, Srivastava S, Gupta J, Thakur IS: **Dibenzofuran induces oxidative stress, disruption of trans-mitochondrial membrane potential (DeltaPsim) and G1 arrest in human hepatoma cell line.** *Toxicol Lett* 2012, **214**(2):137–144.
10. Hoffart E, Ghebreghiorghis L, Nussler AK, Thasler WE, Weiss TS, Schwab M, Burk O: **Effects of atorvastatin metabolites on induction of drug-metabolizing enzymes and membrane transporters through human pregnane X receptor.** *Br J Pharmacol* 2012, **165**(5):1595–1608.
11. Bar-Am O, Weinreb O, Amit T, Youdim MB: **The neuroprotective mechanism of 1-(R)-aminoindan, the major metabolite of the anti-parkinsonian drug rasagiline.** *J Neurochem* 2010, **112**(5):1131–1137.
12. Osborne R, Joel S, Trew D, Slevin M: **Morphine and metabolite behavior after different routes of morphine administration: demonstration of the importance of the active metabolite morphine-6-glucuronide.** *Clin Pharmacol Ther* 1990, **47**(1):12–19.

13. Portenoy RK, Thaler HT, Inturrisi CE, Friedlander-Klar H, Foley KM: **The metabolite morphine-6-glucuronide contributes to the analgesia produced by morphine infusion in patients with pain and normal renal function.** *Clin Pharmacol Ther* 1992, **51**(4):422–431.
14. Choo M, Park E, Han M, Kim D: **Antiallergic activity of ginseng and its ginsenosides.** *Planta Med* 2003, **69**(6):518–522.
15. Kiso Y, Tohkin M, Hikino H, Hattori M, Sakamoto T, Namba T: **Mechanism of antihepatotoxic activity of glycyrrhizin. I: Effect on free radical generation and lipid peroxidation.** *Planta Med* 1984, **50**(4):298–302.
16. *MDL Metabolite Database.* http://www.mdl.com.
17. *Accelrys Metabolism Database.* http://www.accelrys.com.
18. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J: **DrugBank: a comprehensive resource for in silico drug discovery and exploration.** *Nucleic Acids Res* 2006, **34**(Database issue): D668–672.
19. Ye H, Ye L, Kang H, Zhang D, Tao L, Tang K, Liu X, Zhu R, Liu Q, Chen YZ, *et al*: **HIT: linking herbal active ingredients to targets.** *Nucleic Acids Res* 2011, **39**(Database issue):D1055–1059.
20. *PubMed.* http://www.ncbi.nlm.nih.gov/pubmed.
21. Yang X: *Absorption, Distribution, Metabolism, Excretion, Toxicity and Activity of The Chemical Constituents in Traditional Chinese Medicines.* Beijing: China Medical Science Press; 2006.
22. *SciFinder.* https://www.cas.org/products/scifinder.
23. *PubChem.* http://pubchem.ncbi.nlm.nih.gov/.
24. *Marvin Bean.* http://www.chemaxon.com/download/marvin/for-end-users/.
25. *ChemAxon.* http://www.chemaxon.com.
26. Steinbeck C, Han Y, Kuhn S, Horlacher O, Luttmann E, Willighagen E: **The Chemistry Development Kit (CDK): an open-source Java library for Chemo- and Bioinformatics.** *J Chem Inf Comput Sci* 2003, **43**(2):493–500.
27. Martin Y, Kofron J, Traphagen L: **Do structurally similar molecules have similar biological activity?** *J Med Chem* 2002, **45**(19):4350–4358.
28. Xu J, Hagler A: **Chemoinformatics and drug discovery.** *Molecules* 2002, **7**(8):566–600.
29. Merlot C, Domine D, Cleva C, Church DJ: **Chemical substructures in drug discovery.** *Drug Discov Today* 2003, **8**(13):594–602.
30. *JChem.* http://www.chemaxon.com.
31. Blake JF: **Examination of the computed molecular properties of compounds selected for clinical development.** *Biotechniques* 2003, **Suppl**:16–20.
32. Rizzello CG, Coda R, Macias DS, Pinto D, Marzani B, Filannino P, Giuliani G, Paradiso VM, Di Cagno R, Gobbetti M: **Lactic acid fermentation as a tool to enhance the functional features of Echinacea spp.** *Microb Cell Fact* 2013, **12**(1):44.
33. Zhao W, Chen X, Li X: **The application of microbial fermentation in the study of Chinese herbal medicine.** *Life Sci Instrum* 2008, **10**:3–5.