ARTICLES

Phytochemical Databases of Chinese Herbal Constituents and Bioactive Plant Compounds with Known Target Specificities

Thomas M. Ehrman,[†] David J. Barlow,*[,†] and Peter J. Hylands[‡]

Pharmaceutical Sciences Research Division and Centre for Natural Medicines Research, King's College London, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NH, United Kingdom

Received July 11, 2006

Two databases have been constructed to facilitate applications of cheminformatics and molecular modeling to medicinal plants. The first contains data on known chemical constituents of 240 commonly used Chinese herbs, the other contains information on target specificities of bioactive plant compounds. Structures are available for all compounds. In the case of the Chinese herbal constituents database, further details include trivial and systematic names, compound class and skeletal type, botanical and Chinese (pinyin) names of associated herb(s), CAS registry number, chirality, pharmacological and toxicological information, and chemical references. For the bioactive plant compounds database, details of molecular target(s), IC_{50} and related measures, and associated botanical species are given. For Chinese herbs, approximately 7000 unique compounds are listed, though some are found in more than one herb, the total number for all herbs being 8264. For bioactive plant compounds, 2597 compounds active against 78 molecular targets are covered. Statistical relationships within and between the two databases are explored.

INTRODUCTION

Natural products have always represented a significant, though often underappreciated, resource for the development of new medicines. Even now, in an era dominated by combinatorial chemistry, drugs of plant or microbial origin count for more than 30% of worldwide sales, and natural products have been notably successful in the past in opening up new avenues of exploration and in producing entirely new therapeutic classes.[1,2] Though plant compounds exhibit enormous structural diversity, only a small proportion of that diversity has been seriously explored for its pharmacological potential so far, and there is therefore little reason to believe that this potential has now run dry. By comparison with other areas of pharmaceutical research, however, the screening of natural products has suffered from a lack of data in an appropriate format. In particular, electronic information on chemical structure, pharmacological activity, specificity against known molecular targets, and traditional uses of the herbs in which such compounds are found has been insufficient, though, in the case of Chinese herbs, an increasing amount of information has become available in recent years.[3,4]

While such information can serve a wide variety of purposes, it is perhaps in the field of virtual screening that it may have its greatest impact. Structures either can be searched for their similarity to known active compounds or can be docked to receptors of particular molecular targets, providing new information on their therapeutic potential. This may serve the needs of drug development and may also provide new insights into Chinese (and other) systems of herbal medicine, an objective which is of much interest in view of the recent growth of traditional Chinese medicine (TCM) beyond its traditional boundaries and its emergence as a significant contributor to healthcare worldwide, a trend reflected in the increasing number of publications on TCM now available in languages other than Chinese.[5]

Given the long history of pharmacological prospecting from nature, Chinese herbs are relative newcomers to the scene but have nevertheless shown considerable promise as a source of new drugs in recent years, including artemisinin (qinghaosu), the antimarial compound derived from *Artemisia annua*, and the acteylcholinesterase inhibitors, huperzines A and B, from *Huperzia serrata* (*Lycopodium serratum*), for the treatment of Alzheimer's disease.[1] Chinese herbs are among the best investigated plants from a chemical perspective, and considerably more information is available on their chemical constitution than for herbs from most other parts of the world. Concurrent with work on the elucidation of chemical structures have been investigations into their pharmacology, some of it now summarized in a number of recent volumes,[6,7] though these contain relatively little data on targets at the molecular level.

Information on molecular targets of plant compounds has however also increased over the past few years, though it is only recently that the first such compilation has been published.[8] Again, this is of particular significance for virtual screening in that the information can be used to identify other phytochemicals which may be expected to show similar behavior and affords the first opportunity to map the ligand−

* Corresponding author fax: 44 (0) 207 848 4800; e-mail: dave.barlow@kcl.ac.uk.
[†] Pharmaceutical Sciences Research Division.
[‡] Centre for Natural Medicines Research.

10.1021/ci600288m CCC: $37.00
Published on Web 01/09/2007

receptor space of plant compounds and their respective molecular targets.

Here, we report on the construction of two new databases of use in research relating to Chinese herbs. The first covers details on chemical constituents of many of the major herbs used in TCM, the other provides information on phytochemical compounds with known activity against a wide variety of targets, among them many of proven or suspected therapeutic significance. In the case of the latter, the information is largely restricted to experimental data on in vitro systems of mammalian origin. For all entries in both databases, structures are provided.

It is only in recent years that a sufficient volume of information has accumulated to make this possible, and the data are therefore limited. In the case of Chinese herbs, our database contains details on 8264 compounds found in 240 of the most commonly used herbs in Chinese medicine. A number of these, however, are found in more than one herb, with the number of unique compounds being closer to 7000. For the other database, the number of compounds with known targets is 2597. A total of 78 targets is presently covered.

CONTENT AND DETAILS

The data were compiled from both primary and secondary sources, though a number of secondary sources in particular played an important role. In the case of Chinese herbs (CHCD) these were Hsu Hong-Yen, *The Chemical Constituents of Oriental Herbs*, volumes I and II;[9] Zhu You-Ping, *Chinese Materia Medica: Chemistry, Pharmacology & Applications*;[6] Yan X et al., *Traditional Chinese Medicines: Molecular Structures, Natural Sources & Applications*;[10] Duke, *Handbook of Phytochemical Constituents of GRAS Herbs and Other Economic Plants*,[11] and the CRC/Chapman & Hall *Dictionary of Natural Products* (DNP).[12]

In the case of bioactive plant compounds (BPCD), the major reference source was Gideon Polya's *Biochemical Targets of Plant Bioactive Compounds*.[8] In addition, another useful source of information was the *Dictionary of Natural Products*.[12]

Both the CHCD and BPCD were constructed using Microsoft Access 2000 (9.0).

In all cases, structures were downloaded (MDL molfile[13] and SMILES[14] formats) from the DNP (CD-ROM version). In cases where structural modifications were required, these were carried out first within the DNP prior to export.

Both data sets contain the following information:

1. Structure. For each entry, 2D structures are stored in MDL mol format. 3D conformations, built in MOE (Chemical Computing Group, Montreal, Quebec) using the Merck MMFF94 force field, are also available, though these only provide a single minimized conformation per entry.

2. SMILES Code. All structures are accompanied by corresponding SMILES codes, thus allowing for greater flexibility of export and subsequent import into a wide variety of chemical software.

3. Trivial and Systematic Names. The trivial name refers to the common name given to a compound. Often, this gives little clue to its structure, though suffixes such as "-oside" or "-olide" indicate that it is a glycoside or lactone, respectively, and "saponin" is frequently found, indicating that the compound is a triterpene glycoside (usually of the pentacyclic type). Examples of trivial names include *ginsenoside Rg1* (a triterpene glycoside from *Panax* spp.), *apigenin* (a trihydroxylated flavone found in many species), and *tanshinone V* (a diterpene from *Salvia miltiorrhiza*, an herb known as Dan Shen in Chinese or Tanjin in Japanese).

The systematic name gives precise details of chemical structure, though *semisystematic* names, based on the names of natural product skeletal types, are usually less cumbersome and also give similar precision. Where possible these have been preferred. Thus, *2-menthanol* or *menthan-2-ol*, for instance, indicates the menthane monoterpene skeleton with a hydroxyl group attached to carbon 2, and *12-oleanen-3-ol* or *olean-12-en-3-ol* (of which the trivial name is β-amyrin) denotes the oleanane triterpene skeleton, with a hydroxyl group attached to carbon 3 and with a double bond between carbons 12 and 13. In all cases, IUPAC recommendations have been followed.[15] For further details on phytochemical nomenclature, see the *Dictionary of Natural Products*.[12]

Where possible, both trivial and semisystematic (or systematic) names are listed in the database on Chinese herbs. For the bioactive plant compounds database, only a single name (generally the trivial name) is currently given. In the case of Chinese herbs, common synonyms are also listed.

For glycosides, the following convention has been adopted: the parent aglycone is given first, followed by details of substituent sugars. Thus, in the case of ginsenoside Rg_1 (Figure 1), for example, the following is given under systematic name: *dammar-24-ene-3,6,12,20-tetrol*; *6,20-di-O-β-D-glucopyranoside*. This indicates that the hydroxyl groups attached to carbons 6 and 20 of the dammarane skeleton in the aglycone are substituted by two glucopyranoside moieties in the glycoside.

4. Compound Class and Skeletal Type. The classification of compounds follows that in the *Dictionary of Natural Products*.[12] Major classes comprise the following: aliphatics, alkaloids, phenolics, and terpenoids. Phenolics are further subdivided into simple phenolics (phenols, phenylpropanoids, etc.), coumarins and benzofurans, lignans, flavonoids, polycyclic aromatics (largely anthraquinones and related classes), and tannins. Terpenoids comprise monoterpenes, sesquiterpenes, diterpenes, triterpenes, and steroids.

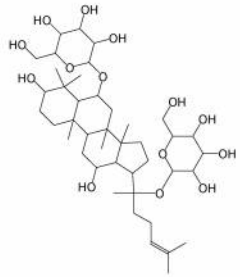
Information on the skeletal type within each class is also included. As indicated above, the name of a skeletal type indicates the scaffold upon which individual compounds are built.

5. Chirality. Information on the chirality of each compound containing one or more chiral centers is given (where known). Where possible, the Cahn-Ingold-Prelog priority rules by which chiral centers are labeled *R* or *S* are followed, though in some cases, older or alternative nomenclatures (such as $\pm$, D/L, or α/β) are given as well.[16]

Default structures do not show chiral centers, but these may be added where such details are available.

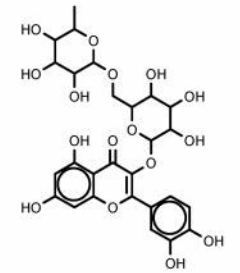
6. CAS Registry Number(s). For each compound, the CAS registry number is given. These provide a reliable common link between different systems of nomenclature and provide access to further information on individual compounds.

7. Pharmacology. Where pharmacological information is available for individual compounds, it has been included,



Trivial name	*Synonyms*	*Systematic name*
Ginsenoside Rg_1	Ginsenoside A_2, Panaxoside A, Sanchinoside C_1	Dammar-24-ene-3,6,12,20-tetrol; 6,20-Di-*O*-β-D-glucopyranoside
Herb (pinyin)	***Species***	***Part of plant***
(1) Ren Shen, (2) San Qi	(1) *Panax ginseng*, (2) *Panax notoginseng*	Root
CAS number	***Class***	***Skeletal type***	***Chirality***
22427-39-0	Triterpenoid	Dammarane	(3*R*,6*S*,12*R*,20*S*) or (3β,6α,12β,20*S*)
Pharmacology	***Toxicology***	***References***
Adaptogenic*, CNS stimulant*, Immunomodulator*, Tumour-inhibitory activity	LD_{50} (mus, ipr) 405 mg/kg, LD_{50} (mus, ipr) 1,600 mg/kg*, RTECS No. LY9537200	Sanada S et al. *Shoyakugaku Zasshi*, **1978**, *32*: 96 [*P. notoginseng*]; Yahara S et al. *Chem. Pharm. Bull.*, **1976**, *24*: 2204; Yahara S. et al. *Chem. Pharm. Bull.*, **1979**, *27*: 88 [*P. ginseng*]; *Duke, J.A., *Biologically Active Phytochemicals and Their Activities, CRC Press*, 1992, p *70*.

Figure 1. Entry for ginsenoside Rg_1 from the CHCD.



Trivial name	*Species*	*CAS number*
Rutin	Widespread	153-18-4
Targets	***IC_{50}***	***Further details***
(1) Ca^{2+}-calmodulin myosin light chain kinase, (2) protein kinase A, (3) HIV-1 protease, (4) 5-LOX, (5) Aldose reductase	(1) 320 μM, (2) 32 μM, (3) <82 μM, (4) 2.5 mM (IC_{75}), (5) no details	None

Class	*Skeletal type*	*Pharmacology*	*Toxicology*	*Chirality*
Flavonoid	Flavonol (*O*-glycoside)	Anti-HIV agent, Antihypotensive, Anti-inflammatory, Antispasmodic, Antithrombotic, Antiviral, Haemostatic.	LD_{50} (rat, ipr) 2000 mg/kg, RTECS no. VM2975000.	-

Figure 2. Entry for rutin from the BPCD.

though there is no information at present for herbal extracts containing more than one compound.

8. Toxicology. Where known, toxicological information (mainly LD_{50} values) is included for individual compounds. RTECS (Registry of Toxic Effects of Chemical Substances)[17] numbers are also included to facilitate linking to further sources of information. Again, there are no data currently available for plant extracts as opposed to individual compounds.

In the case of the CHCD, the following information is also available:

9. Botanical Species and Common Herb Name. Latin binomials of the herb(s) in which the compound is found are listed as well as the Chinese (pinyin) name. Nomenclature follows that of the references listed above. Where two or more Chinese names are given for a herb, priority was given to the most common name found in Bensky et al.,[18] the major reference source in English on the Chinese materia medica.

10. Part of the Plant in which the Compound is Found. This is among the most difficult information to establish with any certainty as the part of the plant in which a compound is preferentially found can vary considerably dependent on genotype, season, and growing conditions. For many compounds, broad guidelines can nevertheless be established in this respect, though the information given should always be treated cautiously. Fortunately, much of the work carried out in China and Japan on chemical characterization has been undertaken from a primarily medical perspective, so the part of the plant chosen for investigation often matches that found in traditional materia medica.

11. Chemical Reference(s). In the case of Chinese herbs, references to the primary literature, concerning the chemical characterization of compound structures, are included.

12. TCM Categories. For each herb, the major category of TCM into which it falls is given in addition to other traditional uses. Where possible, TCM categorization follows Bensky et al.[18] In other cases, the references listed above have been used.

For the BPCD, the following additional information is included:

13. Target(s). The molecular targets for each compound are listed. These are shown in Table 1. They fall within a number of therapeutic categories such as ion channels,

Table 1. Details of Molecular Targets and the Number of Ligands for Each in the BPCD (NC = Number of Compounds)

category	target	NC	category	target	NC
ion channels	nicotinic acetylcholine receptor agonists	19	gene expression	protein synthesis	54
	nicotinic acetylcholine receptor antagonists	49		DNA ligands	38
	GABA(A) agonists/ligands	21		DNA helicase	6
	GABA(A) antagonists	66		DNA ligase	12
	Ca^{2+}-ATPase	21		DNA polymerase	43
	H^+,K^+-ATPase	4		topoisomerase I (TOPI)	31
	Na^+,K^+-ATPase	64		topoisomerase II (TOPII)	59
	voltage-gated Na^+ channel	62		apoptotic	95
	voltage-gated Ca^{2+} channel	2	HIV enzymes	HIV-1 integrase	41
G protein-coupled receptors	adenosine receptor	34		HIV-1 reverse transcriptase	91
	muscarinic acetylcholine receptor agonists	14		HIV-1 protease	58
	muscarinic acetylcholine receptor antagonists	28	cytosolic hormone receptors & enzymes	androgen receptor	12
	α1-adrenergic receptor	43		testosterone 5α-reductase	25
	α2-adrenergic receptor	36		oestrogen receptor	49
	β-adrenergic receptor	28		oestrogen aromatase	44
	dopamine receptor	47		17β-hydroxysteroid oxidoreductase	18
	acetylcholinesterase	44		ACE	28
	butyrylcholinesterase	8		chymotrypsin	23
	monoamine oxidase	58		trypsin	18
	prolyl endopeptidase	40		prolyl endopeptidase	40
neurotransmitter converters	DOPA decarboxylase	2		F_1-ATPase	27
	choline acetyltransferase	1		electron transport chain	54
	dopamine-β-hydroxylase	1−2		oxidative phosphorylation uncouplers and inhibitors	17
	succinic semialdehyde dehydrogenase and reductase	4	metabolism	glucose transporter	12
	tyrosinase	12		multidrug resistance transporter	61
	tyrosine hydroxylase	4		cytochrome P450 oxygenase	23
	adenylyl cyclase activators	5		glutathione-S-transferase	21
cyclic nucleotides	adenylyl cyclase inhibitors	6		**nucleotidase/CAB nucleotidase	15
	cAMP phosphodiesterases	222		phospholipase C	3
	cGMP phosphodiesterases	8		squalene epoxidase	11
	iNOS expression	81		xanthine oxidase	27
nitric oxide	NO production in vivo	97		cyclooxygenase/COX (general)	169
	NOS	4		COX-1	24
	CDPK (Ca^{2+}-dependent PK)	56		COX-2	25
	MLCK (myosin light chain kinase)	65		lipoxygenase/LOX (general)	199
	PKA (cAMP-dependent PK)	101		5-LOX	153
	PKC (Ca^{2+} and phospholipid activated protein kinase)	155	inflammation	12-LOX	18
protein kinases	PKC activators	>25		15-LOX	3
	EGF-RTK (epidermal growth factor receptor tyrosine kinase)	25		sLOX (soybean)	25
	PDGF-RTK (platelet-derived growth factor receptor tyrosine kinase)	2		phospholipase A2	31
	RTK (receptor tyrosine kinase)	25	antioxidants	antioxidants/free radical scavengers	109
diabetes	aldose reductase	187	taste	taste receptors	121

G-protein-coupled receptors (GPCRs), nitric oxide, HIV, inflammation, and so forth.

14. Ligand Type. Information on whether the compound is an agonist, antagonist, inhibitor, or ligand is given.

15. Inhibition. For target inhibitors, which make up the great majority of compounds, quantitative data are included where known. Three commonly used measures are found: (a) IC_{50}, which measures the concentration (μM or nM units) for 50% inhibition of an enzyme, 50% displacement of a known ligand from the target molecule, or 50% inhibition of an in vivo process; (b) K_d, the compound-target dissociation constant (μM or nM units); and (c) K_i, inhibitor-target dissociation constant (μM or nM units), another measure of the tightness of association. Of these, the most common measure is IC_{50}.

16. Further Details. In some cases, other details may be found concerning the target in question, such as information on the protein subunit to which the compound binds and additional features of interest.

17. Botanical Species. The name(s) of some of the species in which the compound has so far been found are given.

STATISTICS

1. Sample Entries from Both Databases. Figures 1 and 2 show examples taken from the CHCD and BPCD, respectively. In the case of Chinese herbs, Figure 1 shows the entry for *ginsenoside Rg1*, a well-studied dammarane triterpene from ginseng species.

In Figure 2, details from the BPCD are given for *rutin*, a widely distributed flavonol glycoside with a number of documented target affinities.

2. Differences in Phytochemical Distribution between Databases. If information from one database, such as patterns of bioactivity, is to be extrapolated to the other, then it is important that there should be a sufficient degree of structural overlap between the two. Figure 3 gives details of compound distribution in the two data sets in terms of the

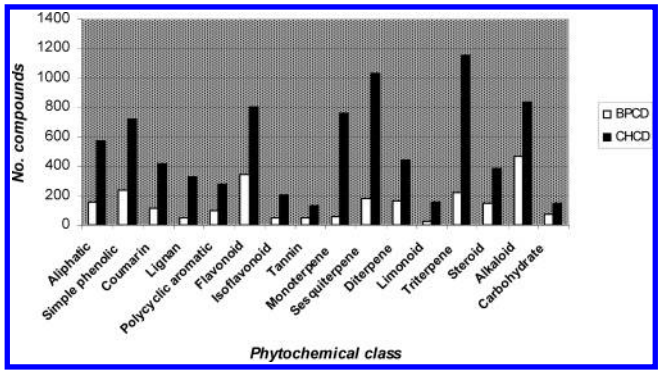


Figure 3. Number of compounds in both databases by phytochemical class.

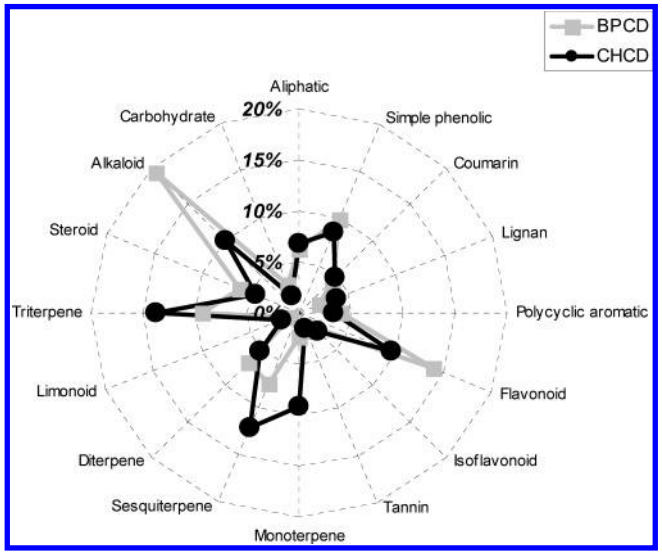


Figure 4. Comparison of both databases in terms of percentage of compounds contributing to each phytochemical class.

major phytochemical classes involved, where the total number of compounds belonging to each class is shown.

In the case of the BPCD, the two largest classes are alkaloids and flavonoids, though simple phenolics (e.g., phenols, phenylpropanoids, etc.) and to a lesser extent triterpenes are also reasonably well-represented. For the CHCD, triterpenes are the largest class, though alkaloids, flavonoids, sesquiterpenes, monoterpenes, alkaloids, and simple phenolics are also present with significant frequency.

Figure 4 shows the relative contribution, in terms of the percentage of compounds, which each class makes to each database. In this way, the degree of structural overlap between the two is rendered more apparent.

We see that aliphatics, coumarins, polycyclic aromatics, and tannins make almost equal contributions to both databases. Flavonoids and, particularly, alkaloids make a larger contribution to the BPCD, whereas most terpenoids, particularly mono-, sesqui-, and triterpenes, contribute a higher proportion of compounds to the CHCD. Simple phenolics, carbohydrates, steroids, and diterpenes are marginally higher, proportionally, among plant bioactive compounds, whereas lignans are marginally higher in Chinese herbs.

3. Compound Distribution in Chinese Herbs. Analysis of the publication dates for each decade from 1911 to 2000 of compounds found in the CHCD reveals that up until 1990 there was an exponential rise in the number of new structures found, as shown in Figure 5. In the ensuing decade, the number did not continue to rise exponentially but was only marginally greater than that found over the period 1980−90. Numbers for the present decade are incomplete, and a trend is therefore difficult to establish. However, it is clear that much information has become available since the first compilation of data on Chinese herbal constituents was published at the start of the 1980s.[9] The apparent decline in publication since 1990 may be partly due to the fact that fewer scientists in China and Japan, where the great majority of studies have been conducted, are now interested in the structural determination of natural products,[19] though it may also reflect the fact that many of the major structural classes have now been elucidated for a large number of herbs.

The numbers of compounds found in 240 Chinese herbs are listed in Table 2. It is often the case that different, though usually closely related, species refer to the same herb, and where this is the case, details for the major species involved are given. In cases where more than three species from the same genus are found, the term "spp." is used. This does not imply that *any* species from that genus can be used but is simply intended to save space. The number of compounds

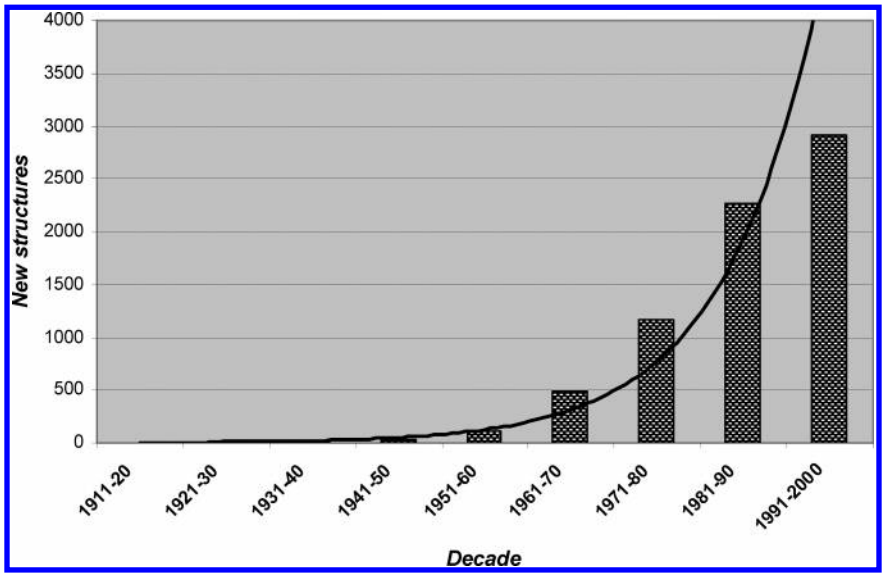


Figure 5. Publication dates for structures in the CHCD up to 2000 (exponential relationship as indicated by curve: $y = 1.325\ e^{0.91x}$).

Table 2. Details of Herbs Currently Covered in the CHCD (NC = Number of Compounds)

species	herb name	NC	species	herb name	NC
Abrus cantoniensis, *A. precatorius*	Ji Gu Cao	71	*Impatiens balsamina*	Tou Gu Cao	33
Achyranthes aspera	Tu Niu Xi	14	*Imperata cylindrica*	Bai Mao Gen	14
Aconitum carmichaeli	Fu Zi	23	*Inula britannica*, *I. japonica*	Xuan Fu Hua	86
Aconitum koreanum	Bai Fu Zi	12	*Inula helenium*	Mu Xiang	21
Aconitum kusnezoffii	Cao Wu	7	*Isatis indigotica*, *I. tinctoria* (leaf)	Ban Lan Gen	12
Acontium japonicum	Cao Wu Tou	20	*Isatis indigotica*, *I. tinctoria* (root)	Da Qing Ye	9
Acoris calamus, *A. gramineus*, *A. tatarinowii*	Shi Chang Pu	92	*Isatis indigotica*, *I. tinctoria* (processed)	Qing Dai	7
Agastache rugosa	Tu Huo Xiang	29	*Juncus effusus*	Deng Xin Cao	44
Agrimonia pilosa	Xian He Cao	26	*Justicia procumbens*	Jue Chuang Cao	19
Ailanthus altissima	Chun Pi	37	*Kochia scoparia*	Di Fu Zi	12
Akebia quinata, *A. trifoliata*	Mu Tong	22	*Ledebouriella divaricata*, *L. seseloides*	Fang Feng	23
Albizzia julibrissin, *A. lebbek*	He Huan Pi	41	*Leonurus cardiaca*, *L. heterophyllus*	Yi Mu Cao	22
Alisma orientale	Ze Xie	45	*Ligusticum chuanxiong*, *L. wallichii*	Chuan Xiong	42
Allium fistulosum	Cong Bai	9	*Ligusticum jeholense*, *L. sinense*	Gao Ben	13
Aloe ferox, *A. vera*	Lu Hui	42	*Lilium brownii*, *L. lancifolium*, *L. longiflorum*	Bai He	31
Alpinia galanga	Hong Dou Kou	4	*Lindera strychnifolia*	Wu Yao	29
Alpinia katsumadai	Cao Dou Kou	25	*Lippia nodiflora*	Peng Lai Cao	18
Alpinia officinarum	Gao Liang Jiang	27	*Liquidambar formosana*, *L. orientalis*	Lu Lu Tong, Su He Xiang	48
Andrographis paniculata	Chuan Xin Lian	37			
Anemarrhena asphodeloides	Zhi Mu	20	*Lithospermum erythrorhizon*	Zi Cao	25
Angelica acutiloba, *A. gigas*, *A. sinensis*	Dang Gui	37	*Litsea cubeba*	Dou Chi Jiang	8
Angelica dahurica	Bai Zhi	30	*Lobelia inflata*	Ban Bian Lian	20
Angelica pubescens	Du Huo	42	*Lonicera japonica*	Jin Yin Hua	24
Aquilaria agallocha, *A. sinensis*	Chen Xiang	86	*Luffa cylindrica*	Si Gua Lou	22
Arctium lappa	Niu Bang Zi	56	*Lycopodium serratum*	Jin Bu Huan	33
Areca catechu	Da Fu Pi	12	*Magnolia obovata*, *M. officinalis*	Hou Po	54
Arisaema amurense, *A. erubescens*	Tian Nan Xing	13	*Magnolia* spp.	Xin Yi Hua	85
Aristolochia debilis	Qing Mu Xiang	14	*Melandrium firmum*	Wang Bu Liu Xing	4
Aristolochia fanchi	Guang Fang Ji	5	*Melia azedarach*, *M. toosendan*	Chuan Lian Zi	10
Aristolochia manshuriensis	Mu Tong	8	*Mentha arvensis*, *M. haplocalyx*	Bo He	39
Arnebia euchroma	Zi Cao	12	*Morinda citrifolia*, *M. officinalis*	Ba Ji Tian	22
Artemisia annua	Qing Hao	73	*Morus alba* (leaf)	Sang Ye	34
Artemisia capillaris, *A. scoparia*	Yin Chen Hao	76	*Morus alba* (twig)	Sang Zhi	42
Artemisia vulgaris, *A. argyi*, *A. princeps*	Ai Ye	106	*Myristica fragrans*	Rou Dou Kou	85
Asarum heterotropoides, *A. sieboldii*	Xi Xin	32	*Nandina domestica*	Tian Zhu Zi	23
Asparagus cochinchinensis	Tian Men Dong	11	*Nardostachys chinensis*, *N. jatamansi*	Gan Song Xiang	54
Aster tataricus	Zi Wan	29	*Nelumbo nucifera* (leaf)	He Ye	19
Astragalus membranaceus, *A. mongholicus*	Huang Qi	53	*Nelumbo nucifera* (seed)	Lian Zi	20
Atractylodes japonica, *A. macrocephala*	Bai Zhu	22	*Notopterygium forbesii*, *N. incisum*	Qiang Huo	34
Atractylodes lancea	Cang Zhu	43	*Ophiopogon japonicus*, *O. ohwii*	Mai Men Dong	42
Belamcanda chinensis	She Gan	14	*Osmunda japonica*	Guan Zhong	5

Table 2. (Continued)

species	herb name	NC	species	herb name	NC
Bidens bipinnata, *B. pilosa*	Xian Feng Cao	38	*Paeonia albiflora*	Bai Shao Yao	5
Blechnum orientale	Guan Zhong	4	*Paeonia lactiflora*	Chi Shao Yao	10
Bletilla striata	Bai Ji	42	*Paeonia suffruticosa*	Mu Dan Pi	33
Boschniakia rossica	Rou Cong Rong	9	*Panax ginseng*	Ren Shen	126
Boswellia carteri, *B. serrata*	Ru Xiang	70	*Panax notoginseng*	San Qi	36
Brucea javanica	Ya Dan Zi	41	*Papaver somniferum*	Ying Su Ke	66
Bupleurum falcatum, *B. longiradiatum*, *B. scorzonerifolium*	Chai Hu	46	*Patrinia scabiosaefolia*, *P. villosa*	Bai Jiang	20
Caesalpinia sappan	Su Mu	30	*Perilla frutescens*	Zi Su Ye	34
Camptotheca acuminata	Xi Shu	16	*Periploca sepium*	Xiang Jia Pi	17
Carthamus tinctorius	Hong Hua	114	*Peucedanum decursivum*, *P. praeruptorium*	Qian Hu	36
Cassia angustifolia	Fan Xie Ye	17	*Pharbitis nil*, *P. purpurea*	Qian Niu Zi	23
Cassia obtusifolia, *C. tora*	Jue Ming Zi	37	*Phellodendron amurense*, *P. chinense*	Huang Bai	33
Chrysanthemum indicum, *C. sinense*	Ye Ju Hua	30	*Phryma leptostachya*	Tou Gu Cao	10
Chrysanthemum morifolium	Ju Hua	22	*Phytolacca acinosa*, *P. esculenta*	Shang Lu	35
Cimicifuga spp.	Sheng Ma	79	*Picrorhiza kurrooa*	He Huang Lian	27
Cinnamomum cassia	Gui zhi	97	*Pinellia pedatisecta*	Tian Nan Xing	12
Cirsium japonicum	Da Ji	22	*Pinus massoniana*	Song Jie	32
Cistanche deserticola, *C. salsa*	Rou Dou Kou	78	*Piper longum*	Bi Bo	16
Clematis chinensis, *C. hexapetala*, *C. manshurica*	Wei Ling Xian	13	*Piper nigrum*	Hu Jiao	38
Clematis armandii, *C. montana*	Mu Tong	5	*Plantago asiatica*	Che Qian Zi	18
Clerodendron trichotomum	Chou Wu Tong	13	*Platycladus orientalis* (leaf)	Ce Bai Ye	21
Cnidium monnieri	She Chuang Zi	43	*Platycladus orientalis* (seed)	Bai Zi Ren	23
Cocculus trilobus	Guang Fang Ji	12	*Platycodon grandiflorum*	Jie Geng	15
Codonopsis pilulosa, *C. tanshen*	Dang Shen	47	*Pogostemon cablin*, *P. heyneanus*	Guang Huo Xiang	56
Commiphora spp.	Mo Yao	66	*Polygala sibirica*, *P. tenuifolia*	Yuan Zhi	56
Coptis chinensis, *C. japonica*	Huang Lian	26	*Polygonatum odoratum*	Yu Zhu	8
Coriandrum sativum	Yan Sui Zi	78	*Polygonum multiflorum*	He Shou Wu	27
Cornus officinalis	Shan Zhu Yu	34	*Polyporus umbellatus*	Zhu Ling	9
Corydalis spp.	Yan Hu Suo	20	*Poria cocos*	Fu Ling	21
Crocus sativus	Xi Hong Hua	25	*Prunella vulgaris*	Xia Ku Cao	23
Croton tiglium	Ba Dou	21	*Psoralea corylifolia*	Bu Gu Zhi	36
Curculigo orchioides	Xian Mao	22	*Pueraria lobata*, *P. thomsonii*	Ge Gen	49
Curcuma aromatica, *C. longa*	Yu Jin	49	*Pulsatilla chinensis*, *P. dahurica*	Bai Tou Weng	15
Curcuma aromatica, *C. zedoaria*	E Zhu	64	*Punica granatum*	Shi Liu Pi	28
Cuscuta chinensis	Tu Si Zi	26	*Pyrrosia lingua*, *P. petiolosa*	Shi Wei	22
Cyathula capitata	Tu Niu Xi	10	*Raphanus sativus*	Lai Fu Zi	21
Cynanchum atratum, *C. versicolor*	Bai Wei	20	*Rehmannia glutinosa*	Di Huang	42
Cyperus rotundus	Xiang Fu	30	*Rheum palmatum* (*Rheum* spp.)	Da Huang	109
Dalbergia odorifera	Jiang Xiang	36	*Rhus vernicifera*	Gan Qi	9
Daphne genkwa	Yuan Hua	11	*Rosa laevigata*	Jin Ying Zi	19
Dendrobium spp.	Shi Hu	28	*Rubia cordifolia*	Qian Cao Gen	67
Dianthus chinensis, *D. superbus*	Qu Mai	26	*Salvia miltiorrhiza*	Dan Shen	82
Dictamnus dasycarpus	Bai Xian Pi	20	*Sanguisorba officinalis*	Di Yu	34
Dioscorea batatas, *D. opposita*	Shan Yao	14	*Santalum album*	Tan Xiang	50
Dioscorea bulbifera	Huang Yao Zi	15	*Saussurea lappa*	Mu Xiang	61
Dioscorea collettii, *D. tokoro*	Bei Xie	26	*Schisandra chinensis*, *S. sphenanthera*	Wu Wei Zi	67
Dipsacus asperoides	Xu Duan	7	*Schizonepeta tenuifolia*	Jing Jie	21

Table 2. (Continued)

species	herb name	NC	species	herb name	NC
Dolichos lablab	Bai Bian Dou	18	*Scrophularia ningpoensis*	Xuan Shen	9
Dracaena cinnabari, *D. draco*	Xue Jie	64	*Scutellaria baicalensis*	Huang Qin	57
Dryopteris crassirhizoma	Guan Zhong	13	*Scutellaria barbata*, *S. rivularis*	Ban Zhi Lian	30
Eclipta alba, *E. prostrata*	Han Lian Cao	33	*Siegesbeckia orientalis*, *S. pubescens*	Xi Xian Cao	42
Elscholtzia ciliata	Xiang Ru	10	*Sinomenium acutum*	Han Fang Ji	10
Ephedra sinica	Ma Huang	10	*Smilax aristolochiaefolia*, *S. glabra*, *S. officinalis*	Tu Fu Ling	13
Epimedium grandiflorum, *E. koreanum*, *E. sagittatum*	Yin Yang Huo	83	*Smilax sieboldii*	Wei Ling Xian	11
Equisetum arvense, *E. hyemale*	Mu Zei	17	*Sophora flavescens*	Ku Shen	63
Eriobotrya japonica	Pi Pa Ye	44	*Sophora subprostrata*	Shan Dou Gen	39
Erythrina arborescens, *E. variegata*	Hai Tong Pi	55	*Sparganium stoloniferum*	San Leng	8
Eucommia ulmoides	Du Zhong	40	*Stemona japonica*, *S. sessilifolia*, *S. tuberosa*	Bai Bu	32
Eugenia caryophyllata	Ding Xiang	16	*Stephania tetrandra*	Han Fang Ji	18
Eupatorium chinense	Tu Niu Xi	13	*Swertia mileensis*	Qing Ye Dan	11
Euphorbia helioscopia	Ze Qi	48	*Syzygium aromaticum*	Ding Xiang	13
Euphorbia kansui	Gan Sui	13	*Taraxacum japonicum*, *T. officinale*	Pu Gong Ying	24
Evodia rutaecarpa	Wu Zhu Yu	40	*Terminalia chebula*	He Zi	32
Foeniculum vulgare	Xiao Hui Xiang	144	*Tetrapanax papyriferum*	Tong Cao	28
Forsythia suspensa, *F. viridissima*	Lian Qiao	31	*Tribulus terrestris*	Bai Ji Li	37
Fritillaria cirrhosa, *F. delavayi*, *F. unibracteata*	Chuan Bei Mu	13	*Trichosanthes kirilowii*	Tian Hua Fen	15
Fritillaria thunbergii	Zhe Bei Mu	38	*Trigonella foenum-graecum*	Hu Lu Ba	39
Ganoderma lucidum	Ling Zhi	137	*Tripterygium hypoglaucum*	Zi Jin Pi	13
Gardenia jasminoides	Zhi Zi	30	*Tripterygium wilfordii*	Lei Gong Teng	139
Gastrodia elata	Tian Ma	22	*Tussilago farfara*	Kuan Dong Hua	54
Gentiana lutea	Long Dan Cao	19	*Uncaria macrophylla*, *U. rhynchophylla*, *U. sinensis*	Gou Teng	38
Gentiana macrophylla	Qin Jiao	11	*Vaccaria segetalis*	Wang Bu Liu Xing	13
Ginkgo biloba	Bai Guo Ye	58	*Veratrum grandiflorum*, *V. nigrum*	Li Lu	39
Glehnia littoralis	Bei Sha Shen	40	*Vitex rotundifolia*, *V. trifoliata*	Man Jing Zi	20
Glycine max	Dou Chi	42	*Xanthium sibiricum*, *X. strumarium*	Cang Er Zi	19
Glycyrrhiza inflata, *G. uralensis*	Gan Cao	136	*Zanthoxylum bungeanum*, *Z. piperitum*, *Z. schinifolium*	Chuan Jiao	47
Hedyotis diffusa	Bai Hua She She Cao	11	*Zea mays*	Yu Mi Xu	68
Houttouynia cordata	Yu Xing Cao	9	*Zingiber officinale*	Sheng Jiang	50
Hypericum japonicum	Tian Ji Huang	31	*Zizyphus spinosus*	Suan Zao Ren	35

accounted for is 8165. The remaining compounds, 99 in all, are omitted as they are found in herbs the taxonomic status of which requires further verification.

Figure 6 shows the distribution of reported compounds among the various Chinese herbs. It can be seen that, for most herbs, between 11 and 50 compounds have been reported, though the number with more than 50 known compounds is still appreciable. Only 23 herbs have less than 10 compounds reported so far.

It should be appreciated that these numbers only represent a fraction of the total number of constituents for each herb, though secondary metabolites of pharmacological interest are well-represented for most herbs.

4. Targets of Bioactive Plant Compounds. Table 1 gives details of the 78 targets presently covered in the BPCD, the number of plant compounds associated with each, and the therapeutic classes into which they fall. The compounds are inhibitors unless otherwise stated.

There is considerable variation in the numbers of compounds associated with each target. Those with the highest number include ion channels, cAMP phosphodiesterases, nitric oxide, protein kinases, HIV enzymes, anti-inflammatory targets [cyclooxygenases (COX) and lipoxygenases (LOX)], and aldose reductase. Data on compounds which affect taste receptors have also been included, though these

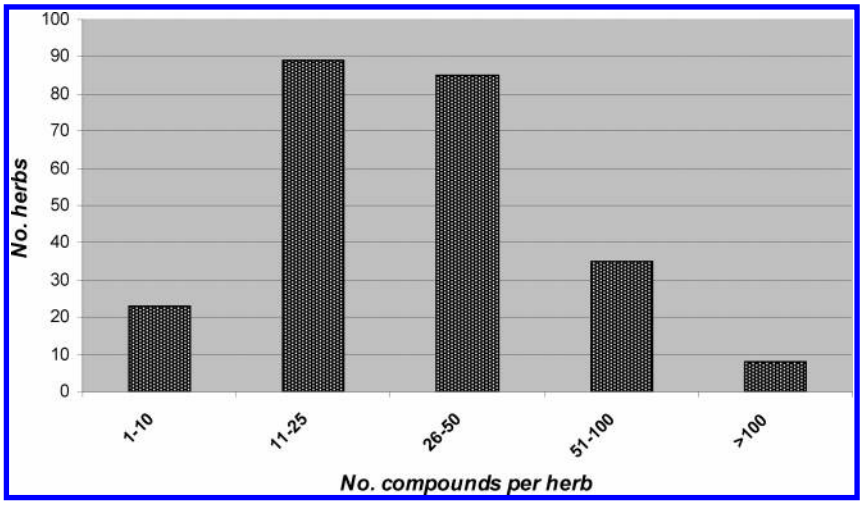


Figure 6. Distribution of compounds isolated from Chinese herbs.

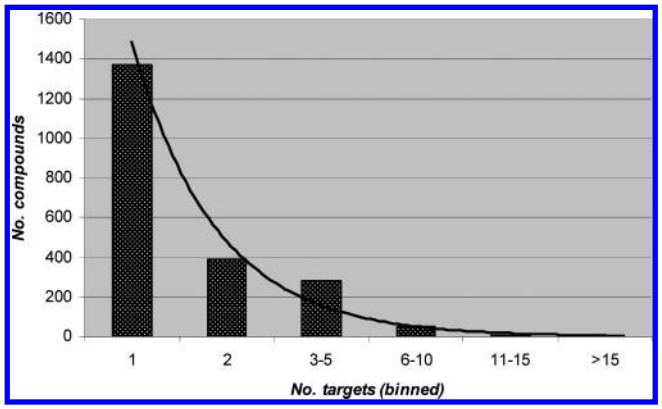


Figure 7. Compound-target relationships in the BPCD (exponential relationship as indicated by curve: $y = 4561.6\ e^{-1.124x}$).

may be considered preliminary and are not considered further.

A variety of targets for which significant numbers of inhibitors are known are not currently covered. These include metabotropic receptors (serotonin and glutamate) and the platelet-activating factor receptor.

4.a. Multiple Targets and Compound−Target Relationships. Many plant compounds are known to inhibit multiple targets, and a breakdown of this is shown in Figure 7. The number of targets (excluding taste receptors) inhibited by any one compound were binned into the five categories shown, the number of compounds found in each category decreasing in an approximately exponential fashion. Given that little is still known about the receptor affinities of plant compounds, it is likely that this relationship significantly underestimates the true extent of multiple-target inhibition.

4.b. Therapeutic Categories and Phytochemical Classes. Figure 8 gives details of the proportion of compounds in 12 phytochemical classes against targets in the main therapeutic categories listed in Table 1. Taste receptors and the heterogeneous collection of targets listed under metabolism are omitted. Limonoids, lignans, monoterpenes, and carbohydrates are also omitted because of low numbers.

Aliphatics are notable for the large proportion of compounds active against COX and LOX as well as a smaller proportion active against targets involved in iNOS expression, another inflammation-related process. Simple phenolics also contain a high proportion of compounds active against

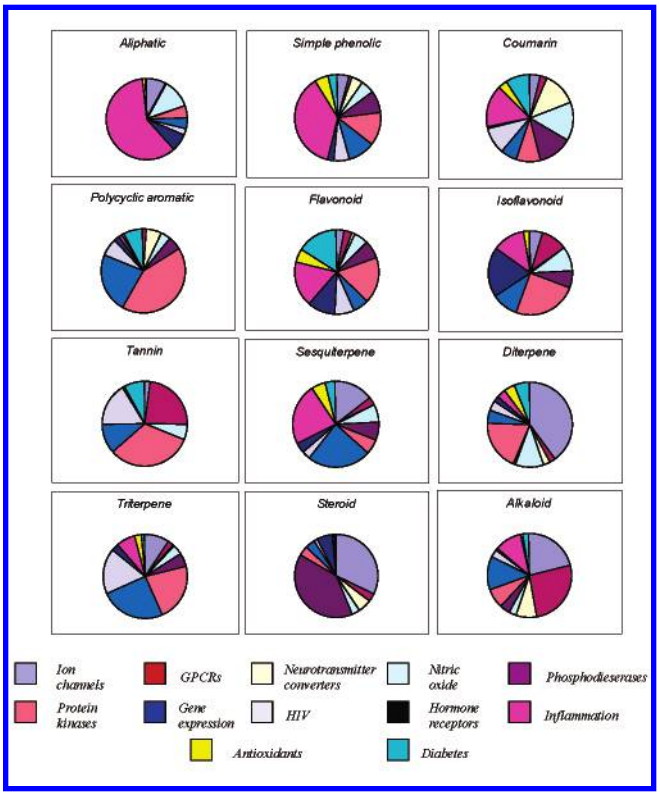


Figure 8. Patterns of target distribution for 12 phytochemical classes in the BPCD.

COX and LOX, in addition to a smaller number active against protein kinases. While coumarins are spread rather evenly between the different categories, polycyclic aromatics are highly uneven and show a strong association with protein kinases, though targets associated with gene expression are also prominent.

Flavonoids and isoflavonoids exhibit similar spectra, though aldose reductase inhibitors (diabetes), which are prominent among flavonoids, are lacking among isoflavonoids, as are HIV inhibitors. Like anthraquinones, tannins have a high proportion of protein kinase inhibitors and also inhibit HIV targets and GPCRs to an appreciable extent.

Patterns of association for terpenoid compounds should be treated with caution as, with the possible exception of triterpenes, these are currently under-represented in the BPCD. Diterpenes and steroids, however, both have a high

proportion of compounds which affect ion channels. The latter are also notable for the large number of compounds which inhibit cAMP phosphodiesterases. In the case of triterpenes, inhibitors of protein kinases, gene expression, and HIV targets (largely HIV protease), are represented more fully than other categories.

Finally, alkaloids are high in compounds which affect ion channels, GPCRs, and neurotransmitter converters (mainly acteylcholinesterase and monoamine oxidase in the latter category), reflecting their important effects on the nervous system. A smaller number influence gene expression and inflammation.

Further data are required before more detailed statistical analyses can be undertaken, particularly in the case of terpenoid compounds.

CONCLUSIONS

Sources of information on plant compounds which integrate chemical structure, pharmacology, molecular targets, and botanical sources including patterns of traditional usage, are of interest both in terms of drug discovery and in furthering our understanding of herbal medicine. Though there are considerable relevant data available, the great majority are not in a format suitable for data mining and molecular modeling. Recent years have seen improvements in this respect, though much still remains to be done. The work outlined here represents one such contribution to this effort. Information concerning the use and availability of the databases reported here is available from the authors on request.

REFERENCES AND NOTES

(1) Grabley, S.; Thiericke, R. In *Drug Discovery from Nature*; Grabley, S., Thiericke, R., Eds.; Springer-Verlag: Berlin, 1999; Chapter 1, pp 1−33.
(2) Sneader, W. *Drug Discovery: A History.* Wiley: Chichester, 2005; Chapters 9−11, pp 88−150.
(3) *Traditional Chinese Medicines 2.1*; CambridgeSoft Inc.: Cambridge, MA.
(4) *China Natural Products Database*; NeoTrident Technology Ltd.: Beijing, P. R. China.
(5) Leung S.; Chan, K.; Song, L. Publishing Trends in Chinese Medicine and Related Subjects Documented in WorldCat. *Health Inf. Libraries J.* **2006**, *23*, 13−22.
(6) Zhu, Y.-P. *Chinese Materia Medica: Chemistry, Pharmacology and Applications*; Harwood Academic Publishers: Netherlands, 1998.
(7) Huang, K.-C. *The Pharmacology of Chinese Herbs*; CRC Press: Boca Raton, FL, 1999.
(8) Polya, G. *Biochemical Targets of Plant Bioactive Compounds: A Pharmacological Reference Guide to Sites of Action and Biological Effects*; Taylor & Francis: London, 2003.
(9) Hsu, H.-Y.; Chen, Y.-P.; Hong, M. *The Chemical Constituents of Oriental Herbs*; Oriental Healing Arts Institute: Taiwan, 1982−83; Vols. I and II.
(10) Yan, X.; Zhou, J.; Xie, G. *Traditional Chinese Medicines: Molecular Structures, Natural Sources, and Applications*; Milne, G. W. A., Ed.; Ashgate Publishers Ltd.: Aldershot, U. K., 1999.
(11) Duke, J. A. *Handbook of Phytochemical Comstituents of GRAS Herbs and Other Economic Plants*; CRC Press: Boca Raton, FL, 1992.
(12) *Dictionary of Natural Products* [CD-ROM]; Chapman & Hall/CRC: New York.
(13) Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J. Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 244−255.
(14) Weininger, D. SMILES, A Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.
(15) Giles, P. M. Revised Section F: Natural Products and Related Compounds (IUPAC Recommendations 1999). *Pure Appl. Chem.* **1999**, *71*, 587−643.
(16) Rhodes, P. H. *The Organic Chemist's Desk Reference*; Chapman & Hall/CRC: London, 1995.
(17) Sweet, D. V.; Anderson, V. P.; Fang, J. C. F. An Overview of the Registry of Toxic Effects of Chemical Substances (RTECS): Critical Information on Chemical Hazards. *Chem. Health Saf.* **1999**, *6* (6), 12−16.
(18) Bensky, D.; Clavey, S.; Stöger, E. *Chinese Herbal Medicine: Materia Medica*; Eastland Press: Seattle, WA, 2004.
(19) Nakanishi, K. *A Wandering Natural Products Chemist*; American Chemical Society: Washington, DC, 1991; p 87.

CI600288M