British Journal of Pharmacology (2006) 149, 1092–1103
 0007–1188/06 $30.00
www.brjpharmacol.org

RESEARCH PAPER

Database of traditional Chinese medicine and its application to studies of mechanism and to prescription validation

X Chen[1], H Zhou[2], YB Liu[1], JF Wang[2], H Li[2], CY Ung[2], LY Han[2], ZW Cao[3] and YZ Chen[2]

[1]*College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang, China;* [2]*Department of Pharmacy and Department of Computational Science, National University of Singapore, Singapore, Singapore and* [3]*Shanghai Center for Bioinformation Technology, Shanghai, China*

Background and purpose: Traditional Chinese Medicine (TCM) is widely practised and is viewed as an attractive alternative to conventional medicine. Quantitative information about TCM prescriptions, constituent herbs and herbal ingredients is necessary for studying and exploring TCM.
Experimental approach: We manually collected information on TCM in books and other printed sources in Medline. The Traditional Chinese Medicine Information Database TCM-ID, at http://tcm.cz3.nus.edu.sg/group/tcm-id/tcmid.asp, was introduced for providing comprehensive information about all aspects of TCM including prescriptions, constituent herbs, herbal ingredients, molecular structure and functional properties of active ingredients, therapeutic and side effects, clinical indication and application and related matters.
Results: TCM-ID currently contains information for 1,588 prescriptions, 1,313 herbs, 5,669 herbal ingredients, and the 3D structure of 3,725 herbal ingredients. The value of the data in TCM-ID was illustrated by using some of the data for an *in-silico* study of molecular mechanism of the therapeutic effects of herbal ingredients and for developing a computer program to validate TCM multi-herb preparations.
Conclusions and Implications: The development of systems biology has led to a new design principle for therapeutic intervention strategy, the concept of 'magic shrapnel' (rather than the 'magic bullet'), involving many drugs against multiple targets, administered in a single treatment. TCM offers an extensive source of examples of this concept in which several active ingredients in one prescription are aimed at numerous targets and work together to provide therapeutic benefit. The database and its mining applications described here represent early efforts toward exploring TCM for new theories in drug discovery.
British Journal of Pharmacology (2006) **149,** 1092–1103. doi:10.1038/sj.bjp.0706945; published online 6 November 2006

Keywords: Chinese medicinal plant; drug; herbal medicine; herbal formula; natural product; pharmaceuticals

Abbreviations: AI, artificial intelligence; *k*NN, *k* nearest neighbor; PDB, protein data bank; SARS, severe acute respiratory syndrome; SVM, support vector machine; TCM, traditional Chinese medicine; TCM-ID, traditional Chinese medicine information database; TDHPs, traditionally defined herbal properties

Introduction

Traditional Chinese medicine (TCM) has been widely used in ethnic communities for the treatment of a variety of diseases and it is recognized as an attractive alternative to conventional medicine (Tang and Eisenbrand, 1992; Chan, 1995; Cheng, 2000; Yuan and Lin, 2000; Bhuiyan *et al*., 2003; Wang *et al*., 2003; Lazar, 2004). A major therapeutic approach of TCM is the use of a mixture of herbs that collectively exert therapeutic actions and modulate other effects. It is hypothesized that the principal constituents provide the main therapeutic actions, secondary principal constituents enhance or assist the effects of the principal ones and the rest serve modulation roles such as treatment of accompanying symptoms, moderation of harshness and toxicity, enhancement of the delivery of herbal ingredients and harmonization (Yuan and Lin, 2000).

Because of a growing interest in therapeutic agents based on TCM, increasing effort has been directed towards scientific proof, clinical evaluation and molecular analysis of TCM (Tang and Eisenbrand, 1992; Chan, 1995; Cheng, 2000; Yuan and Lin, 2000; Chen and Ung, 2001; Chen *et al*.,

Correspondence: Associate Professor X Chen, College of Life Sciences, Zhejiang University, Zijinghua Road, 368, Hangzhou, Zhejiang 310058, China or YZ Chen, Department of Pharmacy and Department of Computational Science, National University of Singapore, Singapore 117543, Singapore.
E-mail: xinchen@zju.edu.cn or yzchen@cz3.nus.edu.sg
Received 20 June 2006; revised 11 August 2006; accepted 22 September 2006; published online 6 November 2006

2003). A Medline survey in 2001 (Wheeler *et al*., 2003) showed that there are 6504 Chinese herb-related articles published in 662 journals in the period of 1966–2001 (Pach *et al*., 2002). Our own Medline search using two separate keywords 'traditional Chinese medicine' and 'Chinese herbal' found 1996 and 3839 related publications in the period of 2001–2006, respectively, suggesting a significant level of effort directed at TCM research and clinical studies.

The scientific evaluation of TCM can be made easier by making available information about all aspects of TCM including herbal formulations, constituent herbs, herbal ingredients, molecular structure and functional properties of active ingredients, therapeutic and toxic effects, clinical indications and applications. Several databases have been developed for providing information about different aspects of TCM. For instance, the TCM database provides information about Chinese medicinal plants and constituent chemicals (He *et al*., 2001), the Chinese herbal medicine toxicology database describes scientific evidence on the toxicity of Chinese herbal medicine (Bensoussan *et al*., 2002) and the 3D structure database of components from Chinese traditional medicinal herbs gives the basic molecular properties and optimized 3D structure of the selected herbal ingredients (Qiao *et al*., 2002). The Chinese medicine sampler (http://www.chinesemedicinesampler.com/) and TCM (http://www.healthy.net/CLINIC/therapy/Chinmed/Index.asp) databases provide general information about the history, theory, diagnostics and examples of formulas of TCM. However, only two of these databases are freely accessible at present.

Traditional Chinese medicine information database, TCM-ID, was introduced (Wang *et al*., 2005b; Ji *et al*., 2006) as a resource to provide comprehensive information about all aspects of TCM including prescriptions, constituent herbs in each prescription, herbal ingredients, molecular structure and functional properties of active compounds, clinical indication and application of each formula and each herb, therapeutic and toxicity effects of herbal ingredients and related literatures. The information in TCM-ID is from a comprehensive search of available literatures about TCM (Ji, 1994; Chen, 1998; Zhang, 1998; Huang, 1999) and the abstracts of Medline (Wheeler *et al*., 2003).

The value of the data in TCM-ID can be illustrated by assessing to what extent problems such as the study of mechanism of TCM herbs and validation of TCM multi-herb preparations can be solved based on the available data. To address this question, some of the data in TCM-ID were used in two separate studies. The first was to use the 3D structure of specific herbal ingredients to conduct an *in silico* search for their molecular targets. The identified targets were further examined to find out whether the known therapeutic effects of these ingredients can be interpreted by the expected effects of the interference with these targets. In the second, we used known TCM prescriptions to develop an artificial intelligent (AI) system to validate new TCM multi-herb preparations. The developed AI system was tested by using a number of newly published TCM prescriptions not yet included in TCM-ID.

Methods

Data collection

Information in the TCM-ID database was manually collected from TCM books (Zhang, 1988, 1998; Ji, 1994; Chen, 1998; Huang, 1999) and papers in Medline (Wheeler *et al*., 2003), particularly those publications in relevant journals such as *Complementary Therapies in Medicine*, *Journal of Alternative and Complementary Medicine*, *American Journal of Chinese Medicine*, *Chinese Traditional and Herbal Drugs*, *Planta Medica*, *Journal of Pharmaceutical Sciences*, *Acta Pharmaceutical Sinica*, *Phytochemistry*, *Journal of Chinese medicinal materials and Chinese Journal of Medicinal Chemistry*.

In silico *search for molecular targets of specific herbal ingredients*

The TCM herbal ingredients studied were genistein, ginsenoside Rg1 and baicalin. The *in silico* method for identifying their molecular targets is INVDOCK (Chen and Zhi, 2001; Chen and Ung, 2002; Chen *et al*., 2003), which conducts an automated search of every human and mammalian protein entry in the protein 3D structure database, Protein Data Bank (PDB) (Berman *et al*., 2000) to identify proteins to which each of these TCM herbal ingredient can bind. INVDOCK is based on a ligand–protein inverse docking strategy such that a compound is docked to known ligand-binding pockets of each of the protein entries in the PDB database. A protein is considered as a potential target of a compound if that compound can be docked into the protein and the binding satisfies a molecular-mechanics-based criterion for chemical complementarity (Chen and Zhi, 2001). Because of their capacity to identify potential ligands and binding conformations, docking algorithms are expected to be equally applicable to the inverse docking procedure for finding multiple protein targets to which a compound can bind, strongly or weakly (Chen and Ung, 2001; Chen and Zhi, 2001). Predicted molecular targets of herbal ingredients were manually correlated with published material.

Classification of TCM prescriptions and non-TCM recipes

In silico methods were developed for determining whether or not a multi-herb preparation is a valid TCM prescription (Wang *et al*., 2005a), in which AI systems were trained and tested by using all of the 1588 TCM prescriptions in the TCM-ID database and non-TCM recipes generated by random combination of multiple herbs and the modification of existing TCM prescriptions. The detailed approach and the underpinning TCM prescription principle for generating non-TCM recipes are elaborated in the discussion section.

In training and testing the AI systems, digital forms of these prescriptions and recipes were needed. TCM practitioners formulate TCM multi-herb preparations according to the patient's condition and traditionally defined herbal properties (TDHPs) (Chan, 1995). In this work, we used a method originally developed by Su (1997) to digitize the TDHPs of all of the constituent herbs in these TCM prescriptions and non-TCM recipes. The resulting vector representations of the prescriptions and recipes were used as training and testing sets for AI systems.

Two AI systems were explored for their usefulness in validating TCM prescriptions. One was '*k* nearest neighbour' (*k*NN) (Johnson and Wichern, 2002), which uses the distance between a multi-herb preparation and each of the TCM prescriptions and non-TCM recipes in the training sets to determine whether it is a valid TCM prescription. The other was 'support vector machine' (SVM), which projects a multi-herb preparation into a hyperspace where TCM prescriptions are separated from non-TCM recipes by a hyper-plane (Byvatov and Schneider, 2003; Wang *et al*., 2005a). The two systems both have parameters to be fine-tuned with the data to perform optimally, such as the 'number of votes' parameter k in *k*NN approach and the 'kernel width' parameter σ in SVM approach with a Gaussian kernel, as in our case.

In order to find the optimal parameter and, at the same time, obtain an unbiased estimation of the classification accuracies of the AI systems, we employed two commonly applied methods (Majumder *et al*., 2005; Yu and Chen, 2005; Xu *et al*., 2006) in classification researches. One is the threefold cross validation method and the other is the use of the independent evaluation set. With the cross validation approach, we first randomly divided the dataset into three groups, and trained the AI systems three times with the same parameter value, each time leaving out one of the groups from training, but using only the omitted group to compute the accuracy measures. The system performance with the specific parameter value was then calculated as the average of the three measurements obtained. Finally, the optimal parameter value was selected from an empirical range to maximize the average system accuracy. The average classification accuracy obtained with the optimal parameter value was considered as the unbiased estimation of system accuracy on unseen data. On the other hand, by using the independent evaluation set, we first randomly divided the data into training set, testing set and independent evaluation set. Parameter selection with this approach was done only with the training and testing sets. With any parameter value taken from an empirical range, an AI system was trained with the training set and its accuracy was measured against the testing set. Once the optimal parameter value was chosen to maximize the system performance on the testing set, the independent evaluation set was then used to calculate the unbiased estimate of the accuracy of the system on unseen data.

Results

TCM-ID has a web interface at http://tcm.cz3.nus.edu.sg/group/tcm-id/tcmid.asp. It currently contains 1588 TCM prescriptions covering 4111 disease conditions, 1313 medicinal herbs used in known TCM prescriptions and 5669 ingredients known to be contained in TCM herbs. It also provides the 3D structure of 3725 ingredients. This database is searchable by prescription, herb or ingredient name. It can also be accessed by input of a particular disease condition or selection of a specific therapeutic effect. Table 1 gives the categories and examples of typical search terms of the TCM-ID database.

Searches involving any combination of these options or selection fields are also supported. The search words are case insensitive. In a query, a user can specify the full name or any part of the name in a text field, or choose one item from a selection field. Wild characters of '*' and '?' are supported in the text field. Here, '?' represents any one character and '*' represents a string of any characters of any length. For example, input of 'ginseng' in the herb name field finds

Table 1 Categories of typical search terms of TCM-ID database

Category	*Type of information*	*Search term*	*Examples*
Prescription	Name	Chinese name	Wu Ji Bai Feng Wan (乌鸡白凤丸); Liu Wei Di Huang Wan (六味地黄丸)
		Common name	White Phoenix Bolus of Black-bone Chicken; Six Ingredient Pill with Rehmannia
	Function	Traditionally described functions	Activate the flow of qi; nourish qi; promote qi; clear heat; expel phlegm; reduce fever; dispel cold
	Clinical manifestations	Keywords of symptom, disease, therapeutic effect	Headache; stomach pain; hypertension; diarrhoea; cases of fever;
Herb	Name	English name	Ginseng; fresh rehmannia root
		Latin name	Radix Ginseng; Radix Rehmanniae
		Chinese name	Ren Shen (人参); Sheng Di Huang (生地黄)
	Function	Traditionally described functions	Promote qi circulation; clear heat and cool body
	Therapeutic class	Traditionally described classes	For tonifying weakness; heat clearance
Ingredient	Name	Compound name	Palmitic acid; Ginsenoside Rg1; Rehmaglutin A; Campesterol
	ID	CAS registry number	57-10-3; 22427-39-0; 103744-82-7; 474-62-4
Disease condition (via keyword search)	Disease name	Traditional and modern names	Hypertension; diabetes; loss of qi
Therapeutic effect (via keyword search)	Effect description	Traditional and modern terms	Lower blood sugar; inhibit gut motility; bronchodilatation; increase qi

entries, of which either 'Chinese Name', 'Latin Name', 'Common Name' or 'English Name' contains 'ginseng' in the text description, on the other hand, input of 'gin*' finds entries with names like ginseng, ginkgo, Panax notoginseng (Burk.), and so on.

The result of a typical search is illustrated in Figure 1, which was obtained by using either the Chinese name 'Wu Ji Bai Feng Wan' or the common name 'White Phoenix Bolus of Black-bone Chicken'. In this interface, all of the entries for the TCM prescription formula, herbs or ingredients that match the search criteria are listed. Detailed information of a particular entry can be obtained by clicking the corresponding recipe, herb or ingredient name. The typical information page is shown in Figure 2. From this page, one finds comprehensive information about the Latin name, indigenous name, medicinal name, plant collection site, condition of plant, plant part used for content analysis, known therapeutic effects, analysis method, compound class and the contents of the compound class and individual compounds.

In order to illustrate the usefulness of this database, we present two applications as examples, the *in silico* investigation of the molecular mechanism underpinning herbal therapeutics and the validation of TCM multi-herb recipe by AI approach. In the first application, we used the INVDOCK (Chen and Zhi, 2001) program to identify potential therapeutic targets for three well-studied active herbal ingredients selected from TCM-ID, namely genistein, ginsenoside Rg1 and baicalin. In Tables 2–4, are given all of the INVDOCK identified targets of these three ingredients, respectively.

Out of the 30 putative therapeutic targets of genistein identified by INVDOCK (Table 2), seven targets had relevant literature support. Moreover, there are nine therapeutic effects known to result from drug binding to these 30 targets, of which eight match the reported beneficial effects of genistein. Similarly, INVDOCK predicted three therapeutic targets of ginsenoside Rg1, two of which have relevant literature support (Table 3). The predicted targets are known to be involved in three therapeutic effects, and the effectiveness of ginsenoside Rg1 to produce these three therapeutic effects was consistently supported by available literatures. INVDOCK predicted 14 targets for baicalin (Table 4). There was literature support for five of these targets and for four possible therapeutic effects. These results suggest the usefulness of TCM-ID and INVDOCK as an *in silico* tool in facilitating the identification of potential therapeutic targets of the herbal ingredients and thus providing valuable clues to the mechanisms of TCM prescriptions and their possible secondary therapeutic effects.

In the other application, we trained two AI systems with the TCM prescriptions collected in the database and non-TCM recipes generated by random perturbation of known prescriptions. Both systems showed more then 80% accuracy in recognizing traditional TCM prescriptions and over 98% accuracy in rejecting non-TCM recipes (Table 5). Moreover, one of the AI systems, the SVM system, recognized 68.7% of the recently published new TCM prescriptions (Table 6), suggesting the usefulness of TCM-ID data and AI systems in facilitating the validation and analysis of TCM

TCMID Result(Formula)	
Chinese Name	Wu Ji Bai Feng Wan (乌鸡白凤丸)
Common Name	White Phoenix Bolus of Black-bone Chicken
Function	"To repenish qi and nourish blood, regulate menstruation and arrest leukorrhagia."
Herbal Components	Wu Ji: Lu Rong Jiao: Bie Jia: Mu Li: Sang Piao Xiao: Ren shen: Huang qi: Dang gui: Bai shao: Xiang fu: Tian Dong: Gan cao: Sheng Di Huang: Shu Di Huang: Chuan xiong: Yin Chai Hu: Dan Shen: huai Shan Yao: Qian shi: Lu Jiao Shuang:

Figure 1 Detailed information on a TCM prescription.

TCMID Result (Herb)	
Chinese Name	Ren shen (人参)
Latin Name	Radix Ginseng
English Name	Ginseng
Plant/Source	Panax ginseng C. A. Mey.
Properties	Minor Warm,Sweet,Slightly Bitter
Meridians	Lung,Spleen,Heart
Function	To reinforce the vital energy, to remedy collapse and restore the normal pulse, to benefit the spleen and lung, to promote the production of body fluid, and to calm the nerves.
Indications	Treatment of prostration with impending collapse marked by cold limbs and faint pulse, diminished function of the spleen with loss of appetite, cough and dyspnea due to diminished function of the lung, thirst due to impairment of body fluid, of diabetes caused by internal heat, general weakness with irritability and insomnia in chronic diseases, impotence or frigidity, heart failure, cardiogenic shock.
Clinical Manifestations	1. Acting on the central nervous system including excitation and inhibition processes.2. Exerting significant cardiotonic and hypertensive effects on acute circulatory failure cardiotonic and hypertensive effects on acute circulatory failure after heavy blood loss.3. Decreasing the level of blood sugar.4. Promoting phgocytosis and enhancing lymphocyte-blastogenesis rate.
Collection	Ginseng is the dried root of Panax ginseng C. A. Mey. (Fam. Araliaceae). The drug derived from the cultivated form is known as " Yuanshen " (Garden Ginseng) and the drug derived from the wild origin is known as "Shanshen" (Wild Ginseng). The drug is collected in autumn and washed clean. Sun dried or bake-dried Yuanshen is known as "Shengshaishen" (Sun-dried Ginseng). Sun-dried Shanshen is known as "Sheng-shaishanshen" (Sun-dried Wild Ginseng).
Processing	Sun-dried Ginseng Soften thoroughly, cut into slices, and dry.Sun-dried Wild Ginseng Pulverize or break to pieces before use.
Description	Sun-dried Ginseng Main roots fusiform or cylindrical, 3-15 cm long, 1-2 cm in diameter; externally greyish-yellow, upper part or entire root exhibiting sparse, shallow, interrupted and coarse transverse-striations and distinct longitudinal wrinkles; lower
Usage And Dosage	3-9 g.
Storage	Preserve in well closed containers, stored in a cool and dry place, protected from moth.
Therapeutic Class	For tonifying weakness (补虚)
Therapeutic Sub Class	For tonifying qi (补气)
Compound Ingredients	methyl palmitate 2-methyl-tetradecane 2-heptadecanone Palmitic acid Patchouli alcohol n-pentadecane Nonadecane Humulene Trans-caryophyllene Ginsenoside Rg1 Ginsenoside Rc Ginsenoside Rd Ginsenoside Rb1 Ginsenoside Rb2 Ginsenoside Rb0
Reference	Shennong Bencao Jing

Figure 2 Detailed information on a TCM herb.

prescriptions. Details of these applications are discussed in the following section.

Discussion and conclusions

Mechanistic study of TCM herbal ingredients

We chose genistein, ginsenoside Rg1 and baicalin to study their putative therapeutic targets as predicted by INVDOCK. Based on a comprehensive MEDLINE search of related publications, there was a relatively large number of references available for these selected herbal ingredients. It is essential to have a reasonable amount of experimental findings for each compound, so that this search procedure can provide a meaningful evaluation.

Genistein is a soy-derived isoflavone of therapeutic interest. Dietary intake of soy is associated with a decreased risk of both hormone-dependent and hormone-independent cancers (Barnes and Peterson, 1995; Castle and Thrasher, 2002). At molecular level, genistein inhibits the activity of adenosine 5′ triphosphate binding enzymes such as tyrosine-specific protein kinase, topoisomerase II and

Table 2 Potential targets of genistein identified from an INVDOCK search of human and mammalian proteins

PDB ID	*Target name*	*Experimental findings*	*Therapeutic implications*
1a25	Protein kinase C		Vascular disease (Pan *et al*., 2001), Heart failure (Pan *et al*., 2001), Cancer (Theodorescu *et al*., 1998)
1a27	17-beta-hydroxysteroid-dehydrogenase	Genistein inhibits this enzyme (Evans *et al*., 1995).	
1a35	Topoisomerase I	Genistein has anti- topoisomerase I effect (Kikuchi and Hossain, 1999).	Malaria (Dluzewski and Garcia, 1996), Cancer (Theodorescu *et al*., 1998)
1a4r	G25 K GTP-binding protein	Genistein prevents agonist-induced G protein uncoupling (Reyes-Cruz *et al*., 2000).	Cancer (Theodorescu *et al*., 1998)
1a7c	Plasminogen activator inhibitor	Genistein shifts urokinase/plasminogen activator inhibitor proteolytic balance (Fajardo *et al*., 1999).	Angiogenesis (Pan *et al*., 2001)
1aa9	C-HA-RAS	Genistein blocks RAS activation (Takahashi *et al*., 1997).	Cancer (Theodorescu *et al*., 1998)
1akf	Estrogen receptor	Genistein binds to estrogen receptor beta (Barnes *et al*., 2000).	Cancer (Breast) (Morito *et al*., 2001)
1awn	Guanylyl cyclase		Cancer (Theodorescu *et al*., 1998)
1azm	Carbonic anhydrase I		Kidney failure (Tomobe *et al*., 1998), Glaucoma (Yousufzai and Abdel-Latif, 1998), Cancer (Theodorescu *et al*., 1998)
1b3o	Inosine dehydrogen		Malaria (Dluzewski and Garcia, 1996)
1bbz	ABL tyrosine kinase	Tyrosine kinase inhibitor (Nishimura *et al*., 1988).	Cancer (Theodorescu *et al*., 1998)
1bl7	MAP kinase P38	MAP kinase is blocked by genistein (Pecherskaya and Solem, 2000).	Cancer (Theodorescu *et al*., 1998), Arthritis (Martel-Pelletier *et al*., 1999)
1bpx	DNA polymerase		Cancer (Theodorescu *et al*., 1998), Herpes viral infection (Swa *et al*., 2001)
1bup	HSP-70 (70 kDa heat shock protein)		Inflammation (Sadowska-Krowicka *et al*., 1998), Arthritis (Martel-Pelletier *et al*., 1999)
1bx4	Adenosine kinase		Pain
1cpj	Cathepsin B		Arthritis (Martel-Pelletier *et al*., 1999)
1d3g	Dihydroorotate dehydrogenase		Malaria (Dluzewski and Garcia, 1996)
1d6n	Hypoxanthine-guanine phosphoribosyltransferase (HPRT)	Genistein marginally activates HPRT (Kulling and Metzler, 1997).	Malaria (Dluzewski and Garcia, 1996)
1d8d	Farnesyltransferase		Cancer (Theodorescu *et al*., 1998)
1db4	Phospholipase A2		Inflammation (Sadowska-Krowicka *et al*., 1998)
1di8	Cyclin-dependent kinase 2	Genistein suppresses CDK2 activity (Kuzumaki *et al*., 1998).	Cancer (Theodorescu *et al*., 1998)
1e0o	Fibroblast growth factor 1	FGF effects on scleraxis are blocked by genistein (Kawa-uchi *et al*., 1998).	Cancer (Theodorescu *et al*., 1998)
1fgi	FGF receptor 1	Genistein blocks cytoplasmic receptor domain (Munoz *et al*., 1997).	Cancer (Theodorescu *et al*., 1998)
1prg	Peroxisome proliferator		Diabetes (Orie *et al*., 2000)
1rts	Thymidylate synthase		Cancer (Theodorescu *et al*., 1998), Fungal infection
1ula	Purine nucleoside phosphorylase		Cancer (Theodorescu *et al*., 1998)
1vbt	Cyclophilin A		Immune response (Uckun *et al*., 1995)
2dhf	Dihydrofolate reductase		Bacterial infection (Sadowska-Krowicka *et al*., 1998), Cancer (Theodorescu *et al*., 1998), Malaria (Dluzewski and Garcia, 1996), Inflammation (Sadowska-Krowicka *et al*., 1998)
7odc	Ornithine decarboxylase (ODC)	Effective inhibitor of ODC (Flamigni *et al*., 1999).	Cancer (Theodorescu *et al*., 1998)
830c	Matrix metalloproteinase (MMP)-13	Genistein down-regulates the expression of MMP-13 (Kim *et al*., 2001)	Cancer (Theodorescu *et al*., 1998), Arthritis (Martel-Pelletier *et al*., 1999)

PDB ID is the accession number for a protein in the protein databank (PDB).

Table 3 Potential therapeutic targets of ginsenoside Rg1 identified from INVDOCK search of human and mammalian proteins

PDB ID	*Target name*	*Experimental findings*	*Therapeutic implications*
1rpa	DNA polymerase β	Rg1 stimulates DNA synthesis. Rg1 activates DNA polymerase delta (Cho *et al*., 1995; Lee and Lee, 1996).	Cancer (Shin *et al*., 2000)
1rmh	Cyclophilin A		Immune response (Kenarova *et al*., 1990)
3nos	Endothelial nitric-oxide synthase (NOS)	Rg1 inhibits NOS dose dependently (Li *et al*., 1997).	Maintaining optimal oxidative status (Kitts and Hu, 2000)

Table 4 Potential therapeutic targets of baicalin identified from INVDOCK search of human and mammalian proteins

PDB Id	*Target name*	*Experimental findings*	*Therapeutic implications*
121p	H-Ras p21 protein		Anticancer
1ads	Aldose reductase	Reduced RBC sorbitol levels in diabetic rats as inhibitor of aldose reductase (Zhou and Zhang, 1989)	Diabetes treatment
1agw	FGF receptor 1		Anticancer
1ayk	Collagenase		Inhibit metastasis process of cancerous cells
1a25	Protein kinase C		Anticancer
1awk	Adenylyl cyclase		Anticancer
1awn	Guanylyl cyclase		Anticancer
1irb	Phospholipase A2	Inhibition (Kyo *et al*., 1998)	Pharmacological action on glial cells involved in maintaining the function of neural cells.
1p38	MAP kinase p38		Anticancer
1rpa	Prostatic acid phosphatase		Anticancer (prostate cancer)
1ydr	cAMP-dependent protein kinase		Anticancer
2bpf	DNA polymerase β	Weak inhibition (Kitamura *et al*., 1998)	Anticancer
1jsu	Cyclin-dependent kinase-2	Decrease expression level of cyclin-dependent kinase (Liu *et al*., 1998)	Anticancer
1mmb	Matrix metalloproteinase-8	Downregulate expression level of matrix metalloproteinases (Kato *et al*., 1998)	Anticancer

Table 5 TCM prescription and non-TCM recipe classification accuracies of the artificial intelligence classification systems, *k*NN and SVM, evaluated by threefold cross-validation study

Cross validation	*Training set*		*Testing set*						
	TCM prescriptions	*Non-TCM recipes*	*TP*	*FN*	*P+ %*	*TN*	*FP*	*P− %*	*P%*
AI classification system: kNN									
1	388	7453	341	47	87.9	3706	43	98.8	97.8
2	384	7468	314	70	81.8	3699	35	99.1	97.4
3	389	7483	325	64	83.6	3670	49	98.7	97.3
Average Accuracy					**84.4**			**98.9**	**97.5**
AI classification system: SVM									
1	388	7453	360	28	92.8	3692	57	98.4	97.9
2	384	7468	342	42	89.1	3684	50	98.6	97.7
3	389	7483	358	31	92.0	3670	49	98.7	98.1
Average accuracy					**91.3**			**98.6**	**97.9**

P+, P− and P represent the classification accuracy for TCM prescriptions, non-TCM recipes and all recipes, respectively. TP, TN, FP and FN are the number of true positive (correctly classified TCM prescriptions), true negative (correctly classified non-TCM recipes), false positive (TCM prescriptions falsely classified as non-TCM recipes) and false negative (non-TCM recipes falsely classified as TCM prescriptions), respectively, and *N* is the total number of recipes.
Bold numerals are used to highlight the average accuracies of the two AI systems.

enzymes involved in phosphatidylinositol turnover. Moreover, genistein can act via an oestrogen receptor-mediated mechanism (Polkowski and Mazurek, 2000). INVDOCK identified 18 potential therapeutic targets of genistein. Two of these proteins have been reported to be directly inhibited by genistein *in vitro*. These targets are the oestrogen receptor (Barnes *et al*., 2000) and fibroblast growth factor (FGF) receptor 1 (Munoz *et al*., 1997). The oestrogen receptor β is reported to bind genistein with an affinity close to that of 17β-estradiol. However, it remains to be determined whether it is a mediator of genistein's activity *in vivo*. The tyrosine kinase domain of the FGF receptor 1 is essential for the activity of this protein. As a general tyrosine kinase inhibitor, genistein is expected to inhibit the intracellular signalling activity of the FGF receptor 1. These two proteins are known anticancer targets and the reported anticancer effect of genistein (Theodorescu *et al*., 1998) appears to be consistent with its predicted binding to these two anticancer targets.

Experiments have also shown that the activity or expression level of each of the following five INVDOCK identified therapeutic targets is affected by genistein. Ligand binding may influence the activity of a protein, and it is also known to self-regulate protein levels in certain cases (Schmidt and Meyer, 1994). Hence, there is a possibility that these observed phenomena indicate genistein's binding to each of these proteins as predicted by INVDOCK. The activities of cyclin-dependent kinase 2 (Kuzumaki *et al*., 1998), topoisomerase I (Kikuchi and Hossain, 1999) and prostaglandin H2 synthase (Kniss *et al*., 1996) have been reported to be suppressed by genistein. Cyclin-dependent kinase 2 is a therapeutic target for cardiovascular disease and genistein has been found to possess a preventative effect for cardiovascular disease (Pan *et al*., 2001). Topoisomerase I is another therapeutic target for cancer. Prostaglandin H2 synthase (COX) is a major therapeutic target for inflammation. Genistein has been reported to have possible

Table 6 List of 48 TCM prescriptions published in recent years

Correctly predicted as TCM prescription		
Xiao Qing Long He Ji	Liu Jun Zi Wan	Qi Zhi Xiang Fu Wan
Gui Zhi He Ji	Huang Qi Sheng Mai Yin	Xiang Sha Ping Wei Wan
Xiao Chai Hu Wan	Zhi Gan Cao He Ji	Xiang Ru Wan
Liang Ge He Ji	Shi Quan Da Bu Jiu	Tan Yin Wan
Ku Huang Zhu She Ye	Man Shen Bao Ye	Wei De An Jiang
Qing Gan Li Dan Kou Fu Ye	Gui Shao Di Huang Wan	A Wei Wan
Xiao Jian Zhong Chong Ji	Geng Nian Nv Bao Pian	Sheng Hua Tang Wan
Huang Qi Jian Zhong Wan	Da Bu Yuan Jian Wan	Tong Yu Jian
Wen Wei Shu Chong Ji	Zhen Jing An Mian Pian	Bu Yi Ji Li Wan
Xu Han Wei Tong Chong Ji	Ying Xin Dan	Anti-SARS1
Wen Pi Zhi Xie Wan	Xiao Er Fu Xie Ning	Anti-SARS2
Qing Dai San	Guan Xin Sheng Mai Kou Fu Ye	Dian Xian San
Yin Pu Jie Du Pian	Wu Zi Yan Zhong Wan	Xiao Er Hua Zhi San
Yin Zhi Huang Zhu She Ye	Sang Piao Xiao San	Yang Xue Yin
E Jiao Bu Xue Gao	Wei Ling Wan	
Incorrectly predicted as non-TCM recipe		
Shen Shi Tong Chong Ji	Jiang Tang Dan	Gan Shen Zi
Gan Fu Chong Ji		

These are used for testing SVM classification of TCM prescriptions. There are 44 prescriptions correctly predicted as TCM prescriptions and four incorrectly predicted as non-TCM recipes.

association with anti-inflammation (Sadowska-Krowicka *et al*., 1998). Also, genistein has been known to induce a shift towards antiproteolysis on in the balance between urokinase/plasminogen activator inhibitor (Fajardo *et al*., 1999), which seems to suggest that this protein is a target of genistein and the interaction may partially account for genistein's efficacy in cancer therapy. Genistein can marginally induce hypoxanthine-guanine phosphoribosyltransferase (Kulling and Metzler, 1997), which seems to suggest a possible mechanism of genistein's effect on the treatment of malaria (Dluzewski and Garcia, 1996).

Ginsenoside Rg1 is a major bioactive ingredient of ginseng. Ginseng is a highly valued herb in the Far East and has gained popularity in the West during the last decade (Attele *et al*., 1999). It can be used to combat stress, to enhance both the central nervous and immune systems and to help maintain optimal oxidative status against certain chronic disease states and aging (Kitts and Hu, 2000). It is also reported to prevent cancer (Shin *et al*., 2000) The pharmacokinetics of ginsenoside Rg1 is not as well studied as that of genistein. The available data show that ginsenoside Rg1 has a wide distribution and long half-life in the body (Huo *et al*., 1986; Takino, 1994). INVDOCK identified three potential targets of this ingredient. One is endothelial nitric oxide synthase, which is known to be inhibited by ginsenoside Rg1 (Li *et al*., 1997) and this inhibition may contribute to the observed maintenance of optimal oxidative status against chronic disease states and aging (Kitts and Hu, 2000). DNA polymerase β has not been reported to bind ginsenoside Rg1; however, it has been found that ginsenoside Rg1 can stimulate DNA synthesis (Lee and Lee, 1996) and activate DNA polymerase δ (Cho *et al*., 1995). Cyclophilin is also identified as a potential target by INVDOCK, whereas no experimental reports are available to confirm or refute it. This protein is related to immunomodulatory activity, which is one of the well-known therapeutic effects of ginsenosides including ginsenoside Rg1 (Kenarova *et al*., 1990). In addition to these therapeutic targets, INVDOCK also predicted an experimentally confirmed non-therapeutic target, 1,4-galactosyltransferase, an *in vivo* metabolizing enzyme of ginsenoside Rg1 (Danieli *et al*., 2001).

Baicalin is an active ingredient of *Scutellaria baicalensis* or *Oroxylum indicum*. It is reported to have anticancer (Ikemoto *et al*., 2000), anti-inflammatory (Lin and Shieh, 1996), anti-AIDS effects (De Clercq, 2000), and has been used in the treatment of diabetes (Zhou and Zhang, 1989) and liver problems (Nagai *et al*., 1989). Baicalin is converted to its metabolite baicalein by human intestinal flora (Zuo *et al*., 2002). Baicalein is well absorbed and metabolized back to baicalin in human intestinal epithelial Caco-2 cell monolayers. Baicalin is rapidly transported to both the apical side as well as the basolateral side of the small intestine (Zhang *et al*., 2005). Similar phenomenon has also been observed in rats (Wu *et al*., 1999). Several of the INVDOCK-identified targets are supported by experimental findings. Two of these targets were inhibited by baicalin. One is DNA polymerase β, an anti-viral target, which could be weakly inhibited by baicalin (Kitamura *et al*., 1998). The other is aldose reductase (Nagai *et al*., 1989), a target for the treatment of diabetes (Zhou and Zhang, 1989). It has been reported that baicalin has certain effects on two other therapeutic targets suggested by INVDOCK. Baicalin has been found to downregulate the expression level of cyclin-dependent kinase 2 (Liu *et al*., 1998), which is a known anticancer target. This compound has also been reported to have an inhibitory effect on phospholipase A2 (Kyo *et al*., 1998; Nakahata *et al*., 1998), which is a known anti-inflammatory target. The anticancer and anti-inflammatory effects of binding of baicalin to these targets have been observed experimentally (Lin and Shieh, 1996; Ikemoto *et al*., 2000).

The above results suggest the usefulness of INVDOCK, together with the TCM-ID data support, as an *in silico* tool in facilitating the identification of potential therapeutic targets of herbal ingredients and thus assisting the analysis of the molecular mechanisms of TCM prescriptions.

Validation of TCM multi-herb prescriptions

TCM practitioners formulate TCM multi-herb preparations according to the patient's condition and TDHPs(Chan, 1995). The Wu Hsing theory of the five material agents describes a patient's physical state. TDHPs include four characters: cold, hot, warm and cool; five tastes: pungent, sweet, sour, bitter and salty; four toxic states: non-toxic, slightly toxic, toxic and very toxic and 12 meridians: bladder, spleen, large intestine, stomach, small intestine, liver, cardiovascular system, heart, kidney, gallbladder and san jiao (translated as 'triple heater' – the trunk of the body). These properties correlate the physicochemical properties of the constituent herbs' principal ingredients with observed human responses. In modern terms, they represent the pharmacodynamic (the action or effects of drugs), pharmacokinetic (the process by which the body absorbs, distributes, metabolises and eliminates a drug) and toxicological properties of the herbs (Yuan and Lin, 2000).

TDHPs dictate herb combination for enhancement, assistance, restraint, suppression or antagonism. Combinations of the 'Master,' 'Adviser,' 'Soldier' and 'Guide' herbs form standard prescriptions (Chan, 1995). A Master herb is usually non-toxic and used for the principal diseases. An Adviser is non-toxic or slightly toxic and used for boosting the effects of Master herbs and for the treatment of accompanying symptoms. A Soldier is used for enhancing the therapeutic effects and modulating the adverse effects of the Master and Adviser herbs and for restoring the body to its pre-illness equilibrium. A Guide is used for guiding the active ingredients of other herbs to the specific meridian and for harmonizing the actions of other herbs. A Master in one prescription can be used as an Adviser or Soldier or Guide in another prescription. The same applies to an Adviser, Soldier or Guide. In the majority of TCM prescriptions, Masters are listed first, and Advisers are listed closer to the front and in many cases directly following Masters. For instance, in the recipe 'Wu Ji Bai Feng Wan' of Figure 1, the Masters Wu Ji and Lu Rong Jiao are listed in the first two positions, and the Advisers Ren shen, Huang qi, Dang gui and Bai shao are listed in the fifth to eighth positions.

Formulation of TCM prescriptions often relies on a TCM practitioner's experience, intuition and knowledge of TCM herbal properties and formulating principles. Further complicating the task is the personalized nature of TCM prescriptions. Inexperienced TCM practitioners and students find it particularly difficult to determine whether a proposed formulation constitutes a valid TCM prescription that strictly conforms to the TCM formulating principles. Thus, computer methods facilitating the validation and analysis of TCM prescriptions would be very useful.

AI systems were developed for determining whether or not a multi-herb preparation is a valid TCM prescription, in which the training and testing of classification models require examples of known TCM prescriptions and non-TCM recipes. We used all of the 1588 TCM prescriptions in the TCM-ID database for prescription examples and examples of non-TCM recipes were generated by random combination of multiple herbs and the modification of existing TCM prescriptions. In generating non-TCM recipes, 635 commonly used TCM herbs were divided into 13 traditionally defined therapeutic classes described in the TCM literature (Cheng, 2000). For each therapeutic class, two herbs with TCM-HPs closest to the average values of herbs in the corresponding class were selected as the representative herbs for that class. These representative herbs were then randomly combined and subsequently checked to remove hits of known TCM prescriptions. Moreover, existing TCM prescriptions with knowledge of their 'Master' herbs were modified by one of the following three methods. The first is the removal of the 'Master' herbs from the prescription and in some cases the removal of 'Adviser' herbs as well. The second is to replace the 'Master' herbs with those possessing the opposite TCM-HPs to completely disrupt the expected synergy between the original 'Master' herbs and the rest of the herbs in the prescription. The third is to add a specific herb to form a disallowed or unfavoured herb-pair in the prescription to convert it into an invalid one. These 'modified' recipes were subsequently checked to remove hits of known TCM prescriptions.

Two AI systems, the *k*NN (Johnson and Wichern, 2002) approach and the SVM approach were evaluated for their usefulness in validating TCM prescriptions. And two commonly applied validation strategies were used to make unbiased estimations of the classification accuracies of the AI systems, which are the threefold cross-validation method and the use of independent evaluation set, as described in the Methods section. Table 5 gives the computed accuracies of these two AI systems. Based on the average results from the threefold cross-validation method, 84.4 and 91.3% of the TCM prescriptions and 98.9 and 98.6% of the non-TCM recipes are correctly classified by *k*NN and SVM classification system, respectively. Testing results by using the independent evaluation set shows that 83.1 and 97.3% of the TCM prescriptions and 98.6 and 92.3% of the non-TCM recipes are correctly classified by *k*NN and SVM classification system, respectively. These two evaluation methods consistently show that the two AI methods are capable of using TCM-HPs for separating TCM prescriptions from non-TCM recipes.

The ability of the SVM classification system was further tested by using 48 TCM prescriptions published in recent years which are not yet included in the TCM-ID. These include 46 prescriptions from the Handbook of newly compiled TCM prescriptions (Pang, 2000), which are either modified forms of the traditional recipes or new formulae, and two new anti-SARS (SARS: severe acute respiratory syndrome) formulae published in MinBao newspaper of HongKong and LianHeZaoBao newspaper of Singapore on April 14, 2003, and in BeiJing Youth newspaper of China on April 10, 2003. One prescription, tentatively named anti-SARS1, is composed of Huang qi, Bai zhu, Fang feng, Cang zhu, Huo xiang, Sha shen and Jin yin hua. Another prescription, anti-SARS2, comprises Tai zi shen, Guan zong, Jin yin hua, Lian qiao, Da qing ye, Su ye, Ge gen, Huo xiang, Cang zhu and Pei lan. The list of these newer prescriptions is given in Table 6 and 91.7% of these prescriptions were correctly classified as TCM prescriptions. These results suggest the usefulness of TCM-ID data in developing prescription-recognizing AI systems, as well as the potential of using such systems to validate or optimize new TCM prescriptions.

Concluding remarks

TCM-ID database is intended as a convenient and integrated source for information about all aspects of TCM. The usefulness of the data in this database was illustrated by two case studies for probing the mechanism of therapeutic effects of herbal ingredients and for developing AI systems to validate TCM multi-herb preparations. Work is in progress for mining and analysing published information on additional herbs from various literature sources. With continued efforts and advances in TCM and herb research (Bhuiyan *et al*., 2003; Wang *et al*., 2003; Lazar, 2004), more extensive amount of information will be generated. The relevant new information can be incorporated into this or other related databases to provide more comprehensive description about the ingredients and contents of medicinal herbs. Advances in systems biology and computational biology are expected to generate more tools to mine this information for new theories in drug discovery, especially the so-far unknown mechanisms whereby TCM harmonizes the diverse effects of numerous active ingredients to exert the desired therapeutic results.

Acknowledgements

This work is partially supported by the Zhejiang University Seed Grant 308200-243014 and the Ministry of Education Returned Scholar Seed Grant [2006]331.

Conflict of interest

The authors state no conflict of interest.

References

Attele AS, Wu JA, Yuan CS (1999). Ginseng pharmacology: multiple constituents and multiple actions. *Biochem Pharmacol* **58**: 1685–1693.

Barnes S, Boersma B, Patel R, Kirk M, Darley-Usmar VM, Kim H *et al*. (2000). Isoflavonoids and chronic disease: mechanisms of action. *Biofactors* **12**: 209–215.

Barnes S, Peterson TG (1995). Biochemical targets of the isoflavone genistein in tumor cell lines. *Proc Soc Exp Biol Med* **208**: 103–108.

Bensoussan A, Myers SP, Drew AK, Whyte IM, Dawson AH (2002). Development of a Chinese herbal medicine toxicology database. *J Toxicol Clin Toxicol* **40**: 159–167.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H *et al*. (2000). The Protein Data Bank. *Nucleic Acids Res* **28**: 235–242.

Bhuiyan MB, Fant ME, Dasgupta A (2003). Study on mechanism of action of Chinese medicine Chan Su: dose-dependent biphasic production of nitric oxide in trophoblastic BeWo cells. *Clin Chim Acta* **330**: 179–184.

Byvatov E, Schneider G (2003). Support vector machine applications in bioinformatics. *Appl Bioinformatics* **2**: 67–77.

Castle EP, Thrasher JB (2002). The role of soy phytoestrogens in prostate cancer. *Urol Clin North Am* **29**: 71–81, viii–ix.

Chan K (1995). Progress in traditional Chinese medicine. *Trends Pharmacol Sci* **16**: 182–187.

Chen Q (1998). *Pharmacology and Clinical Studies of Well-known Traditional Chinese Medicinal Recipes*. People Health Pub. Co: Beijing.

Chen X, Ung CY, Chen Y (2003). Can an *in silico* drug-target search method be used to probe potential mechanisms of medicinal plant ingredients? *Nat Prod Rep* **20**: 432–444.

Chen Y, Ung C (2002). Computer automated prediction of putative therapeutic and toxicity protein targets of bioactive compounds from Chinese medicinal plants. *Am J Chin Med* **30**: 139–154.

Chen YZ, Ung CY (2001). Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand–protein inverse docking approach. *J Mol Graph Model* **20**: 199–218.

Chen YZ, Zhi DG (2001). Ligand–protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins* **43**: 217–226.

Cheng JT (2000). Review: drug therapy in Chinese traditional medicine. *J Clin Pharmacol* **40**: 445–450.

Cho SW, Cho EH, Choi SY (1995). Ginsenosides activate DNA polymerase delta from bovine placenta. *Life Sci* **57**: 1359–1365.

Danieli B, Falcone L, Monti D, Riva S, Gebhardt S, Schubert-Zsilavecz M (2001). Regioselective enzymatic glycosylation of natural polyhydroxylated compounds: galactosylation and glucosylation of protopanaxatriol ginsenosides. *J Org Chem* **66**: 262–269.

De Clercq E (2000). Current lead natural products for the chemotherapy of human immunodeficiency virus (HIV) infection. *Med Res Rev* **20**: 323–349.

Dluzewski AR, Garcia CR (1996). Inhibition of invasion and intraerythrocytic development of *Plasmodium falciparum* by kinase inhibitors. *Experientia* **52**: 621–623.

Evans BA, Griffiths K, Morton MS (1995). Inhibition of 5 alpha-reductase in genital skin fibroblasts and prostate tissue by dietary lignans and isoflavonoids. *J Endocrinol* **147**: 295–302.

Fajardo I, Quesada AR, Nunez de Castro I, Sanchez-Jimenez F, Medina MA (1999). A comparative study of the effects of genistein and 2-methoxyestradiol on the proteolytic balance and tumour cell proliferation. *Br J Cancer* **80**: 17–24.

Flamigni F, Facchini A, Capanni C, Stefanelli C, Tantini B, Caldarera CM (1999). p44/42 mitogen-activated protein kinase is involved in the expression of ornithine decarboxylase in leukaemia L1210 cells. *Biochem J* **341**: 363–369.

He M, Yan X, Zhou J, Xie G (2001). Traditional Chinese medicine database and application on the Web. *J Chem Inf Comput Sci* **41**: 273–277.

Huang TK (1999). *Handbook of Composition and Pharmacological Action of Commonly Used Traditional Chinese Medicine*. Chin. Med. Sci. Pub Co: Shanghai.

Huo YS, Zhang SC, Zhou D, Yao DL, You GY, Zhang HW *et al*. (1986). Pharmacokinetics and tissue distribution of [3H]ginsenoside Rg1. *Zhongguo Yao Li Xue Bao* **7**: 519–521.

Ikemoto S, Sugimura K, Yoshida N, Yasumoto R, Wada S, Yamamoto K *et al*. (2000). Antitumor effects of *Scutellariae radix* and its components baicalein, baicalin, and wogonin on bladder cancer cell lines. *Urology* **55**: 951–955.

Ji YB (1994). *Pharmacological Action and Application of Available Composition of Traditional Chinese Medicine*. HeiLongJiang Sci. Tech. Pub. Co: Harbin.

Ji ZL, Zhou H, Wang JF, Han LY, Zheng CJ, Chen YZ (2006). Traditional Chinese medicine information database. **103**: 501.

Johnson RA, Wichern DW (2002). *Applied Multivariate Statistical Analysis*. Prentice Hall: New Jersey.

Kato M, Liu W, Yi H, Asai N, Hayakawa A, Kozaki K *et al*. (1998). The herbal medicine Sho-saiko-to inhibits growth and metastasis of malignant melanoma primarily developed in ret-transgenic mice. *J Invest Dermatol* **111**: 640–644.

Kawa-uchi T, Nifuji A, Mataga N, Olson EN, Bonaventure J, Shinomiya K *et al* (1998). Fibroblast growth factor down-regulates expression of a basic helix-loop-helix-type transcription factor, scleraxis, in a chondrocyte-like cell line, TC6. *J Cell Biol* **70**: 468–477.

Kenarova B, Neychev H, Hadjiivanova C, Petkov VD (1990). Immunomodulating activity of ginsenoside Rg1 from Panax ginseng. *Japanese J Pharmacol* **54**: 447–454.

Kikuchi H, Hossain A (1999). Signal transduction-mediated CYP1A1 induction by omeprazole in human HepG2 cells. *Exp Toxicol Pathol* **51**: 342–346.

Kim MH, Albertsson P, Xue Y, Nannmark U, Kitson RP, Goldfarb RH (2001). Expression of neutrophil collagenase (MMP-8) in Jurkat T leukemia cells and its role in invasion. *Anticancer Res* **21**: 45–50.

Kitamura K, Honda M, Yoshizaki H, Yamamoto S, Nakane H, Fukushima M *et al*. (1998). Baicalin, an inhibitor of hiv-1 production *in vitro*. *Antiviral Res* **37**: 131–140.

Kitts D, Hu C (2000). Efficacy and safety of ginseng. *Public Health Nutr* **3**: 473–485.

Kniss DA, Zimmerman PD, Su HC, Fertel RH (1996). Genistein suppresses EGF-induced prostaglandin biosynthesis by a mechanism independent of EGF receptor tyrosine kinase inhibition. *Prostaglandins* **51**: 87–105.

Kulling SE, Metzler M (1997). Induction of micronuclei, DNA strand breaks and HPRT mutations in cultured Chinese hamster V79 cells by the phytoestrogen coumoestrol. *Food Chem Toxicol* **35**: 605–613.

Kuzumaki T, Kobayashi T, Ishikawa K (1998). Genistein induces p21(Cip1/WAF1) expression and blocks the G1 to S phase transition in mouse fibroblast and melanoma cells. *Biochem Biophys Res Commun* **251**: 291–295.

Kyo R, Nakahata N, Sakakibara I, Kubo M, Ohizumi Y (1998). in and baicalein, constituents of an important medicinal plant, inhibit intracellular Ca^{2+} elevation by reducing phospholipase C activity in C6 rat glioma cells. *J Pharmacy Pharmacol* **50**: 1179–1182.

Lazar MA (2004). East meets West: an herbal tea finds a receptor. *J Clin Invest* **113**: 23–25.

Lee KY, Lee SK (1996). Ginsenoside-Rg1 positively regulates cyclin E-dependent kinase activity in human hepatoma SK-HEP-1 cells. *Biochem Mol Biol Int* **39**: 539–546.

Li JQ, Li ZK, Duan H, Zhang JT (1997). Effect of age and ginsenoside Rg1 on nitric oxide content and nitric oxide synthase activity of cerebral cortex in rats. *Yao xue xue bao* **32**: 251–254.

Lin CC, Shieh DE (1996). The anti-inflammatory activity of *Scutellaria rivularis* extracts and its active components, baicalin, baicalein and wogonin. *Am J Chin Med* **24**: 31–36.

Liu W, Kato M, Akhand AA, Hayakawa A, Takemura M, Yoshida S *et al*. (1998). The herbal medicine sho-saiko-to inhibits the growth of malignant melanoma cells by upregulating Fas-mediated apoptosis and arresting cell cycle through downregulation of cyclin dependent kinases. *Int J Oncol* **12**: 1321–1326.

Majumder SK, Ghosh N, Gupta PK (2005). Support vector machine for optical diagnosis of cancer. *J Biomed Opt* **10**: 024034.

Martel-Pelletier J, Mineau F, Jovanovic D, Di Battista JA, Pelletier JP (1999). Mitogen-activated protein kinase and nuclear factor kappaB together regulate interleukin-17-induced nitric oxide production in human osteoarthritic chondrocytes: possible role of transactivating factor mitogen-activated protein kinase-activated protein kinase (MAPKAPK). *Arthritis Rheumat* **42**: 2399–2409.

Morito K, Hirose T, Kinjo J, Hirakawa T, Okawa M, Nohara T *et al*. (2001). Interaction of phytoestrogens with estrogen receptors alpha and beta. *Biol Pharmaceut Bull* **24**: 351–356.

Munoz R, Klingenberg O, Wiedlocha A, Rapak A, Falnes PO, Olsnes S (1997). Effect of mutation of cytoplasmic receptor domain and of genistein on transport of acidic fibroblast growth factor into cells. *Oncogene* **15**: 525–536.

Nagai T, Yamada H, Otsuka Y (1989). Inhibition of mouse liver sialidase by the root of *Scutellaria baicalensis*. *Planta Med* **55**: 27–29.

Nakahata N, Kutsuwa M, Kyo R, Kubo M, Hayashi K, Ohizumi Y (1998). Analysis of inhibitory effects of *scutellariae radix* and baicalein on prostaglandin E2 production in rat C6 glioma cells. *Am J Chin Med* **26**: 311–323.

Nishimura J, Takahira H, Shibata K, Muta K, Yamamoto M, Ideguchi H *et al*. (1988). Regulation of biosynthesis and phosphorylation of P210bcr/abl protein during differentiation induction of K 562 cells. *Leuk Res* **12**: 875–885.

Orie NN, Zidek W, Tepel M (2000). creased intracellular generation of reactive oxygen species in mononuclear leukocytes from patients with diabetes mellitus type 2. *Exp Clin Endocrinol Diabetes* **108**: 175–180.

Pach D, Willich SN, Becker-Witt C (2002). Availability of research results on traditional Chinese pharmacotherapy. *Forsch Komplementarmed Klass Naturheilkd* **9**: 352–358.

Pan W, Ikeda K, Takebe M, Yamori Y (2001). Genistein, daidzein and glycitein inhibit growth and DNA synthesis of aortic smooth muscle cells from stroke-prone spontaneously hypertensive rats. *J Nutr* **131**: 1154–1158.

Pang SD (2000). Handbook of Newly Compiled TCM Prescriptions. Jiangxi Sci. Tech. Pub. Co: Nanchang.

Pecherskaya A, Solem M (2000). IGF1 activates PKC alpha-dependent protein synthesis in adult rat cardiomyocytes. *Molecular Cell Biol Res Commun* **4**: 166–171.

Polkowski K, Mazurek AP (2000). Biological properties of genistein. A review of *in vitro* and *in vivo* data. *Acta Pol Pharm* **57**: 135–155.

Qiao X, Hou T, Zhang W, Guo S, Xu X (2002). A 3D structure database of components from Chinese traditional medicinal herbs. *J Chem Inf Comput Sci* **42**: 481–489.

Reyes-Cruz G, Vazquez-Prado J, Muller-Esterl W, Vaca L (2000). Regulation of the human bradykinin B2 receptor expressed in sf21 insect cells: a possible role for tyrosine kinases. *J Cell Biochem* **76**: 658–673.

Sadowska-Krowicka H, Mannick EE, Oliver PD, Sandoval M, Zhang XJ, Eloby-Childess S *et al*. (1998). Genistein and gut inflammation: role of nitric oxide. *Proc Soc Exp Biol Med* **217**: 351–357.

Schmidt TJ, Meyer AS (1994). Autoregulation of corticosteroid receptors How, when, where, and why? *Receptor* **4**: 229–257.

Shin HR, Kim JY, Yun TK, Morgan G, Vainio H (2000). The cancer-preventive potential of Panax ginseng: a review of human and experimental evidence. *Cancer Causes Control* **11**: 565–576.

Su WW (1997). Computer-aided analysis of Traditional Chinese Medicine. *China J Chin Materia Medica* **22**: 186–188.

Swa S, Wright H, Thomson J, Reid H, Haig D (2001). Constitutive activation of Lck and Fyn tyrosine kinases in large granular lymphocytes infected with the gamma-herpesvirus agents of malignant catarrhal fever. *Immunology* **102**: 44–52.

Takahashi T, Kawahara Y, Okuda M, Ueno H, Takeshita A, Yokoyama M (1997). Angiotensin II stimulates mitogen-activated protein kinases and protein synthesis by a Ras-independent pathway in vascular smooth muscle cells. *J Biol Chem* **272**: 16018–16022.

Takino Y (1994). Studies on the pharmacodynamics of ginsenoside-Rg1, -Rb1 and -Rb2 in rats. *Yakugaku Zasshi* **114**: 550–564.

Tang W, Eisenbrand G (1992). *Chinese Drugs of Plant Origin: Chemistry, Pharmacology and Use in Traditional and Modern Medicine*. Springer-Verlag: Berlin.

Theodorescu D, Laderoute KR, Calaoagan JM, Guilding KM (1998). Inhibition of human bladder cancer cell motility by genistein is dependent on epidermal growth factor receptor but not p21ras gene expression. *Int J Cancer* **78**: 775–782.

Tomobe K, Philbrick DJ, Ogborn MR, Takahashi H, Holub BJ (1998). Effect of dietary soy protein and genistein on disease progression in mice with polycystic kidney disease. *Am J Kid Dis* **31**: 55–61.

Uckun FM, Evans WE, Forsyth CJ, Waddick KG, Ahlgren LT, Chelstrom LM *et al*. (1995). Biotherapy of B-cell precursor leukemia by targeting genistein to CD19-associated tyrosine kinases. *Science* **267**: 886–891.

Wang JF, Cai CZ, Kong CY, Cao ZW, Chen YZ (2005a). A computer method for validating traditional Chinese medicine herbal prescriptions. *Am J Chin Med* **33**: 281–297.

Wang JF, Zhou H, Han LY, Chen X, Chen YZ, Cao ZW (2005b). Traditional Chinese medicine information database. *Clin Pharmacol Therapeut* **78**: 92–93.

Wang L, Higashiura K, Ura N, Miura T, Shimamoto K (2003). Chinese medicine, Jiang-Tang-Ke-Li, improves insulin resistance by modulating muscle fiber composition and muscle tumor necrosis factor-alpha in fructose-fed rats. *Hypertens Res* **26**: 527–532.

Wheeler DL, Church D.M, Federhen S, Lash A E, Madden T L, Pontius J U *et al*. (2003). Database resources of the National Center for Biotechnology. *Nucleic Acids Res* **31**: 28–33.

Wu J, Chen D, Zhang R (1999). Study on the bioavailability of baicalin-phospholipid complex by using HPLC. *Biomed Chromatogr* **13**: 493–495.

Xu H, Markatou M, Dimova R, Liu H, Friedman C (2006). Machine learning and word sense disambiguation in the biomedical domain: design and evaluation issues. *BMC Bioinform* **7**: 334.

Yousufzai SY, Abdel-Latif AA (1998). Tyrosine kinase inhibitors suppress prostaglandin F2alpha-induced phosphoinositide hydrolysis, Ca^{2+} elevation and contraction in iris sphincter smooth muscle. *Eur J Pharmacol* **360**: 185–193.

Yu J, Chen XW (2005). Bayesian neural network approaches to ovarian cancer identification from high-resolution mass spectrometry data. *Bioinformatics* **21** (Suppl 1): i487–i494.

Yuan R, Lin Y (2000). Traditional Chinese medicine: an approach to scientific proof and clinical validation. *Pharmacol Ther* **86**: 191–198.

Zhang E (1988). *Highly Efficacious Chinese Patent Medicines*. Shanghai Univ. Tradi. Chin. Med. Pub. Co: Shanghai.

Zhang E (1998). *Prescriptions of Traditional Chinese Medicine*.. Shanghai Univ. Tradi. Chin. Med. Pub. Co: Shanghai.

Zhang L, Lin G, Chang Q, Zuo Z (2005). Role of intestinal first-pass metabolism of baicalein in its absorption process. *Pharmaceut Res* **22**: 1050–1058.

Zhou YP, Zhang JQ (1989). Oral baicalin and liquid extract of licorice reduce sorbitol levels in red blood cell of diabetics rats. *Chinese Med J* **102**: 203–206.

Zuo F, Zhou ZM, Yan MZ, Liu ML, Xiong YL, Zhang Q *et al* (2002). Metabolism of constituents in Huangqin-Tang, a prescription in traditional Chinese medicine, by human intestinal flora. *Biol Pharmaceut Bull* **25**: 558–563.