Traditional Chinese Medicine Database and Application on the Web†

Min He,‡ Xinjian Yan,* Jiaju Zhou, and Guirong Xie

Laboratory of Computer Chemistry, Institute of Chemical Metallurgy, Chinese Academy of Sciences, P.O. Box 353, Beijing 100080, China

Received November 8, 2000

To study traditional Chinese medicines and exchange related information through the worldwide Web, we developed a traditional Chinese medicine database and a program for searching and displaying data in the database on the Web. In this paper, the traditional Chinese medicine database is briefly introduced; the methods used in developing the program, including ISAPI (Microsoft Internet Server Application Programming Interface), VRML (Virtual Reality Model Language), and JavaScript are described; and three application examples are also given.

1. INTRODUCTION

Traditional Chinese medicine (TCM)[1−5] is derived from thousands of years of observation and empirical evidence and is one of the brightest pearls in the treasure chest of Chinese cultural inheritance, playing, even to this day, a critical role in the well-being of the Chinese people. TCM views a patient as an interacting and mutually influencing system of functional parts and regards disease as the result of abnormal interactions or imbalances in the system. As a consequence, diagnosis and treatment in TCM may be different from those of Western medicine. The excellent effects of many traditional Chinese medicines on curing diseases have been approved by experiences. However, the weaknesses of TCM still exist. Specially, there is a lack of an understanding on progression and treatment of the disease at the molecular level, which impedes worldwide acceptance and development of TCM.

With the rapid development of computer technologies and molecular modeling methods, computer-aided study has become an essential tool for the study and design of new drugs.[6] In view of the huge amount of information in TCM and the currently available computer technologies in database, molecular modeling, and the Web, it is logical to develop a TCM database containing information from both traditional knowledge and modern science and to display it via the Web. Such a system will be valuable in the following respects: (1) an understanding of Chinese medicine theory, clinical experience, and its relation to Western medicine; (2) study of pharmacology, toxicology, clinical indications, and side effects of chemical components; (3) an explanation of the principle of prescriptions in TCM and examination, by means of QSAR (Quantitative Structure−Activity Relationships) of its effective components; and (4) investigation of the distribution of chemical species in plants and the elucidation of curative effects by means of statistics, fuzzy mathematics, and artificial intelligence.

These are some of the research areas that will be greatly facilitated by studies on TCM at the molecular level. For these reasons, we undertook the establishing of a Traditional Chinese Medicine Database (TCMD) and developing a program for search and display of the TCM data. The data in this database are mainly from the book, Traditional Chinese Medicines: Molecular Structures, Natural Sources and Applications,[7] which contains information gathered independently from Western studies of anatomy and physiology and, in fact, largely predated it. This paper is a brief introduction to the Traditional Chinese Medicine Database.

2. SYSTEM DESIGN

2.1. Creating TCMD. The TCM information has been organized in three parts: Chinese medicines, original plants, and bioactive components. The data of Chinese medicines include all kinds of names, codes of TCM, effective species, property, flavor and channel tropism, effects, indications, clinical applications, pharmaceutical functions, annotations, and references. The data related to the original plants contain all kinds of names, parts used as medicines, collection, and preprocess of the plants. The bioactive components are described in terms of the chemical formulas, two-dimensional (2D) and three-dimensional (3D) chemical structures, and physicochemical properties. Based upon the characteristics of TCM information 54 data fields and related codes have been defined.[8] A TCM database, which includes basic data, chemical components, molecular structures, and bioactive data, has been established. There are more than 1540 species of natural plants (or other natural products) used as TCM in the database and 6800 compounds, which were isolated from plants, herbs, animals, fungi, etc., in the molecular structure part of the database.

2.2. The Search and Display Software. An interactive search and display system on the Internet was implemented by using Internet Server Application Programming Interface (ISAPI)[9,10] which has more advantages than the Common

* Corresponding author phone: (303)497-4164; e-mail: xjyan@boulder.nist.gov. Present address: Physical and Chemical Properties Division, Building 2, Room 1022, National Institute of Standards and Technology, 325 Broadway, Boulder, CO 80305-3328.

† The Web address of the database is URL: http://166.111.57.233/. However, at the present time, it works only during the daytime of China.

‡ Present address: Institute of Simulation and Control for Thermal Power, Department of Thermal Engineering, Tsinghua University, Beijing, 100084 P.R. China. Phone: 10-62784737/9-22(O); fax: 10-62782749; e-mail: mhe@te.tsinghua.edu.cn.

10.1021/ci0003101 CCC: $20.00
Published on Web 03/03/2001

Gateway Interface (CGI)[11] in page generating speed, reduced server load, finer-grained control, and Uniform Resource Locator (URL) mapping. ISAPI enables Web developers to tailor their Web sites through the use of ISAPI extensions and ISAPI filters. Extensions offer capabilities such as those of CGI applications. Filters allow preprocessing and post-processing of HTTP requests for services such as authentication, custom logging, and security. ISAPI is the first step toward a standard Internet interface to the Win32 development platform. By using ISAPI, it is possible to support simple-entry and multiple-entry retrievals at client site. The user-friendly interfaces on the Internet were coded using Hypertext Markup Language (HTML) and JavaScript. JavaScript is the Internet browser's cross-platform, object-based event-driven scripting language. It lets developers create applications that run over the Internet. Using JavaScript, developers can create objects and dynamic HTML pages that process user input and maintain persistent data using special objects, files, and relational databases. We have developed a new browser-window for display of 3D chemical structures in searching TCM effective component data, while the 2D molecular structure and other chemical data are being displayed on the first browser-window.[12,13]

2.3. To Transfer and Display of 3D Chemical Structure. In the field of medicinal chemistry, the display of 3D chemical structures is very important. Virtual Reality Modeling Language (VRML)[14,15] is a plain text (standard ASCII text) format file for describing 3D interactive worlds and objects on the Internet. Unlike graphics files, such as GIF[16] or JPEG,[17] VRML file can be transferred in a compressed format and uncompressed automatically by a viewer. This reduces the time and cost of transmission over the Internet. The images formed by VRML are "live", in that they can be magnified and rotated. VRML, like the 3D analogue of HTML, can be used in conjunction with the Web and be used to create 3D representations of complex scenes such as illustrations, product definition, and virtual reality presentations. To transfer and display 3D chemical structures on the Internet smoothly, we developed a program, VRMLMaker, that can convert a molecular MOL2[18] or ML2[19] file to the VRML file for displaying molecules in four different styles: wireframe, capped sticks, ball-and-stick, and CPK space-filling models. VRMLMaker was developed based on molecular model and VRML techniques.[20−22] One can conveniently browse 3D chemical structure in VRML mode with an Internet browser.

3. APPLICATIONS

The rapid expansion of the Internet has provided opportunities for many different applications. We have attempted to provide an easy access to the TCM information for pharmaceutical researchers. The interface of the TCMD on Web can be accessed through an Internet browser that connects to the TCM server. Figure 1 shows a main interface of the database. The applications can be divided into three components: TCM data queries, chemical data queries, and multiple-entry queries.

3.1. First, with the TCM data queries, users can request the TCM data that include TCM identification number, the English or Latin name of the plant, the Latin name of the plant's family, the Chinese name of the TCM, the parts of the plant used in TCM, the indications of TCM, and the effects of TCM. Users can request TCM data by one of the eight query modes, as shown in Figure 1.

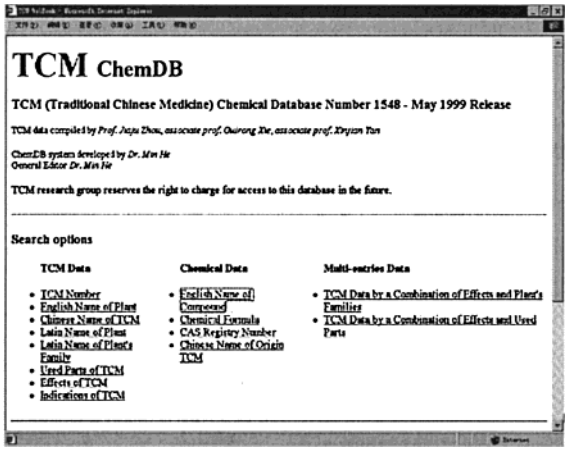


Figure 1. A main interface of TCMD.

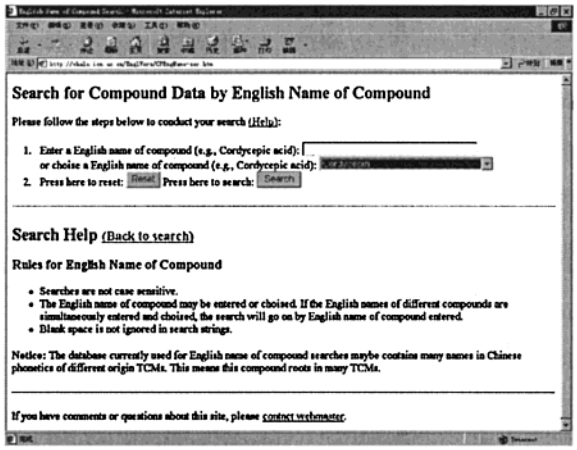


Figure 2. A query interface of TCMD.

Second, in TCM chemical data queries, users can request chemical data, including English names and synonyms, chemical formula, molecular weight, the Chinese name of the original TCM, 2D structures and substructures, 3D structures, and other physicochemical properties. In this query, there are four query modes, namely, the English name of compound, the chemical formula, the CAS Registry Number, and the Chinese name of the original TCM. In any one of these query modes, there are the following three search rules: (1) The query string is case-insensitive. (2) Users can enter or select a query. If a user enters a query and chooses another query simultaneously, the program will search the entered query. (3) Blank spaces in queries are not ignored in searching.

Figure 2 shows an example of a search for chemical data using a compound's English name. The queried English name is "Cordycepin". The searched results are presented in Figures 3 and 4. The interactive application has a handle to the VRML browser object on the hypertext markup language page and communicates with the TCMD. The 3D chemical structure is displayed in VRML modes that have four different styles, including ball-and-stick, a CPK space-filling, wireframe, and capped sticks. When the user clicks one of the four styles of VRML modes, another window will be

Table 1. Statistical Profile of the Numbers of Medicinal Plants' Species and Their Families from the TCM Database

family name	no.	family name	no.	family name	no.	family name	no.
Leguminosae	109	*Araceae*	9	*Acanthaceae*	4	*Polygalaceae*	3
Compositae	94	*Orchidaceae*	9	*Adiantaceae*	4	*Simaroubaceae*	3
Rutaceae	58	*Amaranthaceae*	8	*Agavaceae*	4	*Stemonaceae*	3
Liliaceae	48	*Convolvulaceae*	8	*Balsaminaceae*	4	*Typhaceae*	3
Umbelliferae	46	*Meliaceae*	8	*Betulaceae*	4	*Ulmaceae*	3
Labiatae	44	*Oleaceae*	8	*Blechnaceae*	4	*Usneaceae*	3
Solanaceae	44	*Thymelaeaceae*	8	*Bombycidae*	4	*Agaricaceae*	2
Euphorbiaceae	35	*Amaryllidaceae*	7	*Cephalotaxaceae*	4	*Alariaceae*	2
Ranunculaceae	33	*Caryophyllaceae*	7	*Cladoniaceae*	4	*Alismataceae*	2
Berberidaceae	27	*Cupressaceae*	7	*Combretaceae*	4	*Buxaceae*	2
Cucurbitaceae	26	*Lardizabalaceae*	7	*Linaceae*	4	*Capparidaceae*	2
Polygonaceae	25	*Nymphaeaceae*	7	*Podocarpaceae*	4	*Caricaceae*	2
Rubiaceae	23	*Rhamnaceae*	7	*Pyrolaceae*	4	*Clavicipitaceae*	2
Cruciferae	22	*Sapindaceae*	7	*Salicaceae*	4	*Commelinaceae*	2
Araliaceae	20	*Bignoniaceae*	6	*Tricholomataceae*	4	*Cycadaceae*	2
Verbenaceae	20	*Boraginaceae*	6	*Actinidiaceae*	3	*Cyperaceae*	2
Zingiberaceae	18	*Dioscoreaceae*	6	*Bufonidae*	3	*Daphniphyllaceae*	2
Menispermaceae	17	*Ebenaceae*	6	*Burseraceae*	3	*Fagaceae*	2
Lauraceae	15	*Iridaceae*	6	*Caprifoliaceae*	3	*Huperziaceae*	2
Papaveraceae	15	*Orobanchaceae*	6	*Chenopodiaceae*	3	*Lemnaceae*	2
Anacardiaceae	14	*Polyporaceae*	6	*Cornaceae*	3	*Myricaceae*	2
Gentianaceae	14	*Theaceae*	6	*Dryopteridaceae*	3	*Nyctaginaceae*	2
Apocynaceae	13	*Valerianaceae*	6	*Elaeagnaceae*	3	*Onagraceae*	2
Malvaceae	13	*Aquifoliaceae*	5	*Ephedraceae*	3	*Pinaceae*	2
Ericaceae	12	*Crassulaceae*	5	*Equisetaceae*	3	*Plantaginaceae*	2
Moraceae	12	*Guttiferae*	5	*Hamamelidaceae*	3	*Potamogetonaceae*	2
Scrophulariaceae	12	*Loranthaceae*	5	*Hippocastanaceae*	3	*Primulaceae*	2
Aristolochiaceae	11	*Lythraceae*	5	*Juglandaceae*	3	*Punicaceae*	2
Celastraceae	11	*Palmae*	5	*Loganiaceae*	3	*Saururaceae*	2
Myrtaceae	11	*Piperaceae*	5	*Lycoperdaceae*	3	*Selaginellaceae*	2
Saxifragaceae	11	*Sterculiaceae*	5	*Lycopodiaceae*	3	*Styracaceae*	2
Magnoliaceae	10	*Vitaceae*	5	*Meloidae*	3	*Tiliaceae*	2
Polypodiaceae	10	*Zygophyllaceae*	5	*Pedaliaceae*	3	*Ulvaceae*	2

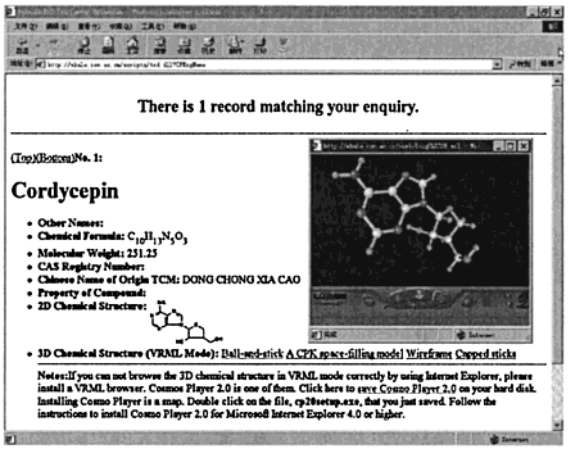


Figure 3. A snapshot of the query result.

Table 2. Statistical Profile of the Numbers of Plant Species and Curative Effects in Some Families

family name	A	B	C	family name	A	B	C
Menispermaceae	17	7	41.18	*Labiatae*	44	6	13.64
Polygonaceae	25	9	36.00	*Polypodiaceae*	10	1	10.00
Scrophulariaceae	12	4	33.33	*Cruciferae*	22	2	9.09
Gentianaceae	14	4	28.57	*Celastraceae*	11	1	9.09
Saxifragaceae	11	3	27.27	*Umbelliferae*	46	4	8.69
Rubiaceae	23	6	26.09	*Ericaceae*	12	1	8.33
Berberidaceae	27	7	25.93	*Moraceae*	12	1	8.33
Compositae	94	24	25.53	*Lauraceae*	15	1	6.67
Solanaceae	44	11	25.00	*Papaveraceae*	15	1	6.67
Malvaceae	13	3	23.08	*Araliaceae*	20	1	5.00
Cucurbitaceae	26	5	19.23	*Verbenaceae*	20	1	5.00
Aristolochiaceae	11	2	18.18	*Rutaceae*	58	2	3.45
Leguminosae	109	19	17.43	*Liliaceae*	48	1	2.08
Apocynaceae	13	2	15.38	*Zingiberaceae*	18	0	0.00
Ranunculaceae	33	5	15.15	*Myrtaceae*	11	0	0.00
Euphorbiaceae	35	5	14.29	*Magnoliaceae*	10	0	0.00

created for displaying the molecule in 3D VRML model. The 2D interface and the 3D molecular view are displayed in the same time. The VRML interface also supports an egocentric view, with facilities for toggling between views using the navigational buttons in an appropriate window. Figure 3 is the search result for "cordycepin" and its 3D chemical structure in ball-and-stick VRML mode interfaces. Figure 4 shows the compound in a CPK space-filling VRML mode. The 3D chemical structure in wireframe or capped sticks VRML mode can also be displayed in the new browser-window.

Third, this program supports diversified data search by multiple-entry queries. A few case studies have been carried out, and some examples are given here to demonstrate the capabilities of this system.[23]

3.2. Table 1 is the distribution profile of 1500 plants in 210 plant families in the present TCM database. Families with only one plant in the database are not listed in the table.

Table 2 lists the statistical profile of the number of plant species in a family and the curative effects in some families. Here, A is the total number of plants for each family in TCMD; B is the number of plants which are known to reduce fever; and C is B divided by A, expressed as a percentage.

3.3. The following is another application example of the TCM database. Flavonoids, which are polyphenolic com-

Table 3. Ten Plants Which Contain Most Flavonoids Based on the TCM Database and Their TCM Names, English Names, the Numbers of Flavonoids Found in Each Plant, and Functions in Antihypertension, Antibacteria, and Antivirus

TCM name	English name	no. of compds	antihypertension	antibacteria	antivirus
Huang Qin	Baikal Skullcap	30	yes	yes	yes
Qing Hao	Celery Wormwood	28	yes	yes	yes
Yin Yang Huo	Largeflower Epimedium	23	yes	yes	yes
Yin Chen Hao	Capillary Wormwood	22	yes	yes	yes
Huo Xiang	Wrinkled Gianthyssop	12		yes	yes
Mu Zei	Common Scouring Rush	10	yes	yes	yes
Pu Huang	Narrowleaf Cattail Pollen	10	yes	yes	
Bai Guo	Ginkgo Nut	9	yes	yes	
Chuan Xin Lian	Common Andrographis	9	yes	yes	
Ku Shen	Lightyellow Sophora	9	yes	yes	yes

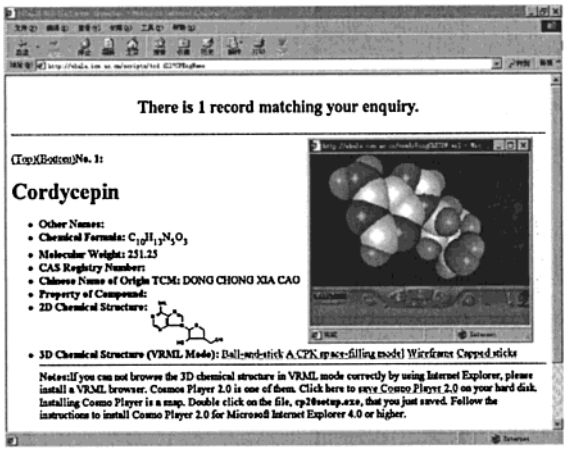
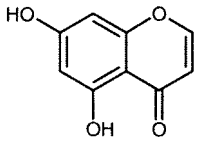


Figure 4. A snapshot of the query result.

pounds that occur ubiquitously in plants, possess many valuable functions, such as anticancer, antihypertension, antibacteria, antivirus, antiinflammation, etc. All flavonoids contain the substructure (**S1**) with replacements



S1

on its carbon atoms and hydroxy groups. Because the amount of flavonoids in a plant may play an important role for the plant's medicinal functions, we performed 2D searches on the TCM database with **S1** as the query structure. The 2D search result shows that in the database there are 485 compounds, which contain **S1** structure, and 335 plants contain flavonoids. Among these plants, there are 10 plants, which contain nine or more flavonoids. Table 3 lists the 10 plants' TCM names, English names, the numbers of flavonoids found in each plant, and their functions[1,24] in antihypertension, antibacteria, and antivirus. Table 3 shows that the 10 plants have similar activities in antihypertension, antibacteria, and antivirus, specially, for the first four plants. The data may imply that the plants that contain similar compounds have similar activities. Here, we must point out that this work is just an application example of the TCM database. It is not a complete study on the relationship between plants' function and their components, because the plants' functions displayed in Table 3 may be due to other compounds instead of flavonoids. More deep studies are needed.

4. CONCLUSION

We developed a database of Traditional Chinese Medicines and put it on the Web by using ISAPI, VRML, JavaScript, and HTML technologies. We also developed a method for 3D chemical structure transfer and display over the Internet by using VRML and JavaScript. Three application examples of the database are provided. They show the value of the database for studying and understanding traditional Chinese medicines.

ACKNOWLEDGMENT

This work was supported by 973 Project of National Key Basic Study of China (No. G1998051115). We are indebted to George W. A. Milne for providing consultation and revising this paper.

REFERENCES AND NOTES

(1) Chinese Materia Medica Editing Committee of the National Chinese Medicine and Pharmacology Bureau. Chinese Materia Medica, Shanghai Science and Technology Press: Shanghai, 1998 (in Chinese).
(2) Shi, L. *The History of Traditional Chinese Medicine*; Science Press: 1996 (in Chinese).
(3) Tierra, M. *The way of Chinese Herbs*; Pocket Book, New York, 1998.
(4) Reid, D. A *Handbook of Chinese Healing Herbs*; Barnes & Noble Books: New York, 1995.
(5) Hesketh, T.; Zhu, W. Health in China: Traditional Chinese Medicine: one country, two systems. *BMJ* July 12, **1997**, 115−117.
(6) Yan, X.; Zhou. J.; Zhi, H. Concept Design of Computer-Aided Study on Traditional Chinese Drug, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 86−89.
(7) Yan, X.; Zhou. J.; Xie. G. *Traditional Chinese Medicines: Molecular Structures, Natural Sources and Applications*; Ashgate: 1999. **Note:** The data from the book is being processed by the authors, Daylight company and Ashgate publisher, and will be available on the Internet.
(8) He, M.; Shen, M.; Zhou, J. Design and Development of Traditional Chinese Medicine Database (in Chinese). *Computers Appl. Chem.* **1999,** *16*, 363−364.
(9) Special Edition Using ISAPI (http://www.mcp.com/que/).
(10) He, M.; Zhou, J. Implementation of Database Searching on the Internet Using ISAPI (in Chinese). *Software World* **1998**, *3*, 72 (http://www.softwareworld.com.cn/).
(11) McCool, R. *Introduction to the Common Gateway Interface*, NCSA: 1994.
(12) He, M.; Zhou, J. The Application of JavaScript to Chemical Database Searching System on the Internet (in Chinese). *HuaXue TongBao* **1999**, *4*, 60−63.
(13) He, M.; Zhou, J. Web-based Chinese Medicine Information Searching System (in Chinese). *Computers Appl. Chem.* **1999**, *16*, 280−282.
(14) VRML 1.0C Specification (http://www.vrml.org/VRML1.0/vrml10c.html).
(15) VRML 2.0 Specification (http://ws05.pc.chemie.th-darmstadt.de/vrml/spec/spec/index.html).
(16) CompuServ, Inc., Graphics Interchange Format, 1992 (http://www.w3.org/hypertext/WWW/Graphics/GIF/spec-gif89a.txt).
(17) Hamilton, E. JPEG File Interchange Format, 1992 (http://www.w3.org/hypertext/WWW/Graphics/JPEG/jfif.txt).
(18) SYBYL Toolkit Manual; Tripos Associates: St. Louis, MO, 1992.

(19) Alchemy 2000; Tripos Inc.: St. Louis, MO, 1996.
(20) He, M.; Chen, H.; Zhou, J. VRML and 3D Chemical Structure Display (in Chinese). *HuaXue TongBao* **1998**, *9*, 55−58.
(21) He, M.; Xie, G.; Zhou, J. VRML for Use in 3D Molecular Models on the WWW (in Chinese). *Computers Appl. Chem.* **2000**, *17*(2), 133−134.
(22) He, M.; Zhou, J. Display of 3D Chemical Structure on the Internet (in Chinese). *J. App. Sci.* **2000**, *18*(2), 95−100.
(23) He, M. Study on the Traditional Chinese Medicine Information System on the Internet. Dissertation of Doctoral Degree, Institute of Chemical Metallurgy, Chinese Academy of Sciences, Beijing, 1999.
(24) *Handbook of Modern Pharmacology of Chinese Medicines*; Mei, Q., Bi, H., Chief Eds.; China Press of Chinese Medicines: Beijing, 1998 (in Chinese).

CI0003101