Inter*net* Electronic Journal of Molecular Design

October 2004, Volume 3, Number 10, Pages 672–683

Editor: Ovidiu Ivanciuc

Proceedings of the Internet Electronic Conference of Molecular Design, IECMD 2003
November 23 – December 6, 2003
Part 5

A Traditional Chinese Medicine Plant–Compound Database and Its Application for Searching

Aijun Lu, Bing Liu, Haibo Liu, Jiaju Zhou, and Guirong Xie

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

Received: November 1, 2003; Revised: March 30, 2004; Accepted: May 30, 2004; Published: October 31, 2004

Citation of the article:
A. Lu, B. Liu, H. Liu, J. Zhou, and G. Xie, A Traditional Chinese Medicine Plant–Compound Database and Its Application for Searching, *Internet Electron. J. Mol. Des.* **2004**, *3*, 672–683, http://www.biochempress.com.

Inter*net* Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

A Traditional Chinese Medicine Plant–Compound Database and Its Application for Searching[#]

Aijun Lu, Bing Liu, Haibo Liu, Jiaju Zhou,* and Guirong Xie

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

Received: November 1, 2003; Revised: March 30, 2004; Accepted: May 30, 2004; Published: October 31, 2004

Internet Electron. J. Mol. Des. **2004**, *3* (10), 672–683

Abstract

This article describes a Traditional Chinese Medicine Plant–Compound Database and an application case of database searching for HIV protease inhibitor. It offers not only basic compound properties such as English name and synonyms, physical properties, bioactivity data, formula, molecular weight, CAS Registry Number and 2D, 3D structure of the molecule but also detailed information on their natural source, including Latin name, English name, pinyin name, used part, indication, family, and curative effect. In order to present a basic profile of the database to the user, some statistic data are provided. The result shows that lots of hot bioactive records are in the database and also that the molecular weight and calculated LogP values are in agreement with previous research of drug database. The plant information in database is helpful to understanding the mechanism of the Traditional Chinese Medicine and relationship between Traditional Chinese Medicine and western medicine. Finally, a pharmacophore searching for HIV protease inhibitor in this database has been carried out. Database search results shows that there are many novel potential HIV–1 protease inhibitors in this database, which will be helpful for designing new HIV protease inhibitor leads.

Keywords. Traditional chinese medicine; structural database; HIV inhibitor; database searching; pharmacophore; AIDS.

Abbreviations and notations

HIV, Human Immunodeficiency Virus	WHO, World Health Organization
TCM, Traditional Chinese Medicine	AIDS, Acquired Immune Deficiency Syndrome
HIVPR, HIV–1 Protease	3DFS, three–Dimensional Flexible Searching

1 INTRODUCTION

Phytomedicine is a part of health care systems around the world. The World Health Organization (WHO) estimated that 80% of the world's people rely on herbs for their primary health care needs [1]. Traditional Chinese Medicine (TCM) [2,3] derived from thousands of years of observation and empirical evidence is one of the brightest pearls in the treasure chest of Chinese cultural inheritance. And it is still practiced alongside western medicine at every level of the healthcare system in China.

[#] Presented in part at the Internet Electronic Conference of Molecular Design, IECMD 2003.

* Correspondence author; phone: 0086–10–62626703; fax: 0086–10–62561822; E–mail: jjzhou@home.ipe.ac.cn.

Traditional treatments include herbal remedies, acupuncture, acupressure and massage, and moxibustion. Recently, Chinese phytomedicine has been of great interest to the international research community [4] due to the global trend of using natural products in prevention and diagnosis of physical and mental diseases.

However, there are many challenges in incorporating Chinese herbs into modern clinical practices. Lack of an understanding on material basis of TCM, difficult to keep quality and consistency of herbal products, hard to identify primary bioactive components, and weak to explain the treatment of the disease in term of western anatomy and physiology, are all barriers of worldwide acceptance and development of TCM.

With rapid development of computer technologies and molecular modeling methods, computer–aided study has become an essential tool of drug design. But the drug discovery process is still a money and time–consuming process and it becomes more and more difficult to develop a new drug. So all kinds of structural databases, especially natural product structural database, were used to shorten the cycle of drug discovery by utilizing the abundant biochemical data. Herbs attract more and more attention, since many natural products, especially herbs, have been used for treating special disease. A growing number of successful drug design cases [5-8] through structural database searching have been reported.

Due to the idea above, we developed a TCM plant – compound database in which not only researchers who study TCM but also scientists in western medicine field will find their exciting points.

Although there are a few successful natural product databases now [9,10], they all have their own limits, e.g. the NAPRALERT [9] lacks structural information and the DNP [10] lacks herbal details. Our database integrates TCM plant and compound data, so it builds a bridge between TCM plant and compound, namely, between traditional and modern medicine. It enables user to browse the botanical information and immediately to view chemical structures and bioactivities. Although currently the scale of commercial natural product databases is far beyond ours, their data have been collected without consideration of the therapeutic effects.

The data in our database are mainly from the books and most publications about TCM in Chinese to 2003 and some in foreign language. The current version of this database is the continuation of previous work in our laboratory and update of the molecular bio–data and source plant information in the book *Traditional Chinese Medicines: Molecular Structures*, *Natural Sources and Applications* [11,12] which contains information from western studies of anatomy and physiology. Naturally, it is expected that our database could, more or less, be more efficient in drug design than common database since all compounds presented here are stemmed from medicinal botanies reputed in ancient TCM. At last, an application of this database using database searching was implemented.

2 MATERIALS AND METHODS

2.1 Scheme of Database Design

Traditional Chinese medicine is a system of health care that has evolved over 3,000 years. Unlike the way orthodox western medicine looks at the human body, TCM views a person's health in a holistic fashion. Our bodies are viewed as being made up of two opposing forces– Yin and Yang. When these forces are out of balances, a person will feel sick and symptoms will arise, that is, 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. So the TCM physician will also diagnose illnesses differently from the western physician. Apart from the usual history of the illness, the TCM doctor will also look at different parts of the body to gain more information about the internal organ and their energies. For example, a person's ears are considered as the window to his kidneys.

Once an illness is diagnosed, a TCM physician will prescribe a treatment that focus on trying to restore the balance of the body's energy. Modalities such as Acupuncture, herbal medicine or exercise will be used. As well, the TCM physician will treat the entire person, including both the physical and the mental aspects to cure patient's illness.

Although the difference of the theory and diagnosis between traditional medicine and western medicine, the treatment actions are both due to the interactions between drug or action and their biological receptors [13]. For example, acupuncture gets more plausible biological support for its action (such as the gate theory and endorphin release). So we collect not only the compound information but also the plant information that the compound is isolated from, which will be valuable in the following respects: (1) an understanding of Chinese medicine theory, clinical practice, and its relation to western medicine; (2) speeding up the searching process of lead compounds and improving the shooting accuracy; (3) setting up new pharmacophore models to guide new drug discovery through finding the common structural characteristic among bioactivity components extracted from medicinal herbs with the same curative effects to specific disease; and (4) identifying the biochemical composition of the active agents in many of the herbal preparations from western standpoint. This approach was successful in research of the antimalarial drug –qing hao su. It's noted that many compounds here are not listed in any commercial database by now since their original papers were published in Chinese.

Based on the characteristics of this database, 24 data fields have been defined. These fields have been organized into three parts described in the following. The first part is molecular information including English name, synonyms, physical properties, bioactivity data, formula, molecular weight, CAS Registry Number and structural information, including 2D and 3D structure. Most of 3D structures are generated by molecular modeling software– Corina [14]. The molecules failed in

generation were minimized to generate 3D coordinate by molecular mechanics module in SYBYL6.5 using Tripos force field. The second part is reference information in which the information of compound is cited. The last part is the information about natural source including Latin name, English name, pinyin name, used part, indication, family, curative effect and other additional information. All compounds in the database were carefully chosen from those isolated from the Chinese traditional medicinal herbs (very few of them is animal). Partial resources used for data collection are listed in Appendix 1.

2.2 Database Administration

The database employs ISIS/Base chemical information management system for storing, searching and retrieving chemical structures and other scientific data with customizable forms, so it can utilize all kind of tools in ISIS/BASE. Though user can browse, retrieve and edit every record in database by graphical interface in system, he have not to learn additional programming language and detailed command syntax. The database runs on MS Windows operating system.

With the help of ISIS/BASE, it is easy to retrieve the database. The only thing that the user has to do is to submit their query with legal syntax, which is very easy. To search the database using ordinary data, the user should build a query with search operators or logical operators. It also enable user to submit structure query by drawing query structure in structural form. Anyway, it is a little case for the ISIS/Base user.

2.3 Database Statistics

The records in the database contain compound information, natural source data and bibliographic information. By now, there are 9127 compound records and 3922 plant (a very few is other natural source) records. The plants in the database distribute in 307 families. 9126 compounds have 2D and 3D structure.

In drug design process, there are many factors that should be considered in order to avoid the pitfall of candidate explosion, that is, the design process can be further streamlined by focusing on "drug–like" molecules. Many molecular physicochemical properties are employed to define the drug–like, such as LogP, the ring number, the number of H–bond donors, the number of H–bond acceptors, the number of rotating bonds and molecular weight. It's easy to understand why these properties are used as drug–like parameters. For example, molecular weight is the molecular size representation. If it's too large, it's difficult to arrive to the active site, on the contrary it will hardly bind effectively to the protein. With this in view, the molecular weight and calculated LogP in this database have been analyzed.

Similar component have similar bioactivity, likewise, related plant will contain near component. We numbered the families in the database, which is helpful for describing the plant distribution. At

the same time, we analyze the frequent bioactivity and calculated LogP value in order to have a general image on our database.

2.4 Pharmacophore Searching

By the end of 2002, the World Health Organization estimated 42 million people worldwide have been already infected with human immunodeficiency virus (HIV), the causative agent of acquired immune deficiency syndrome (AIDS), and projected that number is growing. For example, HIV newly infected 5 million people totally in 2002 [15]. One potential therapeutic target is the HIV protease enzyme, which plays a key role in viral maturation. Although enzyme inhibition and antiviral activity in vitro has been observed with several HIV protease inhibitors, most potent candidates disclosed are peptide–derived compounds to date. Clinical development of such compound is often complicated by unfavorable pharmacokinetic parameter, such as low oral bioavailability, rapid excretion, and lengthy syntheses. One potential solution for these problems is use of low molecular weight, nonpeptidic inhibitor. With this global in mind, a 3D database searching was carried out in our database.

2.4.1 Pharmacophore searching software

3DFS [26] (3D Flexible Searching) developed by our laboratory was used as database searching software, which searches 3D database for compounds matching a given pharmacophore. It can be downloaded in website (http://www.embl–heidelberg.de/~twang/3dfs.html). It supports not only simple atom–based query but also generalized function–based query including detailed definitions of hydrogen bond acceptors/donors, positive/negative charge centers, aromatic ring center and hydrophobic. Its characteristics lie in two aspects: (1) using a set of practical binding site definitions and a rapid hydrophobic recognition algorithm for the function–based query; and (2) using a set of effective searching algorithm different from those used in other 3D searching system.

2.4.2 Searching Pharmacophore

A pharmacophore model represents the necessary 3D orientation of chemical features responsible for biological activity. The pharmacophore model provides information about chemical features that interact with active site residues, and provides a feature template to test new compounds for their capacity to adopt conformations necessary to fit the model features. Pharmacophore models can also be used in 3D database searching for finding potential bioactive compounds and providing new research ideas and directions. Combining pharmacophore and shape queries can be an extremely useful technique for mining databases for likely active structures. In order to search our database, the HIV–1 protease inhibitor pharmacophore [6] shown in Figure 1 is used.

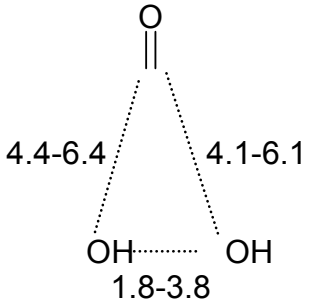

Figure 1. HIV protease inhibitor pharmacophore.

3 RESULTS AND DISCUSSION

Using statistic method, some profiles of database are gained and the results are shown. There are 307 families in our database (Table 1); in other words, most families of plants are covered in our database. The dominant families are Compositae and Leguminosae, which is consistent with the distribution of plant in nature. Because the larger family has more species, the ancient people will be more convenient to get them for their treatment. The plant information will direct the drug design and exploit some relationship between TCM and western medicine by mining the database.

The statistical profile of frequent bioactivity of the compound is listed in Table 2. Many bioactivities in our database are hot points in research in recent years such as antineoplastic and antihypertensive activity. These bioactive recorders are helpful for the drug design research. At this point, the bioactivity in this database is the epitome of recent drug research.

Molecular weight is related to molecular size. Our database has an average molecular weight of 421 and a standard deviation of 233. The proportion of the compounds in which molecular weight is less than 150 is very low, the same as the compounds greater than 600. The molecular weight distribution in this database is shown on figure 2. On the basis of careful analysis of the histogram, we can conclude that the most molecular weight are in range from 100 to 600, which is agreement with previous research [16-18] for drug database. Most outliers having high molecular weight are focusing on classes of sterol, grease, oligose, alkaloid and glucoside. So in the new drug discovery, we will probably not miss much important chance if we avoid too large especially too small molecules.

The logP value was calculated using XlogP [19] method. The average XlogP value of this database is 2.12 with a standard deviation of 3.02. The distribution of XlogP is shown in Figure 3. Most XlogP values are distributed between –2 and 6. By analysis of the compound whose XlogP value is greater than 6, most of them have relatively high molecular weight. The outlier with larger XlogP value has the similar compound classes with set of higher molecular weight. Some of them have a very hydrophobic hydrolysable group. It is possible that some of these compounds have the effect of prodrug. Some of these outliers may resemble some naturally occurring compounds of the body and may have an active transport mechanism over passive transport.

Table 1. Statistical profile of the family frequency larger than 10 of the plant in the database

Family name	No	Family name	No	Family name	No
Agavaceae	10	Bovidae	16	Guttiferae	33
Betulaceae	10	Campanulaceae	16	Myrtaceae	33
Caprifoliaceae	10	Myrsinaceae	16	Scrophulariaceae	36
Ebenaceae	10	Caryophyllaceae	18	Verbenaceae	36
Ephedraceae	10	Podocarpaceae	18	Araliaceae	37
Sterculiaceae	10	Piperaceae	20	Menispermaceae	37
Buxaceae	11	Saxifragaceae	20	Polygonaceae	38
Chenopodiaceae	11	Convolvulaceae	22	Boraginaceae	40
Dryopteridaceae	11	Dioscoreaceae	22	Gramineae	43
Geraniaceae	11	Loganiaceae	22	Lauraceae	44
Juglandaceae	11	Malvaceae	23	Berberidaceae	46
Salicaceae	11	Rhamnaceae	23	Cruciferae	46
Amaranthaceae	12	Amaryllidaceae	24	Gentianaceae	52
Lardizabalaceae	12	Aristolochiaceae	24	Cucurbitaceae	53
Polygalaceae	12	Oleaceae	24	Rubiaceae	61
Taxaceae	12	Zingiberaceae	24	Euphorbiaceae	77
Thymelaeaceae	12	Annonaceae	25	Solanaceae	77
Crassulaceae	13	Magnoliaceae	25	Papaveraceae	82
Polypodiaceae	13	Orchidaceae	25	Apocynaceae	86
Araceae	14	Papilionaceae	25	Rosaceae	107
Simaroubaceae	14	Anacardiaceae	27	Liliaceae	120
Bignoniaceae	15	Pinaceae	27	Labiatae	132
Iridaceae	15	Celastraceae	28	Umbelliferae	134
Meliaceae	15	Cupressaceae	28	Rutaceae	139
Nymphaeaceae	15	Ericaceae	28	Ranunculaceae	146
Primulaceae	15	Moraceae	30	Leguminosae	262
Sapindaceae	15	Asclepiadaceae	33	Compositae	346

Table 2. The most frequent bioactivity in the database

Bioactivity	No
antidiabetic	99
antiviral	108
cardiotonic	126
analgesics	164
sedative–hypnotics	173
m–choline receptor agonist	175
antifungal	271
antihypertensive	286
nonsteroidal antiinflammatory	300
antibacterial	503
antineoplastic	1087

To identity structurally novel HIV–1 protease inhibitor, its pharmacophore model was used as 3D query to search our database and resulted in identification of a total of 820 compounds (hits). Since hydrophobic interactions are known to be important in the binding of inhibitors to HIVPR, the hydrophobic moiety was added in the searching scheme. Among results, some reported HIVPR inhibitors were found. Compound **1** is from the coumarin class which is similar as the warfarin [20],

phenprocouman [21] and PD099560 [22] which all have HIVPR inhibitory bioactivity. Moreover, there are over twenty flavone compounds and **2** which shares most common substructure with resistomycin [23] possessing HIVPR inhibitory bioactivity. On the other hand, we also find many other novel structures which maybe have HIVPR inhibitory bioactivity shown in Figure 5. All the selected compounds satisfy the 3D requirements of HIVPR inhibitor and have at least one hydrophobic moiety which interacts with the HIVPR subsists [24,25] (S1, S2, S1' and S2'). At the same time, we also consider other information in database such as bioactive information and TCM plant information. Then these compounds have been selected, but they all need the bioassay.

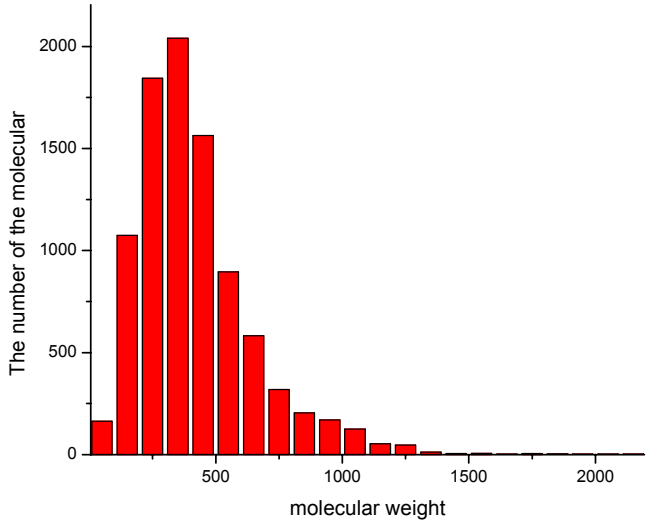

Figure 2. Histogram of molecular weight distribution.

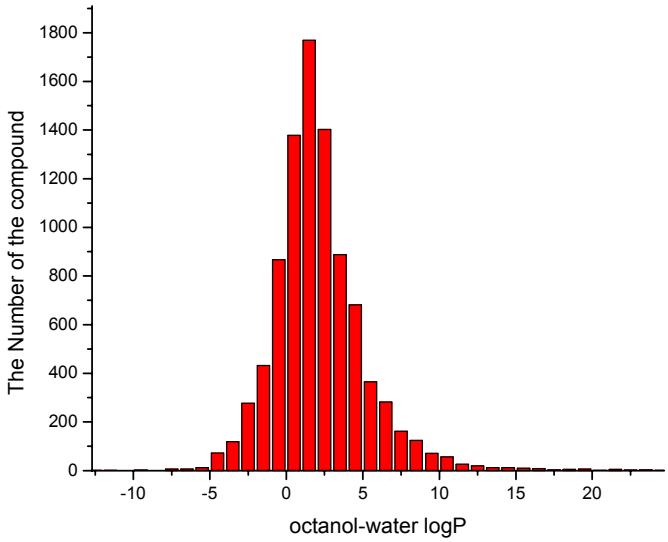

Figure 3. Histogram plots of octanol–water logP (XLOGP) distribution for TCM database.

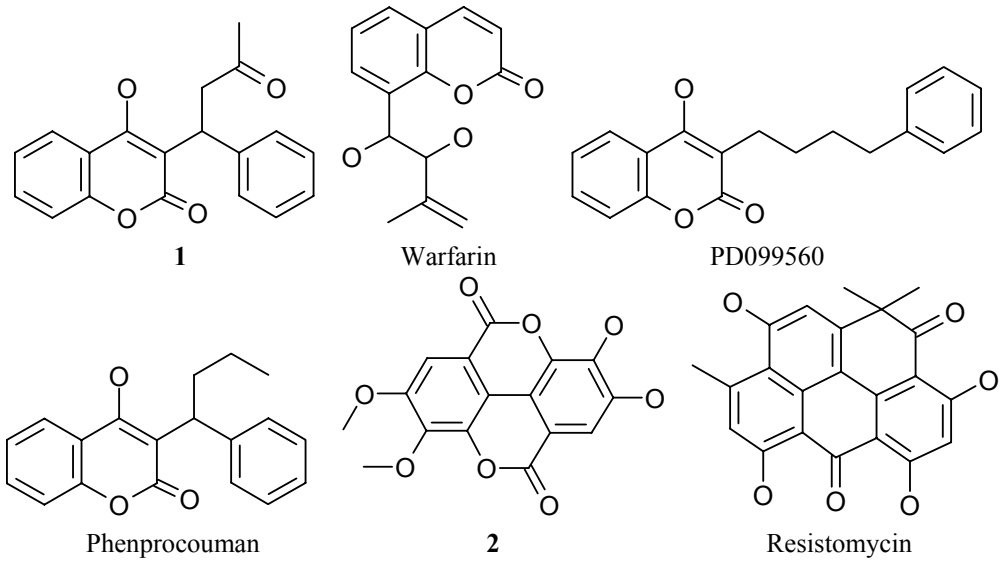

Figure 4. The bioactive molecule reported and the similar molecule in database.

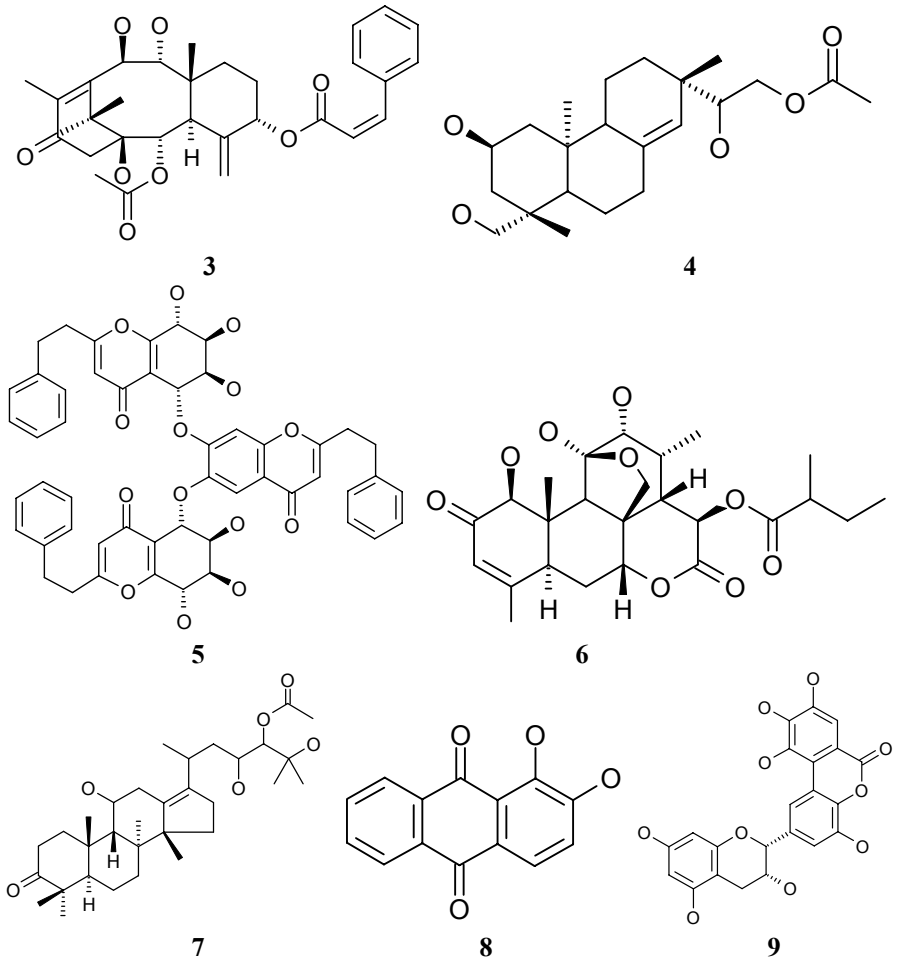

Figure 5. The possible novel bioactive molecules in database by 3D database searching.

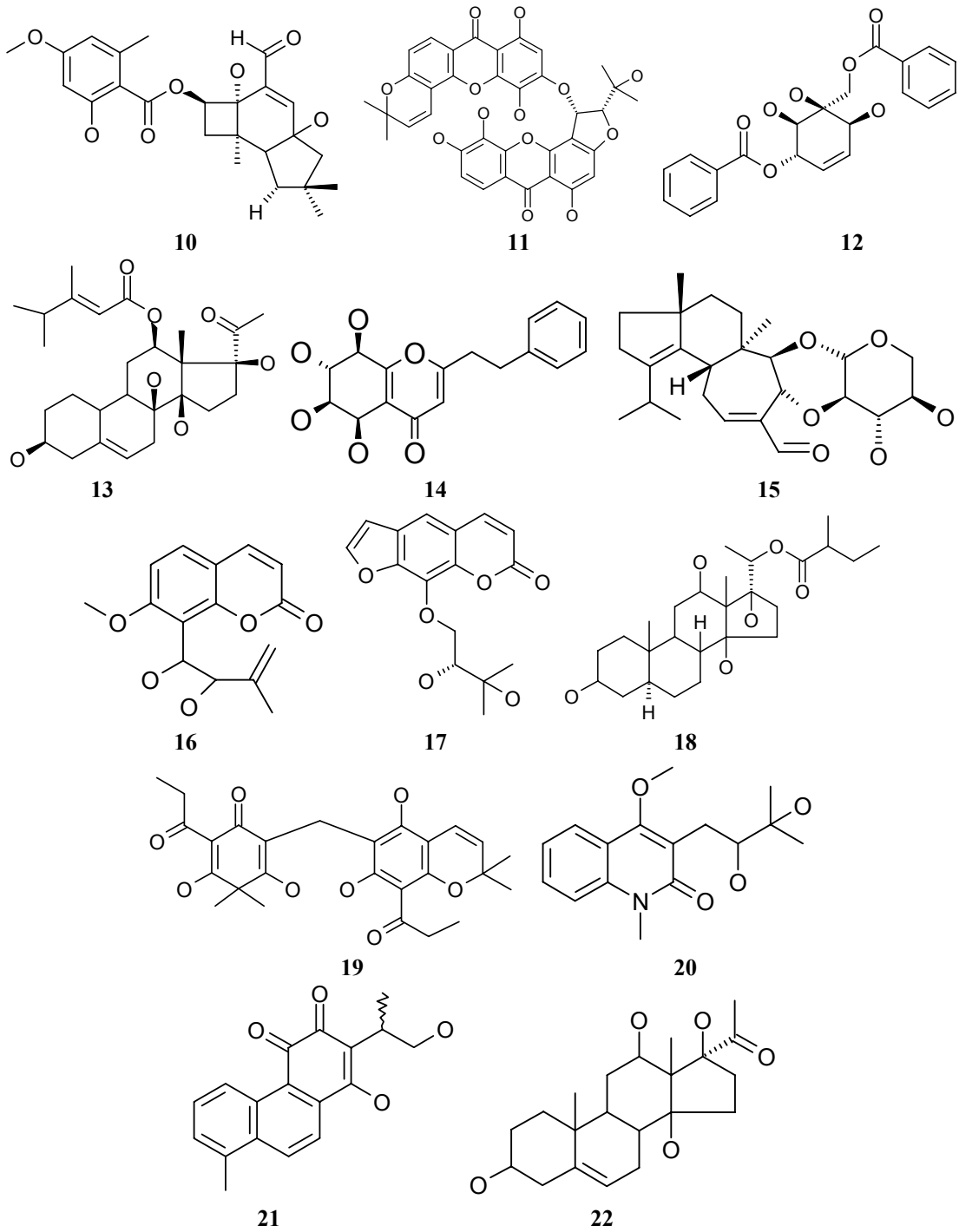

Figure 5. (Continued).

4 CONCLUSIONS

We developed a database named Traditional Chinese Medicines Plant–Compound Database. Some statistic studies have been done, which are helpful for understanding the database properties. Then 3D database pharmacophore searching in this database was carried out. Some structures similar known HIVPR inhibitors in hits have been discussed. The novel structures that may possess

potential HIV–1 protease inhibitory bioactivity were found. This investigation has demonstrated that structural database can be a source for new leads.

Acknowledgment

The authors acknowledge Dr. Zhenming Liu in the Chemistry Department of Peking University, China for converting two–dimensional structures to three–dimensional structures for the database using Corina software.

Appendix 1

Acta Botanica Yunnanica
Acta Pharmaceutica Sinica
China journal of Chinese Materia Medica
Chinese Traditional and Herbal Drugs
Journal of Natural Product
Natural Product R&D
Chinese Pharmaceutical Journal
Phytochemistry.
Planta Med

edited by Institute of Materia Medica, Chinese Academy of Medical Sciences, *Modern Studies of Chinese Herbal Medicine*, *Union Press of Beijing Medical university and Peking Union Medical College*, *Beijing*, 1996 (in Chinese).

Editing Group of the Handbook of Bio–activity Components from Medicinal Plants, *Handbook of Bio–activity Components from Medicinal Plants*, The People's Medical Publishing House, Beijing, 1986 (in Chinese).

Jian Yin and Ligong Guo, *Modern Study of Chinese Drugs and Clinical Applications* (1), Xueyuan Press, Beijing, 1993 (in Chinese).

Yubin Ji (chief editor), *Pharmacological Action and Application of Available Composition of Traditional Chinese Medicine*, Heilongjiang Science and technology Press, Heilongjiang, 1995 (in Chinese).

Yubin Ji and Guangmei Zhang (chief editors), *Pharmacological Action and Application of Available Antitumor Composition of Traditional Chinese Medicine*, Heilongjiang Science and technology Press, Heilongjiang, 1998 (in Chinese).

Jiangsu New Medical College, *Chinese Medicine Dictionary*, Shanghai Science and technology Press, Shanghai, 1979 (in Chinese).

Chinese Materia Medica Editing Committee of the National Chinese Medicine and Pharmacology Bureau, *Chinese Materia Medica* (*selection version*), Shanghai Science and technology Press, Shanghai, 1998 (in Chinese).

Guojun Xu et.al., *Chinese Materia Medica*, Chinese Medicinal Science and Technology Press, 1996 (in Chinese).

Wenji Sun and Jinfang Sneng, *Brief Handbook of Natural Active Compounds*, Medicinal Science and Technology Press of China, Beijing, 1998 (in Chinese).

Jiwu Wang and Qingxiang Xiao, *Handbook of Effective Components in Vegetal Medicines*, People Health Press, Beijing, 1986 (in Chinese).

Huifang Chen, Yong hua Ma and Xuewei Bian, *Lexicon of Active Componentsin in Plants*, *Medicinal Science and Technology Press of China*, Beijing, 2001 (in Chinese).

C. Dierassi, JD. Connolly, DJ. Faulkner, K. Mori, K. Nakanishi, G. Ourisson, RA. Raphael, M. Shamma and Ch. Tamm (International Advisory Board), J. Buckingham (Executive Editor), Dictionary of Natural Products, Chapman & Hall, London, 1994

5 REFERENCES

[1] Olayiwola Akerele, Summary of WHO guidelines for the assessment of herbal medicine, *Herbalgram*, **1993**, *28*, 13–18.

[2] Li S., *The History of Traditional Chinese Medicine (in Chinese)*, Science Press, Beijing, 1996.

[3] Hesketh T. and Zhu W., Health in China: Traditional Chinese Medicine: one country, two systems, *BMJ*, **1997**, *315*, .115 – 117.

[4] Kenner D. and Requena Y., *Botanical medicine: a European professional perspective*, Paradigm Publishing, Brookline, MA, 1996.

[5] Yasuhisa Kurogi, K. M., Takashi Okamura, Kinji Hashimoto, Kazuhiko Tsutsumi, Masahiro Nasu and Matsuko Moriyasu, Discovery of Novel Mesangial Cell Proliferation Inhibitors Using a Three–Dimensional Database Searching Method, *J. Med. Chem*. **2001**, *44*, 2304–2307.
[6] Shaomeng Wang, G. W. A. Milne and Xinjian Yan, Discovery of Novel, Non–Peptide HIV–1 Protease Inhibitors by Pharmacophore Searching, *J. Med. Chem*., **1996**, *39*, 2047–2054.
[7] David P. Marriott, I. G. D., Premji Meghani, Yu–Jiang Liu and Darren R. Flower, Lead Generation Using Pharmacophore Mapping and Three–Dimensional Database Searching: Application to Muscarinic M3 Receptor Antagonists, *J. Med. Chem*. **1999**, *42*, .3210 – 3216.
[8] Grace Shiahuy Chen, C.–S.C., Wai Ming Kan, Chih–Long Chang, K. C. Wang and Ji–Wang Chern, Novel Lead Generation through Hypothetical Pharmacophore Three–Dimensional Database Searching: Discovery of Isoflavonoids as Nonsteroidal Inhibitors of Rat 5α–Reductase, *J. Med. Chem*. **2001**, *44*, 3759–3763.
[9] W. D. Loub, N.R.F., D. D. Soejarto and M. L. Quinn, NAPRALERT: computer handling of natural product research data, *.Chem. Inf. Comput. Sci*., **1985**, *25*, .99–103.
[10] http://www.chemnetbase.com/scripts/dnpweb.exe.
[11] Xinjian Yan, Jiaju Zhou and Guirong Xie, *Traditional Chinese Medicines: Molecular Structures*, *Natural Sources and Applications*, ASHGATE, Burlington, VT, 1999.
[12] Jiaju Zhou, Xinjian Yan and Guirong Xie., *Traditional Chinese Medicines: Molecular Structures*, *Natural Sources and Applications*, ASHGATE, Burlington, VT, 2003.
[13] Xinjian Yan, Jiaju Zhou and Zhihong Xu, Concept Design of Computer – Aided Study on Traditional Chinese Drugs, *J. Chem. Inf. Comput. Sci*., **1999**, *39*, .86–89.
[14] Gasteiger, J., Rudolph, C. and Sadowski, J., Automatic Generation of 3D Atomic coordinates for Organic Molecules, *Tetrahedron Comput. Method*, **1992**, *3*, .537–547.
[15] http://www.who.int/hiv/pub/epidemiology/epi2002/en/.
[16] Arup K. Ghose, V.N.V.a.J.J.W., A Knowledge–Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases, *J. Comb. Chem*., **1999**, *1*, 55 – 68.
[17] Lipinski C. A., L.F., Dominy, B. W. and Feeney P. J, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv. Drug Delivery Rev*. **1997**, *23*, 3–29.
[18] McGregor, M.J.a.P., P. V, Clustering Large Databases of Compounds: Using MDL "Key" as Structural Descriptors, *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 443 – 448.
[19] Renxiao Wang, Ying Fu and Luhua Lai, A New Atom–Additive Method for Calculating Partition Coefficients, *J. Chem. Inf. Comput. Sci*., **1997**, *37*, 615–621.
[20] Bourinbaiar A. S., Tan X. and Nagoruy R., Effect of the Oral Anticoagulant, Warfarin, on HIV–1 Replication and Spread, *AIDS* **1993**, *7*, 129–130.
[21] Tummino, P. J., D. Ferguson and D. Hupe, Competitive inhibition of HIV–1 protease by warfarin derivatives, *Biochem. Biophys. Res. Commun*. **1994**, *201*, 290–294.
[22] Tummino, P. J., Ferguson, D., Hupe, L. and Hupe, D., Competitive Inhibition of HIV–1 Protease by 4–Hydroxy–benzopyran–2–ones and by 4–Hydroxy–6–phenylpyan–2–ones, *Biochem. Biophys. Res. Commun*. **1994**, *200*, 1658–1664.
[23] Roggo B. E., P. F., Delmendo R, Jenny HB, Peter HH and Roesel J., 3–Alkanoyl–5–hydroxymethyl tetronic acid homologues and resistomycin: new inhibitors of HIV–1 protease. I. Fermentation, isolation and biological activity, *J. Antibiot.* (*Tokyo*) **1994**, *47*, 136–142.
[24] Wlodawer, A.a. E., J. W., Structure–based inhibitor of HIV–1 proteas, *Annu. Rev. Biochem* **1993**, *62*, .543–585.
[25] Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasumi, M, The Protein Data Bank: A Computer based Archival File for Macromolecular Structures, *J. Mol. Biol.* **1977**, *112*, 535–542.
[26] Ting Wang, Jiaju Zhou, 3DFS: A New 3D Flexible Searching System for Use in Drug Design. *J. Chem. Inf. Comput. Sci.* **1998**, *3*, 71–77.