A Traditional Chinese Medicine Plant-Compound Database and it's Application for Searching

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

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

Internet Electronic Conference of Molecular Design **2003**, November 23 – December 6

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 names and synonyms, physical properties, natural sources, bioactivity data, formula, molecular weight, CAS Registry Number and 2D, 3D structure of the molecule but also detailed information on their nature source, including Latin name, English name, Pinyin name, used part, indication, family, and curative effects. In order to present a basic profile to the user of the database, some statistic data are provided. The result shows that lots of hot bioactive records are in the database and also the molecular weight and calculated LogP value is rational and in agreement with previous research of drug database. The plant information in database is helpful to understand the mechanism of the Traditional Chinese Medicine and the relationship between Traditional Chinese Medicine and Western medicine too. Finally, a pharmacophore searching for HIV protease inhibitor in this database has been carried out. Result shows that there are many novel structures that may possess potential HIV-1 protease inhibitor bioactivity, which will be helpful for designing new HIV protease inhibitor leads.

Keywords. TCM; 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) estimates 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. Traditional treatments include herbal remedies, acupuncture, acupressure and massage, and moxibustion. Recently, Chinese phytomedicine has become a source of great interest to the international research community [4] due to the global trends 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 bioactivity components, and weak to explain the treatment of the disease in term of Western anatomy and physiology, all of them are barriers of worldwide acceptance and development of TCM.

* Correspondence author; phone: 0086-10-62626703; fax: 0086-10-62561822; E-mail: jjzhou@lcc.icm.ac.cn.

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. But the drug discovery process is still a money and time-consuming process and it becomes more and more difficult for the industry to develop new drugs. So all kinds of structure databases have emerged to utilize the abundant biochemical data for shortening the cycle of drug discovery. Especially, the structure database of compounds derived from natural products and herbs attracts more attention, since many natural products, especially herbs medicine are 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 the researchers who study TCM but also the scientists in west medicine field will find its exiting use. Although there are a few successful natural product databases now [9, 10], they all have their own limitations, 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 will be invaluable to medicine chemists because it can enable them to browser the botanical detail and immediately view chemical structures and their bioactivities. Although currently the scale of many commercial natural product databases is far beyond ours, their data have been collected without respect to the therapeutic effects. The data in our database are mainly from the books and most publications about TCM in Chinese journal 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 of the book *Traditional Chinese Medicines: Molecular Structures, Natural Sources and Applications* [11, 12] which contains information gathered independently 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 the commercial natural product 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, such as that shown in the Traditional Yin-Yang symbol. When these forces are out of balances, a person will then 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 your illnesses differently than the Western physician. Apart from the usual history of the illness, the TCM doctor will also look at different parts of your body to gain more information about your internal organ and their energies. For example, a person's ears are considered a window to their 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 exercises will be used. As well, the TCM physician will treat the entire person, including both the physical and the mental aspects to cure their patient's illness.

Although the difference of the theory and diagnosis between traditional medicine and Western medicine, the concept that the treatment actions are all due to the interactions between drug or action and their biological receptors. [13] For example, acupuncture appeals being increased by plausible biological mechanisms for its action (such as the gate theory and endorphin release). So we collect not only the compound of TCM component information but also the plant information that the compound is isolated, will be valuable in the following respects:

1) Understand Chinese medicine theory, clinical practice, and the relationship to Western medicine and elucidate.

2) Speed up the searching process of lead compounds and improve the shooting accuracy, because the compounds in our database come from Chinese medicinal herbs with definite curative effects. Moreover, many of compounds here have not listed in any commercial database by now since their original papers were published in Chinese.

3) Set up a new pharmacophore model 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.

4) Identify the biochemical composition of the active agents in many of the herbal preparations from a western standpoint. This approach has been successful in research into the antimalarial drug qing hao su.

Based on the characteristics of this database, 24 data fields and related codes have been defined. The fields have been organized in three parts described in the following. The first part is molecular information including English names, synonyms, physical properties, natural sources, bioactivity data, formula, molecular weight, CAS Registry Number and structural information., including 2D structure, 3D structure. Most of the 3D structures are generated by molecular modeling software- corina [14]. The molecules failed in generation were minimized to generate the 3D coordinate by the molecular mechanics module in Sybyl6.5 using the Tripos force field. The second part is references information in which the information of compound is cited. The last part is

the information about natural sources including Latin name, English name, Pinyin name, used part, indication, family, and curative effect and other additional information. All the compounds in the database were carefully chosen from those being reported to isolate from the Chinese traditional medicinal herbs (very few of them is animal). Partial resources used in data collection are listed in Appendix 1.

2.2 Database Administratration

The database employs ISIS/Base chemical information management system for storing; searching and retrieving chemical structures and other associated scientific data with customizable forms. The software possesses many excellent characteristics such as user-friendly graphical interface, no additional programming language learning, easy access to the data especially outstanding structure retrieving attribute without requiring knowledge of detailed command syntax. The database runs on MS Windows operating system.

Because of the powerful function of ISIS/Base, It's easy to retrieve the database. The only thing that the user have to do is to submit their query with the legal syntax which is very easy. To search the database using ordinary data, you should build a query with search operators or at most joint logical operators. It's so facilitate to search it using structure that you can accomplish it by drawing the structure in the structural form. Anyway, it's a little case for the ISIS/Base user.

2.3 Database Statistics

The records in the database contain compound information, nature 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 collected in the database distribute in 307 families. 9126 compounds have 2D and 3D structures.

In drug design process, there are many factors that should be considered in order to avoid the pitfall of the 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 being used as drug-like parameter. For example, molecular weight is the molecular size representation. If it's too larger, 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 has been analyzed.

Similar component have similar bioactivity, likewise, close plant will contain near component. We numbered the families in the database that will describe the plant distribution in database. At the same time, we analyze the frequent bioactivity and calculated LogP value in our database in

order to have a general image on our database.

2.4 Pharmacophore Searching

At the end of 2002, the World Health Organization estimated 42 million people worldwide were already infected with human immunodeficiency virus (HIV), the causative agent of acquired immune deficiency syndrome (AIDS), and projected that number is growing, such as that 5 million people newly infected with HIV in 2002 totally.[15] One potential therapeutic target is the HIV protease enzyme, which plays a key role in viral maturation. Although promising in vitro enzyme inhibition and antiviral activity have been observed with several HIV protease inhibitors, most of the more potent candidates disclosed to date are peptide-derived compound. Clinical development of such compounds is often complicated by unfavorable pharmacokinetic parameter, such as low oral bioavailability and rapid excretion, and by lengthy syntheses. One potential solution for the problems often encountered with peptide-derived inhibitors is use of a 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 a 3D database for compounds matching a given pharmacophore query. It's can be download in the website. It supports not only simple atom-based query but also generalized function-based query including the detailed definitions of hydrogen bond acceptors/donors, positive/negative charge centers, aromatic ring center and hydrophobe. Its characteristics lie in two aspects,(1)Using a set of practical binding site definitions a rapid hydrophobia recognition algorithm for the function-based query ;(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 considered 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 potentially finding biologically active compounds and providing new research ideas and directions for your projects. 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.

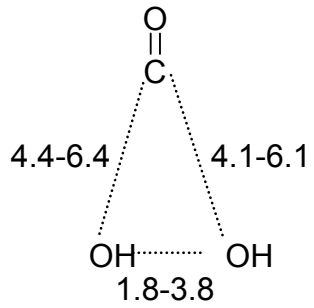


Figure1. HIV protease inhibitor pharmacophore

3 RESULTS AND DISCUSSION

Using statistics method, some of database profiles were gained and the results are shown. There are 307 families in our database; in other words, almost all families of plants are covered in our database. The large family is dominant in the distribution such as Compositae and Leguminosae, which is consistent with the distribution of the plant in nature. Because the larger family has more species, the ancient people will be more convenient to get the herbs for their treatment. In addition, we can get be comprehensive understanding by refer to plant information. Moreover, the plant information will direct the drug design e and exploit some relationship between TCM and western medicine by mining the database.

The statistical profile of the number of the frequent bioactivity of the compound was listed in Table 2. Many of bioactivities in our database are really hot in research in recent years such as antineoplastic and antihypertensive. These bioactive recorders are very 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 which have a molecular weight less than 150 is very low, the same as the compounds larger than 600. The molecular weight distribution in this database is shown on figure 2. On the basis of a careful analysis of the histograms, we can conclude that the most component of molecular weight are in range from 100 to 600, which is agreement with the previous research [16-18] for drug database. Most of the outlier having high molecular weight is focusing on the class of sterol, grease, oligose, glycoalkaloid and glucoside. So in the new drug discovery, we will probably not miss many 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 of the XlogP value are distribute between –2 and 6. By analysis of the compound whose XlogP values were greater than 6, most of these compounds have a relatively high molecular weight. At the same time, the outlier with larger XlogP value has the similar compound class 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 our 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 our database

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, it's pharmacophore model was used as a 3D query to search our database and results in the identification of a total of 820 structurally

compounds (hits). Since hydrophobic interactions are known to the important in the bind of inhibitors to HIVPR, so in the searching scheme, the hydrophobic moiety was added. Among the results, some reported HIVPR inhibitors was found. Such as, 1 was coumarin class which is similar as the Warfarin [20], Phenprocouman [21] and PD099560[22] which all have HIVPR inhibitor bioactivities, moreover, there are over twenties flavone compounds and 2 which shares most common substructure with Resistomycin[23] also having HIVPR inhibitor bioactivity. On the

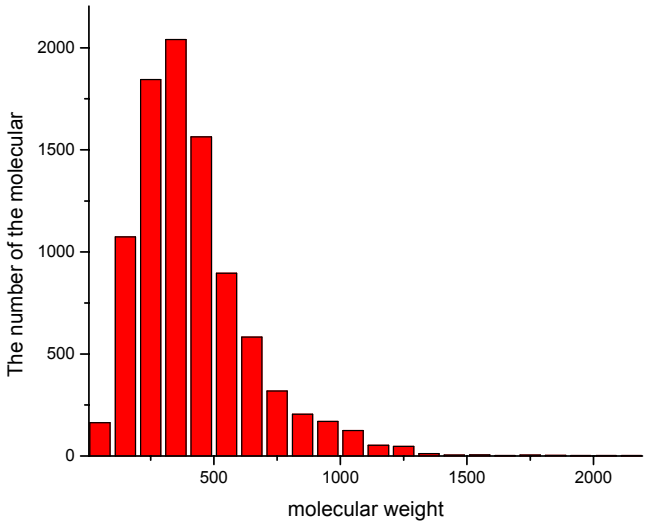


Figure 2. Histogram of molecular weight distribution

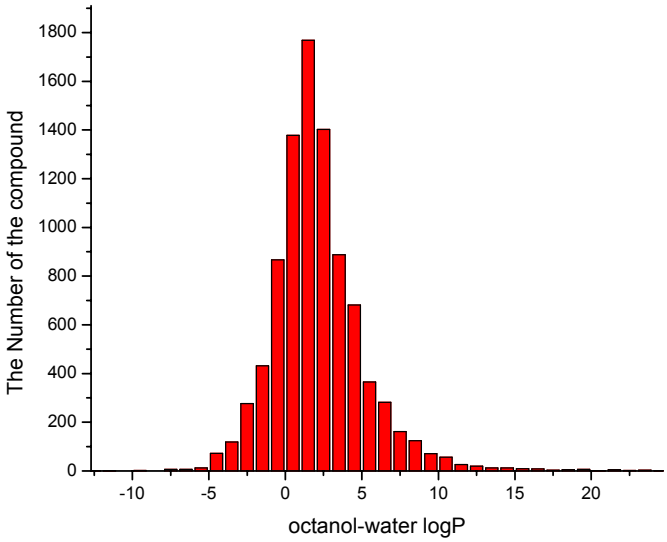
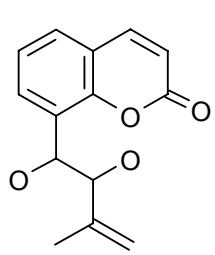
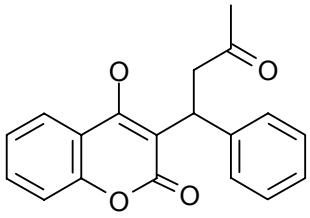


Figure3. Histogram plots of octanol-water logP (XLOGP) distribution for TCM database

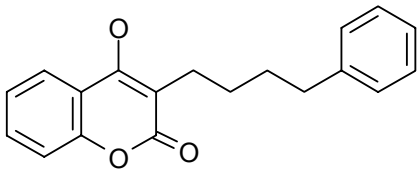
other hand, we also find many other novel structures which maybe have HIVPR inhibitor bioactivity shown in figure 5. All the selected compound have at least hydrophobic moiety which interact with the HIVPR subsists[24, 25] (S1, S2, S1' and S2') and satisfy the 3D requirements of HIVPR inhibitors. At the same time, we also consider other information in database such as the bioactive information and TCM plant information. Then these compounds have been selected. But they all need the bioactivity assay.



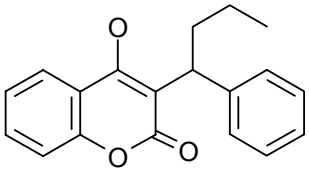
1



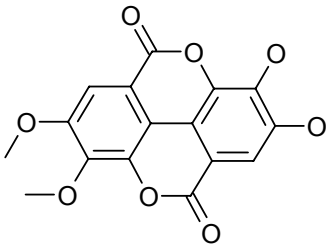
warfarin



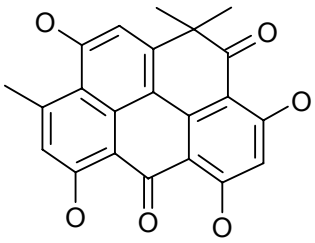
PD099560



Phenprocouman

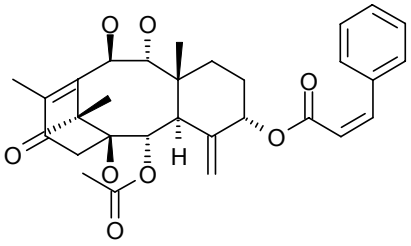


2

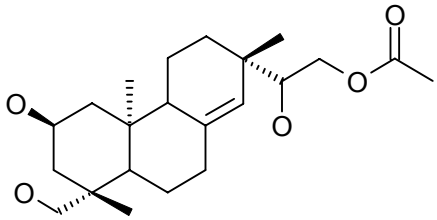


Resistomycin

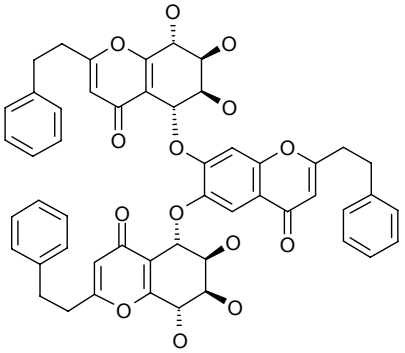
Figure 4. The Bioactive molecule reported and the similar molecule in database



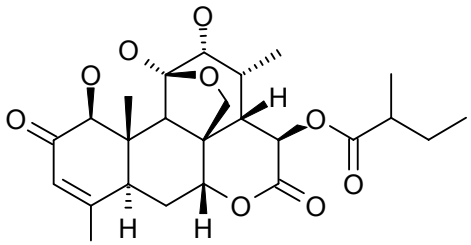
3



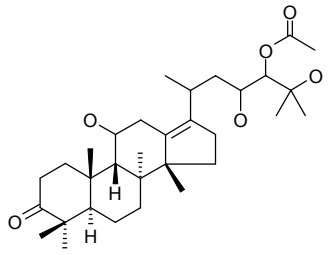
4



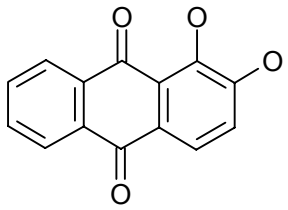
5



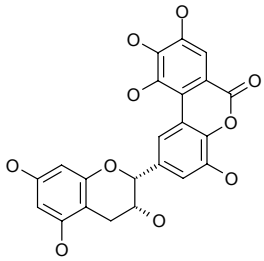
6



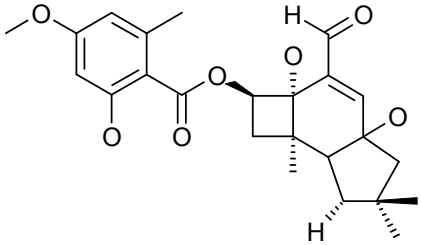
7



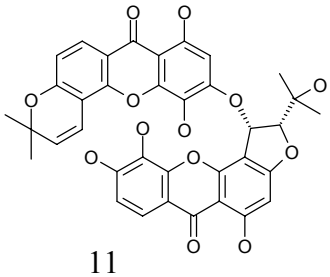
8



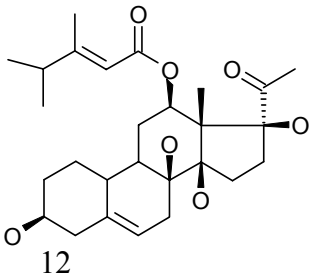
9



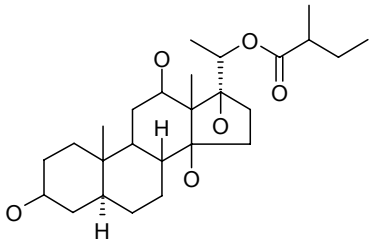
10



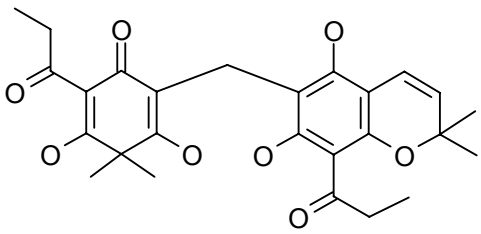
11



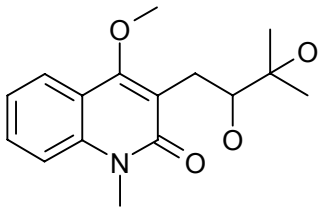
12



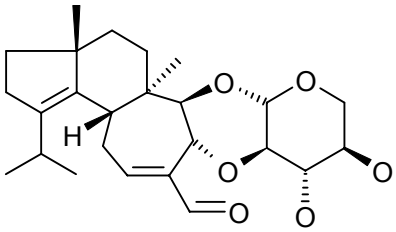
13



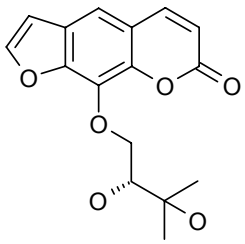
14



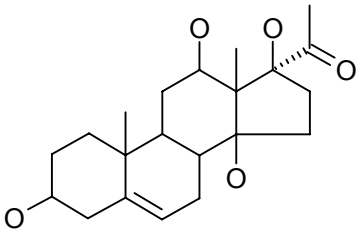
15



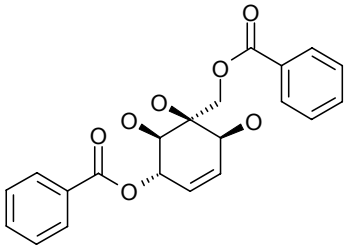
16



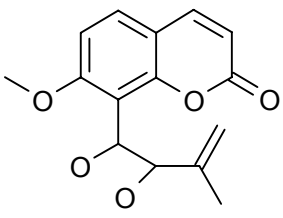
17



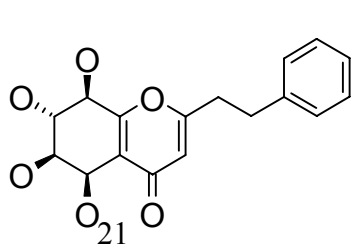
18



19



20



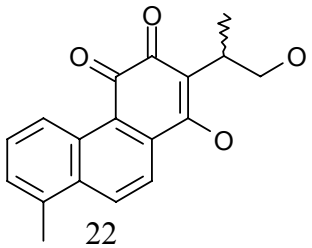


Figure 5. The possible novel bioactive molecules in database by 3D database searching

4 CONCLUSIONS

We developed a database of Traditional Chinese Medicines Plant-Compound Database. Some of statistics analysis has been done which shows the result of the database for understanding the database properties. Then 3D database pharmacophore searching in this database was carried out. Some structures similar know HIVPR inhibitors in hits have been discussed. The novel structures that may possess potential HIV-1 protease inhibitor bioactivity were found. This investigation has demonstrated that structure database can be a source for new leads. Combined with other information, this database is an efficient tool for drug discovery.

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] Jiaju Zhou, Xinjian Yan and Guirong Xie, *Traditional Chinese Medicines: Molecular Structures, Natural Sources and Applications,* ASHGATE, Burlington, VT, 1999

[12] Xinjian Yan, Jiaju Zhou and Guirong Xie., *Traditional Chinese Medicines: Molecular Structures, Natural Sources and Applications,* ASHGATE, Burlington, VT, 1999, 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*, 3-25.

[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 AS., 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 BE, 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*, .136-42.

[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 Computerbased Archival File for Macromolecular Structures, *Journal of Molecular Biology*, **1977**, *112*, .535-542.

[26] Ting Wang, Jiaju Zhou, 3DFS: A New 3D Flexible Searching System for Use in Drug Design. *Journal of Chemical Information and Computer Science,* **1998,** *3,* 71-77