DATABASE DEVELOPMENT AND MECHANISTIC STUDY OF TRADITIONAL CHINESE MEDICINE BY COMPUTER

WANG JIFENG

NATIONAL UNIVERSITY OF SINGAPORE

2003

Founded 1905

DATABASE DEVELOPMENT AND MECHANISTIC STUDY OF TRADITIONAL CHINESE MEDICINE BY COMPUTER

BY

WANG JIFENG

A THESIS SUBMITTED FOR THE DEGREE OF

MASTER OF SCIENCE

NATIONAL UNIVERSITY OF SINGAPORE

2003

Acknowledgment

First and foremost, I would like to express my sincerest appreciation to my supervisor, Associate Professor Chen Yu Zong from Computational Science for his patient guidance, supervision, invaluable advice and suggestions throughout the whole research process.

Sincere gratitude is also expressed to Dr. Cai, Dr. Li, Xue Ying for their patient guidance and œoperation. To Zhiwei, Zhiliang, Chenxin, Lizhi, Chunwei, Lianyi, Chanjuan and Lixia, who are labmates as well as friends, for being ever so willing to share with me their valuable advice on projects, and for sharing with my joy and sorrow at all times.

I would like to thank Ms.Lucee, Ms Lindah, Ms. Hwee sim, Ms Elaine and Ms.Wei Har,for their assistance and friendship.

Most of all, I am eternally grateful to my parents, my GF, for supporting me, and for encouraging me at all times.

Finally, I would like to thank everyone in my department who had helped me in my study.

Wang Jifeng

May 2003

Table of Contents

List of Tables and Figures

Tables **Page**

Figures **Pages**

Summary

Traditional Chinese medicine (TCM) has been used in the treatment of a variety of diseases and is recognized as a valuable alternative to conventional medicine. Increasing effort is being made towards scientific proof, clinical evaluation and molecular study of TCM. To facilitate such an effort, I develop a database which contains the available information about all major aspects of TCM, including herbal formulations, herbal composition, chemical composition, molecular structure and functional properties, therapeutic and toxicity effects, clinical indication and application.

With the rapid development of computer technologies, computational methods have been widely employed in biology. Support Vector Machine (SVM), based on statistical learning theory, is such a method that has been used in a wide range of real-world problems such as text categorization, cancer diagnosis, glaucoma diagnosis, and microarray gene expression data analysis. In this study, SVM is used to facilitate the study of TCM formulae. The results indicate the capability of SVM in recognizing non-effective formulae and it may provide some helpful hints for herbalist doctors to determine the effectiveness of a TCM formula. In addition, the computation provides several potentially effective formulae from the hundreds of randomly mixed formulae. It is unclear whether these formulae have the therapeutic value. The method is expected to facilitate the prescription of new and novel TCM formulae as well as the

validation of existing TCM formulae while more and more formulae are under scientific studies.

The mechanism of action of TCM remains largely unknown, though a large number of active compounds have been isolated from these herbs and their clinical and therapeutic effects have been probed. INVDOCK, a molecular interaction-based method, is employed to study the molecular mechanism of medicinal herbs. This study provides the potential targets of a medicinal herb *Serenoa repens* in the treatment of BPH, parts of which have been demonstrated by previous experiments to be bound by compounds in the extract. Besides these interactions, other bindings between particular compounds and protein targets have not been proven by experiments. It provides a new method for exploration of the mechanism of herb medicine. It is also of importance in drug development based on herbs. In conclusion, as a relatively fast-speed and low-cost tool, this method may find application in systematic study of the molecular mechanism of multiple ingredients of other medicinal plants and has to be further validated by clinical trials.

Chapter 1: Introduction

1.1 Brief History of Traditional Chinese Medicine (TCM)

Traditional Chinese Medicine (TCM) has been used for thousands of years [1-5]. At least it has a recorded history dating back over 2,000 years. Among numerous legends about the origins of traditional Chinese medicine, stories about three legendary emperors/mythical rulers: Fu Xi, Shen Nong and Huang Di have to be mentioned. Historians believe that Shen Nong and Fu Xi were early tribal leaders. Fu Xi is regarded as a cultural hero who developed the trigrams of *Yi Jing*. Ancient books said 'Fu Xi drew the eight trigrams, and created nine needles'. Shen Nong, the legendary emperor who lived 5000 years ago, is regarded as the 'Divine Farmer' by the Chinese people because of his attribution as the founder of herbal medicine. To test and analyze the individual effects of different plant medicine, Shen Nong ingested them by himself. It was said that Shen Nong tested over one hundred herbs including 70 toxic substances in one day in order to find some drug to get rid of people's pain form illness.

Huang Di Nei Jing (Yellow Emperor's Cannon of Internal Medicine) is the first written documentation on traditional Chinese medicine, which was written between 800 BC and 200 BC [6]. This book summarized and systematized the previous

experience of treatment and theories of medicine, such as the meridian theory, as well as many other issues, including physiology, pathology, prevention, diagnosis, treatment and acupuncture etc. It was regarded as the foundation for the theories of Chinese medicine.

During the Zhou dynasty, the theory of TCM was developing very fast. The most important discoveries of medicine were made, including the theoretical foundations of *Yin* and *Yang* [7], the five elements, the pathogenic factors of external environment as a cause of disease and further understanding of the meridians of acupuncture.

The basic theories of acupuncture were established and stone needles became obsolete, being replaced by metal needles. At the time of the Spring and Autumn Warring States Period, one of the most important issues in the development of TCM was the usage of the pulse for diagnosis. Bian Que [8], a very famous doctor/physician was the first man in the world to use this technique. He was reputed to be an excellent diagnostician, excelled in using acupuncture and moxibustion, boiled herbal prescriptions, and massage in internal medicine, external medicine, gynaecology, and paediatrics for the treatment of all kinds of illnesses. Bian Que also recorded his experiences in the book *Nan Jing* (The Classic of Difficult Issues), which developed and explained the fundamental and difficult parts of *Huang Di Nei Jing*.

During this Period another famous classic book *Shen Nong Ben Cao Jing* (Shen Nong's Classic of Materia Medica) [9] was written. This book recorded and described the characters of about 365 herbs and was regarded as the first pharmacopoeia of

traditional Chinese herbal medicine.

The Han dynasty, which lasted almost four and a half centuries, was a period of a thriving Chinese culture. TCM was well developed as well. Zhang Zhongjing (150-219 A.D.), one of the most famous herbalist doctors, was reputed for his remarkable medical skill and his well known medical masterpiece *Shang Han Lun* (Treatise on Febrile Diseases) [10], which was used as a standard reference work for traditional Chinese medicine, including moxibustion, needling and herbal medicine. So far his theory and prescriptions are still of great practical value. In this period there also was one famous physicians/surgeon of traditional Chinese medicine, Hua Tuo (110-207 A.D). He invented the use of anaesthesia called *Ma Fei San* [11] to reduce the pain of patients who was suffering surgery. He also furthered the knowledge of anatomy. He was the first person who used narcotic drug in the world and his skill in this field was ahead of the west about 1600-1700 years.

Li Shizhen was considered to be China's greatest scholar in TCM and made great contribution to the progress of TCM. He spent forty years to collect and taste herbal medicines and wrote down in his most well-known Chinese herbal book *Ben Cao Gang Mu* (Herbal Systematics) [12,13], which contained 1892 different herbs (with 1110 drawings), and was divided into 6 sections, 52 scrolls and 60 different categories and also included 11,096 prescriptions, for treating hundreds of illnesses, ranging from the common cold to drunkenness and food poisoning. Others such as Wang Shuhe, Huang Fumi, Ge Hong, Chao Yuanfang, Sun Simiao and Wang Weiyi also wrote important medical manuals and contributed to the thriving of TCM.

Indeed, TCM has a long history, and in what is regarded as a breakthrough, the World Health Organization (WHO) in 1979 released a list of 43 types of pathologies, which can be effectively treated with acupuncture. Today, there are many hospitals in China, Japan, and the other Asia countries that practice TCM exclusively, and others that combine Eastern and Western healing methods. One of the main reasons that TCM is still used may be their relatively low side effects, compared with western medicine. Another reason may be that it is used as a last resort, when Western medicine is too toxic or unable to provide the expected benefit [14]. In recent years, the effectiveness of TCM has been gaining popularity in US and European countries, such as Germany, France and UK [2,15]. However, there are also negative attitude toward TCM. Some westerners think that TCM is unscientific in its understanding of the human body and the nature of disease and its treatment. They believe that the lack of quality control and the absence of scientific and clinical proof of their effectiveness will impede the adoption of TCMs in industrialized countries [1,16-18].

1.2 Chinese Medicinal Herbs in TCM

As the most important parts of TCM, medicinal herbs and the prescriptions of multiple herbs, which will be discussed in the next chapter, were given much attention in the development of TCM [4,5]. In TCM, people think that the actions of herbs in treatment of diseases is through rectifying the balance of *Yin* or *Yang*, and then consequently helping the body restore its normal physiological functions. Different herbs have different characters and functions. That is why different herbs are used when curing different diseases. To understand the mechanism of them and use them

properly, it is necessary to study and explain their characters. Ancient experts had, from the TCM viewpoint, summarized those characters, which include drugs' Properties and Flavors, Meridians of herbs and Toxicity property, etc. Based on the theories of *Yin* and *Yang*, Viscera, Channels and Collaterals, and treatment principles of traditional Chinese medicine, the information has been summed up throughout a long history of medical practice.

1.2.1 Properties and Flavors

Each herb has its own properties and flavors. Generally, there are four typical types of properties, that is, cold, hot, warm or cool [19]. These properties are experientially summarized according to the actions of the herbs on the human body. Herbs that cure or reduce heat syndrome (*Yang* syndrome) have a cold or cool property, whereas herbs that cure or reduce cold syndrome (*Yin* syndrome) have hot or warm property. Cold or cool properties are quite different from warm or hot. People think that cold and cool are similar, and so are warm and hot. The difference is merely the action abilities. Cold is relatively 'stronger' than cool and hot is 'stronger' than warm. Cool- or cold -natured herbs are thought to have the effects of clearing heat, purging fire, removing toxic substances, and nourishing *Yin*, so they are usually used to cure heat syndromes. On the contrary, warm and hot –natured herbs are believed to have the effects of dispersing cold, warming up the interior, supporting *Yang*, and treating collapse, and are therefore used to treat cold syndromes. In addition to the four properties mentioned above, there is the fifth property: mild. A mild-natured herb can be used for either hot or cold syndromes.

Flavor of herbs is given partly by their tastes. Sometimes they are given according to the actions of herbs rather than tastes. Therefore, the flavors of some herbs are often different from their true tastes. There are a total of seven flavors, including pungent, sweet, sour, bitter, salty, tasteless and astringent [19]. The first five ones are the basic. Herbs with different flavors usually show different pharmacological and therapeutic actions, while the same flavor comes out the similar effects. According to *Yin* and *Yang* theory, herbs with pungent, sweet or tasteless flavor have the attribution of *Yang* and the ones with sour, bitter or salty flavor, of *Yin*. The effects of pungent herbs are to disperse exopathogens from the body and promote the circulation of the vital energy and blood. Therefore, these herbs, such as *Herba Ephedrae* (Ma Huang), *Radix Aucklandiae* (Mu Xiang), are often used in treatment of superficial and mild illnesses. The effects of sweet herbs are to nourish, replenish, or enrich the function of the organs, to normalize the function of the stomach and spleen, to harmonize the properties of different herbs and to relieve spasm and pain, etc. They are usually effective in treating syndromes of deficiency type, dry cough, constipation, such as *Radix Codonopsis pilosulae* (Dang Shen), *Radix Rehmanniae Preparata* (Shu Di) and *Radix Glycyrrhizae* (Gan Cao). The effects of sour herbs are to induce astringency and arrest discharge. They are often used to treat sweating. For example, *Fructus Schisandrae* (Wu Wei Zi), a very important adaptogen can be used to regulate body functions and increase the organism's ability to deal with stress. Herbs with bitter flavor have the effects of clearing heat, purging fire, sending down the adverse flow of Qi to treat cough and vomiting, relaxing the bowels, eliminating dampness, etc. Such

herbs are mostly used for syndromes of pathogenic fire, cough with dyspnea, vomiting, constipation due to heat of excess type, damp-heat syndrome, or cold-damp syndrome and other syndromes. Herbs with salty flavor have the effects of relieving constipation by purgation, and softening and resolving hard mass. They are used in the treatment of dry stool and constipation, scrofula, goiter and mass in the abdomen. Two examples are *Concha Arcae* (Wa Leng Zi) and *Natrii Sulfasl* (Mang Xiao). The effects of tasteless herbs are to excrete dampness and induce diuresis. Therefore these herbs, such as *Polyporus umbellatus* (Zhu Ling) and *Poria* (Fu Ling), are often used for edema and dysuria. Astringent herbs have similar actions as those sour herbs, such as *Os Draconis* (Long Gu), *Concha Ostreae* (Mu Li) and *Halloysitum Rubrum* (Chi Shi Zhi).

Table 1. Properties and Associated Effects of Herb

Property	Effects
Cold	Quells fire (anti-inflammatory/spasmodic, sedative).
Cool	Subdues heat (reduces fever, detoxifies, lowers BP).
Mild	Gentle effects (does not alter Hot or Cold conditions).
Warm	Enhances circulation (alleviate chills, improves organ function).
Hot	Dispels Cold (breaks Qi blockage, warms the center).

Note: Cool and cold herbs do overlap, as do hot and warm herbs.

Herbs that possess the same flavors and properties generally have similar effects. But if only one property or flavor is the same, their actions may be quite different. For example, both *Rhizoma Coptidis* (Huang Lian) and *Radix Rehmanniae Preparat*a (Sheng Di Huang) are cold, however the former is bitter and the latter is sweet. It has been well known about the effects of these two herbs. *Rhizoma Coptidis* has the effects of clearing heat and drying dampness and is therefore used for damp-heat syndrome, while *Radix Rehmanniae Preparat*a has the effects of clearing heat and nourishing *Yin* and is used for the condition of consumption of *Yin* due to febrile diseases. Table 1 and Table 2 give the four typical properties and five typical flavors, respectively, and the corresponding effects.

Table 2. Flavors and Associated Effects of herb

Flavor	Effect
Sweet	Nourishing, tonifying.
Pungent	Dispersing, decongesting, stimulating.
Salty	Diuretic, purgative, softening
Sour	Astringent, absorbing, circulation.
Bitter	Sedating, anti-inflammatory/Fire, soothing.

1.2.2 Meridians of Herb

According to Viscera, Channels and Collaterals theory, the symptoms can reflect the organs that are not in good conditions. So ancient herbalist doctors speculated that a herb might selectively act upon a particular part of the body and this part depended on the corresponding symptoms that the herb can relieve. These parts of the body are named meridians. There are twelve types of meridians [19], including Lung, Bladder, Spleen, Large Intestine, Stomach, Small Intestine, Liver, Pericardium, Heart, Kidney, Gallbladder and San Jiao. The Lung is the most delicate and most exterior of all the organs. Diseases often happened because of external pathogenic invasion and accumulation of Lung heat. The Bladder transforms and excretes fluids from the body. It is extremely sensitive to climatic changes, which can cause induced patterns of Cold-Damp and Damp-Heat. The Spleen is the primary source of nourishment of the body because it governs digestion and the production of Qi. It is most susceptible to the evil of Dampness. The Large Intestine is used to receive food from the Small Intestine, absorb fluids and excrete feces. When excessively exposed to cold, External Cold can invade it resulting in abdominal pain and diarrhea. The Stomach has the function to digest and transform food to make it available to the Spleen. The state of the Stomach is governed by the Hot-Cold nature of ingested food in relationship to the patient's constitution and the environment. The Small Intestine receives the transformation products of food and drink from the Stomach and separate the pure from the impure. It is also affected by the heat or coldness of ingested food and drink. The liver is the central organ of the body, which is primarily responsible for the storage of Blood and the

smooth distribution of Qi throughout the body. Exterior Wind can interfere with the smooth flow of Qi and stir up the Blood stored in the Liver to exacerbate Internal Liver Wind and cause skin rashes of sudden onset that move around the body, often seen in viral exanthema, drug rashes and hives. The Pericardium is membranous sac filled with serous fluid that encloses the heart and the roots of the aorta and other large blood vessels. Disease can be caused by cold, heat and Liver Qi stagnation. The heart controls the blood vessels and regulates the flow of blood in the body. The kidney represents the most important energetic organ in the body with regard to the root of energy and the will to live, develop and reproduce. It is the foundation of *Yin* and *Yang* in the body. Fear, anxiety, shock and prolonged stress are the major emotional issues that deplete Kidney energy. The Gallbladder is used to store and secrete bile. The Excessive ingestion of greasy or fatty foods leads to Dampness that becomes lodged in the Gallbladder. The emotions of anger, repressed resentment and frustration cause Heat and Liver and Gallbladder Fire, and when mixed with Dampness, give Damp-heat. Climatic changes of Heat and Dampness from the exterior will induce elements of Damp-Heat in the Gallbladder. San Jiao refers to Shang Jiao, Zhong Jiao and Xia Jiao, which are the pathway through which Qi and Jing Ye ascend or descend. Disease can be caused by Wind-Damp, Cold-Damp and Heat.

By acting on one or more particular meridians, medicinal herbs can regulate the body to a balance status. For example, *Fructus Zizyphi Jujubae* (Da Zao) can tonify Qi in the spleen and stomach. It is indicated for poor appetite and loose stool due to weakness of the spleen and stomach. Judged by the above indications and analysis, we

say that the herb enters the meridians of the spleen and stomach. *Herba Ephedrae* (Ma Huang) can promote sweating, soothe asthma and benefit urination. It is used for fever, chills and absence of sweating due to invasion by exogenous pathogenic wind and cold, dysuria, edema and so on. So people believe that the herbs can enter the lung and bladder meridians.

The meridian theory studies the physiological function and pathological change on the meridians and the related zangfu organs. The essential functions of the meridian system are to transport Qi and blood, to maintain conductivity and to resist invasion of exogenous pathogenic factors. The meridian system distributes to all parts of the body. The endless circulation of Qi and blood in the meridians is responsible for the maintenance of life and the variety of functions which support it. The meridian theory has been the guiding principle for the clinical practice in the realms of TCM, particularly in those of acupuncture, massage and Qi Gong. By combining the meridian theory with the theories of the zangfu organs and the etiology of TCM, one can thoroughly explain both the physiological activities and pathological changes, which take place in the body. In this manner, a theoretical basis for the principle of treatment in accordance with the differentiation of symptoms and signs was established.

1.2.3 Toxicity and Nontoxicity

Traditionally, people, especially those in Chinese communities, believed that medicinal herbs were weakly toxic or even non-toxic. However, some herbs are toxic

and should also be used carefully. According the difference of their toxicities, they are classified to three classes: toxic, extremely toxic and slightly toxic [19]. Improper use of the toxic herbs may lead to adverse effects, so prescription of toxic herbs should be careful according to the patient's age and the situation of the disease, etc. Nontoxic herbs are moderate in nature and usually do not have any side effects. For example, *Fructus Zizyphi Jujubae* (Da Zao) and *Poria* (Fu Ling) are nontoxic herbs, while *Radix Aconiti lateralis* (Fu Zi) and *Semen Strychni* (Ma Qian Zi) are toxic.

Some toxic herbs are effective because of the other effects on the patient, so when used in combination of other herbs in a prescription, the toxicities are expected to be eliminated or lessened by means of processing, dispensing and preparation. On the contrary, the medical effects of some toxic herbs, on the critical or obstinate diseases, are due directly to their toxicity properties.

1.3. TCM Formulae

1.3.1 Compatibility of Herbs

Two or more herbs are often combined in order to increase or promote their therapeutic effectiveness, to minimize toxicities or side effects, to accommodate complex clinical situations and to alter their actions. Different combinations can cause diverse therapeutic effects. Using only one herb in a formula is called "Single herb formula". The combination of two or more herbs is known as "mutual reinforcement, mutual assistance, mutual restraint, mutual counteraction, mutual suppression and

mutual antagonism" [20]. The details for combining herbs are classified as follows.

1. Single herb formula. The whole formula is composed of only one herb. This formula is used when the disease is not very serious. For example, Formula 'Du Yi Wei Wan', which only contains *Radix Lamiophlomidis rotatae* (Du Yi Wei), is used in treatment of pain caused by traumatic injury, sprain, rheumatism, contusion of muscles, joints and loins, fracture, surgical injury and rheumarthritis.

2. Mutual reinforcement. Two or more herbs with similar properties are combined to reinforce their therapeutic actions. For example, *Radix et Rhizoma Rhei* (Da Huang) is combined with *Natrii Sulfasl* (Mang Xiao) to reinforce the function of purging downward; *Gypsum Fibrosum* (Shi Gao) and *Rhizoma Anemarrhenae* (Zhi Mu) are used together to clear heat and subdue fire.

3. Mutual assistance. Two or more herbs in which one or more is the principle herb and the others play a subsidiary role are combined. For example, *Radix Astragali* (Huang Qi) is combined with *Poria* (Fu Ling) to replenishe Qi, strengthen the spleen and promote urination; *Gypsum Fibrosum* (Shi Gao) is combined with *Radix Achyranthis bidentatae* (Niu Xi) to relieve toothache.

4. Mutual restraint. In the formula, the toxicity or side effects of one herb can be reduced or eliminated by the addition of another. For example, the toxicity of *Rhizoma Pinelliae* (Ban Xia) can be counteracted or restrained by *Rhizoma Zingiberis recens* (Sheng Jiang).

5. Mutual counteraction. It means that one herb in the formula can counteract or

restrain the other herb's toxicities and side effects. In fact, mutual restraint and mutual counteraction refer to the same effect of herb from two different aspects. The former is focused on how to restrain the toxicity of one given herb, while the latter pays attention to the counteraction effects of one herb on the other.

6. Mutual suppression. In the formula, one of the herbs weakens or suppresses the action of the others. For example, *Semen Raphani* (Lai Fu Zi), combined with *Radix Ginseng* (Ren Shen), weakens the function of the latter in replenishing Qi.

7. Mutual antagonism. When two herbs that have no side effects are combined together, severe side effects may result. Traditionally, there existed "eighteen incompatible medicinal herbs" and "nineteen mutual restraining medicinal herbs".

When two herbs are combined together in one formula, they may give rise to interaction with each other. They may be mutual reinforcement, mutual assistance, mutual restraint, mutual counteraction, mutual suppression and mutual antagonism. For this reason, combination should be carefully considered according to the conditions of the patient, and their characters and functions. "Mutual reinforcement" and "mutual assistance" can enhance their effects and therefore should be used as much as possible. "Mutual restraint" and "mutual counteraction" can reduce or eliminate toxicities and side effects of herbs and therefore can be considered when using poisonous herbs. "Mutual suppression" and "mutual antagonism" can weaken efficacy of herbs or make them lose their efficacy or even give rise to toxicities and side effects and therefore should be avoided.

1.3.2 Precautions and Contraindications

The side effects of TCM formula can be reduced by adding one or two herbal ingredients. But sometimes it would be harmful to the patients [20]. So the prescription should be carefully given out according to condition of the patients. Four contraindications should be noticed as following:

1. Contra-syndromes. Each herb or each class of herbs has its own intended functions and effects. Diseases or syndromes other than these intended indications are contra-syndromes or contraindications. For example, *Herba Ephedrae* (Ma Huang) is used to induce diaphoresis and relieve asthma, and its indications are affection by exopathogenic wind-cold, anhidrosis due to exterior syndrome of excess type and cough due to obstruction of the lung Qi, but in case of spontaneous sweating due to exterior syndrome of deficiency type or cough due to lung deficiency, its use should be prohibited.

2. Incompatibility of herbs. Some herbs cannot be used in combination with specific herbs. According to the ancient books and literature of TCM, there are eighteen pairs of incompatible medicinal herbs and nineteen pairs of mutual restraining medicinal herbs.

3. Contraindication for pregnancy. Some herbs cannot be used by pregnant patients. These herbs are mainly with strong actions or toxicities, especially *Fructus Crotonis* (Ba Dou), *Semen Pharbitidis* (Qian Niu Zi), *Herba Cirsii japonici* (Da Ji) and *Rhizoma Sparganii* (San Leng). The other herbs that are pungent and hot should also be used

with caution during pregnancy, such as *Semen Persicae* (Tao Ren), *Flos Carthami* (Hong Hua), *Radix et Rhizoma Rhei* (Da Huang) and *Radix Aconiti lateralis* (Fu Zi).

4. Food taboo. Certain types of food may also influence the action of herbs or induce unnecessary side effects. In general, it is advisable not to eat raw, cold, greasy, strong smelling or spicy food while taking medicine. From the historical medical literature, *Radix Dichroae* (Chang Shan) is contraindicated with onion; *Radix Rehmanniae* (Di Huang) and *Radix Polygoni multiflori* (He Shou Wu) are contraindicated with onion, garlic and turnip; *Herba Menthae* (Bo He) is contraindicated with turtle meat; *Poria* (Fu Ling) is contraindicated with vinegar; and *Carapax Trionycis* (Bie Jia) is contraindicated with three-colored amaranth.

Each herb has its own functions and indications. The herbs should be selected according to their properties, flavors, meridians and toxicity properties. For example, *Herba Ephedrae* (Ma Huang) is pungent, slightly bitter and warm. It is used to treat fever, chills and absence of sweating due to an invasion of exogenous wind and cold. It is contraindicated in deficiency exterior syndrome with the symptoms mentioned above.

In one formula, the different dosage of each herbal component is prescribed according to their effects [20]. Generally speaking, the dosage of the principal component is bigger than the others. When this formula is used for the treatment of patients with different conditions, the dosage of these herbs may be changed in accordance with specific syndrome. In addition, the dosage of those herbs that are very

drastic in nature or extremely poisonous should be strictly controlled to prevent poisoning.

1.4 Methods for Studying TCM

1.4.1 Theory and Practices of TCM

TCM practice is believed to be based on the principle of maintaining the balance of the body. Illness or disease always represents a state of imbalance of *Yin* and *Yang*. We could consider this concept as the need to find the most appropriate and stable chaotic state, with a strong attractor. Over time, TCM has elucidated methods of restoring balance to the body, thus curing disease. The two important methods are herbal therapy and acupuncture. Although various kinds of imbalances can be treated by either method, herbal medicine is excellent at tonification, that is, increasing the energy of a particular organ, and balancing the energy when there is excess in one part of the body and deficiency in another. Of course, in integrated medicine we can also compile acupuncture and herbal medicines together – acupuncture for the acute problem, and herbs to shore up the underlying deficiency that allowed the acute disease to manifest.

Ancient Chinese herbalist doctors have developed their own language and method for diagnosis. The diagnosis is not directly associated with a disease, but the description of a pattern that seems to be preeminent in the whole body, which includes Looking, Smelling, Asking, Pulse diagnosis, etc [21]. The Chinese physical exam is able to show a predominant pattern of imbalance in the patient, which, in the Chinese physician's eyes, is the disease to be treated. Many dissimilar western diagnoses have

similar Chinese patterns of disharmony, and many similar Western diagnoses have different Chinese patterns of disharmony. In Chinese medicine there have been periods in which one theory of harmony and disharmony dominated over another. Thus, during one period, the Five Element Theory of disharmony reigned supreme and during another period, the Yin-Yang and Eight Principles Theory of disharmony reigned supreme. In this paradigm of healing, the western diagnosis is unimportant, as we are treating the patient's explicit response to his or her inner disturbance, which is going to be different for each patient according to his or her initial condition.

1.4.2 Modern Experimental Approach and Clinical Trials for Studying TCM

From the point of view of modern medicine, the effects of herbal drugs result from the interaction between herbal ingredients and organs of body [2,5]. It is necessary to extract these chemicals and then analyze their physiological effects on humans or their action on disease. So far, a considerable body of information on the chemical composition of medicinal herbs, has been found by researchers and institutes, especially those in China and Japan. Many purified compounds are then studied or subjected to be studied *in vitro* or on animal models [22]. For example, *Radix Glycyrrhizae* (Gan Cao) is one of the most extensively studies Chinese herbs [23-28]. The function of *Radix Glycyrrhizae* is to reinforce the function of the spleen and replenish Qi, to remove heat and counteract toxicity, to dispel phlegm and relieve cough, to alleviate spasmodic pain, and to moderate drug actions. It is used in treatment of weakness of the spleen and the stomach marked by lassitude and weakness, cardiac palpitation and shortness of breath, cough with much phlegm,

spasmodic pain in the epigastrium, abdomen and limbs, carbuncles and sores. It is often used for reducing the toxic or drastic actions of other drugs. Its main ingredients are Glycyrrhizin and glycyrrhetinic acid [25,26]. Experiments show that its extract exerts a detoxifying effect [26,27]. Another experiment shows that glycyrrhizin also has this effect [26,27]. In addition, Glycyrrhizin has anti-inflammatory effect similar to hydrocortisone [28]. It was found that its decoction can inhibit intestinal smooth muscles of rabbits in vitro, but glycyrrhizin and glycyrrhetinic acid have no such effects [29]. Thus the effects must be due to the other compounds. *Radix Ginseng* (Ren Shen) being under exploration [30,31], contains more than 200 compounds. The main ingredients are ginsenosides, essential oils and polyacetylenes [30]. *Radix Ginseng* traditionally has been used as a tonic and was thought 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. Experiments have shown that *Radix Ginseng* can act on the central nervous system and exert significant cardiotonic and hypertensive effects on acute circulatory failure cardiotonic and hypertensive effects on acute circulatory failure after heavy blood loss and decrease the level of blood sugar and promote phagocytosis and enhancing lymphocyte-blastogenesis rate [29]. These effects are related to the active compounds. For example, ginsenoside Rg1 can raise blood pressure, while Rb1 can lower it [31,32].

As indicated in the above examples, chemical and biological studies underlined the importance of using both together in the standardization and modernization of TCM.

The chemical profiles of an extract should be matched against the results of biological assays in order to assure lot-to-lot consistency of what are fairly crude extracts of herbs with high natural variability [1]. Chemical analysis, biological assays and animal experiments provide important fundamental information about TCM. In the context of modern biomedical research, they should also be necessary prerequisites for clinical trials. Most of these TCM clinical trials have been undertaken in China, Taiwan, and Japan. A number of commercial treatments based on TCM such as 'Hu Xin Dan", "Puerarin injection', 'Shen Mai injection' and others have been introduced into the market after extensive pharmacological research and clinical trials [33,34].

1.4.3 Computational Methods

Conventional approaches, including chemical standardization, biological assays, animal models and clinical trials, have been extensively used to study the effects and molecular mechanism of medicinal herbs and formulae. More recently, as useful tools, computer-aided methods are rapidly developed to facilitate the understanding of TCM. To control and evaluate the quality of medicinal herbs, the fingerprint analysis technique is set up with the rapid development of instrument analyses and computer pattern interpretation, which includes fuzzy information analysis, artificial neural networks and gray relational grade cluster [35-37]. To identify similar medicinal herbs, computer-aided classification based on back propagation algorithm of artificial neural network pattern recognition is developed [38].

Employing a fuzzy clustering method, Su Weiwei *et al* divided the herbs in one formula into several different parts [39-41]. The results agreed well with the

explanation by TCM theories. Moreover, it provided a feasible computer-aided method to study the formulation of prescription. Aiming at this problem, in this study, we proposed a new method based on Support Vector Machine (SVM) as discussed in Chapter 3 to evaluate the effectiveness of a formula and to generate new potential formulae.

To understand the molecular mechanism and pharmacology of bioactive compounds from Chinese medicinal herbs, which is important in facilitating scientific evaluation of novel therapeutic approaches in traditional Chinese medicine, a newly developed computer software INVDOCK [42] is used for automated identification of potential therapeutic and toxicity targets of bioactive compounds isolated from Chinese medicinal herbs. This method may potentially be used as a relatively fast-speed and low-cost tool for facilitating the study of molecular mechanism and pharmacology of bioactive compounds from Chinese medicinal herbs. And it is also of significance in new drug development based on the mechanism of Chinese medicine.

1.5 Specific Aims of the Project

1.5.1 To Develop a TCM Database

Due to the difficulties in accessing and understanding non-English medicinal records, most scientists are not familiar with traditional Chinese medicine [1]. Moreover, because the TCM theory is very different from that of western medicine, majority of these scientists show little interest in TCM and other traditional medicines.

To attract their attention and help them better understand the merits of TCM, it is necessary to provide the information of TCM first. Hence, a new database, Traditional Chinese Medicine Information Database (TCMID), is developed, to provide, in an integrated manner, comprehensive information about all aspects of TCM, including prescription formulae, constituent herbs, known herbal ingredients and their molecular actions. The therapeutic effects, adverse reactions, clinical indications and applications are provided at the levels of prescription, individual herb and herbal ingredient respectively. Traditional terminologies about medicinal herbs such as properties, flavors, sites of actions, toxicity level, and therapeutic class are also given.

1.5.2 To Develop a Computer-aided Method for Prescription Formulation

In recent years, a great deal of pharmacological research has been undertaken to review and establish reliable composite formulae of TCM [1]. To facilitate such a study, an alternative approach is introduced, based on a computer method, support vector machines (SVM), which is a machine learning algorithm currently considered to be one of the most efficient method in many real-world applications [43]. Recently SVM have been applied to a number of biological problems [44], including gene expression data analysis and protein classification, etc. Using this method, it is expected to find the principle of prescription formulation and then to evaluate the effectiveness of a given formula and even generate new formulae.

1.5.3 To Explore the Molecular Mechanism of Medicinal Herb

The mechanism of action of TCM remains largely unknown, though plenty of

active compounds have been isolated from these herbs and clinical and therapeutic effects are found. This study is proposed to explore the mechanism by a computer-aided method, INVDOCK. In this study, I analyzed the mechanism of *Serenoa repens,* which is widely used in treatment of BPH [45-49]. Although the effect of some active compounds extracted from it has been proven by *in vivo* or *in vitro* experiments, the whole picture is still unknown. The multiple ingredients make the process more difficult. Hopefully, the effects of the predicted therapeutic and toxicity targets will provide us clues of the molecular mechanism of this herb.

Chapter 2: TCM Database Development

2.1 Introduction

Traditional Chinese medicine (TCM) has been used in the treatment of a variety of diseases and is recognized as a valuable alternative to conventional medicine [1,2,16,50]. A major therapeutic approach of TCM is the use of a mixture of herbs, each composed of a number of constituent chemicals, which collectively exert therapeutic actions and modulation of other factors. One or more principal constituents provide main therapeutic actions. Certain secondary principal constituents enhance or assist the effect of the principle ones. The rest serves modulation roles such as treatment of accompanying symptoms, moderation of harshness and toxicity, enhancement of drug delivery, and harmonization etc [1].

Because of growing interest in TCM therapeutics, increasing effort has been made towards scientific proof, clinical evaluation and molecular study of TCM [1,2,42]. Such an effort can be facilitated by making available information about all major aspects of TCM including herbal formulations, herbal composition, chemical composition, molecular structure and functional properties, therapeutic and toxicity effects, clinical indication and application. A few specialized databases have appeared which provide information about different aspects of TCM. For instance, The TCMD

provides information about Chinese medicinal plants and constituent chemicals [51], The Chinese Medicine Sampler (http://www.chinesemedicinesampler.com/) and Traditional Chinese Medicine (http://www.healthy.net/CLINIC/therapy/Chinmed/Index.asp) databases provide general information about the history, theory, diagnostics, and examples of TCM formulae.

However, there is a lack of a database that provides comprehensive information about all major aspects of TCM. Moreover, in the existing databases, information about the composition of constituent chemical compounds in TCM herbs is not included. A new database, Traditional Chinese Medicine Information Database TCMID, is introduced to provide, in an integrated manner, comprehensive information about all aspects of TCM including prescription formulae, constituent herbs, known herbal ingredients and their molecular actions. The therapeutic effects, adverse reactions, clinical indications and applications are provided at the levels of prescription, individual herb and herbal ingredient respectively. Traditional terminologies about medicinal herbs such as tastes, flavors, sites of actions, toxicity level, and therapeutic class are also given.

2.2 Database Development Method

In this work, Oracle 9i is selected as the platform for our database projects and data future analyzing tasks. Oracle 9i is a typical database management software, which has the most variety of modules and development tools, including modules for data mining

and online analytical process, which is essential for high-end data analysis purpose. For many years, it leads the way in indexing and query optimization technology while it is not worse than its competitors in other aspects. Also, Oracle 9i is a fully object-oriented database, which conforms to the world trend toward object, oriented programming (OOP). Oracle has kept the biggest market share for years. According to 2002 statistics (http://www.oracle.com/tellmemore/?1333278), over half of the fortune 100 corporations use Oracle as their database server.

Oracle 9i is based on a relational approach, which means that database developer can create a set of tables (also known as relations) to reflect the inter-relationship between the data stored in the database. The relational approach needs the support of a relational database management system (RDBMS). The RDBMS enables users to create and maintain a relational database. They are designed to control data redundancy, to restrict unauthorized access, to provide persistent storage for program objects and data structures, to permit inference and actions using rules, to provide multiple user interfaces, to representing complex relationships among data, to enforce integrity constraints, and to provide backup and recovery.

2.3 Database Structure and Access

2.3.1 Database and Source of Data

TCMID is freely accessible at http://xin.cz3.nus.edu.sg/group/tcmid/tcmid.asp. The information collected in TCMID is from a search of available literature [29,52-62]. Efforts are made to collect as many of these known formulae, herbs, herbal

constituents and compositions, and other functional information as possible. Intensive research into TCM and TCM herbal constituents are leading to the discovery of new knowledge about different aspects of TCM. It is hoped that new or existing information missing in this database can be gathered by users' submission to a special page in the database as well as by conducting up-dated literature search.

The database currently contains entries for 1343 formulae, 1325 medicinal herbs, 4468 herbal ingredients and description about molecular actions of 502 herbal ingredients. Each entry can be retrieved via the name, alternative name, function or clinical manifestation at the levels of prescription formula, herb and herbal ingredient respectively. This database will be continually updated as new data are found. And the friendly interface will be upgraded as well to meet the user's needs adequately.

2.3.2 Database Access

The TCMID database web interface is shown in Figure 1. Users can easily search the database from three starting points, namely, Formula, Herb and Compound. 'Formula' starting point generates a new interface through which the database can be accessed by keyword search such as Formula Chinese Names, Formula English Names, Formula Functions, Formula Clinical Manifestations, etc. 'Herb' starting point generates another interface to access the herbal information through keyword search such as Herb Latin Names, Herb Chinese Names, Herbal Functions, Herbal Clinical Manifestations, etc. 'Compound' starting point generates the third interface to access the chemical information through keyword search such as Compound Name, CAS No,

etc. Searches involving any combination of these search or selection fields are also supported.

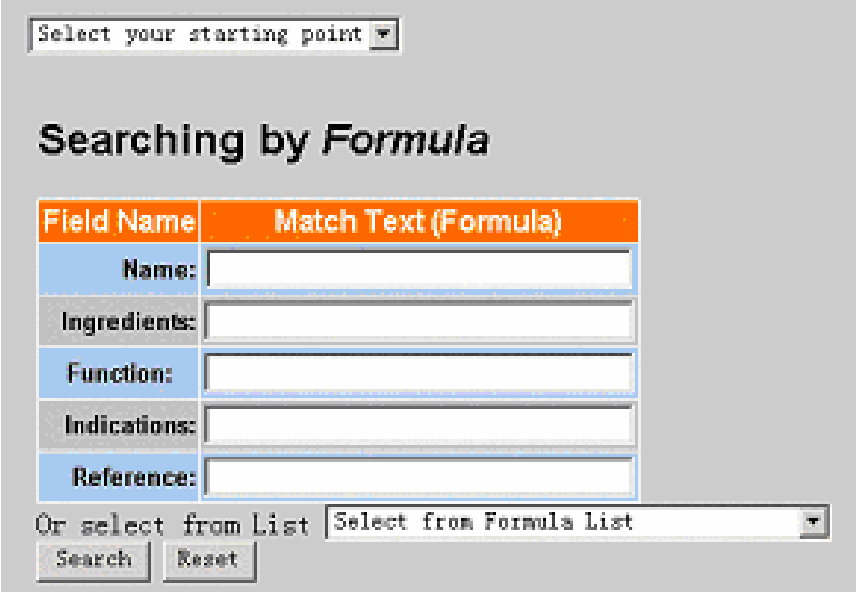


Figure 1. The query interface of TCMID

The full text search is case insensitive and wildcards supported. In a query, a user can specify full name or any part of the name in a text field. Wild character of '%' and '_' is allowed in text field. Here, '_' represents any one character and '%' represents a string of characters of any length. For example, input of 'Ma Huang' in the Formula Chinese Name field finds entries containing 'Ma Huang' in their name, such as Ma Huang Tang or She Gan Ma Huang Tang. On the other hand, input of 'Ma% Tang' finds all the entries with 'Ma' and 'Tang' in their names, such as Ma Huang Tang.

The typical query results are illustrated in Figure 2, 3 and 4. Figure 2 gives details about the formula, including Chinese Name, Common Name, Literature

Reference, Herbal Components, Function, Clinical Manifestation, Procedure, Description, Usage and Dosage, Storage, Precaution and Modifications. From this page, the herbal information can be obtained by clicking each individual herbal component.

TCMID Test Result (Formulae Info)	
Chinese Name	Ge Jie Ding Chuan Wan
Common Name	Gejie Dingchuan Pills, Asthma-Relieving Bolus of Gecko
Herbal Components	Gecko: 11 g; Semen Trichosanthis: 50 g; Radix Asteris: 75 g; Herba Ephedrae: 45 g; Carapax Trionycis: (processed with Vinegar) 50 g; Radix Scutellariae: 50 g; Radix Glycyrrhizae: 50 g; Radix Ophiopogonis: 50 g; Rhizoma Coptidis: 30 g; Bulbus Lilli: 75 g; Fructus Perillae: (stir-fried) 25 g; Gypsum Fibrosum: 25 g; Semen Armeniacae Amarum: (stir-fried) 50 g; Gypsum Fibrosum: (calcined) 25 g.
Procedure	Pulverize the above fourteen ingredients to fine powder, sift and mix well. To each 100 g of the powder add 70-100 g of refined honey to make small honeyed pills or big honeyed pills.
Description	Blackish-brown small honeyed pills or big honeyed pills. Odour,slight;taste,bitter and sweet.
Function	To nourish yin and cool lung, relieving cough and asthma.
Clinical Manifestations	Chronic cough of consumptive disease, asthma in the aged, short breathing, fever, fullness of chest, perspiration, night sweat, anorexia.
Usage & Dose	9 g of small honeyed pills or 1 big honeyed pill, 2 times a day.
Storage	Preserve in tightly closed containers.
Specification	9 g per 60 small honeyed pills, 9 g per big honeyed pill.

Figure 2. The typical query result about formula

TCMID Test Result	
Chinese Name	A wei
Latin Name	Resina Ferulae,
English Name	Chinese Asafetida
Plant Name	Ferula sinkiangensi K. M. Shen;Ferula fukanensis K. M. Shen
Properties and Flavors	Warm,Pungent,Bitter
Meridians	Spleen,Stomach,Liver
Toxicity	Non toxic
Collection	Chinese Asafetida is the resin of Ferula sinkiangensi K. M. Shen or Ferula fukanensis K. M. Shen (Fam. Umbelliferae). The stem is incised from the upper part to the lower part at the flowering and early fruiting stages in the end of spring and early summer. The exuding emulsive resin is collected, and dried in the shade.
Description	Irregular lumps or resinous substance. Colour varying in intensity, externally waxy-yellow to brownish-yellow. Lumps light, texture waxy; fracture somewhat pored; freshly cut surface pale, gradually darken on storage. Resinous substance viscid, greyish-white. odour, strongly and lastingly characteristic and alliaceous; taste, pungent, with the burning sensation on chewing.
Function	To remove food stagnancy. To disintegrate masses. To kill worms.
Indication	To relieve symptoms of stagnation of undigested meat, mass in the abdomen due to blood stasis, abdominal pain due to intestinal parasitonis
Therapeutic Class	Food digestion
Storage	Preserve in well-closed containers, stored in a cool and dry place.
Usage & Dosage	1-1.5 g.
Chemical Ingredients	alpha-pinene, beta-caryophy, lene, beta-eudesmol, beta-myrcene, beta-pinene, camphene, delta(3)-carene, gamma-terpinene, limonene, myrcene, p-cymene

Figure 3. The typical query result about herb

TCMID Test Result	
Name	alpha-pinene,
CAS RN	2437-95-8.
Molecular Formular	C10H16
Molecular Weight	136.2360
Herbs	A Wei, Ai Ye, Bai Zhi
Structure	Download
Proteins inhibited/antagonized	Cytochrome P450 CYP2B1
Proteins activated/agonized	Sensory irritation receptor; Increased androstenedione 16alpha-hydroxylase activity; Increased aniline hydroxylase activity
Enzymes with compound as product	(+)-bornyl diphosphate synthase; 1,8-cineole synthase
Other molecular events	Induction of cytochrome P450s CYP2E1, CYP2C11, CYP2C6 and CYP6B7; Induction of cytochrome b5 mRNAs

Figure 4. The typical query result about compound

Figure 3 gives the details about the herb, including Chinese Name, Latin Name, English Name, Plant Source, Properties and Flavors, Meridians, Toxicity, Chemical Ingredients, Therapeutic Class, Therapeutic Sub-Class, Function, Indications, Clinical Manifestations, Collection, Processing, Description, Usage and Dosage, Storage and Literature Reference. The chemical ingredient can link to the corresponding page containing detail chemical information. They include Chemical Name, CAS No, molecular weight, molecular formula, downloadable structure file and so on (Figure 4).

2.4 Data Submission and Updating

This database is to be updated quarterly, information regarding newly searched TCM formulae, herbs and compounds and additional knowledge them will be added. To facilitate submission of new information concerning TCM by other users, a data submission interface is included in the database (Figure 5). Submitted information will be integrated into the database after validation.

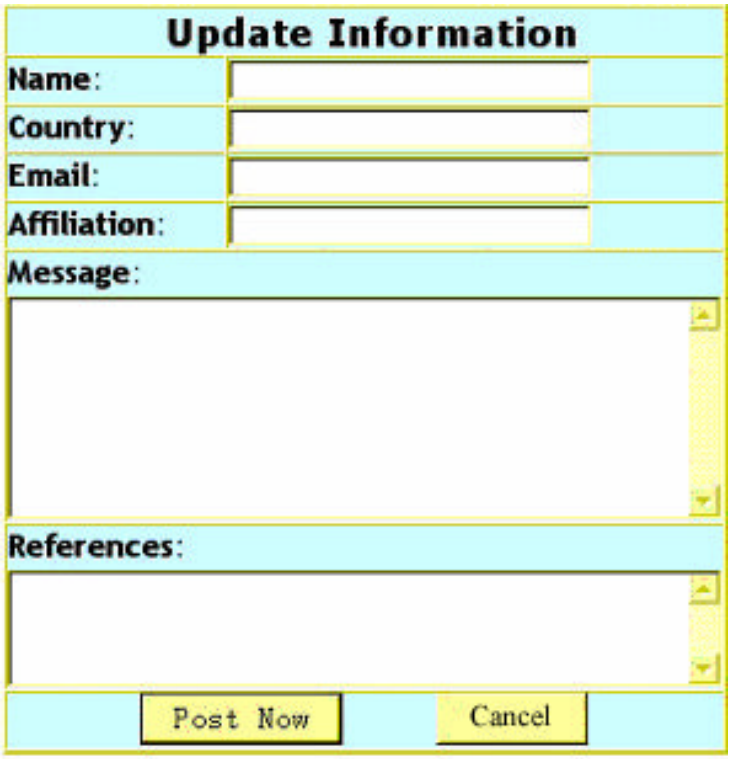


Figure 5. The data submission interface

2.5 Preliminary Analysis of Database

This database currently contains 1343 entries of formulae, 1325 entries of herbs found from the literature along with 4468 entries of compounds which are extracted from the herbs. Among these entries of formulae, 1190 entries are relatively complete. Their herbal components, functions, clinical manifestations are collected into the

database. However, the others lack one or more detail features, at least the information are not found by now.

According to their therapeutic effects, the formulae are classified into 17 Classes, including formulae to release the exterior, purgative formulae, mediation formulae, formulae for clearing heat, formulae to dispel cold, tonic formulae, astringent formulae, formulae for tranquilizing the mind, formulae to promote resuscitation, Qi formulae, blood formulae, formulae for treating wind, formulae for treating dryness, formulae for dispelling dampness, formulae to expel phlegm, digestive formulae, and formulae for dispelling parasitic worms. For example, formula *Xin Jia Xiang Ru Yin* is used to treat exterior syndromes due to invasion by summer-heat, dampness and cold. It was classified under 'formula to release the exterior' in the database.

Among these entries of herbs, 1090 entries are relatively complete. Their functions, clinical manifestations are collected into the database. One or more detail features should be added to the other entries when updating.

According to their therapeutic effects, the herbs are classified into 23 Classes, including anaesthesia, anthelmintic, antitumor, astriction, bleeding control, blood activation and stasis removal, food digestion, for calming liver and containing wind, for dissolving dampness by flavors, for dissolving phlegm, stopping cough, and soothing breathing, for dispelling wind-dampness, for eliminating toxic materials, dissolving rottenness and growing new muscles, for promoting diuresis and penetrating dampness, for relieving exterior syndrome, for tonifying weakness, for

warming interior, heat clearance, mind opening, purging, regulation of Qi, spirit calming, toxication reduction, anthelmintic, dampness removal, and itching control, and vomiting promotion. In addition, each class is divided into several sub-classes. These classifications will facilitate the researchers in collecting and comparing the herbs which have the similar effects. For example, *Semen Lablab album* (Bai Bian Dou) is used in treatment of weakness of the spleen and stomach with loss of appetite and loose bowels, excessive leukorrhea, vomiting, diarrhea, distress in the chest and distension in the abdomen caused by summer-damp. In the database, it belongs to the Class 'for tonifying weakness' and the sub-class 'for tonifying Qi'.

After further investigation, we found that some herb entries are the same as or very similar with the other ones. For example, the database contains 'Ba Yue Zha' and 'Yu Zhi Zi'. These are two different Chinese name of the same herb. A link is provided in each entry to that of the same herb under different name. Thus, user can get sufficient information about the herb even if he only knows one name of this herb. Most of the compounds have the downloadable structure files, which can facilitate the studies of compound-biomolecule interactions and drug design.

2.6 Conclusion and Further Development

TCMID is developed from available information about formulae, herbs and ingredients of herbs in the literature and books, which is a result of collective and persistent effort over the years. It provides a new platform to facilitate mechanistic

study and clinical evaluation of TCM. In contrast to traditional literature and books, the database has the following advantages: reduced data redundancy, data consistency, easier use of data and less storage for larger information. In addition, new information can be easily incorporated or the corresponding new databases can be cross-linked to TCMID to provide more details about TCM.

Chapter 3: Development of a Computer-aided Method for TCM Prescription Formulation

3.1 Introduction

3.1.1 The Principle of TCM Prescription Formulation

A formula of TCM is composed of one or more Chinese medicinal herbs. It is not a simple arbitrary mixture, but a carefully assembled combination of herbal components [1,2,16,50]. Each component in a formula has its particular functions. Moreover, they are mixed together in proper amounts in accordance with specific principles for combining herbs. When giving a prescription, herbalist doctor must take into account the characters of each herbal component in the formula, as well as the stage of the disease and the condition of the patient. According to the different functions of these ingredients, they are classified to four categories [15], in order of importance, *Jun*, *Chen*, *Zuo*, *Shi*.

Jun (King) are one or two herbal ingredients that give the formula its basic characteristic actions. They may not be the greatest percentage of the formula, but the most important clinical herbal components. They play the leading role in the treatment

of the disease. *Chen (Minister)* are the ingredients that strengthen the *Jun*'s action aiming at the principle symptoms. They enhance, augment or broaden the effect of the *Jun*. They also play important role in relieving the secondary symptoms. *Zuo (Assistant)* are these parts that have different effects in different formula. Firstly, they can help the *Jun* and *Chen* components in their activities and can also address secondary symptoms, especially the less important symptoms. Secondly, they will counteract or attenuate any side effects of the first two combinations. Thirdly, they will modify the energy of the formula closer to neutral, when more potential *Jun* and *Chen* components are used in treatment of serious diseases. *Shi (ambassador)* are these parts enable all of the components to be well absorbed, transported and then reach the right place where they will take effects, such as liver, heart and stomach. They may also be used as a dispensing agent or give flavor or texture.

The components in a formula are in close cooperation with one another when exerting their due effects. Theoretically the formulae that have proven effective are precisely formulated in accordance with their definite aim in treatment the disease. For example, *Ma Huang Tang (Decoction of Ephedra*), which is composed of *Herba Ephedrae* (Ma Huang) (9g), *Ramulus Cinnamomi* (Gui zhi) (6g), *Semen Armeniacae amarum* (Ku Xing Ren) (9g), *Radix Glycyrrhizae Preparata* (Zhi Gan Cao) (6g), has the main function to promote sweating and release the exterior. In this condition, the lung is invaded by *cold*; so it is necessary to use herbs with pungent flavor and warm property to expel pathogenic factors from the lung. Therefore, *Herba Ephedrae* is used as the *Jun* ingredient to induce perspiration for dispelling pathogenic wind and cold

from the body and to relieve asthmatic cough by ventilating the lung and smoothing the flow of vital energy; *Gui zhi* is the *Chen* ingredient that aids *Herba Ephedrae* in diaphoresis, warms the channels and stimulates menstrual discharge, reinforces yang, relieves palpitation, and promotes the descending of Qi; *Semen Armeniacae amarum* serves as the *Zuo* ingredient to relieve cough and asthma by ventilating the lung, and to relax bowels; and *Radix Glycyrrhizae Preparata* serves as the *Shi* ingredient to harmonize the action of all the other ingredients and direct them to their proper channels, to reinforce *Qi* and promote blood circulation. Each formula must contain a *Jun* ingredient. However, some formulae may lack one or more of the other categories. For example, *Ren Shen Tang* (Decoction of Ginseng) has only one ingredient: *Radix Ginseng* (Ren Shen). Its function is 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. In some formulae, the *Jun* or *Chen* ingredient may also serve as the *Zuo* or *Shi* ingredient. To prescribe an effective formula, the doctor should have good understanding of the characters of hundreds of herbs.

3.1.2 Modification of TCM Prescription

The composition of a formula is not static. Patients suffering from the same disease may be given different variations of the same formula. Even the same person, during different disease stage, is given modified formula. Factors such as the disease condition, the constitution and age of the patient, the season, and the geographical environment should be considered when composing a formula [20]. A very famous

doctor of Qing Dynasty, Xu Lingtai, said in his book [63] that "before trying to use any old handed-down formula, you must first determine whether all the symptoms and signs of the patient you are treating are consistent with those described in that prescription and whether every ingredient is good for the case. If not you must make the proper modifications or, if it is impossible to do so, choose another one." From what he said, we can see that the formula components are sometimes modified. The *Jun* ingredient must remain the same in the modified formula, but other ingredients are added or deleted according to the signs and symptoms present and the stage of the disease. For example, Formula *Ke Xue Fang* has the effects of clearing fire and resolving phlegmand. But if the condition is associated with severe deficiency of *Yin* and blood, *Radix Ophiopogonis* (Mai Men Dong), *Radix Glehniae* (Sha Shen) and *Radix Trichosanthis* (Tian Hua Fen) should be added. If the condition is associated with severe cough, *Radix Asteris* (Zi Wan), *Flos Farfarae* (Kuan Dong Hua) and *Semen Armeniacae* (Xing Ren) should be added. If the condition is associated with Cough with profuse blood, Bai Bu should be added. Another example is Formula *Huo Xiang Zheng Qi San*, which is to release the exterior and resolve dampness. If the condition is associated with Indigestion, *Massa Fermentata medicinalis* (Shen Qu) and *Semen Raphani* (Lai Fu Zi) should be added. While if the condition is no chill, fever or headache, *Radix Angelicae dahuricae* (Bai Zhi) and *Folium Perillae* (Zi Su Ye) should be deleted. The examples show that it is not always necessary to follow the set formula. On the contrary, proper variations should be made. Similarly, it is not hard to understand that their relative amounts are also varied according to the patient's

condition. Prescriptions consisting of the same substances in different proportions may have different effects. Besides, different forms of the same prescription may have different level of intensities or actions. The major forms of the prescriptions include decoctions, powders, pills, capsules, and honey boluses. Usually in bolus form, the action is slow. But when a decoction of the same ingredients is used instead, the action becomes quicker [20]. Therefore the decoction form is more appropriate than the bolus in serious or critical conditions.

3.1.3 Previous Studies on Prescription Formulation

In TCM, prescriptions, which are composed of various herbs according to the principles, are often considered as the major means of treating diseases. Since ancient times, TCM physicians have summarized their experiences in this field. Many effective prescriptions have thus been tested and documented. However, people were disappointed to find that there was no direct and reliable approach to prescribe new formula toward particular disease. What they can do and are doing is to try to modify and improve the existed formulae. Moreover, how the formula may be modified is also depended on their special experiences and the constitution of the patients. Fortunately, scientists have made progress in developing alternative approach to facilitate the study of the formulation of new prescription. Employing a fuzzy clustering method, Su Weiwei *et al* divided the herbs in one formula into several different parts according to their characters [39-41]. These characters were quantified to generate the feature vectors that can be recognized by computer. From the first chapter, we notice that

different herbs have different characters and functions. That is why different herbs are used for treatment of different diseases. To understand the mechanism of herbs and use them properly, it is necessary to explore their characters. Ancient experts have, from the TCM viewpoint, summarized those basic characters, which include drugs' Properties and Flavors, Meridians and Toxicity, etc. Based on the theories of *Yin* and *Yang*, Viscera, Channels and Collaterals, treatment principles of traditional Chinese medicine have been developed and summed up throughout a long history of medical practice. From the ancient Chinese medicine books, we can also find that the herbs with similar characters have similar effects, while those with different characters have quite different effects. In Su's studies [40,41], they analyze several formulae that have proven effective, including *Wan Shi Niu Huang Qing Xin Wan, Xiao Chai Hu Tang, Hua Tuo Zai Zhao Wan and Nao De Sheng Pian.* The results show that the classification is in accordance with the traditional explanation of Chinese medicine.

3.2 A New Computer-aided Method for Prescription Formulation

Su's studies, introduce in the previous chapter, provide us with a method through which the characters of herbs can be quantified. We have noticed that a TCM prescription is a selective combination of multiple herbs, which indicates the probability of building the quantified characters of the prescription as a whole, according to those of the herbal components. Based on this analysis, we employ Support Vector Machine (SVM), a machine learning method, for the first time, to explore the principle of prescription formulation, to evaluate the effectiveness of a

formula and even generate new formulae.

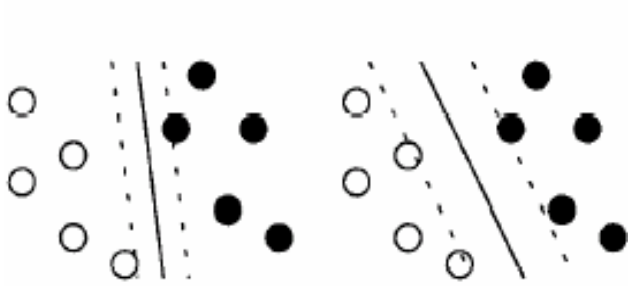
3.2.1 Support Vector Machine (SVM)

SVM is one kind of learning machine based on statistical learning theory. The basic idea of applying SVM to pattern classification is as follows: First, the input vectors are mapped into a feature space, either linearly or non-linearly, which is relevant with the selection of the kernel function. And then a hyperplane of the vectors in the feature space is constructed which separates two classes (this can be extended to multi-class). SVM training always seeks a global optimized solution and avoids over-fitting, so it has the ability to deal with a large number of features. A complete description to the theory of SVMs for pattern recognition is in Vapnik's book [43].

SVMs have been employed in a wide range of real-world problems such as text categorization [64-66], hand-written digit recognition [67,68], tone recognition [69-71], image classification and object detection [72-77], sub-storm and flood stage forecasting [78,79], cancer diagnosis [80-83], glaucoma diagnosis [84], microarray gene expression data analysis [85-87]. It has been shown that SVM is consistently superior to other supervised learning methods [88,89].

When used for classification, SVMs separate a given known set of {+1, -1} labeled training data via a hyperplane that is maximally distant from the positive samples and negative samples (known as Optimal Separating Hyperplane, OSH), then 'plot' the test data at the high dimensional space, distinguish whether it belongs to positive or negative according to the OSH. Figure 6 illustrates this process. The solid lines show

two possible hyperplanes, each of which correctly separates the training data into two classes. The two dashed lines parallel to the separating hyperplane show the boundaries in which one can move the separating hyperplane without misclassification. We call the distance between each parallel dashed lines as margin.



Small Margin Large Margin

Figure 6. Two possible separating hyperplanes

For most of real-world problems, seemingly not to be linearly separable, SVMs can handle non-linear hypotheses with any Kernel function, which automatically realizes a nonlinear mapping onto a feature space. The optimal separating hyperplane found by the SVM in feature space corresponds to a nonlinear decision boundary in the input space.

3.2.2 Linear Classification

Let the training data of two separable classes with n samples be represented by $(\vec{x}_1, y_1), (\vec{x}_2, y_2), \cdots, (\vec{x}_n, y_n), i = 1,2,\cdots,n,$ where $\vec{x}_i \in R^N$ is an N dimensional space,

and $y_i \in \{-1,+1\}$ is the class index. Given a weight vector $\vec{w}$ and bias b (Figure 7), the two classes can be separated by two margins parallel to the hyperplane:

$$\vec{w}^T \bullet \vec{x}_i + b \geq 1, \qquad y_i = +1 \tag{1}$$

$$\vec{w}^T \bullet \vec{x}_i + b \leq -1, \qquad y_i = -1 \tag{2}$$

where $\vec{w} = (w_1, w_2, \cdots, w_n)^T$ is a vector of n elements. Inequalities (1), (2) can be combined into a single inequality:

$$y_i(\vec{w}^T \bullet \vec{x}_i + b) \geq 1, \qquad i = 1,2,\cdots,n \tag{3}$$

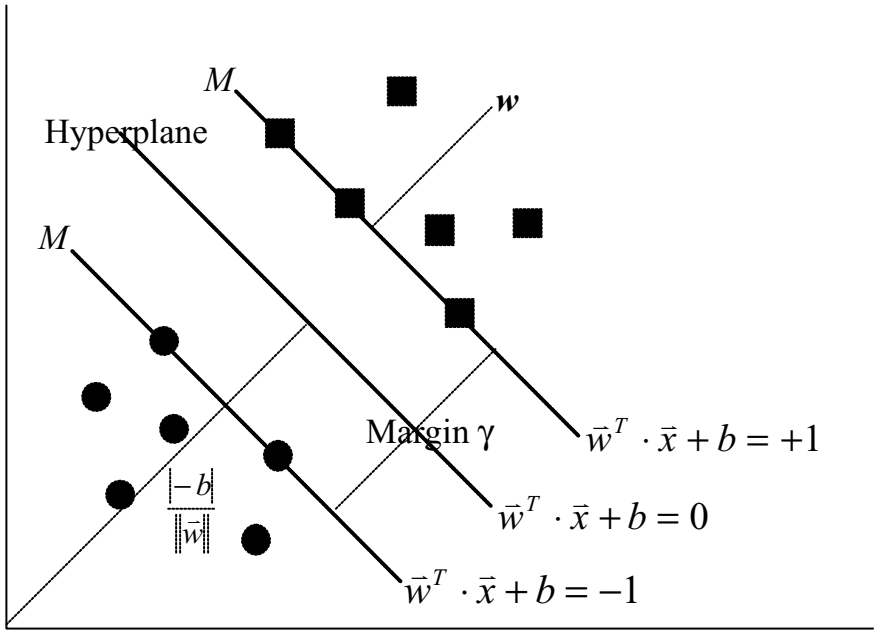


Figure 7. Definition of Hyperplane and Margin. The circular dots and square dots represent samples of class -1 and class +1, respectively.

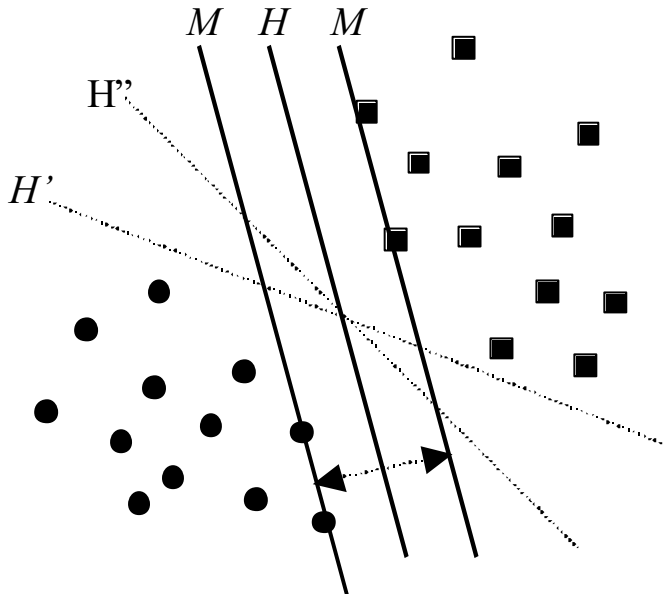


Figure 8. Schematic of the available Hyperplanes H, H', H''…, about a set of training data.

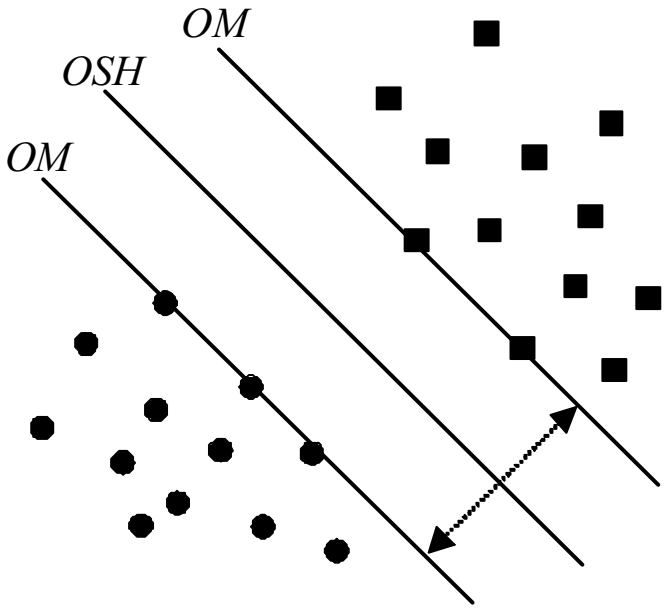


Figure 9. Schematic of unique Optimal Separation Hyperplane of a set of training data.

As shown in Figure 8, there exist a number of hyperplanes for each group of training data. The classification objective of SVM is to determine an optimal weight $\vec{w}_0$ and an optimal bias b_0 such that the selected hyperplane separates the training data with maximum margin. The hyperplane determined by $\vec{w}_0$ and b_0 is called optimal separation hyperplane (OSH) which is illustrated in Figure 9.

The equation of any given hyperplane can be written as:

$$\vec{w}^T \bullet \vec{x}_i + b = 0 \tag{4}$$

and the distance between the two corresponding margins is:

$$\gamma(\vec{w}, b) = \min_{\{\vec{x}|y=+1\}} \frac{\vec{x}^T \bullet \vec{w}}{\| \vec{w} \|} - \max_{\{\vec{x}|y=-1\}} \frac{\vec{x}^T \bullet \vec{w}}{\| \vec{w} \|} \tag{5}$$

The OSH can be found by maximizing the above distance or by minimizing the norm of $\|\vec{w}\|$ under inequality constraints (3), and

$$\gamma_{\max} = \gamma(\vec{w}_0, b_0) = \frac{2}{\|\vec{w}_0\|} \tag{6}$$

The saddle point of the following Lagrangean gives solutions to the above optimization problem:

$$L(\vec{w}, b, \alpha) = \tfrac{1}{2} \vec{w}^T \bullet \vec{w} - \sum_{i=1}^{n} \alpha_i [y_i (\vec{w}^T \bullet \vec{x}_i + b) - 1] \tag{7}$$

where $\alpha_i \geq 0$ are Lagrange multipliers. The solution of this optimization Quadratic Programming (QP) problem requires that the gradient of $L(\vec{w}, b, \alpha)$ with respect to $\vec{w}$ and b vanishes, which gives the following conditions by the calculation

of $\left.\frac{\partial L}{\partial \vec{w}}\right|_{\vec{w}=\vec{w}_0} = 0$ and $\left.\frac{\partial L}{\partial b}\right|_{b=b_0} = 0$:

$$\vec{w}_0 = \sum_{i=1}^{n} \alpha_i y_i \vec{x}_i \tag{8}$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \tag{9}$$

By substituting Eq.(8) and (9) into Eq.(7), the QP problem becomes maximization of the following expression:

$$L(\alpha) = \sum_{i=1}^{n} \alpha_i - \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (\vec{x}_i^T \bullet \vec{x}_j) \tag{10}$$

under the constrains $\sum_{i=1}^{n} \alpha_i y_i = 0$ and $\alpha_i \geq 0, i = 1,2,\cdots,n.$

The points situated at the two optima1 margins have non-zero coefficients α_i among the solutions to Eq.(10), and are called Support Vectors (SV). The bias b_0 can be calculated as follows:

$$b_0 = -\tfrac{1}{2}(\min_{\{\vec{x}_i | y_i = +1\}} \vec{w}_0^T \bullet \vec{x}_i + \max_{\{\vec{x}_i | y_i = -1\}} \vec{w}_0^T \bullet \vec{x}_i) \tag{11}$$

After determination of support vectors and bias, the decision function that separates the two classes can be written as:

$$f(x) = sign[\sum_{i=1}^{n} \alpha_i y_i \vec{x}_i^T \bullet \vec{x} + b_0] = sign[\sum_{SV} \alpha_i y_i \vec{x}_i^T \bullet \vec{x} + b_0] \tag{12}$$

3.2.3 Nonlinear Classification

As real-world problems are usually nonlinear, the following approach has been

introduced into SVM to deal with these problems. The original training data $\vec{x}$ in the input space **X** is projected into a high-dimensional feature space **F** via a Mercer kernel operator K [67], and the OSH is constructed in this feature space. Such a procedure is illustrated in Figure 10.

In mathematical terms, the set of classifiers is transformed into the form:

$$f(x) = sign[\sum_{i\in\{SV\}} \alpha_i y_i K(\vec{x}_i, \vec{x}) + b_0] \tag{13}$$

where K is a symmetric positive definite function, which satisfies Mercer's conditions,

$$K(\vec{x}, \vec{y}) = \sum_{m=1}^{\infty} \alpha_m \phi(\vec{x})^T \bullet \phi(\vec{y}), \qquad \alpha_m \geq 0$$

$$\iint K(\vec{x}, \vec{y}) g(\vec{x}) g(\vec{y}) d\vec{x} d\vec{y} > 0, \qquad \int g^2(\vec{x}) d\vec{x} < \infty \tag{14}$$

The Kernel represents a legitimate inner product in input space:

$$K(\vec{x}, \vec{y}) = \phi(\vec{x})^T \bullet \phi(\vec{y}) \tag{15}$$

In the **F**-space, the dual Lagrangian, given in Eq.(10), is

$$L(\alpha) = \sum_{i=1}^{n} \alpha_i - \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(\vec{x}_i, \vec{x}_j) - \lambda \sum_{i=1}^{n} \alpha_i y_i \tag{16}$$

subject to $\sum_{i=1}^{n} \alpha_i y_i = 0$ and $\alpha_i \geq 0, i = 1,2,\cdots,n.$

and the decision function is:

$$f(x) = sign[\sum_{i\in\{SV\}} \alpha_i y_i K(\vec{x}_i, \vec{x}) + b_0] \tag{17}$$

where

$$b_0 = -\frac{1}{2}\{\min_{\{\vec{x}_i | y_i = +1\}}(\sum_{j \in \{SV\}} \alpha_j y_j K(\vec{x}_i, \vec{x}_j)) + \max_{\{\vec{x}_i | y_i = -1\}}(\sum_{j \in \{SV\}} \alpha_j y_j K(\vec{x}_i, \vec{x}_j))\} \quad (18)$$

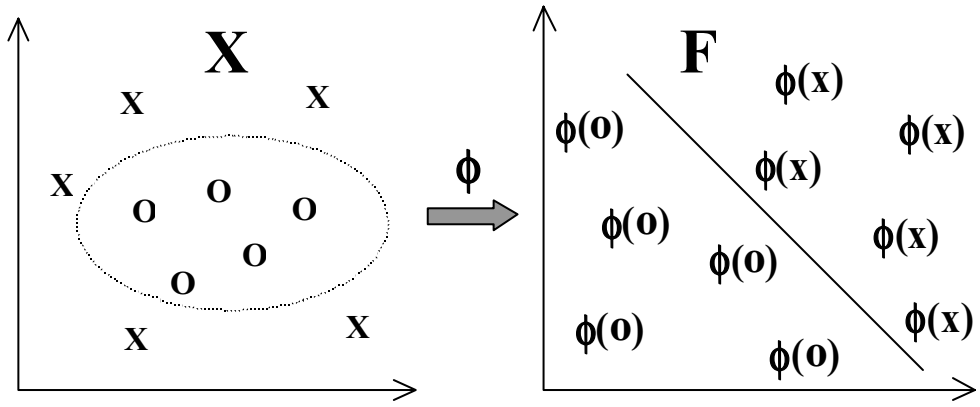


Figure 10. Illustration of basic principle of support vector machines: Step 1: project the training data nonlinearly into a higher-dimensional feature space via ϕ. Step 2: construct a hyperplane to separate positive and negative datasets with maximum margin there.

A number of candidate kernel functions have been used in SVM, including Polynomial $K(\vec{x}, \vec{y}) = (1 + \vec{x} \bullet \vec{y})^d$, Gaussian RBF $K(\vec{x}, \vec{y}) = \exp(-\frac{\|\vec{x} - \vec{y}\|^2}{2\sigma^2})$, Exponential RBF $K(\vec{x}, \vec{y}) = \exp(-\frac{\|\vec{x} - \vec{y}\|}{2\sigma^2})$, and their combinations by summing kernels or tensor products of kernels [90]. In this work, Gaussian RBF is chosen as the kernel function.

3.3 Dataset Preparation

In this study, 644 TCM formulae (from TCMID database, Chapter 2) are used to test the feasibility of this method. Based on the number of herbs in a formula, these formulae are classified to 7 groups, from Group 1 to Group 7, in which the number of herbs is 4, 5, 6, 7, 8, 9 and 10, respectively. These formulae are used as positive samples. 1000 negative samples, which contain the same number of herbs with the positive samples in the same group, are randomly generated. These negative samples are supposed to be therapeutically noneffective because they are only randomly compiled complexes of multiple herbs. The total numbers of samples in each group are given in Table 3. Both positive and negative formulae in each group are further randomly divided into two parts with even number of entries, respectively (If the number of the positive formulae is an odd number (2n+1), they are divided into two parts with n and n+1 formulae, respectively.). One part is used as the training set, while the other as the testing set. Thus, from Group 1 to Group 7, the training sets contain 42, 41, 46, 42, 50, 47 and 55 positive samples, respectively and 500 negative samples which are randomly selected. The testing sets contain 42, 40, 46, 42, 50, 46 and 55 positive samples, respectively and 500 negative samples.

3.4 Feature Vectors

Construction of feature vectors is one of the key steps in successful SVM classification. Because a formula is composed of several different herbs, at first we

must construct the vectors of each herb, and then, construct the feature vectors of the formula according to some rules that can compile the vectors of all herbs. In this study, the feature vectors are simply the concatenation of the vectors of all herbs. I will explain it in the following paragraphs.

Table 3. Number of Positive Formulae and Negative Formulae in Each Group

Formula Group	Number of Herbs in a Formula	Number of Positive Formulae		Number of Negative Formulae	
		Training Set	Testing Set	Training Set	Testing Set
Group 1	4	42	42	500	500
Group 2	5	41	40	500	500
Group 3	6	46	46	500	500
Group 4	7	42	42	500	500
Group 5	8	50	50	500	500
Group 6	9	47	46	500	500
Group 7	10	55	55	500	500

Table 4. Principle for constructing the feature vector

Elements	Characters	Value
1	Cold	1 (Cold), 1.2 (Extremely cold), or 0.8 (Slightly cold) *
2	Hot	1 (Hot) or 1.2 (Extremely hot) *
3	Warm	1 (Warm), 1.2 (Extremely warm) or 0.8 (Slightly warm) *
4	Cool	1 (Cool) or 0.8 (Slightly cool) *
5	Pungent	1 (Pungent) or 0.8 (Slightly Pungent)
6	Sweet	1 (Sweet) or 0.8 (Slightly Sweet) **
7	Sour	1 (Sour) or 0.8 (Slightly Sour) ***
8	Bitter	1 (Bitter) or 0.8 (Slightly Bitter)
9	Salty	1 (Salty) or 0.8 (Slightly Salty)
10	Toxic	1 (Toxic), 1.2 (Extremely Toxic), or 0.8 (Slightly Toxic)
11	Lung	1 (if has)
12	Bladder	1 (if has)
13	Spleen	1 (if has)
14	Large Intestine	1 (if has)
15	Stomach	1 (if has)
16	Small Intestine	1 (if has)
17	Liver	1 (if has)
18	Pericardium	1 (if has)
19	Heart	1 (if has)
20	Kidney	1 (if has)
21	Gallbladder	1 (if has)
22	San Jiao	1 (if has)

* If the herb has mild property, the value of each element should be added 0.25.

** If the herb has tasteless flavor, the value of this element should be added 0.5

*** If the herb has astringent flavor, the value of this element should be added 0.5.

Since the characters of a herb are essential to its effects, the feature vectors should be constructed according to these characters. The characters are quantified and then the values of them are used to compose the feature vector. The vector contains 22 elements, which are the quantified basic characters of herbal properties, flavors, toxicity property and meridians that the herb can enter. Table 4. gives probable value for each element of the feature vector and also the reason for the choice of the value.

As we have known from previous chapters, each herb has its own properties and flavors. There are four typical types of properties, that is, cold, hot, warm or cool. These four characters are used as four elements of feature vector. If the herb has one of these properties, a value '1' is assigned to the element. Otherwise, a value '0' is assigned to this element. The choice of value '1' is to simplify the construction of the vector.

There is also another property: mild. If the herb has this property, the value of the elements, which represent Cold, Hot, Warm and Cool, add a value 0.25. It is because this property does not belong to any of these four typical ones. The value 0.25 is chosen mainly according to our experiences. So are the other values, which are chosen to describe the elements in the following paragraphs. The discussion of the choice of the value is given in Chapter 3.6. Some herbs are extremely cold and some extremely hot, the value '1.2' is assigned to the elements, which represent cold and hot, respectively. To those herbs with minor cold, minor warm or minor cool, a value '0.8' is assigned to the corresponding element.

Table 5. Example: Feature Vector of *Herba Ephedrae* (Ma Huang)

Elements	Characters	Value
1	Cold	0
2	Hot	0
3	Warm	1
4	Cool	0
5	Pungent	1
6	Sweet	0
7	Sour	0
8	Bitter	0.8
9	Salty	0
10	Toxic	0
11	Lung	1
12	Bladder	1
13	Spleen	0
14	Large Intestine	0
15	Stomach	0
16	Small Intestine	0
17	Liver	0
18	Cardiovascular	0
19	Heart	0
20	Kidney	0
21	Gallbladder	0
22	San Jiao	0

There are five major types of herbal flavors, including pungent, sweet, sour, bitter and salty. These characters are also used as elements of the feature vector. If the herb has one of these flavors, a value '1' is assigned to this element. If the flavor is minor pungent, minor sweet, minor sour, minor bitter or minor salty, a value '0.8' is assigned to the corresponding element. There are also another two different flavors: tasteless and astringent. For tasteless, a value '0.5' is added to the element 'sweet' and for astringent, to 'sour'. According to Viscera, Channels and Collaterals theory, the symptoms can reflect the organs that are not in good conditions. So ancient Chinese herbalist doctors speculated that a given herb may selectively act upon a particular part of the body and this part depends on the corresponding symptoms that the herb can relieve. So the vector should also reflect the properties of meridians that herbs affect.

There are a total of 12 meridians, including Lung, Bladder, Spleen, Large Intestine, Stomach, Small Intestine, Liver, Pericardium, Heart, Kidney, Gallbladder and San Jiao. The quantified characters of these meridians are also used as elements of the feature vector. If one herb can enter a meridian, then a value '1' is assigned to the element that represents for the meridian. '0' is assigned to the other elements, which are not related with this herb. Table 5. gives the feature vector of *Herba Ephedrae* (Ma Huang), a useful herb which has warm properties, pungent and slightly bitter flavors, and can enter Lung and Bladder meridians. The related information can be found in TCMID database. So does the information of the other herbs, on which the feature vectors are built.

After building the feature vector of each herb in a formula, we can construct that of

the formula. In this study, I connect the elements of all herbs' vector together to generate a bigger vector. For example, the formula, which contains 4 herbs, has the feature vector with 88 elements, while the vector of a formula (contains 5 herbs) has 110 elements. Because of the difference of feature vector, we separated the formulae into 7 groups and tested them individually (Table 3).

3.5 Accuracy Measure

Accuracy of the results from discriminative methods is commonly measured by the quantity of True Positives (*TP*), True Negatives (*TN*), False Positives (*FP*), and False Negatives (*FN*) [91,92]. In addition to these quantities, standard sensitivity, specificity and overall accuracy (*Q*) performance measures defined by

$$\text{Sensitivity}=TP/(TP+FN) \tag{19}$$

$$\text{Specificity}=TN/(TN+FP) \tag{20}$$

$$Q=\frac{TP+TN}{TP+TN+FP+FN} \tag{21}$$

are also useful in assessing the prediction accuracy [93]. To estimate performance, we used three objective indices: accuracy, specificity and sensitivity. The sensitivity as defined above is the probability of correctly predicting a positive example. The specificity as defined above is the probability of correctly predicting a negative example. All these quantities are used in the evaluation of SVM classification of formulae in this work.

3.6 Results and Discussion

The SVM program can automatically rearrange the training and testing sets. During the training process, the program can calculate the minimum number of the positive samples or negative samples in the training set, which is necessary to reach an optimal training result. If the number of the positive samples or negative samples in the training set is higher than the minimum, the rest samples are automatically transferred to the testing set. On the contrary, if the number is lower than the minimum, some samples in the testing set are taken into the training set and then re-start the training process until an optimal training result is achieved. Tables 6 to12 give the training and testing sets of positive formulae of each groups after the optimal training results are achieved and Table 13 summarizes the number of positive and negative samples in the training set and testing set of each group after the training process. The results show that the number of the positive samples in the training sets of these groups increase (see Table 3). This is perhaps because the numbers of positive samples in the original training set is not enough to cover sufficiently diverse range of positive samples. So a number of positive samples in initial testing set are put into the training set. On the other hand, except Group 1, in the other six groups, the number of the negative samples in the training set decrease from 500 to a certain number between 210 and 290. The decrease of the number of negative samples in the training set indicates that they are enough to achieve the optimal result.

Table 6. List of Positive Formulae in the Training and Testing Set of Group 1

List of positive formulae (contain 4 herbs)		
Training Set		
Bai Tou Weng Tang	Huang Teng Nan Chang Tang	Sheng Ma Huang Qi Tang
Bao Chi San	Huo Xue Tong Mai Pian	Shi Zao Tang
Bei Xie Fen Qing Yin	Ji Jiao Li Huang Wan	Si Jun Zi Tang
Bing Lang Chen Qi Tang	Jia Wei Dang Gui Bu Xue Tang	Si Miao Yong An Tang
Bing Peng San	Jian Shen Fang	Si Ni San
Cai Shi Tui Re Tang	Jiu Fen San	Si Sheng Wan
Cang Er Zi San	Ling Gui Zhu Gan Tang	Si Wei Huang Qi Xuan Fu Pian
Da Cheng Qi Tang	Lou Xing Chen Qi Tang	Si Wei Tu Mu Xiang San
Da Jian Zhong Tang	Ma Huang Tang	Si Wu Tang
Da Xian Xiong Wan	Ma Huang Xing Ren Yi Yi Gan Cao Tang	Tao Hua San
Dao Chi San	Ma Xin Shi Wei Tang	Tong Xie Yao Fang
Ding Xiang Shi Di Tang	Ma Xing Gan Shi Tang	Wei Jing Tang
Ding Zhi Wan	Meng Shi Gun Tan Wan	Wu Long Jian Ji
Fu Zi Xie Xin Tang	Mu Fang Ji Tang	Wu Zhu Yu Tang
Gan Cao Gan Jiang Ling Zhu Tang	Qian Yin Tang Jiang	Xi Jiao Di Haung Tang
Gan Mao Tui Re Chong Ji	Qian Zheng San	Xiang Su San
Gan Mao Tui Re Ke Li	Qing Bing San	Xiao Er Qi Ying Wan
Gui Zhi Gan Cao Long Gu Mu Li Tan	Qiong Yu Gao	Xiao Zhi Shuan
Gun Tan Wan	Re Yan Ning Ke Li	Xie Bai San
He Zi Pi San	San Ao Tang	Xin Shen Yin
Hong Yao Tie Gao	San Liang Ban Yao Jiu	Xuan Bai Cheng Qi Tang
Huai Hua San	San Ye Tang	Yang Xue Run Chang Jian
Huang Lian Jiang Tang San	Shen Fu Tang	Yi Nian Jin
Huang Lian Jie Du Tang	Sheng Ma Ge Gen Tang	Yu Quan Wan
Huang Qin Tang		Zhi Ke Pi Pa Tang Jiang
Testing Set		
Bai Hu Tang	Ge Gen Huang Qin Huang Lian Tang	Li Zhong Wan
Er Chen Wan	Ge Gen Qin Lian Pian	Ling Gui Zhu Gan Tang
Fu Ling Gui Zhi Bai Zhu Gan Cao Tang	Ge Gen Qin Lian Wei Wan	Ma Xing Shi Gan Tang
		Ge Gen Huang Lian Tang

Table 7. List of Positive Formulae in the Training and Testing Set of Group 2

List of positive formulae (contain 5 herbs)		
Training Set		
An Dan Tang	Jian Gu Tang	Tai Yi Gao
Bai Dai Wan	Jian Xin Fu Mai Lin	Tao Ren Cheng Qi Tang
Bai Hu Jia Gui Zhi Tang	Jing Zhi Guan Xin Pian	Tian Jin Gan Mao Pian
Ban Xia Hou Po Tang	Ju Yuan Jian	Tian Jin Zhi Ke Tang Jiang
Bao Long Wan	Kui Yang Pian	Tong Mai Yang Xin Wan
Bi Xie Fen Qing Yin	Ling Bao Hu Xin Dan	Wen Pi Tang
Bu Shen Qiang Shen Pian	Ma Xin Long Xie Xiong Tang	Wu Ling San
Da Huang Mu Dan Tang	Nao De Sheng Pian	Wu Pi San
Dan Qi Tang	Niao Shi 1 Hao	Wu Pi Yin
Dang Gui Zhi Tong Tang	Ning Sou Lu	Wu Shi Tang
Fu Kang Ning Pian	Qing Hao Bie Jia Tang	Xie Huang San
Fu Tu Dan	Qing Hou Yan He Ji	Xin Jia Xiang Ru Yin
Fu Zheng Gu Ben Fang	Qing Nao Jiang Ya Tang	Yi Huang San
Fu Zi Da Huang Tang	Qing Wei San	Yi Huang Tang
Fu Zi Li Zhong Wan	Qing Xuan Wan	Yin Pu Xiao Du Yin
Gua Lou Qu Mai Wan	Qing Yi 1 Hao	Yu Nu Jian
Guan Xin Su He Wan	Qu Tao Tang	Yue Bi Tang
Gui Zhi Fu Ling Wan	Run Chang Tang	Yue Ju Wan
Gui Zhi Fu Zi Tang	Sang Xin Shi Gao Tang	Zeng Ye Cheng Qi Tang
Gui Zhi Tang	Shen Xue Pian	Zhen Jiang Gao Yao
Hou Po Sheng Jiang Ban Xia Gan Cao Ren Shen Tang	Sheng Bai Chong Ji	Zhen Wu Tang
	Shi Qu Gan Ce Da Zao Tang	Zhi Ke Pi Pa Chong Ji
Huan Xin Dan	Si Fo He Ji	Zhu Ling Tang
Huang Lian E Jiao Tang	Suan Zao Ren Tang	Zhu Sha An Shen Wan
Huang Long Tang		Huo Xue Chu Xuan Tang
Testing Set		
Tao He Cheng Qi Tang	Gui Zhi Jia Gui Tang	Huang Qi Gui Zhi Wu Wu Tang
Bai Hu Jia Ren Shen Tang	Ling Gan Wu Wei Jiang Xin Tang	Kang Bai Huo He Ji
Bao Yuan Tang		
Fu Fang Da Qing Ye He Ji		

Table 8. List of Positive Formulae in the Training and Testing Set of Group 3

List of positive formulae (contain 6 herbs)		
Training Set		
An Shen Ding Zhi Wan	Jie Gu Gao	Sheng Hua Tang
Ban Xia Lu	Jie Re He Ji	Sheng Yi Xuan Hua Tang
Bei Mu Gua Lou San	Jin Ji Chong Ji	Si Jun Zi Wan
Bing Lang Si Xiao Wan	Jin Pu Tang	Si Shen Wan
Bu Fei E Jiao Tang	Jin Suo Gu Jing Wan	Suo Pi Yin
Chan Su Ding	Jin Yin Tang	Tao Hong Si Wu Tang
Chuan Xiong Zhi Tong Tang	Ku Shen Tong Lin Tang	Tou Nong San
Da Bu Yin Wan	Lai Zhi San	Tuo Hua Jian
De Sheng Wan	Lian Mei An Hui Tang	Wan Shi Niu Huang Qing Xin Wan
Er Cha Er Huang San	Liang Di Tang	
Er Chen Tang	Ling Yang Gan Mao Pian	Wu Mei Chen Qi Tang
Er Zhu Ling Pi Tang	Liu Shen Wan	Wu Tou Gui Zhi Tang
Fei Lian Fen Qing Tang	Liu Wei Di Huang Wan	Wu Wei Xiao Du Yin
Fu Fang Gan Yan Chong Ji	Ma Pu San	Xi Huang Wan
Gou Teng Yin	Ma Ren Run Chang Wan	Xiao Er Xiao Shi Pian
Gu Jing Wan	Mai Men Dong Tang	Xiao Huo Luo Dan
Guan Xin 2 Hao Yao	Nao Ling Su Pian	Xiao Jian Zhong Tang
Gui Fu Yi Lin Tang	Ping Wei San	Xiao Tong Tang
Hua Ban Tang	Qi Shu Hua Fen Tang	Xu Shi Tong Jing Fang
Hua Shui Zhong Zi Tang	Qin Lian Pian	Xuan Jiang Tang
Huo Xue Zhi Tong San	Qing Gong Tang	Yi Guan Jian
Ji Chuan Jian	Qing Ying Jie Yu Tang	Yi Mu Cao Gao
Jia Liu Wan	Ru Kuai Xiao Pian	Yi Qi Ju Xian Tang
Jia Wei Jiao Hong Yin	Run Chang Wan	Yi Yi Zhu Ye San
Jia Wei Si Ni San	San Cai Feng Sui Dan	Yin Chai Chong Ji
Jiang Can Er Huang San	Shan Zha Hua Zhi Wan	Yin Qiao Tang
Zuo Gui Yin	Shen Zhu San	Yue Bi Jia Shu Tang
Testing set		
Ai Hu Cheng Qi Tang	Fu Ling Pi Tang	Jia Wei Fu Xie Ning
Bai Mai An Shen Yin	Gui Fu Li Zhong Wan	Jia Wei Wu Yao Tang
Fang Ji Huang Qi Tang	Huo Luo Dan	Ju Pi Zhu Ru Tang
Feng Re He Ji	Dang Gui Shao Yao San	Li Huang Tang

Table 9. List of Positive Formulae in the Training and Testing Set of Group 4

List of positive formulae (contain 7 herbs)		
Training Set		
Ba Zhen Yi Mu Wan	Jia Jian Dang Gui Si Ni Tang	Shan Zha Xiao Mi Tang
Ban Long Wan	Jia Wei Gui Zhi Fu Ling Wan	Shang Dong E Jiao Gao
Ban Xia Xie Xin Tang	Jia Wei Zhi Zhuo Gu Ben Tang	She Xiang Zhen Tong Gao
Bao He Wan	Jiang Ya Ping Pian	She Yang Tang
Bu Fei Shen Ji Tang	Jie Biao Tui Re Fang	Shen Jin Dan Jiao Nang
Bu Yang Huan Wu Tang	Jin Lian Mao Gong Yin	Sheng Ge Er Chong Tang
Cai Hu Qing Gan Yin	Ku Shen Tang	Shun Jing Tang
Chai Hu Shu Gan San	Lan Di Tang	Tian Tai Wu Yao San
Chong Cao Bu Tian Jing	Li Pi Su Fei Tang	Tiao Gan Tang
Da Qing Long Tang	Lian Po Yin	Wan Ying Gao
Da Yuan Yin	Liu Yi Bing Zhu San Ma You Hu	Wei Kang Ling Jiao Nang
Dang Gui Liu Huang Tang	Luo Han Guo Zhi Ke Chong Ji	Xiao Chai Hu Tang
Dang Gui Si Ni Tang	Ma Zi Ren Wan	Xiao Er Yi Niao Fang
Di Yu Wan	Mi Zhen Tang	Xiao Yao Wan
Fei Er Wan	Mu Li Ze Xie San	Xing Jun San
Fu Fang Chan Yi Yin	Qi Bao Mei Ran Ke Li	Xuan Fu Dai Zhe Tang
Fu Fang Lei Gong Teng Jian Ji	Qi Wei Du Qi Wan	Yang He Tang
Gan Cao Xie Xin Tang	Qi Wei Pu Tao San	Yang Wei Li Qi Tang
Gan Ji San	Qi Wei Zhi Shen Tang	Yi Qi Huo Xue Pian
Ge Gen Tang	Qian Bai Bi Yan Pian	Yu Zhen San
Gong Wai Yun Fang	Qiang Huo Sheng Shi Tang	Yun Qi San
Hong Ling San	Qin Lian Si Wu Tang	Zhen Xin An Shen Tang
Hua Gan Jian	Qing Li Tang	Zhi Sou San
Huang Lian Tang	Qu Feng Zhi Tong Pian	Zhi Zhi Tang
Huang Qin Hua Shi Tang		Zhu Ye Mai Dong Shi Gao Tang
Huang Tu Tang		Zhu Ye Shi Gao Tang
Ji Ming San		Sha Shen Mai Dong Tang
Sang Xing Tang		
Testing Set		
Dang Gui Si Ni Tang	Dang Gui Jian Zhong Tang	Huang Qi Jian Zhong Tang
Huo Bu San Ren Tang	Ji Xue Teng Gao	

Table 10. List of Positive Formulae in the Training and Testing Set of Group 5

List of positive formulae (contain 8 herbs)		
Training Set		
Ba Wei Qing Xin Chen Xiang San	Kuo Qing Yin	Si Chong Huo Xue Tang
Ba Wei Tan Xiang San	Li Gan Long Pian	Snag Piao Xiao San
Ba Zhen Wan	Li Ling Tang	Su Wei Tang
Ban Xia Bai Zhu Tian Ma Tang	Ling Yi Bai Jiang Tang	Su Xiao Niu Huang Wan
Bei Ling San	Mi Niao Xi Jie Shi Fang Nu Bao	Tan Du Jian
Bi Xue Shen Shi Tang	Qi Gong Tang	Tong Jing Zhi Ke Tang
Bu Dai Wan	Qi Guan Yan Wai Tie Fang	Tong You Tang
Bu Zhong Yi Qi Wan	Qi Ju Di Huang Wan	Tuo Li Ding Tong Tang
Chan Su Wan	Qi Li San	Wen Dan Tang
Chuan Bei Qing Fei Tang Jiang	Qi Shao Ji Cao Tang	Wu Cao Tang
Chuan Xiong Cha Tiao Wan	Qiang Xin Ying	Wu Jia Qiang Xin Tang
Er Long Zuo Ci Wan	Qing Gan He Jie Tang	Xian Qi Yin
Fu Yuan Huo Xue Tang	Qing Gu San	Xiang Sha Liu Jun Zi Wan
He Che Da Zao Wan	Qing Guo Wan	Xiao Ban Tang
Hua Yu Qing San Tang	Qing Lin He Ji	Xiao Qing Long Tang
Hui Tian Zai Zao Wan	Quan Long Si Zi Tang	Xiao Yao San
Huo Xue Run Zao Sheng Jin San	Ren Shen Ding Chuan Tang	Xie Qing Wan
Jia Jian Tuo Hua Jian	Rou Gan Xi Feng Tong Luo Tang	Xie Shi Tang
Jia Jian Wei Rui Tang	Ru Xiang Ding Tong Wan	Yang Yin Li Yan Tang
Jia Wei Ban Xia Hou Po Tang	San Bai Tang	Yang Yin Qing Fei Tang
Jia Wei Fu Zi Li Zhong Tang	San Qi Shang Yao Pian	Yi Chan He Ji
Jia Wei Xiao Yao San	San Ren Tang	You Gui Yin
Jian Wei Cha	San Shen Tang	Zao Du Fu Jiang Tang
Jiao Ai Tang	Sang Ju Yin	Zhen Xin Di Tan Tang
Jie Nue Qi Bao Yin	Shen Kang Ning Pian	Zhi Shi Dao Zhi Wan
	Shen Qi Wan	Zhi Shu He Wei Tang
	Shu Gan Chong Ji	Zhi Xue Tang
	Ku Shen Zhi Yang Tang	Zuo Gui Wan
		Kang Bao Kou Fu Ye
		Kai Xiong Shun Qi Wan
Testing Set		
Da Bu Yuan Jian	Gou Zai Hua He Ji	Juan Bi Tang
Da Chai Hu Tang	Hou Po Wen Zhong Tang	Mai Wei Di Huang Wan
E Jiao Yi Shou Jing	Huang Qi Gui Di Tang	Niu Huang Jie Du Pian
Er Dong Tang	Jia Wei Wu Pi Yin	Niu Huang Jie Du Wan
Er Yin Jian	Jin Gui Shen Qi Wan	Nuan Gan Jian
Fu Fang Yi Bai He Ji		

Table 11. List of Positive Formulae in the Training and Testing Set of Group 6

List of positive formulae (contain 9 herbs)		
Training Set		
Ai Fu Nuan Gong Wan	Ju Fang Zhi Bao San	Ti Qi Yi Zhong Tang
An Kun Zan Yu Wan	Jun Li Tang	Tiao Wei He Ji
Ba Zhen San	Kai Wei Qi Cao	Ting Tao Chen Qi Tang
Chuan Xiong Cha Tiao San	Kang Gu Zeng Sheng Wan	Tong Qiao Huo Xue Tang
Da Zao Wan	Kang Yuan Tang	Wei Hu Tang
Ding Tong Tang	Li Shi Hua Yu Tang	Wei Qi Shuang Jie Tang
Ding Xuan Tang	Liang Ge San	Wei Yan Yan Fang
Fu Fang She Dan Chuan Bei San	Ling Yang Qing Fei San	Wen Yang Tong Mai Tang
Geng Nian An	Lu Jin Yin	Wu Shao She Yin
Gu Ci Wan	Mu Dan Pi Tang	Xi Ling Jie Du Pian
Gu Yin Jian	Niu Bang Jie Ji Tang	Xiao Er Jin Dan Pian
Gui Ling Gan Lu San	Pai Shi Chong Ji	Xiao Tan Hua Jie Tang
Gui Ling Ji	Qi Long Tang	Xiao Ying Tang
Gui Zhi Shao Yao Zhi Mu Tang	Qi Zhen Wan	Xie Fei Tiao Zhong Fang
Hu Qian Wan	Qing Li An Shen Tang	Xuan Bi Tang
Huang Qi Yi Shen Tang	Qing Qi Hua Tan Wan	Xuan Fei Zhi Ke Tang
Huang Qi Zhu Fu Tang	Qing Xin Lian Zi Yin	Xue Xie Gua Lou Tang
Jia Jian Yi Qi Gu Chong Tang	Qing Yan Wan	Yang Xue Dang Gui Jing
Jia Kang Ping Wan	Qing Ying Tang	Yi Jia Jian Zheng Qi San
Jia Wei Xiao Yao Wan	Qing Zao Jiu Fei Tang	Yi Qi Yang Yin Tang
Jian Pi Ding Xian Tang	Ren Shen Bai Du San	Yin Qiao Jie Du Wan
Jiao Mei Bing Guan Tang	Ren Shen Zi Bu Gao	Yu Yin Zhi Beng Tang
Jin Fei Cao San	Ru Mi Niao Tang	Zhi Gan Cao Tang
Jing Zhui Fang	Sha Mai Yu Tian Tang	Zhi Shi Xiao Pi Wan
Jiu Wei Qiang Huo Keli	Shao Fu Zhu Yu Tang	Zhi Zhuo Gu Ben Wan
Jiu Wei Qiang Huo Tang	Shao Yao Tang	Zhu Shi Tang
Jiu Wei Qiang Huo Wan	She Gan Ma Huang Tang	Zhu Ye He Ji
Jiu Wei Zi Shen Wan	Shen Jiang Xie Xin Tang	Zi Shen Rong Jing Wan
	Sheng Li Pian	Zi Yin Xuan Fei Tang
	Si Ling He Qin Shao Tang	Jiu Xian San
Testing set		
Jia Jian Bu Zhong Yi Qi Tang	Jiu Wei Xiao Ding Yin	Ma Xin Qin Ting Tang
	Ding Chuan Tang	Man Gan Jiu Wei Tang

Table 12. List of Positive Formulae in the Training and Testing Set of Group 7

List of positive formulae (contain 10 herbs)		
Training Set		
An Nao Niu Huang Pian	Jia Wei Xiang Su San	Shu Jin Huo Xue Pian
An Shen Bu Xin Wan	Jian Nao An Shen Tang	Shu Zao Yin Zi
Bai He Gu Jin Wan	Jie Du Huo Xue Tang	Su Zi Jiang Qi Wan
Bao Gu Jiu	Jie Xiao Tang	Tai Shan Pan Shi San
Bu Gan Yi Qi Tang	Kui Yang He Ji	Tang Shen Yi Tang
Bu Shen Yi Nao Tang	Lan Wei Xiao Yan Pian	Tian Ma Jiao Nang
Can Shi Tang	Li Dan Pai Shi Pian	Tian Ma Wan
Da Huang Qing Wei Wan	Lian Kui Yu Yang Tang	Wan Dai Tang
Da Huo Luo Wan	Ling Jiao Gou Teng Tang	Wei Chang Fu Yuan Tang
Dan Dao Pai Shi Er Hao Fang	Lu Shang Yu Zhen	Wei Tong San
Dan Zhi Xiao Yan San	Lu Tai Gao	Wen Fei Hua Yin Tang
Fang Nian Lian Tang	Mi Yuan Jian	Wen Jing Huo Xue Tang
Fu Fang Zang Lian Wan	Mie Di Ku Shen Tang	Wen Yang Yi Qi Tong Mai Tang
Fu Gan Tang	Mie Di Tang	Wu Mei Tang
Gan Lu Yin	Mu Xiang Shun Qi Wan	Wu Mei Wan
Gan Wei Bai He Tang	Niu Huang Bao Long Wan	Xi Xian Gou Ji Xian Lin Pi Tang
Gong Ti Wan	Qing Fei Yi Huo Wan	Xiao Ji Yin Zi
Gu Chong Tang	Qing Fei Yin Zi	Xiao Jin Dan
Han Lian Cao Jian	Qing Gan An Tai Yin	Xiao Yan Shen Ji He Ji
Hou Zheng Wan	Qing Shi Yi Qi Tang	Xin Jia Huang Long Tang
Hu Gu Yao Jiu	Qu Feng Zheng Rong Tang	Xin Shuai He Ji
Hu Po Bao Long Wan	Re Du Qing	Yi Qi Chu Tan Fang
Hua Yu Zhi Beng Tang	San Huang Shi Gao Tang	Yin Jun Hu She Tang
Hui Yan Zhu Yu Tang	Shang Gan Tang	Yin Qiao San
Huo Xiang Zheng Qi Kou Fu Ye	Shao Fu Zhu Yu Wan	Zhen Ren Yang Zang Tang
Huo Xiang Zheng Qi Shui	Shen Lu Bu Gao	Zhi Chuang Xun Xi Fang
Huo Xue Tong Luo Tang	Shen Rong Jiu	Zhi Dai Fang
Ji Sheng Shen Qi Wan	Shi Hu Si Jin Tang	Zhou Che Wan
Ji Xue Teng Tang	Shi Quan Da Bu Gao	Zhu Huang Bei Mu Tang
Jia Jian Xiao Cai Hu Tang	Shi Quan Da Bu Wan	Zhuang Yao Jian Shen Wan
Jia Wei Bai He Di Huang Tang	Shi Wei San	
	Shu Gan Huo Xue Tang	
	Jia Wei Jiao Ai Si Wu Tang	
Testing Set		
Ba Zhen Tang	Dao Chi Wan	Jian Nao Bu Shen Wan
Bu Zhong Yi Qi Tang	Fu Ke Shi Wei Pian	Long Dan Xie Gan Tang
Cai Hu Yu Jin Pai Shi Tang	Hua Shuan Tong Mai Tang	Long Dan Xie Gan Wan
Chen Xiang Hua Qi Wan	Jia Jian Zhu Qi Bu Tai Tang	Ren Shen Gui Pi Wan
Dang Gui Yang Xue Wan	Jia Wei Yin Qiao San	Bai He Gu Jin Tang
Jiu Qi Nian Tong Wan		

Table 13. Number of Samples in the Training and Testing sets after Calculation Using SVM

Formula Group	Training Set		Testing Set			
	Positive	Negative	Positive		Negative	
			TP	FN	TN	FP
1	74	501	10	0	498	1
2	73	251	5	3	749	0
3	80	211	10	2	786	3
4	79	267	1	4	733	0
5	84	231	10	6	768	1
6	88	290	3	2	710	0
7	94	210	14	2	790	0

Description of TP, FN, TN and FP in Chapter 3.5

The prediction accuracy of SVM in classifying TCM formulae is given in Table 14. The overall accuracy of SVM for classifying Group 1 was 99.8%; the sensitivity, 100%; the specificity, 99.8%. The overall accuracy of SVM for classifying Group 2 was 99.6%; the sensitivity, 62.5%; the specificity, 100%. The overall accuracy of SVM for classifying Group 3 was 99.4%; the sensitivity, 83.3%; the specificity, 99.6%. The

overall accuracy of SVM for classifying Group 4 was 99.5%; the sensitivity, 20%; the specificity, 100%. The overall accuracy of SVM for classifying Group 5 was 99.1%; the sensitivity, 62.5%; the specificity, 99.9%. The overall accuracy of SVM for classifying Group 6 was 99.7%; the sensitivity, 60%; the specificity, 100%. The overall accuracy of SVM for classifying Group 7 was 99.7%; the sensitivity, 87.5%; the specificity, 100%.

Table 14. Sensitivity, Specificity and Overall Accuracy of the Results

Formula Group	Herbs Number	Sensitivity $TP/(TP+FN)$	Specificity $TN/(TN+FP)$	Q $(TP+TN)/(TP+TN+FN+FP)$
Group 1	4	100%	99.8%	99.8%
Group 2	5	62.5%	100%	99.6%
Group 3	6	83.3%	99.6%	99.4%
Group 4	7	20%	100%	99.5
Group 5	8	62.5%	99.9%	99.1%
Group 6	9	60%	100%	99.7%
Group 7	10	87.5%	100%	99.7%

Description of Sensitivity, Specificity, Overall Accuracy, TP, FN, TN and FP in Chapter 3.5

The high overall accuracies (>99%) may reflect the reliability of this method. At least, it has strong ability to recognize negative samples. The specificities of all 7 groups are bigger than 99.5%. For Group 2, 4, 6 and 7, all of the negative samples are recognized. Thus this SVM system appears to be capable of minimizing the probability of incorrect selection of non-effective formulae. However, the high overall accuracies may not truly reflect its ability to pick up "true positive", that is, the known effective formulae. In this study, the values of overall accuracies largely depend on the number of "true negative" samples, which is much higher than "true positive" samples. That is the reason why, in some groups, Sensitivity is much lower than Specificity and Q. In other words, we can say that recognition ability of "true positive" is different in different groups, or, more seriously, the evaluation of a known effective formula may be incorrect. Table 14 shows satisfactory results for Group 1, Group 3 and Group 7 (sensitivity > 83%). These groups have a common character, that is, in the testing set, the positive samples (TP and FN) are relatively more than, at least not less than those samples in the other groups. In Group 1, Group 3 and Group 7, the positive sample is 10, 12 and 16, respectively. While in Group 4, the recognition ability is particularly low, with only 20% of the formulae correctly classified. The number of the positive testing samples in this group is 5. Similarly, in Group 6, the number of the positive testing formulae is also 5. For this group, the sensitivity is 60%. Considering the small number of these positive testing samples, the results in these groups may be not reliable. Several factors that limit this method are given in the next paragraphs.

Several factors may affect the prediction accuracy. Firstly, the number of positive samples is critical to the predicted results. As shown in Table 14, the lower values of sensitivity than those of specificities may be the consequence of lack of positive data. Secondly, the number of negative samples also needs careful consideration. This number should be higher than, but not much higher than the minimum. In this study, the number of the negative testing samples is too high, while that of the positive testing samples is too low. Further studies should be carried out to reduce the number of the former and to find more positive samples for testing. Moreover, SVM prediction can be further improved by choosing better SVM optimization procedure and feature vector. The selection of feature vector may be more important. This includes the selection of the value of each element and the construction rule of formulae's feature vector from herb' feature vector. Changing to a proper value of an element may generate reasonable results. For example, when a herb is extremely cold, the element that represents "cold" is assigned a value 1.2 (Table 4). This value can be changed to 1.1, 1.2 or whatever, and then a new feature vector is generated. Based on a set of these new feature vectors, an optimal training result may be achieved. This will be carried out in the future study. There may be a number of possible methods to construct the feature vector of the formulae from the herb's feature vector. In this study, I select a simple one (see Chapter 3.4). This method has an obvious disadvantage, that is, it can only train the formulae that contain the same number of herbs. Hence, I have to separate the formulae into several groups and then train and test them respectively. The other methods for constructing the feature vector should be employed in the future

study.

Table 15. False Predicted Negative Formulae (or Potential Formulae)

Herb Number	Formula Name	Herbal Components	Description
4	Negative Formula-4 443	Mu Xiang, Chuan Xin Lian Ye, Zhi Ma Qian Zi, Shuan Hua	FP
6	Negative Formula-6 18	Man Jin Zi, Ji Nei Jin, Deng Xin Cao, Duan Long Gu, Man Jin Ye, Xi He Liu	FP
6	Negative Formula-6 414	Fa Ban Xia, Cang Zhu, Guo Lu Huang, Zhi Cao, Meng Shi, Zi Bei Chi	FP
6	Negative Formula-6 481	Chui Pen Cao, Ye Ming Sha,Ya Tuo Cao, Jiu Da Huang, Ting Li Zi, Shan Zha	FP
8	Negative Formula-8 533	Zhi Jun Tan, Xiao Gan Cao,Gan Song, Su Geng, Su He Xiang You, Rou Cong Rong, Tu Ku Shen, A Wei	FP

Table 15 gives the ineffective formulae that are incorrectly predicted as effective formulae. One possible reason for the incorrect prediction is the inherent deficiency of this method as described above. But we cannot rule out the possibility that these randomly generated formulae do have effective in treatment of some kinds of diseases.

In conclusion, our study suggests the capability of SVM in recognizing non-effective formulae and it may provide some helpful hints for herbalist doctors to determine the effectiveness of a TCM formula. In addition, the computation provides

several potentially effective formulae from the hundreds of randomly mixed formulae. These formulae never appear in previous literature and their effects have never been studied before. It is unclear whether these formulae have the therapeutic value. The method is expected to facilitate the prescription of new and novel TCM formulae as well as the validation of existing TCM formulae while more and more formulae are under scientific studies.

Chapter 4: Exploration of the molecular mechanism of a Medicinal Herb *Serenoa repens* by INVDOCK

4.1 Introduction

The pharmacological use of phytotherapeutic agents in the treatment of benign prostatic hyperplasia (BPH) and associated lower urinary tract symptoms (LUTS) has been growing steadily [45-49,94-96]. The most widely used phytotherapeutic agent is the extract of *Serenoa repens*, one kind of dwarf palm plants [45-49,97-104]. Many clinical trials have been carried out to evaluate the effect of the extract of this herb [98,102,103,104,105]. A recent meta-analysis of published clinical trial data of Permixon, one commercial product of the extract of *Serenoa repens*, (11 randomized clinical trials and 2 open label trials) [100], involving 2859 patients, reveals that in men with symptomatic BPH, the extract yields a significant improvement in peak flow rate and reduction in nocturia compared with placebo. Studies in hormone-treated castrated rats also show inhibition effects on the increase of prostate wet weight by the extract of *Serenoa repens* [107]. When compared with commonly used drugs, an alpha-blocker (Tamsulosin) [98] or 5 alpha-reductase inhibitor (Finasteride) [108], the extract shows nearly the same effects, causing parallel and statistically significant

decreases in symptom scores and increases in maximal flow rates. These results indicate the effects of the extracts of *Serenoa repens* on men with symptomatic BPH.

Its main mechanism of action is still unclear; however, possible activities have been demonstrated by several *in vivo* and *in vitro* studies, including antiandrogenic, antiproliferative and anti-inflammatory activities [96]. 5alpha-reductase catalyzes the reduction of testosterone into dihydrotestosterone (DHT), which is critical in regulating prostate cell growth. Studies demonstrate that the lipido-sterol extract of *Serenoa repens* (LSESr) markedly inhibit 5alpha-reductase [109-111]. In addition, it also exerts an anti-androgenic effect through inhibition of the receptor binding of androgens [111].

In one study, LSESr significantly inhibits proliferation and induces cell death in both epithelium and stroma of BPH issue [112]. It also appears to affect the b-FGF-induced proliferate of prostate cells [113] and decreases the amount of EGF in BPH [105]. Another study demonstrates that *Serenoa repens* may block PRL-induced prostate growth by inhibiting several steps of PRL receptor signal transductions [104]. It exerts an anti-inflammatory effect through the inhibition of the enzyme responsible for prostaglandin and leukotriene synthesis, such as cyclooxygenase and lopoxygenase [105].

The mechanism of this herb is unknown. It is also unclear which of the compounds in the extract of this herb account for its therapeutic effects. Many experiments assume that the lipid and sterol in this herb account for the therapeutic effects

[101,103,104,107,109,112-117]. Shimada *et al* separates two compounds from the extract: 1-monolaurin and 1-monomyristin, and finds that they show moderate biological activities in the brine shrimp lethality test and against renal (A-498) and pancreatic (PACA-2) human tumor cells [118]. A number of compounds have been extracted from this herb. To explore the mechanisms and pharmacology of this herb, these ingredients need to be examined first. And it is more important to explore how they can collectively interact with various proteins and other molecules in the human body. But such an investigation is hindered by the high cost and time spent on synthesis and bioassay of natural products. Therefore alternative methods are needed to understand and analyze molecular mechanism of active ingredients at a higher speed and lower cost.

As one of the strategies for mechanistic study of natural products, the software INVDOCK is developed to calculate the binding between the compounds and the protein targets [119]. To evaluate the capability of INVDOCK in identification of potential therapeutic and toxicity protein targets of compounds, Chen *et al* has tested on a more diverse set of compounds [42,120]. The predicted potential therapeutic and toxicity protein targets for a number of clinical drugs were found to be in fair agreement with available experimental data. Knowledge of these potential targets combined with that of proteomics and pharmacokinetic profile can facilitate the assessment of therapeutic and toxic effects as well as the molecular mechanism of predicted effects.

When this work was done, I did not have the 3D structure of 5 alpha-reductase and

androgen receptor, so I did not evaluate the effects of the compounds on these two targets. In this work, I focus on the mechanism of this herb involved in other processes to provide the experimentalists the most possible targets, which, I believe, will facilitate them in the construction of experiments. 23 active compounds from this herb are selected to search for their potential targets through INVDOCK. Because these compounds are the main ingredients of *Serenoa repens*, the collective effects of them may account for parts of the actions of this herb in the treatment of BPH.

4.2 INVDOCK Method

4.2.1 Protein Cavity Database

According to Kuntz's work [121], a cavity is modeled by a group of overlapping spheres that fill up that cavity. Each cavity entry is derived from the corresponding PDB entry by the following procedure: Ligands and water in a PDB protein structure are first removed. The surface of this protein, as defined by Richards [122], is then generated. The *van der Waals* surface of a solvent-accessible atom is generated with the respective parameter from the AMBER force field [123]. The inward-facing surface covering the interface of *van der Waals* surfaces is computed with a probe sphere 1.4 Å in radius. The whole protein surface is then coated and filled with a cluster of spheres by a method similar to that of Kuntz et al. Each sphere is checked for the extent of its surrounding space covered by protein atoms. The surrounding space is defined as a region within 15 Å of the center of the sphere. A sphere is considered covered if more than 50% of the direction around the sphere is covered by protein

atoms. The covered spheres are then divided into separate cavity groups on the basis of the spatial separation of nearest neighboring spheres.

In addition, non-covered spheres close to a covered sphere in a group are included in that group. This is to ensure that a cavity model sufficiently covers the cavity surface region where a tail of a ligand might be located. In some ligand-protein PDB structures, such as HIV-1 protease complex with its inhibitor, one or both ends of the ligand are found to be located in such a region. The computer-generated model includes both the cavity and the groove to which it is connected. Because a groove may also be a binding site, no attempt has been made to remove grooves from a cavity model. A visual inspection of these and a large number of other proteins showed that all the cavities and their adjacent grooves were reasonably represented.

3D structures of proteins from different species have been deposited in the PDB. Because of species-related variation in protein sequences, ligand binding is often specific to a protein of a particular species, or a species class, and those of close phylogenetic distance. In general, the main interest of target identification is to understand the molecular mechanism of the therapeutic and physiological effect of a ligand with respect to one species or a few particular species. Therefore, cavity entries can be divided into classes defined by disease-related species, sequence similarity, and phylogenetic distance to a particular species or species class such as human and mammal. Additional information such as the known ligand-binding region and the

extent of each sphere being buried inside a cavity is also included in a cavity entry to facilitate the prioritized docking strategy described next.

4.2.2 Inverse-docking Procedure

The inverse-docking procedure INVDOCK is designed for the automated search of every entry of the protein cavity database for potential protein targets of a small molecule. Because of the large number of cavity entries, across-the-board specification of possible binding sites within each cavity is difficult. Hence, all parts of each cavity are subject to docking. Docking to sites inside each cavity starts from known ligand-binding sites, followed by more interior sections and then the remaining part. To save central processing unit (CPU) time, the program proceeds to the next protein entry when the first successful dock is obtained without an exhaustive search for the optimum binding mode within each cavity. Although the optimum binding mode is not specifically sorted, the prioritized search strategy seems to dock a ligand to a site reasonably close to the experimentally observed positions in most of the ligand-protein structures studied. All cavity entries are subject to searching unless the related protein has been identified as a potential target.

Flexible ligand docking is similar to the multistep strategy approach proposed by Wang et al [124] Ligand conformation is sampled at a resolution similar to that of Wang et al. The docking of a particular conformer to a cavity is as follows: First, the ligand is aligned within the selected site by the position of each ligand atom being matched with the center of the spheres. Because of the relatively low-resolution nature

of ligand conformation sampling, a certain degree of structural clash is allowed at this stage. A molecular-mechanics conformation optimization is then conducted by a limited torsion space sampling of rotatable bonds in the ligand and those in the side-chain of the receptor amino acid residues at the binding site. Each rotatable bond is sampled at ±15°. This is followed by 50 iterations of Cartesian coordinate energy minimization on all ligand and protein atoms at the binding site to further optimize the ligand-protein complex. Energy minimization is by a steepest decent method.

In both torsion optimization and energy minimization, AMBER force fields [123] are used for covalent-bond, bond-angle, torsion, and nonbonded *van der Waals* and electrostatic interactions. A large number of crystal structures in PDB contain only non-hydrogen atoms. To save computing time and avoid the difficulty in modeling hydrogens, the Morse potential [125], which is a function of the donor-acceptor distance, is used to represent hydrogen-bond terms. This potential has been shown to give a reasonable description of hydrogen-bond energy and dynamics in biomolecules [126,127]. The published Morse potential parameters [126,127] are used in this work.

Protein flexibility is known to influence ligand binding, and thus methods such as side-chain conformational sampling and protein ensemble generation have been explored in docking studies [128,129]. Limited side-chain conformation sampling is considered in INVDOCK along with additional energy minimization for all atoms at the binding site. However, no attempt is made to explicitly sample the ensemble of protein conformations. A substantial number of proteins in PDB have multiple entries

submitted by different groups; some of them are associated with different binding ligands or different mutants. Moreover, there are multiple conformations in most NMR structures. To a certain extent, INVDOCK searching of these multiple forms of structures/conformations may serve as a partial sampling of the conformation for some proteins.

4.2.3 Scoring

The scoring of docked structures is based on a ligand-protein interaction energy function, $?E_{LP}$, composed of the same hydrogen-bond and nonbonded terms as those used for structure optimization. An analysis of a large number of PDB ligand-protein complexes shows that the computed $?E_{LP}$ is generally less than $?E_{Threshold} = -\alpha N$ kcal/mol, where N is the number of ligand atoms and α is a constant (~1.0). The exact value of α can be determined by the fitting of the linear equation $?E_{Threshold} = -\alpha N$ to the computed $?E_{LP}$ for a large set of PDB structures. This statistically derived energy value can be used empirically as a threshold for screening likely binders. A polynomial form of $?E_{Threshold}$ involving more parameters may also be introduced to derive an energy threshold. $?E_{LP}$ can be required to be less than $?E_{Threshold}$ when successfully docked structures are selected.

Ligand binding is competitive in nature. A drug is less likely to be effective if it binds to its receptor noncompetitively against natural ligands and, to some extent, other drugs that bind to the same receptor site. This binding competitiveness may be partially taken into consideration for cavities known to be ligand-bound in at least one

PDB entry. Ligands in PDB structures are known binders. Therefore, PDB ligands bound to the same receptor site as that of a docked molecule may thus be considered competitors of that molecule. In INVDOCK selection of a potential protein target, the computed $?E_{LP}$ is not only evaluated against $?E_{Threshold}$ but also compared to the ligand-protein interaction energy of the corresponding PDB ligands that bind in the same cavity in this or other relevant PDB entries. The ligand-protein interaction energy for the relevant PDB structures is computed by the same energy functions as that for the docked molecule. In addition to the condition that it be lower than $?E_{threshold}$, $?E_{LP}$ of a docked molecule is required to be lower than a competitor energy threshold $?E_{Competitor}$ when a potential target is selected. $?E_{Competitor}$ can be taken as the highest ligand-protein interaction energy of the corresponding PDB ligands multiplied by a factor β. For finding weak binders as well as strong binders, a factor $\beta \leq 1$ is introduced to scale the ligand-protein interaction energy of PDB ligands. This is because a weak binder may have a slightly higher interaction energy than that of a PDB binder. We have not found experimental data to determine the value of β. Here, β is tentatively determined by an analysis of the computed energy for a number of compounds. Our study suggests that a value of 0.8 for β leads to reasonable results statistically.

4.2.4 Selection of Compounds and Therapeutic and Toxicity Proteins

23 ingredients of *Serenoa repens* (Table 16) are collected according to published literature and books [96,130-132]. The 3D structures of these ingredients are generated through the following steps. First, the 2D structures are got from the Combined

Compound Database (CCD) (http://www.chemnetbase.com). Second, these structures are converted to 3D structures through the software Weblab. The structures are given in Figures 11 to 14.

Therapeutically targeted proteins are well documented in pharmacology literature and in a number of publications. A database of therapeutic targets, TTD (http://xin.cz3.nus.edu.sg/group/TTD), has been developed [133]. TTD currently contains 433 protein entries that have been reported to be the target of either clinical drugs or investigating drugs.

Table 16. Herbal ingredients of *Serenoa repens*

Compound Class	Compound Name
Fatty acid (8)	Capric acid, Caproic acid, Caprylic acid, Lauric acid, Myristic acid, Oleic acid, Palmitic acid, Stearic acid
Sterol (5)	Beta-sitosterol, Cycloartenol, Lupenone, Lupeol, Stigmasterol
Ethyl ether of fatty acid (8)	Ethylcappate, Ethylcaproate, Ethylcaprylate, Ethyllaurate, Ethyloleate, Ethylmyristate, Ethylpalmitate, Ethylstearate,
Monoacylglycerides (2)	Monolaurin, Monomyristin

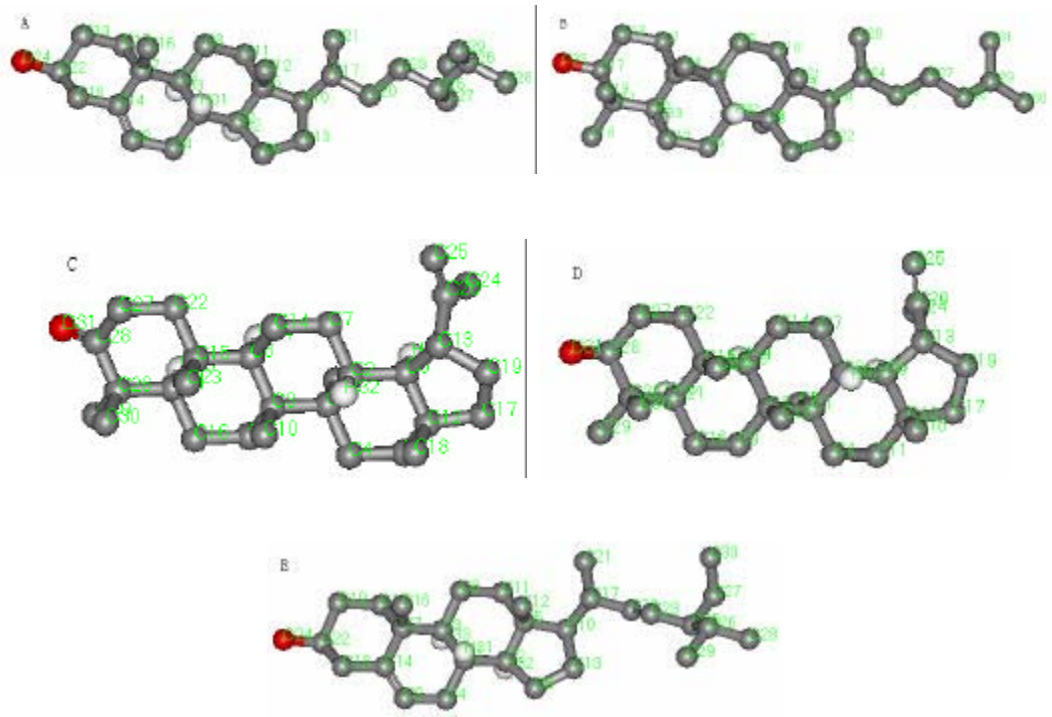


Figure 11. 3D Structure of Phytosterols of *Serenoa repens*
A: Beta-sitosterol, B: Cycloartenol C: Lupenone, D: Lupeol, E: Stigmasterol

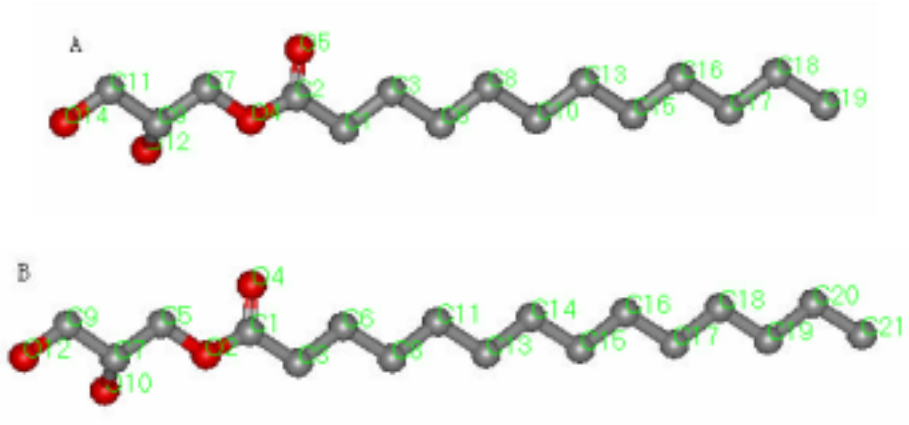


Figure 12. 3D Structure of Monoacylglycerides of *Serenoa repens*
A: Monolaurin, B:Monomyristin

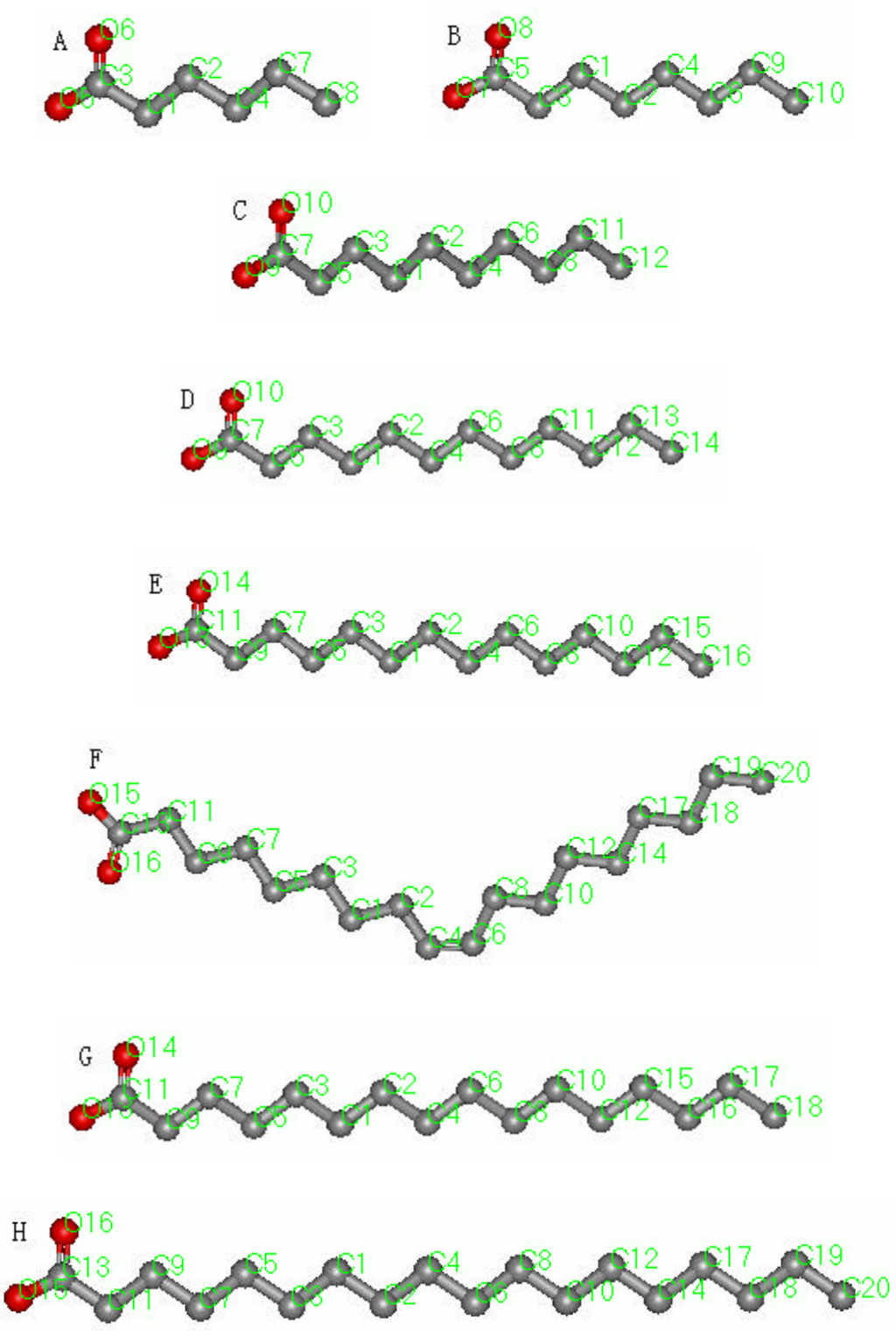


Figure 13. 3D Structure of Fatty acids of *Serenoa repens*
A: Caproic acid, B: Caprylic acid, C: Capric acid, D: Lauric acid, E: Myristic acid, F: Oleic acid, G: Palmitic acid, H: Stearic acid

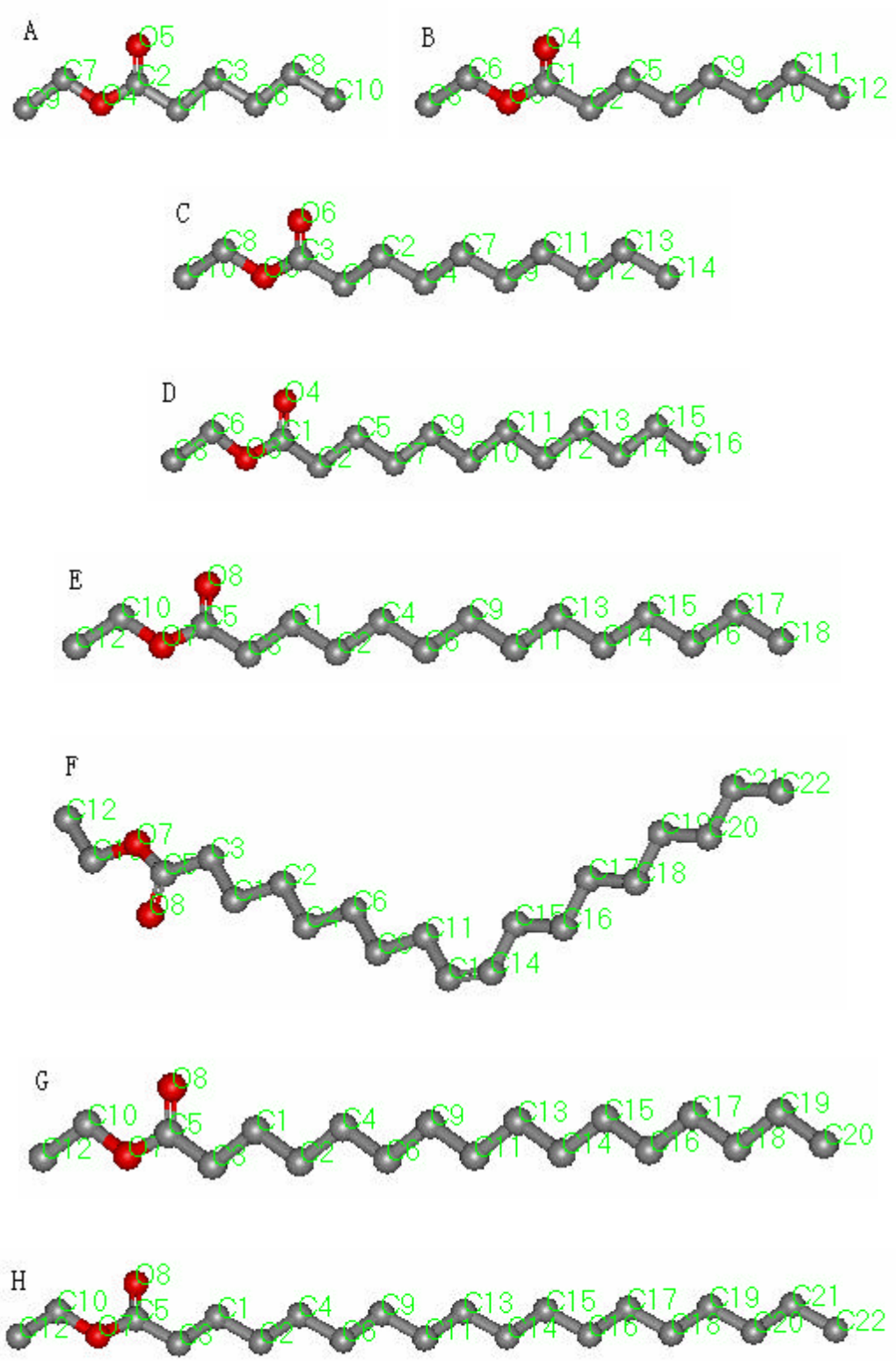


Figure 14. 3D Structure of Ethyl Esters of Fatty acids of *Serenoa repens*
A: Ethyl caproate, B: Ethyl caprylate, C: Ethyl cappate, D: Ethyl laurate, E: Ethyl myristate, F: Ethyl oleate, G: Ethyl palmitate, H: Ethyl stearate

4.3 Results

All the 23 ingredients, including eight fatty acids, eight ethyl esters of these fatty acids, five sterols and two monoacylglycerides, are used to scan the protein library to identify their potential targets. Based on the information, their contribution to the therapeutic and toxic effects of *Serenoa repens* is evaluated (Table 16). These compounds constitute more than 85% of all the ingredients in the extract of *Serenoa repens*, which is used for treatment of BPH and is believed to contain the pharmacologically active components. Although their mechanism of action is still unclear, theoretically, the therapeutic effects of this herb may be due to some of these compounds.

From the computational results, each compound can bind to several protein targets, ranging from one to seven. Taking into account of the overlap of the same targets by different compounds, finally, 25 protein targets are obtained (Table 17 and Table 18), each of which can be bound by several different compounds. To a given protein target, the binding energy of different compounds varies significantly. For example, the results show that 20 compounds can bind to prostaglandin H2 synthase (the data is not included in the thesis), but their binding energy is very different. The highest one is that of stigmasterol and the lowest one is that of capric acid. The energy is 47.6 kcal/mol (Table 17) and 28.4 kcal/mol (this data does not include in the thesis), respectively. Compared with the former, the latter is a very weak binder and may have little contribution to the interaction with the target protein. Hence, in this study, I only consider the interactions with sufficiently strong binding energy. When the binding

energy of a compound to a given protein is 3 kcal/mol less than the maximum, the interaction between the compound and the protein is neglected. The compounds that can bind to the targets are given in Table 17 and Table 18.

The action of the herbal extract has been suggested to be through the following mechanism [45,49,100,134]: anti-inflammatory, direct inhibition of prostate growth and inhibition of growth factor-induced growth, anti-androgenic and anti-estrogenic effects. A number of proteins involved in these processes are in the list of our predicted targets.

4.3.1 Anti-inflammatory Effects

The anti-inflammatory effects are mainly through interference with prostaglandin synthesis [45,100,103]. One critical enzyme involved in this process, Prostaglandin H2 synthase, is found highly expressed in prostate cancer and BPH issue, suggesting an important role in the pathogenesis of BPH. Studies find its expression can be decreased by the extract of *Serenoa repens* [135]. One study suggests phytosterols to be the most importance of inhibition of this enzyme [104]. From our study data, two phytosterols (stigmasterol and cycloartenol) can bind to this enzyme. On the other hand, beta-sitosterol, believed to be the most important ingredient, is found not to bind to this enzyme. Further study need to be carried out to compare their effects.

Table 17. Predicted Proteins related with BPH. **[energy >= (Max-3) kcal/mol]**

PDB ID	Protein Name	Pathway	Ref.	Ligand and Energy
1ebv 1diy	Prostaglandin H2 Synthase (Cyclooxygenase)	Prostaglandin synthesis	[135]	Stigmasterol -47.6 Cycloartenol -47.2
1fgk	FGF Receptor 1	FGF-induced cell proliferate	[136,138]	Stigmasterol -44.8
1e0o	Fibroblast Growth Factor 1	FGF-induced cell proliferate	[137]	Lauric acid -36.0 Monomyristin -34.2 Capric acid -33.7
1exz	Stem Cell Factor	GF-induced cell proliferate	[138]	Stearic acid -43.8
7odc	Ornithine Decarboxylase	Polyamine synthesis and cell growth	[141,147]	Stearic acid -33.7 Oleic acid -33.1 Myristic acid -31.3
1jen	S-Adenosylmethionine Decarboxyla	Polyamine synthesis and cell growth	[142]	Palmitic acid -38.3 Monolaurin -37.7 Caprylic acid -37.5
	Protein Kinase C	Proliferation of stromal cell	[143]	Betasitosterol -46.2 Monolaurin -46.2 Oleic acid -44.5 Stearic acid -43.7
2hdh	L-3-Hydroxyacyl COA Dehydrogenas	Generation of DHT	[146]	Oleic acid -37.4 Lauric acid -36.8
1err	Estrogen Receptor	Estrogenic pathway	[96,106]	Stigmasterol -53.5 betasitosterol -46.6
2nll	Retinoic Acid Receptor	Proliferation by retinal	[145]	Stigmasterol -49.7
1ed4	Nitric Oxide Synthase	NO pathway	[144]	Monolaurin -47.9 Lupeol -45.8 Stigmasterol -45.7

4.3.2 Anti-proliferate Effects

In BPH, bFGF concentration is 2 to 3-fold higher than in normal prostate. Moreover, over-expression of FGFR1 is observed in BPH [136-138]. Studies also demonstrate that the extract of this herb can alter bFGF growth factors-induced growth and proliferation [113]. In their study, it is mentioned that lupenone and the unsaponified fraction of the LSESr markedly inhibit the b-FGF-induced cell proliferation. The inhibition effect may be through reducing the available growth factor receptor and growth factor. Our data show that one sterol (stigmasterol) can bind to fibroblast growth factor receptor. Moreover, fibroblast growth factor (FGF) can be bound by two fatty acid (lauric acid and capric acid) and monomyristin.

Another growth factor, stem cell factor, which is supposed to exert a broad range of biological activities during organogenesis and normal cell development [139], is expressed in Prostate carcinoma cell lines: DU-145 and PC-3 lines [140]. Out data show it is bound by stearic acid.

Besides the inhibition of the growth factor-induced growth of prostate cell, inhibition of other proteins that are critical to the growth of cells may account for the effect of this herb. Ornithine decarboxylase (ODC) is essential for polyamine synthesis and cell growth. Polyamine is believed to participate in cellular proliferation and differentiation. It has been found that not only the expression level of ODC mRNA, but also the activity of this enzyme is higher in BPH tissue than that of normal prostate tissues [141]. Our data show that three fatty acids, stearic acid, oleic acid and myristic

acid can bind to this enzyme. Another rate-limiting enzyme of the polyamine pathway, S-adenosylmethionine decarboxylase, is also in the list of predicted targets, binding by palmitic acid, caprylic acid and monolaurin. Both of them are found to be useful targets in the treatment of prostate cancer [142]. Protein kinase C (PKC) is involved in the proliferative response of human cultured prostatic stromal cells [143]. Our data show that betasitosterol, monolaurin, oleic acid and stearic acid can bind to this enzyme.

One predicted protein target is nitric oxide synthase. This enzyme catalyzes the production of NO, which is involved in several important pathways. One study finds that inducible nitric oxide synthase is expressed in the prostate of all benign prostatic hyperplasia (BPH) patients, suggesting the role of NO in the pathogenesis of BPH [144]. Monolaurin and two sterols (lupeol and stigmasterol) can bind to this enzyme. Vitamin A (retinol) and its derivatives, the retinoids, have been implicated as chemopreventive and differentiating agents in a variety of cancers, including that of the prostate. Very little is known about the physiological role of retinoids in the prostate. Another study finds that the nuclear retinoic acid receptors are expressed in BPH samples [145]. Inhibition of this receptor may neutralize the proliferate effect caused by retinal. Two fatty acids (oleic acid and Caprylic acid) and one sterol (stigmasterol) can bind to this receptor.

4.3.3 Anti-androgenic and Anti-estrogenic Effects

Many studies suggest the anti-androgenic and anti-estrogenic effects of *Serenoa*

repens in the treatment of BPH. L-3-hydroxyacyl-CoA dehydrogenase (SCHAD) has been demonstrated to convert 5alpha-androstane-3alpha, 17beta-diol (3alpha-adiol) to dihydrotestosterone (DHT) [146]. Dihydrotestosterone (DHT) is the major androgen implicated in the pathogenesis of BPH. A number of studies have shown that the *Serenoa repens* extract takes effect through reducing the generation of DHT [109-111]. It is believed that part of the anti-androgenic effects of this herb is due to the inhibition of SCHAD. Our data show that two fatty acids (oleic acid and lauric acid) can bind to this enzyme.

In regard to the anti-estrogenic effects of *Serenoa repens*, it has been found that it can competitively block the translocation of cytosolic estrogen receptors to the nucleus. [96,106]. Two sterols (stigmasterol and betasitosterol) can bind to this receptor. One *in vivo* study demonstrates that beta-sitosterol is capable of producing a weak estrogenic effect only at the lowest dose [148].

Our study suggests that effects of this herb on other targets may be involved in the treatment of this disease. In the list of our predicted proteins, besides the targets that have been mentioned previously, there are still some other targets. So, the next task of our study is to analyze these proteins to make clear the relationship of these proteins with BPH. Among these proteins, many proteins are found to be involved in cell proliferating, cell cycle, tumor metastasis and other process (Table 18). These proteins had higher expression in BPH or prostate cancer tissue. Targeting toward them can at least inhibit the conversion from BPH to advanced prostate diseases.

Table 18. Other predicted important proteins. **[energy >= (Max-3) kcal/mol]**

PDB ID	Protein Name	Pathway or Function	Ref.	Ligand and Energy
1hcl	Cyclin-Dependent Kinase 2	Cell cycle	[149]	Betasitosterol –48.1 Lupenone -47.1 Lupeol -47.1 Cycloartenol -46.3
1cws	Cdc25 B-Type Tyrosine Phosphatase	Cell cycle	[150,151]	Caproic acid -21.2
1doa	GTP-Binding Protein (CDC42)	Cell cycle	[152,153]	Monolaurin -45.2
1dow	Alpha-Catenin	Cell cycle	[154,155]	Capric acid -37.0 Caprylic acid -33.7
1a7c	Plasminogen Activator Inhibitor	Inhibition of apoptosis	[162]	Cycloartenol -49.2 Stigmasterol -47.0
1the	Cathepsin B	Invasion and metastasis	[156]	Stigmasterol -45.9 Monomyristin –44.7
1e5t	Prolyl Endopeptidase	High activity in tumor of prostate	[157]	Ethyloleate -33.2
1b3d	Stromelysin-1 (MMP 3)	Metastasis	[158]	Stigmasterol -49.5 Cycloarteno1 -46.5
	Neutrophil Collagenase (MMP 8)	Metastasis	[158]	Caprylic acid -38.8
	Fibroblast Collagenase (MMP1)	Metastasis	[158]	Betasitosterol -48.7
	MMP-13	Metastasis	[158]	Stigmasterol -43.7
	Adp-Ribosylation Factor 6	Metastasis	[159]	Caprylic acid -24.2 Caproic acid -22.2
	Chymase	Metastasis	[160]	Monolaurin -43.4
1ext	Tumor Necrosis Factor Receptor	Increase in PC	[161]	Ethyllaurate -25.1 Caproic acid -22.8 Ethylcaprylate -22.8

4.3.4 Arrest of Cell Cycle

Abnormalities of cell division occur in the carcinogenesis, due to overexpression of positive regulators of cyclin-dependent kinase (CDK) function. Many drugs have developed toward CDK to induce growth arrest of the cells at G1 phase [149]. Studies find that the extract of this herb can induce apoptosis and inhibit cell proliferation of BPH tissues from patients [112]. In our data, four phytosterols (Beta-sitosterol, lupenone, lupeol and cycloartenol) can bind to this enzyme.

The other two important molecules involved in the cell cycle, Cdc25 B-Type Tyrosine Phosphatase [150,151] and CDC42 [152,153] may be the potential targets. From our data, they are targets of caproic acid and monolaurin, respectively. Alpha-catenin is important in cell adhesion in the process of proliferate. It has been shown that its production is elevated in prostatic carcinoma in comparison with benign prostatic hyperplasia and normal prostate [155]. Our data show two fatty acids (capric acid and caprylic acid) can bind to it. Plasminogen activator inhibitor 1 (PAI-1) is suggested by one study to be involved in the tumor growth through the anti-apoptotic effect in the benign human breast epithelial cell line MCF-10A [162]. Our data show that two phytosterols (cycloartenol and stigmasterol) can bind to this protein.

Awad AB *et al* [162] finds that phytosterol, especially beta-Sitosterol, inhibit cell growth and induce cell cycle arrest using MDA-MB-231 human breast cancer cells to evaluate the effects of phytosterols. Connolly JM's study shows inhibition effects of fatty acids on DU145 human prostate cancer cell growth in athymic nude mice [164].

4.3.5 Anti-metastasis

Matrix metalloproteinases (MMPs) are a family of zinc endopeptidases involved in tissue remodeling and thus in various disease processes including tumor invasion and joint destruction. Correspondingly, these enzymes represent attractive targets for inhibitor design and drug development. In our predicted list, four types of MMPs are found: MMP1 (Fibroblast Collagenase), MMP3, MMP8 (Neutrophil collagenase) and MMP13. Their concentrations are significantly higher in PCa patients with metastases [158]. MMP1 can be bound by beta-sitosterol, MMP3 by stigmasterol and cycloartenol, MMP8 by caprylic acid and MMP13 by stigmasterol. ADP-ribosylation factor 6 plays significant roles in coordinating the transition of epithelial cells from a stationary to a motile state when metastasis [159]. Caprylic acid and Caproic acid can bind to this protein. Chymase is first found in mast cells and seems to be involved in the degradation of the extracelluar matrix (directly and/or by activation of matrix metalloproteases), stimulation of submucosal gland cell secretion, and complement-mediated inflammation. Given the variety of pathophysiological processes in which chymase is involved, inhibitors of this enzyme may be potential drugs in cardiovascular and inflammatory diseases [160]. Monolaurin can bind to this protein. Cathepsin B is among the candidate proteinases believed to participate in invasion and metastasis. Cathepsin B often presents with higher amounts in malignant tumors than in normal tissues or benign tumors. Numerous clinical studies have confirmed a correlation between cathepsin B expression, disease progression, and

clinical outcome for patients with diverse types of tumors [156]. Stigmasterol and monomyristin can bind to this protein. Prolyl oligopeptidase activity is significant higher in tumours of prostate, lung and sigmoid, than in the healthy tissues. It supports the possible involvement of prolyl oligopeptidase in the renin-angiotensin system and in the pathogenesis of hypertension [157]. Out data show that ethyloleate bind to it.

Some studies have demonstrated the inhibitory effects of fatty acids [165,166] and phytosterols [163,167,168] on metastasis of cancer cells. The effects of other compounds, such as ethyloleate, should be investigated further. Although few experiments have been carried out to prove the relationship between these proteins and the pathogenesis of BPH, our study seems to indicate that they play an important role in this process.

4.4 Discussion

The results show that most of these chemical ingredients of *Serenoa repens* can bind to a number of targets, which has been proven by experiments. Moreover, our computational results also provide additional potential protein targets of these compounds. The results show in Table 19 that a compound can only bind to one or a few protein targets, which indicates the role of this herb in the treatment of BPH is due to a collective effect of the compounds.

The results indicate that not all compounds are helpful to the overall action of this herb. 5 of the 23 compounds, including ethylcappate, ethylcaproate, ethylmyristate,

ethylpalmitate and ethylstearate, may have little contribution to the expected actions of this herb. These five compounds have their own target proteins. But compared with the other compounds, they are relatively weak binders to these targets. For example, both ethylstearate and stigmasterol are predicted to interact with Prostaglandin H2 Synthase, but the latter strongly docked into the protein. Hence the effects of ethylstearate can be neglected.

Table 19. Summary of Compounds and their predicted targets

Compound	Predicted Targets
Capric acid	Fibroblast Growth Factor 1
	Alpha-Catenin
Caproic acid	Tumor Necrosis Factor Receptor
	Adp-Ribosylation Factor 6
	Cdc25 B-Type Tyrosine Phosphatase
Caprylic acid	S-Adenosylmethionine Decarboxyla
	Alpha-Catenin
	Neutrophil Collagenase (Mmp 8)
	Adp-Ribosylation Factor 6
Lauric acid	Fibroblast Growth Factor 1
	L-3-Hydroxyacyl Coa Dehydrogenas
Myristic acid	Ornithine Decarboxylase
Oleic acid	L-3-Hydroxyacyl Coa Dehydrogenas
	Protein Kinase C
	Ornithine Decarboxylase
Palmitic acid	S-Adenosylmethionine Decarboxyla
Stearic acid	Ornithine Decarboxylase
	Protein Kinase C
	Stem Cell Factor
Betasitosterol	Fibroblast Collagenase (Mmp1)
	Estrogen Receptor
	Protein Kinase C
	Cyclin-Dependent Kinase 2

Next page

Continued

Compound	Predicted Targets
Cycloartenol	Prostaglandin H2 Synthase (Cyclooxygenase)
	Cyclin-Dependent Kinase 2
	Plasminogen Activator Inhibitor
	Stromelysin-1 (Mmp 3)
Lupenone	Cyclin-Dependent Kinase 2
Lupeol	Nitric Oxide Synthase
	Cyclin-Dependent Kinase 2
Stigmasterol	Plasminogen Activator Inhibitor
	Stromelysin-1 (Mmp 3)
	Adp-Ribosylation Factor 6
	Fgf Receptor 1
	Retinoic Acid Receptor
	Nitric Oxide Synthase
	Prostaglandin H2 Synthase (Cyclooxygenase)
	Estrogen Receptor
	Cathepsin B
Ethylcaprylate	Tumor Necrosis Factor Receptor
Ethyllaurate	Tumor Necrosis Factor Receptor
Ethyloleate	Prolyl Endopeptidase
Monolaurin	Chymase
	Gtp-Binding Protein (Cdc42)
	S-Adenosylmethionine Decarboxyla
	Nitric Oxide Synthase
	Protein Kinase C
Monomyristin	Fibroblast Growth Factor 1
	Cathepsin B

Among the interaction between the predicted targets and the ingredients of *Serenoa repens* that have been mentioned above, many have not been proven or implicated by experiments. The discrepancy arises for a number of reasons.

It is not expected that exhaustive experiments have been done to determine all protein targets of a given compound, nor the multiple compounds of one herb. Because

the protein expression, protein concentration and accessibility vary largely in different cell system, on which many experimental studies of compound are based, observation of the molecular events related to the ligand-protein interaction is also difficult. These targets do not exist or have lower level in these cells. Although the binding energy suggests that the interaction should be possible, but not exist in these cells. The discovery of protein profiles can provide useful information to facilitate the study using our method.

Another discrepancy is caused by the difference of absorption and distribution of these compounds by different cells or issues. Even if a large doze of one chemical drug is taken, when it arrives at its protein target, its concentration may become very low compared with the initial concentration. On the contrary, a chemical with low initial concentration may easily arrive at its targets to take affect efficiently. In this regards, when we select the compounds, we chose not only those with great amount in the extract but also those with little amount. Besides, to determine the potential ligands of one given protein, we take into account of the binding energy. If the binding energy of a compound was 3 kcal/mol less than the maximum, the binding of the compound is neglected. Considered that the binding ability is exponential with the binding energy, this selection role may reasonably pick up the correct interaction.

The capability of the inverse docking approach in identifying potential protein targets of a small molecule is constrained by the relatively limited number of available protein 3D structures. This is particularly true for membrane-bond receptor proteins that are key therapeutic targets for a variety of diseases. Moreover, some of the PDB

structures may be of little relevance to binding study for a particular molecule. These include entries containing an incomplete section or a chain, protein mutants that are structurally different from the corresponding proteins investigated in experiments, ligand-bound proteins whose conformation is relevant only to a specific set of compounds and macromolecular complexes unrelated to a particular biological process studied experimentally. Computer selection of such irrelevant entries may thus generate 'false hits'. Anticipated rapid progress instructural genomics [169] is expected to provide a more diverse set of relevant structures. Knowledge from study of protein functions also facilitates the selection of relevant structures in determination of potential protein targets related to a particular cellular or physiological condition.

In our study, 5 alpha reductase, which has been long suggested to be the most important target in treatment of BPH, is missed because of the lack of 3D structure. Experiments demonstrate that 5 alpha reductase is a very important enzyme involved in the proliferation of the prostate. Studies show that inhibition of this enzyme could be a useful method to cure diseases related with the prostate. This problem can be partially alleviated by rapid progress in high-throughput X-ray crystallography and NMR of protein structures and in the development of new structural determination methods [169].

4.5 Conclusions

Our study provides the potential targets of *Serenoa repens* in the treatment of BPH, part of which have been demonstrated by previous experiments to be bound by some compounds in the extract. Besides these interactions, other predicted bindings between particular compounds and protein targets have not been proven by experiments. Moreover, from our data, other important targets may be involved in the pathogenesis of BPH and can be targeted by *Serenoa repens*. Thus further studies need to be carried out to evaluate these protein targets and the role of these compounds. In conclusion, as a relatively fast-speed and low-cost tool, this software may find application in systematic study of the molecular mechanism of multiple ingredients of other medicinal plants.

Chapter 5: Conclusions

In this work, I develop a new TCM database to provide comprehensive information about all aspects of TCM including prescription formulae, constituent herbs, known herbal ingredients and their molecular actions. The therapeutic effects, adverse reactions, clinical indications and applications are provided at the levels of prescription, individual herb and herbal ingredient respectively. Traditional terminologies about medicinal herbs such as tastes, flavors, site of actions, toxicity level, and therapeutic class are also given.

The study suggests the capability of SVM in recognizing non-effective formulae. Almost all negative formulae are correctly classified. Especially in Group 2,4,6 and 7, all of the negative samples are correctly recognized. The study may also have some helpful hints for herbalist doctors to determine the effectiveness of a TCM formula. Except Group 4 (formulae contain 7 herbs), the sensitivities of the other groups are more than 60%. Especially, this method has satisfactory performance for groups in which the formulae contains 4,6,10 herbs, respectively. In these groups, the sensitivities are more than 83%. In addition, several potential effective TCM formulae are derived from hundreds of randomly mixed recipes. These formulae never appear in previous literature and their effects have never been studied before. It is unclear

whether these formulae have the therapeutic value. It is expected that the method can be an alternative approach to study formulae while more and more formulae are under scientific studies.

The other part of this work is focused on the molecular mechanism study of herbal medicines by a computer-based method INVDOCK. In this study, I examine the interaction between the chemical ingredients of a medicinal herb *Serenoa repens* and a cavity database to generate the potential targets of these compounds. A total of 25 potential targets are obtained, each of which can be bound by several different compounds. Among the list of the targets, a number of proteins have been proven by experiments to be bound by the compounds. Moreover, many protein targets are found to be involved in cell proliferating, cell cycle, tumor metastasis and other processes and have higher expression in BPH or prostate cancer tissue, indicating the important roles in the treatment of BPH. The computational results indicate the feasibility of INVDOCK in systematic study of the molecular mechanism of multiple ingredients of *Serenoa repens*.

References

1. Yuan R and Lin Y. Traditional Chinese medicine: an approach to scientific proof and clinical validation. Pharmacology Therapeutics. 2000, 86:191-198.
2. Chan K. Progress in traditional Chinese medicine. TiPS. 1995, 16:182-187.
3. Porkert M. Chinese Medicine. William Morrow and Company, INC. 1988.
4. Wu DX. The basic principle of traditional Chinese medicine (Chinese). Shanghai Science and Technology Press. 1994.
5. Ling YK and Yan ZH. Traditional Chinese medicine. Shanghai Science and Technology Press. 1984.
6. Xu C. Exploration on the characteristics of classification in *Huangdi nei jing tai su*. Zhonghua Yi Shi Za Zhi. 1999, 29(2):109-112.
7. Shiang E and Li FP. The *Yin-Yang* (cold-hot) theory of disease. JAMA. 1971, 217(8):1108.
8. Jia F. Exploration on the medical life of *Bian Que*, a renowned physician in ancient China. Zhonghua Yi Shi Za Zhi. 1984, 14(4):214-215.
9. Shang Z. Discussion on the date of appearance of the title Shen nong ben caojing (Shennong's Herbal Classic). Zhonghua Yi Shi Za Zhi. 1999, 29(3):135-8.
10. Ouyang B. Studies on *Shang Han Lun* in Ming Dynasty and its influence. Zhonghua Yi Shi Za Zhi. 1995, 25(2):92-94.

11. Song Z. Research on the archaic pronunciation of *Ma Bi San* for *Ma Fei San*, a well-known anaesthetic proposed by *Hua Tuo* of Han Dynasty. Zhonghua Yi Shi Za Zhi. 1987, 17(2):76-78.

12. Liu ZZ. A course in English on Chinese materia medica. (15)Li Shi-Zhen--a great physician and herbalist. Zhong Yao Tong Bao. 1984, 9(4):187-90.

13. Li SZ. *Ben Cao Gang Mu.* People's Medical Publishing House. 1982.

14. Chen JT. Review: drug therapy in Chinese traditional medicine. Journal of Clinical Pharmacology. 2000, 40:445-450.

15. Diamond WJ. The clinical practice of complementary, alternative, and western medicine. CRC Press LLC. 2001.

16. Ang-Lee MK, Moss J and Yuan CS. Herbal medicines and perioperative care. JAMA 2001, 286:208-216.

17. Eisenberg DM, Davis RB, Ettner SL, Appel S, Wilkey S, Van Rompay M and Kessler RC. Trends in alternative medicine use in the United States, 1990-1997. JAMA. 1998, 280:1569-1575.

18. Kam PC and Liew S. Traditional Chinese herbal medicine and anaesthesia. Anaesthesia. 2002, 57(11):1083-1089.

19. Qu J, Zhang S and Xie R. The Chinese Materia Medica. Shanghai University of Traditional Chinese Medicine Press. 1990.

20. Xie Z. Prescriptions of traditional chinese medicine. Foreign Languages Press. 2000.

21. Im PO and Yolton RL. Concepts of traditional Oriental medicine. Optometry.

2000, 71(10):621-629.

22. Zhang L. Studies of active ingredients of Chinese traditional compound medicine. Progress in Chemistry. 1999, 11(2):197-199.

23. Yokozawa T, Liu ZW and Chen CP. Protective effects of Glycyrrhizae radix extract and its compounds in a renal hypoxia (ischemia)-reoxygenation (reperfusion) model. Phytomedicine. 2000, 6(6):439-445.

24. Niwa K, Hashimoto M, Morishita S, Yokoyama Y, Mori H and Tamaya T. Preventive effects of Glycyrrhizae radix extract on estrogen-related endometrial carcinogenesis in mice. Jpn J Cancer Res. 1999, 90(7):726-732.

25. Akao T. Effects of glycyrrhizin and glycyrrhetic acid on the growth, glycyrrhizin beta-D-glucuronidase and 3 beta-hydroxysteroid dehydrogenase of human intestinal bacteria. Biol Pharm Bull. 2000, 23(1):104-107.

26. Sun XY, Huang LJ and Liu X, Chemical ingredients of *Radix Glycyrrhizae.* ACTA Chinese Medicine and Pharmacology. 1994, 5:40-42.

27. Wang F and Su YH, Pharmacological action and clinical application of *Radix Glycyrrhizae*. Lishizhen Medicine and Meteria Medica Research. 2002, 13(5):303-304.

28. Fujisawa Y, Sakamoto M, Matsushita M, Fujita T and Nishioka K. Glycyrrhizin inhibits the lytic pathway of complement--possible mechanism of its anti-inflammatory effect on liver cells in viral hepatitis. Microbiol Immunol. 2000, 44(9):799-804.

29. Ou M. Chinese-English manual of common-used in traditional Chinese medicine.

Guangdong Science and Technology Press. 1992.

30. Banskota AH, Tezuka Y, Le Tran Q and Kadota S. Chemical constituents and biological activities of Vietnamese medicinal plants. Curr Top Med Chem. 2003, 3(2):227-248.

31. Singh VK, George CX, Singh N, Agarwal SS and Gupta BM. Combined treatment of mice with *Panax ginseng* extract and interferon inducer. Amplification of host resistance to Semliki forest virus. Planta Med. 1983, 47:234-236.

32. Singh VK, Agarwal SS and Gupta BM, Immunomodulatory activity of Panax ginseng extract. Planta Med. 1984, 50:462-465.

33. Tianjin Angel Power Allied Pharmaceutical Company. Hu Xing Dan, Tianjin, China, 1996.

34. Chiatai Qingchunbao Pharmaceutical Co. Ltd. Sheng Mai Ingection. Hangzhou, China, 1993.

35. Su W, Wu Z and Quan J. Application of fingerprint analysis and computer pattern interpretation in identification and quality evaluation of traditional Chinese medicines. Zhong Yao Cai. 2001, 24(4):295-298.

36. Wang LX, Xiao HB, Liang XM and Bi KS. Vectorial angle method for evaluating the similarity between two chromatographic fingerprints of Chinese herb. Yao Xue Xue Bao. 2002, 37(9):713-7.

37. Qiao Y, Li P and Wang Y, Journal of Beijing Univ. of TCM (Chinese), 1998, 21(5):17-19.

38. Su W, Wu Z, He X and Chen J. Chemical pattern recognition of traditional Chinese medicine kudingcha (II). Zhong Yao Cai. 1998, 21(4):170-173.

39. Su W, Wu Z, Liang R, Li Q and Huang T. Zhong Yao Cai (Chinese). 1997, 20(11):588-590.

40. Su W, Wu Z and Liang R. World Science and Technology (Chinese). 1999, 3:33-37.

41. Su W. China Journal of Chinese Materia Medica (Chinese). 1997, 22(3):186-188.

42. Chen YZ and Ung CY. Computer automated prediction of potential therapeutic and toxicity protein targets of bioactive compounds from Chinese medicinal plants. Am J Chin Med. 2002, 30(1):139-54.

43. Vapnik VN. Statistical Learning Theory. Wiley-interscience, New York, 1995.

44. Cai CZ, Han LY, Ji ZL, Chen X and Chen YZ. SVM-Prot: web-based support vector machine software for functional classification of a protein from its primary sequence. Nucleic Acids Res. 2003, 31(13):3692-3697.

45. Levin RM and Das AK. A scientific basis for the therapeutic effects of *Pygeum africanum* and *Serenoa repens*. Urol Res. 2000, 28(3):201-209.

46. Dreikorn K. Phytotherapeutic agents in the treatment of benign prostatic hyperplasia. Curr Urol Rep. 2000, 1(2):103-109.

47. Wilt TJ, Ishani A, Rutks I and MacDonald R. Phytotherapy for benign prostatic hyperplasia. Public Health Nutr. 2000, 3(4A):459-472.

48. Gerber GS. Phytotherapy for benign prostatic hyperplasia. Curr Urol Rep. 2002, 3(4):285-291.

49. Lowe FC. Phytotherapy in the management of benign prostatic hyperplasia. Urology. 2001, 58(6 Suppl 1):71-76.

50. Mantle F. The role of alternative medicine in treating postnatal depression. Complement Ther Nurs Midwifery. 2002, 8(4):197-203.

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

52. Ji YB and Zhang GM. Pharmacological action and application of available antitumor composition of traditional Chinese medicine (Chinese). Heilongjiang Science and Technology Press. 1998.

53. Ji YB. Pharmacological action and application of blood-activating and stasis-eliminating available composition of traditional Chinese medicine (Chinese). Heilongjiang Science and Technology Press. 1999.

54. Zhao C and Xu YZ, Antitumor herbal medicines and their active compositions. Traditional Chinese medicine press of China. 1997.

55. Ji YB. Pharmacological action and application of available composition of traditional Chinese medicine. Heilongjiang Science and Technology Press. 1994.

56. Huang KT. Handbook of composition and pharmacological action of common-used traditional Chinese medicine. China Pharmaceutical Science and Technology Press. 1994.

57. Zhang WG, Jia WC, Li SP, Zhang J, Ou YB and Xu XL. Chinese medicated diet. Shanghai University of Traditional Chinese Medicine Press. 1990.

58. Ou M. Chinese-English manual of common-used prescriptions in traditional

Chinese medicine. Guangdong Science and Technology Press. 1991.

59. Liu DL. Manual of common-used prescriptions in traditional Chinese medecine. People's Military Medical Press. 1996.

60. Li DQ, Shi LH, Gao Y and Jiang JX. Prescription of traditional Chinese medicine. Shanghai University of Traditional Chinese Medicine Press. 1990.

61. Shen FW. A comprehensive Chinese-Latin-English dictionary of the names of Chinese herbal medicines. Guangdong World Publishing Co., Ltd. 1998.

62. Zhang EQ. English-Chinese highly efficacious Chinese patent medicines. Shanghai University of Traditional Chinese Medicine Press. 1998.

63. He R. Xu Lingtai and his medical literatures. Journal of Zhejiang College of TCM. 2002, 24(6):13-14.

64. de Vel O, Anderson A, Corney M and Mohay G. Mining E-mail content for author identification forensics. Sigmod Record. 2001, 30,55-64.

65. Kim KI, Jung K, Park SH and Kim HJ. Support vector machine-based text detection in digital video. Pattern Recogn. 2001, 34:527-529.

66. Drucker H, Wu DH and Vapnik VN. Support vector machine for spam categorization. IEEE T. Neural Networ. 1999, 10:1048-1054.

67. Vapnik VN. The nature of statistical learning theory, 2nd Ed. Springer, New York, 1999.

68. Tong S and Koller D. Support vector machine active learning with applications to text classification. J Mach Learn Res. 2001, 2:45-66.

69. Li ZY, Tang SW and Yan SC. Multi-class SVM classifier based on pairwise

coupling. LNCS. 2002, 238:321-333.

70. Thubthong N and Kijsirikul B. Support vector machines for Thai phoneme recognition. Inter J Uncertain Fuzz. 2001, 9:803-813.

71. Gordan M, Kotropoulos C and Pitas I. A temporal network of support vector machine classifiers for the recognition of visual speech. LNAI. 2002, 230:355-365.

72. Ben-Yacoub S, Abdeljaoued Y and Mayoraz E. Fusion of face and speech data for person identity verification. IEEE T. Neural Networ. 1999, 10:1065-1074.

73. Wu CY, Liu C and Zhou J. Eyeglasses verification by support vector machine. LNCS. 2001, 219:1126-1131.

74. Wang YJ, Chua CS and Ho YK. Facial feature detection and face recognition from 2D and 3D images. Pattern Recogn Lett. 2002, 23:1191-1202.

75. Hsieh JW and Huang YS. Multiple-person tracking system for content analysis. Inter.J.Pattern Recognition and Artificial Intelligence. 2002, 16:447-462.

76. Papageorgiou C and Poggio T. A trainable system for object detection. Inter J Comput Vision. 2000, 38:15-33.

77. Karlsen RE, Gorsich DJ and Gerhart GR. Target classification via support vector machines. Opt Eng. 2000, 39:704-711.

78. Zhao Q, Principe JC, Brennan VL, Xu DX and Wang Z. Synthetic aperture radar automatic target recognition with three strategies of learning and representation. Opt Eng. 2000, 39:1230-1244.

79. Gavrishchaka VV and Ganguli SB. Support vector machine as an efficient tool

for high-dimensional data processing: application to substorm forecasting. J Geophysical Res. 2001, 106:29911-29914.

80. Liong SY, Sivapragasam C. Flood stage forecasting with support vector machines. J Am Water Resour As. 2002, 38:173-186.

81. Fritsche HA. Tumor markers and pattern recognition analysis: a new diagnostic tool for cancer. J Clin Ligand Assay. 2002, 25:11-15.

82. Bao L and Sun Z. Identifying genes related to drug anticancer mechanisms using support vector machine. FEBS letters. 2002, 521:109-114.

83. Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES and Golub TR. Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci USA. 2001, 98:15149-15154.

84. Chan K, Lee TW, Sample PA, Goldbaum MH, Weinreb RN and Sejnowski TJ, Comparison of machine learning and traditional classifiers in glaucoma diagnosis. IEEE T Biomed Eng. 2002, 49:963-974.

85. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M and Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics. 2000, 16:906-914.

86. Pavlidis P, Weston J, Cai JS and Noble WS. Learning gene functional classifications from multiple data types. J Comput Biol. 2002, 9:401-411.

87. Brown MPS, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M Jr and Haussler D. Knowledge-based analysis of microarray gene expression data

by using support vector machines. Proc Natl Acad Sci USA. 2000, 97:262-267.

88. Burbidge R, Trotter M, Buxton B and Holden S. Drug design by machine learning: support vector machines for pharmaceutical data analysis. Comput Chem. 2001, 26:5-14.

89. Cai YD, Liu XJ, Xu XB and Chou KC. Support Vector Machines for Predicting HIV Protease Cleavage Sites in Protein. J Comput Chem. 2002, 23:267-274.

90. Wacher VJ, Salphati L and Benet LZ. Active secretion and enterocytic drug metabolism barriers to drug absorption. Adv Drug Del Rev. 1996, 20:99-112.

91. Ding CH and Dubchak I. Multi-class protein fold recognition using support vector machines and neural networks, Bioinformatics. 2001, 17:349-358.

92. Baldi P, Brunak S, Chauvin Y, Anderson CAF and Nielsen H. Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics. 2000, 16:412-424.

93. Roulston JE. Screening with Tumor Markers. Mol Biotechnol. 2002, 20:153-162.

94. Buck A. Phytotherapy for the prostate. Br J Urol. 1996, 78:325-336.

95. Shoskes DA. Phytotherapy and other alternative forms of care for the patient with prostatitis. Curr Urol Rep. 2002, 3(4):330-334.

96. Lowe F and KU J. Phytotherapy in treatment of benign prostatic hyperplasia: a critical review. Urology. 1996, 48(1):12-20.

97. Wilt T, Ishani A and Mac Donald R. *Serenoa repens* for benign prostatic hyperplasia (Cochrane Review). Cochrane Database Syst Rev. 2002, 3:CD001423.

98. Debruyne F, Koch G, Boyle P, Da Silva FC, Gillenwater JG, Hamdy FC, Perrin P, Teillac P, Vela-Navarrete R and Raynaud JP. Comparison of a Phytotherapeutic Agent (Permixon) with an alpha-Blocker (Tamsulosin) in the Treatment of Benign Prostatic Hyperplasia: A 1-Year Randomized International Study. Eur Urol. 2002, 41(5):497-507.

99. Koch E. Extracts from fruits of *saw palmetto (Sabal serrulata*) and roots of stinging nettle (*Urtica dioica*): viable alternatives in the medical treatment of benign prostatic hyperplasia and associated lower urinary tracts symptoms. Planta Med. 2001, 67(6):489-500.

100. Boyle P, Robertson C, Lowe F and Roehrborn C. Meta-analysis of clinical trials of permixon in the treatment of symptomatic benign prostatic hyperplasia. Urology. 2000, 55(4):533-539.

101. Wilt TJ, Ishani A, Stark G, MacDonald R, Lau J and Mulrow C. *Saw palmetto* extracts for treatment of benign prostatic hyperplasia: a systematic review. JAMA. 1998, 280(18):1604-1609.

102. Gerber GS, Zagaja GP, Bales GT, Chodak GW and Contreras BA. *Saw palmetto (Serenoa repens*) in men with lower urinary tract symptoms: effects on urodynamic parameters and voiding symptoms. Urology. 1998, 51(6):1003-1007.

103. Plosker GL and Brogden RN. *Serenoa repens* (Permixon). A review of its pharmacology and therapeutic efficacy in benign prostatic hyperplasia. Drugs Aging. 1996, 9(5):379-395.

104. Di Silverio F, Flammia GP, Sciarra A, Caponera M, Mauro M, Buscarini M,

Tavani M and D'Eramo G. Plant extracts in BPH. Minerva Urol Nefrol. 1993, 45(4):143-149.

105. Di Silverio F, Monti S, Sciarra A, Varasano PA, Martini C, Lanzara S, D'Eramo G, Di Nicola S and Toscano V. Effects of long-term treatment with *Serenoa repens* (Permixon) on the concentrations and regional distribution of androgens and epidermal growth factor in benign prostatic hyperplasia. Prostate. 1998, 37(2):77-83.

106. Di Silverio F, D'Eramo G, Lubrano C, Flammia GP, Sciarra A, Palma E, Caponera M and Sciarra F. Evidence that *Serenoa repens* extract displays an antiestrogenic activity in prostatic tissue of benign prostatic hypertrophy patients. Eur Urol. 1992, 21(4):309-314.

107. Paubert-Braquet M, Richardson FO, Servent-Saez N, Gordon WC, Monge MC, Bazan NG, Authie D and Braquet P. Effect of *Serenoa repens* extract (Permixon) on estradiol/testosterone-induced experimental prostate enlargement in the rat. Pharmacol Res. 1996, 34(3-4):171-179.

108. Carraro JC, Raynaud JP, Koch G, Chisholm GD, Di Silverio F, Teillac P, Da Silva FC, Cauquil J, Chopin DK, Hamdy FC, Hanus M, Hauri D, Kalinteris A, Marencak J, Perier A and Perrin P. Comparison of phytotherapy (Permixon) with finasteride in the treatment of benign prostate hyperplasia: a randomized international study of 1,098 patients. Prostate. 1996, 29(4):231-240.

109. Bayne CW, Donnelly F, Ross M and Habib FK. *Serenoa repens* (Permixon): a 5alpha-reductase types I and II inhibitor-new evidence in a coculture model of

BPH. Prostate. 1999, 40(4):232-241.

110. Delos S, Iehle C, Martin PM and Raynaud JP. Inhibition of the activity of 'basic' 5 alpha-reductase (type 1) detected in DU 145 cells and expressed in insect cells. J Steroid Biochem Mol Biol. 1994, 48(4):347-352.

111. Sultan C, Terraza A, Devillier C, Carilla E, Briley M, Loire C and Descomps B. Inhibition of androgen metabolism and binding by a liposterolic extract of "Serenoa repens B" in human foreskin fibroblasts. J Steroid Biochem. 1984, 20(1):515-519.

112. Vacherot F, Azzouz M, Gil-Diez-De-Medina S, Colombel M, De La Taille A, Lefrere Belda MA, Abbou CC, Raynaud JP and Chopin DK. Induction of apoptosis and inhibition of cell proliferation by the lipido-sterolic extract of *Serenoa repens* (LSESr, Permixon in benign prostatic hyperplasia. Prostate. 2000, 45(3):255-266.

113. Paubert-Braquet M, Cousse H, Raynaud JP, Mencia-Huerta JM and Braquet P. Effect of the lipidosterolic extract of *Serenoa repens* (Permixon) and its major components on basic fibroblast growth factor-induced proliferation of cultures of human prostate biopsies. Eur Urol. 1998, 33(3):340-347.

114. Vacher P, Prevarskaya N, Skryma R, Audy MC, Vacher AM, Odessa MF and Dufy B. The Lipidosterolic Extract fromSerenoa repens Interferes with Prolactin Receptor Signal Transduction. J Biomed Sci. 1995, 2(4):357-365.

115. Paubert-Braquet M, Mencia Huerta JM, Cousse H and Braquet P. Effect of the lipidic lipidosterolic extract of *Serenoa repens* (Permixon) on the ionophore

A23187-stimulated production of leukotriene B4 (LTB4) from human polymorphonuclear neutrophils. Prostaglandins Leukot Essent Fatty Acids. 1997, 57(3):299-304.

116. Ravenna L, Di Silverio F, Russo MA, Salvatori L, Morgante E, Morrone S, Cardillo MR, Russo A, Frati L, Gulino A and Petrangeli E. Effects of the lipidosterolic extract of *Serenoa repens* (Permixon) on human prostatic cell lines. Prostate. 1996, 29(4):219-230.

117. Mitropoulos D, Kyroudi A, Zervas A, Papadoukakis S, Giannopoulos A, Kittas C and Karayannacos P. In vivo effect of the lipido-sterolic extract of *Serenoa repens* (Permixon) on mast cell accumulation and glandular epithelium trophism in the rat prostate. World J Urol. 2002, 19(6):457-561.

118. Shimada H, Tyler V and McLaughlin J. Biologically active acylglycerides from the berries of S*aw-palmetto* (*Serenoa repens*). J Nat Pro. 1997, 60:417-418.

119. Chen YZ and Zhi DG. Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. Proteins. 2001, 43(2):217-226.

120. Chen YZ and Ung CY. Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand-protein inverse docking approach. J Mol Graph Model. 2001, 20(3):199-218.

121. Kuntz ID, Blaney JM, Oatley SJ, Langridge R and Ferrin TE. A geometric approach to macromolecule-ligand recognitions. *J Mol Biol.* 1982, 161:269-288.

122. Richards FM. Areas, volumes, packing and protein structure. *Annu Rev Biophys*

Bioeng. 1977, 6:151-176.

123. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW and Kollman PA. A second generation force field for the simulation of proteins and nucleic acids. *J Am Chem Soc.* 1995, 117:5179-5197.

124. Wang J, Kollman PA and Kuntz ID. Flexible ligand docking: a multistep strategy approach. *Proteins.* 1999, 36:1-19.

125. Baird NC. Simulation of hydrogen bonding in biological systems: ab initio calculations for NH3-NH3 and NH3-NH4+. *Int J Quantum Chem Symp.* 1974, 1:49-53.

126. Chen YZ and Prohofsky EW. The role of a minor groove spine of hydration in stabilizing poly(dA). poly(dT) against fluctuational interbase H-bond disruption in the premelting temperature regime. *Nucleic Acids Res.* 1992, 20:415-419.

127. Chen YZ and Prohofsky EW. Premelting base pair opening probability and drug binding constant of a daunomycin - Poly d(GCAT)-Poly d(ATGC) complex. *Biophys J.* 1994, 66:820-826.

128. Carlson HA and McCammon JA. Accommodating protein flexibility in computational drug design. *Mol Pharmacol.* 2000, 57:213-218.

129. Carlson HA. Protein flexibility is an important component of structure-based drug discovery. Curr Pharm Des. 2002, 8(17):1571-8.

130. Ernst E. The risk-benefit profile of commonly used herbal therapies: Ginkgo, St. John's Wort, Ginseng, Echinacea, Saw Palmetto, and Kava. Ann Intern Med.

2000. 136:42-53.

131. Palin MF, Faguy M, LeHoux JG and Pelletier G. Inhibitory effects of *Serenoa repens* on the kinetic of pig prostatic microsomal 5alpha-reductase activity. Endocrin. 1998, 9(1):65-69.

132. Cristoni A, Pierro FD and Bombardelli E. Botanical derivatives for the prostate. Fitoterapia. 2000, 71:S21-S28.

133. Chen X, Ji ZL and Chen YZ. TTD: Therapeutic Target Database. Nucleic Acids Res. 2002, 30(1):412-415.

134. Dvorkin L and Song KY. Herbs for Benign prostatic hyperplasia. Annals of Pharmacotherapy. 2002, 36:1443-1552.

135. Goldmann WH, Sharma AL, Currier SJ, Johnston PD, Rana A and Sharma CP. *Saw palmetto* berry extract inhibits cell growth and Cox-2 expression in prostatic cancer cells. Cell Biol Int. 2001, 25(11):1117-1124.

136. Boget S, Cereser C, Parvaz P, Leriche A and Revol A. Fibroblast growth factor receptor 1 (FGFR1) is over-expressed in benign prostatic hyperplasia whereas FGFR2-IIIc and FGFR3 are not. Eur J Endocrinol. 2001, 145(3):303-310.

137. Deshmukh N, Scotson J, Dodson AR, Smith PH, Ke Y and Foster CS. Differential expression of acidic and basic fibroblast growth factors in benign prostatic hyperplasia identified by immunohistochemistry. Br J Urol. 1997, 80(6):869-874.

138. Ozen M, Giri D, Ropiquet F, Mansukhani A and Ittmann M. Role of fibroblast growth factor receptor signaling in prostate cancer cell survival. J Natl Cancer

Inst. 2001, 93(23):1783-1790.

139. Simak R, Capodieci P, Cohen DW, Fair WR, Scher H, Melamed J, Drobnjak M, Heston WD, Stix U, Steiner G and Cordon-Cardo C. Expression of c-kit and kit-ligand in benign and malignant prostatic tissues. Histol Histopathol. 2000, 15(2):365-374.

140. Savarese DM, Valinski H, Quesenberry P and Savarese T. Expression and function of colony-stimulating factors and their receptors in human prostate carcinoma cell lines. Prostate. 1998, 34(2):80-91.

141. Liu X, Wang L, Lin Y, Teng Q, Zhao C, Hu H and Chi W. Ornithine decarboxylase activity and its gene expression are increased in benign hyperplastic prostate. Prostate. 2000, 43(2):83-87.

142. Dunzendorfer U and Knoner M. Therapy with inhibitors of polyamine biosynthesis in refractory prostatic carcinoma. An experimental and clinical study. Onkologie. 1985, 8(4):196-200.

143. Haynes JM, Frydenberg M and Majewski H. Testosterone- and phorbol ester-stimulated proliferation in human cultured prostatic stromal cells. Cell Signal. 2001, 13(10):703-709.

144. Gradini R, Realacci M, Ginepri A, Naso G, Santangelo C, Cela O, Sale P, Berardi A, Petrangeli E, Gallucci M, Di Silverio F and Russo MA. Nitric oxide synthases in normal and benign hyperplastic human prostate: immunohistochemistry and molecular biology. J Pathol. 1999, 189(2):224-229.

145. Pasquali D, Thaller C and Eichele G. Abnormal level of retinoic acid in prostate

cancer tissues. J Clin Endocrinol Metab. 1996, 81(6):2186-2191.

146. He XY, Merz G, Yang YZ, Pullakart R, Mehta P, Schulz H and Yang SY. Function of human brain short chain L-3-hydroxyacyl coenzyme A dehydrogenase in androgen metabolism. Biochim Biophys Acta. 2000, 1484(2-3):267-277.

147. Mohan RR, Challa A, Gupta S, Bostwick DG, Ahmad N, Agarwal R, Marengo SR, Amini SB, Paras F, MacLennan GT, Resnick MI and Mukhtar H. Overexpression of ornithine decarboxylase in prostate cancer and prostatic fluid in humans. Clin Cancer Res. 1999, 5(1):143-147.

148. Rosenblum ER, Stauber RE, Van Thiel DH, Campbell IM and Gavaler JS. Assessment of the estrogenic activity of phytoestrogens isolated from bourbon and beer. Alcohol Clin Exp Res. 1993, 17(6):1207-1209.

149. Carnero A. Targeting the cell cycle for cancer therapy. Br J Cancer. 2002, 87(2):129-133.

150. Garner-Hamrick PA and Fisher C. Antisense phosphorothioate oligonucleotides specifically down-regulate cdc25B causing S-phase delay and persistent antiproliferative effects. Int J Cancer. 1998, 76(5):720-728.

151. Draetta G and Eckstein J. Cdc25 protein phosphatases in cell proliferation. Biochim Biophys Acta. 1997, 1332(2):M53-M63.

152. Aznar S and Lacal JC. Rho signals to cell growth and apoptosis. Cancer Lett. 2001, 165(1):1-10.

153. Erickson JW and Cerione RA. Multiple roles for Cdc42 in cell regulation. Curr Opin Cell Biol. 2001, 13(2):153-157.

154. Murant SJ, Handley J, Stower M, Reid N, Cussenot O and Maitland NJ. Co-ordinated changes in expression of cell adhesion molecules in prostate cancer. Eur J Cancer. 1997, 33(2):263-271.

155. Arenas MI, Romo E, Royuela M, Fraile B and Paniagua R. E-, N- and P-cadherin, and alpha-, beta- and gamma-catenin protein expression in normal, hyperplastic and carcinomatous human prostate. Histochem J. 2000, 32(11):659-667.

156. Szpaderska AM and Frankfater A. An Intracellular Form of Cathepsin B Contributes to Invasiveness in Cancer. Cancer Res. 2001, 61(8):3493-3500.

157. Goossens F, De Meester I, Vanhoof G and Scharpe S. Distribution of prolyl oligopeptidase in human peripheral tissues and body fluids. Eur J Clin Chem Clin Biochem. 1996, 34(1):17-22.

158. Jung K, Nowak L, Lein M, Priem F, Schnorr D and Loening SA. Matrix metalloproteinases 1 and 3, tissue inhibitor of metalloproteinase-1 and the complex of metalloproteinase-1/tissue inhibitor in plasma of patients with prostate cancer. Int J Cancer. 1997, 74(2):220-223.

159. Turner CE and Brown MC. Cell motility: ARNO and ARF6 at the cutting edge. Curr Biol. 2001, 11(21):R875-R877

160. Neumann U SNaGM. Inhibition of Human Chymase by 2-Amino-3,1-benzoxazin-4-ones. Bioorganic and Medicinal chemistry. 2002, 9:947-954.

161. de Miguel MP, Royuela M, Bethencourt FR, Santamaria L, Fraile B and Paniagua R. Immunoexpression of tumour necrosis factor-alpha and its receptors

1 and 2 correlates with proliferation/apoptosis equilibrium in normal, hyperplasic and carcinomatous human prostate. Cytokine. 2000, 12(5):535-538.

162. Kwaan HC, Wang J, Svoboda K and Declerck PJ. Plasminogen activator inhibitor 1 may promote tumour growth through inhibition of apoptosis. Br J Cancer. 2000, 82(10):1702-1708.

163. Awad AB, Williams H and Fink CS. Phytosterols reduce in vitro metastatic ability of MDA-MB-231 human breast cancer cells. Nutr Cancer. 2001, 40(2):157-164.

164. Connolly JM, Coleman M and Rose DP. Effects of dietary fatty acids on DU145 human prostate cancer cell growth in athymic nude mice. Nutr Cancer. 1997, 29(2):114-119.

165. Senzaki H, Iwamoto S, Ogura E, Kiyozuka Y, Arita S, Kurebayashi J, Takada H, Hioki K and Tsubura A. Dietary effects of fatty acids on growth and metastasis of KPL-1 human breast cancer cells in vivo and in vitro. Anticancer Res. 1998, 18(3A):1621-1627.

166. Iwamoto S, Senzaki H, Kiyozuka Y, Ogura E, Takada H, Hioki K and Tsubura A. Effects of fatty acids on liver metastasis of ACL-15 rat colon cancer cells. Nutr Cancer. 1998, 31(2):143-150.

167. Awad AB, Fink CS, Williams H and Kim U. In vitro and in vivo (SCID mice) effects of phytosterols on the growth and dissemination of human prostate cancer PC-3 cells. Eur J Cancer Prev. 2002, 10(6):507-513.

168. Awad AB, Downie A, Fink CS and Kim U. Dietary phytosterol inhibits the

growth and metastasis of MDA-MB-231 human breast cancer cells grown in SCID mice. Anticancer Res. 2000, 20(2A):821-824.

169. Sali A. 100,000 protein structures for the biologist. Nat Struct Biol. 1998, 5(12):1029-1032.