Chapter 14
Exploring the Ligand-Protein Networks in Traditional Chinese Medicine: Current Databases, Methods and Applications

Mingzhu Zhao and Dongqing Wei

Abstract While the concept of "single component–single target" in drug discovery seems to have come to an end, "Multi-component–multi-target" is considered to be another promising way out in this field. The Traditional Chinese Medicine (TCM), which has thousands of years' clinical application among China and other Asian countries, is the pioneer of the "Multi-component–multi-target" and network pharmacology. Hundreds of different components in a TCM prescription can cure the diseases or relieve the patients by modulating the network of potential therapeutic targets. Although there is no doubt of the efficacy, it is difficult to elucidate convincing underlying mechanism of TCM due to its complex composition and unclear pharmacology. Without thorough investigation of its potential targets and side effects, TCM is not able to generate large-scale medicinal benefits, especially in the days when scientific reductionism and quantification are dominant. The use of ligand-protein networks has been gaining significant value in the history of drug discovery while its application in TCM is still in its early stage. This article firstly surveys TCM databases for virtual screening that have been greatly expanded in size and data diversity in recent years. On that basis, different screening methods and strategies for identifying active ingredients and targets of TCM are outlined based on the amount of network information available, both on sides of ligand bioactivity and the protein structures. Furthermore, applications of successful *in silico* target identification attempts are discussed in details along with experiments in exploring the ligand-protein networks of TCM. Finally, it will be concluded that the prospective application of ligand-protein networks can be used not only to predict protein targets of a small molecule, but also to explore the mode of action of TCM.

M. Zhao · D. Wei (✉)
State Key Laboratory of Microbial Metabolism, College of Life Sciences
and Biotechnology, Shanghai Jiao Tong University, Shanghai, China
e-mail: dqwei@sjtu.edu.cn

14.1 Introduction

Drug discovery was once an empirical process when the effect of the medicine was purely based on phenotype readout, while the mode of action of drug molecules remained unknown. Later, reductionists began to research on the molecular mechanism of the drug-target interactions, believing that the drug is like a magic bullet towards the functioning targets [1]. This means a drug takes action on the disease by interacting with one specific therapeutic target. The idea of each drug being like a key (or ligand) matching each 'lock' (or protein) has guided the modern drug discovery practice for the last several decades. However, in the recent years, more and more evidence has shown that many drugs exert their activities by modulating multi-targets [2–4]. Besides, some drugs interact with anti-targets and induce strong side effects [5, 6]. Therefore, it is inappropriate to stick to the paradigm that drug interact with only one target. How to modulate a set of targets to achieve efficacy, while avoiding others to reduce the risk of side effects remains a central challenging task for pharmaceutical industry.

The Traditional Chinese Medicine (TCM), which has been widely used in China as well as other Asian countries for a long history, is considered to be the pioneer of the "Multi-component—multi-target" pharmacology [7, 8]. Thousands of years' clinic practices in TCM have accumulated a considerable number of formulae that exhibit reliable in vivo efficacy and safety. Based on the methodology of holism, hundreds of different components in a TCM prescription can cure the diseases or relieve the patients by modulating a serial of potential therapeutic targets [9].

In recent years, great efforts have been made on modernization of TCM, most on identification of effective ingredients and ligands in TCM formulae and functioning targets [10, 11]. Several databases of TCM formulae, ingredients and compounds with chemical structures have been established such as Traditional Chinese Medicine Database (TCMD) [12]. However, the molecular mechanisms responsible for their therapeutic effectiveness are still unclear. On one hand, experimental validation of new drug-target interactions still remains very limiting and expensive, and very few new drugs and targets are identified as clinical applications every year [13, 14]. On the other hand, the complex composition and polypharmacology of TCM make it even harder to conduct a full set of experiments between compounds and targets and elucidate the multi-target mode of action from the holistic view on the biological network level.

On the contrary, *in silico* methods can predict a large number of new drug-target interactions, construct the drug-target networks, and explore the functional mechanism underlying the multi-component drug combinations at the molecular level. In the present stage, there have already been successfully applications in interpreting the action mechanism of TCM from the perspective of drug-target networks, although the quantity is limited. Compared with the huge amounts of TCM formulae and components, only a small portion of drug-target pairs have been validated by the laborious and costly biochemical experiments. This

motivates the needs for constructing models that could predict genuine interacting pairs between ligands and targets, based on the existing small number of known ligand-target bindings.

In this article, we firstly investigate TCM databases for *in silico* methods that have been greatly expanded in size and data diversity in recent years. On that basis, different screening methods and strategies for identifying active ingredients and targets of TCM are outlined based on the amount of information available, both on sides of ligand bioactivity and the protein structures. Finally, successful applications in this area have been summarized and reviewed, including experimental and computational examples. Learning from the methods in modern western medicine (WM), different computing models and strategies can be used to confirm the effective components and related targets in TCM in order to build the ligand-target networks. One of the research directions of the modernization of TCM is to clarify the mode of action of TCM based on ligand-protein networks.

14.2 Databases for TCM

Data availability is the first consideration before any virtual screening or data mining task could be undertaken. The TCM databases can be classified in accordance to several categories, namely formulae, herbs, and compounds. The formula of TCM is a combination of herbs for treating a disease, while compounds are the bioactive molecules within herbs. In this section, we have summarized a list of databases for TCM herbs, formulations and compounds, as shown in Table 14.1.

The elementary units of TCM databases are compounds, the bioactive components that exert efficacy through binding to therapeutic targets. Most of the compounds in TCM databases have two-dimensional structure, while some of them have three-dimensional structures deduced by force filed. In most TCM databases, the information of both herbs and compounds are collected while some even have formulae information as well.

The Traditional Chinese Medicine Database (TCMD) contains 23,033 chemical constituents and over 6,760 herbs that mainly come from Yan et al. [12]. The query keywords for the database include molecular formula, substructure, botanical identity, CAS number, pharmacological activity and traditional indications. Only a small proportion of herbs in TCMD have full coverage of compounds while most have partial coverage. Chinese Herb Constituents Database (CHCD) contains information on 8,264 compounds derived from 240 commonly used herbs with both botanical and Chinese pinyin names, the part of the herbs that contain the compounds, pharmacological and toxicological information, and other useful information [15]. Qiao et al. [16] have developed 3D structural database of biochemical components which covers 10,564 constituents from 2,073 herbs with 3D structures built and optimized using the MMFF94 force field [17]. This database uses MySQL as the data engine and contains detailed information such as basic

Table 14.1 Basic information for main TCM databases

Database	Description	ULR or References
Traditional Chinese Medicine Database (TCMD)	6,760 herbs, 23,033 compounds	[12]
Chinese Herb Constituents Database (CHCD)	240 herbs, 8,264 compounds	[15]
3D structural database of biochemical components	2,073 herbs, 10,564 compounds	[16]
TCM Database@Taiwan	453 herbs, 20,000 compounds	[18]
Traditional Chinese Medicine Information Database (TCM-ID)	1,197 formulae, 1,313 herbs, ∼9,000 compounds	[19]
TCM Drugs Information System	1,712 formulae, 2,738 herbs, 16,500 compounds, 868 dietotherapy prescription	[20]
Comprehensive Herbal Medicine Information System for Cancer (CHMIS)-C	203 formulae, 900 herbs, 8,500 compounds	[21]
China Natural Products Database (CNPD)	45,055 compounds	[22]
Marine Natural Products Database (MNPD)	8,078 compounds, 3,200 with bioactivity data	[23]
Bioactive Plant Compounds Database (BPCD)	2,794 compounds	[15]
Acupuncture.com.au	TCM formulations	http://www.acupuncture.com.au/education/herbs/herbs.html
Dictionary of Chinese Herbs	TCM formulae, toxicity and side effects	http://alternativehealing.org/Chinese herbs dictionary.htm
Plants For a Future	Herb medical usage and potential side effects	http://www.pfaf.org

molecular properties, optimized 3D structures, herb origin and clinical effects. The TCM Database@Taiwan was reported to be the world's largest traditional Chinese medicine database. The web-based database contains more than 20,000 pure compounds isolated from 453 TCM herbs [18]. Both simple and advanced query methods are acceptable in terms of molecular properties, substructures, TCM ingredients and TCM classifications.

In addition to herbs and compounds, Traditional Chinese Medicine Information Database (TCM-ID) [19], TCM Drugs Information System [20], and Comprehensive Herbal Medicine Information System for Cancer (CHMIS-C) [21] also collect the information of TCM formulae. TCM-ID is developed by Zhejiang University together with National University of Singapore on all aspects of TCM

herbs. TCM-ID currently takes in 1197 TCM formulae, 1,313 herbs and around 9,000 compounds. It covers ∼4,000 disease conditions and more than half of the compounds have valid 3D structures. The data are collected from creditable TCM books as well as Journals and the records can be retried by different sets of query keywords. TCM Drugs Information System based on networks of five large databases has also been developed [20]. It includes information of 1,712 formulae, 2,738 herbs, 16,500 compounds, 868 dietotherapy prescriptions from the integration of Chinese herb database, Chinese patent medicine database, effective components database of Chinese herbs, Chinese medical dietotherapy prescription database, and Chinese medical recipe database. Herbal Medicine Information System for Cancer (CHMIS-C) integrates the information of 203 formulae that are commonly used to treat cancer clinically as well as 900 herbs and 8,500 compounds. The compounds in this database are linked to the entries in National Cancer Institute's database and drugs approved by the U.S. Food and Drug Administration.

The China Natural Products Database (CNPD) [22], Marine Natural Products Database (MNPD) [23], and Bioactive Plant Compounds Database (BPCD) [15] only focus on the structures of the compounds in TCM and do not contain pertinent information on formulae and herbs. CNPD is built to meet the needs for drug discovery using natural products including TCM and collects the 2D and 3D structures of more than 45,055 compounds. MNPD has a collection of 8,078 compounds from 10,000 marine natural products, of which 3,200 have bioactivity data. BPCD contains information on 2,794 active compounds against 78 molecular targets, as well as the subunits of the target structures to which the compounds bind.

There are other databases from the internet focusing only on the clinical efficacy or side effects of formulae and herbs, without details of compounds. Acupuncture.com.au collects the TCM formulae according to their clinical action and efficacy. Both the English and Chinese names of TCM herbs are recorded to facilitate studies using both traditional and modern methods. The Dictionary of Chinese Herbs contains information on both clinical usage and side effects of the TCM herbs. It also includes the samples of TCM formulae for treating diseases such as cancer, dengue fever, diabetes, and hepatitis B. Besides, the compatibility of TCM herbs and certain drugs are listed to provide biochemical explanation for drug designers. The Plants for a Future database allows querying of herbs with special medicinal usage, and also lists the potential side effects, medical usage, and physical characteristics.

14.3 In Silico Methods for Ligand-Protein Interactions

The computational methods for drug discovery based on ligand-protein networks have been increasingly developed and applied in the area of TCM and other drugs in recent years [7, 8]. These methods mainly fall into the territories of ligand-based

approach, target-based approach, and machine learning. Of course, these methods of predicting ligand-protein interactions are not isolated, and researchers often use them jointly to achieve better computational results, which can be easily shown in the following case studies.

14.3.1 Ligand-Based Approach

The ligand-based approach, also known as the chemical approach, is to reorganize pharmacological characteristics and protein associations, by means of ligand similarities rather than genomic space such as sequence, structural or pathway information. The basic assumption for ligand-based approach is that regardless that similar chemical structures may interact with proteins in different ways, similar ligands tend to bind to similar targets more than not [24]. The core of ligand-based approach is the calculation of chemical similarities, with the help of chemical descriptors. Before this approach can work, one need to answer how to describe molecular structures in a way that computers can recognize. Currently there are plenty of molecular descriptors to indicate the similarity of two different ligands.

14.3.1.1 Chemical Descriptors

In order to predict the ligand-target interactions, prior knowledge should be acquired in terms of the ligand information for the target [25]. By comparing the chemical structures of the new ligands against the know ligand set of a targeting protein, a threshold is usually set to decide whether the new ligand and the targeting protein can interact.

The most commonly used structure representation is the topological fingerprints that encode the sub-structural information [26]. In these fingerprints, the atom-centred feature pairs have also been proven to be very successful in many applications of virtual screening. The widely used samples of fingerprints are 2D *Daylight* [25] and *Scitegic* extended connectivity fingerprints [27] with atom types and the bond connectivity among them.

Although 2D fingerprints have been proven to be extremely robust and reliable in many chemoinformatics approaches, they seem hard to credit to be informative, therefore, consistent efforts have been made to develop more comprehensive three-dimensional fingerprints [28, 29]. 3D fingerprints encode the 3D geometry or scaffolds of molecular structures. One method to encode a compound is based on geometrical configuration of molecular structures. A common methodology in descriptors *Flexible* is the superposition of molecules onto one or multiple conformations of a reference bioactive ligand [30–32].

There are other topological descriptors based on molecular features that have been developed to compare ligand profiles. The *SHED* (SHannon Entropy Descriptors) is derived from distributions of atom-centered pairs and calculates the

variability in a feature-pair distribution [33]. Gregori-Puigjane then used the *SHED* descriptor to *in silico* profiling of 767 drugs against 684 related targets and revealed the promiscuity of the drugs targeting aminergic G protein-coupled receptors (GPCRs) [34]. On the other hand, *RED* (Renyi entropy descriptors) is another topological descriptor that measures the molecular features generalized Renyi Entropy. The scaffolds can also be used to predict the bioactivity of the compounds on target sets. In this particular research, 24,000 unique scaffolds were extracted from 458 target sets and the external test shows that to the high-priority virtual scaffolds have the predictive activities [35].

Also in the area of chemicogeneric-based predictive methods to screen ligand-target interactions Weil proposed a novel fingerprint encoding both ligand and target properties. The ligand properties are represented by common descriptors, while the cavity information of the target is incorporated by a fixed length bit string. This fingerprint shows preference to support vector machine (SVM) classifiers and the resulting precision is as high as 90 % in separating true and false pairs [36].

14.3.1.2 Similarity Coefficient

The most common way to compare molecular fingerprints for similarity analysis is by means of Tanimoto Coefficient (T_C, also known as Jaccard index) [37–39], which compares the number of bits shared between the two fingerprints to all possible matched bits between them,

$$T_C = N_{AB}/(N_A + N_B - N_{AB})$$

where N_A is number of features (ON bits) in compound A, N_B is the number of features (ON bits) in compound B, and N_{AB} is the number of features (ON bits) common to both A and B. If the Tanimoto Coefficient of two molecules is larger than 0.85, then they are considered to have a higher structurally similarity [40].

14.3.1.3 Ligand-Based Predictions

One advantage of the ligand-based similarity searching approach is that it does not need alignment between multiple molecules. The ligand-based approach describes a protein by the chemicgenomic space of its ligands. With the ligand-based descriptions of a protein, one can predict which targets are likely to be hit by a ligand, given its known structure.

In the area of ligand-based virtual screening, researchers have tried to evaluate whether novel ligand-target pairs could be identified, based on the chemical knowledge of ligands and ligand-target interactions. G protein-coupled receptors (GPCRs) are a family of effective drug targets with significant therapeutic value. Many researchers have built SVM models as well as substructural analysis to describe GPCRs from the perspective of ligand chemicogenerics [41]. Especially,

the de-orphanization of receptors without known ligands was employed using the ligands of the related receptors. For 93 % of the orphan receptors, the prediction results are better than random, while for 35 % the performance was good.

A powerful ligand-based prediction method based on features of protein ligands is the Similarity Ensemble Approach (SEA), which was originally used to investigate protein similarity based on chemical similarity between their ligand sets with the main idea that similar ligands might tend to share same targets [3]. SEA calculates Z-score and E-value by summing up the T_C over a threshold between two ligand sets as indicators to evaluate the possible interaction between two ligand sets in a way similar to BLAST. The similarity threshold for T_C is chosen in a way that the Z-score best observes the extreme value distribution (EVD). This method was then applied to predict new molecular targets for known drugs [42]. The Author investigated 3000 FDA-approved drugs against hundreds of targets and found 23 new cases of drug-target interactions. By *in vitro* experiments, five of them were validated to be positive with affinities less than 100 nM. Besides Keiser's research, SEA was also used to investigate the off-target effect of the some commercial available drugs against the target protein farnesyltransferase (PFTase) [43] and two drug loratadine and miconazole were found to be able to bind to PFTase.

The pharmacophore model is perhaps the most widely used methods that make use of the 3D structure representations of molecules [44]. A pharmacophore is defined to be the molecular features pertinent to bioactivity aligned in three dimensional spaces, including hydrogen bonding, charge transfer, electrostatic and hydrophobic interactions [45]. The underlying methodology of pharmacophore model was defined by different researchers [46]. Recently, this model was successfully applied in mesangial cell proliferation inhibitor discovery and virtual screening of potential ligands for many targets such as HIV integrase and CCR5 antagonist [47–50]. In 3D pharmacophore model, the molecular spatial features and volume constraints represent the intrinsic interactions of small bioactive ligands with protein receptors. Wolber tried to extract ligand pharmacophores from protein cavities based on a define set of six types of chemical structures [51], and develop the algorithms for ligand extraction and interpretation as well as pharmacophore creation for multiple targets.

Pharmacophore screening only considers those compounds who are direct mimics of the ligand from which the pharmacophore has been generated and may neglect the other positive binding modes as well. In fact, the pharmacophore model limits to only one mode of action for small molecules [52]. However, this limitation can be conquered by combining multiple pharmacophore models with different modes of action. This method is called Virtual Parallel Screening and has been successfully applied to the identification of Natural Products' activity [52, 53]. In such work, The PDB-based pharmacophores was firstly used for target fishing for TCM constituents. Results shown 16 constituents of *Ruta graveolens* were screened against a database of pharmacophores and good congruity was found between the potential predictions and their corresponding IC50 values.

Quantitative structure-activity relationships (QSAR) was first established in early 1960s when computational means were used to quantitatively describe

pharmacodynamics and pharmacokinetic effects in biology systems and the chemical structures of compounds [54]. Generally speaking, any mathematical model or statistical method that builds relationship between molecular structures and biological properties may be considered as QSAR. The idea of QSAR is easy while training and application of QSAR is much difficult since similar structures may interact with totally different targets due to the diversity and complexity of biology [55]. Furthermore, the intrinsic noise in data to describe both the chemical space and biological effects brings much trouble in accurate modeling [56]. Despite these difficulties, in case robust biological data is available and few outliers coexist, thousands of QSAR models have been generated and stored in related database in the past 40 years [57, 58].

14.3.2 Target-Based Approach

The target-based approach, predicts ligand-target interactions by the structural information of protein targets as well as ligands. The target-based approach depends highly on the availability of the structural information of targets, either from wet experiments or numerical simulations [59, 60]. On one hand, these methods aim to predict the conformation and orientation of the ligand within the protein cavity. On the other hand, the binding affinity of the ligand and protein is simulated with scoring functions. The main target-based approach is docking, which predicts the preferred orientation of one molecule to another when they bound to each other to form a stable complex [61]. Usually, docking is implemented to search appropriate ligands for known targets with the lowest fitting energy. On the contrary, inverse docking seeks to fish targets from known ligands 'from scratch' and also plays an important role in virtual screening.

Despite more than 20 years' research, docking and scoring ligands with proteins are still challenging processes and the performance is highly dependent on targets [62–64]. Docking cannot be applied to proteins whose 3D structures are not identified [65]. The high-resolution structure of the protein target is preferably obtained from X-ray crystallography and NMR spectroscopy. However, approximately half of the currently approved drugs bind to the membrane proteins, whose structures are extremely difficult to be acquired experimentally. Alternatively, homology modeling is usually adopted to build a putative geometry and docking cavity [66]. Besides, threading and ab initio structure prediction together with molecular dynamics (MD) and Monte Carlo simulations are utilized to predict the target structures. However, the fidelity of homology modeling, threading and ab initio structures is still questioned by many researchers. Other important challenges of docking are the dynamic behavior, the large number of degrees of freedom and the complexity of the potential energy surface. This confines docking to be a low throughput method on a very small scale, which fails to predict interactions on the level of millions of ligands and targets.

To alleviate the situation that docking depends on the nature of targets, multiple active site has been used to compensate the ligand-dependent biases and the Consensus scoring has been also suggested to reduce the false positives in virtual screening [67]. The accuracy of scoring functions still remain the main weakness of docking approach [68]. Also, docking is starting to adopt the conformation information derived from protein-bound ligands as a strategy to overcome the limitations of current scoring functions and can predict the orientation of the ligands into the protein cavity [69]. Besides, molecular dynamics-assisted docking method has been applied in virtual screening against the individual targets in HIV to search for multi-target drug-like agents and KNI-765 was identified to be potential inhibitors [70].

Regardless of the all the limitations, virtual screening based on docking and inverse docking has been successfully utilized to identify and predict novel bio-active compounds in the past 10 years. Using the combinatorial small molecule growth algorithm, Grzybowski applied the docking to the design of picomolar ligands for the human carbonic anhydrase II [71, 72]. Inverse docking was firstly developed to identify multiple proteins to which a small molecule can bind or weakly bind. In some cases, the bioactivity of the TCM compounds is well recognized, while the underlying mode of action is not very clear. In 2001, INV-DOCK [73] has been developed to search for the targets for TCM constitutes, and employed a database of protein cavities derived from PDB entries. The results of inverse docking involving multiple-conformer shape-matching alignment showed that 50 % of the computer-predicted potential protein targets were implicated or experimentally validated. The same approach was used to determine potential drug toxicity and side effects in early stages of drug development and results showed that 83 % of the experimentally known toxicity and side effects were predicted [74]. Zahler tried the inverse docking method to find potential kinase targets for three indirubin derivatives and examined 84 unique protein kinases in total [75]. Recently, one indirubin compound was found to possess therapeutic effects against myelogenous leukemia [76].

Docking is usually used as the second step to further validate the ligand-target binding features after the first round of virtual screening by other ligand-based approaches [77–80]. Wei applied the docking together with similarity search and molecular simulation to search for Anti-SAS drugs [81], find the binding mechanism of H5N1 Influenza Virus with ligands [82], detect possible drug leads for Alzheimer's Disease [83, 84] and identify the binding sites for several novel amide derivatives in the nicotinic acetylcholine receptors (AChRs) [85].

14.3.3 Machine Learning

The ligand-based approach and target-based approach predict potential ligand-target bindings by means of chemical similarity and structural information. Machine learning is a high-throughput method of artificial intelligence that enables

computers to learn from data of knowns, including ligand chemistry, structural information and ligand-protein networks and to predict unknowns, such as new drugs, targets and drug-target pairs. This method gains stability and credibility, and has strong ability for classifications among large numbers of ligand-protein pairs that otherwise would be impossible to be connected based on chemical similarity alone.

Machine learning is to exact features from data automatically by computers [86]. Basically, machine learning can be categorized into unsupervised learning and supervised learning. In unsupervised learning, the objective is to extract and conjecture patterns and interactions among a series of input variables and there is no outcome to train the input variables. The common approaches in unsupervised learning are clustering, data compression and outlier detection, such as principal component based methods [87]. In supervised learning, the objective is to predict the value of an outcome variable based on the input variables [88]. The data is commonly divided into training and validation datasets, which are used in turn to finalize a robust model. The variable the supervised model predicts is typically the binding probability of ligands and targets.

Nidhi trained a multiple-category Laplacian-modified naive Bayesian model from 964 target classes in WOMBAT and predict the top three potential targets for compounds in MDDR with or without known targets information [89]. On average, the prediction accuracy with compounds with known targets is 77 %. Bayesian classifier was usually used in early prediction, while the Winnow algorithm was reported more recently [90]. With the same training datasets, the prediction result is slightly different with the Multiple-category laplacian. This indicates that it is necessary to apply different prediction methods and make comparisons even on the same training dataset.

The Gaussian interaction profile kernels, which represented the drug-target interactions, were used in Regularized Least Squares in combined with chemical and genomic space to achieve the prediction with precision-recall curve (AUPR) up to 92.7 [91]. Based on simple physicochemical properties extracted from protein sequences, the potential drug targets were related to the existing ones by several models [92]. The supervised bipartite graph inference is used to represent the drug interaction networks and can be solely be able to predict new interactions, or together with chemical and genomic space [93, 94]. Besides, semi-supervised learning method (Laplacian regularized least square FLapRLS) was also explored to effectively predict the results by integration of genomic and chemical space [95].

The Support Vector Machine (SVM) is a powerful classification tool in which appropriate kernel functions are selected to map the data space into higher dimensional space without increasing the computational difficulties. The performance of SVM is usually stronger than other probability based models. Wale and Karypis [96] made comparisons between a Bayes Classifier together with binary SVM, cascaded SVM, a ranking-based SVM, Ranking Perception and the combination of SVM and Ranking Perception in terms of the ability to predict the targets for small compound, and found that the cascaded SVM has better

performance than the Bayes models and the combination of SVM and Ranking Perceptron has the best performance of all. Zhao et al. developed a SVM model based on the chemical-protein interactions from STITCH [97] using new features from ligand chemical space and interaction networks. Four new D-amino acid oxidase inhibitors were successfully predicted by this model and validate by wet experiments, and one may have a new application in therapy of psychiatric disorders other than being an antineoplastic agent [98].

Random forest, a form of multiple decision trees, recently has been applied to screen TCM database for potential inhibitors against several therapeutically important targets [99]. With the use of binding information from another database, random forest was performed to find multiple hits out of 8,264 compounds in 240 Chinese Herbs on an unbalanced dataset. Among all the predictions, 83 herb-target predictions were proved by literature search. Three Potential inhibitors of the human, aromatase enzyme (CYP19) myricetin, liquiritigenin and gossypetin, were screened by Random Forest as well as molecular docking studies. The virtual screening results were subsequently confirmed experimentally by *in vitro* assay [100].

Linear regression models have also been applied to predict ligand-target pairs. Zhao developed a computational framework, drugCIPHER to infer drug-target interactions based on pharmacology and genomic space [101]. In this framework, three linear regress models were created to relate drug therapeutic similarity, chemical similarity and target similarity on the basis of a protein-protein interaction network. The drugCIPHER achieved the performance with AUC of 0.988 in the training set and 0.935 in the test set and 501 new drug-target interactions were found, implying potential novel applications or side effects.

Although machine learning has strong performance in classification of protein-ligand interactions, its shortcoming is obvious. The process of some machine learning methods is implicit, like a black box, from which we cannot have an intuitive biological or physical relevance between proteins and ligands. SVM maps the classification problem into higher space, and acquires excellent performance with high computational efficiency. The tradeoff is that it can hardly explicitly create relationship between a protein and a ligand. Therefore, even with a very strong prediction tool, we can hardly move forward with innovations in theory of protein-ligand interactions.

14.3.4 Case Studies

14.3.4.1 Inhibiting Biological Transmethylation Reaction

Wei et al. focused on the discovery of potential inhibitors against S-adenosyl-homocysteine hydrolase (SAH), a key reactant in duplication of virus life cycle. A similarity search in Traditional Chinese Medicine Database was performed and 17 hits with high similarity were retrieved. Followed by docking, they proposed the

potential inhibitors by comparing best docked solutions and possible modification for the best inhibitors [79].

14.3.4.2 New D-Amino Acid Oxidase Inhibitor Discovery

Zhao et al. have developed a support vector machine (SVM) model based on the chemical-protein interactions from STITCH using new features from ligand chemical space and interaction networks. The model is used to search for the potential D-amino acid oxidase inhibitors from STITCH database and the predicted results are finally validated by wet experiments. Out of the ten candidates obtained, seven D-amino acid oxidase inhibitors have been verified, in which four are newly found, and one may have a new application in therapy of psychiatric disorders other than being an antineoplastic agent [98].

14.3.4.3 Drug Discovery for AIDs

From docking experiments for more than 9,000 compounds extracted from various Chinese medicines, Gao found that the compound agaritine distinguished itself from all the others in binding to the HIV protease with the most favorable free energy. It has been observed thru an extensive docking study that some of agaritine derivatives had markedly stronger binding interaction with the HIV protease than agaritine, suggesting that these derivatives might be good candidates for developing drugs for AIDS therapy [77].

14.3.4.4 Treating Alzheimer's Disease

To find new drug candidates for treating Alzheimer's disease, Zheng used the similarity search technique and GTS-21 as a template to search the Traditional Chinese Medicines Database. Then those molecules with higher score were selected for docking studies against the alpha7 nicotinic acetylcholine receptor. Though an in-depth structural analysis, it was found Mol 7,235 might be a promising candidate and need further experimental validation before it becomes an effective drug for treating Alzheimer's disease [78].

14.4 Applications of Ligand-Protein Networks in TCM Pharmacology

Network-based pharmacology explores the possibility to develop a systematic and holistic understanding of the mode of actions of multi-drugs by considering their multi-targets in the context of molecular networks. It has also been suggested that

relatively weak patterns of inhibition of many targets may prove more satisfactory than the highly potent single target inhibitors routinely developed in the course of a drug discovery program [102]. In drug discovery, the use of networks incorporating multiple components and the corresponding multiple targets, is one of the driving force to propel the current development in TCM pharmacology. Several successful examples have been accumulated both in experiments and *in silico* analysis, as shown in Table 14.2.

14.4.1 Experimental Study

Many bioactive compounds in TCM herbs may have synergetic effort with many non-TCM drugs in markets. Tannin, a component derived from a TCM, can be combined with HIV triple cocktail therapy to yield everlasting efforts in preventing HIV virus propagation. The underlying mechanism is that Tannin suppresses the activity of HIV-1 reverse transcriptase, protease and intergrase and cut off virus fusion and virus entry into the host cells [103]. Recently, Li proposed a new idea to induce immunetolerance in T cells by using matrine, a chemical derived from the root of Sophora flavescens AIT, targeting both the PKCy pathway and the NFAT pathway in cocktail preparations for treating AIDS [104].

Lam et al. recently showed in murine colon 38 allograft model that a formula containing 4 herbs (PHY906) has synergetic effect on reducing side effects and enhancing efficacy induced by CPT-11, a power anticancer agent with strong toxicity. The reason is that PHY906 can repair the intestinal epithelium by facilitating the intestinal progenitor or stem cells and several Wnt signaling components and suppress a batch of inflammatory responses like factor kB, cyclooxygenase-2, and inducible nitric oxide synthase [105].

Multi-component and multi-target interactions are the main mode of action for TCM formula, which exerts synergetic effects as a whole preparation rather than the primary active compound in TCM alone. Xie et al. demonstrated that other components in "Qingfu Guanjieshu" (QFGJS) could effectively influence the pharmacokinetic behavior and metabolic profile of paeonol in rats, indicating the synergy of herbal components. This synergy may be the result of enhanced adsorption of paeonol in the gastrointestinal tract induced by P-glycoprotein-mediate efflux change [106]. Another similar study, showed that paeoniflorin from the root of *Paeonia lactiflora* were markedly enhanced when co-administrated with sinomenine, the stem of *Sinomenium acutum*. Sinomenine promotes intestinal transportation via inhibition of P-glycoprotei, and affect the hydrolysis of paeoniflorin via interaction with b-glycosidase [107].

Huang-Lian-Jie-Du-Tang (HLJDT) is a TCM formula with anti-inflammatory efficacy, but the action mechanism is still not very clear. Zeng et al. investigated the effects of its component herbs and pure components on eicosanoid generation and found out the active components and their precise targets on arachidonic acid (AA) cascade. Results showed that Rhizoma coptidis and Radix scutellariae were

Table 14.2 Summary of multi-target drugs/preparations with TCM pharmacology based on ligand-protein networks

Disease	Methods and experiments	Formula, herbs and components	TCM pharmacology	References
AIDS	Experiments	Tannin	Tannin suppresses the activity of HIV-1 reverse transcriptase, protease and intergrase and cut off virus fusion and virus entry into the host cells	[103]
AIDS	Experiments	Matrine from the root of *Sophora flavescens*	Matrine is effective in inducing T cell anergy by targeting both the MAPKs pathway and the NFAT pathway	[104]
Anti-tumor	Experiments	PHY906: *Glycyrrhiza uralensis* Fisch (G), *Paeonia lactiflora* Pall (P), *Scutellaria baicalensis* Georgi (S), and *Ziziphus jujuba* Mill (Z).	PHY906 reduces CPT-11-induced gastrointestinal toxicity in the treatment of colon or rectal cancer by several mechanisms. It both repairs the intestinal epithelium by facilitating the generation of intestinal progenitor or stem cells and several Wnt signaling components and suppresses inflammatory responses like factor kB, cyclooxygenase-2, and inducible nitric oxide. synthase	[105]
Anti-inflammatory and analgesic effects	Experiments	Qingfu Guanjieshu (QFGJS): Paeonol and other components	The pharmacokinetic behavior and metabolites of paeonol are greatly promoted by other components in	[106]

(continued)

Table 14.2 (continued)

Disease	Methods and experiments	Formula, herbs and components	TCM pharmacology	References
			QFGJS. This may be the result of enhanced adsorption of paeonol in the gastrointestinal tract by P-glycoprotein-mediate efflux change	
Inflammatory and arthritic diseases	Experiments	Paeoniflorin from the root of *Paeonia lactiflora* and sinomenine from the stem of Sinomenium acutum	Paeoniflorin is markedly enhanced when co-administrated with Sinomenine, which promotes of intestinal transportation via the inhibition of P-glycoprotein, and affects the hydrolysis of Paeoniflorin via interaction with b-glycosidase	[107]
Anti-inflammatory	Experiments	Huang-Lian-Jie-Du-Tang (HLJDT): Rhizoma coptidis and Radix scutellariae	Baicalein derived from Radix scutellariae showed significant inhibitory effect on 5-LO and 15-LO while coptisine from Rhizoma coptidis showed medium inhibitory effects on LTA(4)H	[108]
Acute promyelocytic leukemia (APL)	Experiments	Realgar-Indigo naturalis: tetraarsenic tetrasulfide (A), indirubin (I), and tanshinone IIA (T)	ATI leads to ubiquitination/ degradation of promyelocytic leukemia (PML)-retinoic acid receptor oncoprotein, reprogramming of myeloid differentiation	[109]

Table 14.2 (continued)

Disease	Methods and experiments	Formula, herbs and components	TCM pharmacology	References
			regulators, and G1/G0 arrest in APL cells by mediating multiple targets. A acts as the principal component of the formula, whereas T and I serve as adjuvant ingredients	
Chronic myeloid leukemia (CML)	Experiments	Imatinib (IM) and arsenic sulfide [As(4)S(4) (AS)]	AS targets BCR/ABL through the ubiquitination of key lysine residues, leading to its proteasomal degradation, whereas IM inhibits the PI3 K/AKT/mTOR pathway	[110]
Inflammation	Pharmacophore-assisted docking	Twelve examples of compounds from CHCD	The screened compounds target cyclo-oxygenases 1 & 2 (COX), p38 MAP kinase (p38), c-Jun terminal-NH(2) kinase (JNK) and type 4 cAMP-specific phosphodiesterase (PDE4)	[111]
Type II diabetes mellitus (T2DM)	Molecular docking (LigandFit), clustering and drug-target network analysis	676 compounds in eleven herbs from Tangminling Pills	Multiple active components in Tangminling Pills interact with multiple targets. The 37 targets were classified into 3 clusters, and proteins in each cluster were highly relevant to each other. 10 known compounds were selected according to their network attribute ranking in drug-target and drug-drug network	[112]

Table 14.2 (continued)

Disease	Methods and experiments	Formula, herbs and components	TCM pharmacology	References
Cardiovascular disease	Similarity search and alignment, docking (LigandFit)	Xuefu Zhuyu decoction (XFZYD): 501 compounds, 489 drug/drug like compounds	Active components in XFZYD mainly target rennin, ACE and ACE2 in Renin-Angiotensin. System (RAS), which modulates the cardiovascular physiological function	[113]
9 types of cancer, 5 diseases with dysfunction, and 2 cardiovascular disorders	Distance-based Mutual Information Model (DMIM)	Liu-wei-di-huang formula (LWDH) Shan-zhu-yu (Fructus Corni), Ze-xie (Rhizoma Alismatis), Dan-pi (Cortex Moutan), Di-huang (Radix Rehmaniae), Fu-ling (Poria Cocos) and Shan-yao (Rhizoma Dioscoreae)	The interactions between TCM drugs and disease genes in cancer pathways and neuro-endocrine-immune pathways were inferred to contribute to the action of LWDH formula	[114]
Cardiovascular diseases	Quantitative composition-activity relationship model (QCAR) (SVM and linear regression)	*Shenmai*, Qi-Xue-Bing-Zhi-Fang (QXBZF)	The proportion of active components of *Shenmai* and QXBZF were optimized based on clinical outcome (collateral and infarct rate of heart) using QCAR. The interactions of multiple weak bindings among different compounds and targets may contribute to the synergetic effect of multi-component drugs	[115, 116]
Anticoagulant	Network-based computational scheme utilizing multi-target	Six argatroban intermediates and a series of components from	A ligand can have impact on multiple targets based on the docking scores, and	[117]

Table 14.2 (continued)

Disease	Methods and experiments	Formula, herbs and components	TCM pharmacology	References
	docking score (Ligandfit and AutoDock)	24 TCMs widely used for cardiac system diseases	those with highest target network efficiency are regarded as potential anticoagulant agents. Factor Xa and thrombin are two critical targets for anticoagulant compounds and the catalytic reactions they mediate were recognized as the most fragile biological matters in the human clotting cascade system	
Alzheimer disease	Systematical target network analysis framework	Ginkgo biloba, Huperzia serrata, Melissa officinalis and Salvia officinalis	AD symptoms-associated pathways, inflammation-associated pathways, cancer-associated pathways, diabetes mellitus associated pathways, Ca2þ-associated pathways and cell proliferation pathways are densely targeted by herbal ingredients	[118]
Depression	Literature search and network analysis	Hyperforin (HP), hypericin (HY), pseudohypericin (PH), amentoflavone (AF) and several flavonoids (FL) from St. John's Wort (SJW)	Active components in SJW mainly intervene with neuroactive ligand-receptor interaction, the calcium signaling pathway, and the gap junction related pathway	[119]
			Pertinent targets include NMDA-receptor, CRF1	

Table 14.2 (continued)

Disease	Methods and experiments	Formula, herbs and components	TCM pharmacology	References
			receptor, 5-hydroxytryptamine receptor 1D, dopamine receptor D1, etc	
Rheumatoid arthritis (RA)	Integrative Platform of TCM Network Pharmacology including drugCIPHER	Qing-Luo-Yin (QLY), including four herbs: Ku-Shen (*Sophora flavescens*), Qing-Feng-Teng (*Sinomenium acutum*), Huang-Bai (*Phellodendron chinensis*) and Bi-Xie (*Dioscorea collettii*), which contain several groups of ingredients such as Saponins and Alkaloids	The target network of QLY is involved in RA-related key processes including angiogenesis, inflammatory response, and immune response. The four herbs in QLY work in concert to promote efficiency and reduce toxicity. Specifically, the synergetic effect of Ku-Shen (*jun* herb) and Qing-Feng-Teng (*chen* herb) may come from the feedback loop and compensatory mechanisms	[120]

the key herbs responsible for the suppressive effect of HLJDT on eicosanoid generation. Further experiments on the pure components of HLJDT revealed that baicalein derived from *Radix scutellariae* has significant inhibitory effect on 5-LO and 15-LO while coptisine from *Rhizoma coptidis* show medium inhibitory effects on LTA(4)H. Besides, baicalein and coptisine were proved to have synergetic inhibition on LTB(4) by the rat peritoneal macrophages [108].

A TCM formula, Realgar-Indigo naturalis formula (RIF), was applied to treat Acute promyelocytic leukemia (APL) and showed a high complete remission (CR rate) [109]. In RIF, multiple agents within one formula were found to work synergistically. A small-scale combinational study using Chou and Talalay combination index method was performed and three main active components of RIF and six core proteins they targets in mediating the auti-tumor effect were identified. The main active ingredients of RIF are tetraarsenic tetrasulfide (A), indirubin (I), and tanshinone IIA (T), from *Realgar*, *Indigo naturalis*, and *Salvia miltiorrhiza*, respectively. A acts as the principal component of the formula, whereas T and I serve as adjuvant ingredients. ATI leads to ubiquitination/degradation of

promyelocytic leukemia (PML)-retinoic acid receptor oncoprotein, reprogramming of myeloid differentiation regulators, and G1/G0 arrest in APL cells by mediating multiple targets. Using multi-omics technologies, Zhang later proved that combined use of imatinib and arsenic sulfide from toxic herbal remedy exerted better therapeutic effects in a BCR/ABL-positive mouse model of chronic myeloid leukemia (CML) than either drug as a single agent. AS targets BCR/ABL through the ubiquitination of key lysine residues, leading to its proteasomal degradation, whereas IM inhibits the PI3 K/AKT/mTOR pathway [110].

14.4.2 Computational Framework

To target the complex, multi-factorial diseases more effectively, the network biology incorporating ligand-protein networks has been applied in multi-target drug development as well as modernization of traditional Chinese medicine in the systematic and holistic way. Zhao reviewed the available disease-associated networks, drug-associated networks that can be used to assist the drug discovery and elaborate the network-based TCM pharmacology [119]. Klipp discussed the possibility to use networks to aid the drug discovery process and focused on networks and pathways in which the components are related by physical interactions or biochemical process [121]. Leung investigated the possibility of network-based intervention for curing system diseases by means of network-based computational models and using medicinal herbs to develop into new wave of network-based multi-target drugs. It was concluded that further integration across various 'omics' platform and computational tools would accelerate the drug discovery based on network [122].

Barlow et al. screened among Chinese herbs for compounds that may be active against 4 targets in inflammation, by means of pharmacophore-assisted docking. The results showed that the twelve examples of compounds from CHCD inhibit multiple targets including cyclo-oxygenases 1 & 2 (COX), p38 MAP kinase (p38), c-Jun terminal-NH(2) kinase (JNK) and type 4 cAMP-specific phosphodiesterase (PDE4).The distribution of herbs containing the predicted active inhibitors was studied in regards to 192 Chinese Formulae and it was found that these herbs were in the formulae that were traditionally used to treat fever, headache and so on [111].

Many Traditional Chinese Medicines (TCMs) are effective to relieve complicated diseases such as type II diabetes mellitus (T2DM). Gu et al. employed the molecular docking and network analysis to elucidate the action mechanism of a medical composition-Tangminling Pills which had clinical efficacy for T2DM. It was found that multiple active components in Tangminling Pills interact with multiple targets in the biological network of T2DM. The 37 targets were classified into 3 clusters, and proteins in each cluster were highly relevant to each other. 10 known compounds were selected according to their network attribute ranking in drug-target and drug-drug network [112].

XFZYD, a recipe derives from Wang Q. R. in Qing dynasty, was widely used in cardiac system disease. From similarity search and alignment, the chemical space of compounds in XFZYD was found to share a lot of similarities with that of drug/drug-like ligands set collected from cardiovascular pharmacology while the chemical pattern in XFZYD are more diverse than drug/drug-like ligands for cardiovascular pharmacology. Docking protocol between compounds in XFZYD and targets related to cardiac system disease using LigandFit show that many molecules have good binding affinity with the targeting enzymes and most have interactions with more than one single target. The active components in XFZYD mainly target rennin, ACE and ACE2 in Renin-Angiotensin System (RAS), which modulates the cardiovascular physiological function. It was proved that promiscuous drugs in TCM might be more effective for treating cardio system diseases, which tends to result from multi-target abnormalities, but not from a single defect [113].

A lot of integrative computational tools and models have been developed and widely used to optimize the combination regimen of multi-components drugs and elucidating the interactive mechanism among ligand-target networks.

Li et al. built a method called Distance-based Mutual Information Model (DMIM) to identify useful relationships among herbs in numerous herbal formulae. DMIM combines mutual information entropy and distance between herbs to score herb interactions and construct herb network. Novel anti-angiogenic herbs, Vitexicarpin and Timosaponin A-III were discovered to have synergistic effects. Based on herb network constructed by DMIM from 3,865 collateral-related herbs, the interactions between TCM drugs and disease genes in cancer pathways and neuro-endocrine-immune pathways were inferred to contribute to the action of Liu-wei-di-huang formula, one of the most well-known TCM formula as potential treatment for a variety of diseases including cancer, dysfunction of the neuro-endocrine-immune-metabolism system and cardiovascular [114].

Wang et al. adopted a new method based upon lattice experimental design and multivariate regression to model the quantitative composition-activity relationship (QCAR) of *Shenmai*, a Chinese medicinal formula. This new strategy for multi-component drug design was then successfully applied in searching optimal combination of three key components (PD, PT and OP) of *Shenmai*. Experimental outcome of infarct rate of heart in mice with different dosage combination of the three components were finally measured and the fitted relationship equation showed that the optimal values of PD, PT and OP were 21.6, 39.2 and 39.2 %, respectively [115]. The proportion of two active components of Qi-Xue-Bing-Zhi-Fang, PF and FP, was also optimized in similar way using several fitting technique like linear regression, artificial neural network and support vector regression [116]. Although the underlying mechanism of drug synergy for the two formulae was still not very clear, the interactions of multiple weak bindings among different compounds and targets might be the contributory factors.

A network-based multi-target computational scheme for the whole efficacy of a compound in a complex disease was develop for screening the anticoagulant activities of a serial of argatroban intermediates and eight natural products respectively. Aimed at the phenotypic data of drugs, this scheme predicted

bioactive compounds by integrating biological network efficiency analysis with multi-target docking score, which evolves from the traditional virtual screening method that usually predicted binding affinity between single drug molecule and target. A ligand can have impact on multiple targets based on the docking scores, and those with highest target network efficiency are regarded as potential anti-coagulant agents. Factor Xa and thrombin are two critical targets for anticoagulant compounds and the catalytic reactions they mediate were recognized as the most fragile biological matters in the human clotting cascade system [117].

Sun et al. [118] presented a systematic target network analysis framework to explore the mode of action of anti-Alzheimer's disease (AD) herb ingredients based on applicable bioinformatics resources and methodologies on clinical anti-AD herbs and their corresponding target proteins. The results showed that just as many FDA-approved anti-AD drugs do the compounds of these herbs binds to targets in AD symptoms-associated pathway. Besides, they also interact closely with many successful therapeutic targets related to diseases such as inflammation, cancer and diameters. This suggests that the possible cross-talks between these complicated diseases are prevalent in TCM anti-AD herbs [123]. Moreover, pathways of Ca(2+) equilibrium maintaining, upstream of cell proliferation and inflammation were found to be were intensively hit by the anti-AD herbal ingredients.

Based on the available experimental results, Zhao analyzed the molecular mechanism with the aid of pathways and networks and theoretically proved the multi-target effect of St. John's Wort [119]. A comprehensive literature search was conducted and the neurotransmitter receptors, transporter proteins, and ion channels on which the SJW active compounds show effects were collected. Three main pathways that SJW intervenes were found by mapping these proteins onto KEGG pathways. Active components in SJW mainly intervene with neuroactive ligand-receptor interaction, the calcium signaling pathway, and the gap junction related pathway, pertinent to targets including NMDA-receptor, CRF1 receptor, 5-hydroxytryptamine receptor 1D, dopamine receptor D1. The networks show that the effect of SJW is similar to that of combinations of different types of antidepressants. However, the inhibitory effects of the SJW on each of the pathway are lower than other individual agents. Accordingly, the significant antidepressant efficacy and lower side effects are due to the synergetic effect of low-dose multi-target actions.

Zhang et al. established an integrative platform of TCM network pharmacology to discover herbal formulae on basis of systematic network. This platform incorporates a set of state-of-the-art network-based methods to explore the action mechanism, identify activate ingredients, and create new synergetic combinations of components. The Qing-Luo-Yin (QLY), an antirheumatoid arthritis (RA) formula was studied comprehensively using the new platform. It is found the target network of QLY is involved on RA-related key processes including angiogenesis, inflammatory response, and immune response. The four herbs in QLY work in concert to promote efficiency and reduce toxicity, as the *jun*, *chen*, *zuo*, s*hi* in Chinese, respectively. Specifically, the synergetic effect of Ku-Shen (*jun* herb) and Qing-Feng-Teng (*chen* herb) may come from the feedback loop and compensatory mechanisms [120].

14.5 Discussion and Conclusion

In recent years, the bottleneck in western medicine has brought unprecedented opportunities in TCM research and development. For decades, the fundamental research has achieved great success, and laid the foundation of modern western medicine and the philosophical idea of "reductionism" was considered to own the credit.

The counterparty of "reductionism" in Chinese medicine is the philosophical idea of holism, which has thousands years' history of practice in China as well as other Asian countries. Using this methodology, the effectiveness of TCM can only be verified from a large number of clinical trials given the unclear composition and unknown relationship among various components. This implicit effect without clear clarification at the molecular level has been hindering the modernization of TCM. How to learn from the accumulative knowledge of western medicine, in order to identify the effective compositions and explore the molecular mechanism of the efficacy is an urgent problem that needs to be solved in TCM.

The hypothesis of "multi-drug, multi-target and multi-gene" in fact bridges the gap between TCM and western medicine and is also a manifestation of unity of opposites on "reductionism" and "holism". TCM uses the holistic method to investigate the effects of multi-component formula across the whole organism, such as the use of a variety of "ZHENG" in TCM theory [124]. However, the only option we have to uncover the underlying mechanism of TCM at the molecular level is to make use of the theory of reductionism. Of course, for complex systems, the reduction method can only reach to a certain depth since it becomes more troublesome as we get deeper. Therefore, some researchers tend to reduce the mechanism of TCM to the level of "multi-drug, multi-target, multi-gene" at present, and for further reduction to the level of "single-drug, single-target, single gene", the problem of emergentism [125] in philosophy needs to be addressed properly. The theory of emergentism believes that some unique features or "ultimate features" of a system can never be reduced to properties at lower levels, nor the former can be predicted or explained by the latter, as shown in Fig. 14.1.

So far, ligand-protein network or "multi-drug, multi-target, multi-gene" is one of the few basic modules that can clearly reveal the pharmacology of TCM and is expected to be the future direction of the modernization of TCM. But just relying on experimental scientists to build ligand-protein interactions non-exhaustively will slow down both the modernization of TCM and the development of its industry. Therefore, the use of cross-platform database (TCM compounds and recipe database, see Sect. 14.2 in this paper) and the improvement on modeling technique (computational method of ligand-protein interactions, see Sect. 14.3 in this paper), will afford the basis of *in silico* research for future modernization and development of TCM. It can be foreseen that, one future direction is to use these TCM databases and predictive models to reveal the pharmacological effect of TCM, through the establishment of ligand-protein networks or, "multi-drug, multi-target, multi-gene" relationships. Nevertheless, the pharmacological mechanism of
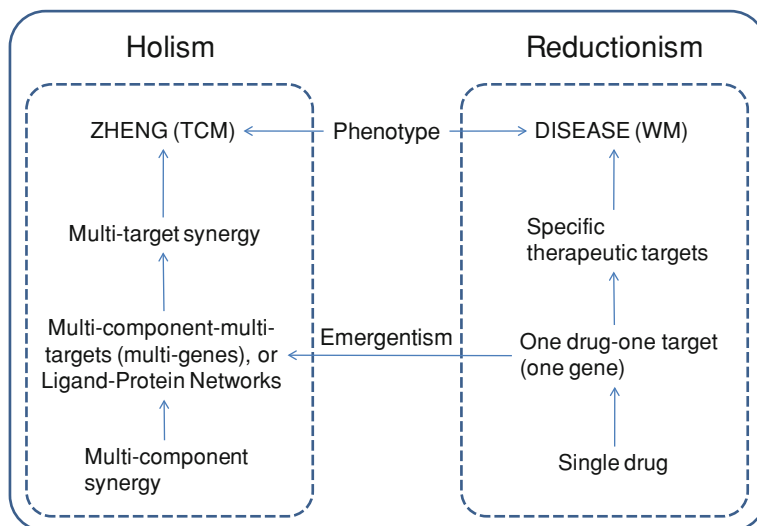
Fig. 14.1 Unity of opposites on holism in traditional Chinese medicine and reductionism in Western medicine. Emergentism constructs the framework of the understanding of holism in TCM via accumulative practice of reductionism in WM

TCM can be very complex and may not be well explained only with the known ligand-protein network. After all, this is a process of reeling silk from cocoons and also one of the best choices we have right now.

The increasing availability of ligand-protein networks is a unique chance to boost success in the modernization of TCM based on the accumulative knowledge of TCM formulae and practices based on the assumption that TCM exerts the pharmacological efficacy in multi-drug-multi-target way. Although preliminary research has been initiated in this area, there is still a long way to go to further leverage these networks and modeling techniques. Virtual screening and informatics in the drug discovery area have already been proven to be quite useful either to predict potential new drug and target candidates for experimentalists or explore the functional mechanism at the molecular level. A large number of drug-target interactions have thus been gained and the resulted drug-target networks will also be quite beneficial to investigate the underlying mechanism of multi-component drugs, such as the TCM. With further applications of these methods in TCM area, we are expecting to reveal the mode of action underlying polypharmacology of TCM. This grants us the possibility to discover novel effective drug leads, understand the synergistic mechanism of drug combinations, and more importantly, develop drug portfolios against epidemic, chronic disease, cancer and other complex disease that are almost incurable by western medicine.

References

1. Kaufmann SH (2008) Paul Ehrlich: founder of chemotherapy. Nat Rev Drug Discov 7(5):373
2. Paolini GV et al (2006) Global mapping of pharmacological space. Nat Biotechnol 24(7):805–815
3. Keiser MJ et al (2007) Relating protein pharmacology by ligand chemistry. Nat Biotechnol 25(2):197–206
4. Hopkins AL (2009) Drug discovery: predicting promiscuity. Nature 462(7270):167–168
5. Vedani A, Dobler M, Lill MA (2006) The challenge of predicting drug toxicity in silico. Basic Clin Pharmacol Toxicol 99(3):195–208
6. Klabunde T, Evers A (2005) GPCR antitarget modeling: pharmacophore models for biogenic amine binding GPCRs to avoid GPCR-mediated side effects. ChemBioChem 6(5):876–889
7. Ekins S, Mestres J, Testa B (2007) In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. Br J Pharmacol 152(1):9–20
8. Ekins S, Mestres J, Testa B (2007) In silico pharmacology for drug discovery: applications to targets and beyond. Br J Pharmacol 152(1):21–37
9. Lukman S, He Y, Hui SC (2007) Computational methods for traditional Chinese medicine: a survey. Comput Methods Programs Biomed 88(3):283–294
10. Ehrman TM, Barlow DJ, Hylands PJ (2010) Phytochemical informatics and virtual screening of herbs used in Chinese medicine. Curr Pharm Des 16(15):1785–1798
11. Feng Y et al (2006) Knowledge discovery in traditional Chinese medicine: state of the art and perspectives. Artif Intell Med 38(3):219–236
12. Yan X, Zhou J, Xu Z (1999) Concept design of computer-aided study on traditional Chinese drugs. J Chem Inf Comput Sci 39(1):86–89
13. Haggarty SJ et al (2003) Multidimensional chemical genetic analysis of diversity-oriented synthesis-derived deacetylase inhibitors using cell-based assays. Chem Biol 10(5):383–396
14. Kuruvilla FG et al (2002) Dissecting glucose signalling with diversity-oriented synthesis and small-molecule microarrays. Nature 416(6881):653–657
15. Ehrman TM, Barlow DJ, Hylands PJ (2007) Phytochemical databases of Chinese herbal constituents and bioactive plant compounds with known target specificities. J Chem Inf Model 47(2):254–263
16. Qiao X et al (2002) A 3D structure database of components from Chinese traditional medicinal herbs. J Chem Inf Comput Sci 42(3):481–489
17. Cheng A et al (2000) GB/SA water model for the Merck molecular force field (MMFF). J Mol Graph Model 18(3):273–282
18. Chen CY (2011) TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. PLoS ONE 6(1):e15939
19. Chen X et al (2006) Database of traditional Chinese medicine and its application to studies of mechanism and to prescription validation. Br J Pharmacol 149(8):1092–1103
20. Qiao X et al (2002) Research and development of traditional Chinese medicine drugs. Acta Phys Chim Sin 18:394–398
21. Fang X et al (2005) CHMIS-C: a comprehensive herbal medicine information system for cancer. J Med Chem 48(5):1481–1488
22. Shen J et al (2003) Virtual screening on natural products for discovering active compounds and target information. Curr Med Chem 10(21):2327–2342
23. Lei J, Zhou J (2002) A marine natural product database. J Chem Inf Comput Sci 42(3):742–748
24. Bender A, Glen RC (2004) Molecular similarity: a key technique in molecular informatics. Org Biomol Chem 2(22):3204–3218
25. Sastry M et al (2010) Large-scale systematic analysis of 2D fingerprint methods and parameters to improve virtual screening enrichments. J Chem Inf Model 50(5):771–784

26. Keiser MJ, Irwin JJ, Shoichet BK (2010) The chemical basis of pharmacology. Biochemistry 49(48):10267–10276
27. Shoichet BK et al (2008) Quantifying the relationships among drug classes. J Chem Inf Model 48(4):755–765
28. Koutsoukas A et al (2011) From in silico target prediction to multi-target drug design: current databases, methods and applications. J Proteomics 74(12):2554–2574
29. Rush TS 3rd et al (2005) A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. J Med Chem 48(5):1489–1495
30. Lemmen C, Lengauer T (2000) Computational methods for the structural alignment of molecules. J Comput Aided Mol Des 14(3):215–232
31. Mestres J, Veeneman GH (2003) Identification of "latent hits" in compound screening collections. J Med Chem 46(16):3441–3444
32. Jain AN (2004) Ligand-based structural hypotheses for virtual screening. J Med Chem 47(4):947–961
33. Gregori-Puigjane E, Mestres J (2006) SHED: Shannon entropy descriptors from topological feature distributions. J Chem Inf Model 46(4):1615–1622
34. Delgado-Soler L et al (2009) RED: a set of molecular descriptors based on Renyi entropy. J Chem Inf Model 49(11):2457–2468
35. Hu Y, Bajorath J (2011) Combining horizontal and vertical substructure relationships in scaffold hierarchies for activity prediction. J Chem Inf Model 51(2):248–257
36. Weill N, Rognan D (2009) Development and validation of a novel protein-ligand fingerprint to mine chemogenomic space: application to G protein-coupled receptors and their ligands. J Chem Inf Model 49(4):1049–1062
37. Willett P (1987) Similarity and clustering in chemical information systems. Chemometrics Series, Letchworth, Hertfordshire, England. Research Studies Press, Wiley, New York, 254 p (xii)
38. Willett P (2005) Searching techniques for databases of two- and three-dimensional chemical structures. J Med Chem 48(13):4183–4199
39. Willett P (2006) Similarity-based virtual screening using 2D fingerprints. Drug Discov Today 11(23–24):1046–1053
40. Brown RD, Martin YC (1996) Use of structure-activity data to compare structure-based clustering methods and descriptors for use in compound selection. J Chem Inf Comput Sci 36(3):572–584
41. van der Horst E et al (2010) A novel chemogenomics analysis of G protein-coupled receptors (GPCRs) and their ligands: a potential strategy for receptor de-orphanization. BMC Bioinform 11:316
42. Keiser MJ et al (2009) Predicting new molecular targets for known drugs. Nature 462(7270):175–181
43. DeGraw AJ et al (2010) Prediction and evaluation of protein farnesyltransferase inhibition by commercial drugs. J Med Chem 53(6):2464–2471
44. Mason JS, Good AC, Martin EJ (2001) 3-D pharmacophores in drug discovery. Curr Pharm Des 7(7):567–597
45. Maclean D et al (2000) Glossary of terms used in combinatorial chemistry. J Comb Chem 2(6):562–578
46. Langer T, Hoffmann RD (2006) Pharmacophores and pharmacophore searches. Wiley-VCH, Weinheim (John Wiley, Chichester: distributor)
47. Nicklaus MC et al (1997) HIV-1 integrase pharmacophore: discovery of inhibitors through three-dimensional database searching. J Med Chem 40(6):920–929
48. Koide Y et al (2002) Development of novel EDG3 antagonists using a 3D database search and their structure-activity relationships. J Med Chem 45(21):4629–4638
49. Debnath AK (2003) Generation of predictive pharmacophore models for CCR5 antagonists: study with piperidine- and piperazine-based compounds as a new class of HIV-1 entry inhibitors. J Med Chem 46(21):4501–4515

50. Kurogi Y et al (2001) Discovery of novel mesangial cell proliferation inhibitors using a three-dimensional database searching method. J Med Chem 44(14):2304–2307
51. Wolber G, Langer T (2005) LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J Chem Inf Model 45(1):160–169
52. Rollinger JM (2009) Accessing target information by virtual parallel screening—The impact on natural product research. Phytochem Lett 2(2):53–58
53. Rollinger JM et al (2009) In silico target fishing for rationalized ligand discovery exemplified on constituents of *Ruta graveolens*. Planta Med 75(3):195–204
54. Sharples D (1976) Factors affecting the binding of tricyclic tranquillizers and antidepressants to human serum albumin. J Pharm Pharmacol 28(2):100–105
55. Verma RP, Hansch C (2005) An approach toward the problem of outliers in QSAR. Bioorg Med Chem 13(15):4597–4621
56. Polanski J et al (2006) Modeling robust QSAR. J Chem Inf Model 46(6):2310–2318
57. Kurup A (2003) C-QSAR: a database of 18,000 QSARs and associated biological and physical data. J Comput Aided Mol Des 17(2–4):187–196
58. Hansch C et al (2002) Chem-bioinformatics: comparative QSAR at the interface between chemistry and biology. Chem Rev Columbus 102(3):783–812
59. Shoichet BK (2004) Virtual screening of chemical libraries. Nature 432(7019):862–865
60. Klebe G (2006) Virtual ligand screening: strategies, perspectives and limitations. Drug Discov Today 11(13–14):580–594
61. Lengauer T, Rarey M (1996) Computational methods for biomolecular docking. Curr Opin Struct Biol 6(3):402–406
62. Kitchen DB et al (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 3(11):935–949
63. Leach AR, Shoichet BK, Peishoff CE (2006) Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. J Med Chem 49(20):5851–5855
64. Ghoshal N, Manoharan P, Vijayan RSK (2010) Rationalizing fragment based drug discovery for BACE1: insights from FB-QSAR, FB-QSSR, multi objective (MO-QSPR) and MIF studies. J Comput Aided Mol Des 24(10):843–864
65. Cheng AC et al (2007) Structure-based maximal affinity model predicts small-molecule druggability. Nat Biotechnol 25(1):71–75
66. Evers A, Gohlke H, Klebe G (2003) Ligand-supported homology modelling of protein binding-sites using knowledge-based potentials. J Mol Biol 334(2):327–345
67. Vigers GP, Rizzi JP (2004) Multiple active site corrections for docking and virtual screening. J Med Chem 47(1):80–89
68. Gao Z et al (2008) PDTD: a web-accessible protein database for drug target identification. BMC Bioinform 9:104
69. Fradera X, Mestres J (2004) Guided docking approaches to structure-based design and screening. Curr Top Med Chem 4(7):687–700
70. Clemente JC et al (2006) Structure of the aspartic protease plasmepsin 4 from the malarial parasite Plasmodium malariae bound to an allophenylnorstatine-based inhibitor. Acta Crystallogr D Biol Crystallogr 62(Pt 3):246–252
71. Grzybowski BA et al (2002) Combinatorial computational method gives new picomolar ligands for a known enzyme. Proc Natl Acad Sci USA 99(3):1270–1273
72. Bissantz C (2003) Conformational changes of G protein-coupled receptors during their activation by agonist binding. J Recept Signal Transduct Res 23(2–3):123–153
73. Chen YZ, Zhi DG (2001) Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. Proteins 43(2):217–226
74. Chen YZ, Ung CY (2001) Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand-protein inverse docking approach. J Mol Graph Model 20(3):199–218
75. Zahler S et al (2007) Inverse in silico screening for identification of kinase inhibitor targets. Chem Biol 14(11):1207–1214

76. MacDonald ML et al (2006) Identifying off-target effects and hidden phenotypes of drugs in human cells. Nat Chem Biol 2(6):329–337
77. Gao WN et al (2007) Agaritine and its derivatives are potential inhibitors against HIV proteases. Med Chem 3(3):221–226
78. Zheng H et al (2007) Screening for new agonists against Alzheimer's disease. Med Chem 3(5):488–493
79. Wei H et al (2007) Molecular insights of SAH enzyme catalysis and implication for inhibitor design. J Theor Biol 244(4):692–702
80. Wang SQ et al (2007) Virtual screening for finding natural inhibitor against cathepsin-L for SARS therapy. Amino Acids 33(1):129–135
81. Wei DQ et al (2006) Anti-SARS drug screening by molecular docking. Amino Acids 31(1):73–80
82. Gong K et al (2009) Binding mechanism of H5N1 influenza virus neuraminidase with ligands and its implication for drug design. Med Chem 5(3):242–249
83. Gu RX et al (2009) Possible drug candidates for Alzheimer's disease deduced from studying their binding interactions with alpha7 nicotinic acetylcholine receptor. Med Chem 5(3):250–262
84. Chen SG et al (2013) Virtual screening for alpha7 nicotinic acetylcholine receptor for treatment of Alzheimer's disease. J Mol Graph Model 39:98–107
85. Arias HR et al (2011) Novel positive allosteric modulators of the human alpha7 nicotinic acetylcholine receptor. Biochemistry 50(23):5263–5278
86. Mitchell TM (1997) Machine learning. McGraw-Hill series in computer science. McGraw-Hill, New York, 414 p (xvii)
87. Strombergsson H, Kleywegt GJ (2009) A chemogenomics view on protein-ligand spaces. BMC Bioinform 10(Suppl 6):S13
88. Jensen LJ, Bateman A (2011) The rise and fall of supervised machine learning techniques. Bioinformatics 27(24):3331–3332
89. Nidhi et al (2006) Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. J Chem Inf Model 46(3):1124–1133
90. Nigsch F et al (2009) Computational toxicology: an overview of the sources of data and of modelling methods. Expert Opin Drug Metab Toxicol 5(1):1–14
91. van Laarhoven T, Nabuurs SB, Marchiori E (2011) Gaussian interaction profile kernels for predicting drug-target interaction. Bioinformatics 27(21):3036–3043
92. Li Q, Lai L (2007) Prediction of potential drug targets based on simple sequence properties. BMC Bioinform 8:353
93. Yamanishi Y et al (2008) Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics 24(13):i232–i240
94. Bleakley K, Yamanishi Y (2009) Supervised prediction of drug-target interactions using bipartite local models. Bioinformatics 25(18):2397–2403
95. Yu W et al (2011) Predicting drug-target interactions based on an improved semi-supervised learning approach. Drug Dev Res 72(2):219–224
96. Wale N, Karypis G (2009) Target fishing for chemical compounds using target-ligand activity data and ranking based methods. J Chem Inf Model 49(10):2190–2201
97. Kuhn M et al (2008) STITCH: interaction networks of chemicals and proteins. Nucleic Acids Res 36(Database issue):D684–D688
98. Zhao M et al (2014) Predicting Protein-Ligand interactions based on chemical preference features with its application to new D-Amino acid oxidase inhibitor discovery. Curr Pharm Des [epub ahead of print]
99. Ehrman TM, Barlow DJ, Hylands PJ (2007) Virtual screening of Chinese herbs with random forest. J Chem Inf Model 47(2):264–278
100. Paoletta S et al (2008) Screening of herbal constituents for aromatase inhibitory activity. Bioorg Med Chem 16(18):8466–8470
101. Zhao S, Li S (2010) Network-based relating pharmacological and genomic spaces for drug target identification. PLoS ONE 5(7):e11764

102. Li S, Zhang B (2013) Traditional Chinese medicine network pharmacology: theory, methodology and application. Chin J Nat Med 11(2):110–120
103. Borkow G, Lapidot A (2005) Multi-targeting the entrance door to block HIV-1. Curr Drug Targets Infect Disord 5(1):3–15
104. Li T et al (2009) Matrine induces cell anergy in human Jurkat T cells through modulation of mitogen-activated protein kinases and nuclear factor of activated T-cells signaling with concomitant up-regulation of anergy-associated genes expression. Biol Pharm Bull 33(1):40–46
105. Lam W et al (2010) The four-herb Chinese medicine PHY906 reduces chemotherapy-induced gastrointestinal toxicity. Sci Transl Med 2(45):45ra59
106. Xie Y et al (2008) Study on the pharmacokinetics and metabolism of paeonol in rats treated with pure paeonol and an herbal preparation containing paeonol by using HPLC-DAD-MS method. J Pharm Biomed Anal 46(4):748–756
107. Liu ZQ et al (2005) Pharmacokinetic interaction of paeoniflorin and sinomenine: pharmacokinetic parameters and tissue distribution characteristics in rats and protein binding ability in vitro. J Pharmacol Sci 99(4):381–391
108. Zeng H et al (2011) The inhibitory activities of the components of Huang-Lian-Jie-Du-Tang (HLJDT) on eicosanoid generation via lipoxygenase pathway. J Ethnopharmacol 135(2):561–568
109. Wang L et al (2008) Dissection of mechanisms of Chinese medicinal formula Realgar-Indigo naturalis as an effective treatment for promyelocytic leukemia. Proc Natl Acad Sci USA 105(12):4826–4831
110. Zhang QY et al (2009) A systems biology understanding of the synergistic effects of arsenic sulfide and Imatinib in BCR/ABL-associated leukemia. Proc Natl Acad Sci USA 106(9):3378–3383
111. Ehrman TM, Barlow DJ, Hylands PJ (2010) In silico search for multi-target anti-inflammatories in Chinese herbs and formulas. Bioorg Med Chem 18(6):2204–2218
112. Gu J et al (2011) Drug-target network and polypharmacology studies of a traditional Chinese medicine for type II diabetes mellitus. Comput Biol Chem 35(5):293–297
113. Huang Q, Qiao X, Xu X (2007) Potential synergism and inhibitors to multiple target enzymes of Xuefu Zhuyu Decoction in cardiac disease therapeutics: a computational approach. Bioorg Med Chem Lett 17(6):1779–1783
114. Li S et al (2010) Herb network construction and co-module analysis for uncovering the combination rule of traditional Chinese herbal formulae. BMC Bioinform 11(Suppl 11):S6
115. Wang Y et al (2010) A novel methodology for multicomponent drug design and its application in optimizing the combination of active components from Chinese medicinal formula Shenmai. Chem Biol Drug Des 75(3):318–324
116. Wang Y, Wang X, Cheng Y (2006) A computational approach to botanical drug design by modeling quantitative composition-activity relationship. Chem Biol Drug Des 68(3):166–172
117. Li Q et al (2011) A network-based multi-target computational estimation scheme for anticoagulant activities of compounds. PLoS ONE 6(3):e14774
118. Sun Y et al (2012) Towards a bioinformatics analysis of anti-Alzheimer's herbal medicines from a target network perspective. Brief Bioinform
119. Zhao J, Jiang P, Zhang W (2009) Molecular networks for the study of TCM pharmacology. Brief Bioinform 11(4):417–430
120. Zhang B, Wang X, Li S (2013) An integrative platform of TCM network pharmacology and its application on a herbal formula, Qing-Luo-Yin. Evid Based Complement Altern Med 2013:12
121. Klipp E, Wade RC, Kummer U (2010) Biochemical network-based drug-target prediction. Curr Opin Biotechnol 21(4):511–516
122. Leung EL et al (2012) Network-based drug discovery by integrating systems biology and computational technologies. Brief Bioinform

123. Wang T et al (2013) Inferring pathway crosstalk networks using gene set co-expression signatures. Mol BioSyst
124. Kanawong R et al (2012) Automated tongue feature extraction for ZHENG classification in traditional Chinese medicine. Evid Based Complement Alternat Med 2012:912852
125. Soto AM, Sonnenschein C (2005) Emergentism as a default: cancer as a problem of tissue organization. J Biosci 30(1):103–118