OPEN ACCESS Freely available online

PLoS one

Predicting Drug-Target Interaction Networks Based on Functional Groups and Biological Features

Zhisong He[2,5☯], Jian Zhang[3☯], Xiao-He Shi[4], Le-Le Hu[1], Xiangyin Kong[4,6*], Yu-Dong Cai[1,7*], Kuo-Chen Chou[7]

1 Institute of System Biology, Shanghai University, Shanghai, China, **2** CAS-MPG Partner Institute of Computational Biology, Shanghai Institutes for Biological Sciences (SIBS), Chinese Academy of Sciences (CAS), Shanghai, China, **3** Department of Ophthalmology, Yangpu District Central Hospital, Shanghai, China, **4** Institute of Health Sciences, Shanghai Institutes for Biological Sciences (SIBS), Chinese Academy of Sciences (CAS) and Shanghai Jiao Tong University School of Medicine (SJTUSM), Shanghai, China, **5** Centre for Computational Systems Biology, Fudan University, Shanghai, China, **6** State Key Laboratory of Medical Genomics, Ruijin Hospital, Shanghai Jiaotong University, Shanghai, China, **7** Gordon Life Science Institute, San Diego, California, United States of America

Abstract

Background: Study of drug-target interaction networks is an important topic for drug development. It is both time-consuming and costly to determine compound-protein interactions or potential drug-target interactions by experiments alone. As a complement, the in silico prediction methods can provide us with very useful information in a timely manner.

Methods/Principal Findings: To realize this, drug compounds are encoded with functional groups and proteins encoded by biological features including biochemical and physicochemical properties. The optimal feature selection procedures are adopted by means of the mRMR (Maximum Relevance Minimum Redundancy) method. Instead of classifying the proteins as a whole family, target proteins are divided into four groups: enzymes, ion channels, G-protein- coupled receptors and nuclear receptors. Thus, four independent predictors are established using the Nearest Neighbor algorithm as their operation engine, with each to predict the interactions between drugs and one of the four protein groups. As a result, the overall success rates by the jackknife cross-validation tests achieved with the four predictors are 85.48%, 80.78%, 78.49%, and 85.66%, respectively.

Conclusion/Significance: Our results indicate that the network prediction system thus established is quite promising and encouraging.

Citation: He Z, Zhang J, Shi X-H, Hu L-L, Kong X, et al. (2010) Predicting Drug-Target Interaction Networks Based on Functional Groups and Biological Features. PLoS ONE 5(3): e9603. doi:10.1371/journal.pone.0009603

Editor: Ramy K. Aziz, Cairo University, Egypt

Received December 13, 2009; **Accepted** February 16, 2010; **Published** March 11, 2010

Funding: This research is supported by the grant of National Basic Research Program of China (2004CB518603), grant from the Key Research Program (CAS) (KSCX2-YW-R-112). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: xykong@sibs.ac.cn (XK); cai_yud@yahoo.com.cn (YDC)

☯ These authors contributed equally to this work.

Introduction

Identification of drug-target interaction networks is an essential step in the drug discovery pipeline [1]. The emergence of molecular medicine and the completion of the human genome project provide more opportunity to discover unknown target proteins of drugs. Many efforts have been made to discover new drugs in the past few years. However, the number of new drug approvals remains quite low (around only 30 per year). This is partially because many compounds or drug candidates have to be withdrawn owing to unacceptable toxicity. Such failures have wasted a lot of money. It would be beneficial to develop computational methods for predicting the sensitivity and toxicity before a drug candidate was synthesized [2,3,4]. However, a number of problems need to be overcome in order to find out the exact effects of a drug. Firstly, drugs could have numerous effects including positive and negative effects, and it is hard to find out and elucidate the possible effects; secondly, different people would have completely different responses to a drug even though the same gene products are only slightly different [5,6,7,8]; thirdly, it is very hard to trace the drug effects since the biological interaction pathways are extremely complicated in human beings. Therefore, it would be very helpful for drug development if the interactions between drugs and target proteins could be predicted more accurately and the underlying mechanisms could be better understood.

Several computational approaches have been developed for analyzing and predicting drug-protein interactions. The most commonly used are docking simulations [9,10,11,12], literature text mining [13], and combining chemical structure, genomic sequence, and 3D structure information [14], among others (see, e.g., [15,16,17]).

Machine learning and data mining methods have been widely used in the computational biology and bioinformatics area. Many researchers have made lots of efforts to develop useful algorithms and softwares to investigate various drug-related biological problems, such as HIV protease cleavage site prediction [18,19], identification of GPCR (G protein-coupled receptors) type [20,21],

protein signal peptide prediction [22], protein subcellular location prediction [23,24,25], analysis of specificity of GalNAc-transferase protein [26], identification of protease type [27,28], membrane protein type prediction [29,30,31,32], and a series of relevant web-server predictors as summarized in a recent review [33].

Here we propose a predictor for drug-target interactions based on the Nearest Neighbor algorithm [34]. Since biochemical and physicochemical features [35] are important for characterizing proteins, in this study they are used to represent proteins as done by many previous investigators (see, e.g., [36,37,38]. To improve the predictor's performance, minimum Redundancy Maximum Relevance (mRMR) algorithm [39] is used to rank the features. Meanwhile, the Incremental Feature Selection and Forward Feature Selection are applied for feature selection. The protein targets for drugs are divided into enzymes, ion channels [40,41,42,43], GPCRs [44,45], and nuclear receptors [14] in this study. Finally, four predictors for predicting the interactions of drugs with each of the four protein families are developed in hopes that they can help provide useful information for drug design.

Materials and Methods

Benchmark Datasets

In addition to the dataset used by Yamanishi et al. [14], information about drug compounds and genes can be obtained from KEGG [46,47] by the FTP operations: ftp://ftp.genome.jp/pub/kegg/ligand/drug/drug for the drugs, and ftp://ftp.genome.jp/pub/kegg/genes/fasta/gene.pep for the genes. After excluding the drug-target pairs that lack experimental information, we finally obtained a total of 4,797 drug-target pairs, of which 2,719 for enzymes, 1,372 for ion channels, 630 for GPCRs, and 82 for nuclear receptors. All these datasets were used as the positive datasets in the current study.

The corresponding negative datasets were derived from the above positive datasets via the following steps: (1) separate the pairs in the above positive dataset into single drugs and proteins; (2) re-couple these singles into pairs in a way that none of them occurs in the corresponding positive dataset; (3) randomly picked the negative pairs thus formed until they reached the number two times as many as the positive pairs.

The drug-target benchmark datasets thus obtained for enzymes, ion-channels, GPCRs, and nuclear receptors are given in Online Supporting Information S1, S2, S3, and S4, respectively.

Feature Vector Construction

Representing drugs with chemical functional groups composition. The number of drugs is extremely large. However, most of them are small organic molecules and are composed of some fixed small structures, called functional groups. Since functional groups usually represent the characteristics of a compound as well as its reaction mechanism with other molecules, features derived from its functional groups could be very effective in characterizing a drug. Moreover, the number of common functional groups is quite small, and hence it is possible to use the functional group composition to uniquely represent a drug [48]. A number of functional groups are available in nature, and we selected the following 28 common groups for the current study: (1) alcohol, (2) aldehyde, (3) amide, (4) amine, (5) hydroxamic acid, (6) phosphorus, (7) carboxylate, (8) methyl, (9) ester, (10) ether, (11) imine, (12) ketone, (13) nitro, (14) halogen, (15) thiol, (16) sulfonic acid, (17) sulfone, (18) sulfonamide, (19) sulfoxide, (20) sulfide, (21) a_5c_ring, (22) ar_6c_ring, (23) non_ar_5c_ring, (24) non_ar_6c_ring, (25) hetero ar_6_ring, (26) hetero non_ar_5_ring, (27) hetero non_ar_6_ring, and (28) hetero ar_5_ring. Thus, following the same treatment as in [23], a drug compound can now be formulated as a 28-D (dimensional) vector given below:

$$\mathbb{D} = [\, g_1 \quad g_2 \quad \cdots \quad g_i \quad \cdots \quad g_{28} \,]^{\mathbf{T}} \tag{1}$$

where g_i ($i = 1, 2, \cdots, 28$) is the occurrence frequency of the i-th functional group in the drug $\mathbb{D}$, and $\mathbf{T}$ the matrix transpose operator.

Representing target proteins with pseudo amino acid composition by incorporating biochemical and physicochemical features. Now the problem is how to effectively represent a target protein. Two kinds of representations are generally used in this regard: the sequential representation and the non-sequential representation. The most typical sequential representation for a protein sample is its entire amino acid sequence, which can contain the most complete information of a protein. To deal with this model, the sequence-similarity-search-based tools, such as BLAST [49], are usually used to find the desired results. Unfortunately, this kind of approach failed to work when the query protein did not have significant homology to the proteins in the training dataset. Thus, various non-sequential representations or discrete models were proposed. The simplest discrete model was based on the amino acid composition (AAC) (see, e.g., [50]). However, if using the AAC model to represent a protein, all its sequence-order information will be lost. To avoid completely losing the sequence-order information, the pseudo amino acid composition (Pse-AAC) was proposed [36] to represent the sample of a protein. The PseAAC can be used to represent a protein sequence with a discrete model yet without completely losing its sequence-order information. For further information about PseAAC, see the web-page by clicking the link http://en.wikipedia.org/wiki/Pseudo_amino_acid_composition. Ever since the concept of PseAAC was introduced, it has been widely used to study various problems in proteins and protein-related systems (see, e.g., [37,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66]). Meanwhile, many different forms of discrete models were also proposed (see, e.g., [20,30,32,51,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82]). However, regardless of how much different these models are, they just belong to different forms of PseAAC, as elucidated in a recent comprehensive review [83]. Here, we are to propose a different PseAAC to represent drug-targeted proteins in terms of their biochemical and physicochemical features [84]. Six different types of features were considered: (1) hydrophobicity, (2) polarizability, (3) polarity, (4) secondary structure, (5) normalized van der Waals volume, and (6) solvent accessibility.

Each amino acid residue in a protein sequence can be represented by a set of different states according to its features. For instance, its hydrophobicity feature can be marked by one of the following three states: "polar", "neutral", or "hydrophobic" [85]; its solvent accessibility feature by one of the two: "buried" or "exposed to solvent", as predicted by PredAcc [35]; its secondary structure feature by one of the three: "helix", "sheet", or "coil", as predicted by the method in [86]; and so forth.

Thus, a protein sequence can be translated to a series of codes according to the biochemical and physicochemical properties of its constituent amino acid residues. For example, if using "P", "N" and "H" to represent the three states of hydrophobicity: "polar", "neutral", and "hydrophobic", the protein sequence "DMAEIMSDKPQAGML" can be translated to "PHNPHHNPPNPNNHH" according to the codes of the hydrophobic property feature. The encoded sequences thus obtained would have different length for proteins of different sizes, which will make the prediction engine difficult to handle.

To make the feature-encoded sequence to be a vector with a fixed number of dimensions, three properties of a sequence was

used: composition (C), transition (T), and distribution (D). C represents the global composition of each letter in the sequence; T, the frequency of a code letter changing from one to another; D, the distribution pattern of the code letters along the sequence, measuring the percentage of the sequence length within which the first, 25%, 50%, 75%, and 100% of the amino acids of each code letter is located. Take the above hydrophobic property sequence as an example: its C feature is 5/15 = 33.3% for all of P, H, and N, while the T feature is 2/10 = 20%, 3/10 = 30% and 5/10 = 50% for the changes between H and P, N and H, N and P, respectively. The measurement of feature D is a little more complicated. For the letter H, the first, 25%, 50%, 75% and 100% of Hs in the sequence is located at the position of 2, 5, 6, 14, and 15. Thus its D feature is (2/15 = 13.3%, 5/15 = 33.3%, 6/15 = 40%, 14/15 = 93.3%, 15/15 = 100%). In the same way, the distributions of letters P and N are (6.7%, 26.7%, 53.3%, 60%, 73.3%) and (20%, 46.7%, 66.7%, 80%, 86.7%), respectively. Accordingly, the three features of the code letter sequence are: C = (33.3%, 33.3%, 33.3%), T = (20%, 30%, 50%), and D = (13.3%, 33.3%, 40%, 93.3%, 100%, 6.7%, 26.7%, 53.3%, 60%, 73.3%, 20%, 46.7%, 66.7%, 80%, 86.7%), with a total of 21 components. Likewise, for the sequences encoded by the other four biochemical properties, each is also corresponding to 21 components. But for the sequence encoded by the solvent accessibility with only two states ("buried" or "exposed to solvent"), the encoded sequence is corresponding to only 14 components. Finally, by adding the 20 components of AAC [87] into the vector concerned, the total number of components thus obtained for a given protein is $5\times21+20+14=139$; i.e., the protein can be formulated as a 139-D vector given by

$$\mathbb{P}=[p_2 \quad p_2 \quad \cdots \quad p_i \quad \cdots \quad p_{239}]^{\mathbf{T}} \tag{2}$$

where p_i ($i=1, 2, \cdots, 139$) is the i-th component of the protein $\mathbb{P}$. Of the 139 components, 119 are derived according to the codes of the above six biochemical and physicochemical features, and 20 are the AAC components of $\mathbb{P}$.

Nearest Neighbor Algorithm

With all samples represented by a feature vector, now it is possible for us to construct our predictor using the machine learning approach. The NN (Nearest Neighbor) algorithm is quite popular in pattern recognition community owing to its good performance and simple-to-use feature. According to the NN rule [88], the query sample should be assigned to the subset represented by its nearest neighbor. In this study, if the drug-target pair with the shortest distance is a positive sample, meaning that they can interact with each other, the sample for test is seen as a positive drug-target pair. Otherwise, the test sample is seen as a negative one.

There are many different definitions to measure the "nearness" for the NN algorithm, such as Euclidean distance, Hamming distance [89], and Mahalanobis distance [50,90,91]. In the current study, the following equation was adopted to measure the nearness between samples $\mathbf{V}_x$ and $\mathbf{V}_y$

$$D(\mathbf{V}_x,\mathbf{V}_y)=1-\frac{\mathbf{V}_x\cdot\mathbf{V}_y}{\|\mathbf{V}_x\|\|\mathbf{V}_y\|} \tag{3}$$

where $\mathbf{V}_x\cdot\mathbf{V}_y$ is the dot product of the two vectors, and $\|\mathbf{V}_x\|$ and $\|\mathbf{V}_y\|$ their modulus, respectively. When $\mathbf{V}_x\equiv\mathbf{V}_y$ we have $D(\mathbf{V}_x,\mathbf{V}_y)=0$, indicating the "distance" between these two sample vectors is zero and hence they have perfect or 100% similarity.

Jackknife Cross-Validation Test

After constructing the drug-target interaction predictor, we have to evaluate its performance. In statistical prediction, the following three cross-validation methods are often used to examine a predictor for its effectiveness in practical application: independent dataset test, subsampling (K-fold cross-validation) test, and jackknife test [92]. However, as elucidated by [24] and demonstrated by Eq.50 in [93], among the three cross-validation methods, the jackknife test is deemed the most objective that can always yield a unique result for a given benchmark dataset, and hence has been increasingly used and widely recognized by investigators to examine the accuracy of various predictors (see, e.g. [51,53,54,55,56,57,59,62,63,64,94,95,96])." Accordingly, in this study the jackknife cross-validation was adopted to calculate the success prediction rates as well.

Maximum Relevance Minimum Redundancy (mRMR)

Although we've constructed the drug-target predictor based on the original feature set described above, it is possible to improve its performance with a better feature set. Apparently, not every feature in the feature set is equally relevant to the drug-target interaction. What's more, features may not be independent with each other. The "bad" will have negative impact on the accuracy and efficiency of the predictor, so it is possible to do the feature selection process to construct a more compact and effective feature set. The first step is using Maximum Relevance Minimum Redundancy (mRMR) [36] to do feature evaluation. Maximum Relevance Minimum Redundancy (mRMR) [39] was firstly developed for analysis of microarray data. It ranks each feature according to its relevance to the target and redundancy to other features. The better a feature is deemed to be, the higher the rank it will be assigned to. Mutual information (MI), denoted by I to indicate the dependence of two features used to quantify the relevance and redundancy. MI is defined as following:

$$I(x,y)=\int\int p(x,y)\log\frac{p(x,y)}{p(x)p(y)}dxdy \tag{4}$$

Based on MI, we can quantify relevance (D) and redundancy (R) as:

$$D=I(f_{\text{candidate}},c) \tag{5}$$

$$R=\frac{1}{m}\sum_{f_i\in\Omega_s} I(f_{\text{candidate}},f_i) \tag{6}$$

where $f_{\text{candidate}}$ is the feature to be calculated, and c is the target variable. By combining the above two equations to maximize relevance and minimize redundancy, the following mRMR function is constructed:

$$\max_{f_j\in\Omega_t}\left[I(f_j,c)-\frac{1}{m}\sum_{f_i\in\Omega_s} I(f_j,f_i)\right](j=1, 2, ..., n) \tag{7}$$

where Ω_s and Ω_t are the already-selected feature set and to-be-selected feature set, respectively, and m and n are the sizes of these two feature sets, respectively. The earlier a feature is selected, the better it would be though of. Finally, we can get an ordered feature list with a rank for every feature to indicate its importance in the feature set. In our study, the mRMR program is obtained from: http://research.janelia.org/peng/proj/mRMR/index.htm.

To calculate MI, the joint probabilistic density and the marginal probabilistic densities of the two vectors were used. A parameter t is introduced here to deal with these variables. Suppose mean to be the average value of one feature in all samples, and std to be the standard deviation, the feature of each sample would be classified into one of the three groups according to the boundaries: $\text{mean} \pm (t \cdot \text{std})$. In our study, t was assigned to be 1.

Incremental Feature Selection

As mentioned above, the importance of each feature is rated according to its rank in the mRMR analysis. The next step is to determine which features should be selected as the optimal feature set for our drug-target predictor. Here the IFS (Incremental Feature Selection) procedure is used to solve the problem. Each feature in the mRMR feature list was added one by one, and N different feature sets are obtained if the total feature number is N, while the i-th feature set is:

$$S_i = \{f_1, f_2, ..., f_i\} \quad (1 \le i \le N) \tag{8}$$

Based on each of the N feature sets, an NN algorithm predictor was constructed and tested with the jackknife cross-validation test. With all the N overall accurate rates calculated, we could draw an IFS curve with the index i to be the x-axis and the corresponding overall accurate rate to be the y-axis. Thus, $S_{\text{opt}} = \{f_1, f_2, ..., f_n\}$ is regarded as the optimal feature set if the curve reach its peak where the value of its x-axis is $n \le N$.

Because four independent predictors are needed for the four different classes of drug-target pairs, the IFS analysis procedure will be processed four times with each for a specific predictor.

Forward Feature Selection

To refine feature selection, the FFS (Forward Feature Selection) procedure based on the result of IFS was used. FFS is a feature selection method based on IFS results which tries every feature in the candidate feature set and adds the feature that achieves the highest prediction accuracy into the already-selected feature set in each goes. Suppose the IFS curve reaches its peak with apex as its x-axis, the initial FFS-selected feature set was constructed as:

$$S_{\text{FFS}} = \{f'_1, f'_1, ..., f'_k\} \quad (1 \le k \le \text{apex}) \tag{9}$$

More features in FFS-to-be-selected feature set would be added into the FFS-selected feature set one by one. The FFS-to-be-selected feature set with M features covers the features with mRMR ranks between $k+1$ and $k+1+M$, where M is a user-defined positive integer smaller than $N-k$ with N to be the size of the original feature set. In each round of FFS, each feature in FFS-to-be-selected feature set would be taken out and added to the FFS-selected feature set. Each predictor based on each new FFS-selected feature set would be tested, and the feature set obtained the highest overall accurate rate would be used as the new FFS-selected feature set. This process would be run for M times, until the FFS-to-be-selected feature set becomes a null set. An FFS curve similar to the IFS curve could be drawn with x-axis as the index and y-axis as the overall accurate rate.

In this study, FFS was run for each of the four benchmark datasets based on the corresponding IFS result. M for all these processes was set to 50, while k for each FFS was set to be the index of the point with the first maximum value (i.e. the maximum point with the smallest index) in the corresponding IFS curve.

Results and Discussion

Results of mRMR

To improve performance of the predictor of drug-target interaction, feature selection process was carried out. The first step of feature selection is feature evaluation. In this study, mRMR was used to evaluate every feature in original feature set. Listed in Online Supporting Information S5 are two kinds of outputs: the first one is the MaxRel list which shows ranks of features for their relevance to the target; the second is mRMR list showing the mRMR ranks according to the feature order satisfying Eq. 3. In this study, only the mRMR list was used as the results of feature evaluation. Since there are four groups of samples, mRMR was run four times with each for one of them.

Results of IFS and FFS

With the four mRMR lists, IFS was processed for each of the four sample groups, generating four IFS curves. Based on these results, we set k in FFS to be 16, 15, 14 and 19 for the data of enzymes, ion channels, GPCRs and nuclear receptors, respectively. Each of these figures is the index of the point of the first maximum value in the corresponding IFS curve. Shown in **Fig. 1** are the four IFS curves with their corresponding FFS curves. The peaks of the four FFS curves finally reach the overall success rates of 85.48% with 32 features, 80.78% with 37 features, 78.49% with 30 features, and 85.66% with 32 features for enzyme group, ion channel group, GPCR group and nuclear receptor groups, respectively.

Features selected by mRMR+FFS for the four different groups are quite different from each other, showing the intrinsic differences between them. Although there are more features for target than those for drug in the original feature set, more drug features were selected, showing the important role of drugs. Many of the selected target features are for protein secondary structure, especially for enzyme group (half of selected target features are for this). All types of features are selected in at least one group, showing that all biochemical and physicochemical features have their irreplaceable positions in drug-target interaction process.

For the details of the optimal feature-set outputs by FFS for the four benchmark datasets, see the Online Supporting Information S6.

Discussions

For the specificity and promiscuity, we divided the drug-protein interactions into four groups according to the targets of drugs: enzymes, ion channels, GPCRs, and nuclear receptors. We used all the known drugs and target proteins in the gold standard data as training data to predict the potential interactions between all human proteins annotated as members of the four classes in KEGG genes and all compounds in KEGG ligands.

Enzyme recognition is the primary event involved in the interaction of proteins with other proteins and with small molecules such as metabolites and therapeutics. Predicting drug-enzyme interactions has direct application for completing genome annotations, finding enzymes for synthetic chemistry, and predicting drug specificity, promiscuity and pharmacology. It is suggested that the secondary structure information plays the major role in determining the drug-enzyme interactions activity. For example, cytochrome P450 (CYP) induction-mediated interaction is one of the major concerns in clinical practice and for the pharmaceutical industry [97]. Induction of CYP1A enzymes with a specific structure-stable state may activate some xenobiotics to their reactive metabolites, leading to toxicity [98,99]. Amino acid composition and hydrophobicity also contribute considerably to

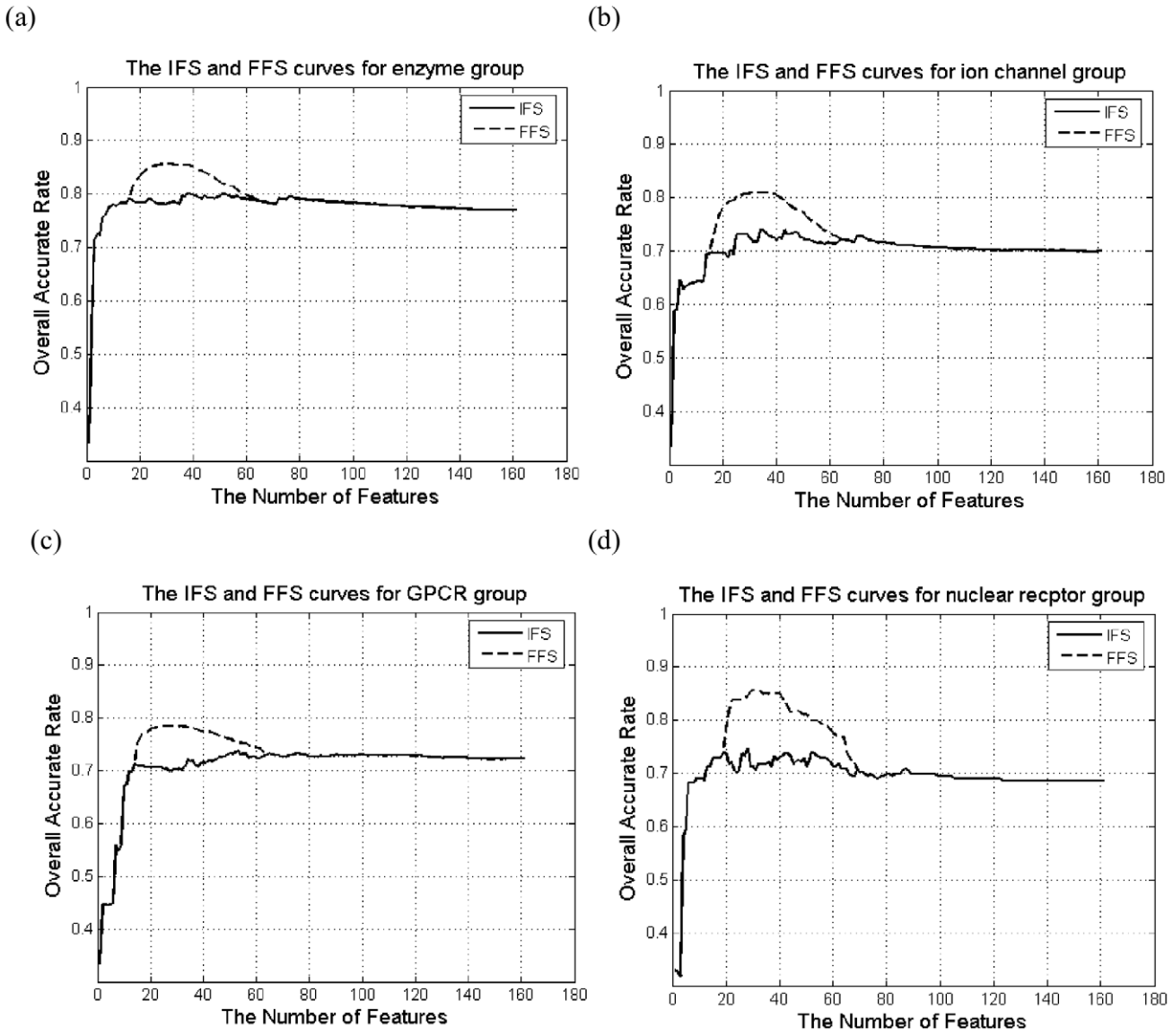

Figure 1. The IFS and FFS curves of the 4 groups. The detailed IFS curve with their corresponding FFS curve for (a) enzyme group, (b) ion channel group, (c) GPCR group, and (d) nuclear receptor group.
doi:10.1371/journal.pone.0009603.g001

these interactions. An insertion/deletion (I/D) polymorphism of the angiotensin I-converting enzyme (ACE) have an influence on the antihypertensive response, particularly when using ACE inhibitors (ACEI) [100], mirroring that the amino acid composition did contribute to the interactions. Hydrophobicity plays a role in determining the coefficients of drug-enzyme interaction energy with the application to drug screening as well as in silico target protein screening [101,102].

The G-protein coupled receptor (GPCR) superfamily, which is comprised of estimated 600–1,000 members, is another largest known class of molecular targets with varieties of physiological activities and proven therapeutic value [103]. They are integral membrane proteins sharing a common global topology that consists of seven transmembrane alpha helices, intracellular C-terminal, an extracellular N-terminal, three intracellular loops and three extracellular loops [33,44]. It is suggested that secondary structure and polarity would play a major role in determining the drug-GPCRs interactions activity. Small secondary structures such as helices and loops are identified as entities potentially involved in stabilizing interactions with ligands [33]. These motifs were situated mainly in the apical region of transmembrane segments and included a few extracellular residues [104]. Crystal structures of engineered human beta 2-adrenergic receptors (ARs) in complex with an inverse agonist ligand, carazolol, provide three-dimensional snapshots of an important G protein-coupled receptor (GPCR) with a beta-sheet structure and forms part of the chromophore-binding site [105]. GLIDA provides interaction data between GPCRs and their ligands, along with chemical information on the ligands, as well as biological information regarding GPCRs [106]. Some of the features reflect physical interactions that are responsible for the structural stability of the transmembrane, the formation of extensive networks of inter-helical H-bonds and sulfur-aromatic clusters that are spatially organized as "polarity", the close packing of side-chains throughout the transmembrane domain. When more experimental 3D structures become available for GPCRs in the future, this will help

building reliable models for a wider range of GPCRs that would be suitable for docking studies. Joint use of ligand-based chemogenomic and docking would certainly improve the prediction.

Ion channels are a large superfamily of membrane proteins that pass ions across membranes and are critical to diverse physiological functions in both excitable and nonexcitable cells and underlie many diseases. As a result, they are an important target class which is proven to be highly "druggable". According to our analysis, secondary structure and polarity play the major role in determining the drug-ion channels interactions activity. Secondary structure controls the membrane potential and interrogates ion channels in different conformational states. The drug-ion channels interaction needs gated state where they can switch conformation between a closed and an open state [42,43]. Simulations on model nanopores reveal that a narrow hydrophobic region can form a functionally closed gate in the channel and can be opened by either a small increase in pore radius or an increase in polarity [107,108]. Nowadays, intense research is being conducted to develop new drugs acting selectively on ion channel subtypes and aimed at the understanding of the intimate drug–channel interaction [109].

Nuclear receptors (NR) are ligand-activated transcription factors that regulate the activation of a variety of important target genes, which are the most important drug targets in terms of potential therapeutic application. According to our results, secondary structure and polarizability play the major role in determining the drug-NRs interactions. The conservative motif of the NR is typically described as three stacked alpha-helical sheets. The helices that make up the "front" and "back" sheets are aligned parallel to one another. The helices in the middle sheet run across the two outer sheets and only occupy the space in the upper portion of the domain. The space in the lower part of the domain is relatively void of protein, and for most NRs, this creates an internal cavity for small-molecule ligands [110]. Hydrogen bonds with polarizability activity play a crucial role in protein-drug interactions (see, e.g., [11]). Our approaches and the results thus obtained could be used to demonstrate how nuclear hormone receptors form a network of direct interactions. And this network increases in complexity to describe the interactions with target genes as well as small molecules known to bind a receptor, enzyme, or transporter.

A comprehensive drug-target interaction network system has been established that contains four classifiers for predicting the drugable interaction of compounds with enzymes, ion-channels, GPCRs, and nuclear receptors, respectively. It is anticipated that the network predictor system may become a very useful tool for drug development. Particularly it may help us find new or potential drug-target interactions.

Supporting Information

Online Supporting Information S1 The benchmark dataset for the drug-target enzyme interaction system. It contains 8,157 gene-drug pair samples, of which 2,719 are positive and 5,438 negative. The 1st column of the table indicates the nature of samples with 1 for positive and 2 for negative; the 2nd column shows the code of target gene; and the 3rd column shows the code of drug. All the detailed information for the genes and drugs listed here can be found in KEGG via their codes (Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., Hirakawa, M. From genomics to chemical genomics: new developments in KEGG Nucleic Acids Research, 2006, 34: D354-357).
Found at: doi:10.1371/journal.pone.0009603.s001 (6.30 MB DOC)

Online Supporting Information S2 The benchmark dataset for the drug-target ion channel interaction system. It contains 4,116 gene-drug pair samples, of which 1,372 are positive and 2,744 negative. The 1st column of the table indicates the nature of samples with 1 for positive and 2 for negative; the 2nd column shows the code of target gene; and the 3rd column shows the code of drug. All the detailed information for the genes and drugs listed here can be found in KEGG via their codes (see the caption of Online Supporting Information A for further explanation).
Found at: doi:10.1371/journal.pone.0009603.s002 (3.35 MB DOC)

Online Supporting Information S3 The benchmark dataset for the drug-target GPCR interaction system. It contains 1,860 gene-drug pair samples, of which 620 are positive and 1,240 negative. The 1st column of the table indicates the nature of samples with 1 for positive and 2 for negative; the 2nd column shows the code of target gene; and the 3rd column shows the code of drug. All the detailed information for the genes and drugs listed here can be found in KEGG via their codes (see the caption of Online Supporting Information A for further explanation).
Found at: doi:10.1371/journal.pone.0009603.s003 (1.53 MB DOC)

Online Supporting Information S4 The benchmark dataset for the drug-target nuclear receptor interaction system. It contains 258 gene-drug pair samples, of which 86 are positive and 172 negative. The 1st column of the table indicates the nature of samples with 1 for positive and 2 for negative; the 2nd column shows the code of target gene; and the 3rd column shows the code of drug. All the detailed information for the genes and drugs listed here can be found in KEGG via their codes (see the caption of Online Supporting Information A for further explanation).
Found at: doi:10.1371/journal.pone.0009603.s004 (0.22 MB DOC)

Online Supporting Information S5 Output of Maximum Relevancy Minimum Redundancy (mRMR).
Found at: doi:10.1371/journal.pone.0009603.s005 (1.02 MB DOC)

Online Supporting Information S6 The Results of Forward Feature Selection (FFS).
Found at: doi:10.1371/journal.pone.0009603.s006 (0.12 MB DOC)

Author Contributions

Conceived and designed the experiments: ZH JZ LH XK YDC. Performed the experiments: ZH JZ LH. Analyzed the data: XHS. Contributed reagents/materials/analysis tools: JZ YDC. Wrote the paper: ZH XHS KCC.

References

1. Knowles J, Gromo G (2003) A guide to drug discovery: Target selection in drug discovery. Nat Rev Drug Discov 2: 63–69.
2. Johnson DE, Wolfgang GH (2000) Predicting human safety: screening and computational approaches. Drug Discov Today 5: 445–454.
3. Sirois S, Hatzakis GE, Wei DQ, Du QS, Chou KC (2005) Assessment of chemical libraries for their druggability. Computational Biology & Chemistry 29: 55–67.
4. Chou KC, Wei DQ, Du QS, Sirois S, Zhong WZ (2006) Review: Progress in computational approach to drug development against SARS. Current Medicinal Chemistry 13: 3263–3270.
5. Wang JF, Wei DQ, Li L, Zheng SY, Li YX, et al. (2007) 3D structure modeling of cytochrome P450 2C19 and its implication for personalized drug design. Biochem Biophys Res Commun (Corrigendum: ibid, 2007, Vol357, 330) 355: 513–519.

6. Wang JF, Wei DQ, Chen C, Li Y, Chou KC (2008) Molecular modeling of two CYP2C19 SNPs and its implications for personalized drug design. Protein & Peptide Letters 15: 27–32.
7. Wang JF, Wei DQ, Li L, Chou KC (2008) Review: Pharmacogenomics and personalized use of drugs. Current Topics of Medicinal Chemistry 8: 1573–1579.
8. Wang JF, Zhang CC, Chou KC, Wei DQ (2009) Review: Structure of cytochrome P450s and personalized drug. Current Medicinal Chemistry 16: 232–244.
9. Cheng AC, Coleman RG, Smyth KT, Cao Q, Soulard P, et al. (2007) Structure-based maximal affinity model predicts small-molecule druggability. Nat Biotechnol 25: 71–75.
10. Rarey M, Kramer B, Lengauer T, Klebe G (1996) A fast flexible docking method using an incremental construction algorithm. J Mol Biol 261: 470–489.
11. Chou KC (2004) Review: Structural bioinformatics and its impact to biomedical science. Current Medicinal Chemistry 11: 2105–2134.
12. Chou KC, Wei DQ, Zhong WZ (2003) Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS. (Erratum: ibid., 2003, Vol.310, 675). Biochem Biophys Res Comm 308: 148–151.
13. Zhu S, Okuno Y, Tsujimoto G, Mamitsuka H (2005) A probabilistic model for mining implicit 'chemical compound-gene' relations from literature. Bioinformatics 21 Suppl 2: ii245–251.
14. Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M (2008) Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics 24: i232–240.
15. Nagamine N, Sakakibara Y (2007) Statistical prediction of protein chemical interactions based on chemical structure and mass spectrometry data. Bioinformatics 23: 2004–2012.
16. Nagamine N, Shirakawa T, Minato Y, Torii K, Kobayashi H, et al. (2009) Integrating statistical predictions and experimental verifications for enhancing protein-chemical interaction predictions in virtual screening. PLoS Comput Biol 5: e1000397.
17. Vina D, Uriarte E, Orallo F, Gonzalez-Diaz H (2009) Alignment-free prediction of a drug-target complex network based on parameters of drug connectivity and protein sequence of receptors. Mol Pharm 6: 825–835.
18. Chou KC (1993) A vectorized sequence-coupling model for predicting HIV protease cleavage sites in proteins. Journal of Biological Chemistry 268: 16938–16948.
19. Chou KC (1996) Review: Prediction of HIV protease cleavage sites in proteins. Analytical Biochemistry 233: 1–14.
20. Xiao X, Wang P, Chou KC (2009) GPCR-CA: A cellular automaton image approach for predicting G-protein-coupled receptor functional classes. Journal of Computational Chemistry 30: 1414–1423.
21. Lin WZ, Xiao X, Chou KC (2009) GPCR-GIA: a web-server for identifying G-protein coupled receptors and their families with grey incidence analysis. Protein Eng Des Sel 22: 699–705.
22. Chou KC, Shen HB (2007) Signal-CF: a subsite-coupled and window-fusing approach for predicting signal peptides. Biochem Biophys Res Comm 357: 633–640.
23. Chou KC, Cai YD (2002) Using functional domain composition and support vector machines for prediction of protein subcellular location. Journal of Biological Chemistry 277: 45765–45769.
24. Chou KC, Shen HB (2008) Cell-PLoc: A package of web-servers for predicting subcellular localization of proteins in various organisms. Nature Protocols 3: 153–162.
25. Chou KC, Shen HB (2007) Euk-mPLoc: a fusion classifier for large-scale eukaryotic protein subcellular location prediction by incorporating multiple sites. Journal of Proteome Research 6: 1728–1734.
26. Chou KC (1995) A sequence-coupled vector-projection model for predicting the specificity of GalNAc-transferase. Protein Science 4: 1365–1383.
27. Chou KC, Shen HB (2008) ProtIdent: A web server for identifying proteases and their types by fusing functional domain and sequential evolution information. Biochem Biophys Res Comm 376: 321–325.
28. Chou KC, Cai YD (2006) Prediction of protease types in a hybridization space. Biochem Biophys Res Comm 339: 1015–1020.
29. Chou KC, Elrod DW (1999) Prediction of membrane protein types and subcellular locations. PROTEINS: Structure, Function, and Genetics 34: 137–153.
30. Liu H, Wang M, Chou KC (2005) Low-frequency Fourier spectrum for predicting membrane protein types. Biochem Biophys Res Commun 336: 737–739.
31. Chou KC, Shen HB (2007) MemType-2L: A Web server for predicting membrane proteins and their types by incorporating evolution information through Pse-PSSM. Biochem Biophys Res Comm 360: 339–345.
32. Cai YD, Zhou GP, Chou KC (2003) Support vector machines for predicting membrane protein types by using functional domain composition. Biophysical Journal 84: 3257–3263.
33. Chou KC, Shen HB (2009) Review: recent advances in developing web-servers for predicting protein attributes. Natural Science 2: 63–92. (openly accessible at http://www.scirp.org/journal/NS/).
34. Denoeux T (1995) A k-nearest neighbor classification rule based on Dempster-Shafer theory. IEEE Transactions on Systems, Man and Cybernetics 25: 804–813.
35. Mucchielli-Giorgi MH, Hazout S, Tuffery P (1999) PredAcc: prediction of solvent accessibility. Bioinformatics 15: 176–177.
36. Chou KC (2001) Prediction of protein cellular attributes using pseudo amino acid composition. PROTEINS: Structure, Function, and Genetics (Erratum: ibid, 2001, Vol44, 60) 43: 246–255.
37. Chou KC (2005) Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes. Bioinformatics 21: 10–19.
38. Xiao X, Chou KC (2007) Digital coding of amino acids based on hydrophobic index. Protein & Peptide Letters 14: 871–875.
39. Peng H, Long F, Ding C (2005) Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 27: 1226–1238.
40. Chou KC (2004) Insights from modelling three-dimensional structures of the human potassium and sodium channels. Journal of Proteome Research 3: 856–861.
41. Oxenoid K, Chou JJ (2005) The structure of phospholamban pentamer reveals a channel-like architecture in membranes. Proc Natl Acad Sci U S A 102: 10870–10875.
42. Pielak RM, Jason R, Schnell JR, Chou JJ (2009) Mechanism of drug inhibition and drug resistance of influenza A M2 channel. Proceedings of National Academy of Science, USA 106: 7379–7384.
43. Schnell JR, Chou JJ (2008) Structure and mechanism of the M2 proton channel of influenza A virus. Nature 451: 591–595.
44. Chou KC (2005) Prediction of G-protein-coupled receptor classes. Journal of Proteome Research 4: 1413–1418.
45. Chou KC (2005) Coupling interaction between thromboxane A2 receptor and alpha-13 subunit of guanine nucleotide-binding protein. Journal of Proteome Research 4: 1681–1686.
46. Goto S, Nishioka T, Kanehisa M (1998) LIGAND: chemical database for enzyme reactions. Bioinformatics 14: 591–599.
47. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, et al. (2006) From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res 34: D354–357.
48. Chou KC, Cai YD, Zhong WZ (2006) Predicting networking couples for metabolic pathways of Arabidopsis. EXCLI Journal (Experimental and Clinical Sciences International Online Journal for Advances in Science) 5: 55–65.
49. Altschul SF (1997) Evaluating the statistical significance of multiple distinct local alignments. In: Suhai S, ed. Theoretical and Computational Methods in Genome Research. New York: Plenum. pp 1–14.
50. Chou KC (1995) A novel approach to predicting protein structural classes in a (20-1)-D amino acid composition space. Proteins: Structure, Function & Genetics 21: 319–344.
51. Chen C, Chen L, Zou X, Cai P (2009) Prediction of protein secondary structure content by using the concept of Chou's pseudo amino acid composition and support vector machine. Protein & Peptide Letters 16: 27–31.
52. Georgiou DN, Karakasidis TE, Nieto JJ, Torres A (2009) Use of fuzzy clustering technique and matrices to classify amino acids and its impact to Chou's pseudo amino acid composition. Journal of Theoretical Biology 257: 17–26.
53. Jiang X, Wei R, Zhang TL, Gu Q (2008) Using the concept of Chou's pseudo amino acid composition to predict apoptosis proteins subcellular location: an approach by approximate entropy. Protein & Peptide Letters 15: 392–396.
54. Li FM, Li QZ (2008) Predicting protein subcellular location using Chou's pseudo amino acid composition and improved hybrid approach. Protein & Peptide Letters 15: 612–616.
55. Lin H (2008) The modified Mahalanobis discriminant for predicting outer membrane proteins by using Chou's pseudo amino acid composition. Journal of Theoretical Biology 252: 350–356.
56. Lin H, Ding H, Feng-Biao Guo FB, Zhang AY, Huang J (2008) Predicting subcellular localization of mycobacterial proteins by using Chou's pseudo amino acid composition. Protein & Peptide Letters 15: 739–744.
57. Lin H, Wang H, Ding H, Chen YL, Li QZ (2009) Prediction of Subcellular Localization of Apoptosis Protein Using Chou's Pseudo Amino Acid Composition. Acta Biotheor 57: 321–330.
58. Qiu JD, Huang JH, Liang RP, Lu XQ (2009) Prediction of G-protein-coupled receptor classes based on the concept of Chou's pseudo amino acid composition: an approach from discrete wavelet transform. Analytical Biochemistry 390: 68–73.
59. Zeng YH, Guo YZ, Xiao RQ, Yang L, Yu LZ, et al. (2009) Using the augmented Chou's pseudo amino acid composition for predicting protein submitochondria locations based on auto covariance approach. Journal of Theoretical Biology 259: 366–372.
60. Zhang GY, Fang BS (2008) Predicting the cofactors of oxidoreductases based on amino acid composition distribution and Chou's amphiphilic pseudo amino acid composition. Journal of Theoretical Biology 253: 310–315.
61. Zhang GY, Li HC, Fang BS (2008) Predicting lipase types by improved Chou's pseudo-amino acid composition. Protein & Peptide Letters 15: 1132–1137.
62. Zhou XB, Chen C, Li ZC, Zou XY (2007) Using Chou's amphiphilic pseudo-amino acid composition and support vector machine for prediction of enzyme subfamily classes. Journal of Theoretical Biology 248: 546–551.
63. Ding YS, Zhang TL (2008) Using Chou's pseudo amino acid composition to predict subcellular localization of apoptosis proteins: an approach with immune genetic algorithm-based ensemble classifier. Pattern Recognition Letters 29: 1887–1892.

64. Ding H, Luo L, Lin H (2009) Prediction of cell wall lytic enzymes using Chou's amphiphilic pseudo amino acid composition. Protein & Peptide Letters 16: 351–355.
65. Gonzalez-Diaz H, Vilar S, Santana L, Uriarte E (2007) Medicinal chemistry and bioinformatics - current trends in drugs discovery with networks topological indices. Curr Top Med Chem 10: 1015–1029.
66. Gonzalez-Diaz H, Gonzalez-Diaz Y, Santana L, Ubeira FM, Uriarte E (2008) Proteomics, networks, and connectivity indices. Proteomics 8: 750–778.
67. Pan YX, Zhang ZZ, Guo ZM, Feng GY, Huang ZD, et al. (2003) Application of pseudo amino acid composition for predicting protein subcellular location: stochastic signal processing approach. Journal of Protein Chemistry 22: 395–402.
68. Wang M, Yang J, Liu GP, Xu ZJ, Chou KC (2004) Weighted-support vector machines for predicting membrane protein types based on pseudo amino acid composition. Protein Engineering, Design, and Selection 17: 509–516.
69. Wang M, Yang J, Xu ZJ, Chou KC (2005) SLLE for predicting membrane protein types. Journal of Theoretical Biology 232: 7–15.
70. Xiao X, Shao SH, Huang ZD, Chou KC (2006) Using pseudo amino acid composition to predict protein structural classes: approached with complexity measure factor. Journal of Computational Chemistry 27: 478–482.
71. Gao Y, Shao SH, Xiao X, Ding YS, Huang YS, et al. (2005) Using pseudo amino acid composition to predict protein subcellular location: approached with Lyapunov index, Bessel function, and Chebyshev filter. Amino Acids 28: 373–376.
72. Xiao X, Shao S, Ding Y, Huang Z, Chen X, et al. (2005) Using cellular automata to generate Image representation for biological sequences. Amino Acids 28: 29–35.
73. Xiao X, Shao SH, Ding YS, Huang ZD, Chou KC (2006) Using cellular automata images and pseudo amino acid composition to predict protein subcellular location. Amino Acids 30: 49–54.
74. Diao Y, Ma D, Wen Z, Yin J, Xiang J, et al. (2008) Using pseudo amino acid composition to predict transmembrane regions in protein: cellular automata and Lempel-Ziv complexity. Amino Acids 34: 111–117.
75. Lin H, Li QZ (2007) Using Pseudo Amino Acid Composition to Predict Protein Structural Class: Approached by Incorporating 400 Dipeptide Components. Journal of Computational Chemistry 28: 1463–1466.
76. Xiao X, Wang P, Chou KC (2008) Predicting protein structural classes with pseudo amino acid composition: an approach using geometric moments of cellular automaton image. Journal of Theoretical Biology 254: 691–696.
77. Xiao X, Lin WZ, Chou KC (2008) Using grey dynamic modeling and pseudo amino acid composition to predict protein structural classes. Journal of Computational Chemistry 29: 2018–2024.
78. Cai YD, Chou KC (2006) Predicting membrane protein type by functional domain composition and pseudo amino acid composition. Journal of Theoretical Biology 238: 395–400.
79. Chou KC, Shen HB (2006) Hum-PLoc: A novel ensemble classifier for predicting human protein subcellular localization. Biochem Biophys Res Commun 347: 150–157.
80. Chou KC, Shen HB (2007) Large-scale plant protein subcellular location prediction. Journal of Cellular Biochemistry 100: 665–678.
81. Wang T, Yang J, Shen HB, Chou KC (2008) Predicting membrane protein types by the LLDA algorithm. Protein & Peptide Letters 15: 915–921.
82. Chou KC, Shen HB (2006) Predicting eukaryotic protein subcellular location by fusing optimized evidence-theoretic K-nearest neighbor classifiers. Journal of Proteome Research 5: 1888–1897.
83. Chou KC (2009) Pseudo amino acid composition and its applications in bioinformatics, proteomics and system biology. Current Proteomics 6: 262–274.
84. Dubchak I, Muchnik I, Mayor C, Dralyuk I, Kim SH (1999) Recognition of a protein fold in the context of the Structural Classification of Proteins (SCOP) classification. PROTEINS: Structure, Function, and Genetics 35: 401–407.
85. Chothia C, Finkelstein AV (1990) The classification and origins of protein folding patterns. Annu Rev Biochem 59: 1007–1039.
86. Frishman D, Argos P (1997) Seventy-five percent accuracy in protein secondary structure prediction. Proteins 27: 329–335.
87. Chou KC, Zhang CT (1994) Predicting protein folding types by distance functions that make allowances for amino acid interactions. Journal of Biological Chemistry 269: 22014–22020.
88. Keller JM, Gray MR, Givens JA (1985) A fuzzy k-nearest neighbours algorithm. IEEE Trans Syst Man Cybern 15: 580–585.
89. Mardia KV, Kent JT, Bibby JM (1979) Multivariate Analysis: Chapter 11 Discriminant Analysis; Chapter 12 Multivariate analysis of variance; Chapter 13 cluster analysis (pp. 322–381). London: Academic Press. pp 322–381.
90. Mahalanobis PC (1936) On the generalized distance in statistics. Proc Natl Inst Sci India 2: 49–55.
91. Pillai KCS (1985) Mahalanobis D2. In: Kotz S, Johnson NL, eds. Encyclopedia of Statistical Sciences. New York: John Wiley & Sons, This reference also presents a brief biography of Mahalanobis who was a man of great originality and who made considerable contributions to statistics. pp 176–181.
92. Chou KC, Zhang CT (1995) Review: Prediction of protein structural classes. Critical Reviews in Biochemistry and Molecular Biology 30: 275–349.
93. Chou KC, Shen HB (2007) Review: Recent progresses in protein subcellular location prediction. Analytical Biochemistry 370: 1–16.
94. Zhou GP (1998) An intriguing controversy over protein structural class prediction. Journal of Protein Chemistry 17: 729–738.
95. Zhou GP, Assa-Munt N (2001) Some insights into protein structural class prediction. PROTEINS: Structure, Function, and Genetics 44: 57–59.
96. Zhou GP, Doctor K (2003) Subcellular location prediction of apoptosis proteins. PROTEINS: Structure, Function, and Genetics 50: 44–48.
97. Lin JH (2006) CYP induction-mediated drug interactions: in vitro assessment and clinical implications. Pharm Res 23: 1089–1116.
98. Beresford AP (1993) CYP1A1: friend or foe? Drug Metab Rev 25: 503–517.
99. Pelkonen O, Turpeinen M, Hakkola J, Honkakoski P, Hukkanen J, et al. (2008) Inhibition and induction of human cytochrome P450 enzymes: current status. Arch Toxicol 82: 667–715.
100. Baudin B (2000) Angiotensin I-converting enzyme gene polymorphism and drug response. Clin Chem Lab Med 38: 853–856.
101. Faulon JL, Misra M, Martin S, Sale K, Sapra R (2008) Genome scale enzyme-metabolite and drug-target interaction predictions using the signature molecular descriptor. Bioinformatics 24: 225–233.
102. Cai CZ, Han LY, Ji ZL, Chen X, Chen YZ (2003) SVM-Prot: Web-based support vector machine software for functional classification of a protein from its primary sequence. Nucleic Acids Res 31: 3692–3697.
103. Bockaert J, Pin JP (1999) Molecular tinkering of G protein-coupled receptors: an evolutionary success. Embo J 18: 1723–1729.
104. Avlani VA, Gregory KJ, Morton CJ, Parker MW, Sexton PM, et al. (2007) Critical role for the second extracellular loop in the binding of both orthosteric and allosteric G protein-coupled receptor ligands. J Biol Chem 282: 25677–25686.
105. Huber T, Menon S, Sakmar TP (2008) Structural basis for ligand binding and specificity in adrenergic receptors: implications for GPCR-targeted drug discovery. Biochemistry 47: 11013–11023.
106. Okuno Y, Tamon A, Yabuuchi H, Niijima S, Minowa Y, et al. (2008) GLIDA: GPCR–ligand database for chemical genomics drug discovery–database and tools update. Nucleic Acids Res 36: D907–912.
107. Wei H, Wang CH, Du QS, Meng J, Chou KC (2009) Investigation into adamantane-based M2 inhibitors with FB-QSAR. Medicinal Chemistry 5: 305–317.
108. Huang RB, Du QS, Wang CH, Chou KC (2008) An in-depth analysis of the biological functional studies based on the NMR M2 channel structure of influenza A virus. Biochem Biophys Res Comm 377: 1243–1247.
109. Camerino DC, Tricarico D, Desaphy JF (2007) Ion channel pharmacology. Neurotherapeutics 4: 184–198.
110. Moore JT, Collins JL, Pearce KH (2006) The nuclear receptor superfamily and drug discovery. ChemMedChem 1: 504–523.