Virtual Screening of Chinese Herbs with Random Forest

Thomas M. Ehrman,[†] David J. Barlow,*[,†] and Peter J. Hylands[‡]

Pharmaceutical Sciences Research Division and Centre for Natural Medicines Research, King's College London, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NH, United Kingdom

Received July 11, 2006

Random Forest, a form of multiple decision trees, has been used to screen a database of Chinese herbal constituents for potential inhibitors against several therapeutically important molecular targets. These comprise cyclic adenosine 3′-5′-monophosphate phosphodiesterases, protein kinase A, cyclooxygenases, lipoxygenases, aldose reductase, and three HIV targets-integrase, protease, and reverse transcriptase. In addition, compounds were identified which may inhibit the expression of inducible nitric oxide synthase and/or nitric oxide production in vivo. A total of 240 Chinese herbs containing 8264 compounds were screened in silico, including many used on a regular basis in traditional Chinese medicine. Active compounds were selected from another database of 2597 phytochemicals and related natural products with known target affinities and covered a wide range of structural classes. Random Forest was found to perform well, even on highly unbalanced data characteristic of ligand-based screening where the compounds to be screened are far more numerous than the number of active compounds used in training. Despite a conservative screening protocol, a wide variety of compounds from Chinese herbs were hit. Of particular interest were the relatively large number of herbs predicted to inhibit multiple targets, as well as a number which appeared to contain inhibitors of the same target from different phytochemical classes. The latter point to the possibility that individual species may make use of alternative phytochemical strategies in target inhibition. A literature search provided evidence to support 83 herb−target predictions.

INTRODUCTION

The screening of chemical libraries in silico has in recent years become an increasingly effective tool in the search for bioactive compounds. Indeed, given the vast numbers of compounds generated both by nature and by synthesis, the use of computers in identifying such compounds is now an essential prerequisite to an increasingly wide range of applications. Despite the potential that such methods represent for increased understanding of the pharmacological potential of plant compounds, there have, however, been relatively few attempts so far to apply them in this context.

There are a number of reasons for this. One concerns the pre-eminence of combinatorial chemistry as the preferred method for generating new drug leads, which has in turn led to the comparative neglect of this valuable resource; though, with increased calls in recent years for further research on natural products,[1,2] there are again signs that they may play a more active role in future drug discovery. In addition, data suitable for the virtual screening of plant compounds have until recently been scarce, and there have been only limited attempts to present them in an appropriate format. In the case of Chinese herbs, at least, the situation has improved with the recent appearance of three compilations of structural data on herbal constituents.[3−5] However, though a useful resource, these data contain little information on known target affinities of Chinese herbal constituents, and in the first instance, it is necessary to resort to virtual screening in an attempt to explore and identify regions of chemical space in Chinese herbs which are most likely to possess appropriate activities, or to provide leads for further development.

Virtual screening falls into two major categories, protein-based, in which compounds are docked to the receptor of interest and a "score" then calculated from component energy terms, and ligand-based as here, in which one or more compounds with reported activity against a target are used, in the absence of a model of the receptor, to search for potentially active compounds on the basis of structural similarity.

Where there is more than one compound with known activity, as is usually the case, and those compounds are of arbitrary structural diversity, then decision trees have much to offer in identifying compounds with similar activities. Decision trees constitute a form of *discriminant analysis* in which the objective is to find as good a distinction as possible between active compounds and others, on the basis of a set of molecular descriptors. They identify features shared by different subsets of active compounds and accordingly filter out compounds within the target data set in which these combinations are lacking. Though single trees afford a simple, logical description of the model, for purposes of prediction, multiple decision trees, based on an *ensemble* of trees as opposed to a single tree, tend to be the preferred option as their use of consensus voting among trees gives them higher predictive accuracy. Though detailed discussion of their features may be found elsewhere,[6−9] their other strengths can be summarized as follows: (1) they are quick to build, (2) cross-validation, in addition to consensus,

* Corresponding author fax: 44 (0) 207 848 4800; e-mail: dave.barlow@kcl.ac.uk.
[†] Pharmaceutical Sciences Research Division.
[‡] Centre for Natural Medicines Research.

10.1021/ci600289v CCC: $37.00
Published on Web 01/09/2007

produces robust models which can be reliably extrapolated to new data, (3) they can handle an arbitrary degree of nonlinearity, and are thus suitable for modeling structurally diverse compounds, and (4) they are suitable for highly unbalanced data, given certain modifications, where members of one class (e.g., active compounds) are considerably less frequent than those of the other (e.g., inactive or unknown), which is usually the case in virtual screening.

In this paper, Random Forest (RF), one of the two main types of multiple decision trees, is applied to searching Chinese herbs for compounds active against a total of 10 targets, on the basis of both the amount of data available for each and their biological and therapeutic significance. Selected targets fall into the following categories:

(1) Cell Signal Regulators. Cyclic adenosine 3′-5′-monophosphate (cAMP) -dependent protein kinase and cAMP phosphodiesterase (cAMP PDE) play opposing yet complementary roles in many cellular processes and are potential targets for a number of conditions, notably cancer, asthma, chronic obstructive pulmonary disease (COPD), and depression or memory-related disorders.

(2) Inducible Nitric Oxide Synthase (iNOS) Expression and NO Production in vivo. Overproduction of NO has been implicated in a wide variety of pathologies including atherosclerosis, inflammation, and neurodegenerative disease.

(3) Cyclooxygenase (COX) and Lipoxygenase (LOX). These are targets of nonsteroidal anti-inflammatory drugs (NSAIDS) and of great significance in the prevention and treatment of a wide range of inflammation-related disorders including cancer and Alzheimer's disease as well as arthritis.

(4) Aldose Reductase (AR). This represents the rate-limiting step in the polyol pathway and is of major significance in diabetic complications.

(5) HIV-1 Integrase, Protease, and Reverse Transcriptase (RT). These represent three of the major targets in HIV/AIDS therapy.

In addition to identifying potential inhibitors of these targets, the performance of RF with four different sets of molecular descriptors is assessed. Misclassification rates and related measures are used to ascertain the suitability of any particular set of descriptors.

The results are of interest from a number of perspectives. In addition to their potential use in drug discovery, target profiles are also of value in herbal medicine, especially given the recent and rapid dissemination of Chinese medicine worldwide. In particular, they can throw light both on herbs which may serve as reservoirs of particular target ligands and on the phytochemical classes involved. They can also be used to search for links between the putative targets of Chinese herbal constituents and the traditional usage of herbs which contain them. In this way, ethnopharmacological and molecular data may be brought together to throw possible light on such relationships.

As is shown below, most plant compounds tend to have *moderate to weak* affinities for their targets and may therefore serve more as *modulators* of target function rather than strong inhibitors as is the case with most pharmaceuticals.[10] The role of target modulation as a strategy in the treatment of disease remains largely unexplored to date, though in many cases, such a strategy may prove useful, particularly in the treatment of chronic or refractory conditions, or where conventional drugs are deemed unsatisfactory. Target modulation may also play an equally important role in disease prevention. In the treatment of type II diabetes, for instance, long-term modulation of aldose reductase may help prevent diabetic complications. As a potential source of such compounds, herbs are probably outstanding in this respect.

MATERIALS AND METHODS

A. Materials. Two databases were initially constructed, one containing data on known constituents of Chinese herbs, the other on bioactive plant compounds and their molecular targets.

The data were compiled from both primary and secondary sources, though a number of secondary sources in particular played an important role. In the case of Chinese herbs, these were Hsu et al., *The Chemical Constituents of Oriental Herbs*, volumes I and II;[11] Zhu, *Chinese Materia Medica: Chemistry, Pharmacology & Applications*;[12] Yan et al., *Traditional Chinese Medicines; Molecular Structures, Natural Sources & Applications*;[13] Duke, *Handbook of Phytochemical Constituents of GRAS Herbs and Other Economic Plants*;[14] and the CRC/Chapman & Hall *Dictionary of Natural Products* (DNP).[15]

In the case of bioactive plant compounds, the major reference source was Gideon Polya's *Biochemical Targets of Plant Bioactive Compounds*,[16] the only comprehensive reference work of its kind so far published. In addition, another useful source of information was the DNP.

In all cases, structures were downloaded (MDL mol and SMILES formats) from the DNP. In cases where structural modifications were required, this was carried out first prior to export.

For each entry, both data sets contain the following information: structure, SMILES code, trivial name, scientific name, compound class, skeletal type, chirality, CAS registry number, pharmacology (where known), toxicology (e.g., LD_{50} values and RTECS number), and reference(s).

In the case of Chinese herbs, the botanical species, herb name (pinyin), and part of the plant in which the compound is primarily found (where known) are given. For the bioactive plant compounds data set, additional information comprises the target(s) of each compound, $IC_{50}/K_d/K_i$ values, and other details specific to each target.

The Chinese herbs data set contains a total of 8264 compounds found in 230 herbs. However, some of these are found in more than one herb, particularly constituents of essential oils. The number of unique compounds is approximately 7000. The bioactive plant compounds data set contains information on 2597 compounds. A total of 78 targets is presently covered.

B. Methods. 1. Descriptors. Four sets of descriptors were computed for all entries in both data sets as follows (detailed explanations may be found in the references listed):

B.1.a. MOE 2D Descriptors. The full set of 2D descriptors available in MOE (Chemical Computing Group, Montreal, Quebec) were computed, excluding Kier-Hall and Labute VSA descriptors (see below). These largely comprise a mixture of constitutional, topological, and BCUT/GCUT descriptors in addition to various measures of lipophilicity and electronic charge and numbered 150 in all.

B.1.b. Kier−Hall Descriptors. The well-known descriptors of Kier and Hall are made up of three subsets: (a) Chi

(χ) descriptors which enumerate connectivity,[17] (b) Kappa (κ) descriptors which capture features of 2D shape in addition to a descriptor enumerating molecular "flexibility",[17] and (c) the electrotopological (E-state) indices which encode the electronegativity of each "atom type" (atom and associated bonds) within the context of its molecular neighborhood.[18] The full set of Kier-Hall descriptors numbered 56 in all.

B.1.c. Labute VSA Descriptors. Labute's virtual surface area (VSA) descriptors comprise the atomic contributions for each molecule to lipophilicity (log P), molar refractivity (MR), and partial charge. Each of these quantities is subdivided into a range of values (i.e., a set of intervals), and the surface area of the molecule which contributes to each of these intervals is then computed. Each VSA descriptor can thus be characterized as the amount of surface area where the feature in question (log P, MR, or partial charge) falls within a certain interval.

There are 32 VSA descriptors in all, 10 for log P, 8 for MR, and 14 for partial charge.[19]

B.1.d. MACCS Descriptors. These are based on the set of 166 MACCS (Molecular Access System) keys originally designed as queries for database retrieval but since used in a variety of applications.[20] For the most part, they enumerate very simple features but can provide very specific and detailed queries when used in combination. Two examples might be the number of benzene rings found in a molecule or the number of disulfide (−S−S−) groups. They are not generally recognized as having great utility in QSAR and were deliberately chosen here as it was suspected that they would give poor results in RF screening and could thus demonstrate how RF can be used to indicate the suitability or otherwise of a particular descriptor set.

B.2. Random Forest and Unbalanced Data. RF represents one of the two best known forms of multiple decision trees (the other being stochastic gradient boosting, also known as boosted trees[6,7]) and constitutes an ensemble of unpruned classification or regression trees, induced from bootstrap samples of the training data, and using random feature selection in the tree-building process.[8,9] It has been found to yield generalization error rates similar to those of the popular boosting algorithm Adaboost[21] yet is more robust to noise.

In many classification problems, the data are often extremely unbalanced, with one or more classes far outnumbering the other(s). In virtual screening applications, this problem is of considerable significance because a small set of active compounds is almost always used to search a much larger target data set for new candidates. Because RF seeks to minimize error, it will tend to focus more on the majority class, resulting in an unacceptably low degree of accuracy for the minority class.

Two remedies have been suggested as a means of equalizing error rates of highly unbalanced classes. One is cost-sensitive learning, where misclassification of the minority class is assigned a high cost which the algorithm then attempts to minimize. The other is based on selective sampling in which equal samples are drawn from both classes prior to modeling. Each tree is thus built on a more balanced set of data. In the case of the latter, RF adopts the following approach: (a) For each tree in the ensemble, take a bootstrap sample from the *minority* class. Randomly draw the same number of cases from the majority class. (b) Build a classification tree to maximum size without pruning. The tree-building procedure follows the CART algorithm using the Gini splitting criterion,[22] with the following modification: that a random subset of descriptors (D_{rand}) is searched at each node in the tree as opposed to the full set. (c) Repeat steps a and b for every tree built. Make the final predictions on the basis of the full ensemble.

RF also makes use of a form of cross validation termed "out of bag" (OOB) in which a subset of cases (approximately 33%) is withheld for each tree trained and constitutes an independent test sample. To measure the generalization error of the full forest, the OOB sample for each tree is run through that tree and the error rate of prediction measured. The error rates for all trees are then averaged to give the overall generalization error for the entire forest.

In our analysis, compounds active against each target from the bioactive plant compounds data set were added to the target data set of Chinese herbal constituents, the former constituting the minority, the latter the majority class. No account was taken of information on the relative potency of different active compounds, so the analysis constitutes a simple binary classification. A total of 500 trees were built for each target, with D_{rand} set to the square root of the total number of descriptors. The analysis was repeated for each of the four descriptor sets. Modeling was carried out using *Random Forests*, version 1.0 (Salford Systems, San Diego, California).

Care must be taken in interpreting RF results for highly unbalanced data. The overall misclassification, for instance, is of limited use because even if the minority class is badly misclassified there might still be a low overall value. A number of more appropriate measures have been suggested,[23] including the prediction success for *each* class and the geometric mean for both classes (minority prediction × majority prediction$^{0.5}$) among others. In this instance, these were used to compare performance of different descriptor sets.

The number of active compounds used in training were as follows: cAMP PDE (222), cAMP-dependent protein kinase or PKA (101), iNOS expression (81), NO prod^{n} in vivo (97), COX (169), LOX (199), AR (187), HIV-1 integrase (41), HIV-1 protease (58), and HIV-1 RT (91). In all cases, the number of target compounds constituted the full set of 8264 compounds from Chinese herbs.

RESULTS

1. Distribution of IC_{50} Values among Plant Bioactive Compounds. Though IC_{50} values or related measures (such as K_i and K_d) are not listed for most compounds in our bioactive plant compounds database, such values are nevertheless present for an appreciable number (approximately 39%), and their distribution thus provides a useful overall picture of the relative potency of phytochemical inhibitors compared to typical pharmaceutical drugs.

Given that the data have been pooled from many different sources, these should be treated as approximate figures and have been assigned to seven categories, as follows: (1) $IC_{50} < 0.01$ μM (<10 nM), a potency characteristic of a pharmaceutical inhibitor; (2) $IC_{50} = 0.01-0.1$ μM, a weak pharmaceutical-type inhibitor; (3) $IC_{50} = 0.1-1$ μM; (4) IC_{50}

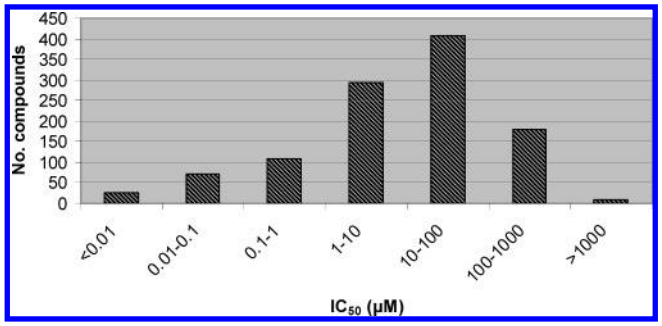

Figure 1. Number of compounds in Bioactive Plant Compounds data set in the seven IC_{50} categories (see text).

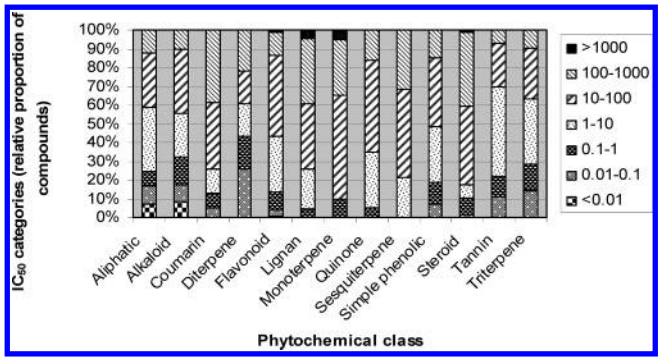

Figure 2. Proportions of compounds falling into the seven IC_{50} categories for different phytochemical classes.

$= 1-10\ \mu M$; (5) $IC_{50} = 10-100\ \mu M$; (6) $IC_{50} = 100-1000\ \mu M$; and (7) $IC_{50} > 1000\ \mu M$.

Figure 1 shows the number of compounds (for all targets) falling into these seven categories.

Despite the relative paucity of the data, the proportion of compounds falling into each category may nevertheless provide a reasonable assessment of the range of IC_{50} values characteristic of plant compounds and shows that the great majority are 2−3 orders of magnitude less potent than typical pharmaceuticals.

The analysis was refined by considering different phytochemical classes separately. As Figure 2 shows, we do not see any outstanding differences between different classes, though alkaloids, as one might expect, contain a relatively high proportion of compounds with values characteristic of drug compounds. Aliphatics, diterpenes, triterpenes, and tannins also contain compounds with relatively low values compared to other plant compounds.

Monoterpenes, sesquiterpenes, and steroids, by contrast, tend to show high values, though it should be stressed that examples of mono- and sesquiterpenes in particular are poorly represented in the data set at present, and further statistical analysis might thus prove misleading at this stage.

It should also be noted that, in keeping with these results, the compounds identified as potentially active in the target data are likely to show similar behavior, with the great majority expected to have moderate to weak activity against the target in question.

2. Performance of Random Forest. Despite the highly unbalanced ratios between active compounds and the target data set (comprising 8264 compounds), results, for three out of the four descriptor sets, proved satisfactory, as judged by misclassification rates which, with few exceptions, were reasonably well-balanced for both categories and usually less than 30% (taking the geometric mean for both). The one exception was the MACCS descriptor set which gave misclassification rates which were both higher on average than the others and highly unbalanced for both categories. Misclassification thus appears to provide a useful measure of the suitability of a particular set of descriptors for screening. For each of the 10 targets and four descriptor sets used, the percentage correctly predicted for active compounds (A), target compounds (T), and their geometric mean (G*) are shown in Table 1. The number of active compounds used in training is shown in brackets.

In general, the Kier-Hall descriptor set gave the best results, and one could therefore simply choose to use this to identify hits in the target data. With the exception of the MACCS set, however, the other descriptors gave results which are reasonably good given the highly unbalanced nature of the data.

As has become increasingly clear in recent years, there are advantages to using consensus based on several different analyses in many virtual screening applications,[24] and accordingly, we chose to use all descriptor sets to identify hits, excluding the MACCS set. Furthermore, to qualify as a hit, a compound had to satisfy the following two constraints: (1) it had to be classified as active by 80% or more of the votes from the full ensemble of trees, and (2) it had to achieve this score for at least two of the three descriptor sets used.

This has the effect of restricting hits to a relatively small set of compounds (usually on the order of 200 or less) for each target. The analysis thus errs on the side of caution but has the advantage of only identifying compounds which cannot be satisfactorily distinguished, by more than one set of descriptors, from examples in the training set. In addition, despite this limitation, a wide diversity of phytochemical classes and herbs were hit, as shown in the following tables, suggesting that inhibitors for these targets may be widespread in Chinese herbs.

3. Distribution of Hits in Chinese Herbs in Relation to Specific Targets. The total number of species showing at least two hits for one or more targets amounted to 150 out of the 240 found in the Chinese herbs database, just over 62%. Of these, a large proportion were predicted to inhibit two or more targets, 80/150 or 53%. As shown below, it is also not uncommon to find a range of phytochemicals from more than one class in any one species which may inhibit a single target.

In Tables 2−7, the term "multiple hits" indicates that there were at least two compounds found in the herbs listed for that phytochemical class. Herbs only hit once are not shown both for reasons of space and because they were considered less likely, on the basis of current information, to possess appreciable numbers of compounds from the class in question. Herbs in bold script denote those in which more than five hits were found. Those with a superscript "*a*" contain multiple hits from more than one phytochemical class. For each herb, both the Latin binomial and Chinese name (pinyin) are provided.

For each phytochemical class, the skeletal types identified among hits are shown in parentheses. An asterisk (*) denotes a skeletal type not found among the active compounds used for training and, thus, represents an example of "scaffold hopping" (see below). It will be noticed that for some tables certain phytochemical classes are listed in more than one

Table 1. RF Prediction Success of Active and Target Compounds and Their Geometric Means for Four Sets of Descriptors (A = Active Compounds, T = Target Compounds, G* = Geometric Mean)

target/process (no. active compounds)	MOE			Kier-Hall			Labute			MACCS		
	A	T	G*	A	T	G*	A	T	G*	A	T	G*
cAMP PDE (222)	84.4	86.4	**85.4**	98.1	85.3	**91.5**	81.1	86.3	**83.7**	77.8	87.6	**82.6**
PKA (101)	73	87.5	**79.9**	89	82	**85.4**	70	85.7	**77.5**	59	85.8	**71.2**
iNOS expression (81)	73.8	83.8	**78.6**	85	80.9	**82.9**	63.8	84.7	**73.5**	41.3	83	**58.5**
NO prod^n in vivo (97)	81.1	86.7	**83.8**	92.6	82.6	**87.5**	72.6	87	**79.5**	35.8	92.6	**57.6**
COX (169)	75	86.5	**80.5**	85.1	84.3	**84.7**	71.4	85.7	**78.2**	83.9	66.6	**74.8**
LOX (199)	77.2	85.4	**81.2**	87.3	85	**86.1**	75.1	85.2	**80**	83.8	67.2	**75**
Aldose reductase (187)	83.7	88.2	**85.9**	91.9	81.4	**86.5**	78.8	86.1	**82.4**	87	67.1	**76.4**
HIV-1 integrase (41)	78.1	86	**81.9**	90.2	81.2	**85.6**	75.6	85.2	**80.2**	43.9	81.6	**59.8**
HIV-1 protease (58)	72.4	87.2	**79.5**	87.9	81	**84.4**	62.1	86.4	**73.2**	29.3	94.3	**52.6**
HIV-1 reverse transcriptase (91)	74.7	84.5	**79.5**	83.9	80.4	**82.1**	78.2	85.9	**81.9**	63.2	88.8	**74.9**

Table 2. Distribution of Predicted CAMP PDE Inhibitors in Chinese Herbs

phytochemical class (skeletal types)	TCM category	herbs with multiple hits
Alkaloid (aporphine, benzylisoquinoline, protopine)	astringent,[1] invigorate blood[2]	*Papaver somniferum* (Ying Su Ke),[1] *Corydalis* spp. (Yan Hu Suo)[2]
coumarin	phlegm heat,[1] tonify qi,[2] tonify yang[3]	***Peucedanum decursivum, P. praeruptorium* (Qian Hu),[1]** *Glycyrrhiza* spp.[a] (Gan Cao),[2] *Cnidium monnieri* (She Chuang Zi)[3]
Flavonoid (biflavonoid, chalcone, dihydroflavonol, flavanol, flavanone, flavone, flavonol, isoflavan, isoflavone)	antitussive,[1] damp heat,[2] diuretic,[3] fire poison,[4] stop bleeding,[5] tonify qi,[6] wind cold[7]	***Ginkgo biloba* (Bai Guo Ye),[1]** *Scutellaria baicalensis* (Huang Qin),[2] *Sophora flavescens* (Ku Shen),[2] *Zea mays* (Yu Mi Xu),[3] *Sophora subprostrata* (Shen Dou Gen),[4] *Platycladus orientalis* (Ce Bai Ye),[5] ***Glycyrrhiza* spp.[a] (Gan Cao),[6]** *Coriandrum sativum* (Yan Sui Zi)[7]
Lignan (furofuranoid)	fire poison,[1] stop bleeding,[2] tonify yang[3]	*Forsythia suspensa* (Lian Qiao),[1] *Imperata cyclindrica* (Bai Mao Gen),[2] ***Eucommia ulmoides* (Du Zhong)[3]**
Phenolic (xanthone)	damp heat	*Gentiana lutea* (Long Dan Cao)
Steroid (furostan, spirostan)	diuretic,[1] fire poison,[2] heat (qi),[3] internal wind,[4] invigorate blood,[5] tonify yang,[6] tonify yin,[7] wind damp[8]	***Dioscorea collettii* (Bei Xie),[1] *Smilax* spp. (Tu Fu Ling),[2] *Anemarrhena asphodeloides* (Zhi Mu),[3] *Tribulus terrestris* (Bai Ji Li),[4] *Dracaena draco* (Xue Jie),[5] *Trigonella foenum-graecum* (Hu Lu Ba),[6] *Asparagus cochinchinensis* (Tian Men Dong),[7] *Lilium brownii* (Bai He),[7] *Ophiopogon japonicus* (Mai Men Dong),[7]** *Polygonatum odoratum* (Yu Zhu),[7] ***Smilax sieboldii* (Wei Ling Xian)[8]**

[a] Herbs found in >1 phytochemical class.

row. This reflects the importance of that class and helps to clarify the nature of the different skeletal types found in different herbs.

The acronym TCM stands for traditional Chinese medicine, and the column entitled "TCM categories" is intended to give some idea of the nature and diversity of component herbs in terms of their roles in Chinese medicine. The categorization follows that found in most Chinese materia medica.[25] Further details on TCM categories may be found in the Supporting Information. There is only limited discussion of TCM in the text, though in some cases, interesting patterns of association are seen.

Other features specific to individual tables are explained below.

3.a. Cell Signaling: cAMP PDE and cAMP-Dependent PKA. cAMP, the ubiquitous second messenger of hormone action, is generated by adenylyl cyclases which in turn are activated by G-protein-coupled receptors (GPCRs). cAMP diffuses through the cell and affects a variety of targets. One of the principal ones is cAMP-dependent PKA, which in the absence of cAMP is inhibited by regulatory subunits but dissociates, releasing now-active catalytic subunits, when cAMP levels rise. PKA-mediated phosphorylation affects a number of pathways, the effects being transient as PKA targets are in turn dephosphorylated by protein phosphatases (though PKA can also phosphorylate protein phosphatase inhibitor proteins thus slowing the rate at which its effects are reversed). PKA targets are numerous, and over 25 have been discovered to date, among them a variety of enzymes involved in carbohydrate metabolism (e.g., glycogen synthases and acetyl CoA carboxylase).[26] Equally significant are PKA's positive effects on steroidogenesis, in part by its regulatory role in cholesterol transport.[27]

One of the major factors controlling the flux of cellular cAMP is the cAMP phosphodiesterases. This family of enzymes hydrolyses cAMP, reducing its availability to PKA and others. Thus, cAMP phosphodiesterases and PKA are together intimately involved in regulating metabolic pathways which are activated by cAMP, and recent evidence has shown that they are often compartmentalized together via the A-kinase anchoring proteins.[28]

Inhibitors of both cAMP PDEs and PKA are also of considerable interest from a therapeutic perspective. PKA inhibitors may have potential in the treatment of obesity,[29]

Table 3. Distribution of Predicted PKA Inhibitors in Chinese Herbs

phytochemical class (skeletal types)	TCM category	herbs with multiple hits
Alkaloid (quinazoline*)	fire poison	*Isatis indigotica* (Ban Zhi Lian, Da Qing Ye, Qing Dai)
flavonoid (dihydroflavonol, flavanol, flavanone, flavone, homoisoflavonoid, isoflavone)	antitussive,[1] diuretic,[2] damp heat,[3] fire poison,[4] heat (qi),[5] internal cold,[6] invigorate blood,[7] phlegm cold,[8] regulate qi,[9] summer heat,[10] tonify blood,[11] tonify qi,[12] wind cold,[13] wind damp,[14] wind heat[15]	***Ginkgo biloba* (Bai Guo Ye),[1]** *Erythrina variegata* (Hai Tong Pi),[2] *Pyrrosia petiolosa* (Shi Wei),[2] *Zea mays* (Yu Mi Xu),[2] ***Scutellaria baicalensis* (Huang Qin),[3]** *Sophora flavescens* (Ku Shen),[3] *Scutellaria barbata* (Ban Zhi Lian),[4] *Sophora subprostrata*[a] (Shan Dou Gen),[4] *Cassia tora* (Jue Ming Zi),[5] *Prunella vulgaris*[a] (Xia Ku Cao),[5] *Alpinia galanga, A. officinarum* (Gao Liang Jiang),[6] *Caesalpinia sappan*[a] (Su Mu),[7] *Dracaena draco*[a] (Xue Jie),[7] *Leonurus cardiaca* (Yi Mu Cao),[7] *Inula brittanica* (Xuan Fu Hua),[8] *Dalbergia odorifera* (Jiang Xiang),[9] *Dolichos lablab* (Bai Bian Dou),[10] *Polygonum multiflorum* (He Shou Wu),[11] *Glycyrrhiza* spp. (Gan Cao),[12] *Schizonopeta tenuifolia* (Jing Jie),[13] *Morus alba* (Sang Zhi),[14] *Mentha* spp. (Bo He)[15]
Phenolic (benzophenone*, biphenyl*, stilbene*, xanthone)	damp heat,[1] invigorate blood,[2] wind damp[3]	*Gentiana lutea* (Long Dan Cao),[1] *Caesalpinia sappan*[a] (Su Mu),[2] ***Morus alba*[a] (Sang Zhi)[3]**
policyclic aromatic (anthracene*, anthraquinone, phenanthrene*, naphthalene*)	diuretic,[1] heat (qi),[2] laxative,[3] stop bleeding,[4] tonify blood,[5] tonify yang[6]	*Juncus effusus* (Deng Xin Cao),[1] *Cassia tora* (Jue Ming Zi),[2] ***Rheum* spp. (Da Huang),[3]** *Aloe ferox* (Lu Hui),[3] ***Rubia cordifolia* (Qian Cao Gen),[4]** *Polygonum multiflorum*[a] (He Shou Wu),[5] ***Morinda officinalis* (Ba Ji Tian)[6]**
tannin (condensed, hydrolizable)	astringent,[1] internal cold,[2] nourtish heart,[3] phlegm heat,[4] regulate qi,[5] stop bleeding[6]	***Cornus officinalis* (Shan Zhu Yu),[1] *Punica granatum* (Shi Liu Pi),[1] *Terminalia chebula* (He Zi),[1]** *Eugenia caryophyllata, Syzygium aromaticum* (Ding Xiang),[2] *Albizzia lebbek* (He Huan Pi),[3] *Euphorbia helioscopia* (Ze Qi),[4] *Liquidambar formosana* (Su He Xiang),[4] *Areca catechu* (Da Fu Pi),[5] ***Sanguisorba officinalis*[a] (Di Yu)[6]**
Triterpene (cycloartane*, dammarane, hopane*, glutinane*, oleanane, ursane)	diuretic,[1] fire poison,[2] heat (qi),[3] invigorate blood,[4] parasites,[5] stop bleeding,[6] tonify qi,[7] wind damp[8]	*Juncus effusus* (Deng Xin Cao),[1] *Pyrrosia petiolosa* (Shi Wei),[1] *Sophora subprostrata*[a] (Shan Dou Gen),[2] *Prunella vulgaris*[a] (Xia Ku Cao),[3] *Boswellia carteri, B. serrata* (Ru Xiang),[4] *Dracaena draco*[a] (Xue Jie),[4] *Dryopteris crassirhizoma* (Guan Zhong),[5] *Sanguisorba officinalis*[a] (Di Yu),[6] ***Glycyrrhiza uralensis*[a] (Gan Cao),[7] *Trypterygium wilfordii* (Lei Gong Teng)[8]**

[a] Herbs found in >1 phytochemical class.

HIV-induced T-cell dysfunction,[30] and cancer.[31] cAMP PDE is a prominent target for intervention in asthma, COPD, and certain types of cancer.[32] More recently, cAMP PDE inhibitors have also been suggested in the treatment of depression.[33]

Tables 2 and 3 show the distribution of hits within Chinese herbs for cAMP PDE and PKA, respectively. With the exception of certain flavonoid-containing herbs, and the xanthone-containing *Gentiana lutea*, there is no phytochemical overlap. The majority of herbs predicted to show inhibitory effects on cAMP PDE are steroid-containing species, a number from the Liliaceae (such as *Asparagus cochinchinensis*, *Ophipogon japonicus*, *Lilium brownii*, and *Anemarrhena asphodeloides*). Furofuranoid lignans, coumarins, and a variety of alkaloids (aporphine, benzylisoquinoline, and protopine) are also involved to a more modest extent.

In view of the important role of PKA in steroidogenesis, it is interesting that steroidal plant compounds are among the major inhibitors of cAMP PDE because such compounds would be expected to promote PKA activity. Though phytosterols are not considered to be precursors of mammalian steroid hormones in vivo, it is possible that they may nevertheless exert steroidogenic effects due to their inhibition of cAMP PDE.

Considering that cAMP PDE inhibitors are considered among the most promising targets for lung-related disorders as well as depression, it is also perhaps significant that many of the steroid-containing herbs listed in Table 2 are used as "yin tonics" in TCM with particular effects on the yin of the stomach, lungs, and heart. Though they would not always be used in treating conditions such as asthma or COPD, they would certainly be foremost among those considered, depending on the TCM diagnosis.

In the case of PKA, flavonoids are predicted to play a greater role as inhibitors than is the case with cAMP PDE, and polycyclic aromatics, notably anthraquinones, are also prominent. A number of tannin-containing species, particularly from the "astringent" category, are identified in addition. Triterpenes are hit, usually though not exclusively the five-ring types, such as glutinane, hopane, lupane, oleanane, and ursane triterpenes (all very similar structurally).

Another feature of interest concerns the relatively large number of new skeletal types identified (*). For the most part, these are extremely similar to known inhibitors, particularly in the case of triterpenes. A larger number of herbs are found with hits from more than one phytochemical class than is the case with cAMP PDE inhibitors.

3.b. Nitric Oxide. The 1980s and early 1990s witnessed the discovery of one of the most important chemical messengers found so far, nitric oxide (NO), released by the endothelium and other tissues in mammals. This small molecule has since been linked to a host of physiological

Table 4. Distribution of Compounds Predicted to Modulate Production of Nitric Oxide in Chinese Herbs

phytochemical class (skeletal types)	TCM category	herbs with multiple hits
Aliphatic (acetylene)	diuretic,[1] stop bleeding,[2] tonify blood,[3] tonify qi,[4] tonify yin,[5] wind heat[6]	*Artemisia capillaris* (Yin Chen Hao),[1] *Cirsium japonicum* (Da Ji),[2] *Panax notoginseng*[a] (San Qi),[2] *Angelica sinensis*[a] (Dang Gui),[3] ***Panax ginseng*[a] (Ren Shen),[4]** *Glehnia littoralis*[a] (Bei Sha Shen),[5] *Bidens pilosa* (Xian Feng Cao),[6+] *Bupleurum longiradiatum* (Chai Hu)[6]
Alkaloid (aporphine, furanoquinoline*, quinazoline)	antitussive,[1] fire poison,[2] invigorate blood[3]	*Nandina domestica* (Tian Zhu Zi),[1] *Dictamnus dasycarpus* (Bai Xian Pi),[2+] *Isatis indigotica* (Ban Lan Gen, Da Qing Ye, Qing Dai),[2] *Corydalis* spp. (Yan Hu Suo)[3]
coumarin	internal cold,[1] phlegm heat,[2] regulate qi,[3] tonify blood,[4] tonify yang[5], tonify yin[6], wind cold,[7] wind damp[8]	*Foeniculum vulgare*[a] (Xiao Hui Xiang),[1+] ***Zanthoxylum schinifolium* (Chuan Jiao),[1+]** ***Peucedanum praeruptorium, P. decursivum* (Qian Hu),[2]** *Nardostachys jatmansi* (Gan Song Xiang),[3+] *Angelica sinensis*[a] (Dang Gui),[4+] *Cnidium monnieri* (She Chuang Zi),[5] *Glehnia littoralis*[a] (Bei Sha Shen),[6] *Angelica dahurica* (Bai Zhi),[7+] *Ledebouriella divaricata* (Fang Feng),[7] *Notopterygium incisum* (Qiang Huo),[7+] *Angelica pubescens*[a] (Du Huo)[8+]
Diterpene (abietane, labdane, pimarane*)	invigorate blood,[1] wind damp[2]	*Salvia miltorrhiza* (Dan Shen),[1] *Leonurus sibiricus* (Yi Mu Cao),[1] *Siegesbeckia pubescens*[a] (Xi Xian Cao),[2] *Trypterygium wilfordii* (Lei Gong Teng)[2]
Flavonoid (biflavonoid, flavone, flavonol, isoflavone)	antitussive,[1] damp heat,[2] fire poison,[3] invigorate blood,[4] stop bleeding,[5] tonify qi[6]	***Ginkgo biloba*[a] (Bai Guo Ye),[1]** ***Scutellaria baicalensis* (Huang Qin),[2]** *Andrographis paniculata* (Chuan Xin Lian),[3] *Scutellaria rivularis* (Ban Zhi Lian),[3] *Smilax glabra* (Tu Fu Ling),[3] *Sophora subprostrata* (Shan Dou Gen),[3] *Crocus sativus* (Xi Hong Hua),[4+] *Dracaena draco* (Xue Jie),[4] *Platycladus orientalis* (Ce Bai Ye),[5] *Glycyrrhiza uralensis* (Gan Cao)[5+]
Lignan (neolignan)	aromatic digestive,[1] damp heat[2]	*Magnolia officinalis, M. obovata* (Hou Po),[1] ***Coptis japonica* (Huang Lian)[2+]**
Meroterpene	heat (blood)	*Lithospermum erythrorhizon, Arnebia euchroma* (Zi Cao)[+]
phenolic (diarylalkyl, dibenzyl, stilbene)	internal cold,[1] invigorate blood,[2] tonify qi,[3] wind damp[4]	*Foeniculum vulgare*[a] (Xiao Hui Xiang),[1+] ***Curcuma* spp.[a] (E. Zhu, Yu Jin),[2+]** *Dioscorea batatas* (Shan Yao),[3+] *Morus alba* (Sang Zhi)[4]
Polycyclic aromatic (anthracene*, naphthalene)	laxative	*Rheum* spp. (Da Huang), *Cassia angustifolia* (Fan Xie Ye)
Sesquiterpene (bisabolane*, clovane*, eudesmane, germacrane, guaiane, xanthane*)	antitussive,[1] fire poison,[2] invigorate blood,[3] regulate qi,[4] stop bleeding,[5] tonify qi,[6] wind damp,[7] wind heat[8]	*Ginkgo biloba*[a] (Bai Guo Ye),[1+] *Chrysanthemum indicum* (Ye Ju Hua),[2+] *Eupatorium chinense* (Tu Niu Xi),[2] ***Curcuma* spp.[a] (E Zhu, Yu Jin),[3+]** ***Inula helenium, Saussurea lappa* (Mu Xiang),[4]** *Lindera strychnifolia* (Wu Yao),[4+] *Nardostachys jatmansi* (Gan Song Xiang),[4+] *Artemisia argyi* (Ai Ye),[5+] *Atractylodes macrocephala* (Bai Zhu),[6+] *Angelica pubescens*[a] (Du Huo),[7+] *Siegesbeckia orientalis*[a] (Xi Xian Cao),[7] *Xanthium strumarium* (Cang Er Zi),[7] *Arctium lappa* (Niu Bang Zi),[8+] *Chrysanthemum morifolium* (Ju Hua),[8+]
Tannin (hydrolizable)	astringent,[1] stop bleeding[2]	*Cornus officinalis* (Shan Zhu Yu),[1] *Punica granatum* (Shi Liu Pi),[1] *Terminalia chebula* (He Zi),[1] *Agrimonia pilosa* (Xian He Cao),[2] *Sanguisorba officinalis* (Di Yu)[2]
Triterpene (dammarane)	stop bleeding,[1] tonify qi[2]	***Panax notoginseng*[a] (San Qi),[1+]** ***Panax ginseng*[a] (Ren Shen)[2+]**

[a] Herbs found in >1 phytochemical class.

and pathological states. It is now known to be involved in the control of blood pressure, neurotransmission, and immune defense. Excessive production is thought to be involved in atherosclerosis, inflammation, hypotension, Huntington's and Alzheimer's diseases, AIDS dementia, and cancer, whereas underproduction has been implicated in thrombosis, vasospasm, and erectile dysfunction.[34]

Nitric oxide is a relatively stable free radical with one unpaired electron, though it gives rise to a number of other chemical species, such as nitrogen peroxide, nitrogen dioxide, and peroxynitrite, which can cause tissue damage, particularly to fatty acids and proteins.

The enzymes responsible for its synthesis, nitric oxide synthases, are found in two forms, the one constitutively expressed (cNOS), and largely found in endothelial and neural tissue. Activation of cNOS results in the production of short bursts of NO at low concentrations. The inducible form, iNOS, when activated, by contrast, produces NO at high concentration and is found in a variety of cells including mast cells, macrophages, Kuppfer cells, and neutrophils, as well as endothelial cells. Activated immune cells produce NO in quantities sufficient to kill a variety of harmful target cells such as those found in cancers and parasites. On the negative side, overproduction of NO, largely through exces-

Table 5. Predicted COX/LOX Inhibitors in Chinese Herbs (C = COX Inhibitors, L = LOX Inhibitors)

Phytochemical class (skeletal types)	TCM category	herbs with multiple hits
aliphatic (acetylene)	diuretic,[1] invigorate blood,[2] tonify blood,[3] tonify qi,[4] tonify yin,[5] wind damp,[6] wind heat[7]	*Artemisia capillaris*[a] (Yin Chen Hao),[1] *Carthamus tinctorius* (Hong Hua),[2] *Angelica sinensis*[a] (Dang Gui),[3] *Panax ginseng*[a] (Ren Shen),[4] *Glehnia littoralis*[a] (Bei Sha Shen),[5L] *Angelica pubescens*[a] (Du Huo),[6L] ***Bupleurum falcatum* (Chai Hu),[7]** *Bidens bipinnata* (Xian Feng Cao)[7L]
Aliphatic (branched*,[2L] sulfide,[3L] monocarbocyclic[1L])	heat (qi),[1] internal cold,[2] wind cold[3]	*Gardenia jasminoides* (Zhi Zi),[1L] *Zingiber officinale*[a] (Sheng/Gan Jiang,),[2L] *Allium fistulosum* (Cong Bai)[3L]
aliphatic/phenolic (long-chain aromatic)	antitussive,[1] fire poison,[2] internal cold,[3] invigorate blood[4]	***Ginkgo biloba* (Bai Guo, fruit),[1]** *Belamcanda chinensis* (She Gan),[2] ***Zingiber officinale*[a] (Sheng/Gan Jiang),[3]** *Rhus vernicifera* (Gan Qi)[4]
Aliphatic (unsaturated fatty acid[I])	nourish blood,[1] tonify qi,[2] tonify yin,[3] wind damp[4]	*Angelica sinensis*[a] (Dang Gui),[1L] *Panax ginseng*[a] (Ren Shen),[2L] *Glehnia littoralis*[a] (Bei Sha Shen),[3L] *Angelica pubescens*[a] (Du Huo)[4L]
Alkaloid (isobutylamide, quinazoline,[C] protoberberine[L])	fire poison,[1] internal cold,[2] invigorate blood,[3] wind cold[4]	*Isatis indigotica* (Ban Lan Gen, Qing Dai, Da Qing Ye),[1C] ***Piper* spp. (Bi Bo, Hu Jiao),[2]** ***Zanthoxylum piperitum* (Chuan Jiao),[2]** *Corydalis* spp. (Yan Hu Suo),[3L] *Asarum heterotropoides* (Xi Xin)[4]
Benzopyran/coumarin	diuretic,[1] wind damp[2]	*Artemisia capillaris*[a] (Yin Chen Hao),[1C] *Angelica pubescens*[a] (Du Huo)[2L]
Carotenoid	invigorate blood	*Crocus sativus* (Xi Hong Hua)
flavonoid (biflavonoid,[C] chalcone, flavan, flavanol, flavanone,[L] flavone, flavonol, isoflavan,[C] isoflavone,[C] pterocarpan)	antitussive,[1] damp heat,[2] diuretic,[3] fire poison,[4] internal cold,[5] invigorate blood,[6] regulate qi,[7] stop bleeding,[8] tonify blood,[9] tonify qi,[10] wind cold,[11] wind damp,[12] wind heat[13]	***Ginkgo biloba* (Bai Guo Ye, leaf),[1C]** ***Scutellaria baicalensis* (Huang Qin),[2]** *Sophora flavescens* (Ku Shen),[2L] *Artemisia capillaris*[a] (Yin Chen Hao),[3] *Erythrina variegata* (Hai Tong Pi),[3L] *Zea mays* (Yu Mi Xu),[3C] *Andrographis paniculata* (Chuan Xin Lian),[4C] *Sophora subprostrata* (Shan Dou Gen),[4C] ***Alpinia officinarum*[a] (Gao Liang Jiang),[5C]** ***Dracaena draco* (Xue Jie),[6C]** ***Dalbergia odorifera* (Jiang Xiang),[7]** *Artemisia* spp. (Ai Ye),[8C] *Polygonum multiflorum* (He Shou Wu),[9C] *Glycyrrhiza glabra* (Gan Cao),[10C] *Astragalus membranaceus* (Huang Qi),[10C] *Cinnamomum cassia* (Gui Zhi),[11C] ***Morus alba*[a] (Sang Zhi),[12]** *Mentha* spp. (Bo He)[13C]
Lignan (neolignan)	aromatic digestive,[1] invigorate blood[2]	*Magnolia obovata* (Hou Po),[1] *Salvia miltiorrhiza* (Dan Shen)[2L]
Pheolic (phenol)	internal cold,[1] open heart,[2] phlegm cold[3]	***Zingiber officinale*[a] (Gan/Sheng Jiang),[1]** *Liquidambar orientalis* (Su He Xiang),[2] *Inula brittanica*[a] (Xuan Fu Hua)[3]
Phenolic (diarylalkyl)	aromatic digestive,[1] internal cold,[2] invigorate blood[3]	***Alpinia katsumadai*[a] (Cao Dou Kou),[1]** *Pogostemon heyneanus* (Guang Huo Xiang),[1] ***Alpinia officinarum*[a] (Gao Liang Jiang),[2]** *Zingiber officinale*[a] (Sheng/Gan Jiang),[2L] *Curcuma* spp. (E Zhu, Yu Jin)[3]
Phenolic (dibenzyl,[C] diphenyl ether[L])	antitussive,[1] stop bleeding[2]	*Stemona tuberosa* (Bai Bu),[1C] *Imperata cylindrica* (Bai Mao Gen)[2L]
phenolic (stilbene)	aromatic digestive,[1] laxative,[2] wind damp[3]	***Alpinia katsumadai*[a] (Cao Dou Kou),[1C]** ***Rheum* spp. (Da Huang),[2]** *Morus alba*[a] (Sang Zhi)[3L]
Sesquiterpene (germacranolide, guaianolide)	regulate qi,[1] phlegm cold[2]	*Inula helenium* (Mu Xiang),[1C] *Inula brittanica*[a] (Xuan Fu Hua)[2C]

[a] Herbs found in >1 phytochemical class.

sive expression of the inducible form, is implicated in the pathologies mentioned above, and there is therefore considerable interest in compounds which inhibit this process.[35]

Table 4 lists those phytochemical classes predicted to modulate iNOS expression and the herbs containing them. Herbs marked with a plus sign (+) are predicted, on the basis of RF analysis, to contain compounds which inhibit production of NO in vivo but where the actual mechanism is unknown (though in most cases inhibition of iNOS expression may be suspected). Of special interest are the large number of hits among herbs containing acetylenes, coumarins, flavonoids, and sesquiterpenes. The latter in particular are poorly predicted for other targets, probably because of insufficient information. Many skeletal types are covered within this latter class, indicating that activity may be due less to the particular scaffold than to the nature of the substituents involved.

Also of interest are the large number of ginsenosides (dammarane triterpenes) identified as possible inhibitors, mainly from *Panax notoginseng* and *P. ginseng*. In both species, acetylenes are also prominently identified, suggesting that these herbs may exert inhibitory effects due to the activities of more than one phytochemical class. Such effects may also play a role in the case of various other herbs, such

Table 6. Predicted Inhibitors of Aldose Reductase in Chinese Herbs

phytochemical class (skeletal types)	TCM category	herbs with multiple hits
Benzopyran/coumarin	diuretic	*Artemisia capillaris*[a] (Yin Chen Hao)
diterpene (abietane)	invigorate blood,[1] wind damp[2]	***Salvia miltiorrhiza* (Dan Shen),[1]** *Pinus massoniana* (Song Jie)[2]
Flavonoid (anthocyanin, biflavonoid, chalcone, flavanol, flavone, flavonol, isoflavone*)	antitussive,[1] damp heat,[2] diuretic,[3] fire poison,[4] heat (deficient),[5] heat (qi),[6] internal cold,[7] invigorate blood,[8] nourish heart,[9] phlegm cold,[10] stop bleeding,[11] summer heat,[12] tonify qi,[13] tonify yang,[14] wind cold,[15] wind damp,[16] wind heat[17]	***Ginkgo biloba* (Bai Guo Ye),[1]** ***Scutellaria baicalensis* (Huang Qin),[2]** *Artemisia capillaris*[a] (Yin Chen Hao),[3] *Clematis armandii* (Mu Tong),[3] *Plantago* spp. (Che Qian Zi),[3] *Pyrrosia petiolosa* (Shi Wei),[3] *Zea mays* (Yu Mi Xu),[3] *Andrographis paniculata* (Chuan Xin Lian),[4] *Belamcanda chinensis* (She Gan),[4] *Chrysanthemum indicum* (Ye Ju Hua),[4] *Hypericum japonicum* (Tian Ji Huang),[4] *Lonicera japonica* (Jin Yin Hua),[4] ***Scutellaria rivularis* (Ban Zhi Lian),[4]** *Smilax glabra* (Tu Fu Ling),[4] *Sophora subprostrata* (Shan Dou Gen),[4] *Artemisia annua* (Qing Hao),[5] *Cassia tora*[a] (Jue Ming Zi),[6] *Gardenia jasminoides* (Zhi Zi),[6] *Prunella vulgaris* (Xia Ku Cao),[6] *Ailanthus altissima* (Chun Pi),[7] *Alpinia* spp. (Gao Liang Jiang),[7] *Zanthoxylum bungeanum* (Chuan Jiao),[7] *Dracaena draco* (Xue Jie),[8] *Leonurus cardiaca* (Yi Mu Cao),[8] *Carthamus tinctorius* (Hong Hua),[8] *Albizzia lebbek* (He Huan Pi),[9] *Inula brittanica* (Xuan Fu Hua),[10] *Artemisia argyi* (Ai Ye),[11] *Cirsium japonicum* (Da Ji),[11] *Dolichos lablab* (Bai Bian Dou),[12] *Nelumbo nucifera* (He Ye),[12] *Glycyrrhiza uralensis* (Gan Cao),[13] *Epimedium* spp. (Yin Yang Huo),[14] *Schizonopeta tenuifolia*[a] (Jing Jie),[15] *Coriandrum sativum* (Yan Sui Zi),[15] *Abrus precatorius* (Ji Gu Cao),[16] *Impatiens balsamina* (Tou Gu Cao),[16] *Pinus massoniana* (Song Jie),[16] *Bidens pilosa* (Xian Feng Cao),[17] *Equisetum* spp. (Mu Zei),[17] ***Lippia nodiflora* (Peng Lai Cao),[17]** ***Mentha* spp. (Bo He),[17]** *Vitex* spp. (Man Jing Zi)[17]
Lignan (neolignan)	wind cold	*Schizonopeta tenuifolia*[a] (Jing Jie)
monoterpene (iridoid, menthane)	heat (blood),[1] wind cold[2]	*Rehmannia glutinosa* (Di Huang),[1] *Perilla frutescens* (Zi Su Ye)[2]
Polycylic aromatic (anthraquinone)	heat (qi),[1] laxative[2]	*Cassia obtusifolia* (Jue Ming Zi),[1] *Rheum* spp. (Da Huang)[2]
tannin (hydrolizable)	stop bleeding	*Sanguisorba officinalis* (Di Yu)

[a] Herbs found in >1 phytochemical class.

as *Angelica* species (involving acetylenes, coumarins, and sesquiterpenes), *Ginkgo biloba* (flavonoids and sesquiterpenes), and *Curcuma* species (diarylalkyl phenolics and sesquiterpenes).

3.c. COX and LOX. The role of eicosanoid metabolism in inflammation needs no introduction, and inhibitors of both COX and LOX, the NSAIDS, are among the most powerful therapeutic agents in inflammation-related disorders, though selective inhibitors of COX-2, originally introduced to offset unacceptable side effects of nonselective COX inhibition, have proved problematic in their own right.[36] More recently, attention has been focused on dual COX/LOX inhibitors, because a number of side effects, for example, gastric erosion, are thought to be due to overproduction of gastrotoxic leukotrienes via the LOX pathway as a result of COX inhibition.[37]

Plants are already known to be reservoirs of compounds which inhibit both COX and LOX,[38] and it is likely that a variety of plant compounds may serve either as leads for the development of dual inhibitors or as valuable therapeutic agents in their own right.

In Table 5, skeletal types and herbs marked with either a C or L superscript contained hits against either COX or LOX. Unmarked entries contained hits against both. It should be stressed, however, that many herbs may show some degree of affinity for both, owing to the close degree of structural similarity between COX and LOX binding sites.

A wide variety of aliphatics and phenolics, in particular, are identified and are thus categorized separately in Table 5. Of particular interest are long-chain aromatics (which combine features of both), acetylenes, unsaturated fatty acids, and diarylalkyl phenolics, though flavonoids also feature to an appreciable extent, and a small number of alkaloids (particularly isobutylamide) are also of probable significance. Simple phenolics and stilbenes are also found.

The number of herbs with hits from more than one phytochemical class is particularly noticeable. These include *Zingiber officinale* (ginger) in which long-chain aromatics, phenols, and diarylalkyl phenolics are identified; *Alpinia* species where flavonoids, diarylalkyl phenolics, and stilbenes may all play a role; and *Artemisia capillaris* in which acetylenes, coumarins, and flavonoids are selected.

The similarity between phytochemical classes involved in NO inhibition and those predicted as COX/LOX inhibitors is also striking, suggesting that anti-inflammatory compounds in plants may have a broad spectrum of activity against a number of major targets. The following classes are predicted to inhibit both: acetylenes, quinazoline alkaloids, coumarins, flavonoids, lignans (neolignan), diarylalkyl and stilbene phenolics, and sesquiterpenes. Furthermore, a total of 21 species are identified as both NO and COX/LOX inhibitors, including *Panax ginseng*, *Bupleurum* spp., *Artemisia capil-*

Table 7. Predicted Inhibitors of HIV-1 Integrase, Protease, and Reverse Transcriptase in Chinese Herbs (I = Integrase, P = Protease, T = Reverse Transcriptase)

phytochemical class (skeletal types)	TCM category	herbs with multiple hits
Aliphatic (aldaric acid,[I] cyclitol,[I] monocarbocyclic*[I])	fire poison,[1] wind heat[2]	*Lonicera japonica* (Jin Yin Hua),[1I] *Arctium lappa*[a] (Niu Bang Zi),[2I] *Cimicifuga simplex* (Sheng Ma)[2I]
alkaloid (protoberberine[T])	antitussive,[1] damp heat,[2] invigorate blood[3]	*Nandina domestica* (Tian Zhu Zi),[1T] *Coptis* spp. (Huang Lian),[2T] *Phellodendron* spp. (Huang Bai),[2T] *Corydalis* spp. (Yan Hu Suo)[3T]
Carotenoid[T]	fire poison	*Taraxacum* spp.[a] (Pu Gong Ying)[T]
diterpene (abietane,[PT] ingenane*,[P] kaurane[T])	aromatic digestive,[1PT] cathartic*,[2P] wind damp,[3PT]	*Agastache rugosa* (Tu Huo Xiang),[1PT] *Siegesbeckia pubescens* (Xi Xian Cao),[3T] *Euphorbia kansui* (Gan Sui)*,[2P] *Trypterygium wilfordii*[a] (Lei Gong Teng)[3P]
Flavonoid (biflavonoid,[T] chalcone*,[I] flavanol,[LP] flavone,[IPT] flavonol,[IPT] isoflavone*,[P])	antitussive,[1I] damp heat,[2I] diuretic,[3I] fire poison,[4P] heat (deficient),[5I] phlegm cold,[6I] regulate qi,[7P] stop bleeding,[8T] tonify qi,[9I] tonify yang,[10I] wind damp,[11I] wind heat,[12I]	***Ginkgo biloba* (Bai Guo Ye),[1IPT]** ***Scutellaria baicalensis* (Huang Qin),[2IPT]** *Pyrrosia petiolosa*[a] (Shi Wei),[3I] *Erythrina variegata* (Hai Tong Pi),[3T] *Sophora subprostrata*[a] (Shan Dou Gen),[4P] *Belamcanda chinensis* (She Gan),[4P] *Artemisia annua* (Qing Hao),[5I] *Inula brittanica*[a] (Xuan Fu Hua),[6I] *Dalbergia odorifera* (Jiang Xiang),[7P] *Artemisia vulgaris*[a] (Ai Ye),[8T] *Platycladus orientalis* (Ce Bai Ye),[8T] ***Glycyrrhiza* spp. (Gan Cao),[9I]** *Epimedium* spp. (Yin Yang Huo),[10I] ***Morus alba* (Sang Zhi),[11IT]** *Bidens pilosa* (Xian Feng Cao),[12I] *Equisetum arvense* (Mu Zei),[12I] *Lippia nodiflora* (Peng Lai Cao)[12I]
Lignan (dibenzylbutane,[IT] dibenzocyclooctadiene,[T] neolignan*[IT])	aromatic digestive,[1T] astringent,[2T] fire poison,[3T] invigorate blood,[4I] wind cold,[5I] wind heat[6IT]	*Magnolia* spp. (Hou Po),[1T] *Myristica fragrans* (Rou Dou Kou),[2T] *Schisandra chinensis* (Wu Wei Zi),[2T] *Justicia procumbens* (Jue Chuang Cao),[3T] *Salvia miltiorrhiza*[a] (Dan Shen),[4I] *Schizonopeta tenuifolia* (Jing Jie),[5I] ***Arctium lappa*[a] (Niu Bang Zi)[6IT]**
Phenolic (acylphloroglucinol,[T] diarylalkyl,[I] phenylpropanoid[I])	diuretic,[1I] fire poison,[2T] invigorate blood,[3I] phlegm cold,[4I] wind cold[5I]	*Plantago asiatica* (Che Qian Zi),[1I] *Pyrrosia petiolosa*[a] (Shi Wei),[1I] *Hypericum japonicum*[a] (Tian Ji Huang),[2T] *Curcuma* spp. (E Zhu, Yu Jin),[3I] *Salvia miltiorrhiza*[a] (Dan Shen),[3I] *Inula brittanica*[a] (Xuan Fu Hua),[4I] *Coriandrum sativum* (Yan Sui Zi)[5I]
Phenolic (benzophenone*,[I] stilbene*,[I] xanthone*[I])	damp heat,[1I] fire poison,[2I] laxative[3I]	*Gentiana lutea* (Long Dan Cao),[1I] *Hypericum japonicum*[a] (Tian Ji Huang),[2I] *Rheum* spp.[a] (Da Huang)[3I]
polycyclic aromatic (anthracene*,[IT] anthraquinone[I])	heat (qi),[1I] laxative,[2IT] stop bleeding,[3I] tonify yang[4I]	*Cassia obtusifolia, C. tora* (Jue Ming Zi),[1I] ***Rheum* spp.[a] (Da Huang),[2IT]** *Rubia cordifolia* (Qian Cao Gen),[3I] *Morinda* spp. (Ba Ji Tian)[4I]
Tannin (condensed*,[P] hydrolizable[IPT])	astringent,[1IPT] internal cold,[2IT] laxative,[3IT] phlegm heat,[4P] open heart,[5T] stop bleeding[6IPT]	***Cornus officinalis* (Shan Zhu Yu),[1PT]** *Ephedra sinica* (Ma Huang Gen, root),[1I] ***Punica granatum* (Shiu Liu Pi),[1IPT]** *Rosa laevigata* (Jin Ying Zi),[1P] ***Terminalia chebula* (He Zi),[1IPT]** ***Syzygium aromaticum* (Ding Xiang),[2IT]** ***Rheum* spp.[a] (Da Huang),[3IT]** *Euphorbia helioscopia* (Ze Qi),[4P] *Liquidambar formosana* (Su He Xiang),[5T] *Platycladus orientalis* (Ce Bai Ye),[6T] *Agrimonia pilosa* (Xian He Cao),[6I] ***Sanguisorba officinalis* (Di Yu)[6IT]**
Triterpene (bauerane*,[P] friedelane,[PT] lanostane,[P] lupane,[PT] oleanane,[P] ursane[PT])	antitussive,[1P] diuretic,[2P] fire poison,[3P] heat (blood),[4P] invigorate blood,[5P] nourish heart,[6P] stop bleeding,[7T] wind damp[8PT]	*Tussilago farfara* (Kuan Dong Hua),[1P] *Poria cocos* (Fu Ling),[2P] *Forsythia suspensa* (Lian Qiao),[3P] *Sophora subprostrata*[a] (Shan Dou Gen),[3P] *Taraxacum japonicum*[a] (Pu Gong Ying),[3P] *Paeonia suffruticosa* (Mu Dan Pi),[4P] *Boswellia carteri* (Ru Xiang),[5P] *Commiphora* spp. (Mo Yao),[5P] ***Ganoderma lucidum* (Ling Zhi),[6P]** *Zizyphus spinosus* (Suan Zao Ren),[6P] *Abrus cantoniensis* (Ji Gu Cao),[8P] *Artemisia vulgaris*[a] (Ai Ye),[7T] ***Trypterygium wilfordii*[a] (Lei Gong Teng)[8PT]**

[a] Herbs found in >1 phytochemical class.

laris, Ginkgo biloba, Scutellaria baicalensis, Angelica sinensis and *A. pubescens, Glehnia littoralis, Inula helenium, Magnolia obovata,* and *Rheum* spp.

3.d. Aldose Reductase. AR is a key enzyme in the polyol pathway which catalyzes the NADPH-dependent reduction of glucose to sorbitol. The latter is then converted to fructose by sorbitol dehydrogenase. Normally, this pathway plays a minor role in glucose metabolism. In diabetic hyperglycaemia, however, cells which undergo the insulin-independent uptake of glucose accumulate significant quantities of sorbitol. This is due to poor penetration of sorbitol through cell membranes coupled to slow metabolism by sorbitol

dehydrogenase, both of which favor its intracellular accumulation. The resulting osmotic stress is thought to be the primary cause of diabetic complications which include cataracts, retinopathy, nephropathy, and neuropathy.[39] This in turn suggests that aldose reductase inhibitors can prevent such complications, though the performance of a number of pharmaceutical AR inhibitors has proved disappointing.[40]

A wide variety of plant extracts have been tested for their effects on aldose reductase,[41] and individual compounds with known inhibitory activity to date are mainly flavonoids or are flavonoid-related. This is clearly reflected in Table 6 where by far the largest number of hits are found in flavonoid-containing herbs. Those herbs hit with greatest frequency include *Ginkgo biloba*, *Scutellaria* spp, *Lippia nodiflora*, and *Mentha* species. Over 40 flavonoid-containing species exhibited multiple hits, indicating that AR inhibitors are likely to be widespread in Chinese herbs and furthermore fall into many traditional categories (17 so far of the 30 or so major categories). From the viewpoint of TCM, it is perhaps significant that a number of "wind heat" herbs feature prominently because such herbs are sometimes used to treat eye-related disorders.

Other phytochemical classes are, by contrast, very poorly represented, though the diterpenes found in *Salvia miltiorrhiza* may merit further attention. Iridoids and menthane monoterpenes are also of potential significance.

3.e. HIV-1 Integrase, Protease, and Reverse Transcriptase. Among the most important targets for HIV therapy are those viral enzymes which catalyze a number of essential steps in its lifecycle. HIV-1 reverse transcriptase is responsible for the transcription of viral RNA into DNA, which is then integrated into host nuclear DNA by the action of HIV-1 integrase. The HIV proteins subsequently made in host cells are synthesized as large polyproteins and must first be cleaved by HIV-1 protease to give the smaller proteins found in mature, infectious virions.

A wide variety of phytochemicals are now known to inhibit one or more of these targets, including di- and triterpenes, lignans, flavonoids, tannins, aliphatics, xanthones, and others.[42] A number of Chinese herbal extracts have also been tested in vitro and show promise,[43] though a recent meta-analysis of existing data proved inconclusive and suggested that larger and more rigorous trials were needed.[44]

The results for all three HIV targets are presented in Table 7. The following superscripts are used to indicate the target(s) for each skeletal type and herb: I = integrase, P = protease, and T = reverse transcriptase.

In the case of HIV-1 integrase, flavones, lignans (dibenzylbutane and neolignan), anthraquinones, tannins, and a limited number of phenolics were identified as possible inhibitors, though a number of rarer classes, such as organic acids and cyclitols, were also selected.

For HIV-1 protease, a smaller number of classes were hit, comprising diterpenes, triterpenes, flavonoids (again mainly flavones), and tannins. Of the three targets, HIV-1 protease received by far the largest number of hits for terpenoid compounds, though they were not entirely absent for HIV-1 reverse transcriptase. Unlike the other two targets lignans were absent.

A fairly wide variety of classes are implicated as potential HIV-1 reverse transcriptase inhibitors, including protoberberine alkaloids, diterpenes and triterpenes (though to a lesser extent than protease), flavonoids, tannins, and lignans.

A number of species have more than one target (about 19%), and these are particularly prevalent among tannin-containing herbs. Examples include *Cornus officinalis*, *Punica granatum*, *Terminalia chebula*, *Rheum* species, *Syzygium aromaticum*, and *Sanguisorba officinalis*. Herbs containing flavonoids are also predicted to inhibit more than one target, such as *Ginkgo biloba*, *Scutellaria baicalensis*, and *Morus alba*.

As with PKA, nitric oxide, and COX/LOX, single herbs with compounds from more than one phytochemical class are not uncommon and include the following: *Trypterygium wilfordii* (di- and triterpenes), *Rheum* species (tannins and anthraquinones), *Arctium lappa* (monocarbocyclics and lignans), *Artemisia vulgaris* and *Sophora subprostrata* (triterpenes and flavonoids), and *Inula britannica* (phenolics and flavonoids).

4. Scaffold Hopping. A potentially valuable aspect of virtual screening concerns its ability to identify new "scaffolds" for drug discovery on the basis of features such as pharmacophores, descriptors, and docked poses, a process known as *scaffold hopping*. Given the relatively stringent criteria applied here, we would not expect Random Forest to identify substantially different scaffolds from those in the training sets employed in this study. For all targets, the great majority of hits were found to belong to phytochemical classes and skeletal types already present in the training data. No entirely new phytochemical classes were identified for any target; though, as the asterisks in the tables show, skeletal types *within* each class not already present in the training data were identified an appreciable number of times.

In most cases, these are very similar to existing types with known activity. Unfortunately, space does not permit a detailed exploration of this phenomenon within the context of the present study, and in any event, the conservative nature of our screening protocol served to identify very similar compounds in target data to those used for training and is thus not perhaps the best tool for exploring this important aspect of computational research. Nevertheless, this is an area which deserves more detailed study.

Structural details of all skeletal types listed in Tables 2−7 may be found in the Supporting Information.

5. Corroborative Evidence from the Scientific Literature. Though virtual screening has numerous applications and its efficacy cannot be judged solely on the basis of predictive accuracy, in many cases, its utility is considerably enhanced if predictions can be either verified by subsequent experiment or supported by evidence from the existing literature. Accordingly, an electronic literature search in PubMed was carried out to determine to what extent the lists of herbs (or related species) presented in Tables 2−7 have been observed to possess the activities predicted by our analysis. Clearly, there is a potential element of circularity involved in this, in that the data sets used for training have been derived from the literature in the first place, and may therefore give rise to false estimates of the accuracy of prediction unless this is taken into account.

An analysis of the number of compounds found in both of the data sets used for training and the hit lists identified from Chinese herbs revealed that in the case of most targets the percentage of shared compounds, however, was low.

Table 8. Plant Species (and Constituent Phytochemical Classes Where Known) from the Scientific Literature Which Support the Predictions Made in Tables 2−7 (See Text for Details)

target/process	herbs (phytochemical classes where known)
cAMP PDE	*Glycyrrhiza* sp (coumarins, flavonoids),[1a] *Ginkgo biloba* (biflavones),[2] *Scutellaria baicalensis* (flavonoids),[3] *Sophora flavescens* (flavanones, isoflavones),[3] *Eucommia ulmoides* (lignans),[4a] *Lilium* spp. (steroidal saponins),[5] *Polygonatum sibiricum*,[6] *Smilax sieboldii* (steroidal saponins)[7a]
PKA	***Erythrina addisoniae*** (isoflavonoids),[8] *Rheum palmatum* (anthraquinone)[9]
iNOS expression *or* NO production*	*Panax ginseng* (acetylene),[10] ***Angelica furcijuga*** (acetylenes, coumarins),[11a] *Zanthoxylum bungeanum* (phenylpropanoid),[12] *Salvia miltiorrhiza* (diterpenes),[13] *Tripterygium wilfordii* (diterpene),[14] *Ginkgo biloba* (flavonoids),[15] *Magnolia obovata* (lignans),[16] *Curcuma zedoaria* (diarylheptanoids, sesquiterpenes),[17,18a] *Inula brittanica* (sesquiterpenes),[19,20] *Saussurea lappa* (sesquiterpenes),[21a] ***Artemisia sylvatica*** (sesquiterpenes),[22] ***Atractylodes japonica***,[23] *Xanthium* sp.,[24] *Bidens pilosa**,[25] ***Bupleurum chinense****,[26] *Corydalis* sp (alkaloids)*,[27] *Cnidium officinale**,[28] *Scutellaria baicalensis* (flavonoids)*,[29] *Smilax* spp.*,[30,31] *Sophora* spp. (isoflavonoids)*,[32] *Glycyrrhiza* sp (flavonoid)*,[33] *Coptis japonica* (lignans)*,[34a] ***Terminalia myriocarpa*** (ellagitannins)*,[35] ***Chrysanthemum indicum*** (sesquiterpenes)*[36]
COX,[C] LOX,[L] *or* both*	***Artemisia monosperma*** (acetylenes),[L37] *Angelica pubescens* (acetylenes, coumarins)*,[38a] ***Bupleurum fruticescens****,[39] ***Bidens campylotheca*** (acetylenes)*,[40] *Zingiber officinale**,[41a] *Allium* spp. (sulfides)*,[42,43] *Isatis* spp. (alkaloids)*,[44a,45] *Piper* spp. (amine alkaloids)*,[46] *Ginkgo biloba* (biflavone)*,[47a] *Scutellaria baicalensis* (flavone),[L48,49] *Sophora flavescens* (flavanone)*,[50] ***Erythrina spp. (pterocarpene, flavanones),***[L51,52] *Alpinia* spp. (chalcone, diarylheptanoids)*,[53,54a] *Glycyrrhiza* sp (chalcone)*,[55] *Astragalus membranaceus*,[L56] *Morus alba* (stilbenes)*,[57a] ***Mentha longifolia*** (flavonoids?),[L58] *Magnolia officinalis* (lignan)*,[59a] *Curcuma* spp. (cucuminoids)[L60]
Aldose reductase *or* diabetic complications*	*Perilla frutescens* (monoterpene glucosides),[61a] *Glycyrrhiza uralensis*,[62] *Nelumbo nucifera* (flavonoids),[63] ***Artemisia dracunculus*** (flavonoids),[64] *Scutellaria baicalensis* (flavonoids),[62] *Belamcanda chinensis* (isoflavonoids),[65] *Salvia miltiorrhiza*,[62] *Ginkgo biloba* [nephropathy*],[66a] *Zea mays* [nephropathy*],[67] *Scutellaria baicalensis* [microvascular complications*],[68] ***Pinus pinaster*** (flavonoids) [retinopathy*],[69] *Rehmannia glutinosa* [nephropathy*],[70] *Rheum officinale* [nephrophathy*][71]
HIV-1 reverse transcriptase,[T] protease,[P] integrase,[I] *or* HIV replication (unknown target)*	*Corydalis yanhusuo* (alkaloids),[T72] *Agastache rugosa* (diterpenes),[P73] ***Euphorbia paralias*** (diterpenes)*,[74] *Tripterygium wilfordii* (diterpenes),[T75] *Scutellaria baicalensis* (flavone),[T76] ***Sophora spp.*** (flavanones)*,[77] *Glycyrrhiza* sp (flavonoids)*,[78] *Justicia gendarussa*,[T79] *Salvia miltiorrhiza* (benzofurans),[I80] *Curcuma longa* (curcuminoids),[P81,I82] ***Cornus kousa***,[T83] *Agrimonia pilosa*,[TP83] *Terminalia* spp. (tannins),[I84,T85] ***Euphorbia pekinensis*** (tannins),[I84] ***Poria spp.***,[T86] *Paeonia suffruticosa* (terpenes?),[I87] *Ganoderma lucidum* (triterpenes)[P88a]

Thus, for HIV targets, the percentages of compounds hit which are identical with those in the training set were 5.77% (reverse transcriptase), 3.49% (protease), and 0.79% (integrase). In the case of aldose reductase, the figure was 4.14%. For PKA, it was 2.87%, and for nitric oxide 2.19% (iNOS expression) and 9.57% (NO production in vivo). In the cases of cAMP PDE, COX, and LOX, the percentage of shared compounds was appreciably higher, namely, 16.67% (cAMP PDE), 20.65% (COX), and 22.13% (LOX). These figures indicate that, for the last three targets, Chinese herbs have either been used to furnish a reasonable proportion of the evidence concerning known inhibitors in the first place or the compounds involved are widely distributed. Nevertheless, even here, approximately 80% or more compounds are newly predicted in the case of each target. Had our selection criteria been less stringent and the number of hits consequently larger, the percentages above would be much reduced. However, this might come at the price of lower predictive accuracy.

Table 8 gives details of the herbs or related species in the same genus identified in the literature search, where the reported activities accord with the predictions made in Tables 2−7. Those herbs with compounds identical to those used in training have not been excluded, because, with two or more hits required for inclusion in Tables 2−7, they were found to contain other compounds not previously identified as potentially active. They are, however, given a superscript "*a*". Species closely related to, but not identical with, those in Tables 2−7 are underlined. In cases where the chemical classes involved are known (shown in parentheses in Table 8), only those which accord with our predictions have been included.

In the case of aldose reductase, other information suggestive of a possible association is given. Known inhibitors of AR are listed first, while herbs with effects on diabetic complications are also included (though their direct effects on AR are unknown) because it is likely that such effects may be due in part to AR inhibition.

The large number of references in Table 8 (indicated in superscript following each herb) may be found in the Supporting Information. It should be noted that many of these references are to very recently published papers, the information of which has not as yet been incorporated into our databases. Approximately 50% (45 of the 88 papers listed), for instance, have been published since 2003.

DISCUSSION

1. Advantages of Random Forest. Among the various methods that have been proposed for ligand-based virtual

screening, we suggest that Random Forest, and other related methods such as boosted trees or support vector machines, have much to offer. Their ease of use, flexibility, insensitivity to correlated descriptors, and ability to handle highly unbalanced data are especially appealing, particularly for training on limited numbers of active compounds which may be of arbitrary structural diversity.

Another useful feature concerns their ability to assess the suitability of different sets of descriptors on the basis of misclassification and related measures. In this instance, it was found that Kier-Hall descriptors outperformed the others, though only marginally in the case of MOE and Labute descriptors, while the MACCS set performed poorly overall. If consensus based on results from different analyses is sought, then the highest-scoring sets can be easily identified and the resulting scores fused.

Additionally, the degree of rigor required in identifying hits can be adjusted according to the votes given by the full ensemble, as well as introducing constraints on data fusion when consensus is applied.

2. Patterns of Prediction in Relation to Individual Targets. Considering the rather stringent selection criteria applied and the relatively small number of targets investigated, it may seem surprising that such a large proportion of species (approximately 62%) was hit two or more times. This is probably due to a number of factors, the first being the high degree of structural overlap between bioactive compound and Chinese herb data sets. In addition, many phytochemicals are known for their ability to inhibit a range of targets,[16] which suggests that the ligand-receptor space of plant compounds is likely to be densely populated for a number of classes and that even limited data would tend to yield high hit rates. If so, then we would expect many species to show multiple target affinities, which again appears to be the case, with 33% of the total number of species in the Chinese herb database predicted to contain compounds which inhibit two or more targets. Of further interest are the different phytochemical classes within the same herb predicted to inhibit the same target, suggesting that plants may sometimes make use of alternative phytochemical strategies in target inhibition. Informatics may prove a useful way of identifying such effects, an area which is difficult and time-consuming from an experimental or clinical perspective.

As far as individual targets are concerned, a number of our results merit further attention. In the cases of cAMP PDE and PKA, which have opposing and complementary roles in cell signal regulation, it was found that predicted inhibitors were structurally quite distinct, with the exception of flavonoids, and belong to different phytochemical classes. To adopt the language of TCM, the "yin and yang" of cAMP-driven cellular flux may thus be influenced by different groups of plant compounds, on one hand, those which promote higher levels of cAMP and consequent PKA activity, comprising mainly steroidal compounds and coumarins, as well as a number of flavonoids and lignans, while on the other hand others inhibit PKA activity and belong more to classes such as the phenolics, anthraquinones, tannins, and nonsteroidal triterpenes as well, again, as several flavonoids.

It was further postulated that plant steroidal compounds, despite that fact that they are not precursors in vivo of mammalian steroidal hormones, might nevertheless exert some indirect influence in this direction by virtue of their ability to inhibit cAMP PDE and thus promote the steroidogenic activity of PKA. This, however, is purely speculative and awaits further investigation.

In the case of nitric oxide, one of the more interesting features concerned the large number of sesquiterpenes identified, from a variety of skeletal types, a number of them such as bisabolane, clovane, and xanthane not present among examples used for training. In addition, acetylenes, coumarins, and flavonoids appear to play a potential role as well as dammarane triterpenes from ginseng species.

The COX and LOX inhibitors share much in common with those involved in nitric oxide, and a variety of species were identified which may affect both systems. However, long-chain aromatics and fatty acids were not identified in the case of the former. There is, nevertheless, considerable phytochemical overlap between the two, and it seems likely that there may be many more species which have inhibitory effects on both eicosanoid and NO synthesis than are listed above, though in most cases, such effects may be relatively mild. In the search for anti-inflammatory drugs with multiple targets, plant compounds nevertheless represent a rich resource.

The great majority of compounds predicted to inhibit aldose reductase are flavonoids, and little is known about other inhibitors. This is clearly an area where further information is required, and a promising theoretical approach might involve protein-based virtual screening via ligand−receptor docking. This would help to identify other classes to which future experimental work could be directed.

Potential inhibitors of HIV targets appear to be widespread in Chinese herbs in terms of both phytochemical classes and herbs. Approximately 28% of the herbs listed in our database are implicated. It is also of interest that about 20% of these appear to show inhibitory potential against more than one HIV target and that these herbs fall into a fairly broad range of TCM categories which suggests that, *if* phytochemicals prove effective in the control of HIV/AIDS, then different formulaic prescriptions could be "customized" on the basis of the patient's TCM diagnosis. It should be noted however that the "warming" categories of Chinese herbal medicine (internal cold, aromatic digestive, wind cold, and tonify yang) are notably less frequent than the categories which clear "heat" in various ways. Again, this is of interest in that heat-related signs and symptoms tend to be prevalent in the TCM diagnosis of HIV infection and its complications,[45] and it is therefore intriguing to find that the majority of predicted inhibitors are found among herbs from categories that would tend to be used preferentially from a TCM perspective.

3. Relationship of Predictions to the Scientific Literature. Though the evidence to support our predictions in Table 8 cannot be considered as "validation" of the method presented (the scientific literature cannot be used to provide information on false predictions for instance), the number of herbs or related species where experimental work has provided corroborative evidence is nevertheless striking (83 out of a total of 284 herb-target pairs, or approximately 29%). Furthermore, as discussed above, the degree of identity between compounds used for training and those identified as potential inhibitors in Chinese herbs is low in most cases, and much of the supporting evidence is taken from very

recently published papers, where the information in question has not yet been added to our data.

In the case of COX, LOX, and cAMP PDE, where overlap between training data and hit lists are higher, a larger proportion of herbs in Table 8 are given a superscripted "*a*" compared to other targets, where compounds known to be found in these or closely related species were used for training. Even so, only 17 examples (out of 83) were detected, indicating that about 80% of the species in Table 8 provide "new" evidence to support our predictions.

These findings therefore strongly suggest that virtual screening may prove effective as a means of identifying new species and chemical classes with appropriate activities and may be used to focus experimental investigations in a more effective fashion.

CONCLUSION

It is hoped that the results of this study demonstrate something of the potential of virtual screening both in identifying potential candidates for drug development from natural products and, equally importantly, in addressing some of the broader issues which are of interest in phytotherapy and related fields. The relatively limited amount of phytochemical information available, in an appropriate format, still places appreciable constraints on informatics-based research in this area, though much has been done to address this in recent years.

More focused studies, using techniques such as ligand−receptor docking, are also required, as are theoretical approaches to questions such as the possible role of weak inhibitors in modulating metabolic or signal transduction pathways, something which impacts directly on the use of many herbs and may furthermore provide new insights into the development of next-generation pharmaceuticals without some of the problems currently associated with strong, selective inhibitors.

Predictions are made here largely on the basis of in vitro experiments. Further work, particularly on the absorption, distribution, metabolism, and excretion of plant compounds, is needed to assess the role such predictions may have in vivo.

Supporting Information Available: The following are available: (1) a description of the major therapeutic categories of TCM, (2) diagrams of the full set of skeletal types referred to in Tables 2−7, and (3) the references to the literature given in Table 8. This information is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Clardy, J.; Walsh C. Lessons from Natural Molecules. *Nature Insight (Chem. Space)* **2004**, *432*, 829−837.

(2) Jones, W. P.; Chin, Y.-W.; Kinghorn, A. D. The Role of Pharmacognosy in Modern Medicine and Pharmacy. *Curr. Drug Targets* **2006**, *7* (3), 247−264.

(3) *Traditional Chinese Medicines 2.1*; CambridgeSoft Inc.: Cambridge, MA.

(4) *China Natural Products Database*; NeoTrident Technology Ltd.: Beijing, P. R. China.

(5) Ehrman, T. M.; Barlow, D. J.; Hylands, P. J. Phytochemical Databases of Chinese Herbal Constituents and Bioactive Plant Compounds with Known Target Specificities. *J. Chem. Inf. Model.* **2007**, *47*, 254−263.

(6) Friedman, J. H. *Stochastic Gradient Boosting*; Technical Report; Department of Statistics, Stanford University, 1999.

(7) Svetnik, V.; Wang, T.; Tong, D.; Liaw, A.; Sheridan, R. P.; Song, Q. Boosting: An Ensemble Learning Tool for Compound Classification and QSAR Modeling. *J. Chem. Inf. Model.* **2005**, *45*, 786−799.

(8) Breiman, L. Decision Tree Forests. *Mach. Learning* **2001**, *45*, 5−32.

(9) Svetnik, V.; Liaw, A.; Tong, D.; Culberson, C.; Sheridan, R. P.; Feuston, B. P. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *J. Chem. Inf. Comput. Sci*. **2003**, *43*, 1947−58.

(10) Dashwood, R. H.; Myzak, M. C.; Ho E. Dietary HDAC Inhibitors: Time to Rethink Weak Ligands in Cancer Chemoprevention? *Carcinogenesis* **2006**, *27* (2), 344−349.

(11) Hsu, H.-Y.; Chen, Y.-P.; Hong, M. *The Chemical Constituents of Oriental Herbs*; Oriental Healing Arts Institute: Taiwan, 1982−83; Vols. I and II.

(12) Zhu, Y.-P. *Chinese Materia Medica: Chemistry, Pharmacology & Applications*; Harwood Acadademic Publishers: Amsterdam, 1998.

(13) Yan, X.; Zhou, J.; Xie, G. *Traditional Chinese Medicines: Molecular Structures, Natural Sources, and Applications*; Milne, G. W. A., Ed.; Ashgate Publishers Ltd.: Aldershot, U. K., 1999.

(14) Duke, J. A. *Handbook of Phytochemical Comstituents of GRAS Herbs and Other Economic Plant*; CRC Press: Boca Raton, FL, 1992.

(15) *Dictionary of Natural Products* [CD-ROM]; Chapman & Hall/CRC: Boca Raton, FL.

(16) Polya, G. *Biochemical Targets of Plant Bioactive Compounds: A Pharmacological Reference Guide to Sites of Action and Biological Effects*; Taylor & Francis: London, 2003.

(17) Hall, L. H.; Kier, L. B. The Molecular Connectivity Chi Indices and Kappa Shape Indices in Structure−Property Modeling. *Rev. Comput. Chem.* **1991**, *2*, 367−422.

(18) Kier, L. B.; Hall, L. H. *Molecular Structure Description: The Electrotopological State*; Academic Press: London, 1999.

(19) Labute, P. A Widely Applicable Set of Descriptors. *J. Mol. Graphics Modell.* **2000**, *18*, 464−77.

(20) Brown, R. D.; Martin, Y. C. Use of Structure-Activity Data to Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 572−584.

(21) Freund, Y.; Schapire, R. A. Decision-Theoretic Generalization of Online Learning and an Application to Boosting. *J. Comput. Syst. Sci*. **1997**, *55*, 119−139.

(22) Breiman, L.; Friedman, J. H.; Olshen R. A.; Stone C. J. *Classification and Regression Trees*; Chapman & Hall/CRC: Boca Raton, FL, 2000; Chapter 4, pp 103−104.

(23) Chen, C.; Liaw, A.; Breiman, L. *Using Random Forest to Learn Imbalanced Data*; Technical Report 666; Statistics Department, University of California at Berkeley, Berkeley, CA, 2004.

(24) Baurin, N.; Mozziconacci, J.-C.; Arnoult, E.; Chavatte, P.; Marot, C.; Morin-Allory, L. 2D QSAR Consensus Prediction for High-Throughput Virtual Screening. An Application to COX-2 Inhibition and Screening of the NCI Database. *J. Chem. Inf. Comput. Sci*. **2004**, *44*, 276−285.

(25) Bensky, D.; Clavey, S.; Stöger, E. *Chinese Herbal Medicine: Materia Medica*; Eastland Press: Seattle, WA, 2004.

(26) Walsh, D. A.; Van Patten, S. M. Multiple Pathway Signal Transduction by the cAMP-Dependent Protein Kinase. *FASEB J.* **1994**, *8*, 1227−1236.

(27) Hauet, T.; Liu, J.; Li, H.; Gazouli, M.; Culty, M.; Papadopoulos, V. PBR, StAR, and PKA: Partners in Cholesterol Transport in Steroidogenic Cells. *Endocr. Res.* **2002**, *28* (4), 395−401.

(28) Baillie, G. S.; Scott, J. D.; Houslay, M. D. Compartmentalisation of Phosphodiesterases and Protein Kinase A: Opposites Attract. *FEBS Lett.* **2005**, *579*, 3264−3270.

(29) Cummings, D. E.; Brandon, E. P.; Planas, J. V.; Motamed, K.; Idzerda, R. L.; McKnight G. S. Genetically Lean Mice Result from Targeted Disruption of the RII Beta Subunit of Protein Kinase A. *Nature* **1996**, *382* (6592), 622−626.

(30) Aandahl, E. M.; Ankrust, P.; Skalhegg, B. S.; Muller, F.; Froland, S. S.; Hansson, V.; Tasken K. Protein Kinase A Type I Antagonist Restores Immune Responses of T Cells from HIV-Infected Patients. *FASEB J.* **1998**, *12* (10), 855−862.

(31) Tortora, G.; Ciardiello, F. Protein Kinase A as Target for Novel Integrated Strategies of Cancer Therapy. *Ann. N. Y. Acad. Sci.* **2002**, *968*, 139−147.

(32) Houslay, M. D.; Schafer, P.; Zhang K. Y. J. Phosphodiesterase-4 as a Therapeutic Target. *Drug Discovery Today* **2005**, *10*, 1503−1519.

(33) O'Donnell, J. M.; Zhang H−T. Antidepressant Effects of Inhibitors of cAMP Phosphodiesterase (PDE4). *Trends Pharmacol. Sci.* **2004**, *25* (3), 158−163.

(34) Thomas, G. *Medicinal Chemistry: An Introduction*; Wiley: Chichester, U. K., 2000; Chapter 11, pp 431−463.

(35) Lirk, G.; Hoffmann, G.; Rieder, J. Inducible Nitric Oxide Synthase − Time for Reappraisal. *Curr. Drug Targets: Inflammation Allergy* **2002**, *1*, 89−108.

(36) Drazen, J. M. COX-2 Inhibitors − A Lesson in Unexpected Problems. *N. Engl. J. Med.* **2005**, *352* (11), 1131−1132.

(37) Hatmi, M.; Samama, M. M.; Flalamy, I. Prevention of Thrombosis and Vascular Inflammation: Importance of Combined Cyclooxygenase and 5-Lipoxygenase Inhibitors. *J. Mal. Vasc*. **2006**, *31* (1), 4−9.

(38) Newmark, T. M.; Schulick, P. *Beyond Aspirin: Nature's Challenge to Arthritis, Cancer and Alzheimer's Disease*; Hohm Press: Prescott, AZ, 2000.

(39) Williamson, J.; Kilo, C.; Tilton, R. G. Mechanisms of Glucose- and Diabetes-Induced Vascular Dysfunction. In *Hyperglycemia, Diabetes, and Vascular Disease*; Ruderman, N., Williamson, J., Brownlee, M., Eds; American Physics Society: New York, 1992; pp 107−132.

(40) Chung, S. S.; Chung, S. K. Aldose Reductase in Diabetic Microvascular Complications. *Curr. Drug Targets* **2005**, *6* (4), 475−486.

(41) Kawanishi, K.; Ueda, H.; Moriyasu, M. Aldose Reductase Inhibitors from the Nature. *Curr. Med. Chem*. **2003**, *10* (15), 1353−1374.

(42) Ng, T. B.; Huang, B.; Fong, W. P.; Yeung H. W. Anti-Human Immunodeficiency Virus (Anti-HIV) Natural Products with Special Emphasis on HIV Reverse Transcriptase Inhibitors. *Life Sci.* **1997**, *61* (10), 933−949.

(43) Wu, J. A.; Attele, A. S.; Zhang, L.; Yuan C. S. Anti-HIV Activity of Medicinal Herbs: Usage and Potential Development. *Am. J. Chin. Med.* **2001**, *29* (1), 69−81.

(44) Liu, J. P.; Manheimer, E.; Yang, M. Herbal Medicines for Treating HIV Infection and AIDS. *Cochrane Database Syst. Rev.* **2005**, 1−23.

(45) Zhang, Q.-C.; Hsu, H.-Y. *AIDS and Chinese Medicine*; Keats Publishing Inc.: New Canaan, CT, 1995.

CI600289V