基于官能团和生物特征预测药物 - 靶点相互作用网络

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. 上海大学系统生物学研究所，中国上海2. 中国科学院上海生命科学研究院计算生物学伙伴研究所（中国科学院与马克斯·普朗克学会合作），中国上海3. 上海市杨浦区中心医院眼科，中国上海4. 中国科学院上海生命科学研究院健康科学研究所和上海交通大学医学院，中国上海5. 复旦大学计算系统生物学中心，中国上海6. 上海交通大学附属瑞金医院医学基因组学国家重点实验室，中国上海7. 美国加利福尼亚州圣地亚哥市戈登生命科学研究所

摘要

背景：药物 - 靶点相互作用网络的研究是药物开发中的一个重要课题。仅通过实验来确定化合物 - 蛋白质相互作用或潜在的药物 - 靶点相互作用既耗时又费钱。作为补充，计算机模拟预测方法能够及时为我们提供非常有用的信息。

方法/主要发现：为实现这一目标，药物化合物通过官能团进行编码，蛋白质则通过包括生化和物理化学性质在内的生物学特征进行编码。采用 mRMR（最大相关性最小冗余性）方法进行最优特征选择。不是将蛋白质作为一个整体家族进行分类，而是将靶蛋白分为四组：酶、离子通道、G 蛋白偶联受体和核受体。因此，使用最近邻算法作为运算引擎建立了四个独立的预测器，每个预测器用于预测药物与四组蛋白质之一的相互作用。结果，通过刀切交叉验证测试，四个预测器的总体成功率分别为 85.48%、80.78%、78.49% 和 85.66%。

结论/意义：我们的研究结果表明，由此建立的网络预测系统前景广阔，令人鼓舞。

引用：何 Z，张 J，石 X-H，胡 L-L，孔 X 等（2010）基于官能团和生物特征预测药物-靶点相互作用网络。《公共科学图书馆·综合》5(3)：e9603。doi:10.1371/journal.pone.0009603

编辑：拉米·K·阿齐兹，埃及开罗大学

收稿日期：2009 年 12 月 13 日；接受日期：2010 年 2 月 16 日；发表日期：2010 年 3 月 11 日

资金来源：本研究得到了中国国家基础研究计划（2004CB518603）和中国科学院重点研究项目（KSCX2-YW-R-112）的资助。资助方未参与研究设计、数据收集与分析、发表决定或论文撰写。

利益冲突：作者已声明不存在利益冲突。

* 电子邮件：xykong@sibs.ac.cn（孔晓燕）；cai_yud@yahoo.com.cn（蔡玉德）

请提供需要翻译的原文。这些作者对这项工作贡献相同。

介绍

药物靶点相互作用网络的识别是药物研发流程中的关键步骤[1]。分子医学的兴起以及人类基因组计划的完成为发现药物未知靶点蛋白提供了更多机会。过去几年中，人们为发现新药付出了诸多努力。然而，新药获批的数量仍然相当低（每年仅约 30 种）。部分原因在于许多化合物或候选药物因毒性不可接受而不得不被撤回。这些失败浪费了大量资金。在候选药物合成之前开发计算方法来预测其敏感性和毒性将大有裨益[2,3,4]。然而，要确切了解药物的作用，仍需克服诸多问题。首先，药物可能具有多种作用，包括正面和负面作用，很难找出并阐明其可能的作用；其次，即便药物相同，不同的人对其反应也可能完全不同，即便他们患有相同的疾病。

先，相同的基因产物仅存在细微差异[5,6,7,8]；其次，由于人体内的生物相互作用途径极其复杂，因此很难追踪药物的效果。因此，如果能更准确地预测药物与靶蛋白之间的相互作用，并更好地理解其潜在机制，这对药物研发将大有帮助。

已经开发出了多种计算方法来分析和预测药物与蛋白质的相互作用。其中最常用的有对接模拟[9,10,11,12]、文献文本挖掘[13]以及结合化学结构、基因组序列和三维结构信息[14]等（例如，参见[15,16,17]）。

机器学习和数据挖掘方法已在计算生物学和生物信息学领域得到广泛应用。许多研究人员为开发有用的算法和软件以探究各种与药物相关的生物学问题付出了大量努力，例如 HIV 蛋白酶切割位点预测[18,19]、G 蛋白偶联受体（GPCR）类型的识别[20,21]等。

蛋白质信号肽预测[22]、蛋白质亚细胞定位预测[23,24,25]、GalNAc-转移酶蛋白特异性分析[26]、蛋白酶类型鉴定[27,28]、膜蛋白类型预测[29,30,31,32]，以及一系列相关网络服务器预测工具，如近期综述[33]中所总结的。

在此，我们提出了一种基于最近邻算法[34]的药物-靶点相互作用预测器。由于生化和物理化学特征[35]对于表征蛋白质十分重要，因此在本研究中，我们采用这些特征来表示蛋白质，正如许多先前的研究者所做的那样（例如[36,37,38]）。为了提高预测器的性能，我们使用最小冗余最大相关性（mRMR）算法[39]对特征进行排序。同时，应用增量特征选择和前向特征选择进行特征选择。在本研究中，将药物的蛋白质靶点分为酶、离子通道[40,41,42,43]、G 蛋白偶联受体（GPCRs）[44,45]和核受体[14]。最后，开发了四个预测器，分别用于预测药物与这四个蛋白质家族的相互作用，期望它们能为药物设计提供有用的信息。

材料与方法

基准数据集

除了 Yamanishi 等人[14]所使用的数据集之外，还可以通过 FTP 操作从 KEGG [46,47]获取有关药物化合物和基因的信息：药物信息从 ftp://ftp.genome.jp/pub/kegg/ligand/drug/drug 获取，基因信息从 ftp://ftp.genome.jp/pub/kegg/genes/fasta/gene.pep 获取。在排除缺乏实验信息的药物-靶点对之后，我们最终获得了总共 4797 个药物-靶点对，其中针对酶的有 2719 个，针对离子通道的有 1372 个，针对 G 蛋白偶联受体（GPCRs）的有 630 个，针对核受体的有 82 个。所有这些数据集在本研究中均被用作正数据集。

相应的负数据集是通过以下步骤从上述正数据集中得出的：（1）将上述正数据集中的成对组合拆分为单个药物和蛋白质；（2）将这些单个药物和蛋白质重新组合成对，且确保这些组合在对应的正数据集中均未出现；（3）随机选取这样形成的负对，直至其数量达到正对数量的两倍。

针对酶、离子通道、G 蛋白偶联受体（GPCRs）和核受体所获得的药物靶点基准数据集分别见在线支持信息 S1、S2、S3 和 S4。

特征向量构建

用化学官能团组成来表示药物。药物的数量极其庞大。然而，其中大多数是小分子有机物，由一些固定的较小结构组成，这些结构被称为官能团。由于官能团通常能代表化合物的特性以及其与其他分子的反应机制，因此从官能团中提取的特征在表征药物方面可能非常有效。此外，常见的官能团数量相当少，因此可以利用官能团组成来唯一地表示一种药物[48]。自然界中存在许多官能团，我们为当前研究选择了以下 28 种常见官能团：（1）醇；（2）醛；（3）酰胺；（4）胺；（5）羟肟酸；（6）磷；（7）羧酸盐；（8）甲基；（9）酯；（10）醚；（11）亚胺；（12）酮；（13）硝基；（14）卤素；（15）硫醇；（16）磺酸；（17）砜；（18）磺酰胺；（19）亚砜；（20）硫醚；（21）5 碳环；（22）6 碳环；（23）非芳 5 碳环；（24）非芳 6 碳环；（25）杂芳 6 环；（26）杂环非芳香五元环，（27）杂环非芳香六元环，以及（28）杂环芳香五元环。因此，依照同样的文献[23]所述，现在可以将一种药物化合物表示为如下所示的 28 维向量：

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

其中 g_i（i = 1, 2, ， 28）是药物 D 中第 i 个官能团的出现频率，T 为矩阵转置运算符。

通过结合生化和物理化学特征，用伪氨基酸组成来表示目标蛋白质。目前的问题是如何有效地表示目标蛋白质。在这方面通常使用两种表示方法：顺序表示法和非顺序表示法。对于蛋白质样本来说，最典型的顺序表示法是其完整的氨基酸序列，这可以包含蛋白质最完整的信息。为了处理这种模型，通常使用基于序列相似性搜索的工具，如 BLAST [49]，来找到所需的结果。不幸的是，当查询蛋白质与训练数据集中的蛋白质没有显著同源性时，这种方法就不起作用了。因此，提出了各种非顺序表示法或离散模型。最简单的离散模型是基于氨基酸组成（AAC）的（例如，见 [50]）。然而，如果使用 AAC 模型来表示蛋白质，其所有的序列顺序信息都会丢失。为避免完全丢失序列顺序信息，有人提出了伪氨基酸组成（Pse-AAC）来表示蛋白质样本[36]。PseAAC 可以用离散模型来表示蛋白质序列，同时又不会完全丢失其序列顺序信息。有关 PseAAC 的更多信息，请点击链接 http://en.wikipedia.org/wiki/Pseudo_amino_acid_composition 访问网页。自从 PseAAC 这一概念被提出以来，它已被广泛用于研究蛋白质及蛋白质相关系统的各种问题（例如，[37,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66]）。与此同时，还提出了许多不同形式的离散模型（例如，[20,30,32,51,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82]）。然而，无论这些模型有多么不同，它们都只是 PseAAC 的不同形式，这一点在最近的一篇综述中已有阐明[83]。在此，我们将提出一种不同的 PseAAC 来表示药物靶向蛋白质的生化和物理化学特征[84]。考虑了六种不同类型的特征：（1）疏水性，（2）极化性，（3）极性，（4）二级结构，（5）归一化范德华体积，以及（6）溶剂可及性。

蛋白质序列中的每个氨基酸残基都可以根据其特征用一组不同的状态来表示。例如，其疏水性特征可以用以下三种状态之一来标记：“极性”、“中性”或“疏水性”[85]；其溶剂可及性特征可以用以下两种状态之一来表示：“埋藏”或“暴露于溶剂”，这是由 PredAcc 预测得出的[35]；其二级结构特征可以用以下三种状态之一来表示：“螺旋”、“折叠片”或“卷曲”，这是由[86]中的方法预测得出的；等等。

因此，蛋白质序列可以根据其组成氨基酸残基的生化和物理化学性质转换为一系列代码。例如，如果用“P”、“N”和“H”分别代表疏水性的三种状态：“极性”、“中性”和“疏水性”，那么蛋白质序列“DMAEIMSDKP-QAGML”就可以根据疏水性特征的编码转换为“PHNPHHNPPNPNNHH”。这样得到的编码序列对于不同大小的蛋白质来说长度会有所不同，这会使预测引擎难以处理。

为了将特征编码序列转换为具有固定维度数量的向量，对序列的三个属性进行了处理。

使用了组成（C）、转换（T）和分布（D）这三种特征。C 表示序列中每个字母的全局组成；T 表示一个编码字母转换为另一个编码字母的频率；D 表示编码字母在序列中的分布模式，测量的是序列长度的百分比，在此百分比范围内包含了每个编码字母的氨基酸的 1%、25%、50%、75% 和 100%。以上述疏水性序列为例：其 C 特征为 P、H 和 N 各占 5/15 = 33.3%，而 T 特征为 H 到 P、N 到 H、N 到 P 的转换频率分别为 2/10 = 20%、3/10 = 30% 和 5/10 = 50%。特征 D 的测量稍微复杂一些。对于字母 H，序列中 H 的 1%、25%、50%、75% 和 100% 分别位于 2、5、6、14 和 15 位。因此其 D 特征为（2/15 = 13.3%, 5/15 = 33.3%，6/15 = 40%，14/15 = 93.3%，15/15 = 100%）。同样地，字母 P 和 N 的分布分别为（6.7%，26.7%，53.3%，60%，73.3%）和（20%，46.7%，66.7%，80%，86.7%）。因此，代码字母序列的三个特征为：C = (33.3%, 33.3%, 33.3%), T = (20%, 30%, 50%)，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%), 共有 21 个成分。同样，对于由其他四个生化特性编码的序列，每个也对应 21 个成分。但对于仅具有两种状态（“埋藏”或“暴露于溶剂”）的溶剂可及性编码的序列，编码序列仅对应 14 个成分。最后，通过将 AAC 的 20 个成分[87]添加到相关向量中，对于给定的蛋白质，所获得的成分总数为 5|21 + 20 + 14 ≈ 139；即，该蛋白质可以表示为一个 139 维的向量，由以下给出：

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

其中 p_i （i = 1, 2, ， 139）是蛋白质 P 的第 i 个成分。在这 139 个成分中，119 个是根据上述六种生化和物理化学特征的编码得出的，另外 20 个是蛋白质 P 的氨基酸组成成分。

最近邻算法

由于所有样本都由特征向量表示，现在我们可以使用机器学习方法构建预测器。最近邻（NN）算法在模式识别领域颇受欢迎，因为它性能良好且易于使用。根据 NN 规则[88]，查询样本应被分配到与其最近邻所代表的子集。在本研究中，如果距离最短的药物 - 靶点对是正样本，即它们可以相互作用，则测试样本被视为正药物 - 靶点对。否则，测试样本被视为负样本。

对于近邻算法而言，衡量“接近度”的定义有很多，例如欧几里得距离、汉明距离[89]以及马氏距离[50,90,91]。在本研究中，采用以下公式来衡量样本 V_x 和 V_y 之间的接近度。

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

where V_x :V_y 是这两个向量的点积，而 kV_xk 和 ⊠V_y 分别为它们的模。当 V_x:V_y 我们有 D(V_x,V_y)~0 时，表明这两个样本向量之间的“距离”为零，因此它们具有完美的或 100% 的相似性。

刀切交叉验证检验

构建药物 - 靶点相互作用预测器之后，我们需要对其性能进行评估。在统计预测中，通常采用以下三种交叉验证方法来检验预测器在实际应用中的有效性：独立数据集测试、子抽样（K 折交叉验证）测试和刀切法测试[92]。然而，正如[24]所阐明的，并且如[93]中的公式 50 所示，在这三种交叉验证方法中，刀切法测试被认为是最客观的，对于给定的基准数据集，它总能得出唯一的结果，因此越来越多的研究人员采用并广泛认可该方法来检验各种预测器的准确性（例如[51,53,54,55,56,57,59,62,63,64,94,95,96]）。因此，在本研究中，我们也采用了刀切法交叉验证来计算成功预测率。

最大相关性最小冗余（mRMR）

尽管我们基于上述原始特征集构建了药物靶点预测器，但通过采用更优的特征集有可能提升其性能。显然，并非特征集中每个特征都与药物靶点相互作用具有同等的相关性。而且，特征之间可能并非相互独立。“不良”特征会对预测器的准确性和效率产生负面影响，因此可以进行特征选择过程以构建更紧凑且有效的特征集。第一步是使用最大相关最小冗余（mRMR）[36] 进行特征评估。最大相关最小冗余（mRMR）[39] 最初是为微阵列数据分析而开发的。它根据每个特征与目标的相关性以及与其他特征的冗余性对其进行排序。特征被认为越好，其排名就越高。互信息（MI），用 I 表示，用于量化相关性和冗余性，其定义如下：

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

基于互信息（MI），我们可以将相关性（D）和冗余度（R）量化为：

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

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

其中 $f_{candidate}$ 为待计算的特征，c 为目标变量。通过将上述两个方程结合起来以最大化相关性并最小化冗余，构建了以下的 mRMR 函数：

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

其中 V_s 和 V_t 分别表示已选特征集和待选特征集，m 和 n 分别表示这两个特征集的大小。特征被选择得越早，其重要性就越高。最终，我们可以得到一个有序的特征列表，每个特征都有一个排名，以表明其在特征集中的重要性。在我们的研究中，mRMR 程序是从以下网址获取的：http://research.janelia.org/peng/proj/mRMR/index.htm 。

为了计算互信息（MI），使用了两个向量的联合概率密度和边缘概率密度。这里引入一个参数 t 来处理这些变量。假设均值为所有样本中某一特征的平均值，标准差为 std，那么每个样本的特征将根据以下边界被分为三组：均值 + (t:std)。在我们的研究中，t 被设定为 1。

增量特征选择

如上所述，每个特征的重要性是根据其在 mRMR 分析中的排名来评定的。接下来要确定应选择哪些特征作为我们药物靶点预测器的最优特征集。这里使用 IFS（增量特征选择）过程来解决此问题。mRMR 特征列表中的每个特征依次添加，如果总特征数为 N，则会得到 N 个不同的特征集，而第 i 个特征集为：

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

基于每个 N 个特征集，构建了一个神经网络算法预测器，并通过刀切交叉验证测试进行了测试。计算出所有 N 个总体准确率后，我们可以绘制出 IFS 曲线，以索引 i 作为 x 轴，相应的总体准确率作为 y 轴。因此，如果曲线在 x 轴值为 n*f*N 处达到峰值，则 S_{opt}~ff$_1$， f_2，：：：， f_ng 被视为最优特征集。

由于四种不同类别的药物 - 靶点配对需要四个独立的预测因子，因此 IFS 分析过程将进行四次，每次针对一个特定的预测因子。

前向特征选择

为了优化特征选择，基于 IFS 结果采用了 FFS（前向特征选择）过程。FFS 是一种基于 IFS 结果的特征选择方法，它在每次迭代中尝试候选特征集中的每个特征，并将能实现最高预测准确率的特征添加到已选特征集中。假设 IFS 曲线达到峰值时，其顶点为 x 轴坐标，则初始的 FFS 选择特征集构建为：

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

在特征选择过程中，FFS 待选特征集中更多的特征会逐个被添加到 FFS 已选特征集中。具有 M 个特征的 FFS 待选特征集涵盖了 mRMR 排名在 k + 1 到 k + 1 + M 之间的特征，其中 M 是用户定义的正整数且小于原始特征集的大小 N，k 为 N 的大小。在每一轮 FFS 中，FFS 待选特征集中的每个特征都会被取出并添加到 FFS 已选特征集中。基于每个新 FFS 已选特征集的预测器都会被测试，获得最高总体准确率的特征集将被用作新的 FFS 已选特征集。此过程会运行 M 次，直到 FFS 待选特征集为空集。可以绘制出类似于 IFS 曲线的 FFS 曲线，其中 x 轴为索引，y 轴为总体准确率。

在本研究中，针对四个基准数据集中的每一个，均基于相应的 IFS 结果运行了 FFS。所有这些过程中的 M 均设为 50，而每个 FFS 中的 k 则设为相应 IFS 曲线中第一个最大值点（即索引最小的最大值点）的索引。

结果与讨论

mRMR结果

为了提高药物靶点相互作用预测器的性能，进行了特征选择过程。特征选择的第一步是特征评估。在本研究中，使用 mRMR 对原始特征集中的每个特征进行评估。在线补充信息 ~~S5 中列出了两种输出：第一种是~~ MaxRel 列表，显示了特征与目标相关性的排名；第二种是 mRMR 列表，根据满足公式 3 的特征顺序显示 mRMR 排名。在本研究中，仅使用 mRMR 列表作为特征评估的结果。由于有四组样本，mRMR 运行了四次，每次针对一组样本。

内部框架结构（IFS）和外部框架结构（FFS）的结果

利用这四个 mRMR 列表，对四个样本组分别进行了 IFS 处理，生成了四条 IFS 曲线。根据这些结果，我们将 FFS 中的 k 值分别设为 16、15、14 和 19，分别对应酶类、离子通道、GPCR 和核受体的数据。这些数值分别是相应 IFS 曲线中第一个最大值点的索引。图 1 展示了这四条 IFS 曲线及其对应的 FFS 曲线。四个 FFS 曲线的峰值最终分别达到了 85.48%（32 个特征）、80.78%（37 个特征）、78.49%（30 个特征）和 85.66%（32 个特征），分别对应酶类组、离子通道组、GPCR 组和核受体组。

mRMR+FFS 方法为四个不同组别所选的特征彼此差异很大，这表明它们之间存在内在差异。尽管在原始特征集中，针对靶点的特征比针对药物的特征多，但最终选中的药物特征更多，这表明药物具有重要作用。所选的许多靶点特征与蛋白质二级结构有关，尤其是对于酶组（所选靶点特征的一半都与此相关）。所有类型的特征在至少一个组中均被选中，这表明所有生化和物理化学特征在药物-靶点相互作用过程中都具有不可替代的作用。

关于 FFS 对四个基准数据集输出的最优特征集的详细信息，请参阅在线支持信息 S6。

讨论

针对特异性和多效性，我们根据药物作用的靶点将药物 - 蛋白质相互作用分为四组：酶、离子通道、G 蛋白偶联受体（GPCR）和核受体。我们使用金标准数据中所有已知的药物和靶蛋白作为训练数据，来预测所有在 KEGG 基因中被注释为这四类成员的人类蛋白质与 KEGG 配体中所有化合物之间的潜在相互作用。

酶识别是蛋白质与其他蛋白质以及诸如代谢物和治疗药物等小分子相互作用中的首要事件。预测药物与酶的相互作用在完成基因组注释、寻找用于合成化学的酶以及预测药物的特异性、多效性和药理学方面具有直接应用价值。有研究表明，二级结构信息在决定药物与酶的相互作用活性方面起着主要作用。例如，细胞色素 P450（CYP）诱导介导的相互作用是临床实践和制药行业的主要关注点之一[97]。具有特定结构稳定状态的 CYP1A 酶的诱导可能会使某些外源性物质活化为其活性代谢物，从而导致毒性[98,99]。氨基酸组成和疏水性也对这一过程有相当大的影响。

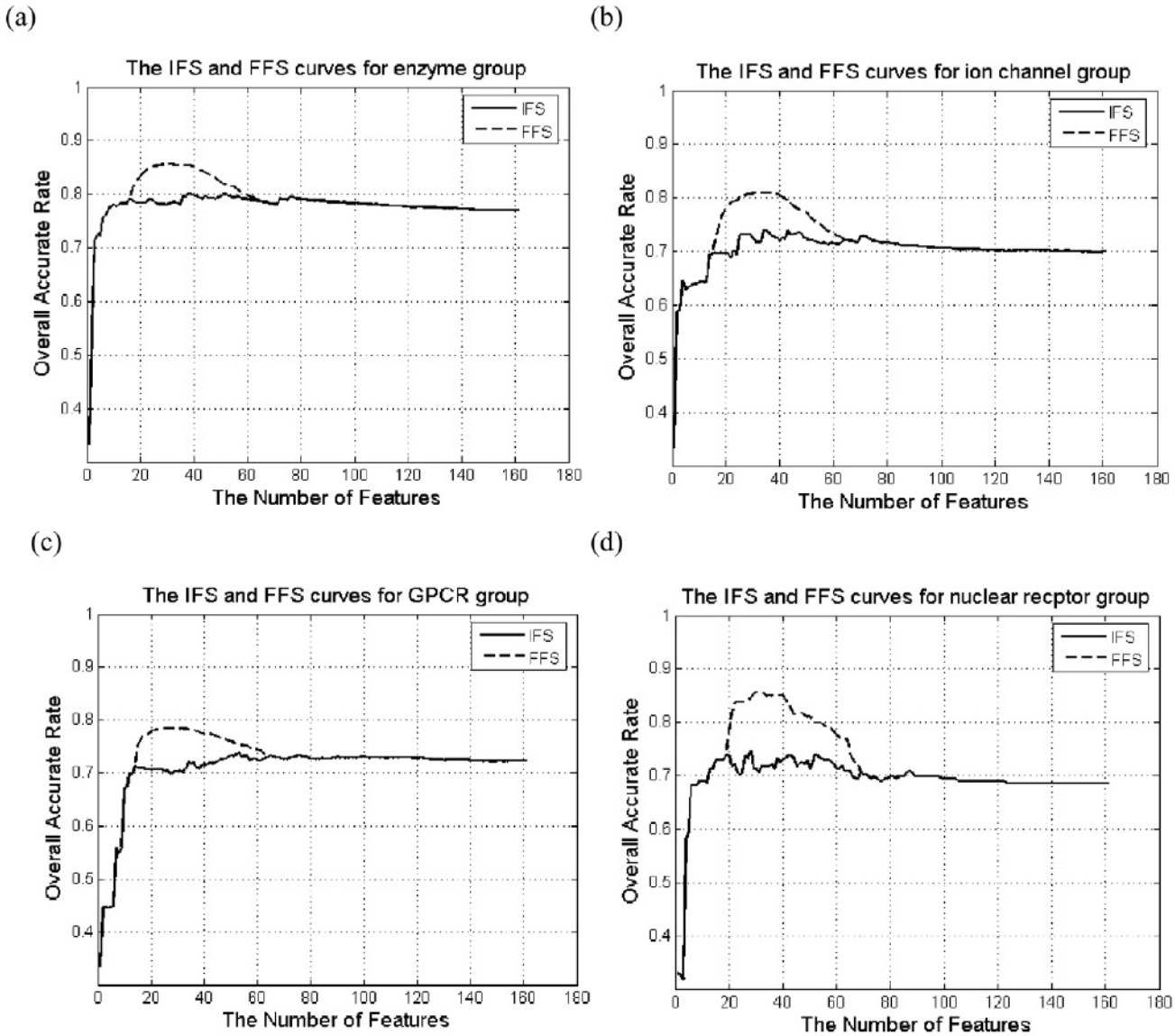

图 1.4 组的 IFS 和 FFS 曲线。分别为（a）酶组、（b）离子通道组、（c）GPCR 组和（d）核受体组的详细 IFS 曲线及其对应的 FFS 曲线。

图 1 doi:10.1371/journal.pone.0009603.g001

这些相互作用。血管紧张素 I 转换酶（ACE）的插入/缺失（I/D）多态性对降压反应有影响，尤其是在使用血管紧张素转换酶抑制剂（ACEI）时[100]，这表明氨基酸组成确实对相互作用有贡献。疏水性在确定药物与酶相互作用能的系数方面起作用，可用于药物筛选以及计算机辅助靶蛋白筛选[101,102]。

G 蛋白偶联受体（GPCR）超家族由约 600 至 1000 个成员组成，是已知最大的分子靶点类别之一，具有多种生理活性和已证实的治疗价值[103]。它们是跨膜整合蛋白，具有共同的总体拓扑结构，包括七个跨膜α螺旋、胞内 C 端、胞外 N 端、三个胞内环和三个胞外环[33,44]。研究表明，二级结构和极性在决定药物与 GPCR 之间的相互作用活性方面起着重要作用。诸如小的二级结构等特征

由于螺旋和环被认为可能是与配体稳定相互作用的实体[33]。这些基序主要位于跨膜段的顶端区域，并包含少量的胞外残基[104]。人β2 肾上腺素能受体（AR）与反激动剂配体卡替洛尔复合物的晶体结构提供了重要 G 蛋白偶联受体（GPCR）的三维快照，该受体具有β折叠结构，并构成发色团结合位点的一部分[105]。GLIDA 提供了 GPCR 与其配体之间的相互作用数据，以及配体的化学信息和关于 GPCR 的生物学信息[106]。其中一些特征反映了负责跨膜结构稳定性的物理相互作用，包括广泛的螺旋间氢键网络和硫-芳香族簇的空间组织，这些簇以“极性”形式存在，以及跨膜区域中侧链的紧密堆积。未来如果能获得更多的 GPCR 实验三维结构，这将有助于

构建适用于更广泛 G 蛋白偶联受体（GPCRs）的可靠模型，使其适合对接研究。基于配体的化学基因组学与对接的联合使用无疑会提高预测的准确性。

离子通道是一个庞大的膜蛋白超家族，能够使离子穿过膜，对可兴奋细胞和不可兴奋细胞的多种生理功能至关重要，并且是许多疾病的发病基础。因此，它们是一个重要的药物靶点类别，已被证明具有高度的“可成药性”。根据我们的分析，二级结构和极性在决定药物与离子通道的相互作用活性方面起着主要作用。二级结构控制膜电位，并在不同构象状态下对离子通道进行检测。药物与离子通道的相互作用需要门控状态，使它们能够在关闭和打开状态之间转换[42,43]。对模型纳米孔的模拟表明，狭窄的疏水区域可以在通道中形成功能性的关闭门，并且可以通过孔径的微小增加或极性的增加而打开[107,108]。如今，人们正在积极开展研究，以开发对离子通道亚型具有选择性作用的新药，并致力于深入理解药物与通道之间的相互作用[109]。

核受体（NR）是配体激活的转录因子，可调节多种重要靶基因的激活，是潜在治疗应用方面最重要的药物靶点。根据我们的研究结果，二级结构和极化率在决定药物与核受体的相互作用中起主要作用。核受体的保守基序通常被描述为三个堆叠的α-螺旋片。构成“前”和“后”片的螺旋彼此平行排列。中间片的螺旋横跨两个外片，仅占据该结构域的上部空间。该结构域的下部空间相对缺乏蛋白质，对于大多数核受体而言，这会形成一个供小分子配体进入的内部空腔[110]。具有极化率活性的氢键在蛋白质-药物相互作用中起着关键作用（例如，见[11]）。我们所采用的方法以及由此获得的结果可用于展示核激素受体如何形成直接相互作用的网络。而且，这一网络的复杂性不断增加，以描述与靶基因的相互作用以及已知能与受体、酶或转运蛋白结合的小分子的情况。

已建立了一个全面的药物 - 靶点相互作用网络系统，其中包含四个分类器，分别用于预测化合物与酶、离子通道、G 蛋白偶联受体（GPCR）和核受体的可成药相互作用。预计该网络预测系统将成为药物开发中非常有用的工具。特别是它可能帮助我们发现新的或潜在的药物 - 靶点相互作用。

补充信息

在线支持信息 S1 药物 - 靶标酶相互作用系统的基准数据集。它包含 8157 个基因 - 药物配对样本，其中 2719 个为阳性，5438 个为阴性。表格的第一列表示样本的性质，1 表示阳性，2 表示阴性；第二列显示靶基因的代码；第三列显示药物的代码。此处列出的所有基因和药物的详细信息均可在通过其代码获取 KEGG 数据（Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., Hirakawa, M. 从基因组学到化学基因组学：KEGG 的新发展，《核酸研究》，2006 年，34 卷：D354 - D357）。获取地址：doi:10.1371/journal.pone.0009603.s001（6.30MB DOC）

在线支持信息 S2 药物 - 离子通道靶点相互作用系统的基准数据集。它包含 4116 个基因 - 药物配对样本，其中 1372 个为阳性，2744 个为阴性。表格的第一列表示样本的性质，1 表示阳性，2 表示阴性；第二列显示靶基因的代码；第三列显示药物的代码。此处列出的所有基因和药物的详细信息均可通过其代码在 KEGG 中找到（有关进一步解释，请参阅在线支持信息 A 的说明）。获取网址：doi:10.1371/journal.pone.0009603.s002（3.35MB DOC）

在线支持信息 S3 药物 - GPCR 相互作用系统的基准数据集。它包含 1860 个基因 - 药物配对样本，其中 620 个为阳性，1240 个为阴性。表格的第一列表示样本的性质，1 表示阳性，2 表示阴性；第二列显示目标基因的代码；第三列显示药物的代码。此处列出的所有基因和药物的详细信息均可通过其代码在 KEGG 中找到（有关进一步解释，请参阅在线支持信息 A 的说明）。获取网址：doi:10.1371/journal.pone.0009603.s003（1.53MB DOC）

在线补充信息 S4 药物 - 核受体靶点相互作用系统的基准数据集。它包含 258 个基因 - 药物配对样本，其中 86 个为阳性，172 个为阴性。表格的第一列表示样本的性质，1 表示阳性，2 表示阴性；第二列显示靶基因的代码；第三列显示药物的代码。此处列出的所有基因和药物的详细信息均可通过其代码在 KEGG 中找到（有关进一步解释，请参阅在线补充信息 A 的说明）。

获取网址：doi:10.1371/journal.pone.0009603.s004 （0.22MB）（文档）

在线支持信息 S5 最大相关性最小冗余（mRMR）输出。

获取地址：doi:10.1371/journal.pone.0009603.s005 （1.02MB DOC）

在线支持信息 S6 前向特征选择（FFS）的结果。

获取地址：doi:10.1371/journal.pone.0009603.s006 （0.12MB DOC）

作者贡献

实验的构思与设计：ZH、JZ、LH、XK、YDC。实验操作：ZH、JZ、LH。数据分析：XHS。试剂/材料/分析工具的提供：JZ、YDC。论文撰写：ZH、XHS、KCC。

参考文献

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.