

CYP1A2 抑制剂预测模型的建立及评价

李曼华¹, 孙昊鹏^{1,2,3}, 尤启冬^{2,3*}

(中国药科大学¹药物化学教研室;²江苏省药物设计与优化重点实验室;
³天然药物活性组分与药效国家重点实验室, 南京 210009)

摘要 CYP1A2 酶在药物代谢中起着重要作用, 抑制 CYP1A2 会引起被 CYP1A2 代谢的其他药物代谢率降低, 从而导致这些药物的血浆浓度增加, 进而使药物的生物效应增强, 可能产生药物毒性。因此识别区分 CYP1A2 抑制剂成为新药早期评选及药物安全性评估的研究重点。本研究利用 674 个已知 CYP1A2 抑制活性的化合物构建 CYP1A2 抑制剂配体库, 从基于受体和基于配体的角度, 采用分子对接和药效团的方法, 利用 Pipeline Pilot 软件建立自动化筛选预测流程, 简单全面地从蛋白-配体结合角度快速准确预测出 CYP1A2 的抑制剂分子。最终从配体库中共预测出 16 个目标化合物, 其中 14 个化合物具有 CYP1A2 抑制活性。研究最后对美国成药数据库进行综合预测, 共发现 4 个药物是已报道的 CYP1A2 抑制剂。说明本模型对 CYP1A2 抑制剂具有很好的预测能力, 可以应用于 CYP1A2 抑制剂的预测。

关键词 CYP1A2 抑制剂; 预测模型; 对接; 药效团模型; 自动化流程

中图分类号 R914.2 文献标志码 A 文章编号 1000-5048(2013)05-0401-09
doi:10.11665/j.issn.1000-5048.20130504

Establishment and evaluation of prediction models for the discovery of CYP1A2 inhibitors

LI Manhua³, SUN Haopeng^{1,2,3}, YOU Qidong^{1,2*}

¹Department of Medicinal Chemistry; ²Jiangsu Key Laboratory of Drug Design & Optimization; ³State Key Laboratory of Natural Medicines, China Pharmaceutical University, Nanjing 210009, China

Abstract CYP1A2 enzyme plays a crucial role in drug metabolism, and its inhibition may cause low metabolic rates and increased plasma concentrations of the drugs metabolized by CYP1A2, thus leading to drug toxicity. Therefore, distinction between CYP1A2 inhibitors and non-inhibitors becomes important topic on the early selection of new drug candidates and drug safety assessment. In this study, a CYP1A2 inhibitor-ligand library was built with 674 compounds known to possess CYP1A2 inhibitory activity. From the point of receptor-based and ligand-based view, we have built an automatic screening protocol with Pipeline Pilot, using molecular docking and pharmacophore modeling methods, so as to predict inhibitors from CYP1A2 inhibitor library quickly and accurately. The final model predicted 16 target compounds from the library, and 14 of which were CYP1A2 inhibitors. At last, we used the final model to screen the American Medicine Database, and four drugs were found to possess CYP1A2 inhibitory activity. The combination of prediction models can improve the efficiency of CYP1A2 inhibitor discovery significantly.

Key words CYP1A2 inhibitors; prediction models; docking; pharmacophore modeling; automatic pipeline

细胞色素 P450 (cytochrome P450, CYP450) 是一个由血红蛋白-巯基盐 (heme-thiolate) 所组成的超家族, 在内源物及外源物代谢方面起着十分重要的作用^[1]。在人体中, CYP450 参与约 70% ~ 80%

已上市药物的 I 期代谢。其中, 涉及大多数药物代谢的 P450 酶系主要是由如下几种酶参与催化: CYP1A2 约 5%, CYP2C9 和 CYP2C19 约 25%, CYP2D6 约 15%, CYP3A4 约 50%。有研究发

现^[2],在药物的吸收、分布、代谢与排泄(absorption, distribution, metabolism, excretion, ADME)性质中,由于药物对 CYP450 的诱导或抑制而导致的药物-药物相互作用(drug-drug interactions, DDIs)是多种药物被撤出市场的主要原因。而其中酶抑制所致 DDIs 的临床意义又远大于酶诱导作用,约占代谢性相互作用的 70%,酶诱导引起的 DDIs 约占 23%。目前,抑制 CYP450 的研究已列为新药早期评选及临床前药代动力学研究中必须开展的一项内容^[3-7]。

CYP1A2 约占整个细胞色素 P450 酶系的 13%~15%,参与约 5% 上市药物的代谢,主要分布于肝脏中,另外在肠道、脑、肺等组织中也有少量 CYP1A2 分布,主要参与多环芳烃和芳香胺类化合物的代谢。参与代谢的药物有非那西丁、咖啡因、茶碱、华法林、美西律、维拉帕米、硝苯地平、对乙酸氨基酚等,同时 CYP1A2 也参与代谢褪黑激素、胆红素、雌二醇等内源性激素的代谢。此外,CYP1A2 在一些前致癌物和前毒性物质的体内活化过程中也起到重要作用^[8-9]。鉴于 CYP1A2 在药物代谢中的重要性,提早将 CYP1A2 的抑制剂从开发行列中寻找出来,对提高新药开发成功率,设计安全有效的药物,避免产生严重的 DDIs 尤为重要。

目前对 CYP1A2 抑制剂的计算机辅助(in silico)研究多基于单一角度,或从 CYP1A2 抑制剂配体分子出发研究其特点,或从基于受体角度发现 CYP1A2 蛋白空间特征。Korhonen 等^[10]利用 52 个化合物建立了基于 CoMFA 的 3D-QSAR 模型,该研究发现分子形状、分子的芳香性以及氢键结合能力会影响化合物对 CYP1A2 的抑制作用。分子越小、芳香性越高、与氢键结合能力越强,化合物成为 CYP1A2 抑制剂的可能性越高。但 CoMFA 模型要求训练集分子具有相似的母核,不能对结构多样的化合物作出很好预测。Vasanthanathan 等^[11]通过支持向量机、最近邻域分析、随机森林和决策树等 4 种 2D-QSAR 模型对 CYP1A2 抑制剂与非抑制剂加以区分,结果表明模型的准确区分度达 73%~76%。该研究并不能清楚说明 CYP1A2 抑制剂具有何种特征。De Rienzo 等^[12]建立了 CYP1A2 同源模型,发现 CYP1A2 的结合口袋较小,且成平面状,这些发现都与之前基于配体发现的 CYP1A2 底物特征相符。但该研究并未说明 CYP1A2 抑制剂

对蛋白有何种影响。

本研究从基于受体和基于配体两种角度出发,首先采用 674 个已知 CYP1A2 抑制活性的化合物构建 CYP1A2 抑制剂配体库。然后采用分子对接和药效团的方法,确定最优对接模型和最优药效团模型。最后利用 Pipeline Pilot 软件建立 CYP1A2 抑制剂筛选流程,使预测流程更为简单快捷,易于操作。研究最后利用建立的 CYP1A2 抑制剂预测模型,综合对美国成药数据库进行预测,共发现 4 个药物为 CYP1A2 已知抑制剂。说明本模型能够全面地从蛋白-配体结合角度快速准确筛选预测出 CYP1A2 的抑制剂分子,可以应用于 CYP1A2 抑制剂的预测和区分。

1 实验数据和方法

本文所有操作均通过 Accelry 公司提供的 Discovery Studio 2.5 软件和 Pipeline Pilot 软件完成。除特殊说明外,所有参数均为系统默认值。

1.1 CYP1A2 抑制剂配体库的构建

数据来源于 Binding Database 数据库,以 cytochrome P450 1A2 为关键字,共收集 674 个有 CYP1A2 抑制活性的化合物,其 IC_{50} 最小为 3 nmol/L,最大为 570 μ mol/L,具有明显差异。应用“Prepare Ligands”和“Minimize Ligands”模块,对化合物经过删除重复结构、生成最优互变异构体、加氢、能量最小化等操作。将 IC_{50} 小于等于 5 μ mol/L 的化合物定义为活性化合物^[13],将 IC_{50} 大于 5 μ mol/L 的化合物定义为非活性化合物,共得到 167 个活性化合物和 507 个非活性化合物,以用于接下来模型中进行模型构建和评价。

1.2 对接模型构建

1.2.1 蛋白前处理 CYP1A2 蛋白晶体(分辨率 1.95 Å, PDB 编号:2HI4)结构数据来源于 PDB 蛋白质晶体结构数据库,该蛋白晶体结构为 α -萘黄酮(α -naphthoflavone, ANF)。应用“Prepare Protein”模块,对 CYP1A2 晶体结构进行预处理。操作包括插入缺失原子、修正蛋白晶体结构的化学键、补充缺失 loop 区域、优化蛋白结构、添加氨基酸残基的质子化状态等,并为 CYP1A2 蛋白加氢,赋予 CHARMM 力场。

1.2.2 对接参数设置 选中未处理前蛋白配体,利用“Define Sphere from Selection”工具,以晶体配

体分子为中心,其所在 9 Å 范围内的所有结构作为对接口袋。本研究对接工具采用“Dock Ligands (Gold)”模块,其中“Input Control Ligands”设置为蛋白晶体配体,“Maximum Poses Retained”设置为 1,其余参数均采用默认值。打分函数分别采用 ChemScore 与 GOLDScore 以建立不同对接模型。

1.2.3 对接模型验证 本研究应用一系列统计学方法对模型做出评价,并以此选出最优对接模型。具体的评价方法^[14-15]见表 1。

Table 1 Statistical evaluation methods for the CYP1A2 inhibitors' model

Methods	Formula
Prevalence	$(TP + FN)/n$
Correct classification rate	$(TP + TN)/n$
Sensitivity	$TP/(TP + FN)$
Specificity	$TN/(FP + TN)$
False positive rate	$FP/(FP + TN)$
False negative rate	$FN/(TP + FN)$
Positive predictive power	$TP/(TP + FP)$
Negative predictive power	$TN/(FN + TN)$
κ	$\frac{\{(TP + TN) - [((TP + FN)(TP + FP) + (FP + TN)(FN + TN))/n]\}}{n - [((TP + FN)(TP + FP) + (FP + TN)(FN + TN))/n]}$

TP: True positive; FP: False positive; FN: False negative; TN: True negative; n : the number of all compounds

流行率(prevalence)评价了某模型中活性化合物占全部化合物的比例。例如在某模型中测试集的流行率为 90%,那么很大程度上模型的正确分类率就可能高达 82%。故流行率不宜太高;正确分类率(correct classification rate)是模型区分活性

化合物与非活性化合物的能力;灵敏度(sensitivity)评价了正确预测活性化合物占全部活性化合物的比例,特异性(specificity)则评价了正确预测为非活性化合物占全部非活性化合物的比例;假阳性率(false positive rate)是错误预测为活性化合物占全部非活性化合物的比例,假阴性率(false negative rate)是错误预测为非活性化合物占全部活性化合物的比例;正向/逆向预测力(positive/negative predictive power)是正确预测为活性/非活性化合物占全部预测为活性/非活性化合物的比例。 κ 指数^[14]是一种测量置信度的方法,表现了重复测量的一致性,见公式(1):

$$\kappa = \frac{P_o - P_c}{1 - P_c} \quad (1)$$

式中, P_o :观测一致性(observed agreement),前后(两种)测量结果一致的百分比; P_c :期望一致性(chance agreement),前后(两种)测量结果预期相同的机率; κ 计算的结果为 $-1 \sim 1$,但通常 κ 是落在 $0 \sim 1$ 之间,可分为五组来表示不同等级的吻合度:0.0 ~ 0.20 极低的吻合度(slight);0.21 ~ 0.40 一般的吻合度(fair);0.41 ~ 0.60 中等的吻合度(moderate);0.61 ~ 0.80 高度的吻合度(substantial)和 0.81 ~ 1 几乎完全吻合(almost perfect)。

1.3 药效团模型构建

1.3.1 训练集分子 本小节从文献中收集了 5 个结构多样的常见 CYP1A2 抑制剂分子^[16],用于定性药效团的建立。具体结构见图 1。

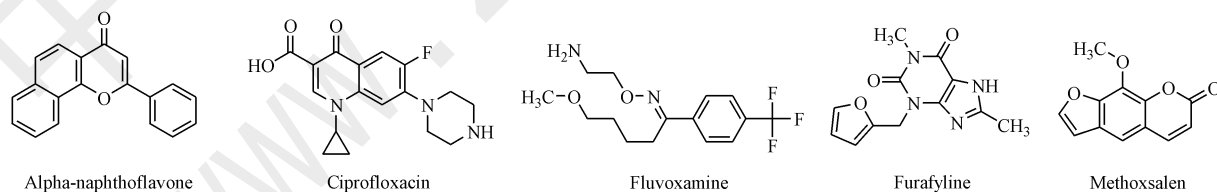


Figure 1 Structures of CYP1A2 inhibitors

1.3.2 药效团模型参数设置 应用“Common Feature Pharmacophore Generation”模块,寻找训练集分子的公共药效团特征。其中药效团特征(Features)选用 5 种,分别为氢键受体(HB_ACCEPTOR, HBA)、氢键供体(HB_DONOR, HBD)、疏水中心(HYDROPHOBIC, HYD)、正电荷中心(POS_IONIZABLE, POS)和芳环中心(RING_AROMATIC,

RA);“Minimum Features”设置为 4,以确保生成的药效团至少有 4 个特征;构象生成方法“Conformation Generation”选择“FAST”生成方法。

对训练集分子添加 Principal 和 MaxOmitFeat 属性值。Principal 表明该化合物是否具有活性,共有 3 个值可供选择。“2”表示化合物有活性,“1”表示化合物有中度活性,“0”表示化合物没有活

性。MaxOmitFeat 表明该化合物允许多少个特征被忽略。“2”表示都可以忽略,即不需要该化合物与产生的药效团完全匹配。“1”表示至少有1个特征需要与产生的药效团匹配,“0”表示所有特征都必须与产生的药效团匹配。本研究中构建 CYP1A2 抑制剂药效团的训练集分子 Principal 值和 MaxOmitFeat 值全部设置为2。

药效团形状(shape)生成时所需分子结构选用最为常见的 CYP1A2 抑制剂 ANF。

1.3.3 药效团模型验证 本研究中药效团模型的主要评价参数为 Rank 值与 Max Fit 值。Rank 值是用于评价分子与药效团匹配程度。Rank 值越高,药效团与活性分子的匹配度越高,与非活性分子匹配度越低。Max Fit 值则说明共有几个训练集分子可与药效团很好的匹配。

另外还采用 CYP1A2 抑制剂配体库来验证药效团模型的质量,在“Ligand Pharmacophore Mapping”模块中进行化合物的活性预测。其中“Best Mapping Only”设为“True”,分子必须匹配药效团所有的特征(“Maximum Omitted Features”=0)。匹配方式(Fitting Method)设置为柔性匹配(Flexible)。

1.4 预测模型构建

在前文的研究基础上,利用 Pipeline Pilot 软件将化合物前处理,对接预测,药效团预测相结合,建立统一预测流程。首先读入待预测分子配体(SD Reader),对其做配体准备(Ligand Preparation)和能量最小化(Minimize Molecule)处理。然后进行对接操作(GOLD Docking),并利用“Property Value Threshold Filter”模块将分子按一定的打分值筛选出来,进行下一步药效团预测(Ligand Pharmacophore Map)。最终预测结果将以 sd 文件形式输出(SD Writer),且可以以网页格式可视化查看预测结果(HTML Molecular Table Viewer)。所有模块的参数设置均与前文参数设置相同。

2 结果与讨论

2.1 对接模型

2.1.1 打分函数对模型的影响及预测标准的确定

打分函数是对配体化合物与蛋白受体结合亲和力的一种数据化体现,配体与受体蛋白结合亲和力越强,则表明两者结合越紧密,配体对蛋白的影响就愈大。本研究尝试通过打分函数对 CYP1A2 抑

制剂与非抑制剂作出区分。分别计算了 Gold 对接的两种打分函数 ChemScore 与 GOLDScore 对抑制剂和非抑制剂分子的影响,希望寻找出更适合区分 CYP1A2 抑制剂与非抑制剂的打分函数及模型预测的标准。

以打分函数为横坐标,化合物在不同打分值时占全部化合物的百分比为纵坐标,见图2和图3。

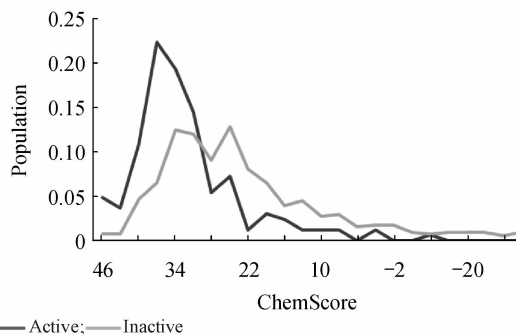


Figure 2 Distribution of active and inactive compounds by ChemScore

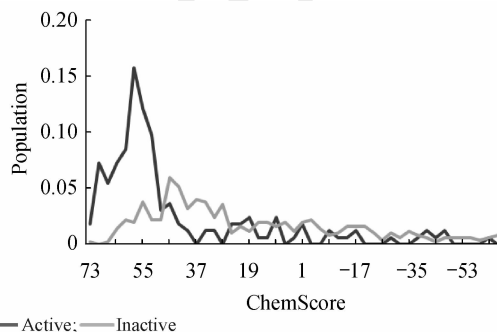


Figure 3 Distribution of active and inactive compounds by GOLDScore

图中,黑色曲线表示抑制剂在不同打分值时占全部抑制剂分子的百分比分布,灰色曲线表示非抑制剂在不同打分值时占全部非抑制剂分子的百分比分布。由图可知,打分值较高区域(在 ChemScore 模型中为打分值大于等于29的区域,在 GOLDScore 模型中为打分值大于等于47的区域),在此区域内抑制剂的分布要明显高于非抑制剂的分布,且两者并无交叉重叠,说明利用打分函数区分抑制剂与非抑制剂是可行的。故以抑制剂与非抑制剂曲线交点对应的打分值作为预测筛选标准,其中 ChemScore 模型以打分值29为预测标准,将得分达到29及以上的化合物视为命中化合物; GOLDScore 模型以打分值47为预测标准,将得分达到47及以上的化合物视为命中化合物。仔细观察两图还可发现,相较于 GOLDScore 模型,ChemScore 模型中抑制剂与非抑制剂曲线除区分点外无

其他交叉部分,说明 ChemScore 打分值对抑制剂的区分度要优于 GOLDScore 打分值;但 GOLDScore 模型中非抑制剂的曲线要低于 ChemScore 模型中非抑制剂的曲线,说明 GOLDScore 模型中假阳性低,抑制剂与非抑制剂的区分效率高于 ChemScore 模型。

综上所述,当数据量较大,需要很好区分抑制剂与非抑制剂时,可以使用 ChemScore 打分函数;当数据需要精确计算,且假阳性率偏低的情况下, GOLDScore 打分函数较优。

2.1.2 对接模型统计学验证和最优对接模型选择

ChemScore 模型中预测获得的命中化合物共 313 个, GOLDScore 模型预测获得的命中化合物共 188 个。以化合物 IC_{50} 作为活性参考标准,利用前文提及的统计学方法,分别对 ChemScore 与 GOLDScore 打分函数建立的模型计算统计学参数,具体见表 2。

Table 2 Summary of classification into active and inactive compounds

Method	TP	FN	FP	TN
ChemScore	125	42	188	319
GOLDScore	116	51	72	435

Table 3 Results of docking modeling using statistical methods

Item	ChemScore	GOLDScore
Prevalence	0.25	0.25
Correct classification rate	0.66	0.82
Sensitivity	0.75	0.69
Specificity	0.63	0.86
False positive rate	0.37	0.14
False negative rate	0.25	0.31
Positive predictive power	0.40	0.62
Negative predictive power	0.88	0.90
κ	0.29	0.53

These values are given in bold for better scores

由表 3 可知,两个模型的流行率均较低,排除了模型的偶然性,表明模型的可信度均良好。除灵敏度 and 假阴性率这两项参数, GOLDScore 模型均优于 ChemScore 模型。 GOLDScore 模型的正确分类率可达到 82%,已具备很好的区分抑制剂与非抑制剂的能力,其假阳性率也很低(14%),保证了数据的低失误差率。 GOLDScore 模型对活性化合物的预测能力为 62%,对非活性化合物的预测能力则

高达 90%。相较而言, ChemScore 模型对活性化合物和非活性化合物的预测能力皆不如 GOLDScore 模型。观察未成功被预测为活性的化合物发现,绝大部分化合物的相对分子质量都超过 400,较不符合 CYP1A2 结合口袋特征。值得一提的是 GOLDScore 的 κ 指数高达 0.53,与文献[11]中 2D-QSAR 模型的预测能力相当,说明 GOLDScore 模型具有很好的可重复性,模型可靠性较高。从各项统计学评价参数来看, GOLDScore 模型表现更优,故本研究选取 GOLDScore 模型参与 CYP1A2 抑制剂的预测。

2.1.3 对接构象分析 GOLD 可以选择原蛋白晶体结构中的配体,作为对接参考分子,将其重新与蛋白对接,保证对接结果的可靠性。计算原蛋白结构中的配体分子 ANF 与重新对接 ANF 构象的 RMSD,预测最优结合构象非常接近于晶体结构中的真实构象(RMSD = 0.59 Å)。这说明 GOLD 能良好的重现已知配体在晶体结构中的结合模式,当其应用在未知的化合物的结合模式预测上时,具有更加令人信服的准确性。

CYP1A2 蛋白的活性位点是由螺旋 F 和螺旋 I 上的氨基酸残基,联合血红蛋白(heme)共同构成。该活性口袋是一个窄小、平面、疏水、几乎闭合的空腔,无明显的溶剂或底物通道,其体积较小,约为 375 Å³。该蛋白晶体结构还包含两个水分子。 CYP1A2 关键作用的氨基酸残基主要包括: Thr124、Phe125、Phe226、Gly316、Ala317 等^[17]。从关键作用的氨基酸残基也可看出 CYP1A2 的活性作用口袋疏水性较大,这与 CYP1A2 多参与代谢芳烃类药物相吻合。

选择打分值最高的化合物 CHEBI_236408 分析其与 CYP1A2 的对接构象。由图 4 可见, CHEBI_236408 的芳环与 Phe226 产生 π 共轭相互作用,其中末端苯环与血红蛋白的 Fe 离子也发生 π - π 共轭作用;羰基氧与水分子发生氢键结合;甲基与 Phe125 产生碰撞(bump)作用。基本符合 CYP1A2 活性口袋特征,说明本模型可以满足化合物与 CYP1A2 蛋白对接的需求。

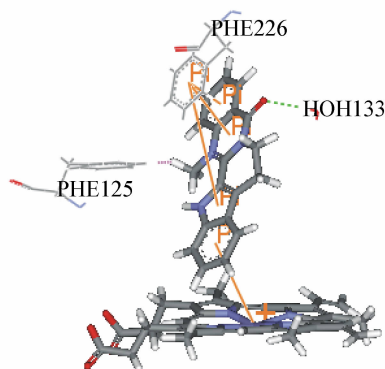


Figure 4 Binding model of CHEBI_236408 with CYP1A2

2.2 药效团模型

2.2.1 药效团模型 由于 CYP1A2 的抑制剂分子结构差异较大,且对抑制活性测量的标准不统一,采用定量药效团方法不能很好地模拟出 CYP1A2 药效团的特征,故本文中的药效团构建利用 5 个常见 CYP1A2 抑制剂作为训练集,利用定性药效团方

法寻找 CYP1A2 的典型药效团特征。当药效团特征至少有四个特征时,模型筛选能力将更具有特异性,避免预测的假阳性率过大,故建立药效团时,将“Minimum Features”设置为 4,保证 CYP1A2 的药效团至少含有 4 个特征。因此本次模型构建结果只得到 2 个药效团模型,将生成的两个药效团模型命名为 HHAA1 和 HHAA2。两个药效团模型参数一致,特征皆为 HHAA,两个氢键受体、两个疏水中心。两者的不同在于其中一个氢键受体的矢量方向不同,具体见图 5-A 和图 5-B。绿色表示氢键受体(HBA),蓝色表示疏水中心(HYD)。两个氢键受体间的距离为 3.9 Å,两个疏水中心间的距离为 5.8 Å,氢键受体与疏水中心间的距离从 4.0 ~ 7.0 Å 不等。4 个特征几乎保持在同一平面内,符合 CYP1A2 蛋白空腔特征和底物结构特征^[16]。从药效团特征的距离来看,发现 CYP1A2 抑制剂多为小分子化合物。

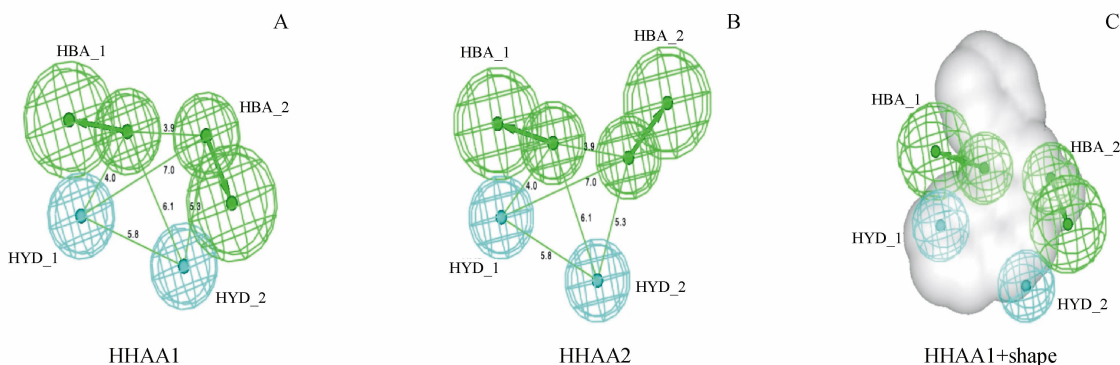


Figure 5 Pharmacophore of CYP1A2 inhibitors

2.2.2 药效团模型评价和最优药效团模型选择

HHAA1 和 HHAA2 的 Rank 值皆为 27.366, Max Fit 为 4,表示所有训练集分子都可以与药效团特征全部匹配。图 6 是呋啦茶碱(furalylline)与药效团匹配的示意图,图中显示呋啦茶碱结构中嘧啶环上的两个羰基氧分别与两个氢键受体特征匹配,咪唑环上的甲基与其中一个疏水中心匹配,呋喃环则与另一个疏水中心匹配,说明所构建的药效团模型可以很好的满足呋啦茶碱所有特征。利用“Ligand Pharmacophore Mapping”模块,对 CYP1A2 抑制剂配体库进行预测。两个模型均预测出 424 个化合物,其中 HHAA1 模型预测出 89 个有抑制活性分

子,HHAA2 模型预测出 90 个有抑制活性分子。可见通过简单的药效团特征对化合物库进行抑制剂与非抑制剂的预测区分,会带来很大的假阳性率,特异性不高。故对 HHAA1 和 HHAA2 模型添加形状约束(见图 5-C),增加药效团的特异性。增加形状约束后,应用相同参数的“Ligand Pharmacophore Mapping”模块,HHAA1 模型共预测出 22 个分子,其中 16 个有抑制活性;HHAA2 模型共预测出 25 个分子,其中 16 个有抑制活性。可以看出添加有形状约束的 HHAA1 模型更能准确对有 CYP1A2 抑制活性的化合物做出预测,故选择有形状约束的 HHAA1 模型参与后续 CYP1A2 抑制剂的预测。

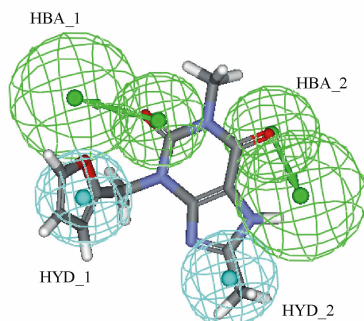


Figure 6 Mapping of CYP1A2 inhibitor pharmacophore model (HHAA1) and furalylline

2.3 预测模型

计算机辅助的预测方法作为一种近年发展起来的药物筛选预测方法,在药物研究中发挥的作用越来越大。其与传统药物筛选预测方法相比,所需时间大为缩短,劳动强度和成本也大大降低。本研究利用 Pipeline Pilot 软件建立一个全自动的流程化协议,只需输入待预测的化合物库,就可直接预测出 CYP1A2 的抑制剂分子。非专业用户也能方便地使用该模型。且该 CYP1A2 抑制剂预测流程的建立是从基于受体和基于配体两个角度出发,全面对 CYP1A2 抑制剂做出预测,避免基于配体的药物设计方法中只能关注配体分子局部,而不能兼顾分子整体性质的缺陷。

本研究在开始 SD Reader 模块中输入 CYP1A2 抑制剂配体库共 674 个化合物,共获得命中化合物共 16 个,其中 14 个化合物具有 CYP1A2 抑制活性 ($IC_{50} \leq 5 \mu\text{mol/L}$),所有数据都经文献验证准确无误^[18-23]。该 14 个化合物分别属于甲氧基黄酮类(2 个)、吴茱萸次碱及其衍生物(7 个)、吡唑和异噻唑衍生物(2 个)、喹啉类衍生物(1 个)及胺类和杂环化合物(2 个)。其中甲氧基黄酮类和吴茱萸次碱类化合物都是常见的 CYP1A2 抑制剂。说明此次基于对接模型(GOLDScore 模型)和药效团模型(“HHAA1 + shape”模型)的流程化筛选预测,可以精确的预测出 CYP1A2 抑制剂分子,可用来进行抑制剂的预测,发现具有潜在 CYP1A2 抑制活性的化合物,避免因抑制 CYP1A2 而产生的不良 DDI_s。该操作过程方便快捷,适合专业及非专业人员

应用。

本文将对接模型和药效团模型结合起来建立综合预测模型,相较于单一的对接模型或药效团模型,综合模型对抑制剂分子的预测率得到了极大地提升,同时也降低了假阳性率。利用表 1 对对接模型(GOLDScore),药效团模型(HHAA1 + shape)以及综合模型的预测能力做出评价,具体见表 4。

Table 4 Results of consensus modeling using statistical methods

Methods	GOLDScore	HHAA1 + shape	Consensus model
Prevalence	0.25	0.25	0.25
Correct classification rate	0.82	0.77	0.75
Sensitivity	0.69	0.10	0.01
Specificity	0.86	0.99	0.998
False positive rate	0.14	0.01	0.001
False negative rate	0.31	0.90	0.99
Positive predictive power	0.62	0.73	0.67
Negative predictive power	0.90	0.77	0.75
κ	0.53	0.12	0.01

由表 4 可见,综合模型极大地提升了对于 CYP1A2 抑制剂分子的预测能力,特异性和假阳性率都接近最大值和最小值,说明综合模型可以准确的从大量分子中预测出 CYP1A2 的抑制剂分子。从特异性与敏感性,假阳性率与假阴性率,正向预测能力与逆向预测能力的公式可以发现,这 3 种评价参数每每相互对立的,提高了某一项的同时,另一项的数值必然有所降低。本文中研究主要目的是从大量无生物活性数据的分子中,准确发现预测出 CYP1A2 的抑制剂,避免这些抑制剂分子对后续的药物研发或临床药物使用产生干扰或不利影响。因此基于该种目的,本文的综合模型具有很好的预测能力。

2.4 应用

利用上文建立的 CYP1A2 抑制剂筛选预测流程,综合对美国成药数据库(US Drug Collection Database, 2009)的 1 280 个药物进行预测,共得到 55 个化合物。通过查询文献及相关资料^[24-26],发现 17 个是已报道经 CYP450 酶代谢的化合物,其中 4 个药物为 CYP1A2 常见抑制剂,这些药物的商品名,通用名, CAS 号以及治疗范围见表 5。

Table 4 Results of virtual screening by the prediction model established

Trade name	Generic name	CAS#	Indication
Avandia	Rosiglitazone maleate	122320-73-4	Antidiabetic
Prilosec	Omeprazole	73590-58-6	Gastric acid depressant
Lopid	Gemfibrozil	25812-30-0	Antihyperlipoproteinemic
Inderal	Propranolol hydrochloride	318-98-9, 525-66-6 [propranolol]	Antihypertensive, antianginal, antiarrhythmic

3 结论与展望

计算机辅助分子设计的建立,为药物研究开拓了新的领域和方法,从分子和蛋白水平研究药物-受体相互作用。在目前已知的有关化学结构信息基础上,利用计算机虚拟药物筛选技术预测化合物,可以快速高效地提高药物设计的命中率,同时减少大量人力、物力的投入。CYP1A2 酶在药物代谢中起着重要的作用,抑制 CYP1A2 可能会引起代谢紊乱,或被 CYP1A2 代谢的其他药物代谢率降低,从而导致这些药物的血浆浓度增加,进而使药物的生物效应增强,可能会使其毒性增加。因此识别区分 CYP1A2 的抑制剂成为新药早期筛选及药物安全性评估的研究重点。本研究利用 674 个已知 CYP1A2 抑制活性的化合物构建 CYP1A2 抑制剂配体库,从基于受体和基于配体的角度,采用分子对接和药效团的方法,利用 Pipeline Pilot 软件建立自动化筛选预测流程,全面地从蛋白-配体结合角度快速准确预测出 CYP1A2 的抑制剂分子。

在分子对接模型中,本文首先应用抑制剂与非抑制剂在不同打分值时的占全部抑制剂或非抑制剂的比例作图,确定模型预测标准,其中 ChemScore 模型以打分值 29 为预测标准,将得分达到 29 及以上的化合物视为命中化合物;GOLDScore 模型以打分值 47 为预测标准,将得分达到 47 及以上的化合物视为命中化合物。并应用一系列统计学评价手段对接模型做出评价,从不同参数可以发现 GOLDScore 模型要优于 ChemScore 模型,且 GOLDScore 模型的 κ 为 0.53,说明该模型具有很好的可重现性。

在药效团模型中,利用 5 个常见的 CYP1A2 抑制剂构建了 CYP1A2 抑制剂的药效团模型。该药效团特征为两个氢键受体和两个疏水中心,且特征几乎位于同一平面,特征间距离也较为接近,与文献中报道的 CYP1A2 抑制剂特征相符。研究还对 CYP1A2 抑制剂的药效团加以形状约束,并利用 CYP1A2 抑制剂配体库对药效团模型加以评价,发

现“HHAA1 + shape”模型可以很好的对 CYP1A2 抑制剂作出预测。

最后利用 Pipeline Pilot 软件建立一个全自动的流程化协议,只需输入待预测的化合物库,就可直接筛选预测出 CYP1A2 的抑制剂分子。对 CYP1A2 抑制剂分子库进行预测,共预测出 16 个化合物,其中 14 个化合物具有 CYP1A2 抑制活性,验证了模型的预测能力。应用该 CYP1A2 抑制剂筛选预测流程对美国成药数据库进行预测应用,得到可能具有 CYP1A2 抑制活性的 55 个化合物。通过文献及相关资料调研发现,4 个药物已报道为 CYP1A2 的抑制剂。说明本模型对 CYP1A2 抑制剂具有很好的预测能力,可以应用于 CYP1A2 抑制剂的预测和区分,且方便快捷,适合专业及非专业人员应用。

本研究通过建立了基于 CYP1A2 抑制剂的自动化筛选预测流程,从基于受体和基于配体的角度,全面准确而又快速便捷地发现 CYP1A2 的抑制剂分子。在此,期望能够借助该预测模型的应用,针对不同用户的需求,可方便快捷地对 CYP1A2 的抑制剂作出筛选预测,以避免由此所产生的代谢不良反应或因严重的 DDIs 引起的药物毒性,同时也可以为新化学实体的设计提供有益帮助。

参考文献

- [1] Nettleton DO, Einolf HJ. Assessment of cytochrome p450 enzyme inhibition and inactivation in drug discovery and development [J]. *Curr Top Med Chem*, 2011, **11**(4): 382 - 403.
- [2] Foti RS, Wienkers LC, Wahlstrom JL. Application of cytochrome P450 drug interaction screening in drug discovery [J]. *Comb Chem High Throughput Screen*, 2010, **13**(2): 145 - 158.
- [3] DeLisle RK, Otten J, Rhodes S. *In silico* modeling of p450 substrates, inhibitors, activators, and inducers [J]. *Comb Chem High Throughput Screen*, 2011, **14**(5): 396 - 416.
- [4] Gleeson MP. Generation of *in-silico* cytochrome P450 1A2, 2C9, 2C19, 2D6, and 3A4 inhibition QSAR models [J]. *J Comput Aided Mol Des*, 2007, **21**(10/11): 559 - 573.
- [5] Kontijevskis A, Komorowski J, Wikberg JES. Generalized proteo-chemometric model of multiple cytochrome P450 enzymes and

- their inhibitors[J]. *J Chem Inf Model*, 2008, **48**(9): 1 840 – 1 850.
- [6] Mishra NK. Computational modeling of P450s for toxicity prediction[J]. *Expert Opin Drug Metab Toxicol*, 2011, **7**(10): 1 211 – 1 231.
- [7] Roy K, Roy PP. QSAR of cytochrome inhibitors[J]. *Expert Opin Drug Metab Toxicol*, 2009, **5**(10): 1 245 – 1 266.
- [8] Zhou SF, Yang LP, Zhou ZW, *et al.* Insights into the substrate specificity, inhibitors, regulation, and polymorphisms and the clinical impact of human cytochrome P450 1A2 [J]. *AAPS J*, 2009, **11**(3): 481 – 494.
- [9] Gunes A, Dahl ML. Variation in CYP1A2 activity and its clinical implications: influence of environmental factors and genetic polymorphisms[J]. *Pharmacogenomics*, 2008, **9**(5): 625 – 637.
- [10] Korhonen LE, Rahnasto M, Mahonen NJ, *et al.* Predictive three-dimensional quantitative structure-activity relationship of cytochrome P450 1A2 inhibitors[J]. *J Med Chem*, 2005, **48**(11): 3 808 – 3 815.
- [11] Vasanathan P, Taboureau O, Oostenbrink C, *et al.* Classification of cytochrome P450 1A2 inhibitors and noninhibitors by machine learning techniques[J]. *Drug Metab Dispos*, 2009, **37**(3): 658 – 664.
- [12] De Rienzo F, Fanelli F, Menziani MC, *et al.* Theoretical investigation of substrate specificity for cytochromes P450 1A2, P450 IID6 and P450 IIIA4[J]. *J Comput Aided Mol Des*, 2000, **14**(1): 93 – 116.
- [13] Didziapetris R, Dapkunas J, Sazonovas A, *et al.* Trainable structure-activity relationship model for virtual screening of CYP3A4 inhibition[J]. *J Comput Aided Mol Des*, 2010, **24**(11): 891 – 906.
- [14] Landis JR, Koch GG. The measurement of observer agreement for categorical data[J]. *Biometrics*, 1977, **33**(1): 159 – 174.
- [15] Chohan KK, Paine SW, Mistry J, *et al.* A Rapid computational filter for cytochrome P450 1A2 inhibition potential of compound libraries[J]. *J Med Chem*, 2005, **48**(16): 5 154 – 5 161.
- [16] Schuster D, Laggner C, Steindl TM, *et al.* Development and validation of an *in silico* P450 profiler based on pharmacophore models[J]. *Curr Drug Discov Technol*, 2006, **3**(1): 1 – 48.
- [17] Cho US, Park EY, Dong MS, *et al.* Tight-binding inhibition by alpha-naphthoflavone of human cytochrome P450 1A2 [J]. *Biochim Biophys Acta*, 2003, **1648**(1/2): 195 – 202.
- [18] Takemura H, Itoh T, Yamamoto K, *et al.* Selective inhibition of methoxyflavonoids on human CYP1B1 activity [J]. *Bioorg Med Chem*, 2010, **18**(17): 6 310 – 6 315.
- [19] Don MJ, Lewis DF, Wang SY, *et al.* Effect of structural modification on the inhibitory selectivity of rutaecarpine derivatives on human CYP1A1, CYP1A2, and CYP1B1 [J]. *Bioorg Med Chem Lett*, 2003, **13**(15): 2 535 – 2 538.
- [20] Nakamura T, Sato M, Kakinuma H, *et al.* Pyrazole and isoxazole derivatives as new, potent, and selective 20-hydroxy-5, 8, 11, 14-eicosatetraenoic acid synthase inhibitors[J]. *J Med Chem*, 2003, **46**(25): 5 416 – 5 427.
- [21] Kaila N, Janz K, DeBernardo S, *et al.* Synthesis and biological evaluation of quinoline salicylic acids as P-selectin antagonists [J]. *J Med Chem*, 2007, **50**(1): 21 – 39.
- [22] Panchal T, Bailey N, Bamford M, *et al.* Evaluation of basic, heterocyclic ring systems as templates for use as potassium competitive acid blockers (pCABs) [J]. *Bioorg Med Chem Lett*, 2009, **19**(23): 6 813 – 6 817.
- [23] Micheli F, Arista L, Bertani B, *et al.* Exploration of the Amine terminus in a novel series of 1,2,4-triazolo-3-yl-azabicyclo[3.1.0]hexanes as selective dopamine D3 receptor antagonists [J]. *J Med Chem*, 2010, **53**(19): 7 129 – 7 139.
- [24] Preissner S, Kroll K, Dunkel M, *et al.* SuperCYP: a comprehensive database on cytochrome P450 enzymes including a tool for analysis of CYP-drug interactions[J]. *Nucleic Acids Res*, 2010, **38**(Database issue): D237 – 243.
- [25] Carrillo JA, Benitez J. Clinically significant pharmacokinetic interactions between dietary caffeine and medications [J]. *Clin Pharmacokinet*, 2000, **39**(2): 127 – 153.
- [26] Masubuchi Y, Hosokawa S, Horie T, *et al.* Cytochrome P450 isozymes involved in propranolol metabolism in human liver microsomes. The role of CYP2D6 as ring-hydroxylase and CYP1A2 as *N*-desisopropylase[J]. *Drug Metab Dispos*, 1994, **22**(6): 909 – 915.