

抗病毒类化合物结构与离解速率的关系

谭建军, 周梓耘, 韩晓丁, 韩 丹, 王存新, 李春华, 张小轶
(北京工业大学生命科学与生物工程学院, 北京 100124)

摘要: 为了解决如何筛选出更加有效的抗病毒类化合物的问题,研究了离解速率(k_{off})与抗病毒类药物结构的关系,理论依据是 k_{off} 常用于评价药物在人体开放性系统中的活性. 首先,应用分子描述符软件计算出每个抗病毒类化合物的分子描述符,并使用多元逐步回归分析法、偏最小二乘法 and 遗传算法 3 种方法对描述符进行筛选. 然后,分别采用支持向量机和 BP 神经网络方法建立抗病毒类化合物的 k_{off} 的预测模型,并用测试集对模型进行了验证. 结果表明:筛选出了具有良好预测能力的描述符,建立的 2 个预测模型经验证均合理,对未来抗病毒类药物的研制具有指导意义.

关键词: 抗病毒化合物; 结合动力学; 离解速率; 支持向量机; BP 神经网络

中图分类号: R 914. 2

文献标志码: A

文章编号: 0254 - 0037(2017)12 - 1857 - 08

doi: 10. 11936/bjtxb2017040050

Relationship Between Structure and Dissociation Rates of Antiviral Compounds

TAN Jianjun, ZHOU Ziyun, HAN Xiaoding, HAN Dan, WANG Cunxin, LI Chunhua, ZHANG Xiaoyi
(College of Life Science and Bio-engineering, Beijing University of Technology, Beijing 100124, China)

Abstract: To solve the problem of how to screen out more effective antiviral compounds, the relationship between dissociation rate (k_{off}) and antiviral drug structure was studied in this research. The theoretical basis is that k_{off} is often used to evaluate the activity of the drug in the open system of the human body. The molecular descriptor of each antiviral compound was calculated by using molecular descriptor software, and the descriptors were screened by multiple stepwise regression analysis, partial least squares method and genetic algorithm. Then, the support vector machine (SVM) and BP neural network were used to establish the prediction model of antiviral compound structure and dissociation rate k_{off} value, and the model was verified. Results show that this experiment screens out the descriptors with good predictive power, and the two predictive models are proved to be reasonable and have guiding significance for the future development of antiviral drugs.

Key words: antiviral chemicals; binding kinetics; dissociation rate; support vector machine; back propagation neural network

药物设计早期多以体外评测出的复合物离解平衡常数(k_d)或 IC_{50} 值评价化合物的活性,但药物

实际上是在人体内的复杂环境下发挥作用. 人体是一个开放性系统,药物浓度在不断发生变化,药

收稿日期: 2017-04-27

基金项目: 国家自然科学基金资助项目 (21173014); 北京高等学校高水平人才交叉培养“实培计划”项目

作者简介: 谭建军(1965—), 男, 教授, 主要从事生物信息学方面的研究, E-mail: tanjianjun@bjut.edu.cn

物与受体结合形成复合物是药物呈现活性的基础,而这个过程受离解速率(k_{off})的影响^[1]. k_{off} 随着 pH 的增大逐渐增大,复合物的 k_{off} 越慢,意味着它占据受体的时间越长,维持体内药理作用的时间越长. 在新药研发早期,单靠化合物的 k_{d} 或 IC_{50} 作为取舍的标准,可能忽视一些亲和力稍逊但是 k_{off} 慢的化合物,这些化合物的优点在于作用时间长、选择性好,却被过早地淘汰了^[2]. 因此将动力学参数 k_{off} 考虑在内,有助于筛选出更加有效的抗病毒类化合物.

关于 k_{off} 与化合物结构的研究已有很多,例如:Carroll 等^[3] 研究了 2,4-二氨基嘧啶类化合物与二氢叶酸还原酶的结合-离解过程,证明这些抑制剂的离解速率与酶活性中心的构象改变相关. Dror 等^[4] 分别模拟了药物与 G 蛋白偶联受体中的 β_1 -肾上腺素受体 ($\beta_1\text{AR}$) 和 β_2 -肾上腺素受体 ($\beta_2\text{AR}$) 的结合以及与酪氨酸激酶抑制剂 PP1 ^[5] 结合的动力学过程,结果表明预测值与实验值一致^[1].

本文首先从参考文献中收集 k_{off} 值的实验数据;然后进行分子描述符的筛选;之后使用 2 种方法建立预测模型,分别是支持向量机方法和 BP 神经网络方法;最后评估 2 个预测模型的合理性.

定量构动关系 (quantitative structure-property relationship, QSPR) 研究是以分子描述符为基础,即将药物分子表达为若干个分子结构参数,然后建立此参数与药动学性质相关联的数学模型,探讨药物分子结构及性质与其在人体内的吸收、分布、代谢、排泄等药动学参数之间的定量关系,从而为优化药物分子结构以获得最佳人体药动学性质提供依据^[6]. 分子描述符在 QSPR 研究中经常被用于定量描述分子的结构和物理化学特性,预测各种药物及小分子的毒理学与药理学特征^[7]. 因此描述符的选择可以直接影响到预测模型的好坏. 支持向量机最早由 Vapnik 于 1992 年提出^[8],是一种具有某些优良特性的“线性分类器”,其理论基础是统计学习理论. 人工神经网络 (artificial neural networks) 是一种复杂网络系统,它由大量简单的处理单元广泛相互连接,可在基于一定算法的基础上对生物神经网络进行模仿. 神经网络学习训练中误差反向传播算法 (error back propagation, EBP) 的提出,解决了多层网络中隐含层神经连接权值的学习问题^[9].

1 原理与方法

1.1 分子描述符的筛选

本文收集了 120 个抗病毒类化合物及其 k_{off} ,其中包括抗人类免疫缺陷病毒化合物 (74 个)^[10-15]、抗丙型肝炎病毒化合物 (19 个)^[16-24]、抗流感病毒化合物 (27 个)^[25-32]. 运用 Sybyl 软件绘制出 120 个化合物的二维结构,再通过最陡下降法和共轭梯度法进行能量优化获取三维结构. 同时,由于 $\text{pk}_{\text{off}} (-\lg k_{\text{off}})$ 与自由能线性相关,因此将抑制剂的 k_{off} 转换成 pk_{off} ,作为实验的因变量. 之后,将抗病毒类化合物按结构和 pk_{off} 分成 2 组,分别为包含 99 个化合物的训练集和 21 个化合物的测试集,用于后续实验中 2 种模型的建立,旨在为今后预测抗病毒类化合物的活性、药物设计、分子结构改造等提供帮助.

用 Sybyl 软件包构建出 120 个抗病毒类化合物分子后,采用能量梯度法以其自带的分子力学程序 Minimize 进行能量优化,得到化合物的最低能量构象. 然后应用软件 MOE3.0 进行预处理,去掉晶体结构中的水分子并为蛋白加氢,对抑制剂分子能量进行最小化处理. 预处理结束后,将小分子抑制剂放置于受体的活性位点处,应用 AutoDock 软件对配体小分子和受体进行逐个对接,得到所有抗病毒类抑制剂结合自由能最低的构象,这些构象被认为是配体与受体生物大分子结合时可能的活性构象,即最优构象.

用 DRAGON 软件计算每个抗病毒类化合物分子的描述符^[33]. 先进行预处理,剔除方差为 0 的分子描述符;之后进行初步相关处理,计算 2 个描述符之间的相关系数,并去除其中与 k_{off} 的相关系数低的一个.

使用多元逐步回归分析法筛选描述符. 该方法的 2 个重要步骤为引进解释变量和剔除解释变量,本文利用 SPSS 软件进行逐步回归分析^[34]. 首先,引入偏回归平方和最大的解释变量,并对其进行 F 检验,若检验结果显著,则引进此解释变量,反之则剔除此解释变量. 直到不再有新变量被引入,同时也不再变量被剔除,则结束筛选过程^[35].

遗传算法是一种计算生物进化过程的模型,可以通过自然选择和遗传学建立自然进化的模型,再通过此模型来寻求全局最优解;偏最小二乘回归被用来进行多元统计数据分析^[36]. 与其他回归方法

不同的是, 偏最小二乘法回归是针对多因变量对多自变量的回归建模方法, 当输入变量之间存在高度相关性而无法剔除时, 偏最小二乘回归能提供可靠的回归模型^[37].

1.2 支持向量机

支持向量机的基础是统计学理论和 Vapnik-Chervonenkis (VC) 理论^[38], 最初提出是针对分类问题, 对于回归问题的研究还没有分类问题完善. 它将输入的样本集合变换到一个高维空间, 构造出一个最优超平面, 使得超平面到 2 类样本集之间的距离达到最大, 从而取得最好的泛化能力. 支持向量回归机对应的模型也是一个简单形式的二次规划问题, 通过引入新的损失函数, 使该方法能像支持向量分类机一样具有较好的推广能力, 不仅使回归估计具有鲁棒性, 而且使它的解是稀疏的. 留一法检验、交叉验证检验和独立测试集检验是检验模型预测能力的常用方法. 支持向量机包括 2 类参数: 一类是其固有的参数, 如惩罚参数 c ; 另一类是核函数中的参数, 如伽马参数 g . 支持向量机的分类准确率和拟合效果都由这些参数决定, 参数的选择决定着支持向量机模型的预测能力.

建模前需先进行分子描述符的归一化, 由于通过 DRAGON 计算所得的分子描述符数据在数量级上相差很大, 因此建模时要统一参数标准以避免不同参数之间数量级不同而造成的影响^[39].

对模型进行评估时, 描述测试集的线性回归参数 r_0^2 的计算公式^[40]为

$$r_0^2 = 1 - \frac{\sum (y_{\text{实}} - y_{\text{预}})^2}{\sum (y_{\text{预}} - y_{\text{均}})^2} \quad (1)$$

式中: $y_{\text{实}}$ 为测试集中分子的 k_{off} 值; $y_{\text{预}}$ 为根据模型进行预测得到的 k_{off} 值; $y_{\text{均}}$ 为测试集所有分子 k_{off} 值的均值.

据 Roy 等^[41]研究发现, r_m^2 是在 r^2 和 r_0^2 的基础上计算得到的, 它也是一个值得考虑的外部预测指示值. r_m^2 由式 (2) ~ (7) 计算得到

$$r_m^2 = r^2 [1 - (r^2 - r_0^2)^{1/2}] \quad (2)$$

一个好的模型应充分满足以下所有条件^[42-43]:

$$k = \sum (y_{\text{实}} y_{\text{预}}) / (y_{\text{均}})^2 \quad (3)$$

$$r^2 > 0.6 \quad (4)$$

$$[(r^2 - r_0^2) / r^2] < 0 \quad (5)$$

$$0.85 \leq k \leq 1.10 \quad (6)$$

$$r_m^2 > 0.5 \quad (7)$$

式中: k 为通过原点的回归直线斜率; r^2 为测试集交叉验证的相关系数, 多数样本集中的 r^2 高于 0.5, 就表明模型构建是有效的.

1.3 BP 神经网络

BP 神经网络是典型的多层前馈型网络, 由输入层、隐含层和输出层组成, 层与层之间多采用全部连接方式, 同一层单元之间不存在相互连接. BP 网络算法的基本思想是通过网络输出误差的反向传播, 不断调整和修改网络的连接权值, 从而使网络误差达到最小. BP 神经网络的训练过程包括前向计算和误差反向传播 2 个过程^[44]. 按有导师方式进行训练, 在已知输入模式以及期望响应的情况下, 根据减小实际响应和期望响应之间误差的原则, 从输出层经过各隐含层最后回到输入层, 逐层修正各连接权值; 不断进行这种误差反向传播训练, 网络对输入模式响应的正确率会不断提高. 利用 MATLAB 神经网络工具进行模式识别时, 输入层、中间层和输出层的神经元数目及相应的神经元传递函数需要根据样本特征维数和分类数确定, 选取网络训练方法和设置训练参数并对神经网络进行训练.

通过查阅文献^[45], 本文获得的人工神经网络预测模型的评判指标如下.

均方根偏差 (mean squared error, MSE):

$$\text{MSE} = \frac{\sum (e_j)^2}{N} \quad (8)$$

平均绝对误差 (mean absolute error, MAE):

$$\text{MAE} = \frac{\sum |e_j|}{N} \quad (9)$$

平方差 (sum of squares for error, SSE):

$$\text{SSE} = \sum (e_j)^2 \quad (10)$$

式中: e_j 为预测误差 ($e_j = y_{\text{预}} - y_{\text{实}}$); N 为样本数.

2 结果与讨论

2.1 筛选描述符

实验经过初步筛选获得 55 个分子描述符, 其中包括 11 个简单的分子性质描述符、9 个电拓扑分子描述符、9 个连接度指数描述符、17 个分子几何特性描述符和 9 个分子特性描述符. 运用 MATLAB 自带的偏最小二乘-遗传算法 (GA-PLS) 工具箱来筛选最优的分子描述符集^[46]. 实验选用了 GA-PLS 工具箱里的 3 个函数: 1) GAPLSOPT (dataset, 1), 用来评测样本数据的可行性; 2) GAPLSOPT (dataset,

2),用来选取最佳评价次数;3) GAPLS(dataset, the number of evaluationg, precision),用来选定最优分子描述符。

GAPLSOPT(dataset, 1) 测试结果很明显在 0 ~ 1.296 5,说明本实验的数据可以用 GA-PLS 方法进行筛选。GAPLSOPT(dataset, 2) 函数输出的结果,折线的最高点在 111 处,因此选择 111 为最优评价次数(the number of evaluation),而选定好最优评价次数后的训练过程中就不会出现过拟合现象。为了避免增大随机误差,重复 10 次 GAPLS 实验以获取平均结果。在 55 个分子描述符中,GAPLS 函数最终选出了 8 个分子描述符(见表 1)用于下一步的预测分析。

表 1 筛选出的分子描述符

Table 1 Screened molecular descriptors

描述符	英文名称及简写
平均相对分子质量	average molecular weight, AMW
平均电拓扑状态	mean electrotopological state, Ms
氮原子数	number of nitrogen atoms, nN
径向基函数	rotatable bond fraction, RBF
平均分子连接性指数 1	average valence connectivity index chi-1, X1Av
不饱和指数	unsaturation index, Ui
3D-Wiener 指数	3D-Wiener index, W3D
流动性	sphericity, SPH

表 2 BP-ANN 隐含层数变化下的网络误差

Table 2 Network error after changing hidden layers of BP-ANN

隐含层数	10	11	12	13	14	15	16	17	18
网络误差	2.71	1.66	0.99	0.81	0.84	1.65	2.83	1.20	1.53

在 BP 网络中,增加隐含层数可提高 BP 网络的处理能力,但同时也会导致训练时间延长。为了找到合适的隐含层数,最好的办法是在网络学习过程中,根据环境要求,自主地学习、调整结构,最后得到一个大小合适的神经网络模型,即在训练的过程中合理地合并或删除隐含层数^[49]。通过实验过程的调整,将隐含层数定为 13,其网络误差值最小为 0.81。之后选取不同的训练函数进行网络训练,再通过平均绝对误差的比较确定训练函数,见表 3。

由表 3 可知,Trainlm 函数的平均绝对误差最小为 0.000 7,因此选定 Trainlm 函数为训练函数,学习

2.2 建立 2 种模型

2.2.1 支持向量机模型的建立

选用 MATLAB 中的函数 mapminmax 来对数据进行归一化处理,数据范围在 1 ~ 2。选用 SVM 方法搭配径向基 RBF 函数,对 $p_{k_{off}}$ 建立 SVM 模型。模型基于 99 个化合物的训练集建立,并用剩下 21 个化合物组成的测试集进行评估^[47]。使用 SVM 程序中内置的自动参数寻优工具 SVMcgForRegress 来寻找最佳惩罚参数 c 和最佳核函数参数 g ^[48]。首先,对 120 个化合物进行 5 倍交叉检验,检验正确率的高低决定对 (c, g) 的选择,并通过手动的方式进行微调,直到寻找到最优参数。最终确定建模使用的参数为:Best $c = 0.50$, Best $g = 4$ 。

2.2.2 BP 神经网络模型的建立

首先,也需要先将分子描述符进行归一化处理,方法与 SVM 建模时相同。利用 MATLAB 中的函数 mapminmax 将分子描述符和目标值都进行归一化处理,将所有数据都限定在 $-1 \sim 1$,这样对下面 BP 神经网络的建模有利。

建模时为保证实验结果的可比性,BP 神经网络选取的参数与 SVM 的保持一致,运用的训练集和测试集也相同。BP 网络的输入节点数为 $p = 8$,输出节点数为 $t = 1$,也就是抗病毒类化合物分子描述符 AWM、Ms、nN、RBF、X1Av、Ui、W3D 和 SPH。本实验为了不让网络计算变得过于复杂,在保证误差率很小的前提下,将网络误差作为选取隐含层的依据,数据见表 2。

率 $lr = 0.5$,最大迭代次数 $epoch = 1000$,训练目标误差 $goal = 0.0004$,动力因子 $mc = 0.8$ 。

表 3 BP-ANN 训练函数变化下的平均绝对误差

Table 3 Average absolute error after changing training function of BP-ANN

训练函数	Traingdm	Traingdx	Trainlm	Traingda
平均绝对误差	0.002 5	0.001 9	0.000 7	0.015 2

2.3 模型的评估

2.3.1 支持向量机模型的评估

为判断模型的预测能力,下面引入测试集来验

证 2 种方法建立的模型. 表 4 中总结了 SVM 方法建立预测模型的统计数据. 其中测试集交叉验证的相关系数 $r^2 = 0.79 > 0.60$, $r_m^2 = 0.56 > 0.50$, 回归直线的斜率 $k = 0.98$ 在 $0.85 \sim 1.10$, $[(r^2 - r_0^2)/r^2] = 0.01 < 0.10$, 结果符合评估模型的各项标准, 因此认为 SVM 方法建立的模型合理.

表 4 SVM 预测模型的统计结果

Table 4 Statistics and predictive power of SVM model

项目	$r_{\text{训}}^2$	r^2	$(r^2 - r_0^2)/r^2$	k	r_m^2
SVM	0.85	0.79	0.01	0.98	0.56

以实验值为横坐标、预测值为纵坐标做散点图, 可由线性关系直观看出模型的优劣. 如果预测值能均匀分布在 $Y = X$ 两侧, 就证明该模型的预测能力较好. 得到模型训练集和测试集的散点图见图 2.

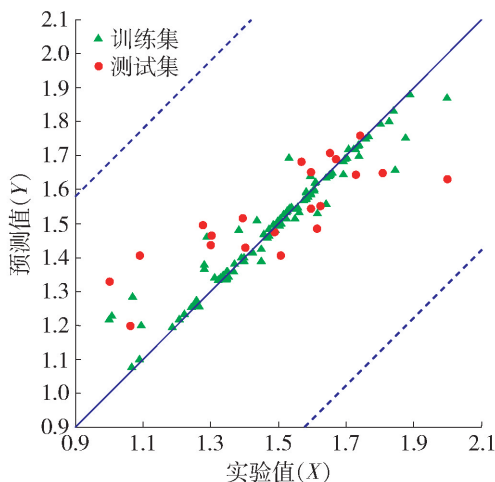


图 1 SVM 模型训练集和测试集离解速率实验值与预测值之间的相关性

Fig. 1 Relevance between the experimental dissociation rate and predicted set of training set and testing set of SVM model

所得模型的样本估算标准偏差 $SD = 0.193$, 其值的 3 倍为 0.578 , 由该值做图 1 中虚线, 可知所有实验值和预测值都在合理偏差范围内. 模型的交叉验证相关系数平方 r^2 和均方根误差 MSE 分别是: 训练集 $r_{\text{训}}^2 = 0.85$, $MSE = 0.006$; 测试集 $r_{\text{测}}^2 = 0.79$, $MSE = 0.046$. 综上, 实验建立的 SVM 预测模型满足所有模型成立条件, 而且可以对抗病毒类化合物的离解速率进行较为准确的预测.

2.3.2 BP 神经网络模型的评估

对 BP 神经网络建立的预测模型进行训练集和测试集的回归预测, 所得模型的交叉验证相关系数

平方 r^2 和均方根误差 MSE 分别是: 训练集 $r_{\text{训}}^2 = 0.79$, $MSE = 0.015$; 测试集 $r_{\text{测}}^2 = 0.67$, $MSE = 0.053$. 预测模型测试集的散点图如图 2 所示.

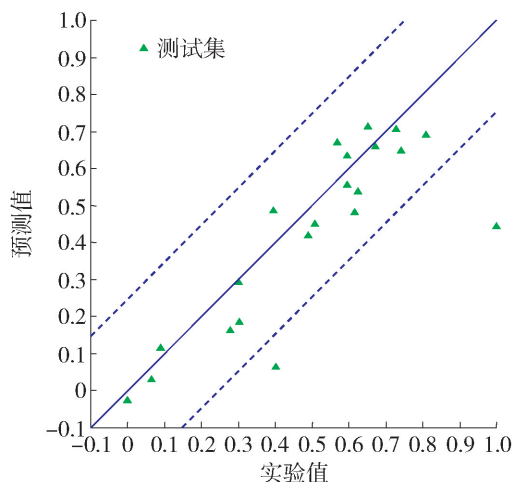


图 2 BP-ANN 模型测试集离解速率实验值与预测值的相关性

Fig. 2 Relevance between the experimental dissociation rate and predicted set of testing set of BP-ANN model

所得模型的样本估算标准偏差 $SD = 0.251$, 其值的 3 倍为 0.753 , 由该值做图 2 中虚线, 可知大部分点落在虚线区域内, 2 个未落在区域内的点, 可能来自测试集数据的误差, 因所占比例很小, 可忽略影响, 认为该模型合理. 与 SVM 相同, 为验证模型的预测性, 实验引入外部数据集来测试模型的预测能力. 通过计算可得 $MSE = 0.053$, $MAE = 0.040$, $SSE = 0.076$. BP 神经网络所建立的预测模型经上述验证证实是可靠的, 但在抗病毒类化合物离解速率的预测方面精确度还不够.

3 结论

1) 本文分别使用 SVM 和 BP 神经网络 2 种方法对抗病毒类化合物离解速率的预测进行了研究. 由实验数据分析可以发现, 2 种预测方法存在一定差异. SVM 方法建立的模型训练集 $r_{\text{训}}^2 = 0.85$, $MSE = 0.006$; 测试集 $r_{\text{测}}^2 = 0.79$, $MSE = 0.046$. 而 BP 神经网络建立的模型训练集 $r_{\text{训}}^2 = 0.79$, $MSE = 0.015$; 测试集 $r_{\text{测}}^2 = 0.67$, $MSE = 0.053$. 经评估可知, 实验建立的 2 种预测模型均可靠, 并且基于 SVM 对抗病毒类抑制剂离解速率 k_{off} 进行预测的方法成功申请了专利^[50], 可为抗病毒类药物的研制提供依据.

2) 从预测能力上看, SVM 方法建立的模型更好. 另外, 对 2 种方法建模的实验过程进行对比, SVM 的实验过程基本上属于自动完成, 其内部参数也是根据训练的进度自动生成的, 各种权值在训练过程中自动调整; 而 BP 神经网络的各种参数都需设置, 并且目前还没有科学依据能够确定神经网络隐含层的数量, 方便程度和准确性都不如 SVM 方法. 定量构效关系研究的关键步骤是选择合适的计量学方法建立结构参数与活性数据之间的定量关系模型, 而化合物结构与生物活性之间存在的大多数是非线性相关关系. 因此, 适合解决非线性问题的 SVM 方法更有优势.

3) 本研究对 SVM 和 BP 神经网络在分析化合物生物活性及其结构之间的定量变化规律进行了探索, 模型筛选出的化合物拥有更高的成药可能性, 希望能为抗病毒类药物的研制提供新方法, 并为更多相关研究提供依据.

参考文献:

- [1] 郭颖, 郭宗儒. 药物与受体靶标结合动力学对体内药效的影响[J]. 中国药物化学杂志, 2014(1): 70-77.
GUO Y, GUO Z R. Significance of binding kinetics in drug efficacy [J]. Chinese Journal of Medicinal Chemistry, 2014(1): 70-77. (in Chinese)
- [2] 郭颖, 郭宗儒. 从噻托溴铵谈药物与受体的结合动力学[J]. 中国新药杂志, 2013(20): 2353-2357.
GUO Y, GUO Z R. Drug-receptor binding kinetics: tiotropium bromide as an example[J]. Chinese Journal of New Drugs, 2013(20): 2353-2357. (in Chinese)
- [3] CARROLL M J, MAULDIN R V, GROMOVA A V, et al. Evidence for dynamics in proteins as a mechanism for ligand dissociation [J]. Natl Chen Biol, 2012, 8(3): 246-252.
- [4] DROR R O, PAN A C, ARLOW D H, et al. Pathway and mechanism of drug binding to G-protein-coupled receptors [J]. Proc Natl Acad Sci USA, 2011, 108(32): 13118-13123.
- [5] SHAN Y, KIM E T, STWOOD M P, et al. How does a drug molecule find its target binding site? [J]. J Am Chem Soc, 2011, 133(24): 9181-9183.
- [6] 陈宪. 基于 OECD 准则对 QSAR/QSPR 模型几个重要问题的研究[D]. 长沙: 中南大学, 2013.
CHEN X. Studies on a few key problems of QSAR/QSPR modeling based on the OECD principles[D]. Changsha: Zhongnan University, 2013. (in Chinese)
- [7] MILLER D C. Investigation of the effect of molecular

properties on the binding kinetics of a ligand to its biological target [J]. Medicinal Chemistry Communication, 2012, 3(3): 449-452.

- [8] 于晓庆. 基于支持向量机的生物数据分析方法研究[D]. 上海: 上海师范大学, 2012.
YU X Q. The study of biological data analysis method based on support vector machines [D]. Shanghai: Shanghai Normal University, 2012. (in Chinese)
- [9] 章建文. 基于压电阻抗技术的焊接结构健康监测研究[D]. 南京: 南京航空航天大学, 2010.
ZHANG J W. Research onwelding structure health monitoring using piezoelectric impedance technology[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2010. (in Chinese)
- [10] JOHAN W, ELDAR A, VERA B, et al. Resolution of the interaction mechanisms and characteristics of non-nucleoside inhibitors of hepatitis C virus polymerase. Antiviral Research, 2013, 97: 356-368.
- [11] GUANGHUI Y, JEROME D, BAOCHANG F, et al. Biochemical study of the comparative inhibition of hepatitis C virus RNA polymerase by VX-222 and filibuvir [J]. Antimicrobial Agents and Chemotherapy, 2012, 56(2): 830-837.
- [12] SOFIA S G, ANGELICA E, BENJAMIN S, et al. Identification of weak points of hepatitis C virus NS3 protease inhibitors using surface plasmon resonance biosensor-based interaction kinetic analysis and genetic variants[J]. J Med Chem, 2014, 57: 1802-1811.
- [13] RAVI R, SHAWN M, SARAH K S, et al. Inhibition and binding kinetics of the hepatitis C virus NS3 protease inhibitor ITMN-191 reveals tight binding and slow dissociative behavior [J]. Biochemistry, 2009, 48: 2559-2568.
- [14] ONA B, AMOREENA C C, RUTH W, et al. Binding kinetics, potency, and selectivity of the hepatitis C virus NS3 protease inhibitors GS-9256 and vedroprevir [J]. Biochimica Et Biophysica Acta, 2014, 1840(12): 3292-3298.
- [15] ORIGENE N, BENOIT D, LEEN V, et al. 1a/1b subtype profiling of nonnucleoside polymerase inhibitors of hepatitis C virus[J]. Journal of Virology, 2010, 84(6): 2923-2934.
- [16] MATTHIS G, GORAN D, HELENA D. Mechanistic and kinetic characterization of hepatitis C virus NS3 protein interactions with NS4A and protease inhibitors[J]. J Mol Recognit, 2011, 24: 60-70.
- [17] JOHAN W, ELDAR A, VERA B, et al. Resolution of the interaction mechanisms and characteristics of non-

- nucleoside inhibitors of hepatitis C virus polymerase[J]. *Antiviral Research*, 2013, 97: 356-368.
- [18] MARIA V F, JOANNE S, GIUSEPPE C, et al. HCV-NS3 inhibitors: determination of their kinetic parameters and mechanism [J]. *Biochimica Et Biophysica Acta*, 2009, 1794: 1441-1448.
- [19] ASHA V, CAROLNE M P. Simeprevir: first global approval[J]. *Drugs*, 2013, 73: 2093-2106.
- [20] NATHALIE G, ZHAO S P, OLIVER H, et al. Combined X-ray, NMR, and kinetic analyses reveal uncommon binding characteristics of the hepatitis C virus NS3-NS4A protease inhibitor BI 201335 [J]. *Journal of Biological Chemistry*, 2011, 286(13): 11434-11443.
- [21] HAO L, PETER J T. Drug-target residence time: critical information for lead optimization [J]. *Curr Opin Chem Biol*, 2010, 14(4): 467-474.
- [22] DAVID R. L, HIMADRI K. S, LIN Z Y, et al. The terminal (catalytic) adenosine of the HIV LTR controls the kinetics of binding and dissociation of HIV integrase strand transfer inhibitors [J]. *Biochemistry*, 2008, 47: 13481-13488.
- [23] EDWARD P G, BENJAMIN S, MARGARET J G, et al. Potent inhibitors of HIV-1 integrase display a two-step, slow-binding inhibition mechanism which is absent in a drug-resistant T66I/M154I mutant [J]. *Biochemistry*, 2009, 48: 1644-1653.
- [24] KENDRA E H, WANG R L, FELIX D, et al. Dolutegravir (S/GSK1349572) exhibits significantly slower dissociation than raltegravir and elvitegravir from wild-type and integrase inhibitor-resistant HIV-1 integrase-DNA complexes [J]. *Antimicrobial Agents and Chemotherapy*, 2011, 55(10): 4552-4559.
- [25] CYNTHIA F. S, LOTTA V, U. HELENA D. Improved structure-activity relationship analysis of HIV-1 protease inhibitors using interaction kinetic data [J]. *J Med Chem*, 2004, 47: 5953-5961.
- [26] WESLEY S, MARKKU H, ANDERS K, et al. Relationships between structure and interaction kinetics for HIV-1 protease inhibitors [J]. *J Med Chem*, 2002, 45: 5430-5439.
- [27] DAN B, MICHEL M, U. HELENA D. Biosensor-based screening and characterization of HIV-1 inhibitor interactions with Sap 1, Sap 2, and Sap 3 from *Candida albicans* [J]. *Journal of Biomolecular Screening*, 2006, 11(2): 165.
- [28] PAUL L D, SUSAN P J, DAWN L H, et al. Dissociation and association of the HIV-1 protease dimer subunits; equilibria and rates [J]. *Biochemistry*, 1994, 33: 98-105.
- [29] PER-OLOF M, WESLEY S, MARKKU H, et al. Relationships between structure and interaction kinetics for HIV-1 protease inhibitors [J]. *J Med Chem*, 2002, 45: 5430-5439.
- [30] ALBERTA S, EMMANUELE C, SAMANTA V, et al. Slow binding-tight binding interaction between benzimidazol-2-one inhibitors and HIV-1 reverse transcriptase containing the lysine 103 to asparagine mutation [J]. *Antiviral Research*, 2010, 86: 268-275.
- [31] MATTHIS G, TORSTEN U, U. HELENA D. Interaction kinetic characterization of HIV-1 reverse transcriptase non-nucleoside inhibitor resistance [J]. *J Med Chem*, 2006, 49: 2375-2387.
- [32] ROBERT A C, DAVID L P, THOMAS D M. Drug-target residence time and its implications for lead optimization [J]. *Nature Reviews Drug Discovery*, 2006, 5(9): 730-739.
- [33] WANG T, CHEN S, SUN J, et al. Identification of factors influencing the pharmacokinetics of voriconazole and the optimization of dosage regimens based on Monte Carlo simulation in patients with invasive fungal infections [J]. *Journal of Antimicrobial Chemotherapy*, 2013, 69(2): 463-470.
- [34] NAVEA S, TAULER R, JUAN A. Application of the local regression method interval partial least-squares to the elucidation of protein secondary structure. [J]. *Analytical Biochemistry*, 2004, 336(2): 231-242.
- [35] 李鹏霞. 基于偏最小二乘和支持向量机的生物分子的定量结构—活性/性质关系研究[D]. 兰州: 西北师范大学, 2011.
- LI P X. Quantitative structure-activity/property relationship studies in biomolecules based on partial least squares and support vector machine [D]. Lanzhou: Northwest Normal University, 2011. (in Chinese)
- [36] 张娟. 潜变量统计建模方法的发展演化过程及其演化规律研究[D]. 济南: 山东财经大学, 2010.
- ZHANG J. Development of latent variable statistical modeling techniques and its evolution of rule [D]. Jinan: Shandong Economic University, 2010. (in Chinese)
- [37] 张晓丽. 偏最小二乘回归分析若干问题的研究[D]. 长春: 东北师范大学, 2010.
- ZHANG X L. Research on a number of issues related to partial least squares regression analysis [D]. Changchun: Northeast Normal University, 2010. (in Chinese)
- [38] GARUTI L, ROBERTI M, BOTTEGONI G.

- Benzimidazole derivatives as kinase inhibitors [J]. *Current Medicinal Chemistry*, 2014, 21 (20): 2284-2298.
- [39] NOBLE W S. What is a support vector machine? [J]. *Nature Biotechnology*, 2006, 24(12): 1565-1567.
- [40] MASAKI B, YUSUKE K, WEI C, et al. Development of a sugar-binding residue prediction system from protein sequences using support vector machine [J]. *Computational Biology and Chemistry*, 2016, 66: 36-43.
- [41] SUN X D, HUANG R B. Prediction of protein structural classes using support vector machines[J]. *Amino Acids*, 2006, 30(4): 469-475.
- [42] ADRIANA M, BOGDAN Z, ADRIAN O, et al. Mapping vegetation communities of the karkonosze national park using APEX hyperspectral data and support vector machines[J]. *Miscellanea Geographica*, 2014, 18(2): 23-29.
- [43] 李佳庆. 最小二乘支持向量机在软测量建模中的应用[D]. 南京: 南京邮电大学, 2015.
- LI J Q. Application of LSSVM in soft sensor modeling [D]. Nanjing: Nanjing University of Posts and Telecommunications, 2015. (in Chinese)
- [44] 杨光, 刘俏, 代蕊, 等. BP神经网络预测 *Bacillus amyloliquefaciens* Q-426 发酵产物活性[J]. *计算机与应用化学*, 2013(9): 1055-1058.
- YANG G, LIU Q, DAI R, et al. Predicting antibacterial activity of *Bacillus amyloliquefaciens* Q-426 in shake flask culture based on BP neural network [J]. *Computers and Applied Chemistry*, 2013(9): 1055-1058. (in Chinese)
- [45] 王炳萱. LM优化算法和神经网络预测控制在非线性系统中的研究[D]. 太原: 太原理工大学, 2016.
- WANG B X. Research on LM optimization algorithm and neural network predictive control in nonlinear system [D]. Taiyuan: Taiyuan University of Technology, 2016. (in Chinese)
- [46] 吕娟. 蛋白酶体抑制剂的定量构效关系及分子设计研究[D]. 重庆: 重庆大学, 2012.
- LÜ J. QSAR and molecular design studies of proteasome inhibitors [D]. Chongqing: Chongqing University, 2012. (in Chinese)
- [47] 肖晓, 张敏. 支持向量机多分类问题研究 [J]. *淮海工学院学报*, 2014(3): 28-31.
- XIAO X, ZHANG M. Research on the multi-class classification problem of support vector machine [J]. *Journal of Huaihai Institute of Technology*, 2014(3): 28-31. (in Chinese)
- [48] 周源. 蛋白质-蛋白质相互作用的简易预测工具开发与实例研究[D]. 北京: 中国农业大学, 2015.
- ZHOU Y. Protein-protein interaction: simple prediction tool development and studies on specific cases [D]. Beijing: China Agricultural University, 2015. (in Chinese)
- [49] 刘维群, 李元臣. BP网络中隐含层节点优化的研究 [J]. *交通与计算机*, 2005(2): 83-86.
- LIU W Q, LI Y C. Research on optimization of implicit layer node in BP network [J]. *Traffic and Computer*, 2005(2): 83-86. (in Chinese)
- [50] 谭建军, 韩晓丁, 李春华, 等. 一种基于支持向量机 (support vector machine, SVM) 对抗病毒类抑制剂离解速率 k_{off} 值进行预测的方法; CN201510919482.5 [P]. 2016-03-30.

(责任编辑 吕小红)