

使用离子对反相色谱联用串联质谱法进行亲水性

代谢物定量结构-保留关系预测

王欢¹, 胡庆宇¹, 孙玉婷¹, 王玉兰¹, 唐惠儒¹

(1. 复旦大学人类表型组研究院, 上海 201203)

摘要: 亲水性代谢物在传统的反相 (RP) 色谱柱上的保留性较差, 阻碍了 RPLC-MS 方法的有效分析。本研究开发了一种基于己胺的 IPRP-UHPLC-QTOFMS 方法, 并通过实验测量了 183 种亲水性代谢物的保留时间 t_R ; 然后, 本研究系统地研究了定量结构-保留关系 (QSRR), 并利用测得的 t_R 构建了定量结构-保留关系 (QSRR) 模型; 随后, 使用随机森林回归算法用于预测代谢物保留时间, 将其应用于 sMRM 模式下的 IPRP-UHPLC-MS/MS 方法进行定量代谢组学分析。我们的方法可以同时定量 260 多种代谢物。此外, 我们发现该方法适用于多种主要生物基质, 包括生物体液和组织。这种方法为大规模定量亲水性代谢组学分析提供了一种有效的方法。

关键词: UHPLC-MS; 定量结构-保留关系; 代谢组学; 离子对反相色谱

中图分类号: Q591

文献标识码: 期刊[J]

Quantitative structure-retention relationship for reliable metabolite identification and quantification in metabolomics using ion-pair reversed-phase chromatography coupled with tandem mass spectrometry

Qingyu Hu¹, Yuting Sun¹, Yulan Wang¹, Huiru Tang¹

(1. Human phenome Institute, Fudan University, Shanghai 201203, China)

Abstract: Hydrophilic metabolites have poor retention of these metabolites on reversed-phase (RP) chromatographic column hinders their effective analysis with RPLC-MS methods. We first developed a hexylamine-based IPRP-UHPLC-QTOFMS method and experimentally measured retention time (t_R) for 183 hydrophilic metabolites. We then systematically investigated the quantitative structure-retention relationship (QSRR) and constructed QSRR models using the measured t_R . Subsequently, we developed a retention time predictive model using the random-forest regression algorithm for predicting metabolite retention time, which was applied in

IPRP-UHPLC-MS/MS method in sMRM mode for quantitative metabolomic analysis. Our method can simultaneously quantify more than 260 metabolites. Moreover, we found that this method was applicable for multiple major biological matrices including biofluids and tissues. This approach offers an efficient method for large-scale quantitative hydrophilic metabolomic profiling even when metabolite standards are unavailable.

Keywords : UHPLC-MS; Quantitative structure-retention relationship;Metabolomics; Ion-pair reversed-phase chromatography

1 实验仪器和材料

1.1 仪器

Agilent 6530 QTOF 质谱系统配置 1290 UHPLC 系统 (Agilent Technologies, USA)。

1.2 试剂和药品

LC 级的甲醇和乙腈购买自美国 Fisher Scientific 公司(NJ USA)。超纯水(电阻率>18.2 MΩ/cm)使用 Millipore Elix Advantage 超纯水系统制备(German)。正己胺(Hexylamine, HXA)和 LC 级的乙酸, 乙酸铵, 氨水购于 Sigma-Aldrich(MO USA)。。

2 实验方法

2.1 色谱条件优化

色谱条件优化使用 8 种具有代表性的代谢物标准品: L-天冬氨酸, L-苯丙氨酸, α -酮戊二酸, L-谷氨酸, 4-酮戊酸, AMP, CDP 和 ATP。将 1mg/mL 上述标准品配制终浓度为 50 μ g/mL 的混标。上机对离子对试剂的浓度和流动相 pH 值进行条件优化。

表 1 离子对反相色谱的色谱参数

色谱条件	实验参数
色谱柱	Waters BEH C ₁₈ (2.1 mm \times 100 mm,1.7 μ m)
柱温	50°C
流速	0.5 mL/min
流动相 A	5 mM HXA in Water, pH=10, acetic acid adjusted
流动相 B	90:10 MeOH:10 mM NH ₄ Ac,NH ₃ ·H ₂ O adjusted

梯度	0-2 min, 0% B, 2 min-5 min: 0-20% B, 5 min-10 min, 20-50%B; 10-12 min, 50-100%B, 12-14 min, 100%B, 14-14.1 min,0%B, 14.1-17 min, 0%B
----	--

2.2 质谱参数

质谱检测前需要进行负离子调谐和校正。校正后质量精度应小于 5 ppm。由于使用标准品实验, 采样时关闭实时校正, 降低校正液的离子抑制。质谱采集使用负离子模式, MS 模式和 AutoMSMS 模式, 扫描速度为 MS: 1 spectra/s, MSMS: 2 spectra/s; 离子源参数设置如下: 干燥气温度(Gas temperature): 350°C, 干燥气流速(Drying Gas Flow): 8L/min, 毛细管电压(VCap): 3500 V, 离子源传输电压(Fragmentor): 100 V, 雾化器压力(Nebulizer): 30 psi, 锥孔电压 Skimmer: 65V, 八级杆传输电压(OCT1 RCF Vpp):750 V。LC-MS 数据使用 MassHunter Qualification (Agilent, B.06.00)定性软件进行分析。

2.2 量子化学计算方法

量子化学计算使用 Gaussian 9.0 (Wallingford CT USA) 和 Gauss View 6.0(Wallingford CT USA)进行分析。采样密度泛函(Density Function Theory, DFT)方法中的明尼苏达泛函 M06-2X 系列, 可以对分子弱相互作用高精度计算。分子的几何构型优化和频率分析选择基组为 6-31G(d,p),使用隐式溶剂模型(SMD)校正溶液种离子行为。Multiwfn (Beijing China) 软件用于分子静电势能面计算, VMD 1.9 program(Urbana IL,USA) 用于可视化展示, Gaussian 计算由中国科学技术大学超算中心帮助完成。

2.3 定量结构-保留时间预测模型建立

保留时间预测模型的搭建如图 1 所示。首先, 我们测量了 183 种代谢物标准品的保留时间, 并通过数据库获得每种标准品化合物的分子描述符 (Molecular descriptors,MDs)。这些描述符为分子物理化学结构, 拓扑学结构等半经验参数, 相当于一个分子的指纹身份证。利用 MDs 作为自变量和保留时间作为因变量构建输入数据集; 然后, 通过多种机器学习模型搭建保留时间预测模型, 通过网格点搜索算法遍历超参数 (hyperparameters) 的取值范围, 利用 10-折交叉验证进行超参数优化, 获得最佳的预测模型。

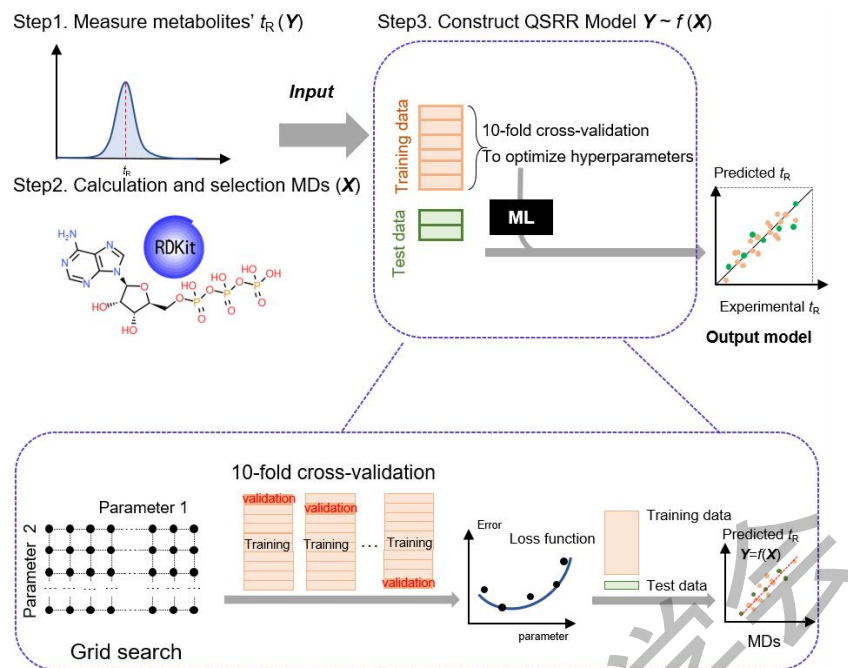


图 1 定量结构-保留时间预测模型搭建和网格点搜索的超参数优化过程总览。

3 结果与讨论

3.1 实验结果

尽管量子化学计算预计保留时间的预测精度很高,但其耗时的性质限制了其在涉及数百种分析物的代谢组学中的应用。研究表明,使用更经济的半经验计算参数进行反相色谱的保留时间预测,依旧可以达到较好的预测精度。因此,在本研究中,我们使用 QSRR 模型来预测 IP-RPLC 代谢物保留时间。我们从 RDKit 库中获得了大约 200 个分子描述符 (MDs) 作为 183 种代谢物中每个代谢物的变量,并使用我们的 IPRP-LC-QTOFMS 方法进行测量了它们的 t_R 。通过排除缺失值过多、标准差低、对模型贡献低和变量间相关性高的变量;利用剩余的 61 个 MDs 来构建 t_R 预测模型。值得注意的是,12 个 MD 对模型的贡献很大(图 2a),包括分子拓扑学参数(如 Kappa1, BalabanJ)、电子拓扑结构参数(Estate 相关参数)、疏水性物化参数(logP),几何结构参数(Chi0, Chi1)等。这表明静电相互作用相关参数在 IP-RPLC-MS 的保留时间预测中起着重要作用。一般而言,疏水参数(如 logP)是主导反相色谱和亲水色谱的保留时间预测模型的主要参数,而这项研究揭示了静电参数主导离子对色谱保留,说明离子对色谱在保留机制上的独特性。

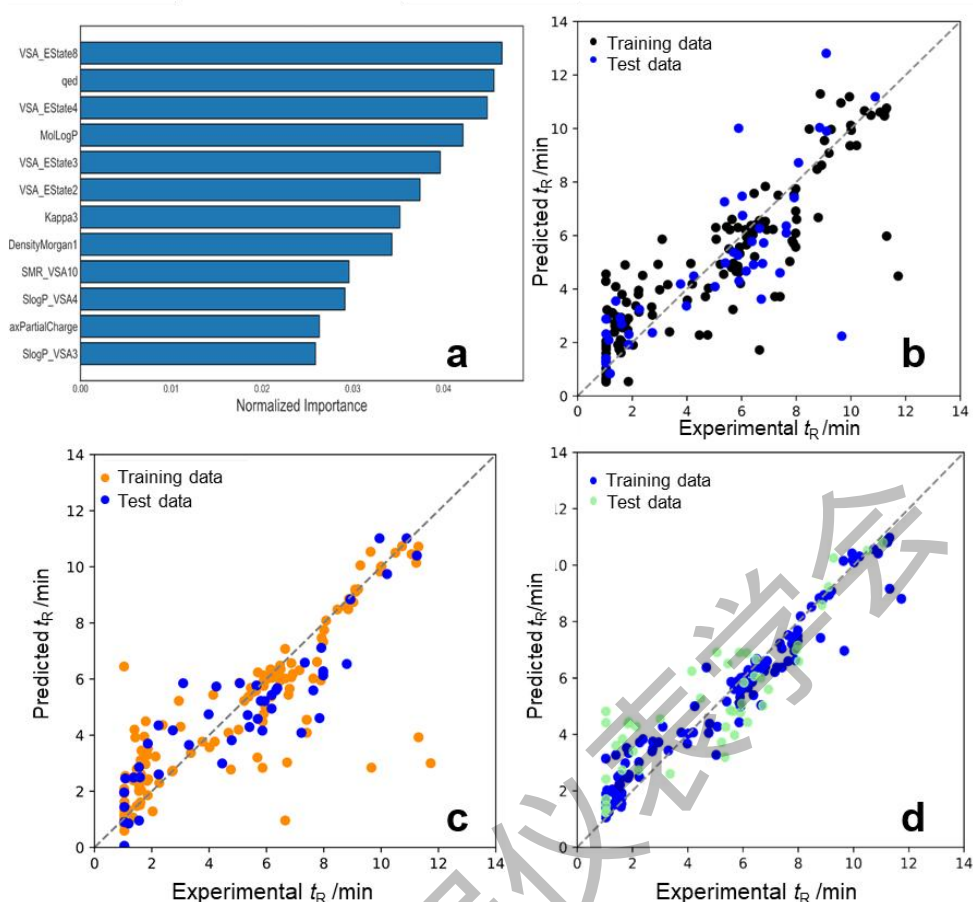


图 2 不同算法构建的 QSRR 模型特征。

(a) 高贡献的分子描述符；(b) 逐步回归分析的拟合曲线；(c) SVR 模型；(d) 随机森林回归模型。

通过采用逐步回归、SVR 和随机森林回归 (RFR) 这三种机器学习算法, 将每个代谢物 61 个筛选的 MDs 作为自变量, 测量的保留时间作为因变量构造输入数据集。在逐步回归建模中, 使用赤池信息准则 (AIC) 和逆向回归策略 (图 2b) 进一步选择参数。在 SVR 建模中, 使用了多项式核函数和高斯径向基核函数 (RBF) 进行建模, 并对 ϵ , C , γ 三个超参数分别进行了网格点搜索优化和 10-折交叉验证 (图 2c)。对于 RFR 回归, 优化得到 100 棵决策树用于构建 RFR 模型, 获得最优回归曲线 (图 2d)。

通过对三种模型性能进行评价 (表 2), 其中, r^2 , q^2 表征了模型整体的预测能力, MAE 和 Med AE 作为绝对预测误差的评价, 选择 Med AE 可以规避异常点对模型整体性能的影响。虽然三种保留时间预测模型的 Med AE 均在 1 min 以内, 随机森林回归模型的性能更佳。

表 2 三种 QSRR 模型的质量评价

QSRR Model	超参数	r^2	q^2	MAE/min	MedAE/min
------------	-----	-------	-------	---------	-----------

逐步回归	—	0.75	0.67	1.32	1.00
RBF-SVR	C=100, $\epsilon=0.1$, $\gamma=0.01$	0.76	0.72	1.16	0.94
RFR	ntree=100	0.93	0.70	1.28	1.06

此外,我们将这种方法分别应用于血浆以及肝脏、脾脏、心脏和肾脏组织的混合 QC 样品的 sMRM 分析。柱上 CE 优化后, 164 个代谢物在 QC 样品的稀释实验中表现出良好的线性 ($r^2 > 0.96$) 因此被认为是可靠的量化结果; 我们分别从兔血浆、肝脏、脾脏、心脏和肾脏样品中可靠地鉴定了 105、156、143、143 和 153 种亲水代谢物。文献调研, 这些样本以前没有报告过此类数据。因此本研究所开发的定量结构-保留时间模型 (QSRR) 模型可以进一步扩大代谢物的覆盖范围, 从而为扩展的代谢组提供定量分析。

3.2 结论

本研究建立了己胺作为离子对试剂的 IP-RPLC-MS 方法, 优化了离子对色谱工作最佳条件, 最终确定了一种适用于大规模亲水性代谢物高覆盖分析的 IP-RPLC-MS 方法。我们使用 180 多种标准品进行保留时间验证, 发现了代谢物在离子对色谱上的保留时间规律: 代谢物的保留与分子净电荷的多少呈现正相关。氨基酸作为两性分子保留较弱, 而氨基被酰基化和磷酸化修饰后保留行为大大增强; 多元羧酸, 磷酸代谢物保留优异, 且保留时间与其携带的负电荷数量成正相关。这些实验现象可以推导出代谢物和己胺的静电相互作用可能在离子对色谱保留中起到了重要作用。因此, 我们利用量子化学计算的方法在理论水平上验证了这一现象背后的原理。分子的静电势最小值和保留时间呈现强的正相关。我们基于此提出利用分子结构参数预测保留时间的策略。考虑到计算经济性和准确性度, 我们利用分子的半经验参数—分子描述符构建了定量结构-保留时间关系模型 (QSRR)。发现利用分子的 61 个结构参数构建的随机森林回归模型具有良好的预测精度。预测能力达到平均绝对误差在 1 min 以内。这项工作为无标准品代谢物鉴定提供了有力工具。

参考文献:

- [1]Hu Q, Sun Y, Yuan P, Lei H, Zhong H, Wang Y, Tang H. Quantitative structure-retention relationship for reliable metabolite identification and quantification in metabolomics using ion-pair reversed-phase chromatography coupled with tandem mass spectrometry. *Talanta*. 2022 Feb 1;238(Pt 2):123059.