



不同相对湿度下药物/聚合物复杂系统相行为的模型研究及机器学习预测

樊钦习, 丁叶薇, 宋昱潼, 吴昊旻, 吉远辉*

东南大学化学化工学院, 南京 211100

*通讯作者, E-mail: yuanhui.ji@seu.edu.cn

收稿日期: 2024-06-30; 接受日期: 2024-09-14; 网络版发表日期: 2024-10-21

国家自然科学基金项目(编号: 22278070, 21978047)资助项目

摘要 研究不同相对湿度下药物/聚合物复杂系统的相行为对药物制剂的理性设计、制备与储存条件选择的至关重要性。本文首先通过构建汽-液、液-固多相平衡热力学模型和Gordon-Taylor方程系统研究了药物/聚合物在不同相对湿度(RH = 0%、60%、75%)下的热力学相行为和玻璃化转变温度。进一步将微扰链统计缔合流体理论模型参数和分子结构描述符与五种不同的机器学习算法相耦合, 发展了不同相对湿度下药物/聚合物复杂系统相图预测的新方法。研究结果表明, 与热力学模型参数和分子结构描述符相耦合的不同机器学习方法中, 随机森林算法性能最佳, 不同相对湿度下药物在聚合物中的溶解度和玻璃化转变温度曲线的预测结果决定系数 R^2 分别为0.970和0.992。研究进一步表明, 机器学习算法耦合机制模型参数以及分子结构描述符可实现不同相对湿度下药物/聚合物相图的精准预测, 有望为药物制剂的理性设计、制备与储存条件的筛选提供有效指导。

关键词 无定形固体分散体, 相对湿度, 热力学相图, 机器学习, 相平衡

1 引言

活性药物成分(active pharmaceutical ingredient, API)的低水溶性常导致口服生物利用度低, 是制约药物制剂开发的关键难题。无定形固体分散体(amorphous solid dispersion, ASD)是改善活性药物成分的低口服生物利用度的有效策略之一^[1-3]。其中, 无定形固体分散体的长期物理稳定性是其质量评价的重要标准, 然而稳定性实验至少需要耗费三至六个月^[4], 若通过粉末X射线衍射(powder X-ray diffraction, PXRD)和差式扫描量热法(differential scanning calorimetry,

DSC)分析固体的固态性质进行稳定性判断无法达到预测目的, 耗时长且不可控^[5]。因此, 亟需发展高效可靠的ASD长期稳定性预测新方法。

目前已有众多无定形固体分散体稳定性影响因素的研究, 从物理化学的角度可以分为热力学、动力学和环境三个因素, 包括ASD中的药物负载量、ASD的玻璃化转变温度(glass transition temperature, T_g)、环境温度和相对湿度(Relative Humidity, RH)^[6]。在不同相对湿度下由玻璃化转变温度曲线和溶解度曲线构成的相图综合了热力学、动力学和环境三种因素成为研究无定形固体分散体物理稳定性的一种有效方法。如

引用格式: Fan Q, Ding Y, Song Y, Wu H, Ji Y. Model study and machine learning prediction of phase behavior of drug/polymer complex systems under different relative humidity. *Sci Sin Chim*, 2024, 54: 2273–2281, doi: [10.1360/SSC-2024-0150](https://doi.org/10.1360/SSC-2024-0150)

图1所示, 实线代表药物在聚合物中的饱和溶解度, 虚线代表无定形制剂的玻璃化转变温度。溶解度曲线上方区域为热力学稳定区域, 当配方中药物的浓度超过溶解度时, 无定形药物将发生结晶^[7]。药物负载量高于其溶解度的制剂在热力学上是不稳定的, 然而, 当该制剂处于玻璃化转变温度曲线下区域时, 其分子运动较慢, 能够在较长时间内维持无定形态, 但最终仍会发生重结晶。因此, 此类相图可以为无定形固体分散体药物制剂配方设计及后期储存条件(温度、相对湿度)的选择提供有效的指导。

构建ASD在不同相对湿度下的相图需要考虑汽液固三相的平衡: 由聚合物、无定形API和吸附水组成的液相, 晶体API固相和水蒸汽(汽相)。在实验中通常借助DSC测定药物在聚合物中的溶解度以及药物/聚合物的玻璃化转变温度, 然而这种手段无法考虑环境温度和相对湿度的影响, 因此, 难以通过实验手段获取不同相对湿度下的相图。当前已经开发了多种基于热力学的方法预测药物在相关储存条件下的溶解度。通常使用的模型是Flory-Huggins理论^[8], 但其在一些系统上的拟合缺乏适用性^[9]。微扰链统计缔合流体理论(perturbed-chain statistical associating fluid theory, PC-SAFT)在药物/聚合物体系的建模方面表现出良好的能力。Prudic等人^[10]在研究相对湿度对API/聚合物相行为的影响时, 基于PC-SAFT模型建立了汽-液相平衡和固-液相平衡模型, 预测了不同相对湿度下API/聚合物热力学的相图, 预测结果与PXRD测量的实验稳定性结果一致。PC-SAFT模型参数常通过实验数据回归获得^[11], 其中纯组分参数通过回归蒸汽压和液体密度获得, 二元交互参数通过回归相平衡数据而获得^[12]。

机器学习作为一种数据驱动的方法, 用于从已知的训练数据集识别复杂关系, 其卓越的建模和预测能力在药物制剂领域发挥了重要作用, 例如药物发现和设计^[13]、临床前研究^[14]等。课题组前期已研究了药物在聚合物中的溶解度, 然而没有考虑到相对湿度这一因素的影响^[15]。因此, 本研究旨在将机器学习算法与热力学机制相耦合, 开发一种预测药物/聚合物在不同相对湿度下相图的新方法。首先建立汽-液、固-液多相平衡热力学模型, 并结合Gordon-Taylor方程构建药物/聚合物在不同相对湿度(RH = 0%、60%、75%)下的相图, 通过模型预测扩充不同相对湿度下药物/聚合物相行为的数据集。进一步地, 基于本文构建的热力学

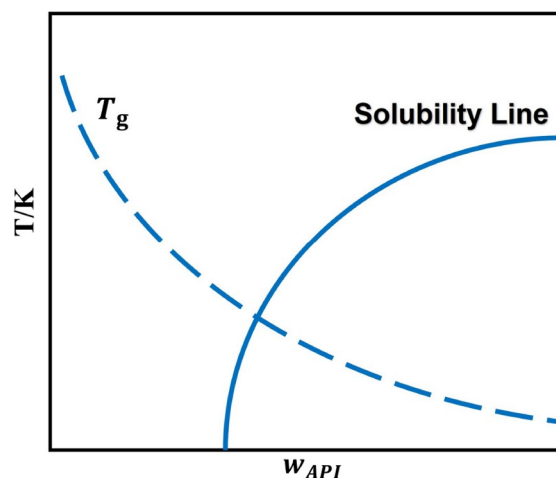


图1 (网络版彩图)药物/聚合物相图示意图, 实线表示溶解度曲线, 虚线表示玻璃化转变温度曲线

Figure 1 (Color online) Schematic of API/polymer phase diagram; the solid line denotes the solubility curve and the dashed line denotes the glass transition temperature curve.

模型所扩充的数据集(7种药物)和课题组前期工作及文献中报道的数据集(8种药物), 使用不同的机器学习算法耦合热力学机制相关参数和分子结构描述符对数据集进行训练。最后, 训练好的机器学习算法用于预测非洛地平 and 氟比洛芬两种药物的相图。研究表明, 本文提出的机器学习算法耦合机制模型参数以及分子结构描述符新方法可实现对不同相对湿度下药物/聚合物相图的精准预测, 有望为药物制剂的理性设计、制备与储存条件的筛选提供有效指导。

2 理论模型和机器学习

2.1 固-液相平衡

对无定形固体分散体建模而言, 聚合物辅料被视为溶剂, 药物被视为溶质。采用固-液相平衡法计算药物在聚合物辅料中的溶解度, 计算公式如式(1)所示, 溶解度与温度有关^[16]。

$$x_{API}^L = \frac{1}{\gamma_{API}^L} \exp \left[\frac{-\frac{\Delta h_{API}^{SL}}{RT} \left(1 - \frac{T}{T_{API}^{SL}} \right) - \frac{\Delta C_{P,API}^{SL}}{R} \left(\ln \left(\frac{T_{API}^{SL}}{T} \right) - \frac{T_{API}^{SL}}{T} + 1 \right)}{1} \right] \quad (1)$$

式中, x_{API}^L 、 R 和 T 分别表示药物以摩尔分数计的溶解

度、理想气体常数(J/mol/K)和温度(K); $T_{\text{API}}^{\text{SL}}$, $\Delta h_{\text{API}}^{\text{SL}}$ 和 $\Delta C_{\text{p,API}}^{\text{SL}}$ 分别表示药物的熔点(K)、熔融焓(J/mol)和药物固相与液相之间的热容差(J/mol/K); $\gamma_{\text{API}}^{\text{L}}$ 是药物在液相中的活度系数。

药物在药物/聚合物中的质量分数可根据式(2)转化为摩尔分数。

$$x_{\text{API}}^{\text{L}} = \frac{C_{\text{API}}/M_{\text{API}}}{C_{\text{API}}/M_{\text{API}} + (1 - C_{\text{API}})/M_{\text{polymer}}} \quad (2)$$

其中, C_{API} 是以质量分数表示药物的溶解度, 单位为g/g, $x_{\text{API}}^{\text{L}}$ 是以摩尔分数表示的药物的溶解度, 单位为mol/mol。

2.2 汽-液相平衡

无定形固体分散体储存在潮湿的环境中, 亲水性聚合物会吸收环境中的水分, 从而影响溶解度曲线。因此, 还需考虑汽-液相平衡。假设汽相仅由水蒸汽组成, 液相包括无定形药物、聚合物和水, 固相为晶体药物。在热力学相平衡中, 液相中水的化学势等于汽相中水蒸汽的化学势。相对湿度被定义为在相同温度下, 汽相中的水蒸汽的分压与纯水饱和蒸汽压之比。根据汽-液相平衡, 相对湿度计算如式(3)所示^[7]。

$$\text{RH} = \frac{P_{\text{H}_2\text{O}}}{P_{\text{H}_2\text{O}}^{\text{LV}}} \times 100\% = x_{\text{H}_2\text{O}}^{\text{L}} \times \gamma_{\text{H}_2\text{O}}^{\text{L}} \times 100\% \quad (3)$$

其中, $P_{\text{H}_2\text{O}}$ 和 $P_{\text{H}_2\text{O}}^{\text{LV}}$ 分别表示水在汽相中的分压和纯水饱和蒸汽压, $x_{\text{H}_2\text{O}}^{\text{L}}$ 和 $\gamma_{\text{H}_2\text{O}}^{\text{L}}$ 分别表示水在液相中的摩尔分数和活度系数, 活度系数通过PC-SAFT状态方程计算得到。

2.3 PC-SAFT状态方程

Gross和Sadowski基于共聚物模型提出的微扰链统计缔合流体理论^[17]已被广泛用于电解质^[18]、聚合物^[19]、超临界流体^[20]和离子液体^[21]等体系。PC-SAFT模型基于统计力学, 通过微扰理论的修正, 成功地建立了流体宏观性质与分子微观参数之间的关系。在PC-SAFT状态方程中, 分子被描述成链段数为 m_i^{seg} , 链段直径为 σ_i 的链段。剩余亥姆霍兹能(a^{res})被表示为链段分子与链段分子之间的硬链斥力(a^{hc}), 范德华力(a^{disp}), 氢键相作用(a^{assoc})^[22]等的贡献之和, 如式(4)所

示。链段之间范德华力(a^{disp})的贡献由色散能参数(u_i/k_{B})描述, k_{B} 为玻尔兹曼常数。对于缔合分子, 根据分子氢键供体和受体定义了分子缔合位点(N_i^{assoc}), 缔合能量参数($\varepsilon^{A_i B_i}/k_{\text{B}}$)和缔合体积参数($k^{A_i B_i}$)。

$$a^{\text{res}} = a^{\text{hc}} + a^{\text{disp}} + a^{\text{assoc}} \quad (4)$$

Bertelot-Lorentz混合规则用于描述混合体系中组分*i*和*j*之间的相互作用, 其中, 二元相互作用参数(k_{ij})用于校正组分之间的色散能参数^[23], 如式(5)~(7)所示。

$$\sigma_{ij} = 0.5 \times (\sigma_i + \sigma_j) \quad (5)$$

$$u_{ij} = (1 - k_{ij}) \times \sqrt{u_i u_j} \quad (6)$$

$$k_{ij} = k_{ij,T} \times T + k_{ij,0} \quad (7)$$

Wolbach和Sandler混合规则用于描述混合体系的缔合能量($\varepsilon^{A_i B_j}$)和缔合体积参数($k^{A_i B_j}$), 如式(8)和(9)所示^[24]。

$$\varepsilon^{A_i B_j} = 0.5 \times (\varepsilon^{A_i B_i} + \varepsilon^{A_j B_j}) \quad (8)$$

$$k^{A_i B_j} = \sqrt{k^{A_i B_i} k^{A_j B_j}} \times \left[\frac{2 \times \sqrt{\sigma_{ii} \sigma_{jj}}}{\sigma_{ii} + \sigma_{jj}} \right]^3 \quad (9)$$

活度系数(γ_i^{L})是混合体系中组分*i*的逸度系数(ϕ_i^{L})与纯组分逸度系数(ϕ_i^{L})之比, 如式(10)所示。同时, 逸度系数(ϕ_i^{L})可通过剩余化学势(μ_i^{res})和压缩因子(*Z*)计算得到, 如式(10)和(11)所示。

$$\gamma_i^{\text{L}} = \phi_i^{\text{L}} / \phi_i^{\text{L}} \quad (10)$$

$$\ln(\phi_i^{\text{L}}) = \mu_i^{\text{res}} / (k_{\text{B}}T) - \ln(Z) \quad (11)$$

此外, 压缩因子(*Z*)和剩余化学势(μ_i^{res})是剩余亥姆霍兹能(a^{res})的函数, 如式(12)和(13)所示。

$$Z = 1 + \rho \left(\partial(a^{\text{res}} / (k_{\text{B}}T)) / \partial \rho \right) \quad (12)$$

$$\frac{\mu_i^{\text{res}}}{k_{\text{B}}T} = \frac{a^{\text{res}}}{k_{\text{B}}T} + Z - 1 + \frac{\partial(a^{\text{res}} / (k_{\text{B}}T))}{\partial x_i} - \sum_{j=1}^N \left[x_j \left(\frac{\partial(a^{\text{res}} / (k_{\text{B}}T))}{\partial x_j} \right) \right] \quad (13)$$

2.4 Gordon-Taylor方程

使用Gordon-Taylor方程预测药物/聚合物/水三元

体系的玻璃化转变温度 $T_{g,ter}$. 该方程使用质量分数对药物、聚合物和水的玻璃化转变温度进行加权^[10].

$$T_{g,ter} = \frac{w_{H_2O} \cdot T_{g,H_2O} + K_1 \cdot w_{API} \cdot T_{g,API} + K_2 \cdot w_{polymer} \cdot T_{g,polymer}}{w_{H_2O} + K_1 \cdot w_{API} + K_2 \cdot w_{polymer}} \quad (14)$$

式中, $T_{g,ter}$ 是三元体系的玻璃化转化温度(K); T_{g,H_2O} 、 $T_{g,API}$ 和 $T_{g,polymer}$ 分别是水、药物和聚合物的玻璃化转变温度, w_{H_2O} 、 w_{API} 和 $w_{polymer}$ 分别是水、药物和聚合物的质量分数; 参数 K_1 、 K_2 分别由式(15)和(16)计算求得^[25].

$$K_1 \approx \frac{\rho_{H_2O} \cdot T_{g,H_2O}}{\rho_{API} \cdot T_{g,API}} \quad (15)$$

$$K_2 \approx \frac{\rho_{H_2O} \cdot T_{g,H_2O}}{\rho_{polymer} \cdot T_{g,polymer}} \quad (16)$$

式中, ρ_{H_2O} 、 ρ_{API} 、 $\rho_{polymer}$ 分别表示水、药物和聚合物的密度(g/cm³).

2.5 机器学习

基于2.1~2.4中建立的汽-液、固-液多相平衡热力学模型系统研究了7种药物: 别嘌呤醇(Allopurinol, ALLO), 美沙拉嗪(Mesalazine, MES), 奥卡西平(Oxcarbazepine, OXCAR), 依托度酸(Etodolac, ETOD), 丙磺舒(Probenecid, PBD), 非洛地平(Felodipine, FEL)和氟比洛芬(Flurbiprofen, FBU)与聚合物在不同相对湿度(RH = 0%、60%、75%)下的相行为, 模型预测结果用于扩充不同相对湿度下药物/聚合物相行为的数据集. 模型扩充数据集结合课题组前期工作及文献中报道的8种药物(萘普生、吲哚美辛、伊曲康唑、灰黄霉素、对乙酰氨基酚、布洛芬、厄贝沙坦、奥沙普秦)的相关数据集, 获得包含15种药物和9种聚合物的3992组溶解度数据集和包含9种药物和7种聚合物的2354组玻璃化转变温度数据集, 数据集包含的药物和聚合物见表S1. 对两个数据集分别进行训练和测试, 训练集和测试集的数据均以7:3的比例划分, 训练好的模型进一步用于预测非洛地平 and 氟比洛芬与聚合物在不同相对湿度下的相图. 本文使用到的机器学习算法包括决策树(decision tree, DT)、随机森林(random forest, RF)、支持向量机(support vector machine, SVM)、k-最近邻(k nearest neighbor, kNN)、多层感知机(multi-layer per-

ceptron, MLP), 评价指标为决定系数(R -square, R^2)和均方根偏差(root mean square error, RMSE).

基于SMILES式表示药物分子和聚合物单体, 在Tayyebi等人^[26]研究的基础上, 使用mordred包生成1613个二维分子描述符, 忽略三维分子描述符以提高预测速度. 修剪分类变量并使用0.1的优化阈值进行相关性过滤, 筛除掉具有低方差、低相关的特征. 随后, 通过描述符相关性矩阵分析消除高度相关的描述符, 得到98个分子结构描述符. 进一步地, 引入PC-SAFT热力学模型参数, 探究机器学习算法耦合分子结构描述符及热力学机制的相图预测能力, 提高模型的可解释性, 同时比较耦合PC-SAFT相关分子结构参数(链段数、链段直径)和相互作用参数(色散能、缔合能、缔合体积和缔合位点数)对机器学习算法性能的提升能力. 进一步, 采用Shapley additive explanations (SHAP)^[27]进行分析, 研究分子结构、热力学机制与相行为之间的关系, 从而增强模型的可解释性.

3 结果与讨论

3.1 不同相对湿度下药物/聚合物相行为的研究及相图构建

本文基于热力学模型系统研究了7种药物(别嘌呤醇, 美沙拉嗪, 奥卡西平, 依托度酸, 丙磺舒, 非洛地平 and 氟比洛芬)与聚合物在不同相对湿度(RH = 0%、60%、75%)下的相行为, 其中别嘌呤醇、美沙拉嗪、奥卡西平、依托度酸、丙磺舒5种药物与2种聚合物(聚乙二醇(polyethylene glycol, PEG 6000)和聚乙烯吡咯烷酮(polyvinyl pyrrolidone, PVP K25))的模型预测结果如图2所示, 非洛地平和氟比洛芬的模型预测结果见图S1. 本文模型研究涉及的药物及聚合物的理化参数、PC-SAFT模型参数及二元相互作用参数均来自于文献^[16,28~30], 见表S2~S4.

图2显示五种药物在不同聚合物中的溶解度曲线. 由图2可见, 药物在聚合物中的溶解度随着温度的升高而增加. 在别嘌呤醇、美沙拉嗪和依托度酸体系中, 药物在聚合物中的溶解度遵循PVP K25 > PEG 6000的规律, 与Prudic等人^[31]研究磺酰胺在聚合物中溶解度的结果一致, 药物在PVP中的溶解度高于PEG, 这可能是由于所选药物与PVP中酰胺基团的相互作用强于与PEG中羟基的相互作用. 对于奥卡西平, 其在PEG 6000

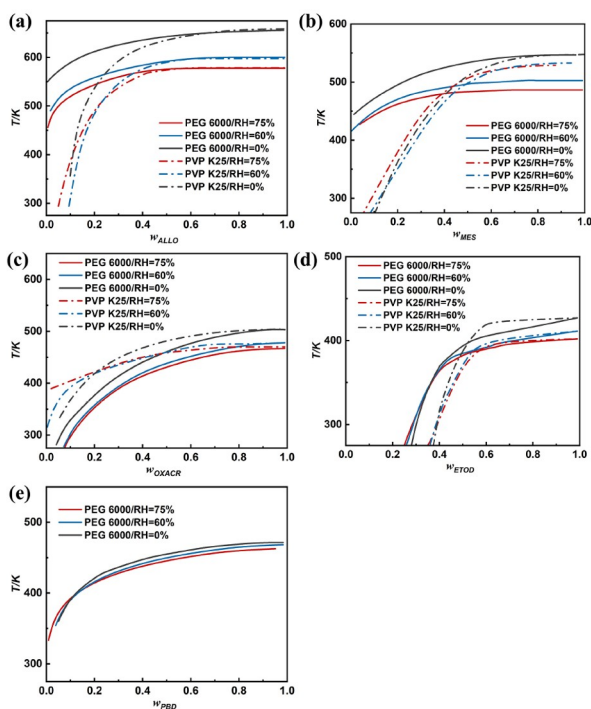


图2 (网络版彩图)别嘌醇(a), 美沙拉嗪(b), 奥卡西平(c), 依托度酸(d), 丙磺舒(e)在RH = 0%(黑色)、60%(蓝色)、75%(红色)时的溶解度曲线, 其中实线表示API/PEG 6000, 虚线表示含API/PVP K25

Figure 2 (Color online) Solubility curves of allopurinol (a), mesalazine (b), oxcarbazepine (c), etodolac (d), and probenecid (e), at RH = 0% (black), 60% (blue), 75% (red), where the solid lines denote API/PEG 6000 and the dash-dot lines denote API/PVP K25.

中的溶解度大于PVP K25, 表现出与上述三种药物不一致的规律, 这可能是药物分子结构的差异造成的。

溶解度曲线上方区域为热力学稳定区域, 位于该区域的无定形固体分散体可以保持稳定, 而位于溶解度曲线下方区域的无定形分散体倾向于重结晶。图2中, ALLO/PEG 6000、MES/PEG 6000、OXCAR/PEG 6000、OXCAR/PVP K25、PBD/PEG 6000的热力学稳定区域温度高于常温(25℃), 因此难以在常温下形成稳定的无定形固体分散体, ALLO/PVP K25、MES/PVP K25可以在药物负载量低的情况下形成稳定的无定形固体分散体, ETOD/PEG 6000、ETOD/PVP K25较容易形成稳定的无定形固体分散体。与Prudic等人^[10]研究的吡罗美辛和萘普生在聚合物中的溶解随相对湿度变化趋势相似, ALLO/PVP K25、MES/PVP K25、ETOD/PEG 6000和ETOD/PVP K25体系在常温下随着相对湿度的增加, 药物在聚合物中的溶解度减

小。聚合物的吸水性导致难溶性药物在聚合物中的溶解度随相对湿度的增加而降低, 同时, 聚合物吸水能力的差异导致了相对湿度对药物在聚合物中溶解度影响程度的不同。

如图S1所示, 随着相对湿度的增加, 药物/聚合物的玻璃化转变温度降低, 与Prudic等人^[10]探究的相对湿度对萘普生、吡罗美辛相图的影响相符。体系中所选用的聚合物均为亲水性聚合物, 具有吸水性, 在相对湿度增加的情况下, 聚合物吸收的水随之增加, 而水具有极低的玻璃化转变温度(136 K)。

综上所述, 相对湿度对药物在聚合物中的溶解度、药物/聚合物玻璃化转变温度有一定的影响, 进而影响无定形固体分散体制剂的长期稳定性。不同相对湿度下药物/聚合物的相图可以帮助指导无定形固体分散体制剂生产时的配方设计以及后续储存条件筛选。在经济条件允许的情况下, 无定形固体分散体制剂应储存于相对湿度较低的环境中。

3.2 机器学习

3.2.1 机器学习算法预测结果

不考虑PC-SAFT热力学机制参数, 仅采用决策树、随机森林、支持向量机、k-最近邻和多层感知机五种机器学习算法耦合mordred包生成的分子结构描述符训练和测试的结果如表1和2所示, 机器学习模型的关键参数见表S5。

由表1和2可知, 机器学习算法耦合mordred包生成的分子结构描述符在不同相对湿度下药物在聚合物中的溶解度以及药物/聚合物的玻璃化转变温度上均表现出良好的预测能力, 其中随机森林算法预测能力最佳, 测试集的 R^2 分别为0.920和0.959。

将考虑热力学机制的PC-SAFT模型参数进一步耦合进机器学习算法中, 结果如表3和4所示。除预测药物/聚合物玻璃化转变温度的k-最近邻算法, PC-SAFT热力学机制参数的耦合在一定程度上提高了机器学习算法的预测能力, 随机森林算法在该计算中仍为最优算法。随机森林算法中, 药物在不同相对湿度下聚合物中溶解度的测试集 R^2 从0.920增加到0.970, 药物/聚合物在不同相对湿度下玻璃化转变温度的测试集 R^2 从0.959增加到0.992。

图3a, b散点图显示了随机森林算法的相图预测值与模型计算值之间的关系。随机森林算法的预测值与

表 1 机器学习预测溶解度的精度(未耦合PC-SAFT参数)**Table 1** The accuracy of machine learning for the solubility dataset (without the PC-SAFT parameters)

机器学习算法	溶解度-训练集		溶解度-测试集	
	R^2	RMSE	R^2	RMSE
决策树	0.875	0.100	0.835	0.114
随机森林	0.946	0.0658	0.920	0.0806
支持向量机	0.903	0.312	0.660	0.583
k-最近邻	0.806	0.126	0.805	0.123
多层感知机	0.850	0.388	0.838	0.403

表 2 机器学习预测玻璃化转变温度的精度(未耦合PC-SAFT参数)**Table 2** The accuracy of machine learning for the glass transition temperature dataset (without the PC-SAFT parameters)

机器学习算法	玻璃化转变温度-训练集		玻璃化转变温度-测试集	
	R^2	RMSE	R^2	RMSE
决策树	0.855	14.7	0.825	15.7
随机森林	0.995	2.53	0.959	7.52
支持向量机	0.913	0.295	0.757	0.493
k-最近邻	0.755	19.3	0.692	20.1
多层感知机	0.958	0.205	0.952	0.218

表 3 机器学习预测溶解度的精度(耦合PC-SAFT参数)**Table 3** The accuracy of machine learning for the solubility dataset (coupling the PC-SAFT parameters)

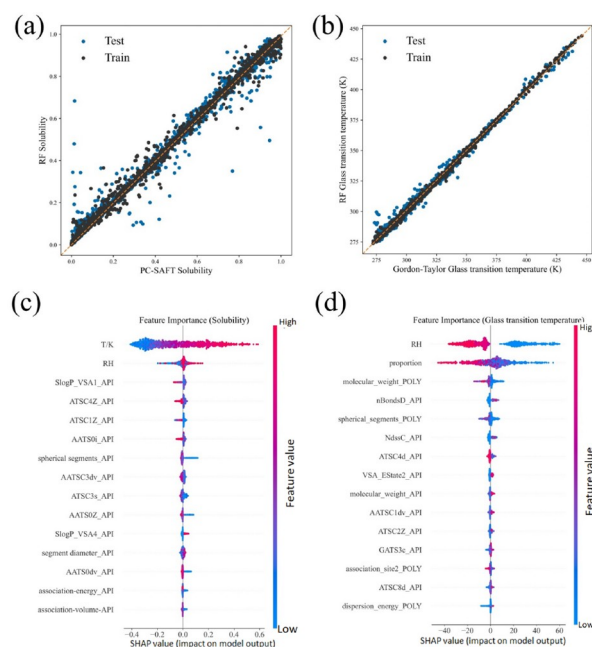
机器学习算法	溶解度-训练集		溶解度-测试集	
	R^2	RMSE	R^2	RMSE
决策树	0.894	0.0919	0.877	0.100
随机森林	0.996	0.0189	0.970	0.0483
支持向量机	0.915	0.291	0.881	0.344
k-最近邻	0.850	0.110	0.858	0.107
多层感知机	0.871	0.359	0.861	0.373

理论模型值表现高度一致, 表明本文所提出的随机森林耦合热力学机制和分子结构描述符的方法在该领域具有可靠的预测能力。

利用SHAP分析研究分子描述符对不同相对湿度下药物/聚合物相图影响的重要性排名, 结果如图3c, d所示。该图展示了不同特征对预测结果影响的大小和方向, 每个点代表不同的样本, 颜色代表SHAP值的大小, SHAP值由高到低可视化为深粉到浅蓝的渐变。从图中可以看出, 温度是影响药物在聚合物中溶解度的

表 4 机器学习预测玻璃化转变温度的精度(耦合PC-SAFT参数)**Table 4** The accuracy of machine learning for the glass transition temperature dataset (coupling the PC-SAFT parameters)

机器学习算法	玻璃化转变温度-训练集		玻璃化转变温度-测试集	
	R^2	RMSE	R^2	RMSE
决策树	0.870	13.6	0.840	15.7
随机森林	0.998	1.53	0.992	3.32
支持向量机	0.993	0.0814	0.976	0.155
k-最近邻	0.688	21.49	0.599	23.9
多层感知机	0.965	0.187	0.952	0.219

**图 3** (网络版彩图)随机森林算法对药物在聚合物中溶解度(a, c)和药物/聚合物玻璃化转变温度(b, d)预测结果及特征重要性排名**Figure 3** (Color online) Random Forest algorithm for drug solubility in polymers (a, c) and API/polymer glass transition temperature (b, d) predicted results and feature importance ranking.

重要因素, 且对输出结果的影响是正向的; 相对湿度和药物占比对药物/聚合物玻璃化转变温度呈现较强负向影响。除此之外, 影响药物/聚合物相图的前15个分子描述中还包括PC-SAFT热力学机制相关参数, 如药物的链段数、链段尺寸与缔合能、聚合物的链段尺寸, 以及与药物亲脂性^[32]、静电相互作用^[32]、原子或分子性质与空间相互作用的强度相关^[33]的描述符, 分子描述符详细解释见表S6。

3.2.2 随机森林算法耦合热力学模型参数预测相图

训练好的随机森林算法预测不同相对湿度下(RH = 0%、60%、75%) FEL/PVPVA64、FEL/PVAC、FBU/PVPVA46、FBU/PVP K30的相图, 结果如图4所示, 模型预测结果见图S1. 结果表明, 随机森林模型成功预测了FEL/PVPVA64、FEL/PVAC、FBU/PVPVA46、FBU/PVP K30在不同相对湿度(RH = 0%、60%、75%)下溶解度和玻璃化转变温度. 目前, 本研究数据集中API和聚合物种类较少, 在后续研究中需不断扩充数据集.

4 结论

本文研究了机器学习耦合热力学机制模型参数及分子结构描述符预测不同相对湿度下(RH = 0%、60%、75%)药物/聚合物相图. PC-SAFT热力学机制相关参数耦合入机器学习算法以增强其可解释性, 同时, 模型建立的相图数据也为机器学习提供数据集扩充. 研究表明, 与热力学模型参数和分子结构描述符相耦合的不同机器学习方法中, 随机森林算法性能最佳, 不同相对湿度下, 药物在聚合物中的溶解度和玻璃化转变温度曲线测试集的 R^2 分别为0.970和0.992. 基于本文的研究结果, 将机器学习算法耦合机制模型参数以及分子结构描述符可实现不同相对湿度下药物/

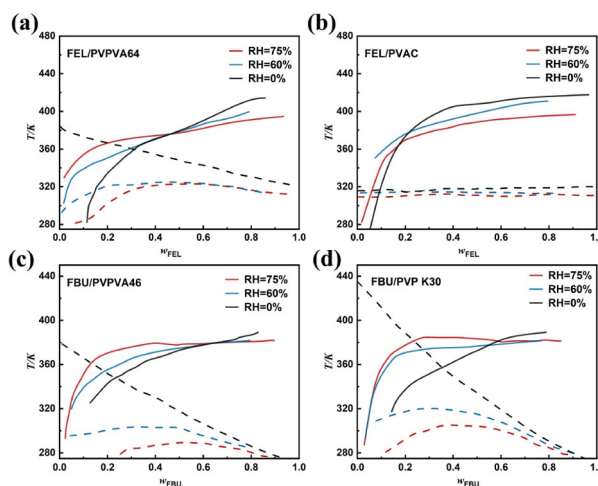


图4 (网络版彩图)随机森林预测不同相对湿度下的相图: FEL/PVPVA 64 (a)、FEL/PVAC (b)、FBU/PVPVA 46 (c)、FBU/PVP K30 (d), 黑色、蓝色和红色实线为RH = 0%、60%、75% 时药物在聚合物中的溶解度, 黑色、蓝色和红色虚线为RH = 0%、60%、75% 时药物/聚合物的玻璃化转变温度

Figure 4 (Color online) Random forest prediction of the phase diagrams under different RH: FEL/PVPVA64 (a), FEL/PVAC (b), FBU/PVPVA46 (c), FBU/PVP K30 (d). The solid lines in black, blue and red denote the solubility of the API in the polymer at RH = 0%, 60% and 75%, and the dashed lines in black, blue and red denote the glass transition temperature of the API/polymer at RH = 0%, 60% and 75%.

聚合物相图的精准预测, 有望为药物制剂的理性设计、制备与储存条件的选择提供有效指导.

致谢 袁权先生深耕化工领域六十余年, 始终以国家需求为己任, 潜心科学研究, 攻坚克难, 视名利淡如水, 将全部智慧和毕生精力奉献给了我国的化工科研事业. 一心为国任甘苦, 劲步科研遍地芳, 谨以此文致敬袁权先生, 我辈当满怀思念, 砥砺前行.

补充材料

本文的补充材料见网络版chemcn.scichina.com. 补充材料为作者提供的原始数据, 作者对其学术质量和内容负责.

参考文献

- Vasconcelos T, Sarmento B, Costa P. *Drug Discov Today*, 2007, 12: 1068–1075
- Altamimi MA, Elzayat EM, Qamar W, Alshehri SM, Sherif AY, Haq N, Shakeel F. *Saudi Pharm J*, 2019, 27: 629–636
- Krstić M, Manić L, Martić N, Vasiljević D, Mračević SĐ, Vukmirović S, Rašković A. *Eur J Pharm Sci*, 2020, 150: 105343
- Han R, Xiong H, Ye Z, Yang Y, Huang T, Jing Q, Lu J, Pan H, Ren F, Ouyang D. *J Control Release*, 2019, 311–312: 16–25
- Dedroog S, Pas T, Vergauwen B, Huygens C, Van den Mooter G. *J Pharm BioMed Anal*, 2020, 178: 112937
- Lin X, Hu Y, Liu L, Su L, Li N, Yu J, Tang B, Yang Z. *Pharm Res*, 2018, 35: 125
- Chen Q, Ji Y. *Ind Eng Chem Res*, 2023, 62: 1596–1605

- 8 Flory PJ. *J Chem Phys*, 1942, 10: 51–61
- 9 Lehmkomper K, Kyeremateng SO, Heinzerling O, Degenhardt M, Sadowski G. *Mol Pharm*, 2017, 14: 4374–4386
- 10 Prudic A, Ji Y, Luebbert C, Sadowski G. *Eur J Pharm BioPharm*, 2015, 94: 352–362
- 11 Klajmon M. *Mol Pharm*, 2022, 19: 4212–4232
- 12 Mejbri K, Taieb A, Bellagi A. *J SuperCrit Fluids*, 2015, 104: 132–144
- 13 Bai QF, Tan SY, Xu TY, Liu HX, Huang JZ, Yao XJ. *Briefings Bioinf*, 2021, 22: 161
- 14 Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, Li B, Madabhushi A, Shah P, Spitzer M, Zhao S. *Nat Rev Drug Discov*, 2019, 18: 463–477
- 15 Ge K, Huang Y, Ji Y. *Chin J Chem Eng*, 2024, 66: 263–272
- 16 Chen Q, Ji Y, Ge K. *Chem Eng Sci*, 2021, 244: 116798
- 17 Gross J, Sadowski G. *Ind Eng Chem Res*, 2001, 40: 1244–1260
- 18 Sun M, Cai J, Shahriari R. *J Solution Chem*, 2022, 51: 1229–1246
- 19 Cheng L, Fan W, Ji Y. *Fluid Phase Equilib*, 2023, 572: 113848
- 20 Shahsavan Markadeh R, Arabkhalaj A, Birouk M. *Fluid Phase Equilib*, 2022, 562: 113560
- 21 Shekaari H, Zafarani-Moattar MT, Faraji S, Mokhtarpour M. *Fluid Phase Equilib*, 2020, 506: 112320
- 22 Dohrn S, Luebbert C, Lehmkomper K, Kyeremateng SO, Degenhardt M, Sadowski G. *Int J Pharm*, 2020, 577: 119065
- 23 Ji YH, Paus R, Prudic A, Luebbert C, Sadowski G. *Pharm Res*, 2015, 32: 2559–2578
- 24 Wolbach JP, Sandler SI. *Ind Eng Chem Res*, 1998, 37: 2917–2928
- 25 Simha R, Boyer RF. *J Chem Phys*, 1962, 37: 1003–1007
- 26 Tayyebi A, Alshami AS, Rabiei Z, Yu X, Ismail N, Talukder MJ, Power J. *J Cheminform*, 2023, 15: 99
- 27 Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. In: Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS). Long Beach, CA, 2017
- 28 Zhang M, Ge K, Chen Y, Ji Y. *Ind Eng Chem Res*, 2023, 63: 731–742
- 29 Fan Q, Zhang M, Ding Y, Victorov AI, Ji Y. *Fluid Phase Equilib*, 2024
- 30 Ge K, Ji Y. *Ind Eng Chem Res*, 2021, 60: 9259–9268
- 31 Prudic A, Ji Y, Sadowski G. *Mol Pharm*, 2014, 11: 2294–2304
- 32 Rzycki M, Gładysiewicz-Kudrawiec M, Kraszewski S. *Sci Rep*, 2024, 14: 4641
- 33 Yue H, Wang JT, Lu MJ. *Crystal Growth & Design*, 2023, 23: 2540–2549

Model study and machine learning prediction of phase behavior of drug/polymer complex systems under different relative humidity

Qinxi Fan, Yewei Ding, Yutong Song, Haomin Wu, Yuanhui Ji*

School of Chemistry and Chemical Engineering, Southeast University, Nanjing 211100, China

**Corresponding author (email: yuanhui.ji@seu.edu.cn)*

Abstract: To rationally design, prepare, and store pharmacological formulations, it is crucial to understand the phase behavior of complicated drug/polymer complexes under varying relative humidity conditions. By creating thermodynamic models of vapor-liquid and liquid-solid multiphase equilibria and Gordon-Taylor equation systems, the thermodynamic phase behavior and glass transition temperature of drug/polymer at different relative humidity (RH = 0%, 60%, and 75%) were studied in this paper. Combining five distinct machine learning algorithms with the perturbed-chain statistical associating fluid theory model parameters and molecular structure descriptors creates a novel approach for predicting the phase diagram of drug/polymer complex systems under different RH. The results demonstrate that the random forest algorithm has the highest prediction accuracy among machine learning techniques combined with thermodynamic model parameters and molecular structure descriptors. The determination coefficients R^2 of the test set results of drug solubility in polymer and glass transition temperature curves at different RH were 0.970 and 0.992, respectively. The study further shows that the machine learning algorithm coupled mechanism model parameters and molecular structure descriptors can achieve accurate prediction of drug/polymer phase diagrams under different RH, which is expected to provide effective guidance for rational design of drug preparations, preparation and storage conditions screening.

Keywords: amorphous solid dispersion, relative humidity, thermodynamic phase diagram, machine learning, phase equilibrium

doi: [10.1360/SSC-2024-0150](https://doi.org/10.1360/SSC-2024-0150)