使用分子动力学模拟和卷积神经网络快速预测液相酸催化反应速率

Alex K.Chew, SHengli Jiang, Weiqi Zhang, Victor M. Zavala, Reid C. Van Lehn

摘要：用于将生物质制造为高价值化学品的液相酸催化反应的反应速率对溶剂的成分非常敏感。而确定合适的混合溶剂在理论和实验上都是较为困难的。我们研究表明，三维卷积神经网络（3D-CNN）可以利用经典的分子动力学模拟数据生成的反应物-溶剂环境复杂构型，来准确预测生物质的Brønsted酸催化反应速率。我们开发了一种三维卷积神经网络（称为SolventNet），使用七种生物质衍生氧化物在水-共溶剂体系内的反应速率实验值和相关的分子动力学模拟数据来训练模型，以达到预测酸催化反应速率的目的。研究表明，SolventNet可以预测更多的反应物和溶剂系统的反应速率，比之前的模拟方法快一个数量级。这种将机器学习和分子动力学模拟结合的方法能快速、高通量地筛选溶剂体系，并且确定改进的生物质转化条件。

1. 引言

木质纤维素生物质得催化转化过程，是从可再生原料中获取交通运输所需燃料、高价值化学品等产品的一种很有前景的方法。生物质衍生分子转化通常由液相、酸催化的反应来促进（如图1.a例），而这些反应经常因为在水中受阻，故需要酸催化。一种加快酸催化反应速率的方法是修改溶剂组成，通过将极性有机共溶剂和水混合，创造混合溶剂环境。这种强化方法相比于在纯水中的速率，可以提高大约100倍。遗憾的事，通过试错实验来确认可以提升反应速率的溶剂环境的时间和金钱消耗都很高，并且对溶剂效应的物理内涵了解相当有限。相对地，计算工具已经被用于了解溶剂对化学反应的影响，并成功指导了溶剂混合物的设计，来实现高效的生物质转化过程。

在过去的十年中，从头算的量子化学方法被用来量化溶剂对生物质转化反应中基元步骤的势垒产生的影响。用从头算的分子动力学模拟(MD)，Mellmer等人发现有机共溶剂可以通过改变酸性质子催化剂、反应物和带电离子周围的溶剂环境，来降低活化能，从而增加生物质转化反应速率。（图1b）。这些模拟显示，亲水反应物会在混合溶剂环境中促进富含水的小区域形成，这些区域会优先溶解酸催化剂，并且稳定随后的碳正离子类似过渡态。这些发现表明，反应物周围水富集的程度和酸催化反应性能直接相关。类似的从头算分子动力学模拟可用于理解溶剂环境如何改变关键酸催化反应的动力学性质，比如将果糖转化为5-羟甲基糠醛（一种用于聚合物前体和交通燃料的中间化学品）的反应。但是，尽管从头算分子动力学方法可以直接探究反应机理的细节，它们的极高计算成本使得筛选多种溶剂变得不可行。

相比从头算分子动力学方法，经典分子动力学模拟可以获取更长时间尺度和更大空间尺度的信息，并且显著降低计算开销，故可以更快速地表征复杂的溶剂环境，甚至是更大的反应物的周围环境。经典分子动力学适用于模拟由反应物-溶剂-共溶剂相互作用相互影响产生的混合溶剂环境中的空间结构。这种环境可能影响反应动力学（例如，由于上述的反应物优先溶解作用），但是这种影响很难被大部分溶剂描述符（例如，介电常数）来捕捉到。从另一方面来说，经典分子动力学模拟的一个重要限制是，它们不能直接建模化学反应。