

机器学习在化学领域中的应用

张思源, 张志成, 李荣金*

天津大学理学院化学系, 有机集成电路教育部重点实验室, 天津市分子光电科学重点实验室, 天津 300072

摘要: 在计算机科学和技术的推动下, 机器学习已成为化学研究的强大工具。本文首先介绍了机器学习的基本概念, 随后探讨了其在化学领域的四个关键应用: 在有机全合成中预测合成路径; 在原子模拟中对势能面进行高效采样和重建; 在异相催化设计中用于反应路径揭示和催化剂筛选; 以及在核磁共振谱学中的信号处理和解析。本文还介绍了移动机器人化学家的概念, 展示了机器学习与自动化技术结合潜力。最后, 本文对机器学习在未来化学研究中的发展前景进行了展望, 指出了其可能带来的变革性影响。

关键词: 机器学习; 深度学习; 机器人化学家

中图分类号: G64; O6-39

Applications of Machine Learning in Chemistry

Siyuan Zhang, Zhicheng Zhang, Rongjin Li*

Key Laboratory of Organic Integrated Circuit, Ministry of Education & Tianjin Key Laboratory of Molecular Optoelectronic Sciences, Department of Chemistry, School of Science, Tianjin University, Tianjin 300072, China.

Abstract: Driven by advancements in computer science and technology, machine learning has emerged as a powerful tool in chemical research. This paper begins by introducing the basic concepts of machine learning, followed by an exploration of its four key applications in the field of chemistry: predicting synthetic pathways in organic total synthesis; conducting efficient sampling and reconstruction of potential energy surfaces in atomic simulations; revealing reaction pathways and screening catalysts in heterogeneous catalysis design; and processing and interpreting signals in nuclear magnetic resonance spectroscopy. Additionally, the concept of the robotic chemist is introduced, illustrating the potential of integrating machine learning with automation technologies. Finally, the paper discusses the future prospects of machine learning in chemical research, highlighting its potential transformative impacts.

Key Words: Machine learning; Deep learning; Robot chemist

随着数字时代的到来, 人工智能(Artificial Intelligence, AI)尤其是机器学习(Machine Learning, ML)已经成为科学探索的强大引擎^[1-3]。在化学领域, AI和ML的应用正在经历一场革命性的发展, 它们不仅改变了传统的化学研究方式, 还为解决复杂的化学问题提供了新的途径^[4,5]。例如, AlphaFold算法在蛋白质结构预测方面的突破性成就^[6], 是机器学习在生物化学领域的巨大成功。

化学作为一门数据密集型学科, 其历史数据的庞大体量^[7]和化学家的深刻洞见^[8]构成了理论体系的基础。然而, 面对机器学习对数据集质量和数量的需求, 化学领域面临的挑战是如何从非全面

性和非平衡性的实验数据中提取有价值的信息。这一挑战要求我们在应用ML技术时采取创新的数据处理和分析策略。

本文概述了机器学习在化学领域的应用，从基础的数据处理和预测方法到高级的应用，如全合成设计、原子模拟、异相催化设计和核磁共振谱解析。我们探讨了如何利用ML算法从化学数据中提取高维特征和模式，以降低研发成本并加速化学发现的进程。此外，我们介绍了AI化学家的概念，这是一种集成了ML算法的自动化化学研究平台，能够自行阅读文献完成研究任务，代表了化学研究的未来方向。在展望部分，本文讨论了机器学习技术在化学领域的未来发展，包括其潜在的技术革新和对化学科学领域的深远影响。同时，我们也指出了ML当前面临的挑战和未来的研究方向，以期为化学领域的研究人员提供指导和启发。

1 机器学习概述

1.1 数据和特征

机器学习的潜力与数据集的质量和所选用的特征紧密相关。数据集中，起到关键作用的变量被称之为“特征(Features)”，也被称为“表述符(Representations)”或“描述符(Descriptors)”。作为机器学习预处理的重要部分，如何从原始数据中抽象出特征对于机器学习的速度、准确性和泛化性至关重要，这被称为“特征工程”。特征工程也包括对缺失值、异常值的识别和处理以及变量的标准化和归一化。

化学数据的特征化是将原始数据转化为模型能够处理的信息的过程。这一步骤至关重要，因为它决定了模型能够捕捉的化学信息的深度和广度^[9]。在化学中，特征通常分为两大类：拓扑信息和坐标信息。拓扑信息描述分子中原子间的连接方式，而坐标信息提供原子在空间中的相对位置。

拓扑信息可以通过多种方式编码，如简化分子线性输入规范(Simplified Molecular Input Line Entry System, SMILES)^[10]、国际化合物标识(International Chemical Identifier, InChI)^[11]和拓展连通性指纹(Extended Connectivity Fingerprints, ECFPs)^[12]等，它们能够以不同的抽象层次捕捉分子的结构信息。为了保证模型的泛化能力和特征的对称性，坐标信息一般会转换为高级构造性特征，例如原子中心对称函数(Atom-centered Symmetry Functions, ACSFs)^[13]和原子位置平滑重叠(Smooth Overlap of Atomic Positions, SOAP)^[14]。

随着机器学习技术的持续进步，自动化特征提取已在化学领域得到应用。这些方法能够从海量数据中自动发掘复杂的特征表示，有助于去除数据中的噪声。然而，这些方法不仅要求对数据有深入的理解，而且某些算法可能会消耗大量的计算资源，并且可能会牺牲模型的可解释性。因此，在选择合适的特征表示方法时，需要综合考虑模型的性能以及实际应用中的可行性。

1.2 机器学习模型和训练

在化学领域，机器学习模型的构建和训练是实现从数据中提取知识和进行预测的关键步骤。通过将化学数据转换为计算机可读的格式，并在选定的特征上进行训练，机器学习模型能够学习数据中的模式，从而对新的输入数据做出准确的预测。如图1所示，机器学习任务通常分为分类(Classification)和回归(Regression)，此外还包括关联规则学习、特征工程和数据降维等技术，这些任务的选择取决于研究目的和模型输出的类型(离散值或连续值)。



图1 机器学习的训练方法和任务

训练机器学习模型涉及多种学习范式, 包括监督学习、无监督学习和强化学习。监督学习在化学数据分析中占据核心地位, 因为化学数据常常具有连续性, 并且研究者通常对特定化学性质(如产率、催化效率、能量和能量梯度、光谱强度等)有明确的预测目标。无监督学习则在探索未知化学现象和发现新物质方面显示出巨大潜力, 它能够帮助研究者在没有明确数据标签的情况下识别数据中的隐藏结构和模式。强化学习则在需要模型做出一系列决策的场景中发挥作用, 如自动化合成路径的探索。

2 机器学习的应用

在机器学习领域, “没有免费午餐”定理(No Free Lunch Theorem)^[15,16]提供了一个重要原则: 没有任何单一算法能够在所有问题和数据集上都表现出最优性能。这一定理强调了算法选择和优化的重要性, 同时也指出了在特定问题上的优化可能会牺牲算法在其他问题上的表现。

在化学领域, 这一原则尤为重要。化学数据的多样性和复杂性^[17]要求我们为不同的问题选择合适的机器学习模型。此外, 机器学习模型的选择和优化过程需要化学专家的先验知识和对数据的深入理解。人类工作者在这一过程中扮演着不可或缺的角色, 他们需要根据问题的特点和数据的特性, 引入先验参数或假设, 并进行模型算法、特征选择和超参数调整。这种基于经验和直觉的决策过程是当前机器学习应用中重要的一部分。因此, 对于不同的任务, 进行专门的模型选择、优化和测试是更为明智的策略, 而不是寻求一个能够解决所有问题的“全能”模型。这要求我们对各种机器学习算法有深入的理解, 并能够根据化学问题的具体需求进行调整。

2.1 全合成设计

全合成设计, 作为有机化学中一项极具挑战性的任务, 不仅对科研工作者的专业知识和技能提出了高要求, 也因其目标导向明确、历史数据丰富和验证方法简便而成为机器学习应用的理想领域。机器学习辅助的全合成设计, 虽未必总能提供最优解, 但已经能够启发实验者的设计思路和提升实验效率。

自20世纪60年代E. J. Corey^[18,19]提出计算机辅助合成设计(Computer-Assisted Synthesis Planning, CASP)以来, 机器学习在全合成设计中的应用逐渐深入^[20-25], 从全局最优化到局部反应条件的精确猜测, 机器学习在化学合成设计中展现出巨大价值。CASP主要分为有模板和无模板两种方法, 有模板CASP依赖于人工生成的反应模板和规则, 如ASKCOS^[26]、Synthia^[20]、RetroPath^[27,28]、RetroBioCat^[29]和AiZynthFinder^[30]; 无模板CASP则直接学习结构间的转化关系, 如IBM RXN^[31,32]和BioNavi-NP^[33]。相较于传统机器学习领域中的图像或文本数据, 化学合成的数据较为有限, 因此综合运用化学见解和反应数据具有明显优势。2022年, L. Yang等^[34]报道了一种“分层学习(Hierarchical Learning)方法”, 该方法将知识和实验设计相结合, 实现了对数据的启发式处理。

全合成设计要求从全局识别分子中可能的成断键模式和官能团转换, 并对可能长达数十步的流程进行最优化。这需要训练数据中包含经过验证的模板化反应, 并排除未经证实的罕见转化。早期, 这一目标似乎难以实现, 但随着机器学习技术的发展, 现在已有可能解决此类复杂问题。2020年, B. Mikulak-Klucznik等人^[35]报道了“Chematica”(也称“Synthia”)系统, 该系统通过学习超过十万条反应规则, 能够对复杂天然产物进行全合成规划, 所预测的路线已在实验室中验证。

利用机器学习进行全合成分析并非完全的黑箱操作, 对于算法和模型进行调整能生成满足特定需求的合成路径。例如, 指定起始原料^[36]或探索废弃化合物的再利用^[37]。2023年, Y. Lin等人^[38]报道了一种结合人工选择和自动生成的方法, 用于最小化生物碱全合成路径的步数。同年, D. Kreutter等人^[39]利用路径惩罚分数(Route Penalty Score, RPScore)筛选CASP生成的可能路径, 获得了偏好商业可得材料起始的短合成路线。

在局部最优化方面, I. Levin等人^[40]在2022年报道了一种合成搜索算法, 该算法利用酶促反应替换成本高昂的过渡金属催化反应或是对映异构体分离过程来优化合成路径, 如图2所示。此外, 针对

单步反应的结果预测和可行性分析,也发展了一些模型,如T. Klucznik等人^[41]在2024年报道的碳正离子重排结果的精细预测模型,这些模型在中间级别的描述水平上提供了广泛而准确的预测,避免了从头算法的高昂成本和粗粒化模型的准确度损失,被期望用于指导复杂产物的合成和分析、受限环境下的重排反应和自动化高通量反应筛选。

通过这些进展,机器学习正在逐步改变全合成设计的面貌,为化学研究提供了新的工具和策略。

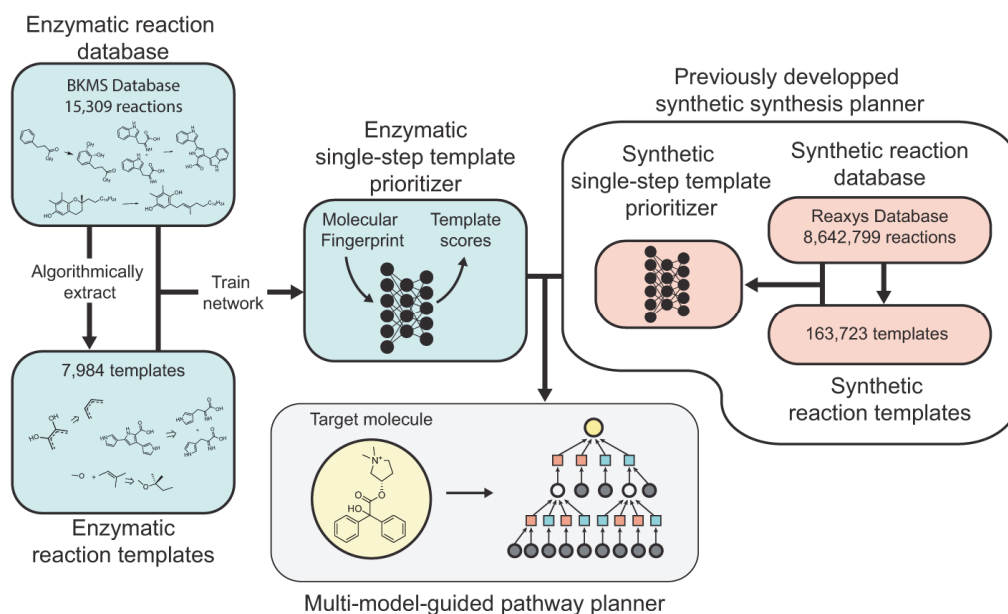


图2 酶促反应/有机合成混合规划器的开发流程

该模型结合了酶促反应数据库和有机合成数据库,因此能够利用罕见反应以取代复杂反应和提升路线转化率

基于CC-BY协议引自I. Levin等, 2022^[40]

2.2 机器学习势

原子模拟通过模拟原子的运动轨迹,包括位置、速度和能量随时间的演化^[42,43],来研究微观尺度的化学过程^[44-47]。实现这一目标的方法主要分为经典分子动力学(Classical Molecular Dynamics, CMD)和从头算分子动力学(*ab initio* Molecular Dynamics, AIMD)。CMD方法使用基于经验的势函数来计算原子间的作用力,计算速度快但精度有限^[48]。AIMD方法通过电子结构计算,如密度泛函理论(Density Functional Theory, DFT)^[49],可获得更精确的轨迹信息,但需要消耗更多的计算资源。

为了结合AIMD的高精度与CMD的高效率,机器学习提供了一种创新的解决方案。机器学习势(Machine Learning Potentials, MLPs)能够在保持AIMD精度的同时,以接近CMD的计算速度预测势能面(Potential Energy Surface, PES)。1995年, T.B. Blank等人^[50]提出了第一个机器学习势,而2007年 J. Behler和M. Parrinello^[51]提出的高维神经网络(High-dimensional Neural Network, HDNN),开启了这一领域的高速发展。使用了神经网络方法拟合势函数的方法,被称为神经网络势(Neural Network Potentials, NNPs)。表1总结了近年来报道的部分神经网络势。

J. Behler^[52,53]将高维神经网络势及其他机器学习势分为四代。第一代模型主要针对低维系统,依赖于少数自由度。第二代模型,以HDNNP为代表,通过原子中心对称函数描述的局部化学环境来构建能量。第三代HDNNP包括了长程相互作用,主要是静电相互作用。第四代HDNNP则能够描述长程电荷转移,例如全局电荷均衡方法,并能够处理体系的不同电荷状态。最近,一系列基于消息传递神经网络(Message Passing Neural Networks)的模型被开发出来,用以捕捉长程作用信息。第二代模型因其在局部环境描述上的高效性和准确性而被广泛使用。第三代和第四代模型虽然发展时间较

短, 但已能为多种系统提供接近电子结构计算精度的模拟, 能够用于大规模体系的计算。

机器学习势因其简便的数据获取、简洁的模型形式以及可验证的精确性和准确性而受到青睐。作为科研训练的一部分, 简单的机器学习势能够引导科研工作者深入理解微观动力学的一些概念^[54]。对于特定体系的训练, 机器学习势提供了一种高效的方法, 使得对特定化学体系的深入研究成为可能^[55]。

表1 神经网络势的发展

时间	模型	代表性应用	时间	模型	代表性应用
1995	D. J. Doren等 ^{[50]*}	CO在Ni(111)表面的吸附	2019	EANN ^[70]	
1998	K. Hermansson等 ^{[56]*}	氧化物晶体		AIMNet ^[71]	
2004	S. Lorenz等 ^{[57]*}	H ₂ /K/Pd(100)表面		PhysNet ^[72]	肽链结构
2006	S. Manzhos等 ^{[58]*}		2020	AP-Net ^[73]	氢键二聚体
2007	2G-HDNNP ^[51]	体相Si		SingleNN ^[74]	
2011	3G-HDNNP ^[59]	ZnO		BpopNN ^[75]	Li _n H _n 团簇
2012	PairHDNNP ^[60]	CH ₃ OH分子; Cu		EPNN ^[76]	生物大分子
2013	PIPNN ^[61]	H + H ₂ ; Cl + H ₂	2021	4G-HDNNP ^[77]	Au ₂ 团簇/MgO(001)
2015	CENT ^[62]	NaCl团簇		ANAKIN-ME ^[78]	开壳层结构
2017	BIM-NN ^[63]			DL_FFLUX ^[79]	柔性分子; 水
	DTNN ^[64]	芳香环结构	2022	DPLR ^[80]	NaCl; 水
2018	ANI-1 ^[65]	有机小分子		QRNN ^[81]	有机小分子
	SchNet ^[66]	C20富勒烯		NequIP ^[82]	
	DeePMD ^[67]				
	Tensormol ^[68]	水团簇			
	HIP-NN ^[69]	有机分子			

*部分报道的模型无特定名称, 此处使用作者姓名表示。

2.3 多相催化

多相催化是化学领域中一个高度复杂的多尺度现象, 它涉及不同时间尺度和空间尺度的相互作用。尽管第一性原理计算和分子动力学模拟等先进计算技术已经对许多相对简单的体系进行了有效的预测和验证, 但现有模拟手段往往与实验条件存在一定距离, 导致理论和实验之间存在差距。随着实验技术和理论计算的发展, 海量数据的获得进一步增加了建立普适性催化理论的难度。

机器学习在多相催化设计中旨在识别数据中的隐藏特征, 并实现三个主要目标: 考察反应位点的性质、揭示反应路径和加速催化剂的发现^[83]。虽然理解非均相催化反应的动力学过程并揭示反应路径对优化催化至关重要, 但考察每一步可能的反应成本高昂, 且相当繁琐。

为了自动化这一过程, 研究者们开发了多种软件工具。U. Gupta等人^[84]开发的Python Reaction Stencil (pReSt)软件, 基于DFT计算数据, 通过考察原子连接关系来识别可能的分子转化路径。S. Rangarajan等人^[85]开发的RING (Rule Input Network Generator)软件, 能够从给定的反应规则构建反应网络。J. T. Margraf等人^[86]构建了一个包含超过100万个反应的网络, 用于分析碳、氢、氧小分子物种间的转化关系和能垒。C. W. Gao等人^[87]提出的Reaction Mechanism Generator (RMG)^[88]从反应速率的角度考察反应途径, 基于这个思路C. F. Goldsmith等人^[89]开发的RMG-Cat软件则用于生成基元反应列表, 并筛选出具有高生成速率的表面物种和中间体。

结合反应路径预测和活性位点微观环境考察的方法, 即“微观动力学引导的机器学习路径搜索

(Microkinetics-Guided Machine Learning Pathway Search)” , 为模型训练数据提供了重要补充。如X. Zhang等人^[90]提出的基于随机表面行走的反应采样方法(Stochastic Surface Walking based Reaction Sampling, SSW-RS)。联合使用机器学习势分子动力学模拟和反应路径搜索, 是全面考察催化反应物网络的有力手段。2019年, P. Kang等^[91]使用SSW技术结合全局神经网络(Global Neural Network, G-NN)势对葡萄糖热解反应过程进行考察, 在提供该过程全面见解的同时, 修正了先前报道的机理, 验证了该类方法在建立反应数据库和探索复杂转化方面的可能性。

2022年, Y. Shi等人^[92]对铜-锌催化剂上CO₂/CO混合物合成甲醇的研究, 提供了有关于进料气体组成、催化剂活性位点和反应活性在这个复杂的异质催化系统中动态耦合的全面图景, 如图3所示。

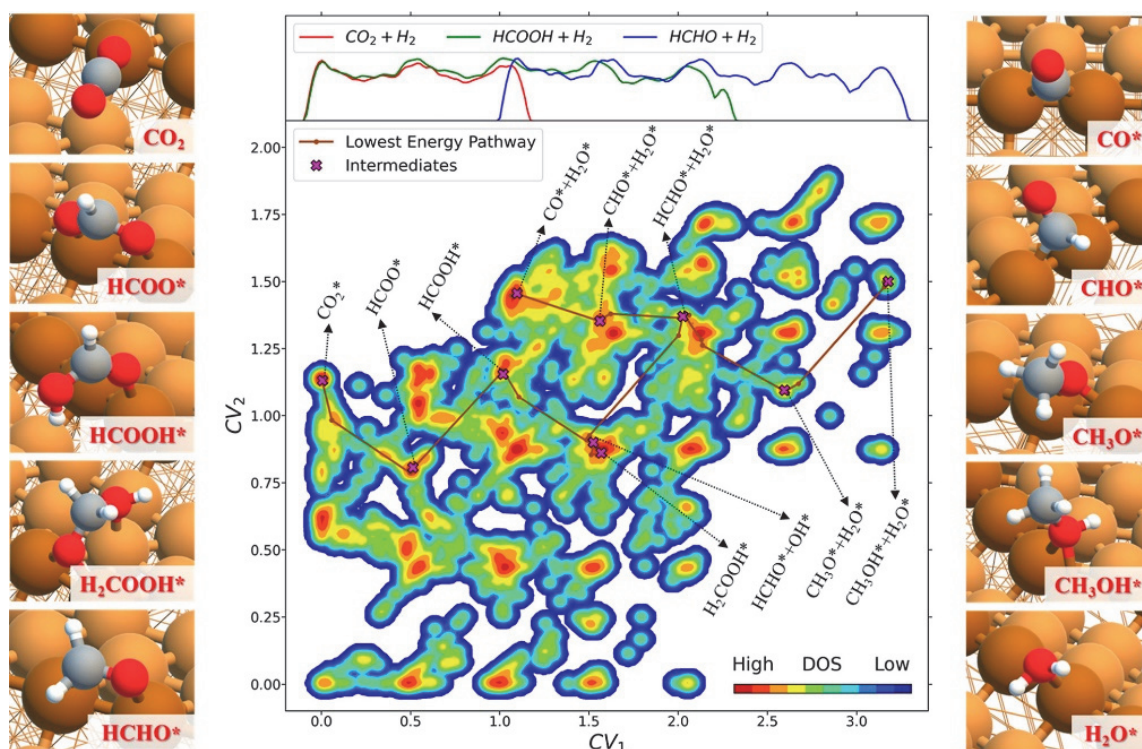


图3 通过微观动力学引导的机器学习路径搜索Cu(211)表面上CO₂-CO混合物的甲醇合成低能路径中的14958个反应对的轮廓图

在参考文献^[92]许可下转载, 美国化学学会2022年版权所有

仅使用机器学习势进行分子动力学模拟, 也可以用于反应历程, 如考察CO₂在Pt(111)上的解离^[93]、N₂在Ru(0001)上的解离^[94]、H₂O在TiO₂(101)上的解离^[95]、H₂O在Pt(110)上的吸附^[96]和Pt(111)上的析氢反应^[97]等反应历程。

从催化反应获得催化历程, 事实上和寻找高效催化剂是互逆的, 因此机器学习在催化剂发现中也展现出巨大潜力^[98]。M. Zhong等人^[99]设计的自动化框架, 通过DFT计算结合机器学习回归和排序, 建立了CO₂吸附能、H₂吸附能与CO₂还原反应催化能力之间的关系, 并筛选出Cu-Al位点作为潜在的催化剂候选。类似的催化剂筛选也包括合金作为CO₂还原和析氢反应电催化剂^[100]及钙钛矿作为析氧反应电催化剂^[101], 并产生了多个有潜力的催化剂候选和经验证的高活性催化剂。

机器学习的特征分析和重要性排序为催化剂的理论洞见提供了新的视角^[102-105]。例如, 对钙钛矿氧化物氧还原反应催化活性的机器学习分析揭示了阳离子间的Lewis酸强度差异是影响催化活性的关键因素^[106]。无监督学习也能够自行从数据中提取关键特征^[107]。

机器学习应用于多相催化设计, 可赋予研究者深入考察复杂体系的能力, 已经产出了多项令人兴奋的结果, 并为未来的研究开辟了新的道路。

2.4 核磁共振谱

核磁共振(Nuclear Magnetic Resonance, NMR)谱学作为重要的化学表征手段之一, 能够揭示分子结构和化学环境的详细信息。传统的解谱方法和基于量子化学的谱图预测已经为化学家提供了丰富的信息, 机器学习技术的引入有望为难以处理的复杂结构或混合物NMR数据处理带来新的突破。

在NMR谱解析的多个环节, 包括采样、去噪声、谱图解析、峰指认, 以及未知物结构的重建, 机器学习都能够发挥着重要作用。X. Qu等人^[108]和D. F. Hansen等人^[109]通过深度学习技术对NMR谱的重建进行了概念性验证。J. Luo等人^[110]报道了深度编解码器高分辨率网络(deep Encoder-Decoder High-Resolution Networks, EDHRN)在非均匀采样(欠采样)的高维NMR谱重建中的应用。D. Li等^[111]开发的DEEP Picker方法, 基于深度神经网络, 用于二维NMR谱的半自动解析, 在分析蛋白质和代谢组学样品方面具有实用性。G. Karunanithy等人^[112]报道了对NMR碳谱的解耦方法; K. Wu等人^[113]报道了去噪声方法DN-Unet以增强信噪比; X. Zheng^[114]和Z. Yang等人^[115]报道了用于获取纯位移核磁共振谱的方法。

在解谱方面, 官能团识别的研究历史可以追溯到1975年^[116]和1977年^[117]。2020年, J. A. Fine等人^[118]训练了神经网络, 以识别未知化合物中的官能团。2021年, T. Specht等人^[119]报道了一种自动识别和定量分析NMR光谱中结构基团的方法。同年, 如图4所示, Z. Huang等人^[120]提出的机器学习框架能够根据NMR氢谱和碳谱预测分子的亚结构, 并为可能的同分异构体提供概率排名: 在测试集上, 正确的结构在67.4%的情况下排名最高, 而在95.8%的情况下是概率最大的十个结构之一。

NMR谱不仅用于分子结构解析, 还可以作为物质的“指纹”来提供其他类型的信息。例如, 使用支持向量机(Support Vector Machine, SVM)分析鲑鱼鱼油的NMR氢谱以鉴别其产地^[121], 使用主成分分析(Principal Component Analysis, PCA)对葡萄酒的产地和类型进行分类^[122]。

上述应用展示了机器学习在NMR谱学中强大的预测能力, 为化学家在解析核磁共振氢谱时提供了新的工具和策略。

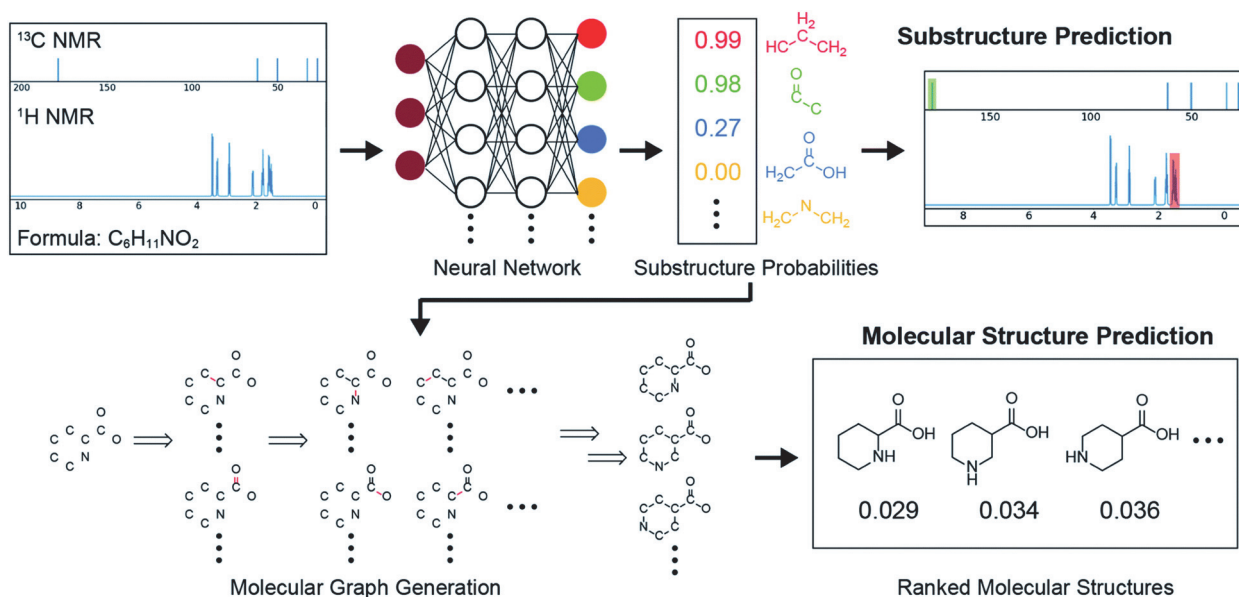


图4 自动结构预测框架, 根据NMR氢谱和碳谱猜测亚结构, 并构造同分异构体作为可能候选的概率排名, 在67.4%的情况下正确的构造结构排名最高, 而在95.8%的情况是最可能的十个构造之一

3 机器人化学家

在化学研究迈向自动化的过程中,机器学习与自动化化学反应的结合正开启新的可能性^[123-126]。这种集成系统,被形象地称为“移动机器人化学家”。移动机器人化学家通过机器学习算法处理大量文献数据,能够独立设计和执行研究计划,实现科学研究的自动化。

2019年,C. W. Coley等^[26]报道了一个专门针对有机合成的先进机器人平台。该平台能够自动分析目标化合物的潜在合成路径,并执行流动合成操作,尽管在实验参数的最终设定上仍需人类化学家的介入。该平台已成功应用于多种目标化合物的合成,包括那些需要精确立体化学控制的复杂分子,如(S)-华法林、沙非酰胺、喹那普利和塞来西布。

2020年,B. Burger等^[127]介绍了一种具备自主移动能力的机器人,它利用激光扫描和触觉反馈技术进行定位,能够执行与人类实验员相媲美的精细实验任务。该机器人系统集成贝叶斯优化算法,在人类化学家提出催化剂活性相关的假设后,机器人能够独立优化实验变量,以获得高活性的催化剂。

2022年,Q. Zhu等^[128]报道了“Chemputer”机器人系统,它通过集成的文献阅读能力,能够自动完成从假设提出到实验执行再到结果分析的科研流程。Chemputer由三个核心模块构成:机器阅读模块负责自动解析科学文献,提取化学知识;移动机器人模块负责根据预设方案执行化学实验;计算脑模块负责理论计算和构建基于物理和理论的预测模型(图5)。该系统在多项科研任务中的成功应用,证明了其在自动化化学研究中的潜力。

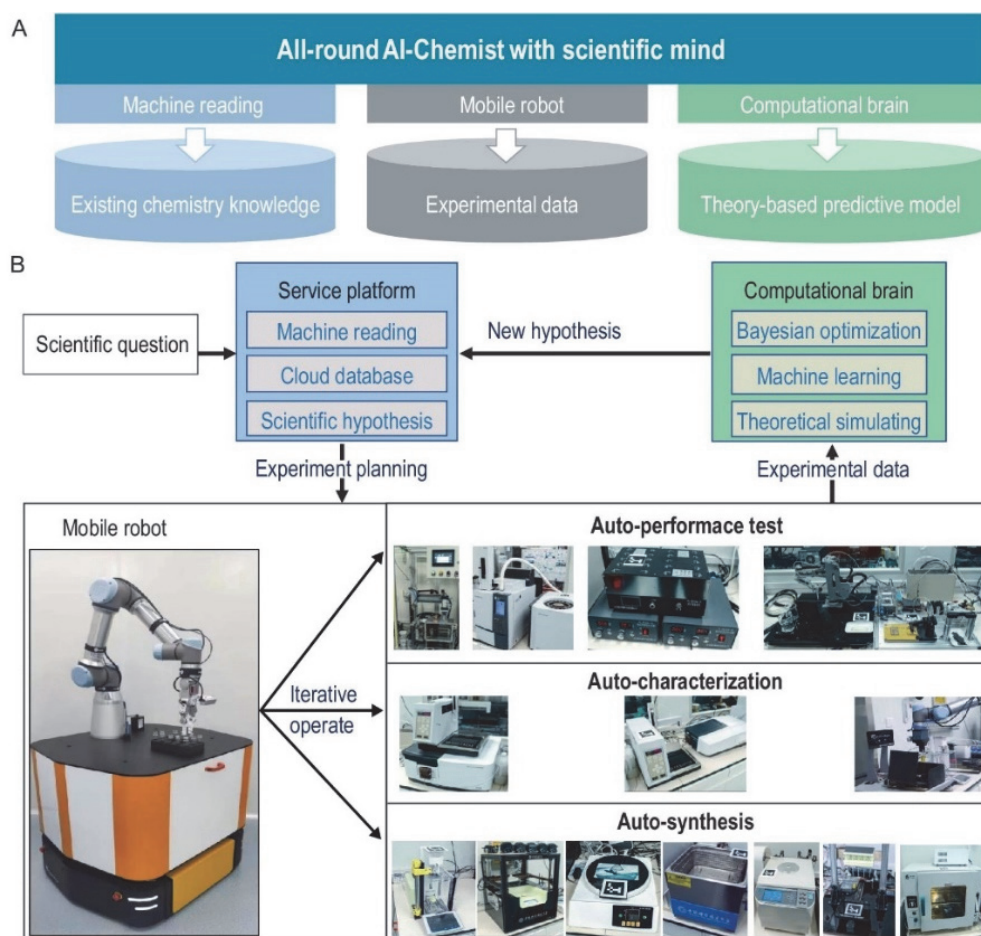


图5 AI化学家“Chemputer”的组成(A)和工作流程(B)

基于CC BY协议引自Z. Qing等, 2022^[120]

2023年, T. Ha等^[129]报道了“Synbot”机器人,它能够独立进行合成规划和批量反应器实验,自主收集实验数据,并优化反应流程以提升转化率。Synbot的设计分为三个层次:人工智能软件层负责主导合成规划,包括逆合成分析、实验设计和优化;机器人软件层负责将合成方案转化为可执行的机器人命令;机器人硬件层则在模块化实验室中执行具体的实验操作。

机器人化学家为当前化学研究范式带来了革新。机器人化学家高效且自动化的文献阅读、实验规划、实验执行能力使人类化学家有机会集中精力于更加抽象而创造性的研究,是化学研究自动化的重要发展方向。

4 展望

本文探讨了机器学习技术在化学领域的应用,包括有机合成设计、机器学习势、异相催化以及核磁共振谱解析。尽管本文已经覆盖了多个应用实例,但由于篇幅和主题的限制,仍有一些相关研究未能讨论,如周期性结构的处理和预测^[130-135]、光谱预测^[136,137]、化学知识模型^[138]以及电子结构计算^[139,140]等。

尽管机器学习在化学领域的应用充满希望,但也面临一些挑战。当前,化学领域的工作者可能对机器学习算法不够熟悉,这可能导致即使面对高度相关的工作,也难以准备出适合机器学习分析的数据集。此外,机器学习作为一个快速发展的跨学科领域,其技术的快速迭代和对计算机能力的需求,对化学研究人员而言是较高的入门门槛。

展望未来,我们有理由相信机器学习将变得更加智能化、泛用化和用户友好。机器学习的进步可能会在多个方面超越传统的化学研究方法,成为化学研究的新范式。通过综合运用多种机器学习算法,如化学知识模型、原子模拟和实验规划,将显著提升化学探索的速度和数据质量,加速科学发现的进程。最终,人类化学家的角色将转变为更加集中于创新思维和战略规划,与机器的优势形成互补,而非简单的竞争关系。这种协同合作将极大提升研究效率,产出更多突破性的科研成果。

参 考 文 献

- [1] Sarker, I. H. *SN Comput. Sci.* **2021**, *2*, 160.
- [2] Moen, E.; Bannon, D.; Kudo, T.; Graf, W.; Covert, M.; Van Valen, D. *Nat. Meth.* **2019**, *16*, 1233.
- [3] Madanian, S.; Chen, T.; Adeleye, O.; Templeton, J. M.; Poellabauer, C.; Parry, D.; Schneider, S. L. *Intell. Syst. Appl.* **2023**, *20*, 200266.
- [4] 刘小平, 刘耀虎, 郑企雨, 朱相丽. 化学通报, **2023**, *86*, 748.
- [5] 孙婕, 李子昊, 张书宇. 上海交通大学学报, **2023**, *57*, 1231.
- [6] Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; *et al.* *Nature* **2021**, *596*, 583.
- [7] Yano, J.; Gaffney, K. J.; Gregoire, J.; Hung, L.; Ourmazd, A.; Schrier, J.; Sethian, J. A.; Toma, F. M. *Nat. Rev. Chem.* **2022**, *6*, 357.
- [8] Holme, T. A.; Luxford, C. J.; Brandriet, A. *J. Chem. Educ.* **2015**, *92*, 1477.
- [9] Langer, M. F.; Goeßmann, A.; Rupp, M. *npj Comput. Mater.* **2022**, *8*, 41.
- [10] Weininger, D. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31.
- [11] Heller, S. R.; Mcnaught, A.; Pletnev, I.; Stein, S.; Tchekhovskoi, D. *J. Cheminf.* **2015**, *7*, 23.
- [12] Rogers, D.; Hahn, M. *J. Chem. Inf. Model.* **2010**, *50*, 742.
- [13] Behler, J. *J. Chem. Phys.* **2011**, *134*, 074106.
- [14] Bartók, A. P.; Kondor, R.; Csányi, G. *Phys. Rev. B* **2013**, *87*, 184115.
- [15] Wolpert, D. H. *Neural Comput.* **1996**, *8*, 1341.

- [16] Wolpert, D. H.; Macready, W. G. *IEEE Trans. Evol. Comput.* **1997**, *1*, 67.
- [17] Williams, W. L.; Zeng, L.; Gensch, T.; Sigman, M. S.; Doyle, A. G.; Anslyn, E. V. *ACS Cent. Sci.* **2021**, *7*, 1622.
- [18] Corey, E. J.; Wipke, W. T. *Science* **1969**, *166*, 178.
- [19] Pensak, D. A.; Corey, E. J. LHASA—Logic and Heuristics Applied to Synthetic Analysis. In *Computer-Assisted Organic Synthesis*; Wipke, W. T., Howe, W. J. Eds.; American Chemical Society: Washington, DC, USA, 1977; pp. 1–32.
- [20] Szymkuć, S.; Gajewska, E. P.; Klucznik, T.; Molga, K.; Dittwald, P.; Startek, M.; Bajczyk, M.; Grzybowski, B. A. *Angew. Chem. Int. Ed.* **2016**, *55*, 5904.
- [21] Klucznik, T.; Mikulak-Klucznik, B.; McCormack, M. P.; Lima, H.; Szymkuć, S.; Bhowmick, M.; Molga, K.; Zhou, Y.; Rickershauser, L.; Gajewska, E. P.; *et al.* *Chem.* **2018**, *4*, 522.
- [22] Kowalik, M.; Gothard, C. M.; Drews, A. M.; Gothard, N. A.; Weckiewicz, A.; Fuller, P. E.; Grzybowski, B. A.; Bishop, K. J. M. *Angew. Chem. Int. Ed.* **2012**, *51*, 7928.
- [23] Beker, W.; Gajewska, E. P.; Badowski, T.; Grzybowski, B. A. *Angew. Chem. Int. Ed.* **2019**, *58*, 4515.
- [24] Badowski, T.; Gajewska, E. P.; Molga, K.; Grzybowski, B. A. *Angew. Chem. Int. Ed.* **2020**, *59*, 725.
- [25] 刘伊迪, 杨骐, 李遥, 张龙, 罗三中. *有机化学*, **2020**, *40*, 3812.
- [26] Coley, C. W.; Thomas, D. A.; Lummiss, J. a. M.; Jaworski, J. N.; Breen, C. P.; Schultz, V.; Hart, T.; Fishman, J. S.; Rogers, L.; Gao, H.; *et al.* *Science* **2019**, *365*, eaax1566.
- [27] Koch, M.; Duigou, T.; Faulon, J.-L. *ACS Synth. Biol.* **2020**, *9*, 157.
- [28] Delépine, B.; Duigou, T.; Carbonell, P.; Faulon, J.-L. *Metab. Eng.* **2018**, *45*, 158.
- [29] Finnigan, W.; Hepworth, L. J.; Flitsch, S. L.; Turner, N. J. *Nat. Catal.* **2021**, *4*, 98.
- [30] Genheden, S.; Thakkar, A.; Chadimová, V.; Reymond, J.-L.; Engkvist, O.; Bjerrum, E. *J. Cheminf.* **2020**, *12*, 70.
- [31] Schwaller, P.; Petraglia, R.; Zullo, V.; Nair, V. H.; Haeuselmann, R. A.; Pisoni, R.; Bekas, C.; Iuliano, A.; Laino, T. *Chem. Sci.* **2020**, *11*, 3316.
- [32] Probst, D.; Manica, M.; Nana Teukam, Y. G.; Castrogiovanni, A.; Paratore, F.; Laino, T. *Nat. Commun.* **2022**, *13*, 964.
- [33] Zheng, S.; Zeng, T.; Li, C.; Chen, B.; Coley, C. W.; Yang, Y.; Wu, R. *Nat. Commun.* **2022**, *13*, 3342.
- [34] Yang, L.-C.; Zhu, L.-J.; Zhang, S.-Q.; Hong, X. *Chin. J. Chem.* **2022**, *40*, 2106.
- [35] Mikulak-Klucznik, B.; Gołębiowska, P.; Bayly, A. A.; Popik, O.; Klucznik, T.; Szymkuć, S.; Gajewska, E. P.; Dittwald, P.; Staszewska-Krajewska O.; Beker W.; *et al.* *Nature* **2020**, *588*, 83.
- [36] Lin, Y.; Zhang, Z.; Mahjour, B.; Wang, D.; Zhang, R.; Shim, E.; Mcgrath, A.; Shen, Y.; Brugger, N.; Turnbull, R.; *et al.* *Nat. Commun.* **2021**, *12*, 7327.
- [37] Wołos, A.; Koszelewski, D.; Roszak, R.; Szymkuć, S.; Moskal, M.; Ostaszewski, R.; Herrera, B. T.; Maier, J. M.; Brezicki, G.; Samuel, J.; *et al.* *Nature* **2022**, *604*, 668.
- [38] Lin, Y.; Zhang, R.; Wang, D.; Cernak, T. *Science* **2023**, *379*, 453.
- [39] Kreutter, D.; Reymond, J.-L. *Chem. Sci.* **2023**, *14*, 9959.
- [40] Levin, I.; Liu M.; Voigt, C. A.; Coley, C. W. *Nat. Commun.* **2022**, *13*, 7747.
- [41] Klucznik, T.; Syntivanis, L.-D.; Baś, S.; Mikulak-Klucznik, B.; Moskal, M.; Szymkuć, S.; Mlynarski, J.; Gadina, L.; Beker, W.; Burke, M. D.; *et al.* *Nature* **2024**, *625*, 508.
- [42] Hollingsworth, S. A.; Dror, R. O. *Neuron* **2018**, *99*, 1129.
- [43] Ciccotti, G.; Dellago, C.; Ferrario, M.; Hernández, E. R.; Tuckerman, M. E. *Eur. Phys. J. B* **2022**, *95*, 3.
- [44] Han, J.; Chen, Y.; Wang, J.; Zhang, G.; Wang, H. *Int. J. Adv. Manuf. Tech.* **2022**, *122*, 1195.
- [45] Wu, X.; Xu, L.-Y.; Li, E.-M.; Dong, G. *Chem. Biol. Drug. Des.* **2022**, *99*, 789.
- [46] Groß, A.; Sakong, S. *Chem. Rev.* **2022**, *122*, 10746.
- [47] Paquet, E.; Viktor, H. L. *Adv. Chem.* **2018**, *2018*, 9839641.
- [48] Van Gunsteren, W. F.; Daura, X.; Hansen, N.; Mark, A. E.; Oostenbrink, C.; Riniker, S.; Smith, L. J. *Angew. Chem. Int. Ed.* **2018**, *57*, 884.

- [49] Mardirossian, N.; Head-Gordon, M. *Mol. Phys.* **2017**, *115*, 2315.
- [50] Blank, T. B.; Brown, S. D.; Calhoun, A. W.; Doren, D. J. *J. Chem. Phys.* **1995**, *103*, 4129.
- [51] Behler, J.; Parrinello, M. *Phys. Rev. Lett.* **2007**, *98*, 146401.
- [52] Kocer, E.; Ko, T. W.; Behler, J. *Annu. Rev. Phys. Chem.* **2022**, *73*, 163.
- [53] Behler, J. *Chem. Rev.* **2021**, *121*, 10037.
- [54] 周佳. 大学化学, **2024**, *39* (3), 351.
- [55] Tokita, A. M.; Behler, J. *J. Chem. Phys.* **2023**, *159*, 121501.
- [56] Gassner, H.; Probst, M.; Lauenstein, A.; Hermansson, K. *J. Phys. Chem.* **1998**, *102*, 4596.
- [57] Lorenz, S.; Groß, A.; Scheffler, M. *Chem. Phys. Lett.* **2004**, *395*, 210.
- [58] Manzhos, S.; Carrington, T., Jr. *J. Chem. Phys.* **2007**, *127*, 014103.
- [59] Artrith, N.; Morawietz, T.; Behler, J. *Phys. Rev. B* **2011**, *83*, 153101.
- [60] Jose, K. V. J.; Artrith, N.; Behler, J. *J. Chem. Phys.* **2012**, *136*, 194111.
- [61] Jiang, B.; Guo, H. *J. Chem. Phys.* **2013**, *139*, 054112.
- [62] Ghasemi, S. A.; Hofstetter, A.; Saha, S.; Goedecker, S. *Phys. Rev. B* **2015**, *92*, 045131.
- [63] Yao, K.; Herr, J. E.; Brown, S. N.; Parkhill, J. *J. Phys. Chem. Lett.* **2017**, *8*, 2689.
- [64] Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. *Nat. Commun.* **2017**, *8*, 13890.
- [65] Smith, J. S.; Isayev, O.; Roitberg, A. E. *Chem. Sci.* **2017**, *8*, 3192.
- [66] Schütt, K. T.; Kindermans, P.-J.; Saucedo, H. E.; Chmiela, S.; Tkatchenko, A.; Müller, K.-R. Schnet: A Continuous-Filter Convolutional Neural Network for Modeling Quantum Interactions. In *NIPS'17*, Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA, Dec 4–9, 2017; Ulrike, v. L., PictureIsabelle G., PictureSamy B., PictureHanna W., PictureRob F., Eds.; Curran Associates Inc.: Long Beach, 2017.
- [67] Zhang, L.; Han, J.; Wang, H.; Car, R.; E. W. *Phys. Rev. Lett.* **2018**, *120*, 143001.
- [68] Yao, K.; Herr, J. E.; Toth, D. W.; Mckintyre, R.; Parkhill, J. *Chem. Sci.* **2018**, *9*, 2261.
- [69] Lubbers, N.; Smith, J. S.; Barros, K. *J. Chem. Phys.* **2018**, *148*, 241715.
- [70] Zhang, Y.; Hu, C.; Jiang, B. *J. Phys. Chem. Lett.* **2019**, *10*, 4962.
- [71] Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O. *Sci. Adv.* **2019**, *5*, eaav6490.
- [72] Unke, O. T.; Meuwly, M. *J. Chem. Theory Comput.* **2019**, *15*, 3678.
- [73] Glick, Z.; Metcalf, D.; Koutsoukas, A.; Spronk, S.; Cheney, D.; Sherrill, D. *J. Chem. Phys.* **2020**, *153*, 044112.
- [74] Liu, M.; Kitchin, J. R. *J. Phys. Chem. C* **2020**, *124*, 17811.
- [75] Xie, X.; Persson, K. A.; Small, D. W. *J. Chem. Theory Comput.* **2020**, *16*, 4256.
- [76] Metcalf, D. P.; Jiang, A.; Spronk, S. A.; Cheney, D. L.; Sherrill, C. D. *J. Chem. Inf. Model.* **2021**, *61*, 115.
- [77] Ko, T. W.; Finkler, J. A.; Goedecker, S.; Behler, J. *Nat. Commun.* **2021**, *12*, 398.
- [78] Zubatiuk, T.; Isayev, O. *Acc. Chem. Res.* **2021**, *54*, 1575.
- [79] Symons, B. C. B.; Bane, M. K.; Popelier, P. L. A. *J. Chem. Theory Comput.* **2021**, *17*, 7043.
- [80] Zhang, L.; Wang, H.; Muniz, M. C.; Panagiotopoulos, A. Z.; Car, R.; E, W. *J. Chem. Phys.* **2022**, *156*, 124107.
- [81] Jacobson, L. D.; Stevenson, J. M.; Ramezanghorbani, F.; Ghoreishi, D.; Leswing, K.; Harder, E. D.; Abel, R. *J. Chem. Theory Comput.* **2022**, *18*, 2354.
- [82] Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. *Nat. Commun.* **2022**, *13*, 2453.
- [83] Mou, T.; Pillai, H. S.; Wang, S.; Wan, M.; Han, X.; Schweitzer, N. M.; Che, F.; Xin, H. *Nat. Catal.* **2023**, *6*, 122.
- [84] Gupta, U.; Vlachos, D. G. *J. Chem. Inf. Model.* **2021**, *61*, 3431.
- [85] Rangarajan, S.; Bhan, A.; Daoutidis, P. *Comput. Chem. Eng.* **2012**, *45*, 114.
- [86] Margraf, J. T.; Reuter, K. *ACS Omega* **2019**, *4*, 3370.

- [87] Susnow, R. G.; Dean, A. M.; Green, W. H.; Peczak, P.; Broadbelt, L. J. *J. Phys. Chem.* **1997**, *101*, 3731.
- [88] Gao, C. W.; Allen, J. W.; Green, W. H.; West, R. H. *Comput. Phys. Commun.* **2016**, *203*, 212.
- [89] Goldsmith, C. F.; West, R. H. *J. Phys. Chem. C* **2017**, *121*, 9970.
- [90] Zhang, X.-J.; Liu, Z.-P. *Phys. Chem. Chem. Phys.* **2015**, *17*, 2757.
- [91] Kang, P.-L.; Shang, C.; Liu, Z.-P. *J. Am. Chem. Soc.* **2019**, *141*, 20525.
- [92] Shi, Y.-F.; Kang, P.-L.; Shang, C.; Liu, Z.-P. *J. Am. Chem. Soc.* **2022**, *144*, 13401.
- [93] Del Cueto M.; Zhou, X.; Zhou, L.; Zhang, Y.; Jiang, B.; Guo, H. *J. Phys. Chem. C* **2020**, *124*, 5174.
- [94] Lee, E. M. Y.; Ludwig, T.; Yu, B.; Singh, A. R.; Gygi, F.; Nørskov, J. K.; De Pablo, J. J. *J. Phys. Chem. Lett.* **2021**, *12*, 2954.
- [95] Calegari Andrade, M. F.; Ko, H.-Y.; Zhang, L.; Car, R.; Selloni, A. *Chem. Sci.* **2020**, *11*, 2335.
- [96] Hu, C.; Zhang, Y.; Jiang, B. *J. Phys. Chem. C* **2020**, *124*, 23190.
- [97] Rice, P. S.; Liu, Z.-P.; Hu, P. *J. Phys. Chem. Lett.* **2021**, *12*, 10637.
- [98] 宗宇杨, 李俊辉, 朱向东, 单光存, 马汝广. 上海大学学报(自然科学版), **2023**, *29*, 859.
- [99] Zhong, M.; Tran, K.; Min, Y.; Wang, C.; Wang, Z.; Dinh, C.-T.; De Luna, P.; Yu, Z.; Rasouli, A. S.; Brodersen, P.; *et al.* *Nature* **2020**, *581*, 178.
- [100] Tran, K.; Ulissi, Z. W. *Nat. Catal.* **2018**, *1*, 696.
- [101] Li, Z.; Achenie, L. E. K.; Xin, H. *ACS Catal.* **2020**, *10*, 4377.
- [102] Andersen, M.; Levchenko, S. V.; Scheffler, M.; Reuter, K. *ACS Catal.* **2019**, *9*, 2752.
- [103] Bartel, C. J.; Sutton, C.; Goldsmith, B. R.; Ouyang, R.; Musgrave, C. B.; Ghiringhelli, L. M.; Scheffler, M. *Sci. Adv.* **2019**, *5*, eaav0693.
- [104] Xu, W.; Andersen, M.; Reuter, K. *ACS Catal.* **2021**, *11*, 734.
- [105] Jiang, C.; Song, H.; Sun, G.; Chang, X.; Zhen, S.; Wu, S.; Zhao, Z.-J.; Gong, J. *Angew. Chem. Int. Ed.* **2022**, *61*, e202206758.
- [106] Zhai, S.; Xie, H.; Cui, P.; Guan, D.; Wang, J.; Zhao, S.; Chen, B.; Song, Y.; Shao, Z.; Ni, M. *Nat. Energy* **2022**, *7*, 866.
- [107] Esterhuizen, J. A.; Goldsmith, B. R.; Linic, S. *Chem. Catal.* **2021**, *1*, 923.
- [108] Qu, X.; Huang, Y.; Lu, H.; Qiu, T.; Guo, D.; Agback, T.; Orekhov, V.; Chen, Z. *Angew. Chem. Int. Ed.* **2020**, *59*, 10297.
- [109] Hansen, D. F. *J. Biomol. NMR* **2019**, *73*, 577.
- [110] Luo, J.; Zeng, Q.; Wu, K.; Lin, Y. *J. Magn. Reson.* **2020**, *317*, 106772.
- [111] Li, D.-W.; Hansen, A. L.; Yuan, C.; Bruschweiler-Li, L.; Bruschweiler, R. *Nat. Commun.* **2021**, *12*, 5229.
- [112] Karunanithy, G.; Mackenzie, H. W.; Hansen, D. F. *J. Am. Chem. Soc.* **2021**, *143*, 16935.
- [113] Wu, K.; Luo, J.; Zeng, Q.; Dong, X.; Chen, J.; Zhan, C.; Chen, Z.; Lin, Y. *Anal. Chem.* **2021**, *93*, 1377.
- [114] Zheng, X.; Yang, Z.; Yang, C.; Shi, X.; Luo, Y.; Luo, J.; Zeng, Q.; Lin, Y.; Chen, Z. *J. Phys. Chem. Lett.* **2022**, *13*, 2101.
- [115] Yang, Z.; Zheng, X.; Gao, X.; Zeng, Q.; Yang, C.; Luo, J.; Zhan, C.; Lin, Y. *J. Phys. Chem. Lett.* **2023**, *14*, 3397.
- [116] Wilkins, C. L.; Isenhour, T. L. *Anal. Chem.* **1975**, *47*, 1849.
- [117] Wilkins, C. L.; Brunner, T. R. *Anal. Chem.* **1977**, *49*, 2136.
- [118] Fine, J. A.; Rajasekar, A. A.; Jethava, K. P.; Chopra, G. *Chem. Sci.* **2020**, *11*, 4618.
- [119] Specht, T.; Münnemann, K.; Hasse, H.; Jirasek, F. *J. Chem. Inf. Model* **2021**, *61*, 143.
- [120] Huang, Z.; Chen, M. S.; Woroch, C. P.; Markland, T. E.; Kanan, M. W. *Chem. Sci.* **2021**, *12*, 15329.
- [121] Masoum, S.; Malabat, C.; Jalali-Heravi, M.; Guillou, C.; Rezzi, S.; Rutledge, D. N. *Anal. Bioanal. Chem.* **2007**, *387*, 1499.
- [122] Du, Y.-Y.; Bai, G.-Y.; Zhang, X.; Liu, M.-L. *Chin. J. Chem.* **2007**, *25*, 930.
- [123] Dragone, V.; Sans, V.; Henson, A. B.; Granda, J. M.; Cronin, L. *Nat. Commun.* **2017**, *8*, 15733.
- [124] Schweidtmann, A. M.; Clayton, A. D.; Holmes, N.; Bradford, E.; Bourne, R. A.; Lapkin, A. A. *Chem. Eng. J.* **2018**, *352*, 277.
- [125] Bédard, A.-C.; Adamo, A.; Aroh, K. C.; Russell, M. G.; Bedermann, A. A.; Torosian, J.; Yue, B.; Jensen, K. F.; Jamison, T. F. *Science* **2018**, *361*, 1220.
- [126] Granda, J. M.; Donina, L.; Dragone, V.; Long, D.-L.; Cronin, L. *Nature* **2018**, *559*, 377.
- [127] Burger, B.; Maffettone, P. M.; Gusev, V. V.; Aitchison, C. M.; Bai, Y.; Wang, X.; Li, X.; Alston, B. M.; Li, B.; Clowes, R.; *et al.* *Nature* **2020**,

583, 237.

- [128] Zhu, Q.; Zhang, F.; Huang, Y.; Xiao, H.; Zhao, L.; Zhang, X.; Song, T.; Tang, X.; Li, X.; He, G.; *et al. Natl. Sci. Rev.* **2022**, *9* (10), nwac190.
- [129] Ha, T.; Lee, D.; Kwon, Y.; Park, M. S.; Lee, S.; Jang, J.; Choi, B.; Jeon, H.; Kim, J.; Choi, H.; *et al. Sci. Adv.* **2023**, *9*, eadj0461.
- [130] Geiger, P.; Dellago, C. *J. Chem. Phys.* **2013**, *139*, 164105.
- [131] Reinhart, W. F.; Long, A. W.; Howard, M. P.; Ferguson, A. L.; Panagiotopoulos, A. Z. *Soft Matter*. **2017**, *13*, 4733.
- [132] Dietz, C.; Kretz, T.; Thoma, M. H. *Phys. Rev. E* **2017**, *96*, 011301.
- [133] Schmidt, J.; Pettersson, L.; Verdozzi, C.; Botti, S.; Marques, M. A. L. *Sci. Adv.* **2021**, *7*, eabi7948.
- [134] Leitherer, A.; Ziletti, A.; Ghiringhelli, L. M. *Nat. Commun.* **2021**, *12*, 6234.
- [135] Banik, S.; Dhabal, D.; Chan, H.; Manna, S.; Cherukara, M.; Molinero, V.; Sankaranarayanan, S. K. R. S. *npj Comput. Mater.* **2023**, *9*, 23.
- [136] Chen, Z.; Yam, V. W.-W. *J. Phys. Chem. Lett.* **2023**, *14*, 1955.
- [137] Ye, S.; Zhong, K.; Zhang, J.; Hu, W.; Hirst, J. D.; Zhang, G.; Mukamel, S.; Jiang, J. *J. Am. Chem. Soc.* **2020**, *142*, 19071.
- [138] Zhao, Z.; Ma, D.; Chen, L.; Sun, L.; Li, Z.; Xu, H.; Zhu, Z.; Zhu, S.; Fan, S.; Shen, G.; *et al. CoRR* **2024**, arXiv:2401.14818. [2024-08-02].
<https://doi.org/10.48550/arXiv.2401.14818>
- [139] Hauser, A. W. *Nat. Comput. Sci.* **2024**, *4*, 163.
- [140] Keith, J. A.; Vassilev-Galindo, V.; Cheng, B.; Chmiela, S.; Gastegger, M.; Müller, K.-R.; Tkatchenko, A. *Chem. Rev.* **2021**, *121*, 9816.