机器学习在分子动力学模拟方面的研究进展

韩雪梅

（华东理工大学）

**摘 要**：分子动力学模拟技术是对“模型分子”进行“计算机实验”的方法的总称。这里，模型分子是指模拟对象的近似性，计算机实验则是指模拟过程的近似性。应用分子模拟技术不仅能够提供定性的描述，而且能给出分子结构与宏观性能间的定量结果；不仅能够模拟分子的静态结构，也能够模拟分子的动态行为（如氢键的缔合与解缔吸附扩散等）。分子动力学模拟甚至还能够模拟现代实验手段尚难以考察的物理现象与过程，从而发展新的理论，研究化学反应途径、过渡态、反应机理等关键问题。分子动力学模拟技术亦可用于分子体系的各种分析测试（如射线衍射光谱）电子显微镜结果的模拟，这不仅可以使实验结果获得更为合理的解释，而且可用于对产品的结构解析。近年来，各种算法工具不断完善，使用机器学习算法的门槛不断降低，越来越多的研究员将机器学习应用于分子动力学研究，开启了分子动力学研究的新篇章。而在分子动力学模拟领域，目前大部分研究人员所使用的方法以监督式学习为主。本文简单总结了现阶段机器学习在分子动力学研究中的应用进展。

**关键词**：机器学习；结构表征；计算模拟；分子动力学模拟；监督式学习

**中图分类号：0645;TQ028 文献标志码**：**A**

**Research progress of machine learning in molecular simulation**

Han Xuemei

(East China University of Science and Technology)

**Abstract:** Molecular simulation technology is the general name of the method of "computer experiment" for "model molecules". Here, the model molecule refers to the approximation of the simulation object, and the computer experiment refers to the approximation of the simulation process. The application of molecular simulation technology can not only provide qualitative description, but also give quantitative results between molecular structure and macroscopic properties. It can not only simulate the static structure of molecules, but also simulate the dynamic behavior of molecules (such as hydrogen bond association and disassociation adsorption diffusion, etc.). Molecular simulation can even simulate physical phenomena and processes that are difficult to be investigated by modern experimental means, thus developing new theories and studying key issues such as chemical reaction pathways, transition states and reaction mechanisms. Molecular simulation technology can also be used to simulate the electron microscope results of various analytical tests (such as X-ray diffraction spectra) of molecular systems, which can not only make the experimental results more reasonable, but also be used to analyze the structure of products. In recent years, various algorithmic tools have been continuously

improved, and the threshold for using machine learning algorithms has been continuously lowered. More and more researchers have applied machine learning to molecular dynamics research, which has opened a new chapter in molecular dynamics research. In the field of molecular dynamics simulation, at present, most researchers mainly use supervised learning. This paper briefly summarizes the application progress of machine learning in molecular dynamics research at present.

**Keywords:** Machine learning; Structural characterization; Computational simulation; Molecular dynamics simulation; Supervised learning

**1 研究背景**

机器学习算法的雏形在20世纪90年代就已出现，但是局限于当时的硬件条件，计算机寻找关联数据间复杂函数的能力有限，机器学习算法在当时并没有得到长足的发展。随着计算机科学的飞速发展，电脑软件与硬件性能得到了巨大的提升,前人设想的基于机器学习的人工智能算法一定程度上得到了实现。近年来，各种算法工具不断完善，使用机器学习算法的门槛不断降低，越来越多的研究员将机器学习应用于分子动力学研究，开启了分子动力学研究的新篇章。

目前，用于构建机器力场学习的模型非常多，如GAP、DPMD、HDNN、PhysNet、ANI等。在分子动力学模拟中，机器学习力场模型可以分为以能量为中心和以力为中心的模型。以能量为中心的模型通过学习势能面，并计算势能面的导数得到力。如梯度域机器学习[[[1]](#endnote-1)](Gradient-Domain MachineLearning, GDML)、对称梯度域机器学习[[[2]](#endnote-2)](Symmetrized Gradient-Domain Machine Learning, sGDML)通过学习能力之间的梯度函数来简化能量和力的表达形式。另外一种以能量为中心的模型如DimeNet、DimeNet++、SchNet采用图神经网络[[[3]](#endnote-3)]的方法，实现对能量的平滑预测。

物质运动具有能量，一般用场来描述物体的运动状态。而力场用来估计分子内原子之间以及分子之间的力。力场是原子间势，可以使用能量来描述场。因为力实际上可以认为是能量的负梯度。

 (1-1)

其中，*Fi* 表示作用在每个原子上的力；*E*为体系内的能量 ；*ri*表示原子*i*的笛卡尔坐标向量。

力是矢量，在空间中的每个原子位置需要用笛卡尔坐标系来描述，而能量是标量，对空间中的每个点只需要用一个值来描述。因此与使用力来描述场比起来，用能量来描述力场更为简单。力场建立了原子坐标和系统相应总能量之间的映射。传统力场是基于晶体结构键长和键角来拟合参数的。与经典力场相比，机器学习力场可以不利用先入为主的理论知识，不对经验函数进行参数化表示，而是通过现有数据来学习原子的能量及其他特征，作为其函数表示，来代替经典力场。该系统中的能量为系统中所有原子能量的总和。

 (1-2)

其中，*E*为体系内总能量；*R*为原子位置。从上述方程也可以看出作用在原子上的力与相距较远的原子的种类和位置并没有很强的相关性。因此，机器学习力场模型可以依托系统内总能量来推测特征与原子能量之间的关系。

另外一种是在以力为中心的模型，在该模型中，力是通过网络模型来直接预测的，然后与实际的力进行比较。同时，加入了残差连接，改善随着网络深度的增加而带来的梯度消失、梯度爆炸等问题。使用力为中心的模型可以有效减少在以能量为中心的模型中计算力所需的额外花销，从而提高计算效率。因为能

量是标量，而力为有3个方向的矢量，所以目前以能量为中心开发的机器学习力场模型居多，而通过以力为中心的模型来计算能量，需要考虑到各个方向上的力的作用，具有一定的复杂性。

分子动力学（Molecular dynamics）是一门结合物理、数学和化学的综合技术。分子动力学（Molecular dynamics）也是一套分子模拟方法，该方法主要是依靠牛顿力学来模拟分子体系的运动，以在由分子体系的不同状态构成的系综中抽取样本，从而计算体系的构型积分，并以构型积分的结果为基础进一步计算体系的热力学量和其他宏观性质。其总是假定原子的运动服从某种确定的描述，这种描叙可以牛顿方程、拉格朗日方程或哈密顿方程所确定的描述，也就是说原子的运动和确定的轨迹联系在一起。在忽略核子的量子效应[[[4]](#endnote-4)]和Born-Oppenheimer[[[5]](#endnote-5)]绝热近似下，分子动力学的这一种假设是可行的。在分子动力学模拟中，一般采用经验势来代替原子间的相互作用势[[[6]](#endnote-6)]，如Lennard-Jones势、Mores势、EAM原子嵌入势、F-S多体势。然而采用经验势必然丢失了局域电子结构之间存在的强相关作用信息，即不能得到原子动力学过程中的电子性质。

分子动力学（Molecular dynamics）是一种计算机辅助模拟工具，用于描述物质或分子中的原子级运动过程。根据研究对象组成原子不同时刻的位置和动量，基于统计力学知识获得想要的物理量，解释对象的性质和行为。其基本过程为：(1)设置研究对象组成原子的初始位置和速度；(2)计算每个原子受到的合力，并基于牛顿第二定律计算原子的加速度；(3)计算原子下一时刻的速度；(4)计算原子下一时刻的位置；(5)循环(2)-(4)的过程，得到一系列时刻原子的位置和速度；(6)基于位置和速度信息得到描述对象性质和行为的物理量。分子动力学模拟的精度依赖于原子间相互作用的描述方法。其中，基于第一性原理的从头算分子动力学模拟（ab initio molecular dynamics），具有突出的精度表现，但其计算精度与电子自由度的3次方成正比，难以应用于原子数量级为103-108、时间尺度为微秒级的实际应用中。而基于empirical force fields (EFF)[[[7]](#endnote-7)]的分子动力学方法，虽然能够处理如此大规模的模拟，但其精度较低。

**2 机器学习在分子动力学模拟方面的研究进展**

# 2.1 机器学习在经典分子动力学中的应用

分子动力学（MD）模拟是研究过程动力学方面的有力工具，因为它可以直接跟踪目标中原子的运动。由于在高性能计算环境中取得的巨大进展，MD模拟现在可以处理具有超过十亿个原子的大型系统。此外，机器学习（ML）方法已经迅速与高性能计算结合起来，并被应用于各种领域。

溶剂化自由能（SFE）对研究化学物质、生物反应以及药物的设计至关重要。因此张[[[8]](#endnote-8)]等人通过研究乙酸、乙酰胺、丙酮、苯、乙醇和甲胺这6个有机小分子，以证明其新颖的ML方法，该法结合附加模型来识别MD过程中遇到的新构型，并采用梯度提升算法，不断随着数据库的增加而更新ML模型，其过程如图1所示。新颖的ML方法采用了修正对称函数和功率谱两种输入方式，以及输出检查模型、均值聚类模型和高斯模型这三种识别新数据点的方式，在MD模拟期间，使用线性回归和修正对称函数的ML方法应

该不断随着数据库的更新而进行更新，以获得稳健而准确的结果。梯度提升算法使更新QM/MM/ML模型的过程更加高效，将这一方法应用于各种复杂的ML模型，如高维NN和高斯核，将有助于利用基于ML的模拟探索新化学。

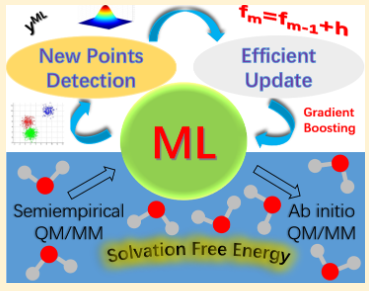


图1 用量子力学/分子力学和机器学习模型计算溶剂化自由能[8]

Fig. 1 Calculation of solvation free energy with quantum mechanics/molecular mechanics and machine learning models

张[[[9]](#endnote-9)]等人的团队提出了基于真实化学空间的SIC-BCPE通用模型，设计了四种设计参数（侧链阴离子直径、侧链长度、电负性基团电荷量与离子分数）以映射实验中可以调控的变量。其团队在合理的设计参数取值范围内进行了MD模拟采样并计算出对应的离子电导率。他们提出的MD模拟与机器学习结合的计算辅助设计方法可以被进一步拓展，用于设计各种结构更复杂，功能更全面的聚电解质材料。

液相酸催化反应的速率对溶剂成分非常敏感，而确定合适的混合溶剂在理论和实验上都具有挑战性。Alex K Chew课题组[[[10]](#endnote-10)]的研究表明，反应物-溶剂环境的复杂原子构型经典分子动力学模拟生成的反应物-溶剂环境的复杂原子配置，可通过三维卷积神经网络准确预测模型生物质化合物的布氏酸催化反应率。其3D CNN 的结构、训练和性能的示意图如2所示。他们开发的这种三维卷积神经网络，称之为SolventNet[[[11]](#endnote-11)]，利用实验反应数据和相应的分子动力学模拟数据对其进行训练，以预测酸催化反应速率。相应的分子动力学模拟数据来预测酸催化反应速率。共溶剂混合物中的七种生物质衍生含氧化合物的实验反应数据和相应的分子动力学模拟数据来预测反应速率。Alex K. Chew课题组的研究表明，SolventNet可以预测额外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。这种机器学习与分子动力学的结合，实现了溶剂系统快速、高通量的筛选，并确定改进的生物质转化条件。

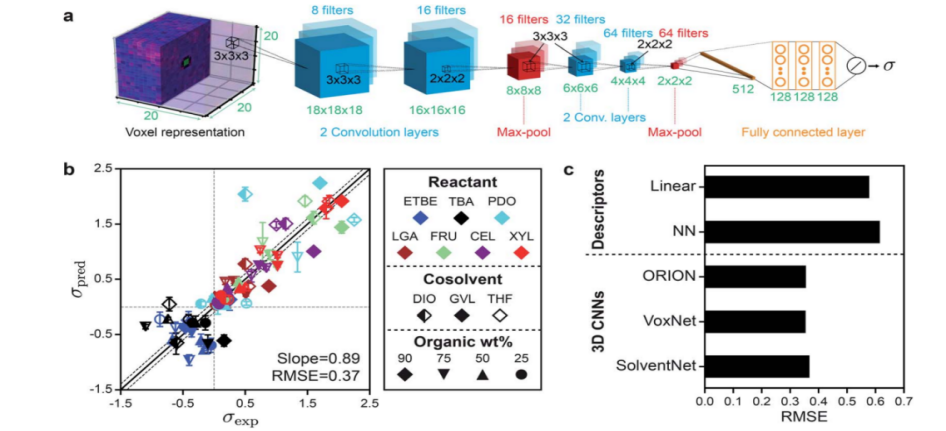


图2 3D CNN 的结构、训练和性能[11]

Fig2 Architecture, training, and performance of 3D CNNs.

近年来，使用最大似然法加速旨在理解物质现象的计算技术出现了激增。ML已被用于预测参数、在材料模拟中生成配置以及对材料属性进行分类。例如，傅[[[12]](#endnote-12)]等人采用人工神经网络选择有效的更新来加速经典伊辛自旋模型在与相变相关的临界参数附近的蒙特卡罗模拟。鉴于集成ML层以增强作为web应用部署的科学仿真框架的性能的探索相对较少，因此[Kadupitiya](https://webofscience.clarivate.cn/wos/author/record/29098598)[[[13]](#endnote-13)]等人开发了一个用于模拟纳米尺度限制中的离子的框架(此处称为纳米限制框架),该框架能够对以不同离子属性(例如，离子化合价)和溶液条件(例如，离子浓度和表面间分离)为特征的离子自组装进行系统研究。其过程如图3所示，该框架阐明了能够决定离子分布的离子相关性和空间效应的微观机制，因此被广泛被用于提取由平面和球面限制的电解质溶液中的离子结构。

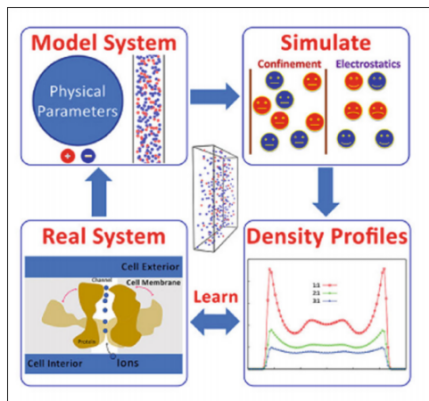


图3 集成ML层与部署在纳米约束框架上的GUI[13]

Fig. 3 Integration of the ML layer with a GUI deployed on a nanoconstrained framework.

分子性质预测是一项基础且重要的研究，准确预测分子性质在化学信息学的各个应用领域中都受到广泛关注，例如工业应用[[[14]](#endnote-14)]、环境工程[[[15]](#endnote-15)]及药物设计[[[16]](#endnote-16)]等。贾等人使用分子动力学模拟方法从微观层面对微波与超临界态水相互作用过程进行信息获取，从分子、原子层面探讨了SCW体系的温升特性、动力学性质；通过微观特性分析反映分子宏观理化性质。其过程如图4[[[17]](#endnote-17)]所示，大致分为三类：首先，通过分子动力学模拟实现微波与超临界水相互作用过程的微观信息获取。其次，使用e-dragon分子描述符作输入变量，构建AdaBoost、XGBoost、LightGBM和CatBoost机器学习算法的定量结构性质模型预水溶性。最后，利用RDKit分子描述符作为输入参数，基于AdaBoost、XGBoost、LightGBM和CatBoost机器学习算法构建定量结构性质模型预测水合自由能。通过分子动力学和机器学习对水的相关性质的预测研究，可以为腐蚀性、爆炸性、高温、高压和微波辐照的极端环境下的分子性质预测提供理论参考，也能为应用于大规模数据量的分子性质准确预测提供一定的帮助。

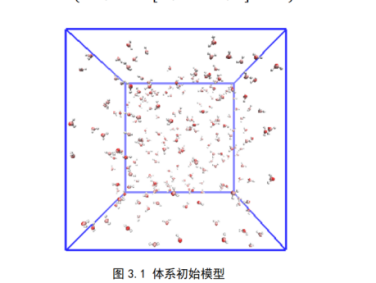
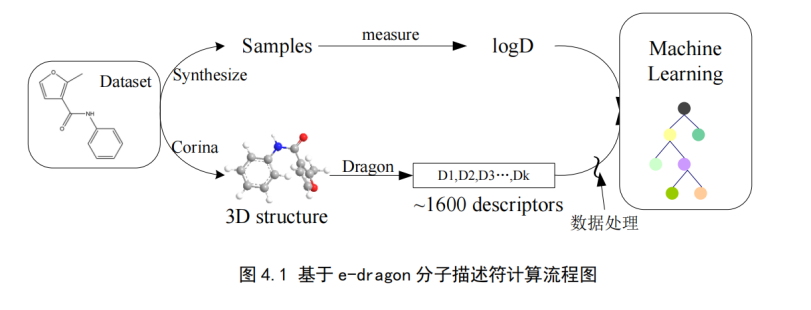
 

图4 初始模型与计算流程图[17]

Fig. 4 Initial model and computational flowchart

Teppei Fukuya[[[18]](#endnote-18)]等人在研究采用基于ML的方法自动分析MD模拟的原子构型。设计了一种3D-CNN架构，用于识别高温下各种材料双相体系中的固态和液态原子。3D-CNN的识别成功地达到了90%以上的准确率，不受晶体结构的影响，而传统技术（CNA）对同一体系的识别准确率最多只有50%。3D-CNN可将所有原子划分为更接近固态特征或液态特征的原子，并根据两种原子的边界统一定义固液界面。因此，3D-CNN可以非常清晰地提取固液界面的形态，包括原子尺度上的粗糙度，即尖锐界面。这是在他们研究中3D-CNN与CNA相比的优势所在，因为在CNA中，边界周围的原子往往被定义为未知结构的原子。

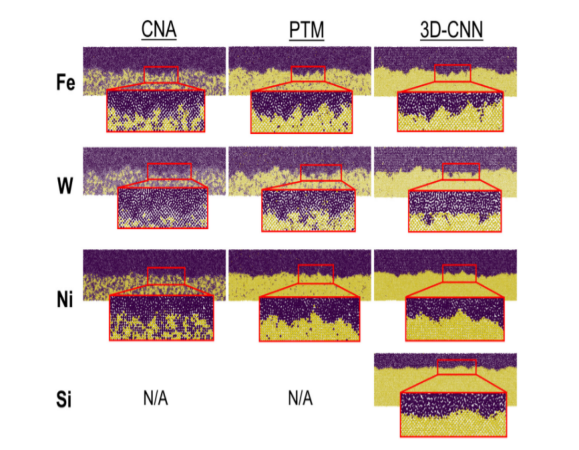
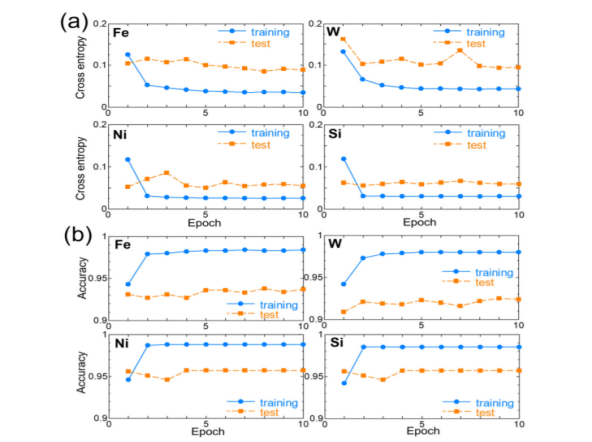
 

图5 三维卷积神经网络的训练和测试[18]

Fig. 5 Training and testing of three-dimensional convolutional neural network

机器学习是建立定量结构-活动关系(QSAR)模型的一种强有力的方法。实际上，ML在发展关于离子液体的各种性质或气体在离子液体中的溶解度(如熔点)、粘度、导电性、热分解温度、关键属性和毒性、H2S溶解度、二氧化碳溶解度和水溶性。分子指纹将离子液体的结构特征分解成二进制向量(0和1)，其适当地应用于虚拟筛选、相似性搜索和聚类的情况。对于分子描述符，通过半经验量子化学计算来计算离子液体的物理化学性质，提供各种量子化学和分子轨道描述符。离子液体的分子描述符结合最大似然法已被成功地用于开发QSAR模型[[[19]](#endnote-19)]。例如，Venkatraman和Alsberg[[[20]](#endnote-20)]基于半经验推导的几何和电荷分子描述符，报告了不同条件下185种离子液体中CO2捕获的ML预测，并获得了与现有实验结果的良好一致。熔点、pKa和离子液体的热分解温度和在离子液体中的水溶性也可以用分子描述符很好地预测。

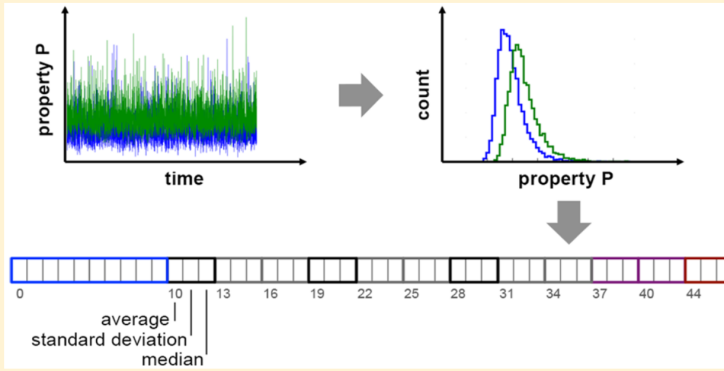


图6 研究策略示意图[19]

Fig. 6 Schematic of the general research strategy

全原子模拟已经证明了基于半经验原子间势的数十亿原子的可扩展性，能够以前所未有的细节揭示材料微观结构的复杂性。然而，原子间势的可转移性破坏了这种方法的基础，仍然受到半经验原子间势的可转移性。因此，彭[[[21]](#endnote-21)]等人提出了一种动态多尺度分子动力学（MD）模拟方法，通过混合经典和机器学习（ML）势，进行高精度和高效的原子间力场计算，他们证明了通过混合经典学习势和机器学习势，在计算原子间力场时可以同时实现较高的精度和效率。图7显示了多晶聚集体的多尺度分子动力学结构的示意图。ML势被用来描述高度扭曲区域中的原子相互作用。采用动态方法，评估MD模拟过程中演化微观结构的中心对称参数，并据此修改ML势所描述的高度畸变区域。在接近完美或完美晶格中的原子力场计算仍然坚持快速的EAM势[[[22]](#endnote-22)]，它精确地捕捉了长程弹性相互作用。

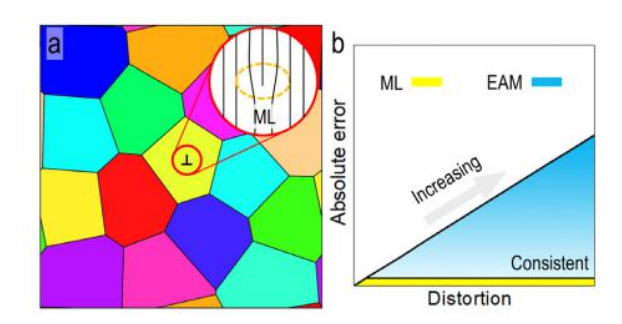


图7 多晶聚集体的多尺度分子动力学结构的示意图[21]

Fig. 7 Schematic representation of the multiscale molecular dynamics structure of polycrystalline aggregates

**2.2 机器学习在从头算分子动力学中的应用**

与经典分子动力学不同，从头算分子动力学(ab initio molecular dynamics, AIMD)模拟不需要提供力场参数，只需要提供原子初始结构，就能根据电子波函数正交化产生的虚拟力，求解牛顿运动方程，从而可以很好地解决分子动力学无法对化学键的断裂描述的问题。计算模拟是学术研究探索和开发的有力工具，基于完善的力场，可以更经济有效地估算材料的内在潜力，并找出这些行为背后的机理。例如Patrick Rowe[[[23]](#endnote-23)]等人探索出了一种训练模式，该模式的结构是由经典分子动力学生成，作为它们主动学习模式的输入，以最大限度地减少昂贵的从头计算的需要，按照主动学习工作流程，Patrick Rowe等人获得了如图8所示的六种不同水相体系的CNNP模型。

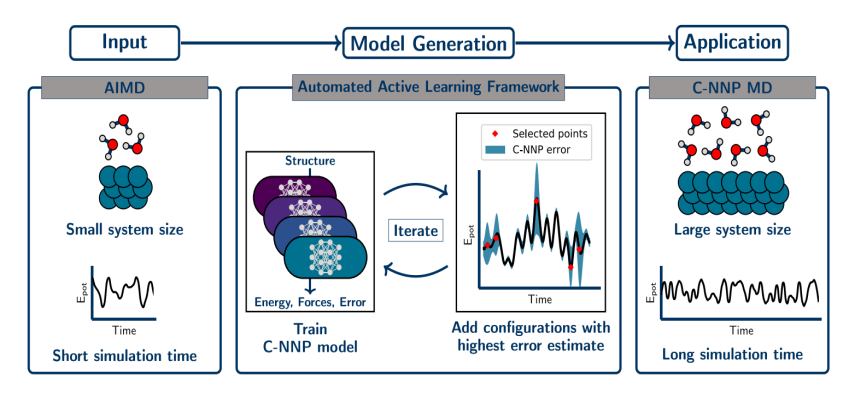
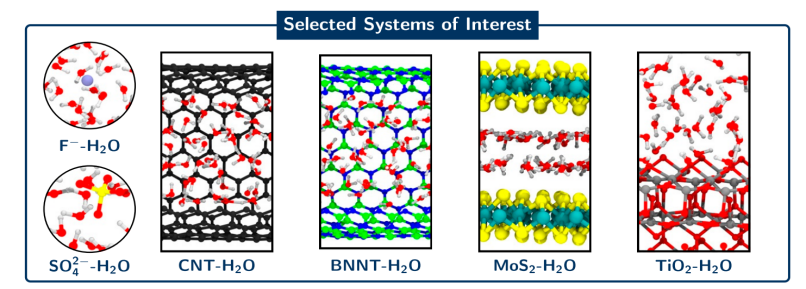
 

图8 不同水相体系的CNNP模型[23]

Fig.8 CNNP modeling of different aqueous phase systems

计算工具与AIMD的结合已被应用于了解溶剂对化学反应性的影响，并指导高效生物质转化过程的溶剂混合物的设计[[[24]](#endnote-24)],如对液相酸催化的快速预测反应速率。例如Mellmer[[[25]](#endnote-25)]等人研究发现，由于酸性质子催化剂、反应物和带电过渡态周围的溶剂化环境的变化，降低了活化能，从而增加了生物质转化反应动力学。模拟结果表明，亲水反应物驱动混合溶剂环境中富水局部域的形成，优先溶剂化酸催化剂，并稳定随后的碳正离子样过渡态如图9所示。反应物周围的富集程度与酸催化反应性能直接相关。类似的研究也使用了从头算分子动力学模拟来了解溶剂环境如何改变关键的酸催化反应的反应动力学，比如将果糖转化为生成5-羟甲基糠醛[[[26]](#endnote-26)]（一种用于聚合物前体和运输燃料的平台化学品）。

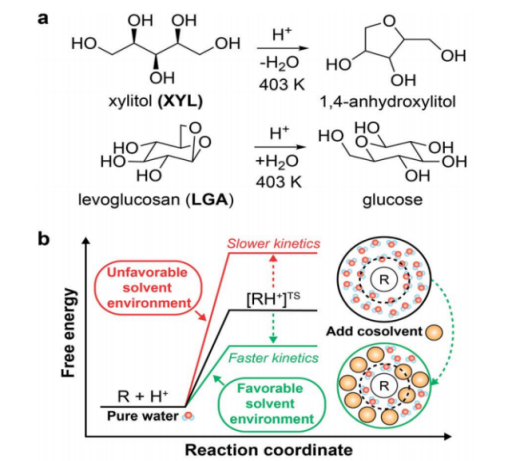
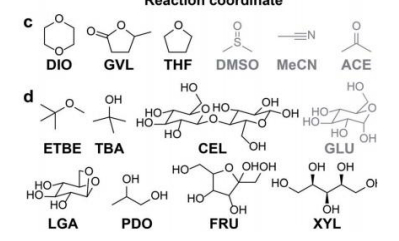
 

图9 溶剂对酸催化反应和模型系统的影响概述[25]

Fig.9 Overview of solvent effects on acid-catalyzed reactions and model systems.

将计算机模拟与深度学习相结合，基于AIMD和新定义的定向氢键布局算符，更有利于研究水分子之间的氢键交换这一典型的动力学过程。Jie[[[27]](#endnote-27)]等人通过递归神经网络(RNN)的模型，设计了一个双向长短期记忆(BLSTM)自动编码器(AE)，使尽可能多地重构输入序列并使用正序列训练AE，重建误差S(x)评估了AE对输入序列的重建，并使用一个最终分类器来区分互换和断裂过程之间的序列和正序列。使用该模型对9个含有N=的64个水分子的大体积水系统在280-360K温度下进行模拟。其结果表明随着温度的升高，交换和断裂过程的数量呈现“先上升，后下降”的趋势，每组交换和破裂过程之间的相对比值几乎不受温度影响，比例大约为1:4(如图10所示)。

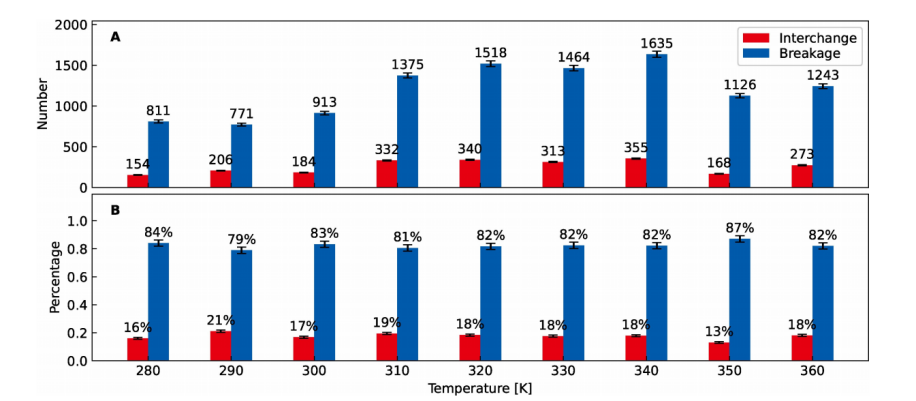


图10 基于 RNN 的分类器在不同温度下确定的互换和断裂过程的数量（A）和比例（B）[27]

Fig.10 The number (A) and proportion (B) of interchange and breakage processes determined by the RNN-based classifier at different temperatures.

Joseph A. Morrone[[[28]](#endnote-28)]课题组对300K时的纯液态甲醇和液态甲醇中的过量质子进行了分子动力学模拟。其结构模型如图11所示，通过对甲醇中过量质子的模拟，他们提出了液体中阳离子缺陷结构扩散的新机制。特别地，他们还发现了在质子化液体中存在比纯液体中典型链长更长的缺陷链结构。

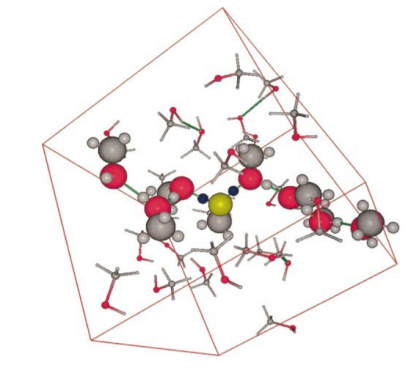
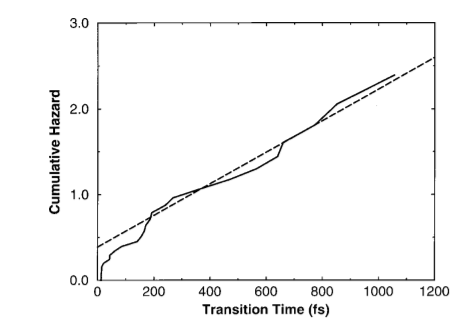
 

图11 液体甲醇中质子迁移率的从头算分子动力学研究[28]

Fig.11 ab initio molecular dynamics study of proton mobility in liquid methanol

1. Han and S. Luber[[[29]](#endnote-29)]提出了一种通过基于轨迹的机器学习即时构建力场的方法（TrajML）如图12所示。其理念是只需要核轨迹就可以在MD运行中进行分子构型的训练和预测，这样，与用于生成训练数据的AIMD相比，可以以更低的计算量获得更长的MD运行时间。TrajML[[[30]](#endnote-30)]可以提供为系统设计的专用力场，核结构信息的准确性原则上可以达到其训练方法的水平。这就减轻了对参数化程度不高或难以参数化的系统的核构型的情况。

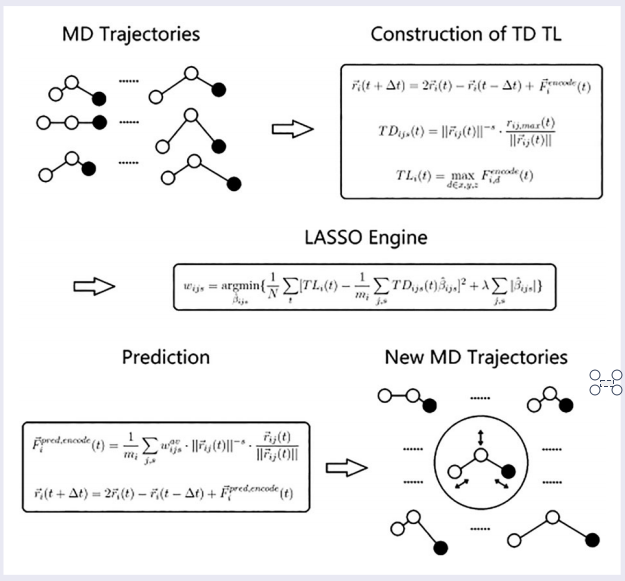


图12 基于轨迹的机器学习方法[29]

Fig.12 Trajectory-based machine learning approach

由于锕系元素的主要离子化学和离子半径的相似性,这就使得在液体中从结构上区分它们极为困难，Manh-Thuong[[[31]](#endnote-31)]等使用AIMD和机器学习来分析并研究了两个MS中的五个An系统中熔融盐中的锕系元素，模拟过程使用Scikit-learn在每个MS的AIMD轨迹的帧上执行了多类支持向量分类(SVC)，并使用具有8

个“优化”径向函数的离散变量表示径向基最大角度为6，为每个MS中的SOAP环境考虑了两种不同的界限:氯化钠为4.3英寸和6.3英寸，FLiBe为3.5英寸和5.5英寸。对于熔融NaCl和FLiBe中的三价锕系元素U3+、Pu3+、Cm3+、Cf3+和Fm3+系列，锕系元素-配体键在NaCl中具有更高程度的共价性(比在FLiBe中)，在FLiBe中具有更高程度的电离度，结果如图13所示。若只需考虑第一溶剂化壳层以外的原子，机器学习的分类模型可以在超过80%的置信度下区分化学上相似的锕系元素的原子环境。

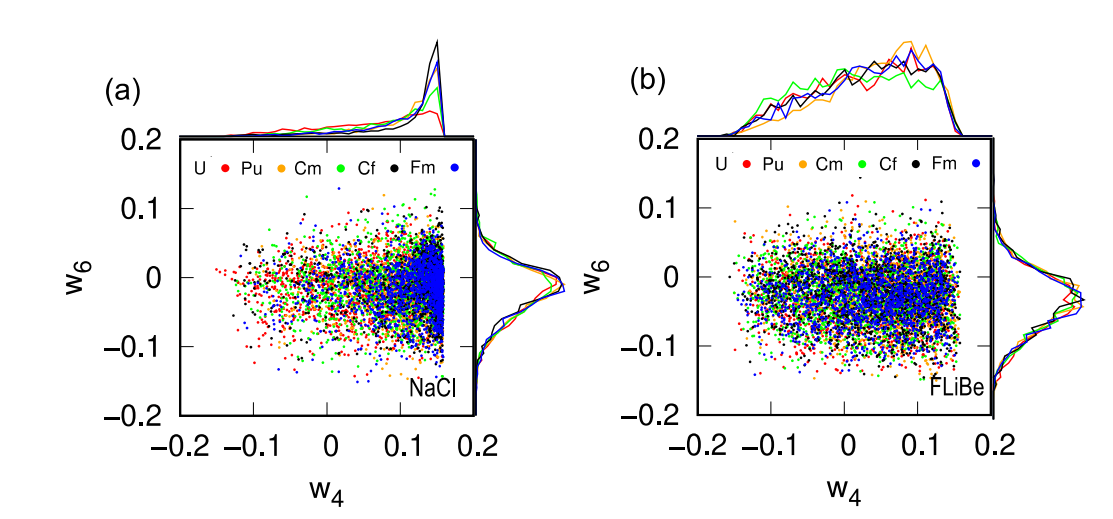


图13 w4和w6在氯化钠(a)和FLiBe (b)中的分布[31]

Fig.13 Distribution of w4 vs w6 in Nacl (a) and FLiBe (b).

闫[[[32]](#endnote-32)]等人主要通过三方面的工作来系统评估ILs的毒性机制，探索出一种环境友好型（抗菌性强，低细胞毒性）的离子液体设计方法，设计过程如图14所示。第一部分主要通过ILs毒性数据的全面搜集构建ILs的毒性数据库，开发一个ILs毒性数据库在线平台ILTox；第二部分基于机器学习的方法使用了MOE和Dragon描述符构建了有效的ILs-AChE抑制性EC50预测模型；第三部分主要利用提出的GPstack-RNN方法，生成环境友好型（抗菌性强、低细胞毒性）的ILs，并用实验方法评估准确性。

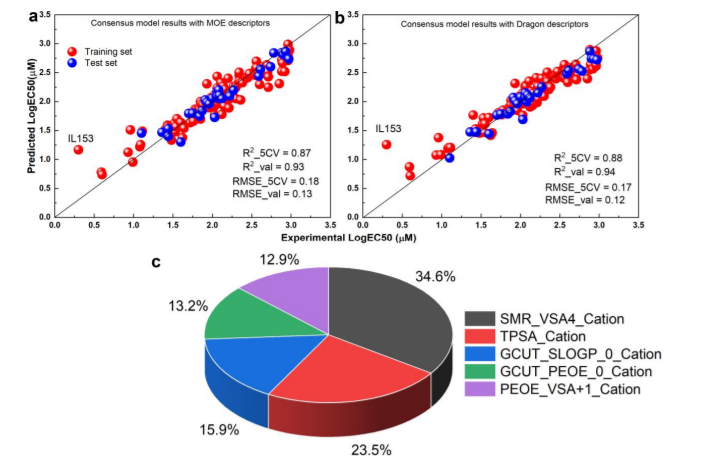
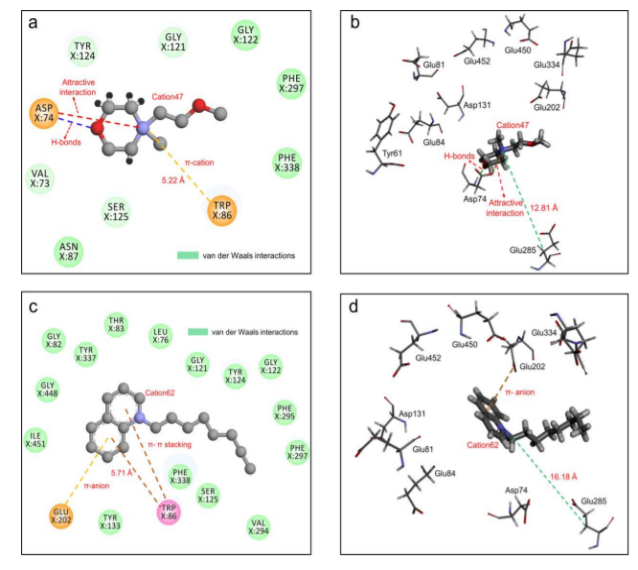
 

图14 基于机器学习和从头算分子动力学的离子液体的设计[32]

Fig.14 Design of ionic liquids based on machine learning and ab initio calculation of molecular dynamics

**三 结论与展望**

传统的分子动力学计算速度快，能一定程度上反映材料的稳定性，热力学性质等多种性质，但由于其使用的是较为简单的经验力场，计算精度有限；而AIMD将分子动力学与密度泛函理论(Density Functional Theory, DFT)相结合，将系统中的粒子划分成原子和电子，原子的质量大且运动速度较慢，可以用经典力学来处理，而电子的质量小且运动速度较快，可以用密度泛函理论来处理。

AIMD方法的提出可以很好地解决分子动力学无法对化学键的断裂描述的问题。由于AIMD考虑了电子相互作用的细节，并且使用量子力学方法对全体原子进行了统一考虑，因此计算精度大大提高，但也因此计算代价高昂，往往只能应用于数十个到数百个原子的小体系，且模拟时长最多到纳秒级。对于体系较大的系统，一般采用传统经典力场来建立模型，传统的科学方法使用一些近似值求解、分析计算，并用更高水平的计算或实验的参考数据验证所获得的结果。但是当所研究的系统过于复杂，仅靠现有知识无法建立可靠的模型时，这种方法并不是十分有效，并且力场往往在牺牲计算效率的条件下才能获得较高的计算精度。虽然可以有效解决传统方法无法解决的复杂问题，但在一定程度上可以有效地平衡计算效率和计算精度之间的制约关系，这就需要将AIMD方法的计算精度和经典力场的计算效率更好地融合，从而使应用于研究更大的分子系统体系成为可能。

虽然说对于某些体系，通过分子动力学模拟能够准确反映宏观物质的某些性质，但由于它是利用有限的盒子大小和有限的粒子数量来进行模拟的，因此它还是存在一定程度的偏差；这种方法一般只能用于少数种类的分子进行模拟计算并只能获得某几种性质，还不能保证能够准确预测大规模分子种类以及多种其他性质。需要对机器学习算法做更深入和详细的分析，在此基础上，提出一个通用的能准确预测多种分子性质的框架。

1. **参考文献：**

   [] BARTÓK A P, PAYNE M C, KONDOR R, et al. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons[J]. Physical review letters, 2010, 104(13): 136-403. [↑](#endnote-ref-1)
2. [] GUO J L, WANG Z G, WANG Y G, et al. A Review of Material Research and Development Methods Based on Computer Technology[J]. Frontiers of Data and Computing, 2021, 3(2): 120-132. [↑](#endnote-ref-2)
3. [] 申林, 贾璐阳, 汤典东, 等. 计算化学中的机器学习[J]. 中国科学:化学, 2022, 52(6): 858-868. [↑](#endnote-ref-3)
4. [] ROSENBERGER D, SMITH J S, GARCIA A E. Modeling of peptides with classical and novel machine learning force fields: A comparison[J]. The Journal of Physical Chemistry B, 2021, 125(14): 3598-3612. [↑](#endnote-ref-4)
5. [] CHMIELA S, TKATCHENKO A, SAUCEDA H E, et al. Machine learning of accurate energy-conserving molecular force fields[J]. Science advances, 2017, 3(5): 1160-3015. [↑](#endnote-ref-5)
6. [] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik and R. P. Adams, Convolutional Networks on Graphs for Learning Molecular Fingerprints, Advances in neural information processing systems, 2015,2224–2232. [↑](#endnote-ref-6)
7. [] N. E. Jackson, A. S. Bowen, L. W. Antony, M. A. Webb, V. Vishwanath and J. J. de Pablo, Sci. Adv., 2019, 5,762-769. [↑](#endnote-ref-7)
8. [] Z. Wu, B. Ramsundar, E. N. Feinberg, J. Gomes, C. Geniesse, A. S. Pappu, K. Leswing and V. Pande, Chem. Sci., 2018, 9, 513–530. [↑](#endnote-ref-8)
9. [] R. D. Singh, A. Mittal and R. K. Bhatia, Multimed. Tool. Appl., 2019, 78, 15951–15995. [↑](#endnote-ref-9)
10. [] J. Jim´enez, S. Doerr, G. Mart´ınez-Rosell, A. S. Rose and G. De Fabritiis, Bioinformatics, 2017, 33, 3036–3042. [↑](#endnote-ref-10)
11. [] H.E. Sauceda, S. Chmiela, I. Poltavsky, K.-R. Mller, A. Tkatchenko, Molecular Force Fields with Gradient-Domain Machine Learning: Construction and Application to Dynamics of Small Molecules with Coupled Cluster Forces, 2019. arXiv:arXiv:190.106-594. [↑](#endnote-ref-11)
12. [] Ai L, Zhou Y, Chen M. A reactive force field molecular dynamics simulation of the dynamic properties of hydrogen bonding in supercritical water [J]. Journal of Molecular Liquids, 2019, 276: 83-92. [↑](#endnote-ref-12)
13. [] Botu, V., Ramprasad, R.: Adaptive machine learning framework to accelerate ab initio molecular dynamics. Int. J. Quantum Chem.2015, 115(16), 1074–1083 [↑](#endnote-ref-13)
14. [] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. J. Chem. Phys. 1983, 79 (2), 926-989. [↑](#endnote-ref-14)
15. [] McKiernan, K. A.; Husic, B. E.; Pande, V. S. Modeling the Mechanism of CLN025 Beta Hairpin Formation. J. Chem. Phys. 2017, 147 (10), 104-107. [↑](#endnote-ref-15)
16. [] Sultan, M. M.; Kiss, G.; Shukla, D.; Pande, V. S. Automatic Selection of Order Parameters in the Analysis of Large Scale Molecular Dynamics Simulations. J. Chem. Theory Comput. 2014, 10 (12),2413–2421. [↑](#endnote-ref-16)
17. [] Ai L, Zhou Y, Chen M. A reactive force field molecular dynamics simulation of the dynamic properties of hydrogen bonding in supercritical water [J]. Journal of Molecular Liquids, 2019, 276: 83-92. [↑](#endnote-ref-17)
18. [] De Luna, P., Wei, J., Bengio, Y., Aspuru-Guzik, A. & Sargent, E. Use machine learning to find energy materials. Nature ,2017,552, 178-184 . [↑](#endnote-ref-18)
19. [] Chinta S, Rengaswamy R. Machine learning derived quantitative structure property relationship (QSPR) to predict drug solubility in binary solvent systems [J]. Industrial & Engineering Chemistry Research, 2019, 58(8): 3082-3092. [↑](#endnote-ref-19)
20. [] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, *Nat. Methods* 2020, 17, 261–272. [↑](#endnote-ref-20)
21. [] Pronobis, W., Tkatchenko, A. & Müller, K.-R. Manybody descriptors for predicting molecular properties with machine learning: analysis of pairwise and three-body interactions in molecules. J. Chem. Theory Comput,2018,14, 2991–3003 . [↑](#endnote-ref-21)
22. [] Wassermann, T. N. & Suhm, M. A. Ethanol monomers and dimers revisited: a Raman study of conformational preferences and argon nanocoating effects. J. Phys. Chem. A,2010,114, 8223–8233 . [↑](#endnote-ref-22)
23. [] Jia W, Wang H, Chen MH, Lu DH, Lin L, Car R, E W, Zhang LF. Pushing the limit of molecular dynamics with ab initio accuracy to 100 million atoms with machine learning. In Proceedings of SC 2020: International Conference for High Performance Computing, Networking, Storage and Analysis. Atlanta: IEEE Computer Society 2020. 2039-2450. [↑](#endnote-ref-23)
24. [] Quaranta, V.; Behler, J. r.; Hellström, M. Structure and dynamics of the liquid−water/zinc-oxide interface from machine learning potential simulations. J. Phys. Chem. C 2019, 123 (2), 1293−1304. [↑](#endnote-ref-24)
25. [] Kang, J.; Noh, S. H.; Hwang, J.; Chun, H.; Kim, H.; Han, B. First-principles database driven computational neural network approach to the discovery of active ternary nanocatalysts for oxygen reduction reaction. Phys. Chem. Chem. Phys. 2018, 20 (38), 24539− 24544. [↑](#endnote-ref-25)
26. [] Van Duin, A. C.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: a reactive force field for hydrocarbons. J. Phys. Chem. A 2001, 105 (41), 9396−9409. [↑](#endnote-ref-26)
27. [] H¨ase, F., Kreisbeck, C., Aspuru-Guzik, A.: Machine learning for quantum dynamics: deep learning of excitation energy transfer properties. Chem. Sci.2017, 8(12), 8419–8426. [↑](#endnote-ref-27)
28. [] Smith, D.G.A.; Burns, L.A.; Patkowski, K.; Sherrill, C.D. Revised Damping Parameters for the D3 Dispersion Correction to Density Functional Theory. J. Phys. Chem. Lett. 2016, 7, 2197–2203. [↑](#endnote-ref-28)
29. [] Brehm, M.; Thomas, M.; Gehrke, S.; Kirchner, B. TRAVIS—A free analyzer for trajectories from molecular simulation. J. Chem. Phys. 2020, 1521-4105. [↑](#endnote-ref-29)
30. [] Brehm, M.; Kirchner, B. TRAVIS—A free Analyzer and Visualizer for Monte Carlo and Molecular Dynamics Trajectories. J. Chem. Inf. Model. 2011, 51, 2007–2023. [↑](#endnote-ref-30)
31. [] Ai L, Zhou Y, Chen M. A reactive force field molecular dynamics simulation of the dynamic properties of hydrogen bonding in supercritical water [J]. Journal of Molecular Liquids, 2019, 276: 83-92. [↑](#endnote-ref-31)
32. [] Saielli, G.; Wang, Y. Role of the Electrostatic Interactions in the Stabilization of Ionic Liquid Crystals: Insights from Coarse-Grained MD Simulations of an Imidazolium Model. J. Phys. Chem. B 2016, 120, 9152–9160. [↑](#endnote-ref-32)