

文章编号: 1006-9941(2024)04-0408-14

# 人工智能辅助含能分子设计的应用与展望

刘锐<sup>1,2</sup>, 刘建<sup>1</sup>, 唐岳川<sup>1</sup>, 张朝阳<sup>1</sup>, 黄静<sup>1</sup>, 黄鑫<sup>1</sup>

(1. 中国工程物理研究院化工材料研究所, 四川 绵阳 621999; 2. 西南石油大学化学化工学院, 四川 成都 610500)

**摘要:** 含能分子研发面临多重挑战, 传统“试错法”效率低下, 计算机辅助分子设计的出现改变了研发模式。本综述回顾了含能分子设计的发展历程, 介绍了计算机辅助含能分子设计的研究现状, 并概述了人工智能技术(AI)在性质预测、分子生成、合成路线和反应条件预测等多个设计环节的最新进展, 讨论了当前含能分子设计模式与其他材料设计方法的差距, 思考差距产生的原因, 并对未来AI辅助含能分子设计的发展方向提出展望。研究发现, AI在含能分子性能预测和分子生成等方面已经有了应用, 但在合成路径规划和反应条件优化等环节的应用仍有待进一步探索, 应用前景巨大。通过数据增强、迁移学习或高通量计算有望能够解决含能分子数据薄弱的问题; 加强AI辅助含能分子合成路线与反应条件探索有望贯通“设计→评估→制备→验证”全流程自动化分子设计模式。AI辅助含能分子设计为提升含能分子设计水平提供新的可能性, 有助提升含能分子研发效率。

**关键词:** 含能分子; 分子设计; 人工智能; 机器学习; 定量构效关系

中图分类号: TJ55; O62

文献标志码: A

DOI: 10.11943/CJEM2023226

## 0 引言

含能分子可被特定的外界刺激触发剧烈的氧化-还原反应, 并通过燃烧和爆炸快速释放其储存的化学能, 广泛用于民爆、航天、军工等领域<sup>[1-2]</sup>。含能分子的微小改进都会带来武器性能的大幅提升, 因此各军事强国都在不遗余力探索性能更优异的新型含能分子<sup>[3]</sup>。然而, 含能分子更新换代的周期极为漫长, 以“钝感”著称的1, 3, 5-Triamino-2, 4, 6-trinitrobenzene (TATB)<sup>[4]</sup>和以“高能”著称的2, 4, 6, 8, 10, 12-Hexanitro-2, 4, 6, 8, 10, 12-Hexaazaisowurtzitane (CL-20)<sup>[5]</sup>分别于1887年和1987年首次被发现, 但迄今为止仍未发现综合性能超越它们的新含能分子。含能分子的探索极具挑战性, 在性能上要兼顾能量、感度、力学强度、相容性等, 在研制流程上要经历化学合

成、理化表征、爆轰实验、安全性实验等<sup>[6]</sup>。单纯采用“试错法”探索新型含能分子, 必将面临巨大工作量, 以更科学的方法发现并合成有潜力的含能分子是必然的选择。

计算机辅助分子设计的出现改变了含能分子的研制模式, 把单纯的实验室“试错法”转变为“设计→评估→制备→验证”<sup>[7]</sup>。计算机辅助分子设计基于计算方法、预测模型等, 预先评估含能分子的爆轰性能、化学稳定性或者与二者密切相关的其它性能。因此, 它可以为实验室研究筛选符合要求的目标分子, 有效地避免在无价值的分子上付出经济成本和时间成本<sup>[8]</sup>。计算机辅助分子设计使含能分子的研发效率得到了提升, 并使一些优秀的含能分子被成功发现, 如1, 1-diamino-2, 2-dinitroethylene (FOX-7)<sup>[9]</sup>、2, 6-diamino-3, 5-dinitro-1, 4-pyrazine 1-oxide (LLM-105)<sup>[10]</sup>、CL-20<sup>[5]</sup>、1, 3, 6, 8-tetraoxide (TTTO)<sup>[11]</sup>、octanitrocubane (ONC)<sup>[12]</sup>等。随着新含能分子的不断涌现, 当前的分子设计的不足开始逐步显现, 包括性能的预测精度、分子结构的产生方式、合成途径的评估等方面。

近年来, 人工智能(Artificial Intelligence, AI)开始在材料领域崭露头角, 成为分子设计的主要发展趋势。AI通过机器学习算法从海量数据中挖掘出潜藏的化

收稿日期: 2023-10-29; 修回日期: 2023-12-07

网络出版日期: 2024-01-08

基金项目: 国家自然科学基金(22173086, 22203081, 22305234)

作者简介: 刘锐(1999-), 女, 博士研究生, 主要从事计算材料学研究。e-mail: ruiliu0501@163.com

通信联系人: 刘建(1977-), 男, 副研究员, 长期从事人工智能辅助含能材料设计研究。e-mail: liujian-12@caep.cn

引用本文: 刘锐, 刘建, 唐岳川, 等. 人工智能辅助含能分子设计的应用与展望[J]. 含能材料, 2024, 32(4):408-421.

LIU Rui, LIU Jian, TANG Yue-chuan, et al. Applications and Prospects of AI-assisted Design of Energetic Molecules[J]. *Chinese Journal of Energetic Materials (Hanneng Cailiao)*, 2024, 32(4):408-421.

学规则,建立用于性能预测<sup>[13]</sup>、分子生成<sup>[14]</sup>、合成路线设计<sup>[15]</sup>甚至化学反应条件预测<sup>[16]</sup>的模型。高通量计算可以生成大规模的计算数据,为机器学习提供了数据支撑。在含能分子设计领域,已有将机器学习用于性能预测的相关报道<sup>[17-22]</sup>,但在分子结构搜索与制备途径等环节发展缓慢。AI在含能分子设计领域的应用存在较大的发展空间,有望为含能分子设计的不足之处提供解决方案,有力提升含能分子的研发效率。当前的含能材料设计已有关于采用高通量计算<sup>[23]</sup>和机器学习<sup>[21-22,24-25]</sup>的研究被报道,具备进一步发展AI辅助含能分子设计的条件。为此,研究回顾了含能分子设计方法的发展与现状,概述了AI在性能预测、分子生成、合成路线及反应条件预测等材料设计领域的最新进展,通过对比含能分子与材料设计方法的差距,思考差距产生的原因,有望借鉴AI辅助材料设计的经验与方法进行含能分子设计,有效提升含能化合物的研制效率。

## 1 含能分子设计方法的发展与现状

### 1.1 含能分子设计的发展历程

含能分子设计可以追溯到二十世纪50年代,研究人员开始理论推导极高的压力和温度下爆炸产物气体的行为,并提出了定量评估爆轰性能的方程,如BKW方程<sup>[26]</sup>、JCZ方程<sup>[27]</sup>、VLW方程<sup>[28]</sup>等。这些方程为评估爆轰性能奠定了理论基础,其中基于BKW和JCZ方程的商业软件EXPLO5<sup>[29]</sup>精度媲美实验值,如爆速预测误差一般小于5%<sup>[29]</sup>。二十世纪80年代,随着计算机硬件和算法的进步,计算机辅助分子设计在含能材料的研究中逐步获得广泛的应用,一大批新型含能分子在此基础上被发现<sup>[5,9-12]</sup>。

含能分子的计算机辅助设计需要综合考虑多方面性能的平衡,需要在量子计算的基础上,通过多步骤完成,如图1所示,主要包括:构建分子结构、量子化学计

算及分析、QSPR模型<sup>[17-22,30]</sup>预测以及爆轰性能预测。

早期的含能分子的计算机辅助设计可以概括为“画加算”<sup>[7]</sup>,主要依靠设计者逐一构建分子结构,分别建立针对不同性质的计算方案完成分子设计。“画加算”方法基本都能达到分子设计的预期性能<sup>[31]</sup>,但贯穿分子设计流程依赖于频繁的人工交互,分子筛选的通量较小,一般处理的对象仅限于数十个分子。研究者需要具备扎实的理论知识和丰富的研究经验,需要敏锐的直觉来保证分子设计的成功率,研究结果往往具有不确定性。

近年来,随着材料基因组<sup>[32]</sup>的提出和超级计算机的发展,高通量计算技术开始用于含能分子设计<sup>[33-36]</sup>。高通量计算通过自动化流程和高效的并行计算,可以在短时间内从海量候选结构中快速筛选出新的目标化合物。高通量计算将繁琐的程序与方法集成到易于使用的系统中,可以面向由“海量”(最高可高达数百万)的分子结构组成的搜索空间进行高通量设计和自动化筛选,可处理的任务量较传统的“画加算”模式高出好几个数量级,可产生海量的化学结构与性能数据<sup>[37]</sup>。通过这种高通量筛选,研究模式从“经验驱动的实验”向“理论预测先导、实验验证”的新模式转变,极大提升了设计效率。例如,Wen等<sup>[36]</sup>通过结合组合库设计与高通量计算的方式用于CHONF含能分子的设计。

在高通量计算平台方面,国际上已经出现了多种面向材料设计的在线平台,例如ASE<sup>[38]</sup>、AiiDA<sup>[39]</sup>、MatCloud<sup>[40]</sup>、ALKEMIE<sup>[41]</sup>、AFlow<sup>[42]</sup>、EM-Studio<sup>[43]</sup>等,其中的EM-Studio是专为含能分子设计研发的高通量计算平台。这些平台实现了建模、计算和数据分析的完全自动化,基本上都提供了超算资源的开放访问权限,个别平台的用户群体已具有相当规模,形成了海量的计算数据,为新材料设计提供了强大的支撑。

海量的高通量计算数据与长期积累的实验数据结合,推动了AI辅助分子设计模式的发展<sup>[13,44]</sup>。在材料

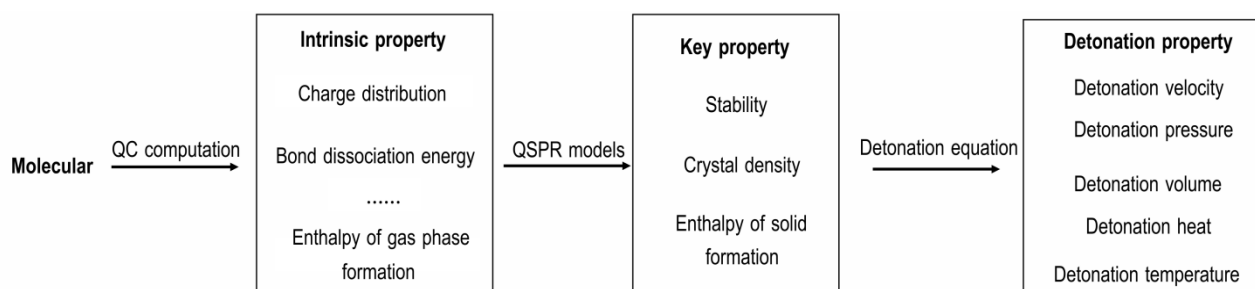


图1 含能分子的计算机辅助设计技术路线

Fig.1 Route of computer-aided material design for energetic compounds

科学和药物分子设计领域,已经有大量关于利用AI技术进行材料性能预测、制备路径预测、化合物结构设计以及各类研究数据的自动采集等方面的研究报道<sup>[45]</sup>。但是由于含能分子的实验数据较为稀少,且高通量计算技术存在限制,这阻碍了AI辅助含能分子设计的开展,目前仅有少量工作采用量子化学和机器学习进行含能分子的性能预测和分子筛选。

## 1.2 含能分子设计的研究进展

含能分子的设计、制备和使用过程中,需要考虑许多关键性能,比如生成焓、升华热和静电势等基本物化性能,以及反映稳定性的键解离能(BDE)和反映爆轰性能的各项参数等<sup>[46]</sup>,可以通过量子化学计算或机器学习等定量构效关系模型(QSPR)预测得到。

一类是基于特征工程的QSPR模型,通常先进行量子化学计算,提取各类结构描述符,然后经过相关性分析、递归特征消除等方法进行特征选择,最后建立反映结构-性能关系的QSPR模型。例如Liu等<sup>[17]</sup>采用支持向量回归(SVR)、随机森林(RF)等机器学习算法用

于构建结构与升华热之间的关系,技术路线如图2所示。Kang等<sup>[18]</sup>利用RF构建爆热模型,从PubChem数据库中筛选出262种潜在的含能化合物。该类模型的特征提取需要人为参与,根据设计领域提取所需特征,这就要求研究者具备一定的化学知识,其模型的准确率与特征工程息息相关。

随着深度学习算法的发展,另一类基于机器学习的QSPR模型开始用于预测含能分子的性能(图3)。Casey等<sup>[19]</sup>基于卷积神经网络(CNN)预测了含能分子8种主要性质,模型的决定系数( $R^2$ )均达到0.95以上,具有出色的模型预测能力。徐雅斌等<sup>[20]</sup>以库伦矩阵为输入,基于CNN和双向长短期记忆网络的融合模型实现了对含能分子生成焓的预测,模型平均绝对误差(MAE)和均方根误差(RMSE)分别为 $0.0374 \text{ kJ}\cdot\text{mol}^{-1}$ 和 $0.0541 \text{ kJ}\cdot\text{mol}^{-1}$ 。Yang等<sup>[21]</sup>采用支持向量机(SVM)、RF及图神经网络(GNN)来建立含能分子结构和密度之间的映射关系,其中GNN模型的预测精度MAE、RMSE和 $R^2$ 分别为0.0398、0.0493和0.9491,

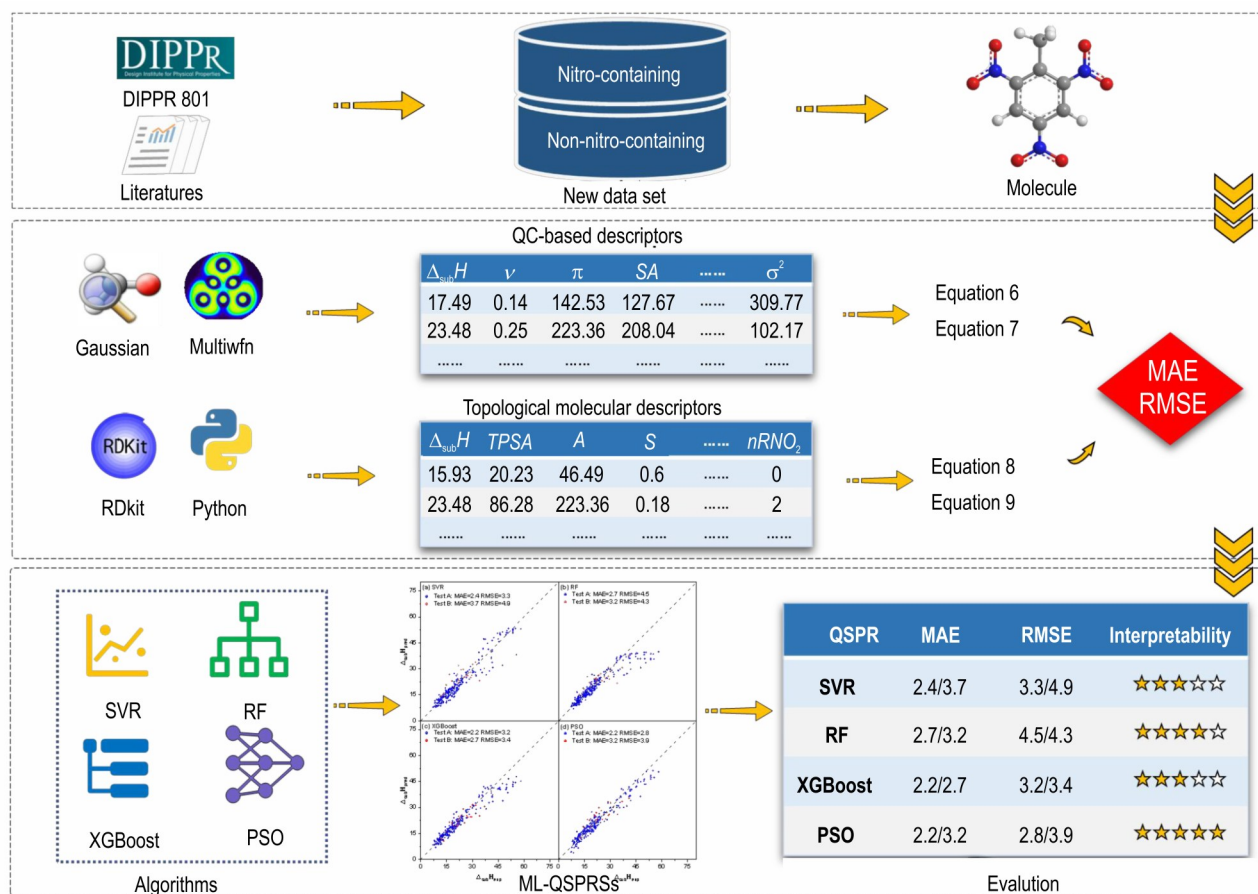


图2 基于特征工程的方法进行含能化合物升华焓预测的QSPR模型<sup>[17]</sup>

Fig.2 QSPR models for prediction of sublimation enthalpy of energetic compounds based on feature engineering method<sup>[17]</sup>

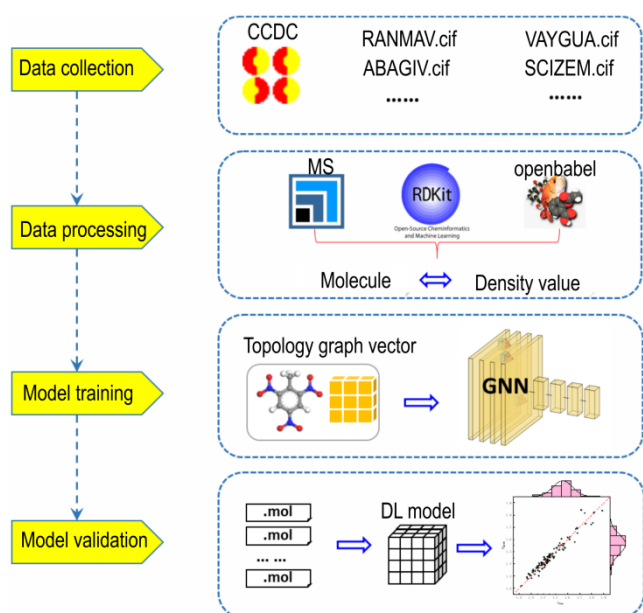


图3 基于深度学习算法预测含能化合物晶体密度的QSPR模型<sup>[21]</sup>

Fig.3 QSPR model for predicting crystal density of energetic compounds based on deep learning algorithm<sup>[21]</sup>

对独立测试集中 88% 的化合物的预测误差小于 5%。Nguyen<sup>[22]</sup>利用消息传递神经网络预测高能炸药的晶体密度,模型的  $R^2$  和 RMSE 分别为 0.914 和  $0.044 \text{ g}\cdot\text{cm}^{-3}$ 。相比于基于特征工程的 QSPR 模型,这类方法可以从大规模的数据中自动学习特征来构建模型,因此对数据量要求很高,其性能在很大程度上取决于输入数据(即样本)的质量和数量<sup>[47]</sup>,相对而言对计算资源消耗较多,对 GPU 要求高。

性能预测模型对快速准确预测含能分子的关键性能具有重要作用,推动了含能分子的高通量设计模式

的发展。利用计算机从大规模化合物库中快速筛选潜在分子,大幅减少进入化学实验的化合物数量,可显著提高材料设计的效率和成功率。高通量虚拟筛选的核心在于构建合理的筛选空间,这决定了后续分子设计的质量。筛选空间的构建源于分子设计,常用的方法有完全枚举法、深度学习生成模型等<sup>[48]</sup>。

完全枚举法<sup>[49]</sup>是当前常用的有机分子虚拟筛选空间构建方法之一,即导入初始的分子骨架为模板,枚举所有可能的分子取代位点并用取代基进行排列和组合,以快速生成数千个候选分子构成分子库。在含能分子领域已有类似设计方案<sup>[33,50-52]</sup>,例如 Xie 等<sup>[50]</sup>枚举了 88 个母环和 13 个高能取代基所有可能的组合,以单取代和双取代的方式得到了 84083 个样本组成的搜索空间,如图 4 所示。Song 等<sup>[51]</sup>以稠合的双杂环主链为研究对象,采用机器学习模型预测密度、爆速、爆压及分解温度等性质,结合实验验证发现了有前途的高能分子 ICM-10。这类方法依据化合价规则对生成分子的化学合理性进行简单判定。尽管最终生成的分子都符合化合价规则,但未从热力学和动力学层面上判断分子的化学可行性<sup>[49]</sup>,导致虚拟筛选对象可能存在结构不合理、分子本身不稳定或者无法合成等情况。

深度学习生成模型的发展为快速设计出理想属性的含能分子提供了新的思路。生成模型可以捕获潜在的概率数据分布(例如化学结构集),以指定如何生成符合分布的新分子<sup>[53]</sup>。含能分子的分子生成研究也开始采用深度学习的生成模型,Li 等<sup>[54]</sup>针对 CHNO 高能分子提出的深度学习生成框架由三个 RNN 模型组

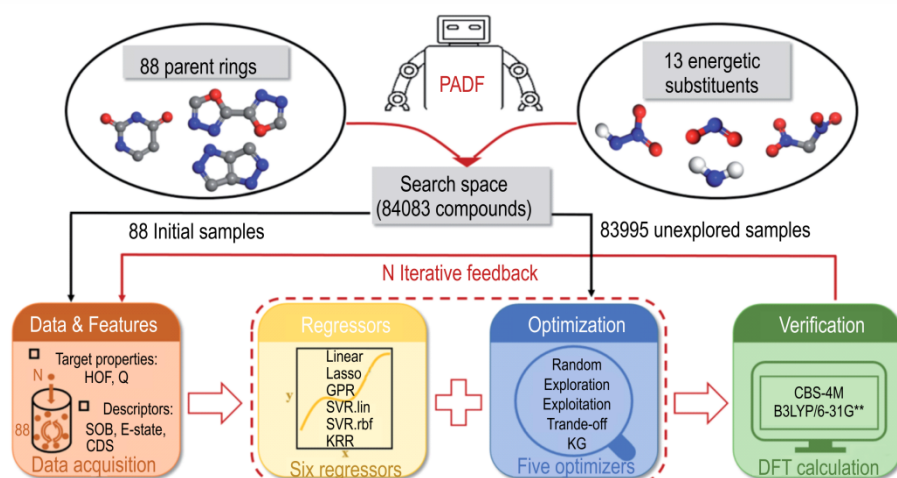


图4 基于完全枚举法构建虚拟筛选空间<sup>[50]</sup>

Fig.4 Virtual screening space constructed based on complete enumeration method<sup>[50]</sup>

成,分别用于预训练、生成以及预测,如图5所示。首先,对303个高能分子采用SMILES枚举和片段洗牌的方式进行数据增强。预训练的RNN对增强后的数据集 $D_s$ 进行学习,通过迁移学习的方式将结果与预测模型和生成模型相关联以便提供更充分的结构信息来提高性能。最后,利用预测RNN用于评估生成式RNN产生的新分子的爆轰性能,经过高精度的量子化学计算筛选出35个性能优异的新分子。

Liu等<sup>[55]</sup>开发的EM-Studio中分子生成板块融合了深度生成模型和完全枚举法如图6所示。其中,生成模型基于RNN模型实现分子生成;完全枚举法通过定义取代位点、枚举位点和取代基的搭配,经去除化学重复结构以及预优化分子结构实现分子生成。完全枚

举法能够覆盖更全面的分子范围,容易因虚拟搜索空间过于庞大而难以完全遍历造成筛选难度大;而深度学习生成模型可以更高效地处理大规模数据,针对性地生成具有特定属性的分子,但该方法对数据需求量大,需要对不同的生成任务进行适应性地参数调整及优化,以便提高分子生成的有效性。

总体而言,计算机辅助设计在含能分子的设计和评估环节已取得重要进展。现有QSPR模型可以较为可靠地预测含能分子的部分关键性能,为分子筛选提供评估标准。例如EXPLO5预测爆速的平均误差一般在5%以内,QSPR模型预测晶体密度平均绝对误差可以达到 $MAD < 0.4 \text{ g} \cdot \text{cm}^{-3}$ <sup>[21,56-58]</sup>,生成焓的预测平均误差可控制在 $10 \text{ kJ} \cdot \text{mol}^{-1}$ 以内<sup>[59-60]</sup>。此外,通过完全枚举

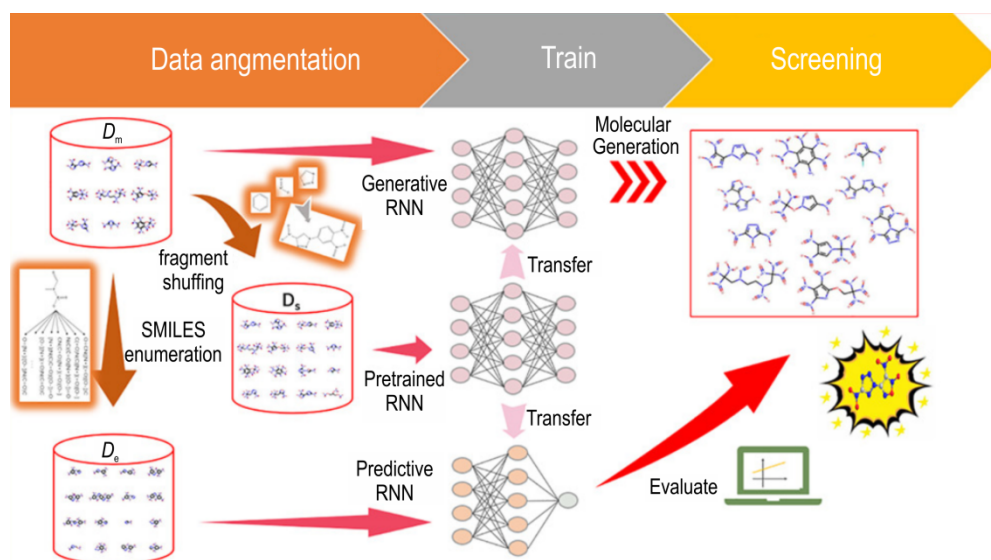


图5 基于RNN模型的含能分子设计工作路线<sup>[54]</sup>

Fig.5 Workflow for the design of energetic molecules based on RNN models<sup>[54]</sup>

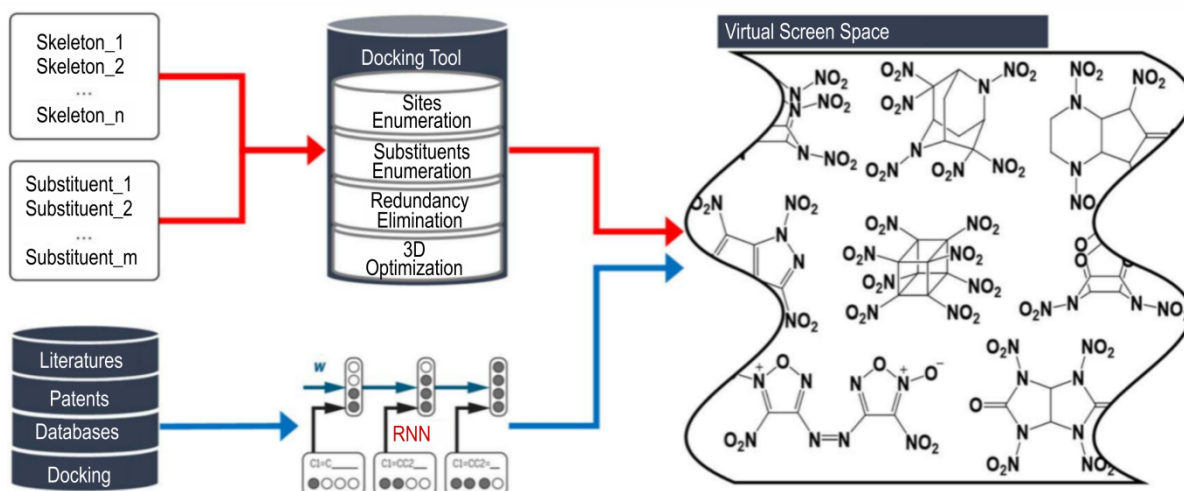


图6 通过片段对接及人工智能辅助生成分子的方案<sup>[55]</sup>

Fig.6 Schemes for generating molecules via fragment docking or AI-assistance<sup>[55]</sup>

法或基于深度学习的生成模型也被用于挖掘含能分子结构与性能之间的内在联系,并生成新的潜力分子结构。结合性能预测模型和高通量计算进行评估,含能分子逐步借助计算机辅助分子设计技术实现“设计-评估”的设计模式,为实验合成潜力分子提供了重要支撑。

## 2 AI辅助分子设计在材料领域的最新发展

近年来,AI在药物和有机光电材料等领域的应用已不限于简单的数据统计和性能预测,在分子生成、合成路线设计和反应条件预测等方面也取得了可喜的成绩。AI辅助分子设计的最新进展为含能分子设计面临的关键性能预测、筛选对象构造和合成路线设计等问题,提供了潜在的解决方案<sup>[61]</sup>。

### 2.1 性能预测

应用AI预测含能化合物性能方面已经有大量文献报道<sup>[17-22]</sup>,所采用的模型水平和预测精度已能与其他材料设计方法媲美。不同于其他材料类别,含能分子的数据稀缺使得深度学习模型的性能受限,模型的预测性能不佳。数据增强<sup>[62]</sup>、迁移学习<sup>[63-64]</sup>以及高通量计算<sup>[33]</sup>已被证明是解决深度学习数据集稀缺问题的有效方法,对提升预测模型的性能至关重要。其中,数据增强通过对原始数据进行各种变换来增强训练数据集的大小与质量,如采用SMILES枚举法<sup>[65-66]</sup>、片段

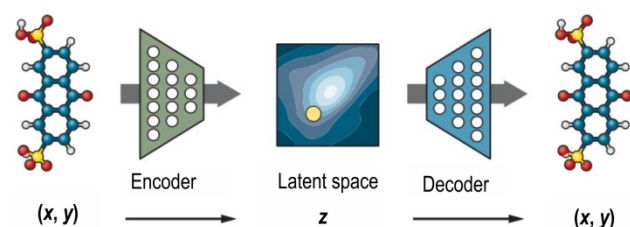
组装法<sup>[54]</sup>进行数据增强;迁移学习常作为一种优化策略用于生成具有所需性质的分子,在药物分子生成中得到广泛应用<sup>[67-69]</sup>;高通量计算通过计算手段补充数据样本以完善虚拟筛选空间,用于满足深度学习数据集需要<sup>[33,51,70]</sup>。通过学习其他材料设计数据增强或数据扩充的经验,对完善含能材料数据集具有重要的启发意义,有望进一步提升AI辅助含能分子性能预测模型的精度,提高材料筛选的效率。

### 2.2 分子生成

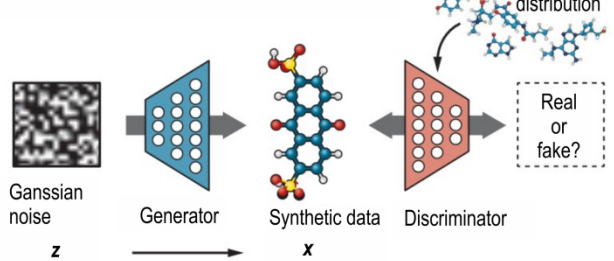
随着现代机器学习方法的进步与发展,AI辅助分子生成方式开始广泛用于分子生成研究<sup>[48]</sup>,促进了分子设计新模式的发展。在最近深度学习的突破下,新一代从头设计<sup>[14]</sup>算法出现,通过学习、分析及提取已知分子的理化性质或者结构特征,以优化为目标从头开始生成符合理想性质的新分子(或化学骨架)<sup>[71]</sup>。已有大量基于深度学习的生成模型在材料设计及药物设计领域的成功应用,如变分自动编码器(VAE)<sup>[72]</sup>、生成对抗网络(GAN)<sup>[73]</sup>、强化学习(RL)<sup>[74]</sup>、RNN<sup>[75]</sup>以及混合框架,对推动分子从头设计发挥着重要作用<sup>[14,72]</sup>,如图7所示。

VAE结合了自编码器和变分判断的思想,通过学习数据的潜在分布来生成新的数据样本,其生成的样本范围更接近训练样本的范围,可以处理以SMILES表

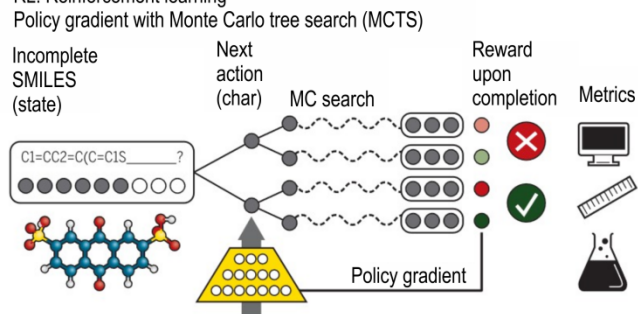
VAE: Variational autoencoders



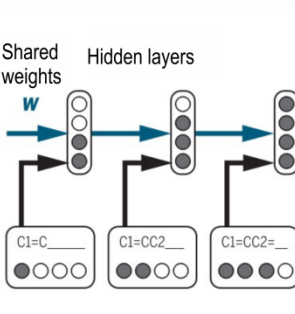
GAN: Generative adversarial networks



RL: Reinforcement learning



RNN: Recurrent neural network



Hybrid approaches

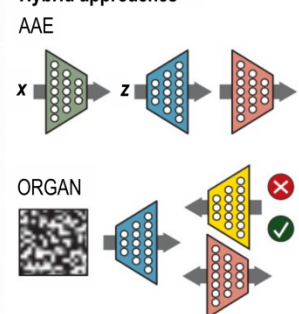


图7 生成模型中发现的几种架构的示意图<sup>[14]</sup>

Fig.7 Schematic representation of several architectures found in generative models<sup>[14]</sup>

示的分子式<sup>[76-80]</sup>、分子图<sup>[81]</sup>、图像<sup>[82]</sup>作为输入数据的生成任务。GAN由生成器和判别器组成的对抗性框架,通过两者相互竞争训练生成器生成与训练样本更接近的样本,在结晶多孔材料<sup>[83]</sup>、储氢材料<sup>[84]</sup>、药物分子<sup>[85-88]</sup>设计中得到应用。RL多作为一种优化生成分子特性的策略,通过与环境的交互反馈调整生成策略以生成所需的分子或材料,主要应用于药物分子的设计中<sup>[89-91]</sup>。RNN具有循环结构,能够捕捉分子 SMILES 序列中原子和键的时间依赖关系,从而获得语法和语义都正确的 SMILES 序列<sup>[92-93]</sup>。混合框架的使用可以充分发挥各类生成模型的优势,实现更高质量和更稳定的分子序列生成<sup>[14,94-96]</sup>,例如 GAN 的生成器-判别器结构提供了对抗训练,可以提升 RNN 的生成质量。

深度生成模型在材料设计和药物分子设计等领域的应用已取得显著成效。然而,针对含能分子的深度生成模型仍处于探索阶段,尚需经历一段漫长的发展过程。目前的挑战主要表现在分子属性相关特征的提取不够充分,导致难以生成具备特定优异性能的新材料;此外,含能分子数据的稀缺和获取困难,使得深度

学习缺乏大规模标注数据集,从而难以提升模型性能。未来的研究需要聚焦于构建高质量的含能材料数据集,为特征工程与生成模型的提升奠定数据基础。此外,需要设计能够综合考虑含能分子多项性能的生成模型框架,可以借助蒙特卡罗树搜索、遗传算法等优化算法来生成具有特定性能的含能分子。

### 2.3 合成路线设计

近年来,人们对利用 AI 技术进行逆合成预测的兴趣急剧上升,AI 驱动的逆合成预测会自动从现成的实验集中学习化学知识预测反应和逆合成路线<sup>[71,97]</sup>。逆合成预测最早由 Corey 和 Wipke<sup>[98]</sup>于 1969 年正式提出,引发了全合成革命,提高了合理性和系统性,其策略是设计一系列反应,以递归方式将目标化合物分解为更简单的结构单元,直到达到商业上可用的起始分子。一般来说,逆合成预测可分为基于模板和不基于模板的,如图 8 所示,采用逆合成预测可以有效解决复杂分子的合成,促进有机合成科学的发展<sup>[99]</sup>。

基于模板的逆反应预测根据模板定义的规则将产物分子转换成反应物分子(图 8a)。模板是指一组描

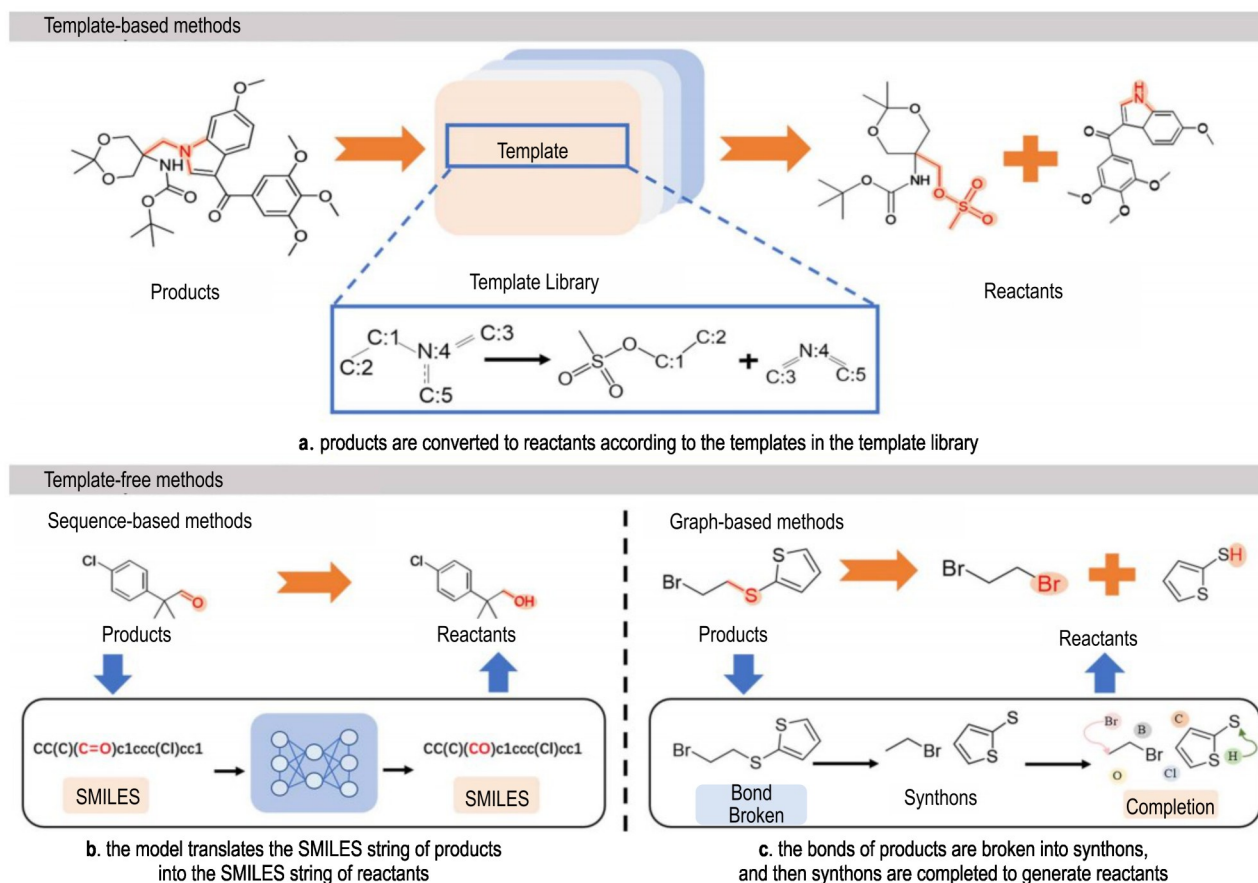


图 8 基于模板和无模板逆合成的基本流程<sup>[99]</sup>

Fig.8 The basic flow of template-based and template-free retrosynthesis<sup>[99]</sup>

述化学反应核心单元的反应规则<sup>[100]</sup>,可以从化学反应数据集中提取各种反应规则(或模板)。依据特定的逆合成反应模板,一些研究已经成功实现了合成路径的搜索,例如Segler、Coley等<sup>[100-106]</sup>的工作。他们利用指定的逆合成模板按照反应反向的顺序遍历产物分子,成功到搜索合成路线。

反应模板是有限的,无法涵盖足够多的反应类型。无模板的逆反应预测方法则克服了模板不足的问题,能够从零开始生成反应物,避免了对模板的依赖。在分子逆合成预测任务中,根据所采用的分子表示方法的不同,可以分为基于SMILES序列和基于分子图的两大类。基于SMILES序列的方法将逆合成建模为序列到序列的转换任务(图8b),代表性工作有Liu、Karpov等<sup>[107-109]</sup>的工作。基于SMILES的方法容易忽略分子图内丰富的结构信息<sup>[110-112]</sup>,而采用分子图可以充分获取分子图传递的结构信息(图8c),代表性的工作有Yan、McDermott等<sup>[111-115]</sup>的工作。

总体来看,基于模板的预测能够提供更可靠的化学解释,但反应模板具有多样性,选择合适的反应模板是需要解决的核心问题。Fortunato等<sup>[116]</sup>通过利用USPTO数据库训练神经网络模型,以提高对合成含能分子所需的稀有反应模板的识别能力。无模板的预测方法则具有更多的可能性,但计算复杂度高且化学可合成性有待评估。陈可辛等<sup>[117]</sup>总结了一些成功用于逆向预测目标分子合成路径的商业软件。目前最先进

的逆合成预测模型主要针对单步反应进行预测,距离独立解决实际合成问题还有一定距离。因为化合物可能存在多种合成路径,很难判断某条逆合成路径的合理性,实现多步逆合成路径预测难度大。尽管对推导有机化合物合成路线取得一些进展,但针对含能分子合成路线预测的相关研究报道仍寥寥无几。其难点在于含能分子的合成涉及众多复杂的反应类型,我们期待在未来能够学习相关算法思想,扩充含能分子合成稀有反应模板库,逐步将该方法应用于含能材料的合成路线预测,以实现该领域分子设计的重大突破。

## 2.4 化学反应条件预测

自动化开发目标化合物合成计划中关键的一环是选择合适的实验条件以获得合理的产率<sup>[117-118]</sup>。机器学习和机器学习工具是实验数据驱动工作流程的强大补充,可以加速化学反应条件的识别<sup>[119-120]</sup>,以确定反应条件中的各种参数,包括化学原料<sup>[121-123]</sup>(例如金属、配体、溶剂和催化剂)和操作条件<sup>[124-126]</sup>(例如温度、压力),如图9所示。

在材料分子设计领域,利用AI进行化学反应条件的自动化预测目前还处于起步阶段,将该新技术应用于含能分子合成反应条件预测对提升研制效率很有必要。但现有数据库并未记录所有可能的反应条件,许多潜在可行的条件仍有待发现。此外,这些数据库通常为手动维护,不可避免地存在数据来源不统一或错误信息的问题,预测模型的准确性和泛化能力还需进

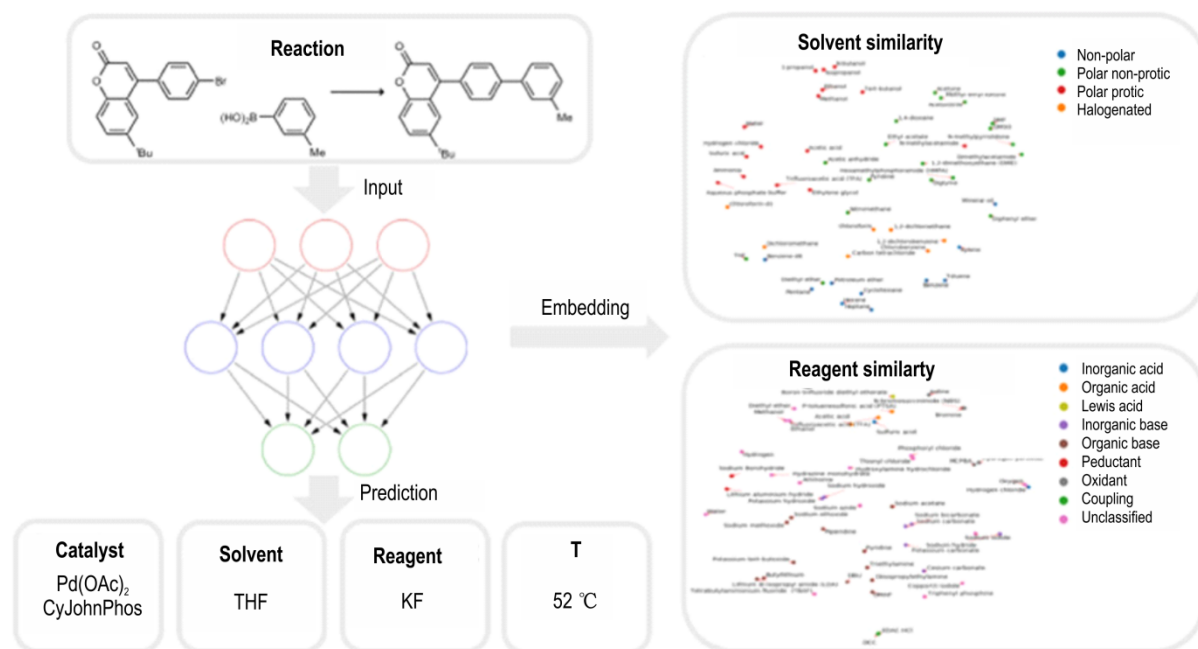


图9 AI辅助预测有机反应的适宜条件<sup>[121]</sup>

Fig.9 AI-assisted prediction of suitable conditions for organic reactions<sup>[121]</sup>



一步评估。如果能成功解决反应数据库缺失的问题, 这项技术将具备巨大的应用潜力。通过实现自动化和高效的反应条件预测, 不仅有助于加速新材料的发现和合成, 还有望推动能源存储和利用领域的创新。

### 3 差距分析

随着高通量虚拟筛选技术的快速发展与广泛应用, 含能分子设计正在向高通量设计阶段和AI辅助分子设计方向发展。在应用AI辅助含能分子设计的过程中面临诸多困难。一方面, 各个设计子环节的技术都有提升的空间。目前, 基于AI的方法已经初步用于性能预测和分子生成, 但样本量不足、特征工程不充分等因素制约着AI技术进一步应用与发展。另一方面, 含能分子合成设计涉及多种类型的化学反应, 反应机理复杂, 为合成路线设计和反应条件优化带来了巨大挑战。总体而言, 与计算机辅助药物或材料分子设计相比, AI在辅助含能分子设计方面的进展与应用还存在明显的差距。

性能预测模型在其他材料设计领域取得一定发展, 但这些方法很难直接应用于含能分子的设计。与大多数的有机材料、药物分子不同, 含能分子需要兼顾稳定性、爆炸能力等多性能指标, 其应用潜力是一个典型的多性能指标综合评估问题。现有模型对单一性能指标的预测精度仍有待提升。例如分解温度的预测目前仍然是个难题, 仅有少量采用机器学习的QSPR模型的报道<sup>[70, 127-128]</sup>且精度不高, 平均误差普遍超过30 °C。撞击感度的预测目前尚未形成可靠的普适性方法, 预测精度低(对于50%爆炸概率的特性落高 $H_{50}$ 平均预测误差超过20 cm)且只在限定条件下对特定类型的化合物有效<sup>[3, 129-130]</sup>。与其他材料性能预测模型的发展相比, AI在含能分子方面的发展受制于实验数据的匮乏。这些数据样本量较少且分子种类分布不均衡, 从而限制了预测模型的准确性和泛化能力。借鉴AI辅助材料设计的经验, 利用数据增强、迁移学习等技术拓展样本空间, 并通过高通量计算丰富数据源, 有望解决含能分子小样本空间的问题。

分子生成方法关乎分子设计的有效性。完全枚举法构造含能分子的方法已经相对成熟, 但其产生的大规模候选分子库中, 真正具有应用价值的目标化合物通常只占很小一部分。基于深度学习的生成模型可以在分子生成的过程中利用模型学习到的结构与性能之间的潜在关联规律, 用于生成新分子。但深度学习模

型更多用于药物分子设计, 用于针对性生成具有特定活性的药物分子, 而在含能分子领域应用还比较少, 仍处于初步探索阶段。该方法的发展受到性能与结构之间特征提取不充分以及数据样本不足以支持数据挖掘需求的限制。未来有望通过构建高质量的数据集来进一步挖掘结构与性能之间的关系, 以设计适用于含能分子的生成模型。

目前尚未建立适用于含能分子合成路线的普适性规律, 实验室合成路线的优化设计是未来的一个重要研究方向。含能化合物的合成涉及多种化学反应类型(氧化、硝化、环化等), 每一步反应的成败都取决于具体的条件(溶剂、温度、反应时间、PH值等), 不存在普适性的规律<sup>[31, 131-135]</sup>。不恰当的合成路线会导致不可预料的时间和资源投入进而影响研制进程。目前的合成路线及反应条件预测方法还无法很好地满足含能化合物的特殊需求, 难以实现个性化设计和优化。分子设计的出发点和归宿是在实验上获得符合预测性能的新化合物, 而实验室合成效率从根本上制约了新型含能化合物的研发。即便采用分子设计策略设计出性能优异的新化合物, 但要在实验室成功合成仍然面临巨大挑战。

与其他材料的自动化分子设计流程相比, 含能分子的分子设计自动化程度仍显不足。有机材料和高分子材料的设计中, 已经实现了利用AI辅助分子生成、性能预测, 到合成路线以及反应条件的预测, 甚至能够实现自动化合成和表征测试<sup>[16, 136]</sup>, 实现了从分子设计到材料制备的全流程自动化。而含能分子的分子设计流程仍停留在单个环节上, 未能形成“设计→评估→制备→验证”全流程自动化分子设计模式。各个分子设计子环节有待进一步融合AI以提升准确度与有效性, 实现全流程自动化设计与材料制备。开发全自动的含能分子研发模式将大大缩短周期, 提高含能分子研发效率, 是值得未来努力探索的重要发展方向。

### 4 结论

研究回顾了含能分子设计的发展历程, 并介绍了计算机辅助含能分子设计的研究现状。当前, 含能分子的设计模式正在向高通量计算和AI辅助分子设计方向发展。其中, AI可以从数据中挖掘出潜藏的化学规则, 建立结构与性能之间的定量关系, 已被用于构建性能预测模型和分子生成。本文概述了AI辅助分子设计在性能预测、分子生成、合成路线以及反应条件预

测中的最新研究进展,讨论了当前含能分子与其他材料设计的方法之间的差距。未来的研究方向应考虑通过数据增强、迁移学习等手段,最大限度地利用现有的实验数据;同时利用高通量计算补充高质量的计算数据,通过足量的含能分子数据样本提升性能预测模型的准确性与分子生成的有效性。同时,在合成路线和反应条件预测等环节引入AI,提升含能分子合成效率,最终有望实现“设计→评估→制备→验证”全流程自动化分子设计模式,加速新型含能分子的发现。

#### 参考文献:

- [1] 欧育湘,刘进全. 高能量密度化合物[M]. 北京:国防工业出版社,2005.  
OU Yu-xiang, LIU Jin-quan. High energy density compounds [M]. Bei Jing: Nation Defense Industry Press,2005.
- [2] 居学海,周素芹,梅争. 量子化学在含能材料中的应用新进展[M]. 北京:国防工业出版社,2017.  
JU Xue-hai, ZHOU Su-qin, MEI Zheng. Advances in quantum chemistry: energetic materials [M]. Bei Jing: Nation Defense Industry Press, 2017.
- [3] XIONG X, HE X, XIONG Y, et al. Correlation between the self-sustaining ignition ability and the impact sensitivity of energetic materials [J]. *Energetic Materials Frontiers*, 2020, 1(1): 40-49.
- [4] JACKSON C L, WING J F. On tribromotrinitrobenzol [J]. 1888, 10: 283.
- [5] NIELSEN A T, NISSAN R A, VANDERAH D J, et al. Polyazapolycyclics by condensation of aldehydes with amines. 2. formation of 2, 4, 6, 8, 10, 12-hexabenzyl-2, 4, 6, 8, 10, 12-hexaazatetracyclo [5.5.0.0.5.9.0.3, 11] dodecanes from glyoxal and benzylamines[J]. 1990, 55: 1459-1466.
- [6] 庞爱民,黎小平. 含能材料——从设计到销毁[M]. 第1版. 北京:国防工业出版社,2022.  
PANG Ai-ming, LI Xiao-ping. Energetic Materials: From Cradle to Grave[M]. Bei Jing: Nation Defense Industry Press, 2022.
- [7] 刘建. 含能分子性能预测模型与高通量计算方法研究[D]. 中国工程物理研究院,2022.  
LIU Jian. Study of Properties Prediction Models and High Throughput Calculation Approaches of Energetic Molecules [D]. China Academy of Engineering Physics, 2022.
- [8] 肖鹤鸣,许晓娟,邱玲. 高能量密度材料的理论设计[M]. 北京:科学出版社,2008.  
XIAO He-ming, XU Xiao-juan, QIU Ling. Theoretical Design of High Energy Density Materials[M]. Beijing: China Science Publishing & Media Ltd, 2008.
- [9] LATYPOV N V, BERGMAN J, LANGLET A, et al. Synthesis and reactions of 1,1-diamino-2,2-dinitroethylene[J]. *Tetrahedron*, 1998, 54(38): 11525-11536.
- [10] PAGORIA P F, MITCHELL A R, SCHMIDT R D. 1, 1, 1-trimethylhydrazinium iodide: A novel, highly reactive reagent for aromatic amination via vicarious nucleophilic substitution of hydrogen[J]. *The Journal of Organic Chemistry*, American Chemical Society, 1996, 61(9): 2934-2935.
- [11] KLENOV M S, GUSKOV A A, ANIKIN O V, et al. Synthesis of tetrazino-tetrazine 1, 3, 6, 8-tetraoxide (tto)[J]. *Angewandte Chemie International Edition*, John Wiley & Sons, Ltd, 2016, 55(38): 11472-11475.
- [12] ZHANG M-X, EATON P E, GILARDI R. Hepta- and octanitrocubanes[J]. *Angewandte Chemie International Edition*, 2000, 39(2): 401-404.
- [13] BUTLER K T, DAVIES D W, CARTWRIGHT H, et al. Machine learning for molecular and materials science [J]. *Nature*, 2018, 559(7715): 547-555.
- [14] SANCHEZ-LENGELING BENJAMIN, ASPURU-GUZIK ALÁN. Inverse molecular design using machine learning: generative models for matter engineering[J]. *Science*, American Association for the Advancement of Science, 2018, 361(6400): 360-365.
- [15] SEGLER M H S, PREUSS M, WALLER M P. Planning chemical syntheses with deep neural networks and symbolic AI[J]. *Nature*, 2018, 555(7698): 604-610.
- [16] COLEY C W, THOMAS D A, LUMMISS J A M, et al. A robotic platform for flow synthesis of organic compounds informed by AI planning[J]. *Science*, 2019, 365(6453): eaax1566.
- [17] LIU R, TANG Y, TIAN J, et al. QSPR models for sublimation enthalpy of energetic compounds [J]. *Chemical Engineering Journal*, 2023, 474: 145725.
- [18] KANG P, LIU Z, ABOU-RACHID H, et al. Machine-learning assisted screening of energetic materials [J]. *The Journal of Physical Chemistry A*, 2020, 124(26): 5341-5351.
- [19] CASEY A D, SON S F, BILIONIS I, et al. Prediction of energetic material properties from electronic structure using 3d convolutional neural networks [J]. *Journal of Chemical Information and Modeling*, 2020, 60(10): 4457-4473.
- [20] 徐雅斌,孙胜杰,武装. 基于深度学习的含能材料生成焓预测方法[J]. 含能材料, 2021, 29(1): 20-28.  
XU Ya-bin, SUN Sheng-jie, WU Zhuang. Enthalpy of formation prediction for energetic materials based on deep learning [J]. *Chinese Journal of Energetic Materials (Hanneng Cailiao)*, 2021, 29(1):20-28.
- [21] YANG C, CHEN J, WANG R, et al. Density prediction models for energetic compounds merely using molecular topology [J]. *Journal of Chemical Information and Modeling*, 2021, 61(6): 2582-2593.
- [22] NGUYEN P, LOVELAND D, KIM J T, et al. Predicting energetic materials' crystalline density from chemical structure by machine learning [J]. *Journal of Chemical Information and Modeling*, American Chemical Society, 2021, 61(5): 2147-2158.
- [23] 刘建,赵士操,张朝阳. 含能材料高通量计算交互式应用系统 EM-studio v1.0: 中国, 2021SR0611735[P]. 2021.  
LIU Jian, ZHAO Shi-cao, ZHANG Chao-yang, et al. EM studio v1.0, an interactive application system for high-throughput computing of energetic materials: China, 2021SR0611735 [P]. 2021
- [24] JIANG Y, YANG Z, GUO J, et al. Coupling complementary strategy to flexible graph neural network for quick discovery of coformer in diverse co-crystal materials[J]. *Nature Communications*, 2021, 12(1): 5950.
- [25] HUANG X, LI C, TAN K, et al. Applying machine learning to balance performance and stability of high energy density materials[J]. *iScience*, 2021, 24(3): 102240.
- [26] COWAN R D, FICKETT W. Calculation of the detonation prop-

- erties of solid explosives with the kistiakowsky-wilson equation of state[J]. *The Journal of Chemical Physics*, American Institute of Physics, 1956, 24(5): 932-939.
- [27] COWPERTHWAIT M, ZWISLER W H. The jcz equation of state of detonation products and their incorporation into the tiger code[C]//Coronado, CA, USA: 1976: 162.
- [28] WU X, LONG X, HE B, et al. VLW equation of state of detonation products [J]. *Science in China Series B: Chemistry*, 2009, 52(5): 605-608.
- [29] SUČESKA M. Evaluation of detonation energy from explo5 computer code results [J]. *Propellants, Explosives, Pyrotechnics*, John Wiley & Sons, Ltd, 1999, 24(5): 280-285.
- [30] BONDARCHUK S V, ZHANG Z, CHEN C, et al. Grammar of impact sensitivity: An incremental theory [J]. *The Journal of Physical Chemistry A*, 2023, 127(49): 10506-10516.
- [31] GAO H, ZHANG Q, SHREEVE J M. Fused heterocycle-based energetic materials (2012-2019) [J]. *Journal of Materials Chemistry A*, 2020, 8(8): 4193-4216.
- [32] HOLDREN J P. Materials genome initiative for global competitiveness[R]. National Science and Technology Council, 2011.
- [33] 王润文, 杨春明, 刘建. 高通量计算与深度学习相结合的稠环含能化合物设计[J]. 含能材料, 2022, 30(12): 1226-1236. WANG Run-wen, YANG Chun-ming, LIU Jian. Exploring novel fused-ring energetic compounds via high-throughput computing and deep learning[J]. *Chinese Journal of Energetic Materials (Hanneng Cailiao)*, 2022, 30(12): 1226-1236.
- [34] WEN L, YU T, LAI W, et al. Transferring the available fused cyclic scaffolds for high-throughput combinatorial design of highly energetic materials via database mining[J]. *Fuel*, 2022, 324: 124591.
- [35] WEN L, YU T, LAI W, et al. Accelerating molecular design of cage energetic materials with zero oxygen balance through large-scale database search[J]. *The Journal of Physical Chemistry Letters*, American Chemical Society, 2021, 12(47): 11591-11597.
- [36] WEN L, WANG B, YU T, et al. Accelerating the search of chonf-containing highly energetic materials by combinatorial library design and high-throughput screening[J]. *Fuel*, 2022, 310: 122241.
- [37] CURTAROLO S, HART G L W, NARDELLI M B, et al. The high-throughput highway to computational materials design [J]. *Nature Materials*, 2013, 12(3): 191-201.
- [38] HJORTH LARSEN A, JØRGEN MORTENSEN J, BLOMQUIST J, et al. The atomic simulation environment—a python library for working with atoms[J]. *Journal of Physics: Condensed Matter*, 2017, 29(27): 273002.
- [39] PIZZI G, CEPPELOTTI A, SABATINI R, et al. AiiDA: automated interactive infrastructure and database for computational science [J]. *Computational Materials Science*, 2016, 111: 218-230.
- [40] YANG X, WANG Z, ZHAO X, et al. MatCloud: a high-throughput computational infrastructure for integrated management of materials simulation, data and resources [J]. *Computational Materials Science*, 2018, 146: 319-333.
- [41] WANG G, PENG L, LI K, et al. ALKEMIE: an intelligent computational platform for accelerating materials discovery and design [J]. *Computational Materials Science*, 2021, 186: 110064.
- [42] CURTAROLO S, SETYAWAN W, WANG S, et al. AFLOWLIB.org: A distributed materials properties repository from high-throughput ab initio calculations [J]. *Computational Materials Science*, 2012, 58: 227-235.
- [43] LIU J, ZHAO S, DUAN B, SU L. Energetic materials studio (2021sr0611735)[EB/OL]. 2021(2021). <http://mathtc.nssc-tj.cn/>.
- [44] LOOKMAN T, BALACHANDRAN P V, XUE D, et al. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design [J]. *npj Computational Materials*, 2019, 5(1): 21.
- [45] XU Y, ELCORO L, SONG Z-D, et al. High-throughput calculations of magnetic topological materials [J]. *Nature*, 2020, 586(7831): 702-707.
- [46] GUO S, LIU J, QIAN W, et al. A review of quantum chemical methods for treating energetic molecules [J]. *Energetic Materials Frontiers*, 2021, 2(4): 292-305.
- [47] LIU Y, YANG Z, ZOU X, et al. Data quantity governance for machine learning in materials science [J]. *National Science Review*, 2023, 10(7): nwad125.
- [48] HU X, BERATAN D N, YANG W. Emergent strategies for inverse molecular design [J]. *Science in China Series B: Chemistry*, 2009, 52(11): 1769-1776.
- [49] GENTILE F, YAACOUB J C, GLEAVE J, et al. Artificial intelligence-enabled virtual screening of ultra-large chemical libraries with deep docking [J]. *Nature Protocols*, 2022, 17(3): 672-697.
- [50] XIE Y, LIU Y, HU R, et al. A property-oriented adaptive design framework for rapid discovery of energetic molecules based on small-scale labeled datasets [J]. *RSC Advances*, 2021, 11(41): 25764-25776.
- [51] SONG S, WANG Y, CHEN F, et al. Machine learning-assisted high-throughput virtual screening for on-demand customization of advanced energetic materials [J]. *Engineering*, 2022, 10: 99-109.
- [52] HOU F, MA Y, HU Z, et al. Machine learning enabled quickly predicting of detonation properties of n-containing molecules for discovering new energetic materials [J]. *Advanced Theory and Simulations*, 2021, 4(6): 2100057.
- [53] JEBARA T. Machine learning: discriminative and generative [M]. Springer Science & Business Media, 2012.
- [54] LI C, WANG C, SUN M, et al. Correlated rnn framework to quickly generate molecules with desired properties for energetic materials in the low data regime [J]. *Journal of Chemical Information and Modeling*, American Chemical Society, 2022, 62(20): 4873-4887.
- [55] LIU J, ZHAO S, DUAN B, et al. High-throughput design of energetic molecules [J]. *Journal of Materials Chemistry A*, 2023, 11(45): 25031-25044.
- [56] RICE B M, BYRD E F C. Evaluation of electrostatic descriptors for predicting crystalline density [J]. *Journal of Computational Chemistry*, 2013, 34(25): 2146-2151.
- [57] WANG L, ZHANG M, CHEN J, et al. Corrections of molecular morphology and hydrogen bond for improved crystal density prediction [J]. *Molecules*, 2019, 25(1): 161.
- [58] POLITZER P, MARTINEZ J, MURRAY J S, et al. An electrostatic interaction correction for improved crystal density prediction [J]. *Molecular Physics*, 2009, 107(19): 2095-2101.

- [59] WHEELER S E, HOUK K N, SCHLEYER P v. R, et al. A hierarchy of homodesmotic reactions for thermochemistry [J]. *Journal of the American Chemical Society*, American Chemical Society, 2009, 131(7): 2547–2560.
- [60] RAMABHADRAN R O, RAGHAVACHARI K. Theoretical thermochemistry for organic molecules: development of the generalized connectivity-based hierarchy [J]. *Journal of Chemical Theory and Computation*, American Chemical Society, 2011, 7(7): 2094–2103.
- [61] REISER P, NEUBERT M, EBERHARD A, et al. Graph neural networks for materials science and chemistry [J]. *Communications Materials*, 2022, 3(1): 93.
- [62] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84–90.
- [63] SOUSA T, CORREIA J, PEREIRA V, et al. Generative deep learning for targeted compound design [J]. *Journal of Chemical Information and Modeling*, American Chemical Society, 2021, 61(11): 5343–5361.
- [64] TONG X, LIU X, TAN X, et al. Generative models for de novo drug design [J]. *Journal of Medicinal Chemistry*, American Chemical Society, 2021, 64(19): 14011–14027.
- [65] ARÚS-POUS J, JOHANSSON S V, PRYKHODKO O, et al. Randomized smiles strings improve the quality of molecular generative models [J]. *Journal of Cheminformatics*, 2019, 11(1): 71.
- [66] MORET M, FRIEDRICH L, GRISONI F, et al. Generative molecular design in low data regimes [J]. *Nature Machine Intelligence*, 2020, 2(3): 171–180.
- [67] MERK D, GRISONI F, FRIEDRICH L, et al. Tuning artificial intelligence on the de novo design of natural-product-inspired retinoid x receptor modulators [J]. *Communications Chemistry*, 2018, 1(1): 68.
- [68] SEGLER M H S, KOGEJ T, TYRCHAN C, et al. Generating focused molecule libraries for drug discovery with recurrent neural networks [J]. *ACS Central Science*, 2018, 4(1): 120–131.
- [69] AWALE M, SIROCKIN F, STIEFL N, et al. Drug analogs from fragment-based long short-term memory generative neural networks [J]. *Journal of Chemical Information and Modeling*, 2019, 59(4): 1347–1356.
- [70] SONG S, CHEN F, WANG Y, et al. Accelerating the discovery of energetic melt-castable materials by a high-throughput virtual screening and experimental approach [J]. *Journal of Materials Chemistry A*, 2021, 9(38): 21723–21731.
- [71] SRIDHARAN B, GOEL M, PRIYAKUMAR U D. Modern machine learning for tackling inverse problems in chemistry: molecular design to realization [J]. *Chemical Communications*, 2022, 58(35): 5316–5331.
- [72] KINGMA D P, WELING M. Auto-encoding variational bayes [J]. *arXiv*, 2022.
- [73] CAI Z, XIONG Z, XU H, et al. Generative adversarial networks: A survey toward private and secure applications [J]. *ACM Computing Surveys*, 2022, 54(6): 1–38.
- [74] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari with deep reinforcement learning [J]. *arXiv*, 2013.
- [75] YOU J, YING R, REN X, et al. GraphRNN: generating realistic graphs with deep auto-regressive models [C]//DY J, KRAUSE A. Proceedings of the 35th International Conference on Machine Learning. PMLR, 2018: 5708–5717.
- [76] GÓMEZ-BOMBARELLI R, WEI J N, DUVENAUD D, et al. Automatic chemical design using a data-driven continuous representation of molecules [J]. *ACS Central Science*, American Chemical Society, 2018, 4(2): 268–276.
- [77] BAGAL V, AGGARWAL R, VINOD P K, et al. MolGPT: molecular generation using a transformer-decoder model [J]. *Journal of Chemical Information and Modeling*, American Chemical Society, 2022, 62(9): 2064–2076.
- [78] YAO Z, SÁNCHEZ-LENGELING B, BOBBITT N S, et al. Inverse design of nanoporous crystalline reticular materials with deep generative models [J]. *Nature Machine Intelligence*, 2021, 3(1): 76–86.
- [79] KIM S, SCHROEDER C M, JACKSON N E. Open macromolecular genome: generative design of synthetically accessible polymers [J]. *ACS Polymers Au*, American Chemical Society, 2023, 3(4): 318–330.
- [80] LIM J, RYU S, KIM J W, et al. Molecular generative model based on conditional variational autoencoder for de novo molecular design [J]. *Journal of Cheminformatics*, 2018, 10(1): 31.
- [81] JIN W, BARZILAY R, JAAKKOLA T. Junction tree variational autoencoder for molecular graph generation [C]//DY J, KRAUSE A. Proceedings of the 35th International Conference on Machine Learning. PMLR, 2018: 2323–2332.
- [82] NOH J, KIM J, STEIN H S, et al. Inverse design of solid-state materials via a continuous representation [J]. *Matter*, 2019, 1(5): 1370–1384.
- [83] KIM B, LEE S, KIM J. Inverse design of porous materials using artificial neural networks [J]. *Science Advances*, 2020, 6(1): eaax9324.
- [84] NOUIRA A, SOKOLOVSKA N, CRIVELLO J-C. CrystalGAN: learning to discover crystallographic structures with generative adversarial networks [J].
- [85] BLASCHKE T, OLIVECRONA M, ENGVIST O, et al. Application of generative autoencoder in de novo molecular design [J]. *Molecular Informatics*, 2018, 37(1–2): 1700123.
- [86] PRYKHODKO O, JOHANSSON S V, KOTSIAS P-C, et al. A de novo molecular generation method using latent vector based generative adversarial network [J]. *Journal of Cheminformatics*, 2019, 11(1): 74.
- [87] MARIN ZAPATA P A, MÉNDEZ-LUCIO O, LE T, et al. Cell morphology-guided de novo hit design by conditioning gans on phenotypic image features [J]. *Digital Discovery*, 2023, 2(1): 91–102.
- [88] KADURIN A, NIKOLENKO S, KHRABROV K, et al. DruGAN: An advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico [J]. *Molecular Pharmaceutics*, American Chemical Society, 2017, 14(9): 3098–3104.
- [89] BUNG N, KRISHNAN S R, BULUSU G, et al. De novo design of new chemical entities for sars-cov-2 using artificial intelligence [J]. *Future Medicinal Chemistry*, 2021, 13(6): 575–585.
- [90] POPOVA M, ISAYEV O, TROPSHA A. Deep reinforcement learning for de novo drug design [J]. *Science Advances*, 2018, 4(7): eaap7885.
- [91] YOU J, LIU B, YING R, et al. Graph convolutional policy net-

- work for goal-directed molecular graph generation [J]. *arXiv*, 2019.
- [92] GOEL M, RAGHUNATHAN S, LAGHUVARAPU S, et al. MoleGuLAR: Molecule generation using reinforcement learning with alternating rewards [J]. *Journal of Chemical Information and Modeling*, American Chemical Society, 2021, 61(12): 5815–5826.
- [93] KIM K, KANG S, YOO J, et al. Deep-learning-based inverse design model for intelligent discovery of organic molecules [J]. *npj Computational Materials*, 2018, 4(1): 67.
- [94] BHADWAL A S, KUMAR K, KUMAR N. GenSMILES: An enhanced validity conscious representation for inverse design of molecules [J]. *Knowledge-Based Systems*, 2023, 268: 110429.
- [95] OLIVECRONA M, BLASCHKE T, ENKVIST O, et al. Molecular de-novo design through deep reinforcement learning [J]. *Journal of Cheminformatics*, 2017, 9(1): 48.
- [96] PUTIN E, ASADULAEV A, IVANENKOV Y, et al. Reinforced adversarial neural computer for de novo molecular design [J]. *Journal of Chemical Information and Modeling*, American Chemical Society, 2018, 58(6): 1194–1204.
- [97] JIANG Y, YU Y, KONG M, et al. Artificial intelligence for retrosynthesis prediction [J]. *Engineering*, 2023, 25: 32–50.
- [98] COREY E J, WIPKE W T. Computer-assisted design of complex organic syntheses [J]. *Science*, American Association for the Advancement of Science, 1969, 166(3902): 178–192.
- [99] DONG J, ZHAO M, LIU Y, et al. Deep learning in retrosynthesis planning: Datasets, models and tools [J]. *Briefings in Bioinformatics*, 2022, 23(1): bbab391.
- [100] BAYLON J L, CILFONE N A, GULCHER J R, et al. Enhancing retrosynthetic reaction prediction with deep learning using multiscale reaction classification [J]. *Journal of Chemical Information and Modeling*, American Chemical Society, 2019, 59(2): 673–688.
- [101] SEGLER M H S, WALLER M P. Modelling chemical reasoning to predict and invent reactions [J]. *Chemistry—A European Journal*, 2017, 23(25): 6118–6128.
- [102] ISHIDA S, TERAYAMA K, KOJIMA R, et al. AI-driven synthetic route design incorporated with retrosynthesis knowledge [J]. *Journal of Chemical Information and Modeling*, 2022, 62(6): 1357–1367.
- [103] COLEY C W, ROGERS L, GREEN W H, et al. Computer-assisted retrosynthesis based on molecular similarity [J]. *ACS Central Science*, 2017, 3(12): 1237–1245.
- [104] DAI H, LI C, COLEY C, et al. Retrosynthesis prediction with conditional graph logic network [C]//WALLACH H, LAROCHELLE H, BEYGELZIMER A, et al. *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2019.
- [105] HAMILTON W, YING Z, LESKOVEC J. Inductive representation learning on large graphs [C]//GUYON I, LUXBURG U V, BENGIO S, et al. *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017.
- [106] LEI T, JIN W, BARZILAY R, et al. Deriving neural architectures from sequence and graph kernels [C]//PRECU D, TEH Y W. *Proceedings of the 34th International Conference on Machine Learning*. PMLR, 2017: 2024–2033.
- [107] LIU B, RAMSUNDAR B, KAWTHEKAR P, et al. Retrosynthetic reaction prediction using neural sequence-to-sequence models [J]. *ACS Central Science*, 2017, 3(10): 1103–1113.
- [108] KARPOV P, GODIN G, TETKO I V. A transformer model for retrosynthesis [G]//TETKO I V, KŮRKOVÁ V, KARPOV P, et al. *Artificial Neural Networks and Machine Learning - ICANN 2019: Workshop and Special Sessions*. Cham: Springer International Publishing, 2019: 817–830.
- [109] ZHENG S, RAO J, ZHANG Z, et al. Predicting retrosynthetic reactions using self-corrected transformer neural networks [J]. *Journal of Chemical Information and Modeling*, American Chemical Society, 2020, 60(1): 47–55.
- [110] UCAK U V, ASHYRMAMATOV I, KO J, et al. Retrosynthetic reaction pathway prediction through neural machine translation of atomic environments [J]. *Nature Communications*, 2022, 13(1): 1186.
- [111] SOMNATH V R, BUNNE C, COLEY C W, et al. Learning graph models for template-free retrosynthesis [C]//International Conference on Machine Learning (ICML) Workshop on Graph Representation Learning and Beyond (GRL+). PMLR, 2020.
- [112] SHI C, XU M, GUO H, et al. A graph to graphs framework for retrosynthesis prediction [C]//H D, SINGH A. *Proceedings of the 37th International Conference on Machine Learning*. PMLR, 2020: 8818–8827.
- [113] YAN C, DING Q, ZHAO P, et al. RetroXpert: decompose retrosynthesis prediction like a chemist [J]. *arXiv*, 2020.
- [114] SACHA M, BŁAŻ M, BYRSKI P, et al. Molecule edit graph attention network: Modeling chemical reactions as sequences of graph edits [J]. *Journal of Chemical Information and Modeling*, American Chemical Society, 2021, 61(7): 3273–3284.
- [115] MCDERMOTT M J, DWARAKNATH S S, PERSSON K A. A graph-based network for predicting chemical reaction pathways in solid-state materials synthesis [J]. *Nature Communications*, 2021, 12(1): 3097.
- [116] FORTUNATO M E, COLEY C W, BARNES B C, et al. Machine learned prediction of reaction template applicability for data-driven retrosynthetic predictions of energetic materials [C]//Portland, OR, USA: 2020: 070014.
- [117] KEXIN CHEN, YAODONG YANG, BOWEN WANG, et al. Black-box optimization algorithms for chemical synthesis and reaction condition optimization [J]. *SCIENTIA SINICA Chimica*, 2022, 53(1): 26–38.
- [118] AFONINA V A, MAZITOV D A, NURMUKHAMETOVA A, et al. Prediction of optimal conditions of hydrogenation reaction using the likelihood ranking approach [J]. *International Journal of Molecular Sciences*, 2021, 23(1): 248.
- [119] SHI Y, PRIETO P L, ZEPEL T, et al. Automated experimentation powers data science in chemistry [J]. *Accounts of Chemical Research*, American Chemical Society, 2021, 54(3): 546–555.
- [120] LIU J, HEIN J E. Automation, analytics and artificial intelligence for chemical synthesis [J]. *Nature Synthesis*, 2023, 2(6): 464–466.
- [121] GAO H, STRUBLE T J, COLEY C W, et al. Using machine learning to predict suitable conditions for organic reactions [J]. *ACS Central Science*, 2018, 4(11): 1465–1476.
- [122] ZHOU Z, LI X, ZARE R N. Optimizing chemical reactions with deep reinforcement learning [J]. *ACS Central Science*, 2017, 3(12): 1337–1344.
- [123] MARCOU G, AIRES DE SOUSA J, LATINO D A R S, et al. Expert system for predicting reaction conditions: the michael re-

- action case[J]. *Journal of Chemical Information and Modeling*, American Chemical Society, 2015, 55(2): 239–250.
- [124] KWON Y, LEE D, KIM J W, et al. Exploring optimal reaction conditions guided by graph neural networks and bayesian optimization[J]. *ACS Omega*, 2022, 7(49): 44939–44950.
- [125] GONG Y, XUE D, CHUAI G, et al. DeepReac+: deep active learning for quantitative modeling of organic chemical reactions[J]. *Chemical Science*, 2021, 12(43): 14459–14472.
- [126] KWON Y, KIM S, CHOI Y-S, et al. Generative modeling to predict multiple suitable conditions for chemical reactions[J]. *Journal of Chemical Information and Modeling*, 2022, 62(23): 5952–5960.
- [127] ZOHARI N, ABRISHAMI F, ZEYNALI V. Prediction of decomposition temperature of azole-based energetic compounds in order to assess of their thermal stability[J]. *Journal of Thermal Analysis and Calorimetry*, 2020, 141(4): 1453–1463.
- [128] FATHOLLAHI M, SAJADY H. QSPR modeling of decomposition temperature of energetic cocrystals using artificial neural network [J]. *Journal of Thermal Analysis and Calorimetry*, 2018, 133(3): 1663–1672.
- [129] BU R, XIONG Y, ZHANG C.  $\pi$ - $\pi$  stacking contributing to the low or reduced impact sensitivity of energetic materials [J]. *Crystal Growth & Design*, American Chemical Society, 2020, 20(5): 2824–2841.
- [130] 钟凯, 刘建, 王林元, 等. 含能材料中“热点”的理论模拟研究进展[J]. *含能材料*, 2018, 26(1): 11–20.
- ZHONG Kai, LIU Jian, WANG Lin-yuan, et al. Lssue of 'hot-spot' in energetic materials: recent progresses of modeling and calculationsg[J]. *Chinese Journal of Energetic Materials (Hanneng Cailiao)*, 2018, 26(1):11–20.
- [131] ZHANG Y, ZHAI L, XU C, et al. Thermal studies of novel heat-resistant 3,6-dinitropyrazolo[4,3-c]pyrazole-based energetic materials[J]. *Journal of Analytical and Applied Pyrolysis*, 2023, 169: 105866.
- [132] ZHANG G, HU W, MA J, et al. Combining 5,6-fused triazolo-triazine with pyrazole: A novel energetic framework for heat-resistant explosive[J]. *Chemical Engineering Journal*, 2021, 426: 131297.
- [133] YAN T, YANG H, YANG C, et al. An advanced and applicable heat-resistant explosive through controllable regiochemical modulation [J]. *Journal of Materials Chemistry A*, The Royal Society of Chemistry, 2020, 8(45): 23857–23865.
- [134] CAI J, XIE C, XIONG J, et al. High performance and heat-resistant pyrazole-1,2,4-triazole energetic materials: tuning the thermal stability by asymmetric framework and azo-bistriazole bridge [J]. *Chemical Engineering Journal*, 2022, 433: 134480.
- [135] DENG M, CHEN F, SONG S, et al. From the sensitive primary explosive icm-103 to insensitive heat-resistant energetic materials through a local azide-to-amino structural modification strategy[J]. *Chemical Engineering Journal*, 2022, 429: 132172.
- [136] SHIELDS B J, STEVENS J, LI J, et al. Bayesian reaction optimization as a tool for chemical synthesis[J]. *Nature*, 2021, 590(7844): 89–96.

## Applications and Prospects of AI-assisted Design of Energetic Molecules

LIU Rui<sup>1,2</sup>, LIU Jian<sup>1</sup>, TANG Yue-chuan<sup>1</sup>, ZHANG Chao-yang<sup>1</sup>, HUANG Jing<sup>1</sup>, HUANG Xin<sup>1</sup>

(1. Institute of Chemical Materials, CAEP, Mianyang 621999, China; 2. School of Chemistry and Chemical Engineering, Southwest Petroleum University, Chengdu 610500, China)

**Abstract:** The explore of energetic molecules faces multiple challenges, and the traditional design method are inefficient. The emergence of computer-aided molecular design has changed the research and development model. This review provides an overview of the development of energetic molecular design and introduces the current research status of computer-aided energetic molecular design. By summarizing the latest advancements in Artificial Intelligence (AI) technology across various design aspects, including performance prediction, molecular generation, retrosynthetic reaction prediction, and reaction condition prediction, we discussed the existing gap between the current approaches in energetic molecular design and other materials design methods. By thinking about the causes of the gap, we present an outlook on the future developmental directions of AI-assisted energetic molecular design. Research indicates that AI has already been applied in property prediction and molecular generation of energetic molecular design, but requires further exploration in retrosynthetic reaction prediction, and reaction conditions prediction. AI-assisted design of energetic molecules holds broad promising application prospects. Data enhancement, transfer learning and high-throughput computing are expected to solve the problem of weak data of energetic molecules. Enhancing AI-assisted prediction of synthesis routes and reaction conditions for energetic molecules shows promise for achieving the automatic molecular design via whole process of “design→evaluation→preparation→verification”. AI-assisted energetic molecular design provides new possibilities for improving the level of energetic molecular design and helps to improve the efficiency of energetic molecule research and development.

**Key words:** energetic molecule; molecular design; artificial intelligence; machine learning; quantitative structure-property relationships

**CLC number:** TJ55;O62

**Document code:** A

**DOI:** 10.11943/CJEM2023226

**Grant Support:** National Natural Science Foundation of China(Nos. 22173086, 22203081, 22305234)

(责编:姜梅)