

• AI 驱动底层技术 •

盛翔 中国科学院天津工业生物技术研究所研究员、博士生导师。研究方向为酶的计算模拟与理性设计：采用物理原理和人工智能融合的方法研究生物体系内物质转化的化学基本原理、蛋白结构的动态变化规律以及新酶工业属性的调控机制，基于结构与功能信息，建立人工酶理性设计新方法，指导实现高效人工酶的精准智能创制。在国内外期刊发表论文 70 余篇，主持国家重点研发计划、国家自然科学基金等多项国家级、省部级基金项目，任中国化工学会生物化工青委会和中国生物工程学会青委会委员。



人工智能与物理原理融合驱动蛋白计算模拟技术

刘保艳¹, 李帅^{1,2}, 苏浩^{1,2}, 盛翔^{1,2*}

1 中国科学院天津工业生物技术研究所 低碳合成工程生物学全国重点实验室, 天津 300308

2 中国科学院大学, 北京 100049

刘保艳, 李帅, 苏浩, 盛翔. 人工智能与物理原理融合驱动蛋白计算模拟技术[J]. 生物工程学报, 2025, 41(3): 917-933.

LIU Baoyan, LI Shuai, SU Hao, SHENG Xiang. Artificial intelligence-enhanced physics-based computational modeling technologies for proteins[J]. Chinese Journal of Biotechnology, 2025, 41(3): 917-933.

摘要: 计算模拟驱动的生物元件、代谢网络乃至细胞系统的机理解析、定向改造和按需设计, 可为解决不同层次的生物学问题提供新的技术方案, 已成为生物制造的核心研究内容。在蛋白元件的计算模拟方面, 基于物理原理的传统方法利用计算机软件和数学模型来模拟生物体系中蛋白行使功能的物理和化学过程, 是理解复杂生物体系和指导实验设计的有力工具。随着生物系统模拟尺度的不断扩大, 传统计算模拟技术面临计算精度和计算速度难以平衡的困境。近年来, 生物数据量呈现爆炸式增长, 使得构建高性能人工智能(artificial intelligence, AI)模型成为可能, 为蛋白计算模拟带来了新思路和新契机, AI 和物理原理融合驱动蛋白计算模拟技术应运而生。本文对基于物理原理的传统蛋白计算模拟方法及其应用进行了详细介绍, 并对融合 AI 和物理原理的最新计算模拟技术进行了梳理和讨论, 进而提出在 AI 模型中结合严谨的化学知识和既定的物理原理, 可有效提升数据处理和模式识别能力, 从而提高计算效率和预测的准确性, 使其具有更强的解释能力、通用性和稳健性。目前, AI 与物理原理融合驱动的计算模拟技术已在生物催化领域展现出巨大的潜力和价值。本文聚焦于主流和先进的蛋白计算模拟技术, 通过对这些技术的系统性回顾和前瞻性分析,

资助项目: 国家重点研发计划(2023YFA0916100); 中国科学院战略性先导科技专项(XDC0110201)

This work was supported by the National Key Research and Development Program of China (2023YFA0916100) and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDC0110201).

*Corresponding author. E-mail: shengx@tib.cas.cn

Received: 2024-07-25; Accepted: 2024-12-24; Published online: 2024-12-25

梳理了蛋白质计算模拟技术的发展脉络，以推动其在酶催化机制解析、蛋白从头设计与理性改造等领域的应用，助力生物制造的发展。

关键词：人工智能；计算模拟；计算生物学；分子对接；分子动力学模拟；量子化学计算

Artificial intelligence-enhanced physics-based computational modeling technologies for proteins

LIU Baoyan¹, LI Shuai^{1,2}, SU Hao^{1,2}, SHENG Xiang^{1,2*}

1 State Key Laboratory of Engineering Biology for Low-Carbon Manufacturing, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

2 University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: Computational modeling is an invaluable tool for mechanism analysis, directed engineering, and rational design of biological parts, metabolic networks, and even cellular systems. It can provide new technological solutions to address biological challenges at different levels and has become a central focus of research in biomanufacturing. In the computational modeling of proteins, which are the key parts in biological systems, the traditional physics-based methods (computer software and mathematical model) have been widely used to study the physical and chemical processes in the functioning of proteins, and have thus been recognized as a powerful tool for understanding complex biological systems and guiding experimental designs. As the scale of computational modeling continues to expand, traditional modeling techniques face difficulties in balancing computational accuracy and speed. In recent years, the explosive growth of biological data has made it possible to construct high-performance artificial intelligence (AI) models, which brings new opportunities to the computational modeling of proteins, and the AI-enhanced physics-based computational modeling technologies have emerged. This combined strategy not only incorporates the chemical knowledge and established physical principles but also is powerful in data processing and pattern recognition, which greatly improves the computational efficiency and prediction accuracy, as well as possesses stronger interpretation ability, transferability, and robustness. The AI-enhanced physics-based computational modeling technologies have already shown great potential and value in biocatalysis, paving a new way for the future development of biomanufacturing.

Keywords: artificial intelligence; computational modeling; computational biology; molecular docking; molecular dynamics simulation; quantum chemistry calculation

传统的计算模拟技术多基于物理原理和经验规则，利用计算机软件和数学模型来模拟生物体系中的物理、化学或生物学过程。其中，分子对接、分子动力学模拟、量子化学团簇、量子力学/分子力学组合等方法被广泛应用于

酶催化机制解析、蛋白结构动态行为分析、蛋白热稳定性评估、天然和人工蛋白设计与优化等研究。这些方法所使用的模拟工具通常包含大量用C/C++或FORTRAN编写的专业代码库，以及用于图形处理器(graphics processing unit,

GPU)加速的自定义计算统一设备架构(compute unified device architecture, CUDA)内核,并且几乎所有软件包都包含大量重复代码和硬编码,这些限制使得这些技术难以跟上不断发展的科学计算技术的步伐^[1]。同时,不同的计算模拟方法具有不同的优缺点,例如量子化学的计算模拟结果精确,但是计算成本高,难以应用于生物大分子的计算;分子动力学模拟一定程度上平衡了模拟精度和速度,但却面临采样不够的问题;分子对接具有很高的计算效率,但缺乏对受体构象柔性变化的充分考虑^[2]。

近年来,人工智能(artificial intelligence, AI)因其强大的海量数据处理能力和学习能力而备受关注,在解决蛋白质折叠、蛋白质从头设计、新酶设计等科学难题方面取得了巨大成功^[3-6]。同时,现代人工智能平台(如 TensorFlow^[7]、PyTorch^[8]和 MindSpore^[9]等)的设计主要以开发、训练和部署深度神经网络模型的效率和便利性为导向,降低了技术门槛和复杂性,而将

传统计算模拟技术整合到人工智能平台,可实现程序间的直接通信(如 MindSPONGE^[1]),进一步提高生产力和创新速度,使得更多的研究者和开发者能够快速实现和应用先进的人工智能技术。分子对接、分子动力学模拟、量子化学等计算模拟技术中复杂且耗时的计算特别适合融合人工智能技术,并且已经受到现有人工智能方法的深刻影响(图 1)。人工智能在计算效率和准确性提升、计算模拟结果分析、构象采样增强和集合变量识别等多方面对传统蛋白计算模拟技术进行了改进。然而,数据驱动的人工智能模型容易出现过拟合,面临着在未训练的数据上无法泛化使用等问题,而基于物理原理的人工智能则可帮助人工智能模型在训练时始终遵循物理规律,从而可以更快、更好地学习和泛化。

为了深入理解 AI 与物理原理融合驱动蛋白计算模拟技术,本文首先介绍了基于物理原理的传统蛋白计算模拟方法,在此基础上,对融合 AI 和物理原理的先进方法展开讨论。目

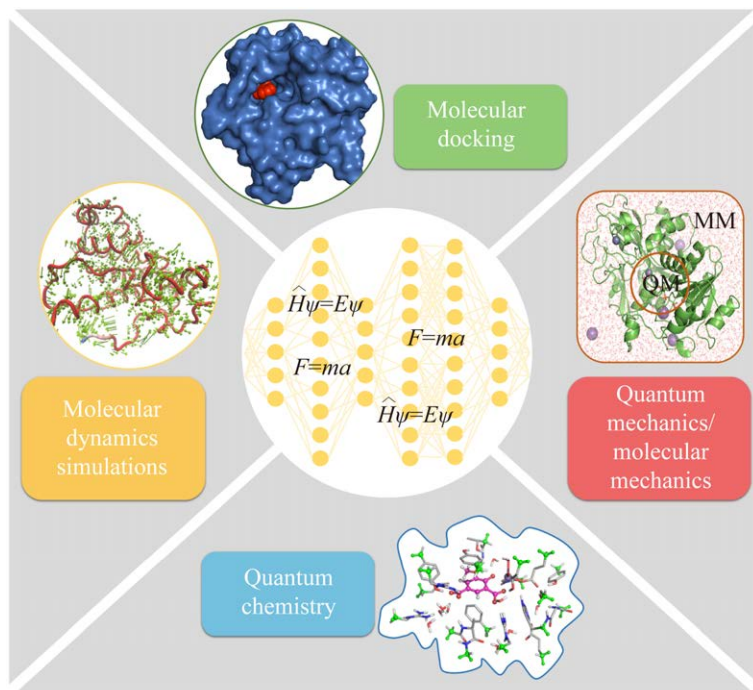


图 1 人工智能与物理原理融合驱动蛋白计算模拟技术概图

Figure 1 Artificial intelligence-enhanced physics-based computational modeling technologies for proteins.

前,常用的传统蛋白计算模拟技术主要包括分子对接、分子动力学模拟、量子化学计算、量子力学/分子力学组合法,而 AI 对这 4 种计算模拟技术的驱动进程及依赖的数据类型显著不同,因此本文以此为分类标准对方法进行介绍,并列举代表性案例。最后,对 AI 与物理原理融合驱动的蛋白计算模拟新技术进行了总结和展望。

1 基于物理原理的传统蛋白计算模拟方法

传统蛋白计算模拟方法,多基于物理原理,可用于模拟蛋白体系不同尺度的分子行为。如,分子对接可以评估受体和配体的相互作用模式,分子动力学模拟可以探究生物大分子的动态运动行为,量子力学方法则可以在电子结构尺度上研究化学反应的热力学或动力学性质。这些方法已被广泛应用于蛋白质工程、药物发现与筛选和蛋白质从头设计等研究,部分模拟工具及研究案例参见表 1。

1.1 分子对接

分子对接(molecular docking)是模拟 2 种或 2 种以上分子间相互识别和结合,并确认最优结合模式的过程,起源于 1984 年 Fischer 提出的“锁钥模型”和 1958 年 Koshland 提出的“诱导契合”学说。分子对接技术主要依赖于评分函数和构象采样,其中评分函数大致分为 4 类:基于物理的(physics-based)、基于经验的(empirical)、基于知识的(knowledge-based)和基于描述符的(descriptor-based);构象采样策略大致分为 3 类:系统搜索(systematic search)、启发式搜索(heuristic search)和确定性搜索(deterministic search)。在生物体系研究领域备受关注的虚拟筛选(virtual screening)就是建立在分子对接基础上的高通量方法,其既可基于配体筛选,又可基于受体筛选,因此在蛋白-配

体相互作用相关研究中被广泛应用,例如酶底物相互作用识别^[20]、酶突变体设计^[21]、蛋白核酸相互作用识别^[22-23]、药物设计^[24]及信号传导途径的研究^[25]等。

1.2 分子动力学模拟

分子动力学(molecular dynamics, MD)模拟通过迭代牛顿运动方程来模拟分子体系的构象在分子力场下随时间的运动变化,可用于研究蛋白质或其他大分子在原子水平的动态行为。通过对蛋白质体系的模拟可以捕获蛋白质的构象变化、蛋白-配体相互作用、蛋白异构调节和蛋白质折叠等各种重要生物分子的动态行为,并以飞秒时间分辨率揭示体系中所有原子的运动轨迹^[63]。经过足够模拟时长的分子动力学模拟,获得具有统计学意义的采样量,模拟产生的构象集合可用于蛋白体系各种性质的研究,例如机理分析、结构和能量等物理化学特性的统计计算以及蛋白质在原子水平上如何响应配体结合与离去、突变、磷酸化、质子化等扰动,有助于设计或定向干扰生物大分子行使功能的过程^[63]。通过与 X 射线衍射(X-ray diffraction, XRD)技术、冷冻电镜(cryogenic electron microscopy, Cryo-EM)技术、核磁共振(nuclear magnetic resonance, NMR)技术等实验手段相结合,分子动力学模拟已被广泛应用于蛋白结构动态热稳定性^[32]、天然和人工蛋白质的设计与优化^[33-34]、基于结构的药物分子设计^[35]、酶的底物特异性^[36]、蛋白别构效应^[37]、酶-底物复合物的近进攻构象分析^[38]等研究。

1.3 量子化学计算方法

量子化学(quantum chemistry, QC)是一门基于量子力学(quantum mechanics, QM)基本原理和方法来研究化学问题的科学,其理论方法主要有分子轨道理论(molecular orbital theory)、价键理论(valence bond theory)和密度泛函理论

表 1 传统计算蛋白模拟技术部分工具汇总及其应用举例^aTable 1 Lists of selected physics-based computational tools and applications^a

Tools	Applications	Cases
Molecular docking		
Glide ^[10-11]	Protein-ligand interactions	Ligand binding in cytochrome P450eryF ^[20] ; Substrate specificity of cytochrome P450cam and its L244A mutant ^[21] ; RNA recognition of tRNA pseudouridine synthase TruB ^[22] ; Dimerization of the damaged DNA induced UV-damaged DNA-binding protein ^[23] ; Discovery of lysine-targeted eIF4E inhibitors ^[24] ; Metal inhibitors of activated Ras protein ^[25]
GOLD ^[12]	Protein-protein interactions	
AutoDock 4.2 ^[13]	Protein-DNA interactions	
AutoDock Vina ^[14-15]	Protein-RNA interactions	
HDOCK ^[16]	Protein-metalloligand interactions	
RosettaDock 4.0 ^[17]		
ZDOCK ^[18]		
MOLS 2.0 ^[19]		
Molecular dynamics simulations		
AMBER ^[26]	Ligand binding	Thermostability prediction of cellobiose 2-epimerases ^[32] ;
CP2K ^[27]	Protein folding	Interaction networks in protein folding ^[33] ; Rational design of peptidic modulators of protein-protein interactions ^[34] ;
GROMACS ^[28]	Structure-based drug design	Structure-based de novo design of picomolar ABL kinase inhibitors ^[35] ; Design of enzymatic C-H hydroxylation ^[36] ;
NAMD ^[29]	Membrane transportation	Dynamic allostery-based molecular workings of kinase ^[37] ;
LAMMPS ^[30]	Allostery in proteins	Analysis of ground and transition states for the S _N 2 displacement in haloalkane dehalogenase ^[38] ; Mg ²⁺ -sensing mechanism of Mg ²⁺ transporter MgtE ^[39]
CHARMM ^[31]	Near-attack conformation analysis	
Quantum chemical calculations		
Gaussian 16 ^[40]	Substrate binding mode	Substrate binding mode in phenolic acid decarboxylase ^[46] ;
Turbomole ^[41]	Reaction mechanisms	Mechanism of nickel-containing carbon monoxide dehydrogenase ^[47] ; Mechanism and enantioselectivity of norcoclaurine synthase ^[48] ; Regioselectivity at each evolutionary stage of an improved mutant ^[49] ; Design of (<i>R</i>)-enantiospecific norcoclaurine synthase ^[50] ; Mechanism and stereospecificity of dihydropyrimidinase ^[51]
ORCA ^[42]	Enantioselectivity	
Q-Chem ^[43]	Regioselectivity	
NWChem ^[44]	Enzyme design	
Molpro ^[45]		
QM/MM calculations^b		
Chemshell ^[52]	Substrate binding mode	Reaction mechanism of P450 enzymes ^[57] ; Reaction mechanism of glycoside hydrolases ^[58] ; Enantioselectivity of the <i>O</i> -acetylation of (<i>R,S</i>)-propranolol catalyzed by lipase B ^[59] ; Regioselectivity of drug metabolism in cytochrome P450 2C9 ^[60] ; Designing the gold-bio interface ^[61] ; Rational redesign of 4-hydroxyphenylpyruvate dioxygenase (HPPD) ^[62]
ASE ^[53]	Reaction mechanisms	
Cuby ^[54]	Enantioselectivity	
MiMiC ^[55]	Regioselectivity	
QMCube ^[56]	Enzyme design	

^a: 表中部分工具包可用于执行不同类型计算模拟, 此处按照主要应用场景分类。^b: QM/MM 计算主要通过接口程序组合不同量子化学计算软件和分子动力学模拟软件实现, 此处也包含部分接口程序。

^a: Some toolkits listed in the table can be used for various types of computational modeling. They are categorized here based on their primary application scenarios for clarity. ^b: QM/MM calculations are primarily facilitated by interface programs that integrate quantum chemistry software with molecular dynamics simulation software. Several such interface programs are also included in this category.

(density functional theory, DFT)等。由于量子化学方法可以在电子水平上研究体系, 因此在酶催化反应过程中的化学键断裂与生成、电子传

递等微观现象的描述方面具有独特优势。基于量子化学的团簇方法(quantum chemical cluster approach)已被广泛应用于阐明酶催化反应不同

方面的性质, 它可以直接比较所有可能反应途径的能量关系, 从而确定酶催化状态下的底物结合模式、催化反应的活化能、反应的决速步骤和关键催化残基等, 这些信息对于酶的理性设计和改造至关重要^[64-65]。近年来, 该技术被广泛应用于研究生物酶元件的各类性质: (1) 底物结合模式, 例如基于量化团簇计算提出了酚酸脱羧酶不同于文献报道底物朝向的全新结合模式^[46]; (2) 酶催化反应机理, 例如含镍一氧化碳脱氢酶的机理研究阐明了镍价态在反应中的关键作用, 并为活性中心晶体结构提供了理论解释^[47]; (3) 酶对映选择性调控机制, 如去甲基乌药碱合成酶的完整反应路径解析与过渡态结构分析阐明了该酶催化天然底物的高 *S*-选择性的控制因素^[48]; (4) 酶区域选择性调控机制, 例如量化团簇计算解析了 Baeyer-Villiger 单加氧酶最优定向进化路径中的系列突变体的选择性影响因素^[49]; (5) 反应机理驱动的酶理性设计, 例如量化计算结果指导设计了自然界中不存在的具有 *R*-选择性偏好的全新去甲基乌药碱合成酶^[50]。

1.4 量子力学/分子力学组合方法

量子力学/分子力学(quantum mechanics/molecular mechanics, QM/MM)组合方法通过结合高水平的量子力学理论和经典分子力学模拟, 平衡计算准确性和速度, 实现对复杂生物系统中化学反应或物理过程的计算模拟。该方法在研究具体体系时, 通常将体系划分为 QM 区域和 MM 区域这 2 部分: QM 区域为关键催化或其他功能区域, 对其使用高水平量子力学理论方法描述; MM 区域主要是为蛋白和溶剂环境, 对其使用经典分子动力学模拟方法描述。二者将计算准确性与计算速度相结合, 可用于全面分析复杂生物系统内的反应机制和其他特性。与量子化学团簇方法一样, QM/MM 组合

方法可以对酶催化反应进行深入解析, 明确反应的关键催化残基、选择性调控机制、势能剖面等关键信息, 加深对相关酶的理解, 例如:

(1) 对系列血红素依赖型铁氧化酶开展的 QM/MM 研究阐明了该家族酶的电子转移机制和催化选择性调控机制, 明确了蛋白环境和铁中心配体对催化活性的关键影响^[57]; (2) 对大量糖苷酶的 QM/MM 计算, 阐明了多糖底物构象在活性位点的动态变化规律, 明确了蛋白环境对糖环构象的影响机制以及构象变化在糖苷酶反应中的关键作用, 建立了对糖苷酶家族普适的催化循环模型^[58]。此外, 该方法已成功应用于酶的对映选择性调控机制^[59], 药物代谢酶的催化机制^[60], 生物传感器的计算设计^[61], 以及酶的理性设计^[62]等。

2 人工智能与物理原理融合驱动的蛋白计算模拟技术

以机器学习(machine learning, ML)为代表的人工智能(artificial intelligence, AI)技术正在与生命科学进行深度交叉融合, 加速生物制造领域向着更快速、更精准和更经济的方向稳步发展。John McCarthy 在 1956 年的达特茅斯会议上, 第一次提出了“人工智能”一词, 并将其定义为制造智能机器的科学和工程^[66]。机器学习则被定义为“人工智能的一个子集”, 不通过显式编程而是通过数据驱动和算法应用来模拟人类学习过程, 同时也被定义为是“关注如何构建能够随着经验自动改进的计算机程序”^[67]。而深度学习(deep learning, DL)则是机器学习的一个分支, 通过多层神经网络(neural network, NN)等架构进行学习, 通常包含比传统机器学习模型更多的参数(可达数十亿个)和超参数, 因此神经网络具有强大的表达能力和设计灵活性, 应用范围涵盖即时文本生成到蛋白质结构

预测等不同领域^[67]。2024年诺贝尔物理学奖授予了 John Hopfield 和 Geoffrey Hinton, 以表彰他们在利用人工神经网络实现机器学习方面的开创性发现和发明。

传统计算模拟技术复杂且耗时, 在应用中面临诸多挑战, 人工智能技术的融合使得其不同方面取得突破。目前, 量子化学计算中采用的计算方法由于需要平衡速度和精度, 普遍存在精度不够的问题, 因为选择更高精度的算法必然意味着计算速度的急剧下降, 而通过融合人工智能技术则可以达到提高精度的同时提高速度的目的。分子对接和分子动力学模拟普遍面临采样不足、势函数不精确等问题, 且在数据分析和处理的过程容易引入人为偏差, 而融合人工智能的方法则可以探索更广阔的构象空间, 学习更复杂的规律, 并实现端到端的结果输出。

需要注意的是, 人工智能模型的预测准确性显著依赖于其训练集的质量与多样性, 不均衡或不充足的训练数据可能导致模型出现过拟合或者欠拟合现象。当前, 实验数据的获取仍面临采集周期长、成本高、通量低以数据格式不一致等挑战, 而理论数据能够在一定程度上弥补这些问题。通过应用融合物理原理的人工智能技术对现有科学问题进行重新建模, 预测研究数据的潜在分布模式, 可在一定程度上弥补因实验数据缺乏而导致的模型训练不准确等问题。例如, 传统分子动力学模拟及量子化学计算模拟可以生成大量有关分子和蛋白结构、能量和动态行为的数据, 这些数据可用于训练机器学习模型来增强采样、预测长时间尺度的蛋白结构变化、预测波函数和高通量筛选等。

随着人工智能技术与基于物理原理的计算模拟技术的逐渐融合, 蛋白结构和功能的研究正在走向新范式。本文对人工智能与物理原理

融合驱动的蛋白计算模拟技术部分工具进行了整理, 并对相应工具的关键方法细节进行了描述, 具体内容参见表 2。

2.1 AI 驱动分子对接技术

传统的分子对接方法利用评分函数和经验力场模拟配体与受体的结合, 筛选过程繁琐、耗时, 常常无法精确描述分子间的复杂相互作用, 并且无法预测受体和配体因动态结合造成的构象改变。AI 驱动分子对接技术可以从大型数据库中学习并理解化学结构和活性关系, 有望突破传统分子对接技术的应用局限。近年来, 基于深度学习的对接方法层出不穷, 这些方法很好地平衡了速度和准确度, 并学习到了复杂的分子间相互作用模式, 因而降低了对评分函数的依赖。同时, 这些方法有助于探索更广阔的化学空间及进一步提高预测精度, 如可从原子水平上预测配体与靶蛋白的相互作用^[90-91]。

评分函数是决定分子对接是否准确的关键, 基于机器学习的评分函数通过构建具有一定物理意义的描述符, 建立机器学习计算模型, 继而由机器学习算法学习配体特征和结合亲和力之间复杂的、非线性的关系并推导能量函数。Ragoza 等^[92]构建了基于卷积神经网络(convolutional neural network, CNN)模型的蛋白与配体相互作用的评分函数, 多个指标测试表明该评分函数优于 AutoDock Vina 经验评分函数。DeepBSP 基于 11 925 个天然复合物结构和 165 000 对接结构对三维卷积神经网络(three-dimensional CNN, 3D-CNN)模型进行训练, 该方法可更准确地预测配体对接姿态与天然结合姿态的均方根偏差(root mean square deviation, RMSD), 从而使得预测姿态更接近天然复合物结构^[93]。GNINA 则使用交叉对接和重新对接的数据训练的 CNN 模型作为评分函数, 其针对每个构象可输出一个配体的 CNNscore 和

表 2 人工智能与物理原理融合驱动下的蛋白计算模拟技术的部分工具

Table 2 Selected toolkits of artificial intelligence-enhanced physics-based computational modeling technologies for proteins

Tools	Methods	Key details of the methods
Molecular docking		
DiffDock ^[68] /DiffBindFR ^[69]	Diffusion generative model	Defining a diffusion process over the degrees of freedom involved in docking
TankBind ^[70]	Trigonometry-aware neural network	Building trigonometry constraint as a vigorous inductive bias into the neural network (NN) model
DeepDock ^[71]	Geometric deep learning	Learning a statistical potential based on the distance likelihood tailor-made for each ligand-target pair
EquiBind ^[72]	SE(3) geometric deep learning	Using graph matching networks (GMN) and E(3)-equivariant graph neural network (GNN)
Uni-Mol ^[73]	SE(3) transformer architecture	Incorporating 3D information by a universal 3D MRL framework and using transformer as a backbone
Molecular dynamics simulation		
GLOW ^[74]	Convolutional neural network	Determining important structural contacts by saliency (attention) maps of the residue contact gradients
TorchMD ^[75]	Deep neural network (DNN)	Extending the bonded and nonbonded force terms used in MD with DNN-based ones of arbitrary complexity
RAVE ^[76]	Variational autoencoders	Using the reweighted autoencoded variational bayes for enhanced sampling
NN2B ^[77]	Artificial neural network	Using the nearest neighbor density estimator (NNDE) and the artificial neural network (ANN) for the bias potential approximation
Quantum chemical method		
Psiformer ^[78]	Transformer	Using wavefunction transformer as an approximation for solving the many-electron Schrödinger equation
PauliNet ^[79]	Deep neural network	Replacing ad-hoc functional forms with DNN representations
Δ -learning ^[80]	Kernel ridge regression	Adding machine learning corrections to computationally inexpensive approximate legacy quantum methods
ML-SQC ^[81]	Kernel ridge regression	Using the kernel ridge regression with a Laplacian kernel, automatic tuning of SQC parameters for molecules
AIQM1 ^[82]	Atomic neural network	Using NN potentials based on torchani ^[83] , leveraging Δ -learning strategy and the power of transfer learning
Fermi Net ^[84]	Deep neural network	Introducing the fermionic neural network as a powerful wave-function Ansatz for many-electron systems
SALTED ^[85-86]	Gaussian process regression	Performing equivariant predictions of the electron density of both molecular and condensed-phase systems
QM/MM (MD) method		
dp-QM/MM ^[87]	Artificial neural network	Using molecular features and a set of collective variables (cvs) as input, chaperone polarizabilities as output
QM/MM-NN ^[88]	Artificial neural network	Using the information on molecular structures as input, and the potential energy as the output layer
QM/MM-NN MD ^[89]	Adaptive neural networks	Introducing adaptations to configuration appending and propagation on the configuration to update NN
(QM) ML/MM MD ^[90]	High dimensional neural network	Integrating the MM environment as an additional element type in the high dimensional neural network potential

一个复合物的 CNNAffinity, 并用于评估复合物的对接效果, 因此评分性能显著提升^[94]。AnnapuRNA 使用实验解析结构中的 RNA 和小分子相互作用的统计数据进行了监督机器学习, 并用于区分实验观察到的 RNA-配体相互作用和可在理论分析中观察到但未在自然界中观察到的相互作用, 该方法可用作评分函数评估 RNA-配体复合物结构的优劣, 且显著优于其他评分函数^[95]。

通过结合使用扩散模型(diffusion model), AI 驱动分子对接方法可以充分、合理地考虑蛋白柔性, 解决传统分子对接中可能存在的构象采样不足问题。Jaakkola 团队开发了基于配体位姿上的非欧几里得流形的扩散生成模型 DiffDock, 该模型以单独的配体和蛋白质作为输入结构, 将流形映射到分子对接中涉及的各种自由度的乘积空间, 在标准测试集 PDBBind 上获得了 38% 的预测成功率, 相比于传统对接方法(23%)和基于深度学习的对接方法(20%), 成功率有明显提高^[68]。DiffBindFR 方法中的扩散模型则从具有随机侧链扭转角度的口袋构象开始, 学习整个系统内变量在由扭转角、旋转和平移组成的积空间中的联合优化, 不仅可以准确预测蛋白质口袋侧链构象, 而且还能生成精确的配体结合姿态^[69]。

AI 驱动分子对接方法通过在机器学习模型中加入几何约束, 可以更准确地预测蛋白质-配体复合物结合构象。为了考虑复杂结构的几何约束及局部功能区域的作用, TANKBind 将三角学约束作为归纳偏差构建到模型中, 并将整个蛋白质分割成功能块以明确关注所有可能的结合位点, 同时利用局部区域负采样构建新的对比损失, 共同优化结合相互作用和亲和力, 最终显著提高了复合物结构和配体亲和力的预测准确性^[70]。DeepDock 基于几何深度学习, 根

据配体-靶标对的距离似然学习统计势垒, 并将势垒与全局优化算法相结合, 用于预测配体与靶标蛋白质的结合构象^[71]。EquiBind 基于 SE(3) 等变几何深度学习模型, 可直接预测受体结合位置、配体的结合姿态和方向, 预测速度和质量都有显著提升^[72]。

与传统分子对接方法相比, AI 驱动分子对接方法的高精度还依赖于合理、充分的分子表示。当分子被当作一维序列标记或二维拓扑图处理时, 会限制三维几何预测或生成等任务中整合其三维信息的能力。针对此问题, Uni-Mol 方法提出了一种通用的 3D 分子表示学习(molecular representation learning, MRL) 框架, 它通过在 MRL 中纳入蛋白和其配体的海量三维结构信息, 显著扩大了 MRL 的表示能力和应用范围, 通过结合分子和蛋白口袋预训练模型, 以及对实验数据进行微调, 可高效完成蛋白三维结构相关的预测任务, 如蛋白质-配体结合姿态和分子构象生成等^[73]。

随着 AI 融合分子对接方法的涌现, 开发公允且准确的独立测试工具包以评估这些方法的性能变得非常有必要。PoseBusters 基于成熟的化学信息学工具包开发, 可用于 AI 分子对接方法的基准测试, 该工具不仅能够检查结合模式的 RMSD 等既定指标, 还能够评估配体结构、化学一致性和物理合理性, 并基于此找出缺失的诱导偏差, 改进基于深度学习的对接方法, 推动更精确、更高效的分子对接预测方法的开发^[96]。

2.2 AI 驱动分子动力学模拟技术

目前, 传统的分子动力学模拟在力场的准确性、轨迹解释的合理性、构象采样的充分性等方面面临一系列挑战。机器学习算法通过从现有轨迹中学习, 引导模拟趋向构象空间的目标区域, 可加速收敛、增加空间探索范围, 并

且在处理高维数据时可提供灵活、易处理的分析,因此可以准确模拟复杂大分子系统的动态行为,同时显著提高模拟速度。

传统分子动力学模拟使用的用于描述原子之间力的力场是物理驱动的,易于解释但不够精确,可移植性差。相较而言,由数据驱动的机器学习势能在保持与经典力场相当的计算效率的同时,达到了越来越高的精度。TorchMD 是一个具有混合经典和机器学习势的分子模拟框架,其将所有力计算(包括键、角度、二面角、Lennard-Jones 和库仑相互作用)均表示为 PyTorch 数组和操作,用于学习和模拟神经网络势(neural network potentials, NNPs),其已成功应用于一系列体系的全原子模拟和蛋白质折叠的粗粒化模拟^[75]。在 TorchMD 基础上引入新颖的等变 transformer (ET) 架构而构建 TorchMD-NET,通过使用集中设计注意力机制,还可以实现量子力学性质的预测^[97]。

在增强采样方面,深度学习模型可以用于构建和识别少量关键集合变量(collective variable, CV),再引入有利于其波动的偏置势,可实现增强采样。NN2B 基于最近邻密度估计器(nearest neighbor density estimator, NNDE)和人工神经网络(artificial neural network, ANN)方法,通过运行短时间偏置分子动力学模拟,获得可以描述不同 CV 间相关性的偏置势,实现多肽模拟的遍历采样和自由能精确计算^[77]。DEEP-VES 将机器学习技术与变异增强采样方法结合使用,利用神经网络的表达能力表示快速变化的自由能面,消除人为边界效应的同时允许操作多个 CV,从而确定引导系统波动的偏置势,该方法可应用于多种体系的不同维度自由能面的计算模拟^[98]。基于强化学习的自适应采样(reinforcement learning based adaptive sampling algorithm, REAP)算法,与单个连续分

子动力学模拟或传统自适应采样技术相比具有更快的构象空间探索速度^[99],由于该方法可以实时评估关键 CV 的重要性,因此在缺乏结构信息体系的模拟方面具有优势。

此外,使用大量短时间的 MD 模拟数据,可构建基于机器学习的马尔可夫状态模型(Markov state model, MSM),并用于预测长时间尺度的蛋白结构变化^[100]。通过应用广义主方程构建准马尔可夫状态模型(quasi-MSMs, qMSMs),可以减少对离散时间差的依赖,提高基于有限 MD 数据的构象动态变化预测精度。另外,结合深度学习进行自动特征选取或者对数据降维,还可将 MSMs 拓展应用于功能构象变化的研究,并对功能构象变化进行合理的解释^[101]。

2.3 AI 驱动的量子化学计算模拟技术

量子化学计算模拟是理解和预测分子化学性质不可或缺的工具。然而,由于求解薛定谔方程近似解非常困难,且随着分子系统模拟尺度的增加,计算成本也显著增加,导致传统方法在应用于生物大分子计算时面临重大挑战。AI 驱动的量子化学计算模拟技术可将量子力学的基本原理与机器学习的高效性融为一体,显著提升计算模拟性能,并加速对生物体系化学空间的高精度探索。

机器学习模型可以用于预测量子化学计算中的哈密顿量。Schütt 等^[102]提出了一个深度学习框架,其在局部原子轨道的基础上预测量子力学波函数,并从中衍生出所有其他基态特性,从而以力场计算相当的效率,对电子结构以可分析、可微分的描述方式进行量子力学计算。Nigam 等^[103]和 Unke 等^[104]在此基础上进一步引入等变机器学习技术以加快训练过程。Zhou 等^[105]通过将半经验哈密顿量与传统神经网络相结合,使得构建的混合半经验模型中哈密顿量的参数可对原子的化学环境发生动态响应。以上

这些模型不仅可以提高计算的准确性和效率，同时还具有很强的可迁移性和通用性。

由于神经网络可被看作一种函数近似方法，因此机器学习方法也已成功应用于波函数的预测与解析。神经网络波函数不同于传统的波函数，其不依赖于任何特定的基组。DeepMind 团队开发的费米子神经网络(fermionic neural network, FermiNet)是一种多电子系统的波函数拟设器，可用于薛定谔方程的近似解计算，其基于原子的位置和电荷这些有限数据，即可预测出强相关体系的解离曲线^[84]。Hermann 等^[79]开发的深度学习波函数拟设器 PauliNet，以内置多参考哈特里-福克(Hartree-Fock)解作为基线，集成有效波函数的物理特性，能够学习电子的复杂运动模式，快速获得电子薛定谔方程的近乎精确解。近期，DeepMind 团队再次开发了基于多头自注意力机制的新型神经网络架构波函数变换器 Psiformer，其可通过学习电子之间复杂的量子力学相关性，基于第一性原理求解多电子薛定谔方程，从而显著提高计算精度，尤其在较大分子基态能量的计算方面表现优异^[78]。

融合深度学习的计算方法可以有效降低分子基态和激发态性质的高精度计算成本。Ng 等^[106]构建了一个低数据深度神经网络模型，获得了与 MP2 和 CCSD 理论水平相当的机器学习电子相关能量，可应用于训练集中不包含的、更复杂的分子。Golub 等^[107]通过密度矩阵重正化群和单点熵计算构建训练集，基于神经网络预测活性轨道，实现活性空间的自动选择，从而减少了量子化学多组态方法中手动选择活性空间时对专业知识和经验的依赖。

在传统量子化学计算方法中，密度泛函理论(density functional theory, DFT)因其在计算准确性和计算成本方面具有良好的平衡而被广泛应用于各类体系的研究，但其精度高度依赖于

对交换-相关(exchange-correlation, XC)泛函的选择。在处理强相关电子系统和范德华相互作用时，标准 DFT 方法可能表现不佳。机器学习模型则可基于特定算法或者学习架构开发相对通用的、精确的泛函。Zheng 等^[108]采用神经网络方法通过考虑高阶贡献来改进 B3LYP 交换相关函数，而 Vargas Hernández^[109]使用贝叶斯优化和高斯过程对特定杂化和长程泛函的杂化参数进行微调，并用于校准不同的密度泛函模型。Nagai 等^[110]则基于参考分子中精确的密度分布和能量通过机器学习来系统地构建泛函，其使用灵活的前馈神经网络将密度和能量联系起来，并通过反向传播算法获取函数导数。

半经验量子化学(semiempirical quantum chemistry, SQC)计算方法可应用于较大体系或较高通量的大规模计算，但其计算精度有限。ML-SQC 方法使用机器学习模型来调整半经验量子化学 SQC 计算方法的参数，并用于反映分子组成和几何形状的变化，从而在不降低分子描述符的迁移性的同时提高准确度，因此可用于高通量筛选^[81]。人工智能-量子力学方法 1 (artificial intelligence-quantum mechanical method 1, AIQM1)基于半经验量子力学方法，结合神经网络势以及 Δ -机器学习(Δ -machine learning)进行参数优化，并使用迁移学习从有限数据中学习，从而以近似低水平半经验量子力学方法的速度、以“黄金标准”耦合簇量子力学(coupled cluster QM)方法的精度计算中性、闭壳层体系的基态能量^[82]。

2.4 AI 驱动的 QM/MM 计算模拟技术

量子力学/分子力学(QM/MM)组合方法由于对体系的化学反应位点和周围环境进行了分区处理，使得生物体系中化学反应的计算模拟变为可能。如前文所述，传统分子动力学模拟和量子化学计算都存在计算精度和计算效率方

面的瓶颈, AI 与物理原理的融合可有效解决这些问题, 因此将融合 AI 的分子力学方法和融合 AI 的量子化学方法组合构建 AI 驱动的 QM/MM 计算具有重要价值^[87-90]。例如, Pan 等^[11]构建了用于生成 MD 模拟能量和力参数的机器学习势和增量机器学习势, 并将分子体系表示拓展至每个 QM 区域原子的外部静电势和静电场, 在具体应用中可获得与传统 QM/MM 计算精度相当的能量和力值, 但计算效率大幅提高。在 QM 区域和 MM 区域的相互作用方面, Bösel 等^[90]在高维神经网络势(high-dimensional neural network potentials, HDNNP)中将 MM 区作为要素类型整合, 拟合的 HDNNP 用静电嵌入方案描述 QM 区的势能面, 因此 MM 区会接收到来自极化的 QM 区的作用, 同时引入 Δ -机器学习策略, 最终在使用较少参数的情况下, 达到密度泛函理论计算的精度。

3 结论与展望

基于物理原理的传统计算模拟方法在复杂生物体系的认识和理解方面发挥着重要作用, 而基于大数据的 AI 技术可有效提升其对参与生命活动不同层次元件的预测和设计能力。AI 与物理原理的融合既能弥补传统计算技术的不足, 又可推动计算模拟技术的发展及革新。

AI 技术在不依赖物理原理、仅依赖化学信息时, 需要大量的训练数据, 而引入物理原理, 不仅可降低非物理行为发生的可能性, 还可以提高模型的泛化和迁移能力, 从而减少对训练数据的依赖。使用微分物理模型, 并使用自动微分技术优化相关物理框架中的参数, 可增加机器学习模型的泛化能力, 从而可利用严谨的化学知识和既定的物理原理提供统计替代模型通常无法实现的解释能力, 并用于解释物理原理指导的化学反应过程。目前, 仍然有必要通

过量子化学计算等传统计算技术来辅助检验机器学习模型的预测结果, 以保证所预测结果的化学稳健性。

总之, AI 与物理原理融合驱动的计算模拟技术近年来取得了突破性进展, 在预测和设计生物大分子结构、功能和相互作用的准确性和效率方面得到显著提升。然而, 这些技术仍然面临许多挑战, 跨学科协作至关重要。计算化学家、物理学家、生物信息学家、人工智能专家和合成生物学家需要共同努力, 打破学科壁垒, 推动这些技术的发展和應用。通过不断实践和优化, AI 与物理原理融合驱动的计算模拟技术有望为生物大分子功能的研究提供更详尽、准确的见解, 加速生物制造所需高效蛋白元件的设计与创制, 从而为合成生物学和生物制造等领域带来深远的影响。

作者贡献声明

刘保艳: 方案设计、文献调研和初稿写作; 李帅: 文献调研和表格整理; 苏浩: 文献调研和初稿写作; 盛翔: 方案设计、监督指导和稿件润色修改。

作者利益冲突公开声明

作者声明没有任何可能会影响本文所报告工作的已知经济利益或个人关系。

REFERENCES

- [1] ZHANG J, CHEN D, XIA YJ, HUANG YP, LIN XH, HAN X, NI NX, WANG ZD, YU F, YANG LJ, YANG YI, GAO YQ. Artificial intelligence enhanced molecular simulations[J]. *Journal of Chemical Theory and Computation*, 2023, 19(14): 4338-4350.
- [2] 王晟, 王泽琛, 陈威华, 陈珂, 彭向达, 欧发芬, 郑良振, 孙璿原, 沈涛, 赵国屏. 基于人工智能和计算生物学的合成生物学元件设计[J]. *合成生物学*, 2023, 4(3): 422-443.
WANG S, WANG ZC, CHEN WH, CHEN K, PENG XD, OU FF, ZHENG LZ, SUN JY, SHEN T, ZHAO GP. Design of synthetic biology components based on

- artificial intelligence and computational biology[J]. *Synthetic Biology Journal*, 2023, 4(3): 422-443 (in Chinese).
- [3] ABRAMSON J, ADLER J, DUNGER J, EVANS R, GREEN T, PRITZEL A, RONNEBERGER O, WILLMORE L, BALLARD AJ, BAMBRICK J, BODENSTEIN SW, EVANS DA, HUNG CC, O'NEILL M, REIMAN D, TUNYASUVUNAKOOL K, WU Z, ŽEMGULYTĖ A, ARVANITI E, BEATTIE C, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3[J]. *Nature*, 2024, 630(8016): 493-500.
- [4] YEH AHW, NORN C, KIPNIS Y, TISCHER D, PELLOCK SJ, EVANS D, MA PC, LEE GR, ZHANG JZ, ANISHCHENKO I, COVENTRY B, CAO LX, DAUPARAS J, HALABIYA S, DeWITT M, CARTER L, HOUK KN, BAKER D. *De novo* design of luciferases using deep learning[J]. *Nature*, 2023, 614(7949): 774-780.
- [5] ANISHCHENKO I, PELLOCK SJ, CHIDYUSIKU TM, RAMELOT TA, OVCHINNIKOV S, HAO JZ, BAFNA K, NORN C, KANG A, BERA AK, DiMAIO F, CARTER L, CHOW CM, MONTELLIONE GT, BAKER D. *De novo* protein design by deep network hallucination[J]. *Nature*, 2021, 600(7889): 547-552.
- [6] DOU JY, VOROBIEVA AA, SHEFFLER W, DOYLE LA, PARK H, BICK MJ, MAO BC, FOIGHT GW, LEE MY, GAGNON LA, CARTER L, SANKARAN B, OVCHINNIKOV S, MARCOS E, HUANG PS, VAUGHAN JC, STODDARD BL, BAKER D. *De novo* design of a fluorescence-activating β -barrel[J]. *Nature*, 2018, 561(7724): 485-491.
- [7] ABADI M, BARHAM P, CHEN J, CHEN ZF, DAVIS A, DEAN J, DEVIN M, GHEMAWAT S, IRVING G, ISARD M, KUDLUR M, LEVENBERG J, MONGA R, MOORE S, MURRAY DG, STEINER B, TUCKER P, VASUDEVAN V, WARDEN P, WICKE M, et al. TensorFlow: a System for Large-scale Machine Learning[C]. 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), 2016: 265-283.
- [8] PASZKE A, GROSS S, MASSA F, LERER A, BRADBURY J, CHANAN G, KILLEEN T, LIN Z, GIMELSHEIN N, ANTIGA L, Desmaison A, KOPF A, YANG E, DEVITO Z, RAISON M, TEJANI A, CHILAMKURTHY S, STEINER B, FANG L, BAI JJ, CHINTALA S. Pytorch: an imperative style, high-performance deep learning library[J]. *Advances in Neural Information Processing Systems*, 2019: 32.
- [9] Huawei Technologies Co., Ltd. Huawei MindSpore AI development framework[M]//Artificial Intelligence Technology. Singapore: Springer Nature Singapore, 2022: 137-162.
- [10] FRIESNER RA, BANKS JL, MURPHY RB, HALGREN TA, KLICIC JJ, MAINZ DT, REPASKY MP, KNOLL EH, SHELLEY M, PERRY JK, SHAW DE, FRANCIS P, SHENKIN PS. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy[J]. *Journal of Medicinal Chemistry*, 2004, 47(7): 1739-1749.
- [11] HALGREN TA, MURPHY RB, FRIESNER RA, BEARD HS, FRYE LL, POLLARD WT, BANKS JL. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening[J]. *Journal of Medicinal Chemistry*, 2004, 47(7): 1750-1759.
- [12] JONES G, WILLETT P, GLEN RC, LEACH AR, TAYLOR R. Development and validation of a genetic algorithm for flexible docking[J]. *Journal of Molecular Biology*, 1997, 267(3): 727-748.
- [13] MORRIS GM, HUEY R, LINDSTROM W, SANNER MF, BELEW RK, GOODSSELL DS, OLSON AJ. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility[J]. *Journal of Computational Chemistry*, 2009, 30(16): 2785-2791.
- [14] TROTT O, OLSON AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading[J]. *Journal of Computational Chemistry*, 2010, 31(2): 455-461.
- [15] EBERHARDT J, SANTOS-MARTINS D, TILLACK AF, FORLI S. AutoDock Vina 1.2.0: new docking methods, expanded force field, and python bindings[J]. *Journal of Chemical Information and Modeling*, 2021, 61(8): 3891-3898.
- [16] YAN YM, ZHANG D, ZHOU P, LI BT, HUANG SY. HDock: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy[J]. *Nucleic Acids Research*, 2017, 45(W1): W365-W373.
- [17] MARZE NA, ROY BURMAN SS, SHEFFLER W, GRAY JJ. Efficient flexible backbone protein-protein docking for challenging targets[J]. *Bioinformatics*, 2018, 34(20): 3461-3469.
- [18] PIERCE BG, WIEHE K, HWANG H, KIM BH, VREVEN T, WENG ZP. ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers[J]. *Bioinformatics*, 2014, 30(12): 1771-1773.
- [19] PAUL DS, GAUTHAM N. MOLS 2.0: software package for peptide modeling and protein-ligand docking[J]. *Journal of Molecular Modeling*, 2016, 22(10): 239.
- [20] MURALIDHARA BK, NEGI SS, HALPERT JR. Dissecting the thermodynamics and cooperativity of ligand binding in cytochrome P450eryF[J]. *Journal of the American Chemical Society*, 2007, 129(7): 2015-2024.
- [21] de VOSS JJ, SIBBESEN O, ZHANG ZP, ORTIZ de MONTELLANO PR. Substrate docking algorithms and prediction of the substrate specificity of cytochrome P450cam and its L244A mutant[J]. *Journal of the American Chemical Society*, 1997, 119(24): 5489-5498.
- [22] PAN H, AGARWALLA S, MOUSTAKAS DT, FINER-MOORE J, STROUD RM. Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2003, 100(22): 12648-12653.
- [23] YEH JI, LEVINE AS, DU SC, CHINTE U, GHODKE H, WANG H, SHI HB, HSIEH CL, CONWAY JF, van HOUTEN B, RAPIĆ-OTRIN V. Damaged DNA induced UV-damaged DNA-binding protein (UV-DDB)

- dimerization and its roles in chromatinized DNA repair[J]. Proceedings of the National Academy of Sciences of the United States of America, 2012, 109(41): E2737-E2746.
- [24] WAN XB, YANG TP, CUESTA A, PANG XM, BALIUS TE, IRWIN JJ, SHOICHET BK, TAUNTON J. Discovery of lysine-targeted eIF4E inhibitors through covalent docking[J]. Journal of the American Chemical Society, 2020, 142(11): 4960-4964.
- [25] ROSNIZECK IC, SPOERNER M, HARSCH T, KREITNER S, FILCHTINSKI D, HERRMANN C, ENGEL D, KÖNIG B, KALBITZER HR. Metal-bis(2-picoyl)amine complexes as state 1(T) inhibitors of activated Ras protein[J]. Angewandte Chemie International Edition, 2012, 51(42): 10647-10651.
- [26] CASE DA, AKTULGA HM, BELFON K, BEN-SHALOM IY, BERRYMAN JT, BROZELL SR, CERUTTI DS, CHEATHAM III TE, CISNEROS GA, CRUZEIRO VWD, DARDEN TA, FOROUZESH N, GHAZIMIRSAEED M, GIAMBAŞU G, GIESE T, GILSON MK, GOHLKE H, GOETZ AW, HARRIS J, HUANG Z, et al. Amber 2024[CP]. University of California, San Francisco, 2024.
- [27] KÜHNE TD, IANNUZZI M, del BEN M, RYBKIN VV, SEEWALD P, STEIN F, LAINO T, KHALIULLIN RZ, SCHÜTT O, SCHIFFMANN F, GOLZE D, WILHELM J, CHULKOV S, BANI-HASHEMIAN MH, WEBER V, BORŠTNIK U, TAILLEFUMIER M, JAKOBOVITS AS, LAZZARO A, PABST H, et al. CP2K: an electronic structure and molecular dynamics software package-quickstep: efficient and accurate electronic structure calculations[J]. Journal of Chemical Physics, 2020, 152(19): 194103.
- [28] van der SPOEL D, LINDAHL E, HESS B, GROENHOF G, MARK AE, BERENDSEN HJC. GROMACS: fast, flexible, and free[J]. Journal of Computational Chemistry, 2005, 26(16): 1701-1718.
- [29] MELO MCR, BERNARDI RC, RUDACK T, SCHEURER M, RIPLINGER C, PHILLIPS JC, MAIA JDC, ROCHA GB, RIBEIRO JV, STONE JE, NEESE F, SCHULTEN K, LUTHEY-SCHULTEN Z. NAMD goes quantum: an integrative suite for hybrid simulations[J]. Nature Methods, 2018, 15(5): 351-354.
- [30] THOMPSON AP, AKTULGA HM, BERGER R, BOLINTINEANU DS, BROWN WM, CROZIER PS, IN 'T VELD PJ, KOHLMAYER A, MOORE SG, NGUYEN TD, SHAN R, STEVENS MJ, TRANCHIDA J, TROTT C, PLIMPTON SJ. LAMMPS-a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales[J]. Computer Physics Communications, 2022, 271: 108171.
- [31] BROOKS BR, BROOKS CL 3rd, MACKERELL AD Jr, NILSSON L, PETRELLA RJ, ROUX B, WON Y, ARCHONTIS G, BARTELS C, BORESCH S, CAFLISCH A, CAVES L, CUI Q, DINNER AR, FEIG M, FISCHER S, GAO J, HODOSCEK M, IM W, KUCZERA K, et al. CHARMM: the biomolecular simulation program[J]. Journal of Computational Chemistry, 2009, 30(10): 1545-1614.
- [32] CHEN QM, XIAO YQ, SHAKHNOVICH EI, ZHANG WL, MU WM. Semi-rational design and molecular dynamics simulations study of the thermostability enhancement of cellobiose 2-epimerases[J]. International Journal of Biological Macromolecules, 2020, 154: 1356-1365.
- [33] SBORGI L, VERMA A, PIANA S, LINDORFF-LARSEN K, CERMINARA M, SANTIVERI CM, SHAW DE, de ALBA E, MUÑOZ V. Interaction networks in protein folding *via* atomic-resolution experiments and long-time-scale molecular dynamics simulations[J]. Journal of the American Chemical Society, 2015, 137(20): 6506-6516.
- [34] TAN YS, SPRING DR, ABELL C, VERMA CS. The application of ligand-mapping molecular dynamics simulations to the rational design of peptidic modulators of protein-protein interactions[J]. Journal of Chemical Theory and Computation, 2015, 11(7): 3199-3210.
- [35] PARK H, HONG S, KIM J, HONG S. Discovery of picomolar ABL kinase inhibitors equipotent for wild type and T315I mutant *via* structure-based *de novo* design[J]. Journal of the American Chemical Society, 2013, 135(22): 8227-8237.
- [36] NARAYAN ARH, JIMÉNEZ-OSÉS G, LIU P, NEGRETTI S, ZHAO WX, GILBERT MM, RAMABHADRAN RO, YANG YF, FURAN LR, LI Z, PODUST LM, MONTGOMERY J, HOUK KN, SHERMAN DH. Enzymatic hydroxylation of an unactivated methylene C-H bond guided by molecular dynamics simulations[J]. Nature Chemistry, 2015, 7(8): 653-660.
- [37] AHUJA LG, AOTO PC, KORNEV AP, VEGLIA G, TAYLOR SS. Dynamic allostery-based molecular workings of kinase: peptide complexes[J]. Proceedings of the National Academy of Sciences of the United States of America, 2019, 116(30): 15052-15061.
- [38] LIGHTSTONE FC, ZHENG YJ, BRUCE TC. Molecular dynamics simulations of ground and transition states for the S_N2 displacement of Cl⁻ from 1,2-dichloroethane at the active site of *Xanthobacter autotrophicus* haloalkane dehalogenase[J]. Journal of the American Chemical Society, 1998, 120(23): 5611-5621.
- [39] ISHITANI R, SUGITA Y, DOHMAE N, FURUYA N, HATTORI M, NUREKI O. Mg²⁺-sensing mechanism of Mg²⁺ transporter MgtE probed by molecular dynamics study[J]. Proceedings of the National Academy of Sciences of the United States of America, 2008, 105(40): 15393-15398.
- [40] FRISCH MJ, TRUCKS GW, SCHLEGEL HB, SCUSERIA GE, ROBB MA, CHEESEMAN JR, SCALMANI G, BARONE V, PETERSSON GA, NAKATSUJI H, LI X, CARICATO M, MARENICH AV, BLOINO J, JANESKO BG, GOMPERS R, MENNUCCI B, HRATCHIAN HP, ORTIZ JV, IZMAYLOV AF, et al. Gaussian 16[Z]. 2016.
- [41] AHLRICHS R, BÄR M, HÄSER M, HORN H, KÖLMEL C. Electronic structure calculations on workstation computers: the program system turbomole[J]. Chemical Physics Letters, 1989, 162(3): 165-169.
- [42] NEESE F, WENMOHS F, BECKER U, RIPLINGER C. The ORCA quantum chemistry program package[J].

- Journal of Chemical Physics, 2020, 152(22): 224108.
- [43] EPIFANOVSKY E, GILBERT ATB, FENG XT, LEE J, MAO YZ, MARDIROSIAN N, POKHILKO P, WHITE AF, COONS MP, DEMPWOLFF AL, GAN ZT, HAIT D, HORN PR, JACOBSON LD, KALIMAN I, KUSSMANN J, LANGE AW, LAO KU, LEVINE DS, LIU J, et al. Software for the frontiers of quantum chemistry: an overview of developments in the Q-chem 5 package[J]. Journal of Chemical Physics, 2021, 155(8): 084801.
- [44] APRÁ E, BYLASKA EJ, de JONG WA, GOVIND N, KOWALSKI K, STRAATSMA TP, VALIEV M, van DAM HJJ, ALEXEEV Y, ANCHELL J, ANISIMOV V, AQUINO FW, ATTA-FYNN R, AUTSCHBACH J, BAUMAN NP, BECCA JC, BERNHOLDT DE, BHASKARAN-NAIR K, BOGATKO S, BOROWSKI P, et al. NWChem: past, present, and future[J]. Journal of Chemical Physics, 2020, 152(18): 184102.
- [45] WERNER HJ, KNOWLES PJ, MANBY FR, BLACK JA, DOLL K, HEBELMANN A, KATS D, KÖHN A, KORONA T, KREPLIN DA, MA QL, MILLER TF 3rd, MITRUSHCHENKOV A, PETERSON KA, POLYAK I, RAUHUT G, SIBAEV M. The Molpro quantum chemistry package[J]. Journal of Chemical Physics, 2020, 152(14): 144107.
- [46] SHENG X, LIND MES, HIMO F. Theoretical study of the reaction mechanism of phenolic acid decarboxylase[J]. The FEBS Journal, 2015, 282(24): 4703-4713.
- [47] LIAO RZ, SIEGBAHN PEM. Energetics for the mechanism of nickel-containing carbon monoxide dehydrogenase[J]. Inorganic Chemistry, 2019, 58(12): 7931-7938.
- [48] SHENG X, HIMO F. Enzymatic pictet-spengler reaction: computational study of the mechanism and enantioselectivity of norcoclaurine synthase[J]. Journal of the American Chemical Society, 2019, 141(28): 11230-11238.
- [49] DONG YJ, LI T, ZHANG SQ, SANCHIS J, YIN H, REN J, SHENG X, LI GY, REETZ MT. Biocatalytic baeyer-villiger reactions: uncovering the source of regioselectivity at each evolutionary stage of a mutant with scrutiny of fleeting chiral intermediates[J]. ACS Catalysis, 2022, 12(6): 3669-3680.
- [50] ZHANG LB, ZHANG SQ, LIAO LJ, TANG HY, WANG WL, YIN FC, HAN LL, ZHU KJ, LIU YS, XU DQ, WANG XB, QIN MJ, XIAO YB, SHENG X, ZHAO YC. Discovery, structure, and mechanism of the (*R,S*)-norcoclaurine synthase for the chiral synthesis of benzylisoquinoline alkaloids[J]. ACS Catalysis, 2023, 13(22): 15164-15174.
- [51] MEELUA W, WANJAI T, THINKUMROB N, OLÁH J, CAIRNS JRK, HANNONGBUA S, RYDE U, JITONNOM J. A computational study of the reaction mechanism and stereospecificity of dihydropyrimidinase[J]. Physical Chemistry Chemical Physics, 2023, 25(12): 8767-8778.
- [52] METZ S, KÄSTNER J, SOKOL AA, KEAL TW, SHERWOOD P. ChemShell: a modular software package for QM/MM simulations[J]. WIREs Computational Molecular Science, 2014, 4(2): 101-110.
- [53] LARSEN AH, MORTENSEN JJ, BLOMQVIST J, CASTELLI IE, CHRISTENSEN R, DUŁAK M, FRIIS J, GROVES MN, HAMMER B, HARGUS C, HERMES ED, JENNINGS PC, JENSEN PB, KERMODE J, KITCHIN JR, KOLSBJERG EL, KUBAL J, KAASBJERG K, LYSGAARD S, MARONSSON JB, et al. The atomic simulation environment—a Python library for working with atoms[J]. Journal of Physics Condensed Matter, 2017, 29(27): 273002.
- [54] ŘEZAČ J. Cuby: an integrative framework for computational chemistry[J]. Journal of Computational Chemistry, 2016, 37(13): 1230-1237.
- [55] OLSEN JMH, BOLNYKH V, MELONI S, IPPOLITI E, BIRCHER MP, CARLONI P, ROTHLSBERGER U. MiMiC: a novel framework for multiscale modeling in computational chemistry[J]. Journal of Chemical Theory and Computation, 2019, 15(6): 3810-3823.
- [56] MARTÍ S. QMCube (QM³): an all-purpose suite for multiscale QM/MM calculations[J]. Journal of Computational Chemistry, 2021, 42(6): 447-457.
- [57] SHAIK S, DUBEY KD. The catalytic cycle of cytochrome P450: a fascinating choreography[J]. Trends in Chemistry, 2021, 3(12): 1027-1044.
- [58] COINES J, RAICH L, ROVIRA C. Modeling catalytic reaction mechanisms in glycoside hydrolases[J]. Current Opinion in Chemical Biology, 2019, 53: 183-191.
- [59] ESCORCIA AM, SEN K, DAZA MC, DOERR M, THIEL W. Quantum mechanics/molecular mechanics insights into the enantioselectivity of the *O*-acetylation of (*R,S*)-propranolol catalyzed by *Candida antarctica* lipase B[J]. ACS Catalysis, 2017, 7(1): 115-127.
- [60] LONSDALE R, HOUGHTON KT, ŽUREK J, BATHELT CM, FOLOPPE N, de GROOT MJ, HARVEY JN, MULHOLLAND AJ. Quantum mechanics/molecular mechanics modeling of regioselectivity of drug metabolism in cytochrome P450 2C9[J]. Journal of the American Chemical Society, 2013, 135(21): 8001-8015.
- [61] CHARCHAR P, CHRISTOFFERSON AJ, TODOROVA N, YAROVSKY I. Understanding and designing the gold-bio interface: insights from simulations[J]. Small, 2016, 12(18): 2395-2418.
- [62] LIN HY, CHEN X, DONG J, YANG JF, XIAO H, YE Y, LI LH, ZHAN CG, YANG WC, YANG GF. Rational redesign of enzyme *via* the combination of quantum mechanics/molecular mechanics, molecular dynamics, and structural biology study[J]. Journal of the American Chemical Society, 2021, 143(38): 15674-15687.
- [63] HOLLINGSWORTH SA, DROR RO. Molecular dynamics simulation for all[J]. Neuron, 2018, 99(6): 1129-1143.
- [64] HIMO F. Recent trends in quantum chemical modeling of enzymatic reactions[J]. Journal of the American Chemical Society, 2017, 139(20): 6780-6786.
- [65] SHENG X, HIMO F. The quantum chemical cluster approach in biocatalysis[J]. Accounts of Chemical Research, 2023, 56(8): 938-947.
- [66] MCCARTHY J, MINSKY ML, ROCHESTER N, SHANNON CE. “A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955”[J]. AI magazine, 2006, 27(4): 12-12.

- [67] 刘海燕. 人工智能: 开启生物体系计算设计的新篇章[J]. 合成生物学, 2023, 4(3): 419-421.
LIU HY. Artificial intelligence: opening a new chapter in computational design of biological systems[J]. Synthetic Biology Journal, 2023, 4(3): 419-421 (in Chinese).
- [68] CORSO G, STÄRK H, JING B, BARZILAY R, JAAKKOLA T. DiffDock: diffusion steps, twists, and turns for molecular docking[Z]. arXiv, 2023.
- [69] ZHU JT, GU ZH, PEI JF, LAI LH. DiffBindFR: an SE(3) equivariant network for flexible protein-ligand docking[J]. Chemical Science, 2024, 15(21): 7926-7942.
- [70] LU W, WU QF, ZHANG JX, RAO JH, LI CT, ZHENG SJ. TANKBind: trigonometry-aware neural networks for drug-protein binding structure prediction[J]. Advances in Neural Information Processing Systems, 2022, 35: 7236-7249.
- [71] MÉNDEZ-LUCIO O, AHMAD M, del RIO-CHANONA EA, WEGNER JK. A geometric deep learning approach to predict binding conformations of bioactive molecules[J]. Nature Machine Intelligence, 2021, 3(12): 1033-1039.
- [72] STÄRK H, GANEA O, PATTANAIK L, BARZILAY DR, JAAKKOLA T. EquiBind: geometric deep learning for drug binding structure prediction[C]. Proceedings of the 39th International Conference on Machine Learning. PMLR, 2022: 20503-20521.
- [73] ZHOU GM, GAO ZF, DING QK, ZHENG H, XU HT, WEI ZW, ZHANG LF, KE GL. Uni-mol: a universal 3D molecular representation learning framework[C]. ChemRxiv. 2023; doi:10.26434/chemrxiv-2022-jjm0j-v4.
- [74] DO HN, WANG JN, BHATTARAI A, MIAO YL. GLOW: a workflow integrating Gaussian-accelerated molecular dynamics and deep learning for free energy profiling[J]. Journal of Chemical Theory and Computation, 2022, 18(3): 1423-1436.
- [75] DOERR S, MAJEWSKI M, PÉREZ A, KRÄMER A, CLEMENTI C, NOE F, GIORGINO T, de FABRITIIS G. TorchMD: a deep learning framework for molecular simulations[J]. Journal of Chemical Theory and Computation, 2021, 17(4): 2355-2363.
- [76] RIBEIRO JML, BRAVO P, WANG YH, TIWARY P. Reweighted autoencoded variational Bayes for enhanced sampling (RAVE)[J]. Journal of Chemical Physics, 2018, 149(7): 072301.
- [77] GALVELIS R, SUGITA Y. Neural network and nearest neighbor algorithms for enhancing sampling of molecular dynamics[J]. Journal of Chemical Theory and Computation, 2017, 13(6): 2489-2500.
- [78] GLEHN IV, SPENCER J, PFAU D. A self-attention Ansatz for ab-initio quantum chemistry[Z]. arXiv, 2023: 2211.13672.
- [79] HERMANN J, SCHÄTZLE Z, NOÉ F. Deep-neural-network solution of the electronic Schrödinger equation[J]. Nature Chemistry, 2020, 12(10): 891-897.
- [80] RAMAKRISHNAN R, DRAL PO, RUPP M, von LILIENFELD OA. Big data meets quantum chemistry approximations: the Δ -machine learning approach[J]. Journal of Chemical Theory and Computation, 2015, 11(5): 2087-2096.
- [81] DRAL PO, von LILIENFELD OA, THIEL W. Machine learning of parameters for accurate semiempirical quantum chemical calculations[J]. Journal of Chemical Theory and Computation, 2015, 11(5): 2120-2125.
- [82] ZHENG PK, ZUBATYUK R, WU W, ISAYEV O, DRAL PO. Artificial intelligence-enhanced quantum chemical method with broad applicability[J]. Nature Communications, 2021, 12(1): 7022.
- [83] GAO X, RAMEZANGHORBANI F, ISAYEV O, SMITH JS, ROITBERG AE. TorchANI: a free and open source PyTorch-based deep learning implementation of the ANI neural network potentials[J]. Journal of Chemical Information and Modeling, 2020, 60(7): 3408-3415.
- [84] PFAU D, SPENCER JS, MATTHEWS AGDG, FOULKES WMC. *Ab initio* solution of the many-electron Schrödinger equation with deep neural networks[J]. Physical Review Research, 2020, 2(3): 033429.
- [85] LEWIS AM, GRISAFI A, CERIOTTI M, ROSSI M. Learning electron densities in the condensed phase[J]. Journal of Chemical Theory and Computation, 2021, 17(11): 7203-7214.
- [86] GRISAFI A, LEWIS AM, ROSSI M, CERIOTTI M. Electronic-structure properties from atom-centered predictions of the electron density[J]. Journal of Chemical Theory and Computation, 2023, 19(14): 4451-4460.
- [87] KIM B, SHAO YH, PU JZ. Doubly polarized QM/MM with machine learning chaperone polarizability[J]. Journal of Chemical Theory and Computation, 2021, 17(12): 7682-7695.
- [88] SHEN L, WU JH, YANG WT. Multiscale quantum mechanics/molecular mechanics simulations with neural networks[J]. Journal of Chemical Theory and Computation, 2016, 12(10): 4934-4946.
- [89] SHEN L, YANG WT. Molecular dynamics simulations with quantum mechanics/molecular mechanics and adaptive neural networks[J]. Journal of Chemical Theory and Computation, 2018, 14(3): 1442-1455.
- [90] BÖSELT L, THÜRLEMANN M, RINIKER S. Machine learning in QM/MM molecular dynamics simulations of condensed-phase systems[J]. Journal of Chemical Theory and Computation, 2021, 17(5): 2641-2658.
- [91] BORAH NK, TRIPATHI Y, TANWAR A, TIWARI D, SINHA A, SHARMA S, JABALIA N, MANI RJ, SANTOSHI S, BANSAL H. Artificial Intelligence-powered Molecular Docking[M]. John Wiley & Sons, Ltd, 2024: 157-188.
- [92] RAGOZA M, HOCHULI J, IDROBO E, SUNSERI J, KOES DR. Protein-ligand scoring with convolutional neural networks[J]. Journal of Chemical Information and Modeling, 2017, 57(4): 942-957.
- [93] BAO JX, HE X, ZHANG JZH. DeepBSP-a machine learning method for accurate prediction of protein-ligand docking structures[J]. Journal of Chemical Information and Modeling, 2021, 61(5): 2231-2240.
- [94] McNUTT AT, FRANCOEUR P, AGGARWAL R, MASUDA T, MELI R, RAGOZA M, SUNSERI J, KOES DR. GNINA 1.0: molecular docking with deep learning[J]. Journal of Cheminformatics, 2021, 13(1): 43.
- [95] STEFANIAK F, BUJNICKI JM. AnnapuRNA: a

- scoring function for predicting RNA-small molecule binding poses[J]. *PLoS Computational Biology*, 2021, 17(2): e1008309.
- [96] BUTTENSCHOEN M, MORRIS GM, DEANE CM. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences[J]. *Chemical Science*, 2023, 15(9): 3130-3139.
- [97] THÖLKE P, FABRITIIS GD. TorchMD-NET: equivariant transformers for neural network based molecular potentials[Z]. arXiv, 2022: 2202.02541.
- [98] BONATI L, ZHANG YY, PARRINELLO M. Neural networks-based variationally enhanced sampling[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2019, 116(36): 17641-17647.
- [99] SHAMSI Z, CHENG KJ, SHUKLA D. Reinforcement learning based adaptive sampling: REAPing rewards by exploring protein conformational landscapes[J]. *The Journal of Physical Chemistry B*, 2018, 122(35): 8386-8395.
- [100] WENG JW, YANG MH, WANG WN, XU X, TIAN ZQ. Revealing thermodynamics and kinetics of lipid self-assembly by Markov state model analysis[J]. *Journal of the American Chemical Society*, 2020, 142(51): 21344-21352.
- [101] KONOVALOV KA, UNARTA IC, CAO SQ, GOONETILLEKE EC, HUANG XH. Markov state models to study the functional dynamics of proteins in the wake of machine learning[J]. *JACS Au*, 2021, 1(9): 1330-1341.
- [102] SCHÜTT KT, GASTEGGER M, TKATCHENKO A, MÜLLER KR, MAURER RJ. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions[J]. *Nature Communications*, 2019, 10(1): 5024.
- [103] NIGAM J, WILLATT MJ, CERIOTTI M. Equivariant representations for molecular Hamiltonians and N-center atomic-scale properties[J]. *Journal of Chemical Physics*, 2022, 156(1): 014115.
- [104] UNKE OT, BOGOJESKI M, GASTEGGER M, GEIGER M, SMIDT T, MÜLLER KR. SE (3)-equivariant prediction of molecular wavefunctions and electronic densities[J]. *Advances in Neural Information Processing Systems*, 2021, 34: 14434-14447.
- [105] ZHOU GQ, LUBBERS N, BARROS K, TRETIK S, NEBGEN B. Deep learning of dynamically responsive chemical Hamiltonians with semiempirical quantum mechanics[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2022, 119(27): e2120333119.
- [106] NG WP, LIANG QJ, YANG J. Low-data deep quantum chemical learning for accurate MP2 and coupled-cluster correlations[J]. *Journal of Chemical Theory and Computation*, 2023, 19(16): 5439-5449.
- [107] GOLUB P, ANTALIK A, VEIS L, BRABEC J. Machine learning-assisted selection of active spaces for strongly correlated transition metal systems[J]. *Journal of Chemical Theory and Computation*, 2021, 17(10): 6053-6072.
- [108] ZHENG X, HU LH, WANG XJ, CHEN GH. A generalized exchange-correlation functional: the neural-networks approach[J]. *Chemical Physics Letters*, 2004, 390(1/2/3): 186-192.
- [109] VARGAS HERNÁNDEZ RA. Bayesian optimization for calibrating and selecting hybrid-density functional models[J]. *The Journal of Physical Chemistry A*, 2020, 124(20): 4053-4061.
- [110] NAGAI R, AKASHI R, SUGINO O. Completing density functional theory by machine learning hidden messages from molecules[J]. *NPJ Computational Materials*, 2020, 6: 43.
- [111] PAN XL, YANG JJ, VAN R, EPIFANOVSKY E, HO J, HUANG J, PU JZ, MEI Y, NAM K, SHAO YH. Machine-learning-assisted free energy simulation of solution-phase and enzyme reactions[J]. *Journal of Chemical Theory and Computation*, 2021, 17(9): 5745-5758.