科学通报 2022 年 第 67 卷 第 35 期: 4180 ~ 4191
污染物与生物大分子互作及效应专题 评 述
《中国科学》杂志社 SCIENCE CHINA PRESS

基于计算毒理的环境污染物-生物大分子的相互作用研究

谭皓月[1,2], 张荣[1], 陈钦畅[1], 张驰[1], 郭婧[1], 张效伟[1], 于红霞[1], 史薇[1,2*]

1. 南京大学环境学院, 污染控制与资源化研究国家重点实验室, 南京 210023;
2. 江苏省生态环境保护化学品安全与健康风险研究重点实验室, 南京 210023
* 联系人, E-mail: njushiwei@nju.edu.cn

2022-05-30 收稿, 2022-07-14 修回, 2022-08-04 接受, 2022-08-05 网络版发表
国家自然科学基金(21922603)、江苏省重点研发计划(社会发展)(BE2022837)、国家重点研发计划(2018YFC1801604)、江苏省环境保护科研项目(2021003, 2018001)、中央高校基础研究经费(021114380139)和污染控制与资源再利用国家重点实验室基金(PCR-ZZ-202103)资助

摘要 计算机技术的革新以及结构生物学和深度学习的爆发式发展, 促使计算毒理学迅速应用到环境新污染物领域, 通过研究环境污染物与生物大分子间的相互作用解析致毒机制并筛查高风险毒物. 本文系统综述了用于解析环境污染物-生物大分子互作过程常用的计算毒理方法, 包括分子对接、分子动力学模拟和机器学习建模; 阐明了近年来3类计算毒理方法在环境毒理学领域的主要用途, 包括环境污染物-生物大分子相互作用的构效关系研究与虚拟筛查应用; 讨论了3类计算毒理方法在配体-受体结合、可解释性、计算效率、计算深度、生物学过程5个方面的优势与局限; 展望了计算毒理技术与新型技术(如人工智能)结合后的未来应用和进一步潜力.

关键词 计算毒理学, 分子对接, 分子动力学模拟, 机器学习, 虚拟筛查, 机制解析

通过生产作业、产品消费、药物服用等途径, 人类和动物在一生中持续不断地接触到环境污染物(environmental contaminants, ECs). 长期暴露于结构各异的ECs会对人体及动物体造成一系列不良影响, 比如生殖障碍、认知缺陷、代谢失调和组织癌变[1]. 面对ECs对人类存在的潜在有害影响, 世界各地的监管机构要求整合流行病学(epidemiology)、体内毒理学(*in vivo* toxicology)和体外机制(*in vitro* mechanism)数据, 以便为化学品的危害分类、标签和风险管理提供必要信息. 许多监管机构都建立了基于动物的体内(*in vivo*)实验对化学品的多种毒性(例如致癌性、生殖毒性等)进行检测. 由于机制信息有限、筛选耗时较长、违背动物伦理、实验成本昂贵等问题, 低通量的体内测试无法应对10万种化学品的毒性筛查需求[2,3].

随着毒理学领域的技术发展, 一些替代动物实验的体外实验应运而生[4]. 体外实验主要基于ECs的致毒机制, 利用非细胞(cell-free)和细胞的(cell-based)受试模型在分子及细胞水平检测ECs的毒性效应. 而有害结局路径(adverse outcome pathway, AOP)可以为ECs的致毒机制提供理论框架[5]. 作为理论框架, AOP包含三大主要部分: 分子启动事件(molecular initiating event, MIE)、有害结局(adverse outcome, AO)以及连接MIE与AO的一系列关键事件(key events, KEs)[6]. 首先, ECs与生物大分子产生相互作用并形成配体-受体复合体. 配体-受体相互作用(ligand-receptor interaction)是ECs产生毒性的关键第一步, 即为MIE. 配体-受体复合体经过变构进而导致基因异常表达等KEs, 最终导致生殖毒性、致癌性等AOs. ECs-生物大分子相互作用过程(本研究针对非共价键的分子间相互作用)作为关键第一步直接决定ECs致毒的可能. 因此, 面向生物大分子进行化学品的批量体外实验测试, 即高通量筛查, 已成为发现环境新污染物(emerging contaminants)的黄金标

引用格式: 谭皓月, 张荣, 陈钦畅, 等. 基于计算毒理的环境污染物-生物大分子的相互作用研究. 科学通报, 2022, 67: 4180–4191
Tan H Y, Zhang R, Chen Q C, et al. Computational toxicology studies on the interactions between environmental contaminants and biomacromolecules (in Chinese). Chin Sci Bull, 2022, 67: 4180–4191, doi: 10.1360/TB-2022-0613

准[7,8].

然而，建立与维护高通量体外筛查平台的高额费用仍然阻碍了该标准在发现环境新污染物方面的应用. 此外，由于近年来结构生物学[9,10]与和深度学习(deep learning, DL, 如AlphaFold2和AF2Complex)[11,12]增量式地解析了大量的生物大分子晶体结构，计算毒理方法(computational toxicology techniques)在环境毒理学领域的使用显著增加. 据估计，针对化合物-生物大分子互作过程，计算毒理可在有限的时间内对数百万个化合物进行虚拟筛查[13,14]. 因此，计算毒理方法可以降低识别新污染物的初始实验成本，增加环境污染物的毒性机制信息，提高监管部门对环境污染物的危害识别效率[15,16].

1 典型计算毒理方法

目前，在ECs-生物大分子相互作用的构效机制研究以及虚拟筛选应用中存在多种计算毒理方法，包括分子对接(molecular docking)、分子动力学(molecular dynamics, MD)模拟、机器学习建模(machine learning-based modelling)、密度泛函理论(density functional theory, DFT)[17]、基于生理的药代动力学(physiologically based pharmacokinetic, PBPK)[18]、量子力学/分子力学(quantum mechanics/molecular mechanics, QM/MM)[19]等. 由于计算原理存在差异，本研究选取了3种典型的基于分子模拟和机器学习(machine learning, ML)的计算毒理方法进行详细介绍.

1.1 分子对接

分子对接是一种通过探究生物大分子和小分子之间的相互作用，进而预测其结合模式(binding position)与结合亲和力(binding affinity)的成熟分子模拟方法(图1)[20]. 分子对接能基于生物大分子与小分子配体的结构特征而给出多样化的分子结合模式信息，并利用这种结合信息给出配体与生物大分子相互作用的机制信息[21]. 目前，分子对接由于高通量、模拟配体-受体静态结合过程的优点，已在环境新污染物的虚拟筛查研

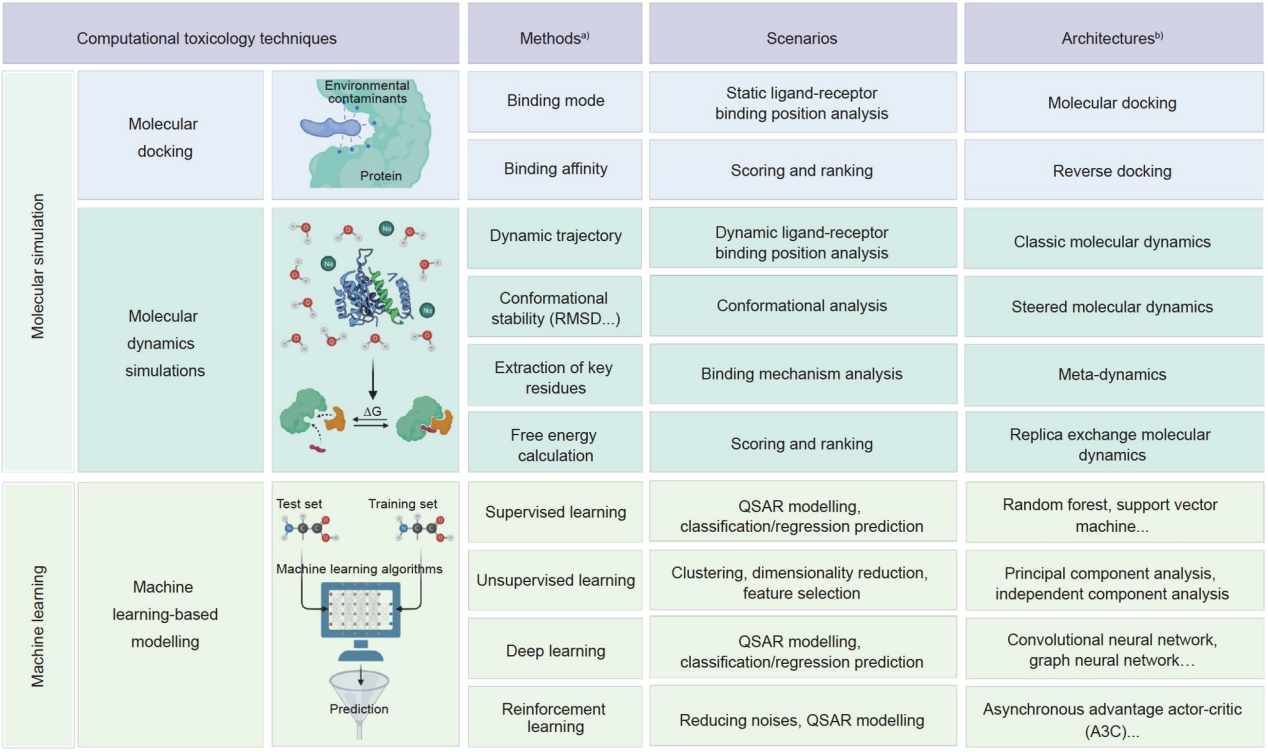

图 1 (网络版彩色)研究环境污染物-生物大分子相互作用过程的3种代表性计算毒理方法. a) 机器学习建模中将传统机器学习算法(监督性、非监督性算法)与深度学习算法、强化学习算法进行了区分; b) 现有构架与方法、使用场景无一一对应关系

Figure 1 (Color online) Three representative computational toxicology techniques for studying environmental contaminant-biomolecule interaction processes. a) Machine learning (ML)-based modelling distinguish traditional ML algorithms (supervised and unsupervised algorithms) from deep learning (DL) algorithms and reinforcement learning (RF) algorithms; b) there is no one-to-one correspondence between methods and scenarios and architectures

究中发挥了重要的作用. 同时, 分子对接也存在着受体刚性、采用近似评分函数、采样不足等技术问题[22~24], 导致其在ECs-生物大分子相互作用的构效机制研究方面存在较为“疲弱”的能力.

1.2 分子动力学模拟

与分子对接的静态结合不同的是, 分子动力学(MD)可以模拟ECs-生物大分子的柔性结合过程, 以及复合物(ligand-receptor complex)的动态变构过程, 提供更为全面的相互作用机制信息[25,26](图1). 因此, MD模拟代表了一种识别并评估ECs-生物大分子复合物结构稳定性的有效分子模拟工具. 虽然MD模拟能尽可能地探究配体-受体结合的动态过程, 为ECs提供致毒机制信息, 但仍然存在两个方面的问题. (1) 低通量. 当分子模拟技术逐渐深入探究化学品与蛋白在分子水平的动态交互过程时, 伴随着机制信息挖掘程度提升的是逐渐降低的通量和逐渐升高的时间消耗. 如何平衡深入挖掘致毒机制和维持较高通量是目前MD模拟遇到的关键挑战之一. (2) 基于通路的毒性预测. 正如前文所述, MD针对某一关键事件(主要是ECs与生物大分子的配体-受体结合过程)进行模拟和研究. 在AOP理论框架下, ECs是通过影响生物大分子调控的通路活性进而导致个体水平的毒害效应. 然而, MD无法基于通路模拟预测ECs的毒害效应.

1.3 机器学习建模

最近, 机器学习(ML)算法以及基于机器学习发展的深度学习(DL)算法在ECs-生物大分子的互作探究中获得了一席之地[27,28]. 机器学习建模(图1)主要利用公开的结构信息和化学(*in chemico*)、体外、体内生物活性数据, 在ML算法的基础上构建定量结构-活性关系(quantitative structure-activity relationship, QSAR)模型, 并利用QSAR模型迅速提高靶向分子靶点ECs的虚拟筛查效率, 且进一步加深复杂生物背景下污染物-生物大分子的作用机制解析[29~31]. 因此, 机器学习建模是一种与分子模拟完全不同的新型计算毒理技术. 然而, 机器学习建模存在着一个关键问题——可解释性较差[30,32,33]. 以往对线性数据具有优良预测能力的统计学方法, 如基于线性的多元线性回归和偏最小二乘回归, 已无法应对爆炸式发展的生物“大数据”. 巨大的数据量和复杂的生物学过程导致更为先进的非线性类数据分析方法——ML算法, 应用于QSAR模型. 高度非线性的决策边界和超参数依赖预测问题使得机器学习建模同时有着高预测精度和低可解释性的特性. 例如, ML/DL无法解释网络中每个节点的生物学意义, 以及每个特征对模型预测性能的重要性占比. 然而, 将非线性的“大数据”转化成人类可理解的线性结论并量化结论的不确定性程度是重中之重[32]. 因为, 在环境毒理学领域, 人们迫切想了解为什么一个模型能够有很好的预测结果. 目前, 在ECs-生物大分子互作研究中, 研究人员试图应用许多, 如基于树的随机森林特征重要性(针对随机森林算法)和特征交互网络分析框架算法(针对深度神经网络算法), 去弥补机器学习预测模型的“黑箱”缺陷[34~38]. 但是, 有限的解释性仍局限了机器学习建模在ECs-生物大分子的互作研究.

下面将系统综述3种代表性的计算毒理方法在环境毒理学领域的主要用途, 包括ECs-生物大分子相互作用的构效关系研究与虚拟筛查应用. 举例讨论了3类计算毒理方法在配体-受体结合、可解释性、计算效率、计算深度以及生物学过程5个方面的优势与局限(图2).

2 计算毒理在污染物-生物大分子互作过程的构效关系研究与虚拟筛选应用

2.1 配体-受体结合

配体-受体结合过程作为污染物-生物大分子相互作用的关键第一步一直是计算毒理学的研究核心[39]. 目前, 针对配体-受体结合进行互作机制探究主要依赖于分子模拟技术. 分子对接利用结合模式与结合亲和力来量化ECs-生物大分子的静态结合过程, 进而为ECs-生物大分子的构效关系研究提供了初步线索(图2). 例如, 2022年, Liang等人[40]利用分子对接研究了对羟基苯甲酸酯(parabens)与甲状腺激素受体(thyroid hormone receptor α/β, TRα/β)的结合过程. 他们发现所有的parabens都以正确的结合模式进入TRα/β的结合口袋, 且结合亲和力范围在–6.36~–6.01 kcal/mol, 证明parabens-TRα/β存在明显的结合活性. 2022年, Li等人[41]利用分子对接研究了有机磷酸酯(organophosphate esters, OPEs)与膜甲状腺激素受体(membrane TR, mTR)和TRβ的结合过程, 发现目标OPEs的结合位点与参考配体一致; 并依据结合亲和力将OPEs进行毒性效力排序, 其模拟结果与体外实验结果显示出强烈的正相关关系(R^2=0.94). 2022年, Zhang等人[42]利用分子对接探

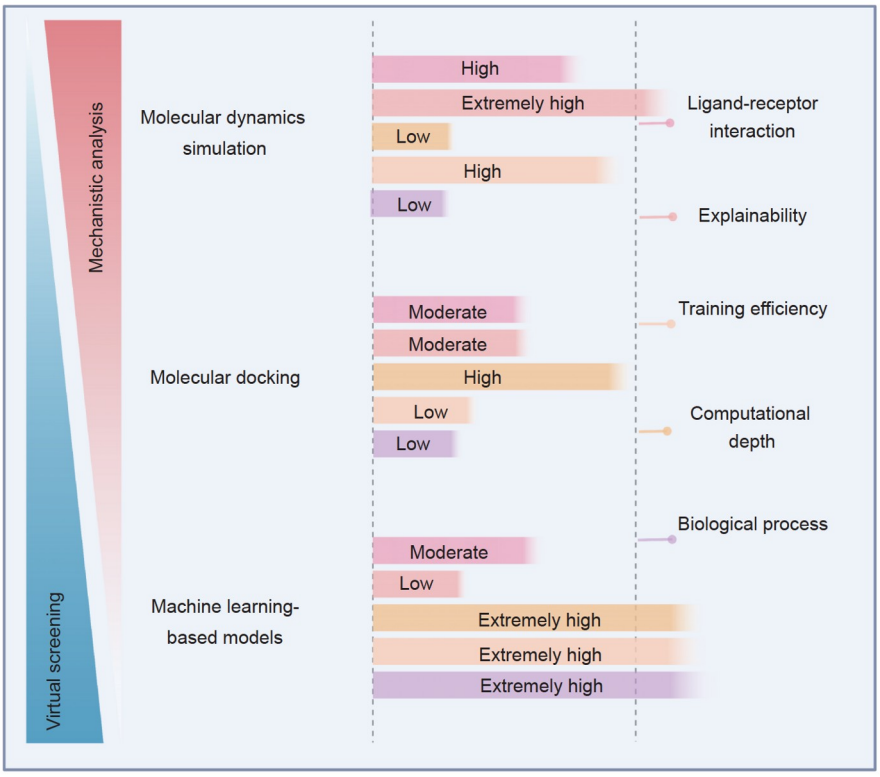

图 2 (网络版彩色)环境污染物-生物大分子相互作用研究中典型计算毒理方法的优势和缺陷
Figure 2 (Color online) Strengths and weaknesses of typical computational toxicology techniques in environmental contaminant-biomolecule interaction studies

究了二硝基呋喃对映体对烟碱乙酰胆碱受体的结合作用, 更高的结合亲和力与更多的氢键作用强调了S对映体比R对映体对受体有更强的结合活性. 目前, 已发表了大量的基于分子对接的配受体结合研究[43~47].

MD通过模拟ECs-生物大分子结合的动态轨迹(dynamic trajectory), 得出ECs与生物大分子结合后形成的复合物在时间轴上的波动(fluctuation)与构象变化(conformational shifts)信息(图2)[48,49]. 例如, 2013年, Wang等人[50]研究多溴联苯醚(polybrominated diphenyl ethers, PBDEs)的代谢产物——羟基化多溴联苯醚(hydroxylated PBDEs, HO-PBDEs)和甲氧基化多溴联苯醚(methoxylated PBDEs, MeO-PBDEs)对TRα/β存在的潜在干扰影响. 通过模拟HO-/MeO-PBDEs与TRα/β的动态结合过程及受体第十二号α螺旋(helix 12, H12)的动态重定位过程, 结果发现污染物通过模仿内源性甲状腺激素(T3)竞争结合到TRα/β结合腔内, 诱导TRα/β的构象变化, 致使H12重新定位, 从而产生激活表面(AF-2), 导致内分泌干扰活性. 接着, 2016年, Chen等人[51]在此基础上模拟发现, HO-PBDEs产生的AF-2活性表面促进了TRs招募共激活因子, 致使产生拟性的甲状腺内分泌干扰活性. 2018年, Lu等人[52]也利用MD模拟发现双酚S及其同系物通过与关键氨基酸His435(H11)和Phe459(H12)作用, 进而诱导TRβ的H12重定位, 产生干扰活性. MD模拟已在研究环境内分泌干扰物(endocrine-disrupting chemicals, EDCs)与核受体(nuclear receptors, NRs)之间的动力学结合机制方面发挥了巨大作用[53~58].

2.2 可解释性

机器学习建模主要在ML/DL算法的基础上构建QSAR模型, 并基于QSAR模型的可解释性对ECs-生物大分子相互作用的结合机制进行探究(图2). 然而, 与具有高度可解释性的分子模拟不同的是, 作为后验性的机器学习建模具有“黑箱”缺陷, 导致只能对ECs的致毒机制进行有限的解释. 例如, 2019年, Cheng和Ng[59]利用26类活性数据结合5种监督性的ML/DL算法(逻辑回归、随机森林、人工神经网络和两种图神经网络), 首先对1012种全氟和多氟烷基化合物(per- and poly-

fluoroalkyl substances, PFASs)进行QSAR建模. 然后, 基于高精确度的预测模型发现大多数具有生物活性PFASs的全氟烷基链长小于12. 2021年, Li等人[60]利用多元线性回归算法发现拓扑极性表面积会部分影响ECs的胎盘转移效率. 2021年, Wang等人[61]利用深度学习框架对靶向雌激素受体(estrogen receptors, ERs)化合物的3D结构进行定量描述, 构建的QSAR模型也进一步发现电荷分布模式可能是化合物产生雌激素干扰活性的关键结构因子. 目前, 已有大量研究利用ML/DL算法对ECs-生物大分子的部分结合机理进行了有效解释[62,63]. 然而, 正如前文所述, 机器学习建模的关键问题就是ML/DL算法的可解释性较差[30,32,33]. 有限的解释性局限了该技术在ECs-生物大分子互作机制研究的应用.

2.3 计算效率

计算毒理技术的计算效率直接决定了该方法在污染物-生物大分子互作研究中的通量, 效率高则通量高, 效率低则通量低. 分子对接由于耗时短、效率高, 该技术在ECs-生物大分子相互作用的高通量机制研究中大放异彩. 例如, 2020年, Tan等人[64]将4000多种化合物高通量对接到雄激素受体(androgen receptor, AR)和ERα的结合口袋, 发现活性与非活性化合物具有完全不同的结合模式. 2021年, Tan等人[65]进一步探索了8000多种化合物与人类12个经典NRs的差异性结合模式, 发现了决定EDCs的关键结合位点. 同时, 由于分子对接的低成本、高通量“潜能”, 促使了反向对接(reverse docking)的兴起(图3). 反向对接是一种基于分子对接的反向虚拟筛选技术, 它可以在ECs活性未知的前提下, 预测其潜在的生物大分子靶点[66]. 因此, 反向对接代表了一种新型的“捕捞”活性靶点的高通量虚拟筛选手段[67]. 例如, 2014年, Kolšek等人[68]开发了一款预测14种NRs介导EDCs的网页预测软件Endocrine Disruptome. Endocrine Disruptome通过模拟配体-受体的结合亲和力大小来预测ECs是否是潜在的14种NRs介导EDCs[68]. 2015年, Vedani等人[69]利用柔性对接开发了一款预测化合物与10种NRs的结合亲和力预测软件VirtualToxLab™. 2017年, Wang等人[70]更是针对目前已有晶体结构的39种NRs构建了基于分子对接的EDCs预测模型. 该研究迅速扩大了基于分子对接的虚拟筛查应用范围[70].

由于耗时耗力, 低通量的MD模拟只能针对典型的ECs(小型数据集)开展研究工作. 例如, 2016年, Zhang等人[71]首先基于经典MD模拟探究了ECs对转甲状腺素蛋白的结合干扰机制, 并基于干扰机制开发了一个新型虚拟筛查流程, 以便更好、更快地筛查出潜在的甲状腺干扰物.

机器学习建模包括传统的ML算法与新兴的DL算法(图1). 一般而言, 传统的ML算法所需数据量适中、计算时间短, 促使基于ML的QSAR建模计算效率高, 适用范围广. 而新兴的DL对超参数的确定具有高度人为依赖性. 因此, 基于DL的QSAR建模所需数据更多、调参时间更长、计算效率更低(图2). 目前, 基于传统ML算法的QSAR建模在配体-受体结合过程已有诸多应用案例. 2016年, Bhhatarai等人[72]利用化合物与ERs/AR的结合数据信息构建了基于ML的QSAR模型, 实现了EDCs的虚拟筛选. 2021年, de Lomana等人[73]首先收集了与甲状腺激素平衡相关的8个关键分子靶点, 并针对每一个靶点构建了基于ML的定性模型, 实现了ECs是否可能通过干扰某一个分子靶点而产生潜在甲状腺激素紊乱的综合虚拟筛选.

2.4 计算深度

ECs-生物大分子互作研究发现, 与针对靶向生物大分子而设计的药物相比, ECs结构各异、结合亲和力普遍较低. 因此, 由于低计算深度(近似评分函数、静态结合预测)导致的低精确结合能预测限制了分子对接在ECs-生物大分子的互作研究[74,75]. 2014年, Da和Kireev[76]为了提高分子对接中配体-受体结合亲和力的评估能力, 开发了一种基于三维结构蛋白-配体相互作用指纹图谱软件SPLIF. SPLIF不仅可以定量评估配体是否与分子靶点相互作用, 还极大地减少了因初始对接得分较低而被忽略的假阴性EDCs[76]. 同时, 研究人员也开发了许多新型的结合能预测方法, 对接以达到提升结合亲和力预测能力的目的. 例如, 柔性(flexible docking)/半柔性对接(semi-flexible docking)[77]、自由能扰动(free energy perturbation)[78]、集成热力学(thermodynamic integration)[79]和漏斗元动力学(funnel metadynamics)[80,81]. 值得注意的是, 尽管这些方法的精确度比近似评分函数更高, 但是提升了计算深度, 降低了计算效率与计算通量.

经典MD只能模拟配体-受体的动态结合过程. 许多研究发现, 配体与受体之间的相互作用不仅包括结合过程, 也包括配体从受体中的逃逸过程, 即结合作用和逃逸效应共同代表了配体在受体中的效力[82~84]. 因

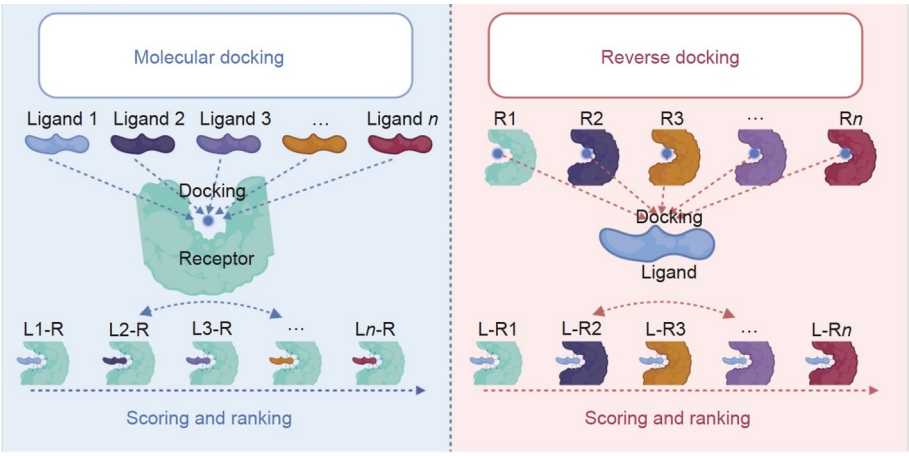

图 3 (网络版彩色)分子对接与反向对接的区别
Figure 3 (Color online) Difference between molecular docking and reverse docking

此, 2018年, Chen等人[85]结合经典MD和拉伸分子动力学(steered molecular dynamics, SMD)同时模拟BPA类化合物与AR的结合与逃逸过程, 进而精确预测结合亲和力. 同年, Hu等人[86]采用SMD发现蜕皮激素受体(ecdysone receptor, EcR)存在4种不同的逃逸路径, 决定了农药与EcR结合的亲和力大小. 此外, ECs往往会同时产生拟性、抗性的内分泌干扰效应, 这种混合效应是由于配体-受体相互作用产生了多种稳态构象[87~89]. 与经典MD模拟相比, 更先进的增强采样技术, 如伞状采样(umbrella sampling)[90]、元动力学(meta-dynamics)[91]和复制交换分子动力学(replica exchange molecular dynamics)[92], 可以用来识别配体-受体复合体的多种构象. 例如, 2020年, Chen等人[93]就采用meta-dynamics发现了BPA类污染物对ERs与糖皮质激素受体的混合效应. 因此, 随着ECs致毒机制的解析程度不断加深, MD模拟技术的计算深度也在不断提升.

2.5 生物学过程

机器学习建模的关键优势在于它可以利用AOP理论框架模拟复杂的生物学过程对污染物的毒性进行高通量预测. 因此, 与针对配体-受体竞争结合过程(MIE)开展研究的分子模拟不同的是, 机器学习建模可以基于AOP理论框架从分子、细胞到个体的整个生物反应过程建立高通量虚拟筛查模型. 例如, 美国环境保护署针对ERs和AR的AOPs, 利用ML算法分别构建了基于通路的EDCs虚拟筛查模型(ToxCast AR pathway模型[94]、ToxCast ER pathway模型[95,96]、CoMPARA模型[97]、CERAPP模型[98]). 2021年, Ciallella等人[99]利用深度神经网络成功开发出了基于通路的ERs介导EDCs高通量虚拟筛查模型. 2021年, Borba等人[100]基于皮肤致敏AOP, 利用深度神经网络结合朴素贝叶斯算法构建了皮肤致敏虚拟筛查模型Pred-skin.

2.6 综合利用多种计算毒理工具的优势

由于ECs-生物大分子互作研究的复杂性, 综合利用多种计算毒理工具进行污染物致毒机制探究与筛选潜在毒物已成趋势. 例如, 同时结合分子对接与MD模拟可以弥补对接/模拟技术的局限性. 首先, 基于对接的工作流中加入MD模拟可以提供更多的机制信息[101~103]. 例如, 2015年, Hirano等人[104]利用对接发现二噁英化合物(dioxin-like compounds, DLCs)对芳烃受体(aryl hydrocarbon receptor, AhR)的结合能随取代的氯原子数量增加而增强; 进一步, MD模拟发现了DLCs与AhR结合的关键氨基酸(Ile324和Ser380). 2018年, Cheng和Ng[105]结合分子对接与经典MD模拟估计15种PFASs对人与大鼠肝型脂肪酸结合蛋白的结合亲和力, 发现长链PFASs仍然具有潜在的生物累积性. 此外, 在MD模拟中加入分子对接的综合工作流程不仅可以维持MD模拟本身的机制信息, 还可以提高工作流程的通量, 使得其能有效运用于ECs的高通量虚拟筛查. 例如, 2022年, Montes-Grajales等人[106]将50多种BPA类似物和衍生物高通量对接到328个与非传染性疾病相关的分子靶点, 以探究BPA类污染物对人类健康的潜在影响; 进一步利用经典MD模拟探究高结合亲和力的化合物-分子靶点复合物, 筛查出了需要进一步进行实验测试的污染物. 目前, 分子对接与MD模拟结合的方法已

在六氟环氧丙烷[107]、二甲双胍[108]、BPA污染物[109]、三氯乙烯[110]等ECs中开展了大量的研究. 此外, 机器学习建模也可以有效解决分子对接的低预测精度问题. 例如, 2019年, Nogueira和Koch[111]结合分子对接与两种ML算法提升了反向对接的评分精确性.

同时结合3种手段可以多角度、更深入地探索ECs与生物大分子的相互作用机制. 例如, 2020年, Li等人[112]首先基于分子对接的盲对接模式发现PFASs与ERα表面的共调节因子招募区域(别构位点)存在结合作用; 再基于经典MD模拟发现PFAS-ERα的别构结合力极高、别构构象稳定存在; 最后, 基于QSAR建模进一步发现分子体积以及溶质可接触的表面积决定了PFASs对ERα的结合能. 2021年, Zhang等人[113]首先基于对接与经典MD模拟探究了DLCs对鸟类AhR1的结合作用, 并用MD模拟计算出了每个污染物与AhR1氨基酸的分子力学能(molecular mechanics energy, E_{MM}); 进一步结合模拟结果和ML算法构建了机制驱动的广义线性模型, 发现了不同鸟种对DLCs的相对敏感度.

3 展望

计算毒理技术自开发以来主要协助于药物的设计和发现. 然而, 除了最初针对药物的开发与应用外, 计算毒理技术现在也被广泛用于环境新污染物的识别, 包括本文所述的污染物-生物大分子相互作用机制探究和环境新污染物的虚拟筛查应用. 计算毒理技术的蓬勃发展主要存在两方面的原因. (1) 高性能计算领域的进步[13,14]. 特别地, 图形处理单元(graphics processing unit, GPU)的进步大大提升了计算毒理的计算速度[114]. (2) 生物大分子三维构象的获取. 生物大分子的未知结构局限了分子模拟在环境毒理学领域的应用. 因此, 获得生物大分子的有效三维结构是分子模拟的关键. 同源模建(homology modeling)是通过利用具有高度同源序列的结构模板来建立同源模型的模拟技术, 它可以弥补无法获得生物大分子三维结构的问题[115,116]. 同时, 结构生物学[9,10]和基于深度学习的晶体结构预测[11,12]的爆发式进展也大量增加了未知生物大分子的三维结构特征.

计算毒理技术总是在环境污染物与生物大分子相互作用的机制解析和虚拟筛查两方面进行权衡(图2). 能深入探究污染物-生物大分子相互作用的MD模拟无法实现ECs的高通量虚拟筛选; 能反应复杂生物学过程并实现高精度预测的机器学习预测模型却由于其“黑箱”缺陷无法对预测结果进行有效解释; 具有高通量“潜能”的分子对接却由于局部采样、近似评分函数缺陷局限了该技术在污染物-生物大分子相互作用过程更为全面的机制解析能力. 正如本文所述, 结合多种计算毒理技术开发综合的工作流进行机制解析和虚拟筛选可以弥补各自短处, 获得最优结果. 但是如何在保持机制解析的基础上增加计算通量筛查仍然是未来亟须解决的关键问题.

参考文献

1 Axelstad M, Hass U, Scholze M, et al. EDC IMPACT: Reduced sperm counts in rats exposed to human relevant mixtures of endocrine disrupters. Endocr Connect, 2018, 7: 139–148

2 Ford K A. Refinement, reduction, and replacement of animal toxicity tests by computational methods. ILAR J, 2016, 57: 226–233

3 Wang J Y, Wang Z Y, Chen J W, et al. Environmental systems engineering consideration on treatment of emerging pollutants and risk prevention and control of chemicals (in Chinese). Chin Sci Bull, 2022, 67: 267–277 [王佳钰, 王中钰, 陈景文, 等. 环境新污染物治理与化学品环境风险防控的系统工程. 科学通报, 2022, 67: 267–277]

4 Ezendam J, Braakhuis H M, Vandebriel R J. State of the art in non-animal approaches for skin sensitization testing: From individual test methods towards testing strategies. Arch Toxicol, 2016, 90: 2861–2883

5 Ankley G T, Bennett R S, Erickson R J, et al. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem, 2010, 29: 730–741

6 Tian M M, Peng Y, Zhang R, et al. Research advance of quantitative adverse outcome pathway (qAOP) in environmental chemicals toxicity assessment Ⅱ: Dioxin-like compounds and AhR-qAOP (in Chinese). Asian J Ecotoxicol, 2021, 16: 1–16 [田明明, 彭颖, 张睿, 等. 定量有害结局路径(qAOP)评估环境化学物质毒性的研究进展Ⅱ: 类二噁英物质及AhR-qAOP. 生态毒理学报, 2021, 16: 1–16]

7 Richard A M, Huang R, Waidyanatha S, et al. The Tox21 10K compound library: Collaborative chemistry advancing toxicology. Chem Res Toxicol, 2021, 34: 189–216

8 Huang R, Xia M, Sakamuru S, et al. Modelling the Tox21 10 K chemical profiles for *in vivo* toxicity prediction and mechanism characterization. Nat Commun, 2016, 7: 10425

9 Gavira J A. Current trends in protein crystallization. Arch Biochem Biophys, 2016, 602: 3–11
10 Grimes J M, Hall D R, Ashton A W, et al. Where is crystallography going? Acta Crystlogr D Struct Biol, 2018, 74: 152–166
11 Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature, 2021, 596: 583–589
12 Gao M, Nakajima An D, Parks J M, et al. AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. Nat Commun, 2022, 13: 1744
13 Song C M, Lim S J, Tong J C. Recent advances in computer-aided drug design. Brief Bioinform, 2009, 10: 579–591
14 Jorgensen W L. The many roles of computation in drug discovery. Science, 2004, 303: 1813–1818
15 Chen Q C, Tan H Y, Shi W, et al. Application and prospect of computational toxicology in screening of endocrine disrupting chemicals (in Chinese). Asian J Ecotoxicol, 2017, 12: 38–48 [陈钦畅, 谭皓月, 史薇, 等. 计算毒理学在内分泌干扰物筛选上的应用和展望. 生态毒理学报, 2017, 12: 38–48]
16 Wang Z Y, Chen J W, Fu Z Q, et al. Characterization of applicability domains for QSAR models (in Chinese). Chin Sci Bull, 2022, 67: 255–266 [王中钰, 陈景文, 傅志强, 等. QSAR模型应用域的表征方法. 科学通报, 2022, 67: 255–266]
17 Townsend P A, Grayson M N. Density functional theory in the prediction of mutagenicity: A perspective. Chem Res Toxicol, 2021, 34: 179–188
18 Kamiya Y, Handa K, Miura T, et al. An updated *in silico* prediction method for volumes of systemic circulation of 323 disparate chemicals for use in physiologically based pharmacokinetic models to estimate plasma and tissue concentrations after oral doses in rats. Chem Res Toxicol, 2021, 34: 2180–2183
19 Li Y, Shi X, Zhang Q, et al. Computational evidence for the detoxifying mechanism of epsilon class glutathione transferase toward the insecticide DDT. Environ Sci Technol, 2014, 48: 5008–5016
20 Kitchen D B, Decornez H, Furr J R, et al. Docking and scoring in virtual screening for drug discovery: Methods and applications. Nat Rev Drug Discov, 2004, 3: 935–949
21 Kortagere S, Krasowski M D, Reschly E J, et al. Evaluation of computational docking to identify pregnane x receptor agonists in the ToxCast database. Environ Health Perspect, 2010, 118: 1412–1417
22 Elokely K M, Doerksen R J. Docking challenge: Protein sampling and molecular docking performance. J Chem Inf Model, 2013, 53: 1934–1945
23 Guedes I A, Pereira F S S, Dardenne L E. Empirical scoring functions for structure-based virtual screening: Applications, critical aspects, and challenges. Front Pharmacol, 2018, 9: 1089
24 Pantsar T, Poso A. Binding affinity via docking: Fact and fiction. Molecules, 2018, 23: 1899
25 Kumari I, Sandhu P, Ahmed M, et al. Molecular dynamics simulations, challenges and opportunities: A biologist's prospective. Curr Protein Pept Sci, 2017, 18: 1163–1179
26 Hollingsworth S A, Dror R O. Molecular dynamics simulation for all. Neuron, 2018, 99: 1129–1143
27 Chen J W, Li X H, Yu H Y, et al. (Q)SAR technology oriented for ecological risk assessment of hazardous organics: Advances and perspetives (in Chinese). Sci Sin Chim, 2008, 38: 461–474 [陈景文, 李雪花, 于海瀛, 等. 面向毒害有机物生态风险评价的(Q)SAR技术: 进展与展望. 中国科学: 化学, 2008, 38: 461–474]
28 Wang Z Y, Chen J W, Qiao X L, et al. Computational toxicology: Oriented for chemicals risk assessment (in Chinese). Sci Sin Chim, 2016, 46: 222–240 [王中钰, 陈景文, 乔显亮, 等. 面向化学品风险评价的计算(预测)毒理学. 中国科学: 化学, 2016, 46: 222–240]
29 Zhong S, Zhang K, Bagheri M, et al. Machine learning: New ideas and tools in environmental science and engineering. Environ Sci Technol, 2021, 55: 12741–12754
30 Liu X, Lu D, Zhang A, et al. Data-driven machine learning in environmental pollution: Gains and problems. Environ Sci Technol, 2022, 56: 2124–2133
31 Wang M W H, Goodman J M, Allen T E H. Machine learning in predictive toxicology: Recent applications and future directions for classification models. Chem Res Toxicol, 2021, 34: 217–239
32 Sapoval N, Aghazadeh A, Nute M G, et al. Current progress and open challenges for applying deep learning across the biosciences. Nat Commun, 2022, 13: 1728
33 Wu L, Huang R, Tetko I V, et al. Trade-off predictivity and explainability for machine-learning powered predictive toxicology: An in-depth investigation with Tox21 data sets. Chem Res Toxicol, 2021, 34: 541–549
34 Yu F, Wei C, Deng P, et al. Deep exploration of random forest model boosts the interpretability of machine learning studies of complicated immune responses and lung burden of nanoparticles. Sci Adv, 2021, 7: eabf4130
35 Montavon G, Samek W, Müller K R. Methods for interpreting and understanding deep neural networks. Digital Signal Process, 2018, 73: 1–15
36 Ying R, Bourgeois D, You J, et al. GNNExplainer: Generating explanations for graph neural networks. Adv Neural Inf Process Syst, 2019, 32: 9240–9251
37 Li T, Tong W, Roberts R, et al. DeepDILI: Deep learning-powered drug-induced liver injury prediction using model-level representation. Chem Res Toxicol, 2021, 34: 550–565

38 Hsieh J H, Sedykh A, Mutlu E, et al. Harnessing *in silico*, *in vitro*, and *in vivo* data to understand the toxicity landscape of polycyclic aromatic compounds (PACs). Chem Res Toxicol, 2021, 34: 268–285

39 Huang F Y, Chen Q C, Tan H Y, et al. Review on the effects of endocrine disrupting chemicals on dimerization of nuclear receptors (in Chinese). Asian J Ecotoxicol, 2021, 16: 17–31 [黄付晏, 陈钦畅, 谭皓月, 等. 内分泌干扰物对核受体二聚化影响的研究进展. 生态毒理学报, 2021, 16: 17–31]

40 Liang J, Yang X, Liu Q S, et al. Assessment of thyroid endocrine disruption effects of parabens using *in vivo*, *in vitro*, and *in silico* approaches. Environ Sci Technol, 2022, 56: 460–469

41 Li J, Xu Y, Li N, et al. Thyroid hormone disruption by organophosphate esters is mediated by nuclear/membrane thyroid hormone receptors: *In vitro*, *in vivo*, and *in silico* studies. Environ Sci Technol, 2022, 56: 4241–4250

42 Zhang Q, Fu L, Cang T, et al. Toxicological effect and molecular mechanism of the chiral neonicotinoid dinotefuran in honeybees. Environ Sci Technol, 2022, 56: 1104–1112

43 Li S, Zhao J, Huang R, et al. Profiling the Tox21 chemical collection for acetylcholinesterase inhibition. Environ Health Perspect, 2021, 129: 047008

44 Ma D, Lu Y, Liang Y, et al. A critical review on transplacental transfer of per- and polyfluoroalkyl substances: Prenatal exposure levels, characteristics, and mechanisms. Environ Sci Technol, 2022, 56: 6014–6026

45 Zhao L, Chen F, Yin S, et al. Enantioselectivity in transplacental transfer of perfluoro-1-methylheptanesulfonate (1m-PFOS): Human biomonitoring and *in silico* study. Environ Pollut, 2020, 261: 114136

46 Li Y, Yu N, Du L, et al. Transplacental transfer of per- and polyfluoroalkyl substances identified in paired maternal and cord sera using suspect and nontarget screening. Environ Sci Technol, 2020, 54: 3407–3416

47 Ng C A, Hungerbuehler K. Exploring the use of molecular docking to identify bioaccumulative perfluorinated alkyl acids (PFAAs). Environ Sci Technol, 2015, 49: 12306–12314

48 Milardi D, Pappalardo M. Molecular dynamics: New advances in drug discovery. Eur J Med Chem, 2015, 91: 1–3

49 Makarewicz T, Kaźmierkiewicz R. Molecular dynamics simulation by GROMACS using GUI plugin for PyMOL. J Chem Inf Model, 2013, 53: 1229–1234

50 Wang X, Yang H, Hu X, et al. Effects of HO-/MeO-PBDEs on androgen receptor: *In vitro* investigation and helix 12-involved MD simulation. Environ Sci Technol, 2013, 47: 11802–11809

51 Chen Q, Wang X, Shi W, et al. Identification of thyroid hormone disruptors among HO-PBDEs: *In vitro* investigations and coregulator involved simulations. Environ Sci Technol, 2016, 50: 12429–12438

52 Lu L, Zhan T, Ma M, et al. Thyroid disruption by bisphenol S analogues via thyroid hormone receptor β: *In vitro*, *in vivo*, and molecular dynamics simulation study. Environ Sci Technol, 2018, 52: 6617–6625

53 Lai T T, Eken Y, Wilson A K. Binding of per- and polyfluoroalkyl substances to the human pregnane X receptor. Environ Sci Technol, 2020, 54: 15986–15995

54 He S, Xiao H, Luo S, et al. Benzotriazole ultraviolet stabilizers promote breast cancer cell proliferation via activating estrogen-related receptors α and γ at human-relevant levels. Environ Sci Technol, 2022, 56: 2466–2475

55 Zhang R, Wang X, Zhang X, et al. Polychlorinated diphenylsulfides activate aryl hydrocarbon receptor 2 in zebrafish embryos: Potential mechanism of developmental toxicity. Environ Sci Technol, 2018, 52: 4402–4412

56 Lu L, Wu H, Cui S, et al. Pentabromoethylbenzene exposure induces transcriptome aberration and thyroid dysfunction: *In vitro, in silico*, and *in vivo* investigations. Environ Sci Technol, 2020, 54: 12335–12344

57 Cui S, Yu Y, Zhan T, et al. Carcinogenic risk of 2,6-di-*tert*-butylphenol and its quinone metabolite 2,6-DTBQ through their interruption of RARβ: *In vivo*, *in vitro*, and *in silico* investigations. Environ Sci Technol, 2022, 56: 480–490

58 Li Y, Perera L, Coons L A, et al. Differential *in vitro* biological action, coregulator interactions, and molecular dynamic analysis of bisphenol A (BPA), BPAF, and BPS ligand-ERα complexes. Environ Health Perspect, 2018, 126: 017012

59 Cheng W, Ng C A. Using machine learning to classify bioactivity for 3486 per- and polyfluoroalkyl substances (PFASs) from the OECD list. Environ Sci Technol, 2019, 53: 13970–13980

60 Li J, Sun X, Xu J, et al. Transplacental transfer of environmental chemicals: Roles of molecular descriptors and placental transporters. Environ Sci Technol, 2021, 55: 519–528

61 Wang L, Zhao L, Liu X, et al. SepPCNET: Deeping learning on a 3D surface electrostatic potential point cloud for enhanced toxicity classification and its application to suspected environmental estrogens. Environ Sci Technol, 2021, 55: 9958–9967

62 Pandit S, Singh P, Sinha M, et al. Integrated QSAR and adverse outcome pathway analysis of chemicals released on 3D printing using acrylonitrile butadiene styrene. Chem Res Toxicol, 2021, 34: 355–364

63 Borba J V B, Alves V M, Braga R C, et al. STopTox: An *in silico* alternative to animal testing for acute systemic and topical toxicity. Environ

Health Perspect, 2022, 130: 027012
64 Tan H, Wang X, Hong H, et al. Structures of endocrine-disrupting chemicals determine binding to and activation of the estrogen receptor α and androgen receptor. Environ Sci Technol, 2020, 54: 11424–11433
65 Tan H, Chen Q, Hong H, et al. Structures of endocrine-disrupting chemicals correlate with the activation of 12 classic nuclear receptors. Environ Sci Technol, 2021, 55: 16552–16562
66 Xu X, Huang M, Zou X. Docking-based inverse virtual screening: Methods, applications, and challenges. Biophys Rep, 2018, 4: 1–16
67 Gloriam D E. Bigger is better in virtual drug screens. Nature, 2019, 566: 193–194
68 Kolšek K, Mavri J, Sollner Dolenc M, et al. Endocrine disruptome—An open source prediction tool for assessing endocrine disruption potential through nuclear receptor binding. J Chem Inf Model, 2014, 54: 1254–1267
69 Vedani A, Dobler M, Hu Z, et al. OpenVirtualToxLab—A platform for generating and exchanging *in silico* toxicity data. Toxicol Lett, 2015, 232: 519–532
70 Wang X, Zhang X, Xia P, et al. A high-throughput, computational system to predict if environmental contaminants can bind to human nuclear receptors. Sci Total Environ, 2017, 576: 609–616
71 Zhang J, Begum A, Brännström K, et al. Structure-based virtual screening protocol for *in silico* identification of potential thyroid disrupting chemicals targeting transthyretin. Environ Sci Technol, 2016, 50: 11984–11993
72 Bhhatarai B, Wilson D M, Price P S, et al. Evaluation of OASIS QSAR models using ToxCast™ *in vitro* estrogen and androgen receptor binding data and application in an integrated endocrine screening approach. Environ Health Perspect, 2016, 124: 1453–1461
73 de Lomana M G, Weber A G, Birk B, et al. *In silico* models to predict the perturbation of molecular initiating events related to thyroid hormone homeostasis. Chem Res Toxicol, 2021, 34: 396–411
74 Rastelli G, Pinzi L. Refinement and rescoring of virtual screening results. Front Chem, 2019, 7: 498
75 Rastelli G, Degliesposti G, Del Rio A, et al. Binding estimation after refinement, a new automated procedure for the refinement and rescoring of docked ligands in virtual screening. Chem Biol Drug Des, 2009, 73: 283–286
76 Da C, Kireev D. Structural protein-ligand interaction fingerprints (SPLIF) for structure-based virtual screening: Method and benchmark study. J Chem Inf Model, 2014, 54: 2555–2561
77 Kuriata A, Gierut A M, Oleniecki T, et al. CABS-flex 2.0: A web server for fast simulations of flexibility of protein structures. Nucleic Acids Res, 2018, 46: 338–343
78 Jorgensen W L, Thomas L L. Perspective on free-energy perturbation calculations for chemical equilibria. J Chem Theor Comput, 2008, 4: 869–876
79 Bhati A P, Wan S, Wright D W, et al. Rapid, accurate, precise, and reliable relative free energy prediction using ensemble based thermodynamic integration. J Chem Theor Comput, 2017, 13: 210–222
80 Limongelli V, Bonomi M, Parrinello M. Funnel metadynamics as accurate binding free-energy method. Proc Natl Acad Sci USA, 2013, 110: 6358–6363
81 Clark A J, Tiwary P, Borrelli K, et al. Prediction of protein-ligand binding poses via a combination of induced fit docking and metadynamics simulations. J Chem Theor Comput, 2016, 12: 2990–2998
82 Tummino P J, Copeland R A. Residence time of receptor-ligand complexes and its effect on biological function. Biochemistry, 2008, 47: 5481–5492
83 Martínez L, Webb P, Polikarpov I, et al. Molecular dynamics simulations of ligand dissociation from thyroid hormone receptors: Evidence of the likeliest escape pathway and its implications for the design of novel ligands. J Med Chem, 2006, 49: 23–26
84 Shen J, Li W, Liu G, et al. Computational insights into the mechanism of ligand unbinding and selectivity of estrogen receptors. J Phys Chem B, 2009, 113: 10436–10444
85 Chen Q, Wang X, Tan H, et al. Molecular initiating events of bisphenols on androgen receptor-mediated pathways provide guidelines for *in silico* screening and design of substitute compounds. Environ Sci Technol Lett, 2019, 6: 205–210
86 Hu X, Hu S, Wang J, et al. Steered molecular dynamics for studying ligand unbinding of ecdysone receptor. J Biomol Struct Dyn, 2018, 36: 3819–3828
87 Lusher S J, Raaijmakers H C A, Vu-Pham D, et al. X-ray structures of progesterone receptor ligand binding domain in its agonist state reveal differing mechanisms for mixed profiles of 11β-substituted steroids. J Biol Chem, 2012, 287: 20333–20343
88 Duan M, Liu N, Zhou W, et al. Structural diversity of ligand-binding androgen receptors revealed by microsecond long molecular dynamics simulations and enhanced sampling. J Chem Theor Comput, 2016, 12: 4611–4619
89 Fratev F. PPARγ helix 12 exhibits an antagonist conformation. Phys Chem Chem Phys, 2016, 18: 9272–9280
90 Torrie G M, Valleau J P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. J Comput Phys, 1977, 23: 187–199

91 Laio A, Parrinello M. Escaping free-energy minima. Proc Natl Acad Sci USA, 2002, 99: 12562–12566
92 Zhou R. Replica exchange molecular dynamics method for protein folding simulation. Methods Mol Biol, 2007, 350: 205–223
93 Chen Q, Zhou C, Shi W, et al. Mechanistic *in silico* modeling of bisphenols to predict estrogen and glucocorticoid disrupting potentials. Sci Total Environ, 2020, 728: 138854
94 Kleinstreuer N C, Ceger P, Watt E D, et al. Development and validation of a computational model for androgen receptor activity. Chem Res Toxicol, 2016, 30: 946–964
95 Browne P, Judson R S, Casey W M, et al. Screening chemicals for estrogen receptor bioactivity using a computational model. Environ Sci Technol, 2015, 49: 8804–8814
96 Judson R S, Magpantay F M, Chickarmane V, et al. Integrated model of chemical perturbations of a biological pathway using 18 *in vitro* high-throughput screening assays for the estrogen receptor. Toxicol Sci, 2015, 148: 137–154
97 Mansouri K, Kleinstreuer N, Abdelaziz A M, et al. CoMPARA: Collaborative modeling project for androgen receptor activity. Environ Health Perspect, 2020, 128: 027002
98 Mansouri K, Abdelaziz A, Rybacka A, et al. CERAPP: Collaborative estrogen receptor activity prediction project. Environ Health Perspect, 2016, 124: 1023–1033
99 Ciallella H L, Russo D P, Aleksunes L M, et al. Revealing adverse outcome pathways from public high-throughput screening data to evaluate new toxicants by a knowledge-based deep neural network approach. Environ Sci Technol, 2021, 55: 10875–10887
100 Borba J V B, Braga R C, Alves V M, et al. Pred-Skin: A web portal for accurate prediction of human skin sensitizers. Chem Res Toxicol, 2021, 34: 258–267
101 Salmaso V, Moro S. Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview. Front Pharmacol, 2018, 9: 923
102 De Vivo M, Masetti M, Bottegoni G, et al. Role of molecular dynamics and related methods in drug discovery. J Med Chem, 2016, 59: 4035–4061
103 Alonso H, Bliznyuk A A, Gready J E. Combining docking and molecular dynamic simulations in drug design. Med Res Rev, 2006, 26: 531–568
104 Hirano M, Hwang J H, Park H J, et al. *In silico* analysis of the interaction of avian aryl hydrocarbon receptors and dioxins to decipher isoform-, ligand-, and species-specific activations. Environ Sci Technol, 2015, 49: 3795–3804
105 Cheng W, Ng C A. Predicting relative protein affinity of novel per- and polyfluoroalkyl substances (PFASs) by an efficient molecular dynamics approach. Environ Sci Technol, 2018, 52: 7972–7980
106 Montes-Grajales D, Morelos-Cortes X, Olivero-Verbel J. Discovery of new protein targets of BPA analogs and derivatives associated with noncommunicable diseases: A virtual high-throughput screening. Environ Health Perspect, 2021, 129: 037009
107 Xin Y, Ren X M, Wan B, et al. Comparative *in vitro* and *in vivo* evaluation of the estrogenic effect of hexafluoropropylene oxide homologues. Environ Sci Technol, 2019, 53: 8371–8380
108 Lin W, Yan Y, Ping S, et al. Metformin-induced epigenetic toxicity in zebrafish: Experimental and molecular dynamics simulation studies. Environ Sci Technol, 2021, 55: 1672–1681
109 Yu H, Song M, Hu K, et al. Influence of bisphenol compounds at nanomolar concentrations on chromosome damage induced by metabolically activated carcinogens in HepG2 cells. Environ Sci Technol, 2021, 55: 10001–10011
110 Tachachartvanich P, Sangsuwan R, Ruiz H S, et al. Assessment of the endocrine-disrupting effects of trichloroethylene and its metabolites using *in vitro* and *in silico* approaches. Environ Sci Technol, 2018, 52: 1542–1550
111 Nogueira M S, Koch O. The development of target-specific machine learning models as scoring functions for docking-based target prediction. J Chem Inf Model, 2019, 59: 1238–1252
112 Li J, Cao H, Feng H, et al. Evaluation of the estrogenic/antiestrogenic activities of perfluoroalkyl substances and their interactions with the human estrogen receptor by combining *in vitro* assays and *in silico* modeling. Environ Sci Technol, 2020, 54: 14514–14524
113 Zhang R, Wu Q, Qi X, et al. Using *in vitro* and machine learning approaches to determine species-specific dioxin-like potency and congener-specific relative sensitivity among birds for brominated dioxin analogues. Environ Sci Technol, 2021, 55: 16056–16066
114 Gawehn E, Hiss J A, Brown J B, et al. Advancing drug discovery via GPU-based deep learning. Expert Opin Drug Discov, 2018, 13: 579–582
115 Stavrou A, Dayl S, Schmid R. Homology modeling of P2X receptors. Methods Mol Biol, 2020, 2041: 65–75
116 Bitencourt-Ferreira G, de Azevedo W F Jr. Homology modeling of protein targets with MODELLER. Methods Mol Biol, 2019, 2053: 231–249

Summary for “基于计算毒理的环境污染物-生物大分子的相互作用研究”

Computational toxicology studies on the interactions between environmental contaminants and biomacromolecules

Haoyue Tan[1,2], Rong Zhang[1], Qinchang Chen[1], Chi Zhang[1], Jing Guo[1], Xiaowei Zhang[1], Hongxia Yu[1] & Wei Shi[1,2*]

[1] *State Key Laboratory of Pollution Control and Resources Reuse, School of Environment, Nanjing University, Nanjing 210023, China;*
[2] *Jiangsu Province Ecology and Environment Protection Key Laboratory of Chemical Safety and Health Risk, Nanjing 210023, China*
* Corresponding author, E-mail: njushiwei@nju.edu.cn

Through manufacturing operations, product consumption, and drug administration, humans and wildlife are continuously exposed to environmental contaminants (ECs) throughout their lives. Faced with the potentially harmful effects of ECs on humans, regulatory agencies around the world require the integration of epidemiological, *in vivo* toxicological, and *in vitro* mechanistic data to provide the necessary information for hazard classification, labeling, and risk management of chemicals. However, animal studies have time-consuming and high-cost defects and ethical problems. High-throughput *in vitro* assays are also unable to provide systematic toxicological information for chemical hazard classification for over 100000 chemicals in commerce.

Recently, computing resources and artificial intelligence have innovatively improved the accuracy and speed of machine learning (ML) algorithms, and the structural biology and deep learning (DL) algorithms (e.g., AlphaFold2 and AF2Complex) have incrementally resolved a large number of biomolecular crystal structures. Thus, the use of computational toxicology techniques in environmental toxicology has increased significantly. It is estimated that computational toxicology techniques can perform virtual screening of millions of compounds in a limited amount of time for the contaminant-biomolecule interaction process. Thus, computational toxicology techniques can reduce the initial experimental cost of identifying environmental emerging contaminants, increase information on the toxicity mechanisms of ECs, and improve the efficiency of hazard identification of ECs by regulatory authorities.

This study systematically reviews computational toxicology techniques commonly used to analyze ECs-biomolecule interactions, including molecular docking, molecular dynamics (MD) simulation, and machine learning-based modelling. Molecular docking is a well-established molecular simulation method that explores the interactions between biomolecules and small molecules to predict their binding modes and binding affinities. MD can simulate the flexible binding process of contaminant-biomolecule and the dynamic conformational shift process of contaminant-biomolecule complexes, providing more comprehensive information on the interaction mechanism. Machine learning-based modelling is a novel computational toxicology technique that is completely different from molecular simulation. It mainly uses publicly available structural information and *in chemico*, *in vitro*, and *in vivo* bioactivity data to construct (quantitative) structure-activity relationships (Q)SARs based on ML algorithms, and uses (Q)SAR models to rapidly improve the efficiency of virtual screening of ECs targeted at biomolecules, and further deepen the analysis of contaminant-biomolecule interaction mechanisms in complex biological contexts. In this paper, the main applications of these techniques in the field of environmental toxicology in recent years are systematically reviewed, including the mechanistic studies of contaminant-biomolecule interactions and high-throughput virtual screening. The advantages and limitations of these techniques in terms of ligand-receptor interaction, explainability, training efficiency, computational depth, and biological processes are discussed. Results showed that MD simulations, which can deeply explore contaminant-biomolecule interactions, are unable to achieve the high-throughput virtual screening of ECs. Machine learning-based modelling, which can reflect the complex biological processes and achieve high accuracy prediction, unable to effectively interpret the prediction results due to the ‘black box’ defects. Molecular docking, which has the potential to be a high throughput virtual screening method, is limited by local sampling and approximate scoring function deficiencies. These problems limit the ability of molecular docking to analyze more comprehensive mechanisms of ECs-biomolecule interactions. Therefore, only the combination of multiple computational toxicology techniques to develop an integrated workflow for mechanistic analysis and virtual screening can compensate for their respective shortcomings and obtain optimal results. However, how to increase the computational throughput screening while maintaining the mechanism analysis remains a key issue to be addressed in the future.

computational toxicology, molecular docking, molecular dynamics simulation, machine learning, virtual screening, mechanisms analysis

doi: 10.1360/TB-2022-0613