

## 计算生物学在药学研究中的应用

汤佳宁, 孙 洋\*

(南京大学生命科学学院, 医药生物技术全国重点实验室, 江苏 南京 210023)

**摘要:** 计算生物学 (computational biology) 属于生物学的一个分支, 是指开发和应用数据分析及理论的方法、数学建模和计算机仿真技术等, 用于生物学、行为学和社会群体系统的研究的一门学科。近年来, 计算生物学越来越受到重视, 这主要是因为高通量技术大大提高了数据的生成能力。海量计算数据的生成带动了设备计算能力的不断增长, 数据的复杂性和异质性也需要更加系统的分析方法, 它们共同推动着计算生物学的发展。随着大数据时代的发展, 生物学研究和实验的数据呈指数型增长, 单一的观察和实验已经无法支撑高复杂性的数据, 计算生物学成为药物研究的一个重要且有力的工具。目前, 计算生物学已经应用于肽合成与小分子设计、虚拟筛选、蛋白质结构与蛋白质相互作用预测、数据库构建、生物活性物质预测和药物释放监测等多个领域, 从而为药物靶点预测、药物设计、药物筛选及药物的临床应用等多个药物研究的过程提供帮助。计算生物学如今已成为相较于传统药物研究工具更加省时省力的“臂膀”, 同时药物研发的需求也不断推动着计算生物学的发展。二者相辅相成, 共同发展, 计算生物学已成为药物研究不可分割的一部分。

**关键词:** 计算生物学; 药学; 药物靶点预测; 药物筛选; 药物设计; 应用

中图分类号: R966 文献标识码: A 文章编号: 0513-4870(2024)08-2192-07

## Application of computational biology in pharmaceutical research

TANG Jia-ning, SUN Yang\*

(State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing 210023, China)

**Abstract:** Computational biology is a branch of biology that develops and applies data analysis and theoretical methods, mathematical modeling, and computer simulation techniques for the study of biological, behavioral, and social group systems. In recent years, computational biology has received increasing attention, mainly because high-throughput technologies have greatly improved the ability to generate data. The generation of massive computational data drives the increasing computing power of devices, and the complexity and heterogeneity of data also require more systematic analysis methods, which jointly promote the development of computational biology. With the development of the era of big data, the data of biological research and experiments has increased exponentially. A single observation and experiment can no longer support high-complexity data, and computational biology has become an important and powerful tool for drug research. At present, computational biology has been applied in many fields, such as peptide synthesis and small molecule design, virtual screening, protein structure and protein interaction prediction, database construction, bioactive substance prediction and drug release monitoring, so as to provide assistance for multiple drug research processes such as drug target prediction, drug design, drug screening and drug clinical application. Computational biology has become a more time- and labor-saving "arm" than traditional drug research tools, and at the same time, the demand for drug research and development continues to promote the development of computational biology. The two complement each other and

收稿日期: 2023-11-06; 修回日期: 2024-03-01.

基金项目: 国家重点研发计划资助“中医药现代化”重点专项 (2022YFC3500200, 2022YFC3500202).

\*通讯作者 Tel: 86-25-89681312, Fax: 86-25-89687620, E-mail: yangsun@nju.edu.cn

DOI: 10.16438/j.0513-4870.2023-1256

develop together, and computational biology has become an integral part of drug research.

**Key words:** computational biology; pharmacy; drug target prediction; drug screening; drug design; application

计算生物学 (computational biology) 属于生物学的一个分支, 是指开发和应用数据分析及理论的方法、数学建模和计算机仿真技术等, 用于生物学、行为学和社会群体系统的研究的一门学科。近年来, 计算生物学越来越受到重视, 这主要是因为高通量技术大大提高了数据的生成能力。海量计算数据的生成带动了设备计算能力的不断增长, 数据的复杂性和异质也需要更加系统的分析方法。它们共同推动计算生物学的发展。计算生物学的最终目的不仅仅局限于测序, 而是运用计算机的思维解决生物问题。计算生物学为研究者对生命的理解带来了秩序。它使生物学概念变得严谨和可测试, 用计算机的语言和数学的逻辑构建和描述并模拟出生物世界。

与此同时, 对于与生物学密不可分的药物研究过程来说, 计算生物学已经渗透进药物研究的各个阶段和领域。在计算生物学出现之前, 药物的研发大多依靠观察和实验来进行。随着大数据时代的发展, 生物学研究和实验的数据呈指数型增长, 单一的观察和实验已经无法支撑高复杂性的数据, 计算生物学成为药物研究的一个重要且有力的工具。目前, 计算生物学已经应用于肽合成与小分子设计、虚拟筛选、蛋白质结构与蛋白质相互作用预测、数据库构建、生物活性物质预测和药物释放监测等多个领域, 从而为药物靶点预测、药物设计、药物筛选及药物的临床应用等多个药物研究的过程提供帮助。如计算代谢组学可以揭示药物作用的代谢途径和可能的不良反应; 基因组学和遗传学分析信息有助于理解药物作用的分子机制, 开发针对特定遗传背景的个性化治疗方案; 生物信息学数据库和工具为药学研究提供了丰富的数据资源和分析方法; 临床试验数据的计算生物学分析方法帮助研发者更好地评估药物的效果和安全性。如今计算生物学已成为相较于传统药物研究工具更加省时省力的“臂膀”。同时, 药物研发的需求也不断推动计算生物学的发展。二者相辅相成, 共同发展, 计算生物学已成为药物研究不可分割的一部分。

## 1 计算生物学的发展

计算生物学是一门典型的交叉学科。涉及的学科包括数学、统计学、化学、物理学、生物学和计算机科学等。在“计算生物学”这一称谓被正式确定下来之前, 曾有人将这门学科定义为“数学生物学” (mathematical biology) 或“定量生物学” (quantitative biology)。但数

学的统计学分析和定量的测度在整个自然科学领域都广为应用, 无法很好地显示出这门学科将学科相关的各个领域相互联系又彼此统一的特点。因此, 最终“计算生物学”这一概念被大家广泛认可和接受。

1990年, 人类基因组计划 (Human Genome Project, HGP) 拉开帷幕, 与之相伴而来的是大量的数据信息。研究人员发现用传统的方式已经无法存储和处理这些海量信息, 更无法很好地加以利用和研究。因此人们开始尝试使用信息学的方法来处理信息, 生物信息学也就应运而生了。随着信息时代和医药领域的进一步发展, 人们再次意识到传统的信息学处理已经无法满足医学和药学等生命科学相关研究的需求, 人们迫切地需要一个更为系统的研究方法体系来帮助人们处理爆炸增长的信息和了解人体这一复杂神秘的生命系统。计算生物学出现在这样一个大数据时代, 并且成为生命科学研究不可或缺的有力手段<sup>[1]</sup>。

伴随着20世纪80年代计算机科学与技术发展, 以及生物化学、分子生物学的系统论建立, 1975年, 美国的皮帕斯和麦克马洪首先提出运用计算机技术预测RNA二级结构。1980年, 美国《科学》周刊发表了关于计算分子生物学的综述。1997年, 国际计算生物学学会 (ISCB) 在美国成立, 如今发展成为一个拥有来自70多个国家的2500多名会员的庞大组织。由其负责的两本刊物, 《公共科学图书馆计算生物学》(PLoS Computational Biology) 和《生物信息学》(Bioinformatics), 成为当今计算生物学领域主流的专业杂志。1989年, 在美国召开了生物化学系统论与生物数学的国际会议, 讨论了生物系统理论的计算机模型研究方法, 开创了计算生物学的发展。它属于早期计算系统生物学家的研讨会, 后来改为国际分子系统生物学会会议 (ICMSB), 第11届国际分子系统生物学会会议在中国科学院-马普学会计算生物学伙伴研究所成功举办。1999年, 世界上第一个系统生物研究所在美国西雅图成立。2002年秋天, 第一届欧洲计算生物学会会议 (ECCB) 在德国萨尔布吕肯召开, 吸引了来自30个国家的459名参会者。2002年11月, 英国《自然》杂志推出一个关于计算生物学的专辑, 介绍计算生物学的学科概念、研究现状和发展前景。2004年, 美国卫生研究院启动了“生物信息学和计算生物学”计划, 并着手建立了数个“国家生物医学计算中心” (National Centers for Biomedical Computing)。近年来, 加州大学、斯坦

福大学、德克萨斯大学、芝加哥大学、威斯康星大学等机构均成立了计算生物学中心,德国、法国、澳大利亚、意大利等国也纷纷建立了计算生物学研究机构。计算分子生物学国际会议(RECOMB)则是目前计算生物学和生物信息学领域最具影响力的会议。自1997年开办至今,受到了来自计算机科学和应用数学领域参会者的广泛关注和认可。

中国的计算生物学起步相比于国际上较晚,但计算生物学在国内也受到越来越多的关注和重视。2005年10月13日,中国科学院上海生命科学研究院计算生物学研究所(中国科学院-马普学会计算生物学伙伴研究所)正式揭牌,也代表着我国第一个计算生物学研究所正式成立。此后,一些大学也相继成立了系统生物学研究机构。2022年5月13~15日,由浙江大学生命科学学院主办的第十届生物信息学与计算生物学国际会议(The 10th International Conference on Bioinformatics and Computational Biology, ICBCB 2022)成功举办,围绕计算生物学、生物化学与分子生物学、生物信息学与组学分析、生物信息学与计算生物学、序列分析、生物标志物识别与蛋白质结构预测等主题进行了研究成果交流和学术研讨。中国的计算生物学也越来越受到人们的重视,在近年得到突飞猛进的发展,成为生命科学研究必不可少的工具和手段。

## 2 计算生物学与药物靶点预测

药物靶点通常是在代谢或信号通路中与特定疾病或病理状态有关的关键分子,药物靶点的预测在药物发现的过程中起重要作用。传统仅依靠实验来确定药物-靶点相互作用的方法,如核磁共振(NMR)技术、RNAi技术、转染技术和基因敲除动物等,不仅成本昂贵,且整个过程充满着高度的不确定性,需花费实验人员大量的时间和精力<sup>[2]</sup>。因此,迫切地需要一种新颖的药物靶点预测方法来解决传统的药物靶点预测方法的弊端,计算生物学在药物靶点的发现过程中发挥着越来越重要的作用,尤其是大规模多组学的数据分析。目前,大量疾病相关的数据库资源、多基因芯片、蛋白质组、转录组、代谢组等数据的积累为计算生物学方法预测和发现药物潜在靶标提供了可能性。

利用计算生物学预测药物靶点的过程一般包括:首先利用已有数据库或基因组、转录组、代谢组等手段获取疾病与非疾病状态下靶点预测所需要的生物学数据,并利用计算生物学手段对所获得的数据进行处理和分析,接下来对数据处理与筛选所确定的生物分子进行功能分析,从而确定该疾病相关的候选靶标,后续进一步针对该靶标进行实验层面的药理学验证,研究确定靶标的特异性和可药性等<sup>[3]</sup>。计算生物学在这一

过程中可快速、高效地确定疾病相关的可能靶点,并为后续湿实验的研究提供可靠的指导,减少传统的依靠实验确定药物靶标的不确定性。

计算生物学在破译核酸序列中的遗传规律,模拟生命体内的信息流过程,认识生命体的代谢、发育及进化方面都扮演着重要的角色,为药物研发和药物靶点预测奠定了基础。在破译核酸序列中的遗传规律方面,计算生物学通过识别基因组序列中的插入缺失等变异,对这些变异与特定疾病或性状间的关联性进行分析,从而更好地理解核酸序列的遗传规律;随着高通量测序技术的发展,利用计算生物学对生成的大量基因组学数据进行分析处理,从而对组学数据和遗传规律进行更好地解释。在模拟生命体内信息流过程方面,计算生物学通过构建蛋白质-蛋白质相互作用网络(PPI网络)、使用高通量技术进行转录组学和基因表达分析、代谢网络建模、构建信号传递网络等方法促进模拟生命体内信息流过程方面的研究。计算生物学在认识生命体代谢、发育和进化等过程中的应用较为广泛和深入,通过通量平衡分析(FBA)、代谢组学数据分析、基因调控网络建模、比较基因组学等手段来应用于生命体代谢、发育和进化方面的研究。

计算生物学的应用离不开数据库的大量信息,通过对大量数据的分析和模拟来筛选可能的药物靶点是计算生物学参与药物靶点确定的重要途径。为了确定某个基因和疾病的关系,可以通过OMIM、Locus Link、The Human Gene Mutation和储存癌症候选基因COSMIC等数据库获取。Therapeutic Target Database(TTD)、Drug Bank、ChEMBL和Potential Drug Target Database(PDTD)等数据库则可为研究者提供药物靶标的生物信息资源。来自美国的研究团队在《Briefings in Bioinformatics》发表综述,从43个数据库/工具中筛选出提供基因→疾病搜索的数据库,包括ClinVar、Copy Number Variations in Disease、DiseaseEnhancer、miR2Disease等,方便研究人员使用<sup>[4]</sup>。而美国基因组研究中心(NCGR)的GeneX、欧洲生物信息研究所(EBI)的ArrayExpress、美国生物技术信息国家中心(NCBI)的Gene Expression Omnibus(GEO)、MIT的ChipDB等数据库则提供了大量基因的相关数据信息。NCBI GEO(<https://www.ncbi.nlm.nih.gov/geo/>)<sup>[5]</sup>、The Cancer Genome Atlas(TCGA)(<https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>)、Arrayexpress(<https://www.ebi.ac.uk/arrayexpress/>)<sup>[6]</sup>是一些含有基因表达数据的大存储库。这些都是计算生物学用于药物靶标发现的重要工具和资源。通过对基因表达特征的分析,可以找出不同疾病的靶基因。

如 van IJzendoorn 等<sup>[7]</sup>使用机器学习 (ML) 方法和基因表达数据, 在 2019 年发现了新的罕见软组织肉瘤生物标志物和潜在药物靶点。

靶标发现过程涉及多方面的研究, 包括疾病相关基因、信号通路、蛋白质相互作用和小分子-蛋白质相互作用的研究。借助计算生物学这一工具可以高效且有效地识别具有成为药物靶点潜力的生物分子。随着高通量测序技术的进步, 不断产生海量的组学数据。此类大规模组学数据 (基因组学、转录组学、蛋白质组学、代谢组学等) 的处理和数据挖掘分析对生物学、医学和药理学来说是革命性的, 特别是在促进研究人员理解复杂的生物系统和过程方面。如使用组学数据发现了新的候选疾病目标, 如 SETD2 和 VGLL4<sup>[8]</sup>。许多在可能与特定疾病相关的生物过程中发挥重要作用的基因或蛋白质已根据组学数据进行鉴定。研究人员可以利用计算生物学作为工具, 通过对组学和多组学数据等方法进行分析处理, 从而促进药物靶标发现的研究。

计算生物学算法是药物靶标发现及发展中的重要成果之一。文本挖掘驱动数据库, 如 DisGeNET、STITCH、STRING 等为确定一种特定疾病的新型分子治疗靶点提供了一个广阔的途径<sup>[9]</sup>。2020 年, Gu 等<sup>[10]</sup>采用相似组合法确定了 197 种最常用中草药的靶标。后来, DisGeNET 数据库被用于将药物靶标与不同疾病联系起来, 从而将已知靶标的中草药与可用疾病的靶标匹配关联起来。此外, 2019 年, Chen 等<sup>[11]</sup>使用 STITCH 寻找食管癌的潜在药物的靶标。同样, Taha 等<sup>[12]</sup>利用 STITCH 为南天竹 (一种用于治疗各种肿瘤的植物) 的活性成分寻找靶点。

### 3 计算生物学与药物筛选

药物筛选指的是采用适当的方法, 对可能作为药物使用的物质进行生物活性、药理作用及药用价值的评估过程。随着人类基因组计划的完成, 潜在的药物靶点被不断地发现, 新的药物靶点为人类疾病的治疗带来曙光的同时, 也对针对特定靶点的药物筛选提出了挑战。大量的靶点和药物及二者相互作用的信息和数据需要一个完整且系统的体系来进行处理, 计算生物学成为基于靶点的药物筛选过程中的重要工具。

从传统的角度来看, 有两种发现新药物的方式<sup>[13]</sup>: 表型药物发现 (PDD) 和基于靶点的药物发现 (TDD)<sup>[14]</sup>。药物发现研究已经进入 TDD 时代<sup>[15]</sup>, TDD 逐渐取代 PDD, 成为药物发现的主流方式。TDD 以“一个基因、一种药物和一种疾病”的概念为中心<sup>[16]</sup>, 药物的研发也进入分子靶点时代。药物筛选是指从大量化合物中筛选出具有潜在药用价值的候选药物的过程。计算生物

学在药物筛选过程中发挥着越来越重要的作用。研究者可以计算生物学的计算和统计能力为工具, 基于靶标的药物发现有望通过系统的计算机处理技术实现更大的通量<sup>[17]</sup>。因此, 高通量筛选、虚拟筛选 (virtual screening)、分子对接、基于结构的药物筛选等计算生物学相关的技术得到了蓬勃发展。

计算生物学参与药物筛选的方法有许多种, 为药物筛选和先导化合物优化提供了许多有力的工具, 如虚拟筛选、分子对接、药效团模型和吸收、分布、代谢、排泄、毒性 (ADMET) 预测等。计算生物学通过虚拟筛选这一计算方法来从大型化合物库中快速筛选出可能与特定生物靶标相互作用的候选化合物, 包括结构基础的筛选和配体基础的筛选, 来快速筛选出可能的目标化合物。分子对接模拟则是通过计算的方式预测目标化合物与其靶蛋白的结合模式和亲和力, 从而更好地研究药物与靶标的相互作用过程, 并可用于优化化合物的药效。药效团分析 (pharmacophore modeling) 则是建立一种表示化合物必需的空间排列特征以实现与特定生物靶标相互作用的模型, 主要用于新化合物的设计和优化。ADMET 预测则代表吸收、分布、代谢、排泄和毒性, 计算生物学通过预测候选药物的 ADMET 属性从而评估其潜在的药学价值, 对药物进行筛选。接下来将对一些计算生物学辅助药物筛选的工具和方法实例进行简要阐述。

目前, 以计算生物学为工具的 DNA 编码化合物库筛选技术 (DNA Encoded Library-HitGen) 成为药物筛选的一种强有力的工具。DNA 编码化合物库 (DNA Encoded Compound Library, 简称 DEL) 合成与筛选的概念最早由美国 Scripps 研究院的 Sydney Breener (2002 年诺贝尔生理与医学奖获得者) 和 Richard Lerner (时任 Scripps 研究所所长) 于 1992 年提出。DEL 技术将活性靶点蛋白和 DNA 编码化合物库孵育, 亲和力强的化合物与蛋白结合; 亲和力弱或不结合的化合物被除去; 由于化合物与 DNA 编码信息一一对应, 可以通过高通量测序技术得到高亲和力化合物的结构信息; 重新合成不带 DNA 标签的化合物后进行活性验证及结构优化, 得到苗头化合物。DEL 技术是组合化学和分子生物学的完美结晶, 并在计算生物学的迅速发展下得到了巨大的推动, 大幅提高了新药筛选的效率。

此外, 利用计算生物学根据药物的 ADMET 评估进行药物筛选也是计算生物学参与药物筛选的实例。在药物进入临床之前, 药物的毒性评估是 I 期临床试验的主要关注点。在早期药物发现阶段, 有数百种化合物等待评估其 ADMET 特性, 通过广泛的动物研究验证每一种化合物将耗费大量的时间和财力<sup>[18]</sup>。因

此,可以利用计算生物学的方法,通过提前对药物进行 ADMET 测试,从而降低临床试验失败的风险,节省金钱和时间成本<sup>[19]</sup>。如药物毒性是指由于化合物的作用方式或新陈代谢,化学分子对有机体或有机体任何部分的不利影响。随着综合毒性评估方法的研究进一步深入,开发出不同的毒性预测软件和基于计算生物学的工具,如 Tox21 (<https://ntp.niehs.nih.gov/whatwestudy/tox21/index.html>)、SEA (<http://sea.bkslab.org/>)、eToxPred (<https://www.brylinski.org/etoxpred-0>)<sup>[20]</sup> 和 TargeTox (<https://github.com/artem-lysenko/TargeTox>)。Tox21 评估了 12 707 种环境化合物和药物的毒性,而 SEA 基于目标药物生物网络预测了 656 种在售药物对 73 种非预期目标的毒性风险。2016 年, Huang 等<sup>[21]</sup>通过 Tox21 预测了 10 000 多种化合物的体内毒性特征和机制特征。而同年, Zhou 等<sup>[22]</sup>利用 SEA 预测了癌症相关蛋白。Gupta 和 Rana<sup>[23]</sup>在 2019 年应用 eToxPred 预测了雄激素受体小分子的毒性。这不仅节省了大量时间财力成本,也提高了药物筛选的准确性和针对性。

新冠疫情中的疫苗研发过程也成为计算生物学辅助药物筛选研发的重要应用实例:新冠疫情在全球暴发,计算生物学在开发有前途的抗 COVID-19 药物方面显示出比以往任何时候更大的潜力。自 COVID-19 暴发以来,进行了十亿级超高通量虚拟筛选以寻找潜在的治疗药物<sup>[24]</sup>。我国在国产超级计算机上用不到 24 h 完成了 AutoDock-GPU 对 SARS-CoV-2 RdRp 的十亿级虚拟筛选,大大提高了药物筛选的效率。此外,广泛的分子动力学模拟 (MD) 也揭示了聚糖在调节 SARS-COV-2 刺突蛋白的构象动力学中的作用,这为进一步的药物和疫苗设计提供了参考<sup>[25]</sup>。计算生物学在新冠病毒药物和疫苗研发的过程中成为重要的工具和手段,大大提高了药物研发的速度和效率,为人类创造福祉。随着对更为复杂生物体的研究,未来的药物筛选将不可避免地检测更多样和更复杂的数据和表型,这也为计算生物学的发展创造新的机遇和挑战。

#### 4 计算生物学与药物设计

传统的药物设计过程费时费力,一般需要测试数千种物质,耗时 10~20 年,但研发成功的效率却并不高,投入与产出很不平衡。因此,人们开始使用计算机辅助药物设计方法来促进药物设计的进程。随着计算生物学的发展,基因组学、蛋白质组学和代谢组学正在深刻地改变药物研发的传统方法<sup>[26]</sup>。生物学通常是计算密集型的,尤其是在处理庞大的数据集或运行详细的模拟时<sup>[27]</sup>。计算机辅助药物设计不仅极大地减少了药物研发所需要的成分,更大大缩短了新药研发的周期,提高了药物研发的效率,生物信息学成为药物设计

的重要工具<sup>[28]</sup>。目前,使用计算机和计算生物学辅助药物的设计主要包括基于结构的药物设计(也称为合理药物设计, SBDD)、基于配体的药物设计 (LBDD)、高通量虚拟筛选 (HTVS) 等技术,它们突破了传统的先导物发现模式,极大地促进了先导化合物发现和优化。

SBDD 成为一种利用计算生物学剖析结构信息以加快药物研发过程,从而改善候选新药过程的技术。据估计 SBDD 可以从目标识别到调查新药 (IND) 降低成本 50%。该技术要求抑制剂的高分辨率三维结构绑定到使用 X 射线晶体学获取的目标。一旦获得结构,开始分析抑制剂与目标活性部位的相互作用。通过分析相互作用的结果来参与新药的研发和设计,从而缩短候选新药的过程。

另一种计算生物学辅助药物设计的方法则是基于片段的药物设计方法和虚拟配体筛选。分子碎片药物设计是把一个已知的药物分子剪裁成多个碎片,这些分子碎片中的一些可能继承了原有活性分子的全部或部分药理性质,再通过筛选这些分子碎片,有可能找到更好的药物分子。一般是根据实验结果初步筛选出低分子量的核心片段,其与靶点结合比较弱(通常分离常量在高微摩尔至毫摩尔范围),然后通过进一步结构优化得到活性更好的化合物(可达纳摩尔水平)。计算生物学在药物发现及发展中的重要成果之一,是预测及估计疾病网络的整体拓扑结构及动态,或药物与药物之间的相互作用或药物目标关系。

计算对接和高通量虚拟筛选是药物设计过程中的常用方法。计算对接广泛用于研究蛋白质-配体相互作用以及药物发现和开发。通常,该过程从已知结构的目标开始,如具有医学意义的酶的晶体结构,然后使用对接来预测小分子与靶标的结合构象和结合自由能。单一的对接实验可用于探索靶标的功能,而虚拟筛选(对大型化合物库进行对接和排序)可用于鉴定新的药物开发抑制剂<sup>[29]</sup>。在可预见的未来,随着高性能计算和增强采样、深度学习等技术的发展,可以开发出更多的创新算法,进一步加快药物的设计和发现<sup>[30]</sup>。

在药物设计方面,多重药理学是指设计一种能够与疾病相关药物目标生物网络中的多个目标相互作用的单一药物分子。它最适合为更复杂的疾病设计一种有前途的治疗药物,如癌症、神经退行性疾病、糖尿病、心脏衰竭等。计算生物学由于具有较强的挖掘能力和数据分析能力,为药物设计提供了一种理想思路。近年来计算生物学在药物发现和设计方面取得了突破,产生了基于网络的多药理学预测工具和独立的软件包,如 polypharmacology browser (PPB) (<http://www.gdb.unibe.ch/>)、TarPred (<http://www.dddc.ac.cn/tarpred/>)、

Self-Organizing Map Based Prediction of Drug Equivalence Relationship (SPiDER) (<http://modlabcbadd.ethz.ch/software/spider>)、TargetHunter (<https://www.cbligand.org/TargetHunter3D/>) 等。Van Vleet 等<sup>[31]</sup>在 2018 年实现了将 TarPred 工具用于筛选策略和改进脱靶效应预测的方法。在上述基于网络的工具中, PharmMapper 和 ChemMapper 经常用于当前的研究, 如采用 PharmMapper 研究黄芪、黄连对糖尿病的协同作用机制<sup>[32]</sup>、当归补血汤的补血机制、白花蛇舌草的多机制预测等。近期, 利用计算生物学的深度学习算法预测 (如生成对抗网络 GANs 和变分自编码器 VAEs) 设计新的药物分子, 并进一步预测分子的药理属性、生物活性和毒性, 加速药物的筛选和优化过程。在蛋白质三维结构预测方面, AlphaFold 等算法取得了显著的突破和进展, 也极大地提高了预测的准确性。计算生物学还被应用于预测 CRISPR-Cas9 等基因编辑技术的靶点特异性和不良反应, 通过计算生物学的预测, 可以帮助研究者选择更加精确和安全的编辑策略, 为药物的设计和研发过程提供帮助。此外, 使用计算生物学的时间序列分析和动态建模方法来研究生物过程的时间依赖性, 从而对动态的生物过程建模, 在研究细胞周期调控和发育过程中, 这些模型可以揭示关键的调控事件和信号。

## 5 总结

人类基因计划和蛋白质计划的开展, 为生物医药研究提供了丰富的生物学信息<sup>[33]</sup>。大量有价值数据产生的需求大大推动了计算生物学的发展, 而计算生物学的发展又成为药物研究等生命科学领域研究的有力工具, 极大地提高了药物研发的效率和成功率, 计算生物学及其相关的技术已经渗透到药物靶点预测、药物设计、药物筛选、优化先导化合物、构效关系 (SAR) 分析, 并预测 ADMET 特性等药物研发过程的各个阶段, 成为药物研发过程不可缺少的重要工具。

计算生物学是一个快速发展的学科, 它结合了生物学、计算科学、数学和工程学的多学科知识, 来更好地理解 and 模拟生物系统的行为。在未来, 随着数据处理能力的提升, 计算生物学将更加深入地研究复杂的生物系统。通过分析大量的遗传、表观遗传和临床数据, 计算生物学有望为个性化医疗和精准医学包括疾病预测、个性化药物设计和治疗方案的优化在内的多方面提供更有力的支持。与此同时, AI 和机器学习技术的集成与发展将进一步促进计算生物学的发展, 在模式识别、生物数据的解读和新药物发现等方面为人类的药物研发和医学进步提供福祉。

计算生物学的未来发展是机遇与挑战并存的: 计算生物学代表着未来生命科学和生物技术研究的发展

方向, 与此同时, 计算生物学的未来发展必定具有多组学特征, 需要高性能计算和完整的系统性研究, 这也为计算生物学的发展带来了许多挑战。计算生物学作为一门典型的交叉学科, 需要研究人员具有数学、统计学、化学、计算机科学等多学科背景, 能够进行广泛的数学建模和正确利用, 并且将各学科知识有机地整合起来, 从而更好地利用于药物研发和生命科学研究。这也对研究人员提出了更高的要求, 是对计算生物学的应用者的一个更高水平的挑战。除此之外, 为实现精准医学目标, 让计算生物学更好地为医药研究所用, 需要克服一系列关键技术难题。相比生物大数据的不断积累速度, 对不同类型大数据的整合分析严重滞后, 使得数据的价值尚未完全发挥出来。因此, 针对具体的各类科学问题来研发数据分析的新方法新流程成为目前计算生物学所面对的严峻挑战之一。计算生物学的兴起为药学领域的相关研究提供了一个有力的工具。而如何利用计算机生物学这一强大的工具, 为人类创造福祉, 还需研究者不断地深挖和学习。

**作者贡献:** 孙洋选题和构思了这篇文章, 汤佳宁进行了内容检索和手稿编写, 孙洋对手稿进行了修改提升, 所有作者都讨论了手稿的编写并对手稿发表了评论。

**利益冲突:** 无利益冲突。

## References

- [1] Liang D, Bo WH, Jiang LB. Computational biology, a new frontier of biology [J]. For Educ China (中国林业教育), 2017, 35: 139-142.
- [2] Bagherian M, Sabeti E, Wang K, et al. Machine learning approaches and databases for prediction of drug-target interaction: a survey paper [J]. Brief Bioinform, 2021, 22: 247-269.
- [3] Liu W, Xie HW. Discovery of potential drug targets based on bioinformatics methods [J]. Prog Biochem Biophys (生物化学与生物物理进展), 2011, 38: 11-19.
- [4] Zeeshan S, Xiong R, Liang BT, et al. 100 years of evolving gene-disease complexities and scientific debutants [J]. Brief Bioinform, 2020, 21: 885-905.
- [5] Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository [J]. Nucleic Acids Res, 2002, 30: 207-210.
- [6] Wang Z, Jensen MA, Zenklusen JC. A practical guide to The Cancer Genome Atlas (TCGA) [J]. Methods Mol Biol, 2016, 1418: 111-141.
- [7] van IJzendoorn DGP, Szuhai K, Briare-De Bruijn IH, et al. Machine learning analysis of gene expression data reveals novel diagnostic and prognostic biomarkers and identifies therapeutic targets for soft tissue sarcomas [J]. PLoS Comput Biol, 2019, 15: e1006826.

- [8] Rutledge J, Oh H, Wyss-Coray T. Measuring biological age using omics data [J]. *Nat Rev Genet*, 2022, 23: 715-727.
- [9] Adam NR, Wieder R, Ghosh D. Data science, learning, and applications to biomedical and health sciences [J]. *Ann N Y Acad Sci*, 2017, 1387: 5-11.
- [10] Gu S, Lai LH. Associating 197 Chinese herbal medicine with drug targets and diseases using the similarity ensemble approach [J]. *Acta Pharmacol Sin*, 2020, 41: 432-438.
- [11] Chen YT, Xie JY, Sun Q, et al. Novel drug candidates for treating esophageal carcinoma: a study on differentially expressed genes, using connectivity mapping and molecular docking [J]. *Int J Oncol*, 2019, 54: 152-166.
- [12] Taha KF, Khalil M, Abubakr MS, et al. Identifying cancer-related molecular targets of *Nandina domestica* Thunb. by network pharmacology-based analysis in combination with chemical profiling and molecular docking studies [J]. *J Ethnopharmacol*, 2020, 249: 112413.
- [13] Swinney DC, Anthony J. How were new medicines discovered? [J]. *Nat Rev Drug Discov*, 2011, 10: 507-519.
- [14] Vincent F, Nueda A, Lee J, et al. Publisher correction: phenotypic drug discovery: recent successes, lessons learned and new directions [J]. *Nat Rev Drug Discov*, 2022, 21: 541.
- [15] Li YH, Li XX, Hong JJ, et al. Clinical trials, progression-speed differentiating features and swiftness rule of the innovative targets of first-in-class drugs [J]. *Brief Bioinform*, 2020, 21: 649-662.
- [16] Hopkins AL. Network pharmacology: the next paradigm in drug discovery [J]. *Nat Chem Biol*, 2008, 4: 682-690.
- [17] Chua HN, Roth FP. Discovering the targets of drugs *via* computational systems biology [J]. *J Biol Chem*, 2011, 286: 23653-23658.
- [18] Casalino L, Gaieb Z, Goldsmith JA, et al. The roles of glycans in the SARS-CoV-2 spike protein [J]. *ACS Cent Sci*, 2020, 6: 1722-1734.
- [19] Cheng F, Li W, Liu G, et al. ADMET prediction: recent advances, current challenges and future trends [J]. *Curr Top Med Chem*, 2013, 13: 1273-1289.
- [20] Wang Y, Xing J, Xu Y, et al. *In silico* ADME/T modelling for rational drug design [J]. *Q Rev Biophys*, 2015, 48: 488-515.
- [21] Huang R, Xia M, Sakamuru S, et al. Modelling the Tox21 10 K chemical profiles for *in vivo* toxicity prediction and mechanism characterization [J]. *Nat Commun*, 2016, 7: 10425.
- [22] Zhou B, Sun Q, Kong DX. Predicting cancer-relevant proteins using an improved molecular similarity ensemble approach [J]. *Oncotarget*, 2016, 7: 32394-32407.
- [23] Gupta VK, Rana PS. Toxicity prediction of small drug molecules of androgen receptor using multilevel ensemble model [J]. *J Bioinform Comput Biol*, 2019, 17: 1950033.
- [24] Li Z, Li H, Yu K, et al. Perspective of drug design with high-performance computing [J]. *Natl Sci Rev*, 2021, 8: nwab105.
- [25] Xie T, Du GH. Pain of high throughput screening-panactive screening of interfering compounds [J]. *Acta Pharm Sin (药学报)*, 2015, 50: 925-930.
- [26] Zheng M, Zhao J, Cui C, et al. Computational chemical biology and drug design: facilitating protein structure, function, and modulation studies [J]. *Med Res Rev*, 2018, 38: 914-950.
- [27] Computation and biology: a partnership [J]. *Nat Methods*, 2021, 18: 695.
- [28] Geromichalos GD. Importance of molecular computer modeling in anticancer drug development [J]. *J BUON*, 2007, 12 Suppl 1: S101-S118.
- [29] Forli S, Huey R, Pique ME, et al. Computational protein-ligand docking and virtual drug screening with the AutoDock suite [J]. *Nat Protoc*, 2016, 11: 905-919.
- [30] Cao R, Li W, Sun HZ, et al. Basic theory and application of computational chemistry method in drug molecular design based on receptor structure [J]. *Acta Pharm Sin (药学报)*, 2013, 48: 1041-1052.
- [31] Van Vleet TR, Liguori MJ, Lynch JJ, et al. Screening strategies and methods for better off-target liability prediction and identification of small-molecule pharmaceuticals [J]. *SLAS Discov*, 2019, 24: 1-24.
- [32] Yue SJ, Liu J, Feng WW, et al. System pharmacology-based dissection of the synergistic mechanism of Huangqi and Huanglian for diabetes mellitus [J]. *Front Pharmacol*, 2017, 8: 694.
- [33] Wu YY, Xu DC, Wang Z, et al. Progress on applying of bioinformatics in drug discovery [J]. *Chin J Bioinf (生物信息学)*, 2005, 3: 89-92.