

人工智能方法在抗菌肽筛选领域的应用及展望

薛凤¹, 封硕¹, 李菁^{1,2*}

(¹中国药科大学生命科学与技术学院, 南京 210009; ²中国药科大学药物质量与安全预警教育部重点实验室, 南京 210009)

摘要 抗菌肽(antimicrobial peptides, AMPs)是一类具有广谱抗菌活性的小分子肽,其独特抗菌机制能够有效治疗感染性疾病,且不易产生耐药性。然而,利用传统实验方式虽然能够筛选出具有抗菌活性的AMPs,但是筛选过程繁琐,人工智能筛选方法则更加快捷便利,在探索新型天然抗菌肽中展现了巨大的潜力。本文总结并比较了人工智能筛选AMPs的相关策略,包括应用于模型训练的数据来源、人工智能机器模型以及应用于模型筛选新型抗菌肽的组学数据,并对应用的前景和优势进行展望,以期对抗菌肽的鉴定识别、研发改造提供新思路。

关键词 人工智能;机器学习;深度学习;抗菌肽

中图分类号 TP18;TQ465.6 文献标志码 A 文章编号 1000-5048(2023)03-0314-09

doi:10.11665/j.issn.1000-5048.2023030901

引用本文 薛凤,封硕,李菁.人工智能方法在抗菌肽筛选领域的应用及展望[J].中国药科大学学报,2023,54(3):314-322.

Cite this article as: XUE Feng, FENG Shuo, LI Jing. Application and prospect of artificial intelligence in antimicrobial peptides screening [J]. *J China Pharm Univ*, 2023, 54(3): 314 - 322.

Application and prospect of artificial intelligence in antimicrobial peptides screening

XUE Feng¹, FENG Shuo¹, LI Jing^{1,2*}

¹School of Life Science and Technology, China Pharmaceutical University, Nanjing 210009; ²Ministry of Education Key Laboratory of Drug Quality Control and Pharmacovigilance, China Pharmaceutical University, Nanjing 210009, China

Abstract Antimicrobial peptides (AMPs) are a class of small molecule peptides with broad-spectrum antimicrobial activity. Their unique antimicrobial mechanism can effectively treat infectious diseases, with rare drug resistance. However, though AMPs with antimicrobial activity can be screened by traditional methods, the whole process is complicated. The artificial intelligence (AI) screening method is faster and more convenient, with great potential in exploring new natural antimicrobial peptides. In this paper, strategies related to AMPs screening by AI were summarized and compared, including data sources applied to model training, artificial intelligence machine model and omics data applied to model screening of novel antimicrobial peptides. The application prospects and advantages were reviewed, in hope of providing new ideas for identification, research and development of antimicrobial peptides.

Key words artificial intelligence; machine learning; deep learning; antimicrobial peptides

This study was supported by the National Natural Science Foundation of China (No. 32170062); and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. 3322200020)

抗生素的长期使用导致了细菌耐药性增加以及多重耐药菌株出现,因此急需找寻具有传统抗生素功效的新型治疗物质。抗菌肽(antimicro-

bial peptides, AMPs)是一种小分子多肽物质(小于100个氨基酸残基),存在于所有生物体中,是保护宿主免受感染的先天免疫系统的主要基石,在

收稿日期 2023-03-09 *通信作者 Tel:13584069768 E-mail:lj_cpu@126.com

基金项目 国家自然科学基金资助项目(No. 32170062);江苏省研究生科研与实践创新计划资助项目(No. 3322200020)

生物系统中发挥着多种功能^[1]。AMPs 具有广谱抗菌活性,已经被证明具有传统抗生素的一些功效。因 AMPs 特殊的抗菌机制,使其能够快速杀菌,并且不易产生耐药性。因此,不论是新型 AMPs 的合成开发,还是对 AMPs 的鉴别筛选都备受关注。

筛选 AMPs 的传统方式是运用一系列实验技术从某一物种体内或者分泌物中进行提取鉴定,在体外针对候选肽段进行抑菌活性的筛查,最终得到高活性 AMPs。这些实验技术包括物质提取、分离、纯化、质谱表征等传统的实验方法^[2-3]。然而,虽然通过这些传统的实验方法能够筛选出具有抗菌活性的 AMPs,但筛选过程繁琐,会造成人力和物力过高的消耗。

在大数据背景下,随着计算机技术的日益发展,人工智能方法筛选 AMPs 相比于传统方法展现出了独特的优势。例如,人工智能方法可以对

AMPs 的抑菌活性、抵抗蛋白酶的稳定性以及细胞毒性和溶血活性多种理化参数进行分析,对净电荷含量、肽长度、氨基酸的组成、疏水性和结构倾向等多维度参数进行预测^[4],可以在海量潜在的多肽序列中精准有效地发掘新型 AMPs,从而加快 AMPs 开发和应用的进程。本文针对人工智能技术在 AMPs 筛选领域的应用及策略进行归纳和总结,以为抗菌肽的鉴定识别、研发改造提供新思路。

1 抗菌肽的抗菌优势

AMPs 种类丰富且易于合成,其中阳离子抗菌肽(cationic antimicrobial peptides, CAMPs),在抗菌反应中发挥着强大的效能。CAMPs 序列具有多样性^[5],并且大部分具有两亲性结构。CAMPs 的抗菌机制主要表现为两种:生物膜靶向渗透机制和非生物膜靶向机制(图 1)^[6]。

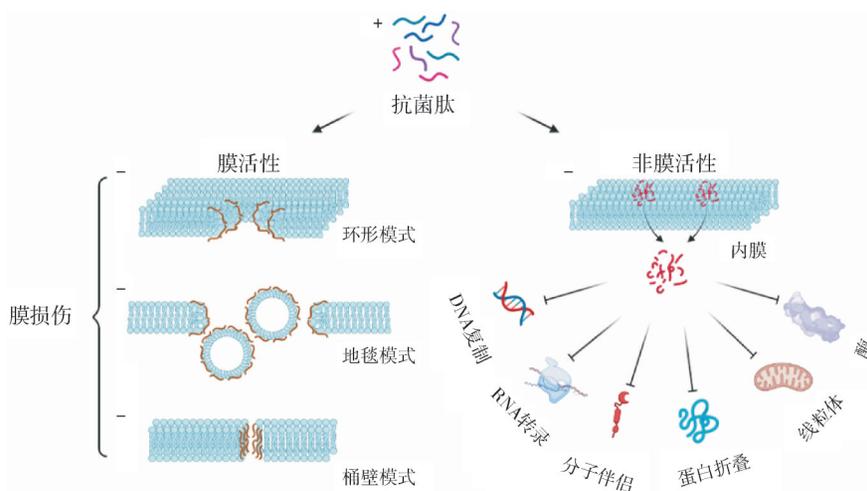


图 1 抗菌肽(AMPs)的抗菌机制

在生物膜靶向渗透机制中,CAMPs 通过静电吸引作用,吸附在带有负电荷的细菌细胞膜表面,使得细胞膜完整性受到破坏,细胞内容物释放出来,导致细菌死亡。该机制中主要有 3 种作用方式:环形模式、地毯模式、桶壁模式^[7]。

在非生物膜靶向机制中,AMPs 进入细菌细胞内部后作用于不同的靶点,发挥抗菌效应。例如抑制 DNA 和 RNA 的合成、降低细胞壁结构蛋白连接所需的酶活性、抑制核糖体功能、蛋白质合成以及阻断伴侣蛋白的正确折叠、靶向线粒体抑制细胞呼吸和诱导活性氧的形成、破坏线粒体细胞膜

的完整性等^[8]。

2 抗菌肽的研究现状

尽管 AMPs 被认为是潜力巨大的抗菌候选药物,但它们也存在一些局限性。例如,抗菌活性低、生物易降解、细胞毒性、特异性差等,这些因素限制了它们临床的应用和发展^[9-10]。目前 98 种 AMPs 已获得美国食品药品监督管理局(FDA)的批准,其中 7 种肽现已商业化,主要用作局部药物^[11]。

通过对现存常用 AMPs 数据库的综合调研,目前 AMPs 数量大约有 3 万条,其中天然来源的 AMPs

仅占总量的十分之一。据 APD₃ 数据库 (Antimicrobial Peptide Database)^[12] 记载,天然来源的抗菌肽主要来自于8个类别,包括人类宿主防御肽、哺乳动物、两栖动物、鱼类、爬行动物、鸟类、节肢动物、软体动物和原生动植物。据中国科学院北京基因组研究所^[13]统计的数据显示,截至2023年3月共计产生原始组学数据 (Genome Sequence Archive, GSA) 21.99 PB。可见,目前发现的 AMPs 数量相比于庞大的组学数据微乎其微,新型 AMPs 序列的开发空间相当广阔。

随着计算机技术的开发与应用,人工智能方法可以从数百万种候选 AMPs 中快速筛选和鉴别潜在的具有生物活性的肽段,提高筛选 AMPs 的效率和准确度。

3 人工智能方法应用于抗菌肽筛选的领域

AMPs 的抑菌活性、抵抗蛋白酶的稳定性以及细胞毒性和溶血活性等功能受多种理化参数的影响,包括净电荷含量、肽长度、氨基酸的组成、疏水性和结构倾向等^[4]。人工智能技术筛选 AMPs,主要依据训练数据集中 AMPs 和非 AMPs 的理化性质,形成独特的判别系统,通过不同特征数据集的训练,模型不断的优化和完善,最终可用于预测和筛选具有活性的 AMPs 的序列。

基因组学、转录组学、蛋白质组学等多组学技术为人工智能方法发现新型抗菌肽提供了庞大的数据来源,利用这些数据可以快速准确地筛选出新型 AMPs (图 2)。

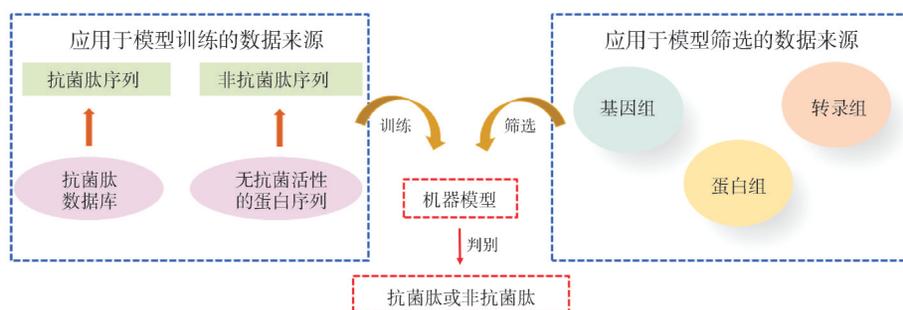


图2 人工智能在抗菌肽筛选领域的应用

3.1 应用于模型训练的抗菌肽数据

人工智能模型筛选候选 AMPs 的准确性,取决于数据集的质量。不同 AMPs 数据库收录的侧重点以及存储数据情况都存在差异,因此对数据库来源和质量的评估对于人工智能模型的训练和筛选十分关键。

目前常用的 18 个 AMPs 数据库包括综合类型抗菌肽数据库 (表 1) 和专有类型数据库 (表 2)。其

中,综合类型抗菌肽数据库可以为模型训练提供广泛的数据来源,最大的优势在于肽段序列覆盖面广,不局限活性或者某一单一来源的数据。专有类型数据库则是为了特定类型的抗菌肽数据的使用和开发而设计,具有方向性,针对性较强。例如,抗癌蛋白存储库、抗病毒多肽数据库、植物来源抗菌肽数据库、细菌素类数据库等。其专一性的优势,可以为某一特定领域的开发和研究所使用。

表1 综合类型抗菌肽数据库

数据库	简写	数据来源	数据量	参考文献
Database of Antimicrobial Peptides	dbAMP ₂	AMPs 数据库、文献	28 709	[14]
Linking Antimicrobial Peptides	LAMP ₂	文献、AMPs 数据库	23 253	[15]
Data Repository of Antimicrobial Peptides	DRAMP _{3.0}	专利抗菌肽	22 259	[16]
A database of Structurally Annotated Therapeutic Peptides	SATPdb	20 个公共领域 AMPs 数据库、两个数据集	19 192	[17]
Database of Antimicrobial Activity and Structure of Peptides	DBAASP _{v3}	人工合成的抗菌肽	18 878	[18]
Collection of Anti-Microbial Peptides	CAMP _{R3}	专利和预测抗菌肽	10 247	[19]
Antimicrobial Peptide Database	APD ₃	天然抗菌肽	3 324	[12]
Dragon Antimicrobial Peptide Database	DAMPD	UniProt 数据库	1 232	[20]

表 2 专有类型抗菌肽数据库

数据库	简写	数据来源	数据类型	数据量	参考文献
Database of Anticancer Peptides & Protein	CancerPPD	文章、专利、AMPs 数据库	抗癌肽	3 491	[21]
Database of Experimentally Determined Hemolytic and Non-hemolytic Peptides	Hemolytik	文献、AMPs 数据库	溶血肽、非溶血肽	2 970	[22]
A Database of Antiviral Peptides	AVPdb	实验验证	抗病毒肽	2 683	[23]
Yet another Database of Antimicrobial Peptides	YADAMP	文献	抗细菌肽	2 525	[24]
Anti-tubercular Peptides Database	AntiTbPdb	文献、专利	抗结核肽	1 010	[25]
Database of FDA-approved Peptide and Protein Therapeutics	THPdb	出版物、专利、DrugBank	治疗性肽	852	[26]
Invertebrate Antimicrobial Peptide Database	InverPep	文献、AMPs 数据库	无脊椎动物、抗菌肽	702	[27]
Antimicrobial Plant Peptides	PhyAMP	文献、UniProt	植物抗菌肽	273	[28]
Bacteriocins Database	BATIBASE	文献、UniProt	细菌素	229	[29]
Biofilm-active AMPs Database	BaAMPs	文献	生物膜活性肽	221	[30]

3.2 人工智能模型

3.2.1 经典机器学习模型 已有许多机器学习算法可以识别具有抗菌功能的候选肽,包括随机森林(RF)^[31]、k-近邻(KNN)^[32]、模糊最近邻(FKNN)^[33]、支持向量机(SVM)^[34-35]、决策树(DT)^[36]、逻辑回归

(LR)^[37]、朴素贝叶斯(NB)^[32,38]、二次判别分析(QDA)^[39]、隐马尔可夫模型(HMM)^[40-42]和集成学习^[43]等。在众多分类器中,RF和SVM作为经典机器学习算法广泛用于新型AMPs的发现(表3)。

表 3 机器学习算法模型

模型	算法	训练数据
IAMPE	NB、KNN、SVM、RF、XGBoost	CAMP, LAMP, ADAM, AntiBP(Server for antibacterial peptide prediction)
AmPEP	RF	APD
Target-AMP	KNN、RF、SVM	APD
AntiBP ₂	SVM	APD
CS-AMPPred	RF、SVM	APD
CAMP _{R3}	RF、SVM、DA	CAMP
C-PAmP	RF、SVM	CAMP、PhytAMP

NB:朴素贝叶斯;KNN:k近邻;SVM:支持向量机;RF:随机森林;XGBoost:极致梯度提升

在各种机器学习算法中,Kavousi等^[32]开发了一个新模型IAMPE,利用氨基酸的¹³C NMR光谱,将AMPs序列分为不同的基团并进行聚类,构建AMPs序列的特征向量并使用NB、KNN、SVM、RF和极限梯度增强(XGBoost)等多种分类器算法筛选和判别具有高活性的AMPs序列。Bhadra等^[44]开发了一项通过随机森林算法进行AMP预测的精确计算方法AmPEP。该算法是基于大量不同的AMPs和非AMPs序列特性建立分类器,从而获得了较高的准确度,可以精准地预测具有AMPs活性的肽段序列。Jan等^[45]开发了一种用于发现候选AMPs的计算方法Target-AMP,采用了KNN、RF、SVM 3种分类技术,可以有效、准确,筛选出抗微生物肽。除此之外还有AntiBP2^[46]、CS-AMPPred^[47]、

CAMP_{R3}^[19]、C-PAmP^[48]等依据AMPs的多种物理化学性质,由RF、SVM机两种经典分类器处理,对每个输入的肽段的氨基酸组成、电荷、极性、等电点、疏水性和二级结构进行分类和判别,筛选出具有活性的AMPs(图3)。

3.2.2 深度学习网络 除了机器学习方法之外,深度学习的算法也应用到AMPs的筛选。例如,人工神经网络(ANN)^[49]、长短期记忆(LSTM)^[50]、注意力(Attention)、卷积神经网络(CNN)^[51]、图神经网络(GNN)等。与经典的机器学习方法相比,深度学习的算法是通过对肽段进行特征提取,具有从原始输入数据中自动提取更高级别的特征的潜力,包括氨基酸序列、图像等用于后续的分类(图4)。例如,CNN可以在图片(二维类型数据)和文

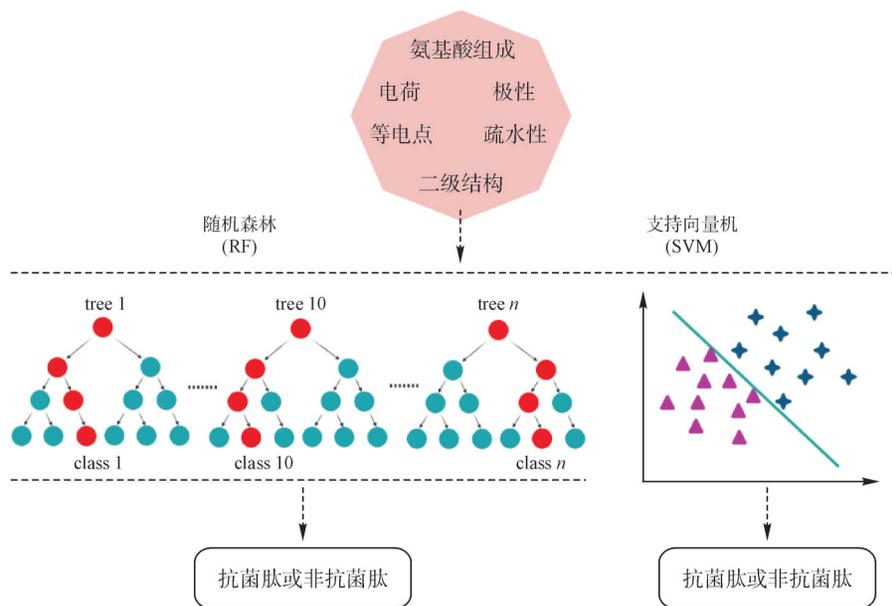


图3 经典机器学习算法应用发掘 AMPs

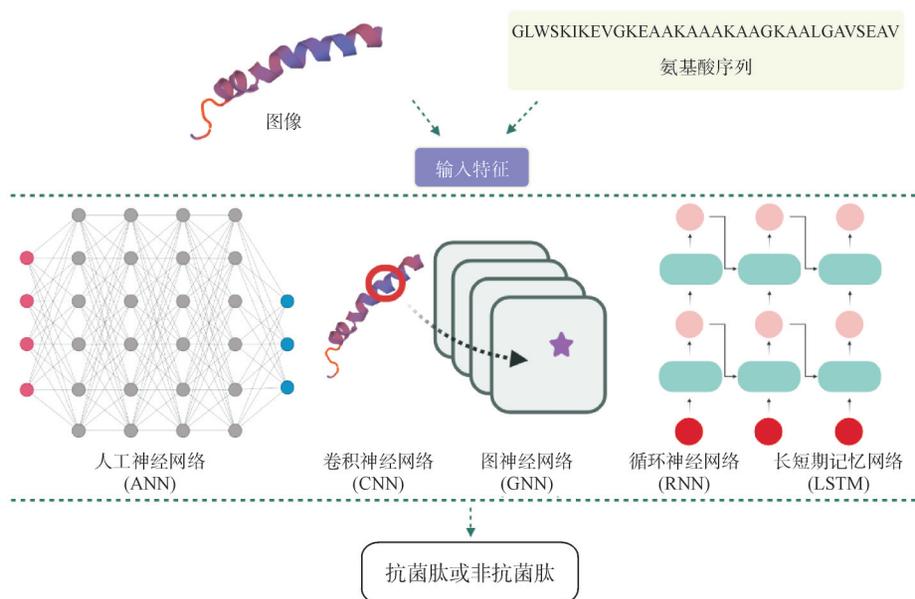


图4 深度学习算法应用发掘 AMPs

本(一维类型数据)提取特征进行采样和分类。GNN通过网络中节点之间的信息传递的方式来获取图中相互联系的信息来筛选候选 AMPs。

在深度学习的算法筛选 AMPs 的研究中, Xiao 等^[52]于 2021 年提出了一种新的算法模型 iAMP-CA2L, 混合了 CNN、LSTM、SVM 3 种算法用于预测 AMPs。iAMP-CA2L 使用 2 级预测器, 不仅筛选出 AMPs 并且可以判别出其具有何种活性。其中, 第 1 级是识别给定的肽段是 AMPs 或非 AMPs, 而第 2

级则按照不同的抗菌活性分为抗菌肽、抗病毒肽、抗真菌肽、抗生物膜肽等 10 个功能类别, 预测它是否具有一种或多种抗菌功能。

中国科学院微生物研究所王军及陈义华课题组结合了 LSTM、Attention 和 BERT 等多个自然语言处理神经网络模型, 从人类肠道微生物组数据中成功识别出候选 AMPs^[53]。而 Wang 等^[54], 针对大肠埃希菌潜在的具有抗菌活性的 AMPs 序列生成短的新序列集, 用 LSTM 算法筛选新序列集中是否有抗

微生物肽活性的序列,发掘出新型 AMPs 序列。

综上所述,经典机器学习模型和深度学习算法广泛应用于发掘新型抗菌肽,人工智能技术成为筛选新型 AMPs 的有效工具。

3.3 应用于模型筛选新型抗菌肽的组学数据

随着测序数据的爆炸式增长,多组学技术产

生了庞大的生物数据,其中包括基因组学、转录组学、蛋白质组学等。多组学技术产生的肽段序列可以作为人工智能筛选 AMPs 的数据来源。人工智能技术将依据抗菌肽序列的特征、理化性质等进行分类和鉴别,从而发掘出新型高效的 AMPs(图 5)。

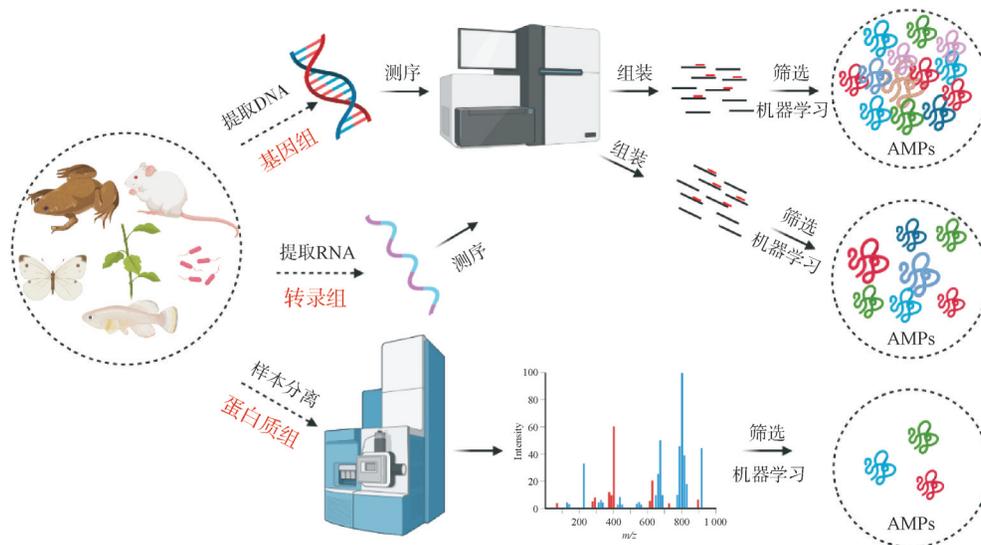


图 5 应用于模型筛选 AMPs 的数据来源

3.3.1 基因组数据 基因决定着 AMPs 的氨基酸序列,氨基酸序列又影响着多种物理参数包括净电荷含量、螺旋结构和疏水性等。因此,基因组数据是筛选出潜在的 AMPs 的重要来源。例如,中国科学院大学华大教育中心对草鱼胃肠道微生物群的整个宏基因组进行了测序,采用同源搜索的方式以预测 AMPs,成功鉴定出了 5 种与先前报道的细菌毒素高度相似的 AMPs^[55]。

虽然通过常规的基因组数据分析可以是筛选出 AMPs,但是标准的同源序列比对方法精度较低,且进行功能分类的召回率也较低。而机器学习算法可以在全基因组范围进行搜索,发现编码新型 AMPs 的基因。例如,Fingerhut 等^[56]开发了 AMPs 分类框架 *ampir*,可以应用于全基因组数据的筛选,比常规的同源搜索的方法具有更高的分类准确度。又如,Sharma 等^[57]开发了 *AniAMPpred*,基于 SVM 算法对具有不同序列长度的 AMPs 和非 AMPs 进行高精度分类,识别动物基因组中可能存在的抗菌蛋白(PAP),成功地从软体动物的泽蛭(*Helobdella robusta*)的基因组中鉴定了 436 个 PAP。

3.3.2 转录组数据 转录组学测序技术可以全

面快速的获取某一物种特定器官或组织在某一状态下的几乎所有转录本,可有效筛选表达的 AMPs 序列。例如, Lee 等^[58]基于 CAMP 数据库,将转录组测序中的肽段映射到 CAMP 数据库中以识别新的 AMPs,产生了 193 种新肽,并且在体外测试了 3 个肽段对革兰氏阴性菌和酵母菌显示出更强的抗菌活性。

常规的转录组学分析可以鉴别出新型 AMPs,但是依旧存在着候选序列过于庞大,无法全部进行体外实验验证其活性,导致出现较高的假阳性。而人工智能机器学习算法可以针对大量候选的 AMPs,依据其抗菌功能重要的理化参数进行深层次的筛选和鉴定。例如,Shelenkov 等^[59]开发了 *Cysmotif*,可以在短时间内获得一个或多个转录组中具有抗菌活性的 AMPs 序列。又如, Grafskaja 等^[60]设计了捕获含有 AMPs 毒素样蛋白质的搜索算法,从海葵触手分泌物的转录组数据中成功筛选到了候选 AMPs。

3.3.3 蛋白质组数据 蛋白质组学技术是在整体水平上研究蛋白质的特征,可以在蛋白层面筛选出有功能性的 AMPs 肽段。例如,华大基因海洋

科学院(深圳海洋基因组学重点实验室)发表的工作,是基于APD数据库,联合使用基因组、转录组、蛋白质组、比较组学等多种组学技术,结合对AMPs肽段进行3D结构预测,从芋螺中筛选出8条具有抗真菌活性的芋螺抗菌肽^[61]。又如, Ebou等^[62]报告了基于蛋白质组和转录组技术从锥形蜗牛中筛选AMPs研究策略。即首先从转录水平对AMPs进行初步筛选,随后使用蛋白质组学对筛选

到的结果进一步印证,再运用机器学习模型预测候选肽段的活性,成功从锥形蜗牛毒液腺中识别出了具有抗菌活性的新型AMPs。

综上所述,多组学技术产生的数据可用于机器学习模型进一步分类和鉴别。人工智能算法针对候选AMPs的不同理化参数进行分类,最终筛选到具有高抑菌活性、高稳定性、高选择性、低细胞毒性的AMPs(图6)。

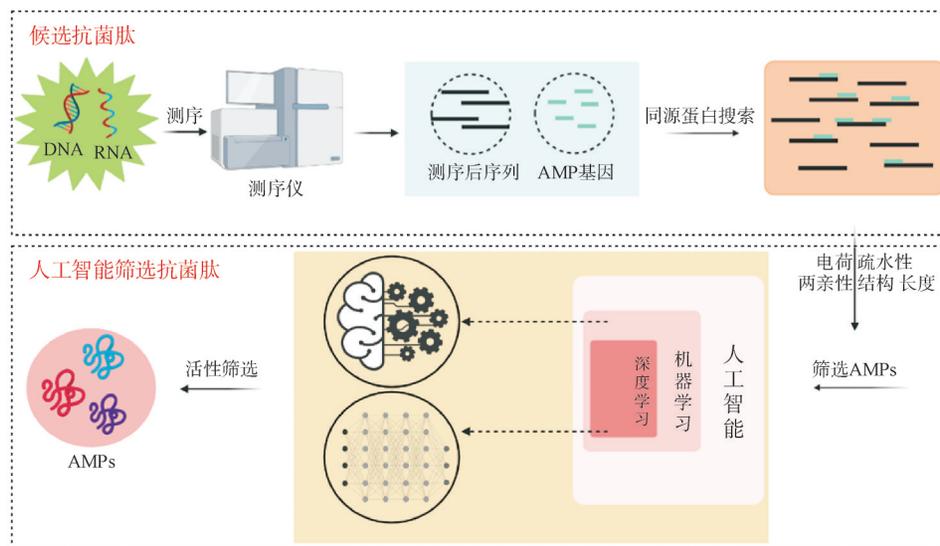


图6 人工智能技术筛选AMPs

4 结论与展望

抗菌肽具有广谱抑菌活性,且不易产生耐药性等特点,有望成为一种应用于临床的新型抗菌剂,解决抗生素耐药性危机。然而现有的抗菌肽存在红细胞溶血毒性、合成成本高等问题,亟待发掘新型抗菌肽。随着人工智能技术发展,精准、有效地筛选出高效抗菌肽成为可能。本文介绍人工智能方法在筛选抗菌肽领域中的应用,重点归纳了多种机器学习模型和深度学习算法在筛选新型抗菌肽中的应用,并且比较了不同预测模型的优缺点以及各自的特性,为该领域研究者提供见解和帮助。

目前对于已发现的抗菌肽的抗菌机制与抗菌肽的结构之间存在的关联性研究还不够深入,因此需要将计算机模型预测的结果与实验相结合。即将人工智能模型识别鉴定出的候选肽段与体外活性检测相结合,以期最终获得高效、优良的抗菌肽。另外,基于人工智能预测抗菌肽的方法还需

要进一步研究,如开发更有针对性的算法对多组学产生的数据进行选取、整合、处理,为高效率筛选抗菌肽提供更多新途径。

References

- [1] Wang GS. The antimicrobial peptide database provides a platform for decoding the design principles of naturally occurring antimicrobial peptides[J]. *Protein Sci*, 2020, **29**(1): 8-18.
- [2] Chaparro E, da Silva Junior PI. Lacrain: the first antimicrobial peptide from the body extract of the Brazilian centipede *Scolopendra viridicornis*[J]. *Int J Antimicrob Agents*, 2016, **48**(3): 277-285.
- [3] Jiao K, Gao J, Zhou T, et al. Isolation and purification of a novel antimicrobial peptide from *Porphyra yezoensis*[J]. *J Food Biochem*, 2019, **43**(7): e12864.
- [4] Wang JJ, Dou XJ, Song J, et al. Antimicrobial peptides: promising alternatives in the post feeding antibiotic era[J]. *Med Res Rev*, 2019, **39**(3): 831-859.
- [5] Passarini I, Rossiter S, Malkinson J, et al. *In silico* structural evaluation of short cationic antimicrobial peptides[J]. *Pharma-*

- ceutics*, 2018, **10**(3): 72.
- [6] Alsaggar M, Al-Hazabreh M, Al Tall Y, *et al.* HAZ, a novel peptide with broad-spectrum antibacterial activity[J]. *Saudi Pharm J*, 2022, **30**(11): 1652-1658.
- [7] Chen N, Jiang C. Antimicrobial peptides: structure, mechanism, and modification[J]. *Eur J Med Chem*, 2023, **255**: 115377.
- [8] da Costa JP, Cova M, Ferreira R, *et al.* Antimicrobial peptides: an alternative for innovative medicines[J]? *Appl Microbiol Biotechnol*, 2015, **99**(5): 2023-2040.
- [9] Czaplewski L, Bax R, Clokie M, *et al.* Alternatives to antibiotics—a pipeline portfolio review[J]. *Lancet Infect Dis*, 2016, **16**(2): 239-251.
- [10] Liu SC, Fan LL, Sun J, *et al.* Computational resources and tools for antimicrobial peptides[J]. *J Pept Sci*, 2017, **23**(1): 4-12.
- [11] Chen CH, Lu TK. Development and challenges of antimicrobial peptides for therapeutic applications[J]. *Antibiotics*, 2020, **9**(1): 24.
- [12] Wang GS, Li X, Wang Z. APD3: the antimicrobial peptide database as a tool for research and education[J]. *Nucleic Acids Res*, 2016, **44**(D1): D1087-D1093.
- [13] Partners CN MA. Database resources of the national genomics data center, China national center for bioinformatics in 2022[J]. *Nucleic Acids Res*, 2022, **50**(D1): D27-D38.
- [14] Jhong JH, Yao LT, Pang YX, *et al.* dbAMP 2.0: updated resource for antimicrobial peptides with an enhanced scanning method for genomic and proteomic data[J]. *Nucleic Acids Res*, 2022, **50**(D1): D460-D470.
- [15] Ye GZ, Wu HY, Huang JJ, *et al.* LAMP2: a major update of the database linking antimicrobial peptides[J]. *Database*, 2020, **2020**: baaa061.
- [16] Shi GB, Kang XY, Dong FY, *et al.* DRAMP 3.0: an enhanced comprehensive data repository of antimicrobial peptides[J]. *Nucleic Acids Res*, 2022, **50**(D1): D488-D496.
- [17] Singh S, Chaudhary K, Dhanda SK, *et al.* SATPdb: a database of structurally annotated therapeutic peptides[J]. *Nucleic Acids Res*, 2016, **44**(D1): D1119-D1126.
- [18] Pirtskhalava M, Amstrong AA, Grigolava M, *et al.* DBAASP v3: database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics[J]. *Nucleic Acids Res*, 2021, **49**(D1): D288-D297.
- [19] Waghugh FH, Barai RS, Gurung P, *et al.* CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides [J]. *Nucleic Acids Res*, 2016, **44**(D1): D1094-D1097.
- [20] Seshadri Sundararajan V, Gabere MN, Pretorius A, *et al.* DAMPD: a manually curated antimicrobial peptide database[J]. *Nucleic Acids Res*, 2012, **40**(Database issue): D1108-D1112.
- [21] Tyagi A, Tuknait A, Anand P, *et al.* CancerPPD: a database of anticancer peptides and proteins[J]. *Nucleic Acids Res*, 2015, **43**(Database issue): D837-D843.
- [22] Gautam A, Chaudhary K, Singh S, *et al.* Hemolytik: a database of experimentally determined hemolytic and non-hemolytic peptides[J]. *Nucleic Acids Res*, 2014, **42**(Database issue): D444-D449.
- [23] Qureshi A, Thakur N, Tandon H, *et al.* AVPdb: a database of experimentally validated antiviral peptides targeting medically important viruses[J]. *Nucleic Acids Res*, 2014, **42**(Database issue): D1147-D1153.
- [24] Piotto SP, Sessa L, Concilio S, *et al.* YADAMP: yet another database of antimicrobial peptides[J]. *Int J Antimicrob Agents*, 2012, **39**(4): 346-351.
- [25] Usmani SS, Kumar R, Kumar V, *et al.* AntiTbPdb: a knowledge-base of anti-tubercular peptides[J]. *Database*, 2018, **2018**: bay025.
- [26] Usmani SS, Bedi G, Samuel JS, *et al.* THPdb: database of FDA-approved peptide and protein therapeutics[J]. *PLoS One*, 2017, **12**(7): e0181748.
- [27] Gómez EA, Giraldo P, Orduz S. InverPep: a database of invertebrate antimicrobial peptides[J]. *J Glob Antimicrob Resist*, 2017, **8**: 13-17.
- [28] Hammami R, Ben Hamida J, Vergoten G, *et al.* PhytAMP: a database dedicated to antimicrobial plant peptides[J]. *Nucleic Acids Res*, 2009, **37**(Database issue): D963-D968.
- [29] Hammami R, Zouhir A, Ben Hamida J, *et al.* BACTIBASE: a new web-accessible database for bacteriocin characterization [J]. *BMC Microbiol*, 2007, **7**: 89.
- [30] Di Luca M, Maccari G, Maisetta G, *et al.* BaAMPs: the database of biofilm-active antimicrobial peptides[J]. *Biofouling*, 2015, **31**(2): 193-199.
- [31] Khabbaz H, Karimi-Jafari MH, Saboury AA, *et al.* Prediction of antimicrobial peptides toxicity based on their physico-chemical properties using machine learning techniques[J]. *BMC Bioinformatics*, 2021, **22**(1): 549.
- [32] Kavousi K, Bagheri M, Behrouzi S, *et al.* IAMPE: NMR-assisted computational prediction of antimicrobial peptides[J]. *J Chem Inf Model*, 2020, **60**(10): 4691-4701.
- [33] Xiao X, Wang P, Lin WZ, *et al.* iAMP-2L: a two-level multi-label classifier for identifying antimicrobial peptides and their functional types[J]. *Anal Biochem*, 2013, **436**(2): 168-177.
- [34] Müller KR, Mika S, Rätsch G, *et al.* An introduction to kernel-based learning algorithms[J]. *IEEE Trans Neural Netw*, 2001, **12**(2): 181-201.
- [35] Jaiswal M, Singh A, Kumar S. PTPAMP: prediction tool for plant-derived antimicrobial peptides[J]. *Amino Acids*, 2023, **55**(1): 1-17.
- [36] Lira F, Perez PS, Baranauskas JA, *et al.* Prediction of antimicrobial activity of synthetic peptides by a decision tree model[J]. *Appl Environ Microbiol*, 2013, **79**(10): 3156-3159.
- [37] Lv HW, Yan K, Guo YC, *et al.* AMPpred-EL: an effective antimicrobial peptide prediction model based on ensemble learning [J]. *Comput Biol Med*, 2022, **146**: 105577.
- [38] Exarchos KP, Exarchos TP, Papaloukas C, *et al.* Predicting peptide bond conformation using feature selection and the Naïve

- Bayes approach[J]. *Annu Int Conf IEEE Eng Med Biol Soc*, 2007, **2007**: 5009-5012.
- [39] Chen W, Luo LF. Classification of antimicrobial peptide using diversity measure with quadratic discriminant analysis[J]. *J Microbiol Methods*, 2009, **78**(1): 94-96.
- [40] Fjell CD, Hancock RE, Cherkasov A. AMPer: a database and an automated discovery tool for antimicrobial peptides[J]. *Bioinformatics*, 2007, **23**(9): 1148-1155.
- [41] de Jong A, van Heel AJ, Kok J, et al. BAGEL2: mining for bacteriocins in genomic data[J]. *Nucleic Acids Res*, 2010, **38**(Web Server issue): W647-W651.
- [42] Polanco C, Samaniego JL. Detection of selective cationic amphipathic antibacterial peptides by Hidden Markov models[J]. *Acta Biochim Pol*, 2009, **56**(1): 167-176.
- [43] Guo YC, Yan K, Lv HW, et al. PreTP-EL: prediction of therapeutic peptides based on ensemble learning[J]. *Brief Bioinform*, 2021, **22**(6): bbab358.
- [44] Bhadra P, Yan JL, Li JY, et al. AmPEP: sequence-based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest[J]. *Sci Rep*, 2018, **8**(1): 1697.
- [45] Jan A, Hayat M, Wedyan M, et al. Target-AMP: computational prediction of antimicrobial peptides by coupling sequential information with evolutionary profile[J]. *Comput Biol Med*, 2022, **151**(Pt A): 106311.
- [46] Lata S, Mishra NK, Raghava GP. AntiBP2: improved version of antibacterial peptide prediction[J]. *BMC Bioinformatics*, 2010, **11**(Suppl 1): S19.
- [47] Porto WF, Pires AS, Franco OL. CS-AMPPred: an updated SVM model for antimicrobial activity prediction in cysteine-stabilized peptides[J]. *PLoS One*, 2012, **7**(12): e51444.
- [48] Niarchou A, Alexandridou A, Athanasiadis E, et al. C-PAmP: large scale analysis and database construction containing high scoring computationally predicted antimicrobial peptides for all the available plant species[J]. *PLoS One*, 2013, **8**(11): e79728.
- [49] Rajkumar M, Bhukya SN, Ahalya N, et al. Impact of ANN in revealing of viral peptides[J]. *Biomed Res Int*, 2022, **2022**: 7760734.
- [50] Zhang HP, Saravanan KM, Wei YJ, et al. Deep learning-based bioactive therapeutic peptide generation and screening[J]. *J Chem Inf Model*, 2023, **63**(3): 835-845.
- [51] Wang HQ, Zhao J, Zhao H, et al. CL-ACP: a parallel combination of CNN and LSTM anticancer peptide recognition model[J]. *BMC Bioinformatics*, 2021, **22**(1): 512.
- [52] Xiao X, Shao YT, Cheng X, et al. iAMP-CA2L: a new CNN-BiLSTM-SVM classifier based on cellular automata image for identifying antimicrobial peptides and their functional types[J]. *Brief Bioinform*, 2021, **22**(6): bbab209.
- [53] Ma Y, Guo ZY, Xia BB, et al. Identification of antimicrobial peptides from the human gut microbiome using deep learning[J]. *Nat Biotechnol*, 2022, **40**(6): 921-931.
- [54] Wang C, Garlick S, Zloh M. Deep learning for novel antimicrobial peptide design[J]. *Biomolecules*, 2021, **11**(3): 471.
- [55] Dong B, Yi YH, Liang LF, et al. High throughput identification of antimicrobial peptides from fish gastrointestinal microbiota[J]. *Toxins*, 2017, **9**(9): 266.
- [56] Fingerhut LCHW, Miller DJ, Strugnell JM, et al. Ampir: an R package for fast genome-wide prediction of antimicrobial peptides[J]. *Bioinformatics*, 2021, **36**(21): 5262-5263.
- [57] Sharma R, Shrivastava S, Kumar Singh S, et al. AniAMPpred: artificial intelligence guided discovery of novel antimicrobial peptides in animal kingdom[J]. *Brief Bioinform*, 2021, **22**(6): bbab242.
- [58] Lee JH, Chung H, Shin YP, et al. Deciphering novel antimicrobial peptides from the transcriptome of *Papilio xuthus*[J]. *Insects*, 2020, **11**(11): 776.
- [59] Shelenvov AA, Slavokhotova AA, Odintsova TI. Cysmotif searcher pipeline for antimicrobial peptide identification in plant transcriptomes[J]. *Biochemistry*, 2018, **83**(11): 1424-1432.
- [60] Grafkaia EN, Polina NF, Babenko VV, et al. Discovery of novel antimicrobial peptides: a transcriptomic study of the sea *Anemone Cnidopus japonicus*[J]. *J Bioinform Comput Biol*, 2018, **16**(2): 1840006.
- [61] Li RH, Huang Y, Peng C, et al. High-throughput prediction and characterization of antimicrobial peptides from multi-omics datasets of Chinese tubular cone snail (*Conus betulinus*) [J]. *Front Mar Sci*, 2022, **9**: 1092731.
- [62] Ebou A, Koua D, Addablah A, et al. Combined proteotranscriptomic-based strategy to discover novel antimicrobial peptides from cone snails[J]. *Biomedicines*, 2021, **9**(4): 344.



[专家介绍] 李菁,教授,博导,无党派人士。江苏省“青蓝工程”中青年学术带头人,近5年以通讯作者或第一作者在 *Gut*, *Gastroenterology*, *Microbiome*, *Sci Adv* 等一流或学科顶尖期刊发表文章。现为 *Frontiers Microbiology* 副编辑,中国生物物理学会肠道菌群分会理事,中国营养学会江苏分子营养学会专委会委员等。主要研究方向为“微生态系统学”,即通过新一代DNA高通量测序技术从遗传多样性、环境适应机制、微生物结构等多方面探索人体微生态系统的基本机制。主持了3项国家自然科学基金面上项目,参与国家科技重大专项重大新药创制项目1项,参与“十三五”中国人民解放军重大研发计划1项,主持湖南省重点实验室开发项目等省级项目多项,主持横向研究项目多项。