AI重塑多糖构效关系研究范式：从经验驱动迈向数据驱动与功能导向设计

作者：writer 来源：科学网

本文原地址：https://www.iikx.com/news/progress/41564.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

AI重塑多糖构效关系研究范式：从经验驱动迈向数据驱动与功能导向设计。近日，河南大学联合APUS AI 实验室在Food Medicine Homology 刊发题 Artificial intelligence-driven paradigms in polysaccharide structure-activity relationships: a new revolution的观点文章。文章围绕人工智能在多糖结构—活性关系研究中的应用与发展趋势，系统梳理了多糖数据标准化、结构数字化表达、多模态谱学解析、构效关系建模、分子相互作用预测、生成式设计以及精准营养等关键方向，并进一步提出从标准化数据建设走向设计—制造—验证—学习闭环研究的新型发展路径，为人工智能与糖科学深度融合提供了系统性思路。

多糖广泛来源于植物、食用菌、藻类、微生物及动物资源，其分子量、单糖组成、糖苷键连接方式、支化程度、化学修饰和高级构象等结构因素共同影响溶解性、流变性质、分子组装、受体结合、肠道菌群利用及生物活性。与具有明确序列编码方式的蛋白质和核酸相比，多糖具有显著的结构异质性和层级复杂性，同一种来源的多糖也可能因品种、采收、提取、纯化和加工条件不同而表现出显著不同的结构和活性。

长期以来，多糖研究主要依赖提取—结构表征—活性验证的实验路线。面对大量相互耦合的结构变量，传统单因素分析和经验判断难以充分解析高度非线性的结构—活性关系，也难以进一步区分真正决定生物效应的关键结构因素与伴随变化因素。人工智能的发展为解决这一问题提供了新的研究工具，同时也推动多糖研究由局部数据分析向完整研究链条的数字化重构转变。

文章认为，人工智能赋能多糖科学的核心并非简单增加一种预测算法，而是推动研究模式实现三方面转变：由经验驱动向数据驱动转变，由孤立分析向全链条整合转变，由被动结构表征向功能导向理性设计转变。

1. 建立AI-ready多糖数据标准，解决人工智能研究的数据底座问题

文章指出，现阶段制约多糖人工智能研究的关键瓶颈首先不是算法复杂度，而是高质量、标准化和可比较数据的不足。

不同研究在原料来源、提取条件、结构检测方法、生物活性评价体系以及数据表达方式方面差异较大，使大量已有研究数据难以直接整合和复用。针对这一问题，文章提出面向AI建模的多糖最小信息记录框架，将生物来源、提取和加工过程、组成与纯度、一级结构、分子量及高级构象、

原始分析证据、生物活性评价、模型来源以及模型验证和预测不确定性等信息纳入统一记录体系。

这一框架旨在建立完整的样品—工艺—结构—检测—活性—模型信息链，使不同实验室、不同批次和不同来源的多糖数据能够被追踪、比较和重新分析，为人工智能模型的外部验证和跨研究应用提供基础。

2. 构建多糖数字化表达体系，让复杂糖链能够被人工智能识别

由于多糖缺少类似DNA或蛋白质那样天然、统一的线性序列表示方式，其数字化表达一直是人工智能应用中的基础难题。

文章总结了当前多糖结构数字化的多种策略，包括将单糖组成、糖苷键类型、支化程度和修饰程度转化为固定长度数值向量；将单糖视为节点、糖苷键视为边建立分子图结构；进一步构建同时描述单糖间拓扑关系和单糖内部原子化学环境的原子级异构图；以及直接将NMR、MS和IR等原始谱学数据输入深度学习模型。

这些数字化表示方式为机器学习、图神经网络及基础模型从复杂多糖结构中学习关键结构特征提供了基础。

3. 推动NMR、质谱和红外光谱由独立分析走向人工智能多模态联合解析

复杂多糖结构解析往往面临分离困难、谱峰重叠和人工归属工作量大等问题。文章提出，可利用多模态融合模型将NMR化学位移、质谱碎裂信息以及红外特征峰等不同类型的数据映射到统一计算空间，通过不同谱学技术之间的互补和交叉验证提升结构解析的准确性。

与此同时，生成式人工智能还可进一步建立反向谱学验证过程：模型首先提出候选多糖结构，再预测候选结构对应的理论谱图，并与实际实验数据进行比较，由此形成结构假设—理论预测—实验比较—结构修正的闭环，为提高复杂多糖结构解析效率提供新的方法学方向。

4. 构效关系研究由相关性预测进一步走向因果机制解析

人工智能可以同时处理分子量、单糖组成、糖苷键、支化度、构象和修饰等多维结构变量，从大量候选特征中寻找与功能密切相关的结构因素。

但文章强调，高预测准确率本身并不能证明结构与活性之间存在真正的因果关系。未来需要进一步引入可解释人工智能、因果表示学习以及反事实推断等技术，回答哪些结构是真正驱动功能变化的关键因素如果单独改变某一结构特征，活性是否随之变化等问题。

这一转变将推动多糖构效关系研究由传统统计相关走向更加可靠的机制解释，为结构优化和功能设计提供理论依据。

5. 生成式AI与合成生物学结合，有望实现按功能设计多糖

文章提出，未来多糖研究可能逐步突破传统先获得天然多糖，再评价功能的研究路线。

在生成式人工智能支持下，可以从免疫调节、益生元、靶向递送等目标功能出发，反向设计具有特定分子量区间、糖苷键组合、支化模式和侧链修饰的候选多糖结构。

合成生物学则可作为数字设计向真实样品转化的重要技术接口。人工智能首先生成候选结构，高通量实验对候选结构进行功能评价，机器学习再根据实验结果更新下一轮分子设计，由此逐步构建AI生成—生物制造—高通量筛选—模型更新的迭代闭环。

这一思路意味着，多糖研究未来可能由发现型研究进一步向设计型研究发展。

6. AI糖组学进一步连接功能食品、肠道菌群与精准营养

在应用层面，文章重点讨论了人工智能在功能食品、益生元和肠道靶向产品、医用多糖辅料与递送体系以及精准营养中的潜在价值。

其中，AI糖组学可将膳食多糖结构与个体肠道菌群、代谢组和宿主表型等数据进行联合分析，构建多糖结构—菌群响应—代谢产物—健康结局的预测模型，从而为个体化膳食方案和功能性益生元组合提供依据。

文章进一步提出 Digital Twin Gut（数字孪生肠道）的前沿设想，通过整合个体肠道菌群组成、代谢特征和宿主信息，在计算机中建立个体化肠道生态模型，并在实际营养干预之前模拟不同多糖可能产生的作用，实现营养配方的虚拟筛选和优化。

从数据标准走向智能闭环：多糖研究迎来新的发展阶段

对于未来发展，文章提出了分阶段的技术路线。近期首先需要推动多糖最小元数据标准、参考物质和可靠基准数据集建设；随后通过多中心数据实现不同实验室、不同仪器和不同生产批次之间的外部验证，并发展主动学习等实验—模型协同策略；更长期来看，有望形成连接功能目标设定、AI结构设计、化学或生物合成、结构确认以及功能验证的闭环研究平台。

文章指出，人工智能的真正价值并不仅在于加快已有实验过程，更重要的是帮助研究者从复杂数据中发现传统经验研究难以识别的新规律和新机制。但AI仍然是服务于科学问题的工具，其预测结果必须接受严格的实验验证和专业判断。

随着标准化数据库、可解释模型、因果推断、多模态分析、合成生物学和自动化实验平台不断发展，人工智能与多糖科学的深入融合有望进一步改变该领域的研究逻辑，推动功能多糖由经验筛选迈向可预测、可设计、可制造和精准应用的新阶段。

图1 AI糖组学驱动精准营养研究框架

图示以 AI Glycomics 为核心，将多糖结构、肠道菌群、宿主表型和数据库等多源信息进行整合，经人工智能和机器学习构建多糖结构—菌群响应—代谢输出模型，并进一步连接个性化营养方案、智能胶囊、纳米载体和水凝胶等配方设计，最终通过健康结局评价和反馈实现精准营养干预的闭环优化。

作者与项目信息

该论文共同第一作者为河南大学食用菌加工技术国家研发专业中心的陈什康及麒麟合盛网络技术股份有限公司APUS AI 实验室人员。

论文通讯作者为南京中医药大学江苏省中药资源产业化过程协同创新中心的段嘉伦副教授；及河南大学食用菌加工技术国家研发专业中心的康文艺教授。

该研究获得 中原领军人才（264200510055）项目支持。

原文链接：https://www.sciopen.com/article/10.26599/FMH.2026.9420165

来源：Food & Medicine Homology

更多 科学进展 请访问 https://www.iikx.com/news/progress/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发