

率,造成重要科学数据无法完整回传;如果错误判断通信窗口或资源配置条件,还可能影响航天器后续任务安排。

因此,未来深空信道推演不能只追求“预测得更准”,还需要回答另外2个问题:模型什么时候值得信任?当它遇到没有见过的情况时,能否及时意识到自己

的预测可能不可靠?

这意味着物理模型与人工智能的结合可能成为重要发展方向。一方面利用人工智能从复杂数据中发现传统模型难以描述的规律;另一方面利用轨道动力学、电磁传播和空间环境模型等物理知识,对人工智能预测结果进行约束和校验。

供的信息量有限。最后,也是最为关键的一点,通过实验测定的RNA结构在数量与多样性上,均比蛋白质结构少数个数量级,严重限制了可用于模型训练的数据规模。

在过去2022—2025年的4年间,学界开发了多种深度学习方法以应对上述挑战。Townley等试图测试这些模型在从头(de novo)RNA设计中的实际表现,即在缺乏天然序列模板的前提下,设计能够折叠成目标结构的RNA分子。作者通过征集“假结”(pseudoknots)的设计方案,进一步提升了测试难度。假结是一类包含相互锁定碱基配对区域的特殊RNA基序。由于其拓扑结构复杂且折叠规则尚未完全阐明,假结成为最难预测的RNA基序之一,且超出了大多数传统设计算法的能力范畴。鉴于假结是功能性RNA中反复出现的核心特征——包括构成核酶的催化中心及配体传感器的特异性结合口袋——实现对假结的精准预测与设计,对于深入理解并工程化改造RNA生物学机制至关重要。

Townley等发现,最新一代的深度学习的(人工智能,AI)系统,能够为57种不同测试的假结架构中的55种,成功设计出可正确折叠的序列。这一表现与经验丰富的人类研究者相当,后者也针对同一组假结结构提交了设计方案。值得注意的是,部分成功设计的假结展现出文献中未曾报道过的全新架构,这不仅凸显了该设计挑战的极高难度,也暗示现有结构数据库可能仅涵盖了潜在

破解RNA结构预测的“结”中之困

——人工智能赋能RNA假结设计

文/Anthony M. Mustoe¹, Junzhe Gu²

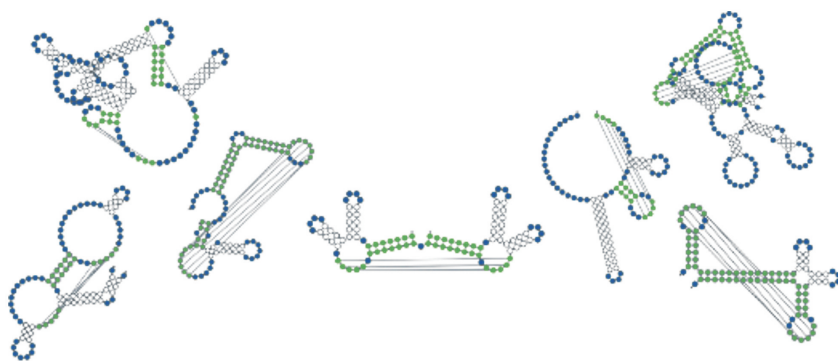
RNA分子在生物学中发挥着极其多样的作用:既可作为编码蛋白质合成指令的信使,也可作为响应配体的调控开关以控制基因表达,还能催化各类生化反应。这些功能的实现,通常取决于每种RNA分子所采用的特定折叠构象或折叠系综,涵盖其二级结构及更高阶的三维(3D)结构。从序列出发准确预测RNA结构,有望加速阐明RNA的作用机制,并推动合成RNA药物与器件的理性设计。在2026年第6814期《Science》上,Townley等人针对旨在实现这一目标的最先进深度学习方法,开展了一项重

要评估。该研究及其他近期发表和未发表成果均表明,尽管该领域正取得令人振奋的进展,但实现稳健的RNA结构预测仍面临重大瓶颈。

以AlphaFold为代表的蛋白质结构预测深度学习模型,已彻底变革了蛋白质结构生物学,但在RNA预测方面仍表现欠佳。为何RNA问题更为棘手?首先,RNA序列仅由4种化学性质相似的核苷酸组成,与构成蛋白质的20种氨基酸相比,其序列字母表提供的信息丰度较低。其次,给定的RNA序列通常可折叠成多种能量近乎相等的构象,这使得难以精准锁定具有生物学相关性的真实折叠状态。再次,蛋白质结构预测方法高度依赖从跨物种全面序列比对中提取的进化共变异信息;然而,大多数RNA的已知同源物较少,导致序列比对提

1. 美国得克萨斯州休斯敦,贝勒医学院生物化学与分子药理学系,治疗创新中心

2. 美国得克萨斯州休斯敦,贝勒医学院分子与人类遗传学系



基于RNA的技术与疗法的发展,长期受限于计算机设计与建模工具精度的不足

(图片来源:《Science》)

RNA折叠空间的一小部分。

该研究的一项关键发现是,当AI系统与人类研究者均能调用一种专门用于从序列预测RNA二级结构的深度学习模型时,设计成功率得到了显著提升。该模型名为RNet,由作者此前利用从化学探测实验中获取的大规模、核苷酸分辨率的RNA折叠模式语料库训练而成。RNet赋能人类与AI系统,使其能够有效甄别出倾向于形成目标假结二级结构的序列设计,并剔除易发生错误折叠的序列。值得注意的是,在引入RNet后,所有设计策略的表现均趋于一致,这表明RNet提供了决定性的先验信息。由于所测试的序列均为全新设计,且部分目标结构可能并未包含在RNet的初始训练集中,这些结果有力地表明,RNet已学习到可迁移的折叠模式,并能够泛化应用于各类RNA分子。

尽管成功设计假结是一项令人瞩目的成就,但需客观指出,若以单序列为基准进行评估,其整体成功率依然偏低,提示该技术

在规模化应用前仍需审慎验证。在通过化学探测评估的设计中,仅有约40%折叠成了所需的二级结构。当对部分设计通过诱变实验进行更严格的验证时,仅有约70%的序列令人信服地展现出目标假结结构。此外,当通过冷冻电子显微镜(cryo-EM)进行高分辨率结构解析时,7个设计中仅有3个显示出预期的空间构象。即使是经cryo-EM确认的设计,也表现出意想不到的结构特征,包括预期碱基对的断裂,并转而形成了未预测到的三级相互作用。外推这些结果表明,即使借助RNet,假结二级结构的预测准确率仍低于20%,而三维结构的预测准确率则更低。

目前尚不明确,合成RNA设计的进展是否能直接泛化至生物RNA的结构预测中。天然生物RNA通常序列更长,且受到复杂的进化约束,其结构塑造并非仅为优化单一、明确的构象。事实上,其他研究发现,与传统的基于热力学的模型相比,深度学习方法在预测长链生物RNA结构时

往往表现欠佳。此外,RNA功能通常高度依赖于结构动力学,且细胞内的RNA折叠往往具有显著的上下文依赖性。例如,共转录折叠、化学修饰与蛋白质结合等过程,会持续驱动并重塑RNA折叠的热力学景观。要全面阐明RNA的生物学机制并设计出具备复杂功能的RNA分子,必须超越单一静态结构的预测,转向对完整结构系综、其相对能量及其上下文依赖性进行动态建模。

分子生物学中一个反复显现的规律是,RNA遵循其独特的物理化学规则,需要不同于蛋白质的专属解决方案。例如,Townley等及2026年4~6月的其他研究进一步强化了这一经典认知:RNA二级结构在很大程度上可独立于三维结构进行建模,而精准的三维建模则严格依赖于准确的二级结构输入。将这一及其他RNA特异性折叠规则更充分地融入深度学习模型中,有望进一步提升模型性能。此外,通过化学探测实验以中等分辨率便捷地测量RNA结构,是RNA区别于蛋白质的又一显著特征。更高效地挖掘这一丰富的数据源,包括在建模过程中将探测数据作为空间约束条件,而不仅限于预训练阶段,代表了提升模型性能与实用价值的另一条有效路径。鉴于RNA生物学发现步伐的不断加快,以及RNA在生物技术领域应用的持续拓展,攻克RNA折叠预测难题从未像今天这样至关重要。

(译自《Science》,2026,393(6814))

(科技新闻栏目编辑 黄文光)