AI技术从零开始生成原始蛋白质

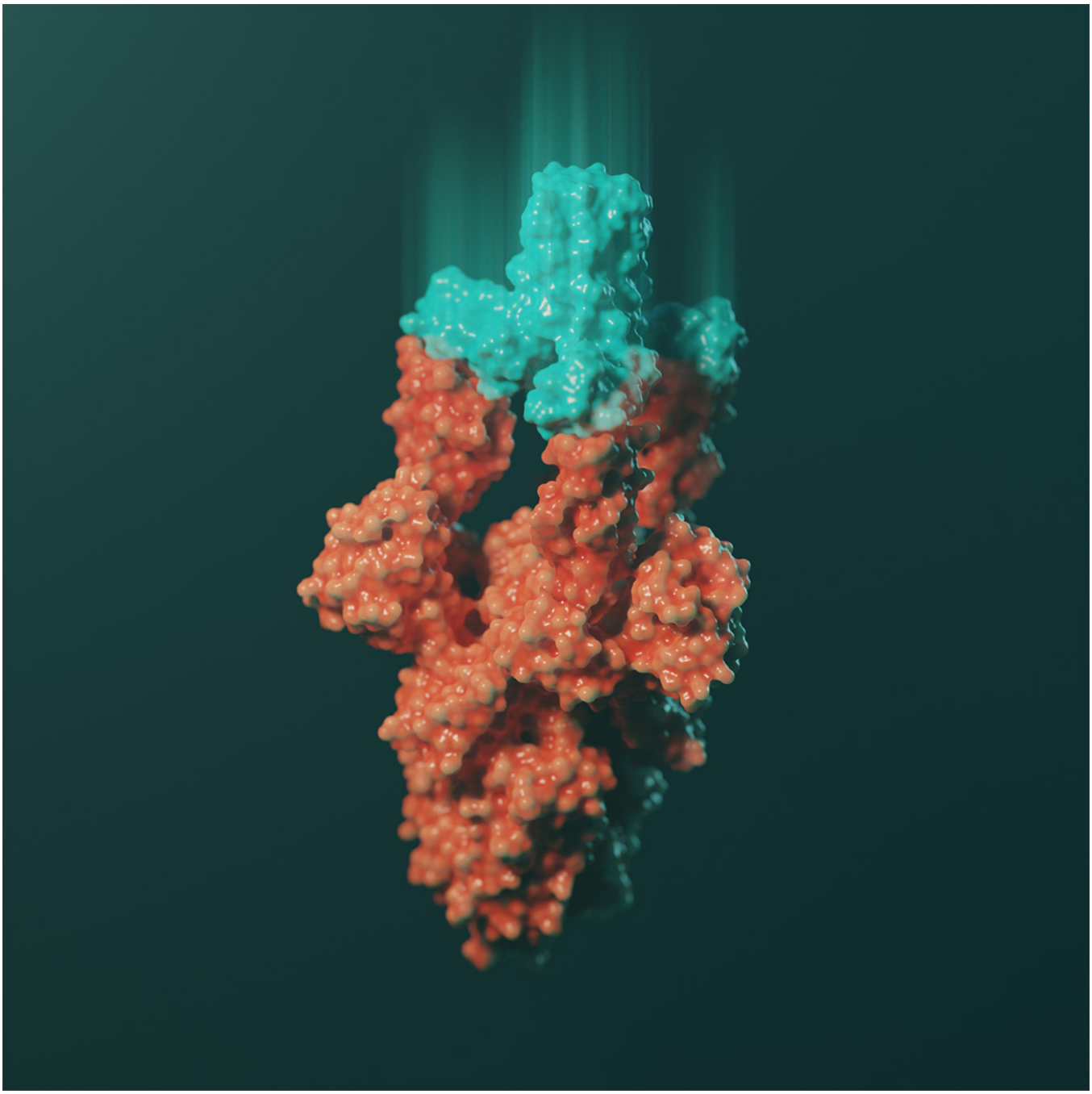
作者：writer 来源：科学网

本文原地址：https://www.iikx.com/news/progress/21815.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技术从零开始生成原始蛋白质。

科学家创建了一个能够从头开始生成人造酶的人工智能系统。在实验室测试中，其中一些酶与自然界中发现的酶一样有效，即使它们的人工生成的氨基酸序列与任何已知的天然蛋白质存在显著差异。相关研究结果发表于《自然—生物技术》。



图片来源：IAN C. HAYDON/UW INSTITUTE FOR PROTEIN DESIGN

该实验表明，自然语言处理虽然是为读写语言文本而开发的，但至少可以学习一些生物学的基本原理。Salesforce Research开发了名为ProGen的AI程序，该程序使用下一个标记预测将氨基酸序列组装成人造蛋白质。

科学家表示，这项新技术可能比获得诺贝尔奖的蛋白质设计技术定向进化更强大，它将加速新蛋白质的开发，为已有50年历史的蛋白质工程领域注入活力。这些新蛋白质几乎可以用于从治疗到降解塑料的任何领域。

人工设计的性能比受进化过程启发的设计好得多。该研究作者之一、加州大学旧金山分校药学院

生物工程和治疗科学教授James Fraser表示，语言模型正在学习进化的各个方面，但它不同于正常的进化过程。我们现在能够针对特定效果调整这些属性的生成。如一种非常热稳定或喜欢酸性环境或不会与其他蛋白质相互作用的酶。

为了创建这个模型，科学家们只需将2.8亿种不同蛋白质的氨基酸序列输入机器学习模型中，并让它消化信息几周。然后，他们通过使用来自五个溶菌酶家族的56,000个序列，以及有关这些蛋白质的一些上下文信息来对模型进行微调。

该模型迅速生成了100万个序列，研究团队根据它们与天然蛋白质序列的相似程度，以及AI蛋白质的潜在氨基酸语法和语义的自然程度，选择了100个进行测试。

在第一批由Tierra生物科学公司体外筛选的100种蛋白质中，该团队制作了5种人工蛋白质以用于细胞测试，并将其活性与鸡蛋清中发现的一种酶(鸡蛋清溶菌酶，HEWL)进行比较。在人类的眼泪、唾液和牛奶中也发现了类似的溶菌酶，它们可以抵御细菌和真菌。

其中两种人工酶能够分解细菌的细胞壁，其活性与HEWL相当。但它们的序列彼此只有约18%相同，这两个序列与任何已知蛋白质的同一性约为90%和70%。

天然蛋白质中只要有一个突变就能使其停止工作，但在另一轮筛选中，研究小组发现，即使只有31.4%的序列与任何已知的天然蛋白质相似，AI生成的酶仍显示出活性。

人工智能甚至能够通过简单地研究原始序列数据来了解酶应该如何形成。通过X射线晶体学测量发现，人造蛋白质的原子结构看起来和它们应该有的样子一样，尽管这些序列是前所未有的。

2020年，Salesforce
Research基于他们的研究人员最初开发用于生成英语文本的一种自然语言编程，开发了 ProGen。他们从之前的工作中了解到，人工智能系统可以自学语法和单词的含义，以及其他使写作井井有条的基本规则。

当你用大量数据训练基于序列的模型时，它们在学习结构和规则方面确实非常强大，能够了解哪些词可以同时出现，以及组合性。该研究通讯作者之一、Salesforce
Research人工智能研究主管Nikhil Naik说。

对于蛋白质，设计选择几乎是无限的。溶菌酶和蛋白质一样小，最多约有300个氨基酸。但是有20种可能的氨基酸，就有20300种可能的组合。这比有史以来所有人类的总和乘以地球上的沙粒数再乘以宇宙中的原子数还要多。

鉴于无限的可能性，该模型能够如此轻松地产生工作酶是非常了不起的。

该研究通讯作者之一、Profluent Bio创始人Ali Madani说：开箱即用的从头开始生成功能性蛋白质的能力表明，我们正在进入蛋白质设计的新时代。对于蛋白质工程师来说，这是可用的多功能新工具，我们期待看到它的治疗应用。(来源：中国科学报 辛雨)

相关论文信息：https://doi.org/10.1038/s41587-022-01618-2

作者：James Fraser 来源：《自然—生物技术》

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