深度学习实现蛋白质序列高成功率从头设计

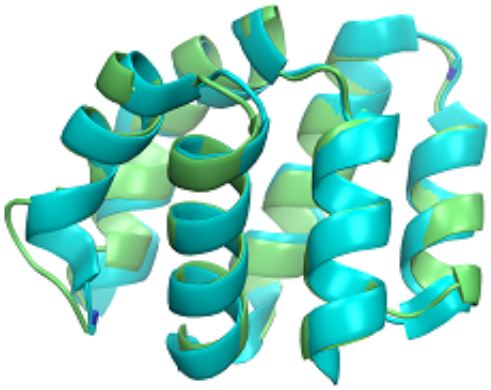
作者：writer 来源：爱科学

本文原地址：https://www.iikx.com/news/progress/19327.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！

深度学习实现蛋白质序列高成功率从头设计。

中国科学技术大学生命科学与医学部教授刘海燕、副教授陈泉团队与信息科学技术学院教授李厚强团队合作，开发了一种基于深度学习为给定主链结构从头设计氨基酸序列的算法ABACUS-R。经过实验验证，ABACUS-R的设计成功率和设计精度超过了原有统计能量模型ABACUS。研究成果北京时间7月21日发表于《自然-计算科学》。



一个来自天然蛋白的目标结构（天蓝色）与相应从头设计蛋白的晶体结构（绿色）叠合图
课题组供图

刘海燕、陈泉团队长期致力于发展数据驱动的蛋白质设计方法，先后建立并实验验证了对给定主链结构设计氨基酸序列的统计能量函数ABACUS模型，以及利用神经网络能量函数从头设计主链结构的SCUBA模型。然而，基于传统统计能量技术的ABACUS模型在成功率、计算效率等方面仍

有不足。

近期有多项研究表明，用深度学习进行氨基酸序列设计，能够在天然氨基酸残基类型恢复率等计算指标上超过能量函数方法。但截至目前已正式发表的工作中，对相关方法的实验验证结果远未达到能量函数方法的成功率。

刘海燕介绍，利用ABACUS-R进行序列设计的方法由两部分组成。

第一部分是一个多任务预训练的编码-解码器网络，用于对单个氨基酸的结构和化学环境进行隐空间编码，再解码为包括中心残基氨基酸类型在内的多种真实特征；第二部分是把该编码解码网络迭代应用于目标主链的每个氨基酸残基，直到获得最大程度自洽的全序列。

在理论验证的基础上，团队尝试用实验表征了ABACUS-R对3个天然主链结构重新设计的57条序列，其中86%的序列（49条）可溶表达并能折叠为稳定单体。实验解析的5个高分辨晶体结构与目标结构高度一致。此外，与此前报道的从头设计蛋白相似，ABACUS-R从头设计的蛋白表现出超高热稳定性，去折叠温度大多可达100℃以上。

总的来说，相较于ABACUS模型，ABACUS-R序列设计更高的成功率和结构精度进一步增强了数据驱动蛋白质从头设计方法的实用性。ABACUS-R还提供了一种对蛋白质局部结构信息的预训练表示方式，可用于序列设计以外的其他任务。

审稿人认为，该研究最新颖的贡献在于对设计的充分实验表征，包括晶体结构，以及可溶表达的高成功率。（来源：中国科学报王敏）

相关论文信息：https://doi.org/10.1038/s43588-022-00273-6

版权声明：凡本网注明来源：中国科学报、科学网、科学新闻杂志的所有作品，网站转载，请在正文上方注明来源和作者，且不得对内容作实质性改动；微信公众号、头条号等新媒体平台，转载请联系授权。邮箱：shouquan@stimes.cn。
作者：刘海燕等 来源：《自然—计算科学》

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