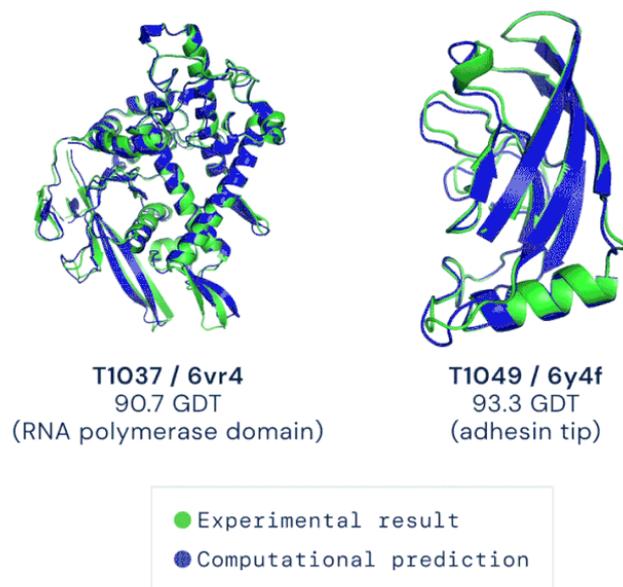


# AI 如何破解蛋白质结构之谜？

今年 11 月 30 日，DeepMind 公布了一项重量级成果，他们用 AI 预测蛋白质分子结构，比传统方法效率提高了数百倍。Nature 在当天的文章中激动地表示：“It will change Everything（所有事情将为之改变）”。

整个生物学界为之震惊。

这是个什么事呢？简单地说，蛋白质的功能是由其结构决定的，知道了结构即知道了蛋白质的功能。然而，已知的结构大概有  $10^{300}$  种，要确定一个蛋白质的结构非常困难。人们设计了核磁共振仪、X 射线、冷冻电镜等一系列极其昂贵的设备，耗费了大量时间和精力对蛋白质结构进行解析。例如，一种膜蛋白结构的检测花了科学家们 10 年。经过 60 年的艰苦奋斗，170,000 种蛋白质的结构已经被确定，但还有 200,000,000 种已知蛋白质等待检测，这是个让人绝望的数字。



**(这是一个动图)**

1972 年诺贝尔化学奖得主 Christian Anfinsen 提出了一个假设，认为**蛋白质结构可以由氨基酸序列来确定**。如果这是真的，那可真是天大的福利，因为氨基酸序列是容易确定的，只是不知道这些序列怎么左折右叠的形成了蛋白质。科学家们开始了信心满满的探索，希望用氨基酸序列来预测蛋白质的结构，但结果都不理想，达不到仪器测量的精度。怀疑情绪开始滋长：Anfinsen 的假设是对的吗？

2018 年，DeepMind 开始了他们的探索，在前人工作的基础上开发出了 AlphaFold 1，达到了当时最好的精度，但还不够。研究团队再接再厉，发现 AlphaFold 1 有些过训练的问题。通过改进建模方案，2020 年，AlphaFold2 横空出世，将预测误差一举降低到 1.6 埃，相当于一个原子的尺度。

Anfinsen 的理论被证明了，困扰人们 50 年的难题解决了，现在科学家们在电脑前输入一个氨基酸序列就可以得到一个蛋白质的结构了。有了这些结构信息，人们就可以深入了解这些蛋白质的特性，从而为生命科学的研究打开了一扇尘封已久的大门，这扇门里，是无穷无尽的宝藏。

而打开这扇门的最大功臣，是人工智能。