



NEW WORLD TIMES
全美首家中文简体字报纸
1997年9月19日创刊
www.newworldtimes.us

新世界时报

2024年
10月11日
第1414期
每星期五出版 本期56版
ISSN 1543-7930

诺贝尔化学奖再颁给AI,我们不再需要科学家了?



2024年诺贝尔化学奖刚刚揭晓,美国科学家大卫·贝克(David Baker)获得了一半奖金,表彰他在蛋白质结构设计领域所做的开创性贡献,另一半则共同授予英国科学家德米斯·哈萨比斯(Demis Hassabis)和美国科学家约翰·江珀(John Jumper),以表彰他俩在蛋白质结构预测方面所做的贡献,那就是如何从蛋白质的一维结构预测其三维结构。

科学家们很早就知道,蛋白质是生命体内最重要的化学物质,从身体结构到生化功能都是蛋白质在发挥作用。化学家们也早就知道蛋白质是由氨基酸首尾相连组成的一条线性分子,并很快就掌握了如何测量蛋白质氨基酸序列的方法。问题在于,生命是三维的,蛋白质要想发挥作用,必须先从一个一维的氨基酸链条转变成三维的立体结构。

一个蛋白质的三维结构应该是由它的一维结构,即氨基酸序列所决定的。上世纪60年代有人发现蛋白质完全不需要任何外力的帮助就可以在溶液里瞬间折叠成特定结构,整个过程用不了一秒钟的时间。

这个过程貌似神奇,但其原理倒是一点都不复杂。大家知道,氨基酸是由碳氢氧氮等原子组成的小分子,这些原子依靠化学键彼此相连,构成了或亲水、或疏水、或中性的各种基团。当蛋白质分子溶于水时,亲水基团很自然地会暴露在外,疏水基团则倾向于躲在内部。除此之外,氨基酸的各个基团有的带正电,有的带负电,彼此之间也会发生或吸引或排斥的反应。如果位置不当,蛋白质内部会有一种紧张感,只有当每个氨基酸都恰好坐落在最合适的位置上时,这个蛋白质才算安定下来,达到能量的最低状态。

蛋白质折叠的原理虽然很容易理解,但从一维结构推测三维结构却难上加难,因为蛋白质是由成百上千个氨基酸组成的。其中一部分氨基酸形成的多肽链被称为蛋白质的初级结构;之后,这些多肽链会先折叠成一些简单的三维结构,好像是螺旋线或者梯子,科学术语称之为二级结构;这些简单的二级结构再会叠加成更加复杂的三维结构,

这就是蛋白质的三级结构;最后,不同的蛋白质还会和一些化学元素组合成更加复杂的蛋白质复合物,这才是最终起作用的蛋白质四级结构。

科学家发现他们只能从氨基酸序列推测出蛋白质的二级结构,再往下就走不动了,因为实在是太过复杂。X射线晶体衍射技术的出现部分地解决了这个问题。简单来说,先把要研究的蛋白质进行提纯,然后让它合适的条件下结晶,之后就可以利用X射线的衍射作用对它进行拍照,从中推测出这个蛋白质的三维结构了。

在这个领域做得最好的人就是中国科学家颜宁,她的实验室是蛋白质三维结构预测领域的国际领先者,发表过无数论文。但是,这个技术需要首先获得蛋白质结晶,这并不是一件很容易的任务,需要耗费大量的时间和精力,成本也极其高昂。更何况很多蛋白质很难在自然条件下结晶,很多跨膜蛋白甚至不可能按照这个方法研究,所以说这个技术路线只能解决一部分问题。

有没有办法直接从氨基酸序列推测出蛋白质的三维结构呢?理论上是有可能的,但因为数据量太大,很多科学家前仆后继地投入这个领域,却均铩羽而归。

最先获得决定性突破的就是获得今年诺贝尔化学奖一半奖金的贝克,他从1993年进入这个领域,开发出了一系列能够根据氨基酸序列预测蛋白质结构的软件,取名为罗塞塔(Rosetta)。这套软件在国际蛋白质结构预测评估大赛(CASP)上保持了多年的领先优势,是这个领域最早的明星。

更可贵的是,贝克没有试图垄断这个具有很大盈利潜力的技术,而是建立了一个罗塞塔社区,让所有感兴趣的人都可以下载软件并参与开发,这种开源精神在当今这个追求金钱的时代显得极其难能可贵。(本文转自四川观察,不代表本报的观点和立场)

2018年的CASP大赛出现了一位神秘的参赛者,这就是大名鼎鼎的DeepMind团队,他们因为开发出了围棋程序阿法狗

(AlphaGo)而享誉全世界。成功之后,团队领导人哈萨比斯很快就把注意力转向了蛋白质结构预测这个生物学领域的珠穆朗玛峰,并雇佣了一批生物学家和计算机科学家一起组成了一个团队集体攻关,于2018年开发出了第一代基于人工智能技术的蛋白质结构预测软件AlphaFold,在2018年的CASP大赛上力压其他97个参赛者。

2020年,DeepMind公司又开发出了AlphaFold2,彻底碾压了此前所有的预测软件,其中就包括罗塞塔。而AlphaFold2团队的领导者就是和哈萨比斯分享另一半奖金的江珀。2021年,AlphaFold2已经能够成功预测98.5%的人类蛋白质结构,其准确性和实验结果相差无几。今年5月,DeepMind团队又推出了AlphaFold3,以前所未有的精确度预测了几乎所有的生命分子,包括蛋白质、DNA和RNA的三维结构,以及它们之间的相互作用模式。

除此之外,科学家们还发现了很多无序蛋白,它们在天然条件下没有确定的三维结构,但却具有相当广泛的功能。

总之,今年的3个理科诺奖其实都和复杂系统有关。获得医学或生理学奖的microRNA虽然本身属于经典的遗传学范畴,但它背后所代表的基因调控是一个庞大到无以复加的复杂系统,很难用简单的因果链条加以解释。获得物理学奖的神经网络其实就是这种复杂系统的一个计算机模拟程序,我们已经在人工智能领域看到了这个模拟程序的巨大潜力。化学奖同样体现了科学家试图利用计算机来模拟蛋白质复杂的三维结构的努力,同样在新药研发领域展现出了巨大潜力。

所以,与其说今年的诺贝尔奖是人工智能的胜利,不如说是复杂系统的胜利。科学研究的范式正在从过去的推导因果关系进化到研究因果关系不明的复杂系统,传统物理学在今天的缺失也从另一个侧面说明科学的范式转变已经来临。(本文转自四川观察,不代表本报的观点和立场)

华府地区中国大专院校校友会
联合会元气青年论坛 2024
——职场新起点-规划你的成功之路

>>详见 17 版

五省西北 多彩文艺
“五彩西北”大型文艺晚会
即将于10/26在JJC上演

>>详见 20 版

2024 CAPA-MC 全国
青少年领导力与创新力展示
大华府地区初赛即将举行

>>详见 23 版

低碳发展、大健康、数字经济
及跨国金融,解读美中经济新
增长极,探寻合作发展优质路径
——第五届华盛顿创新发展
与投资合作论坛成功举办

>>详见 34 版



WWW.newworldtimes.com