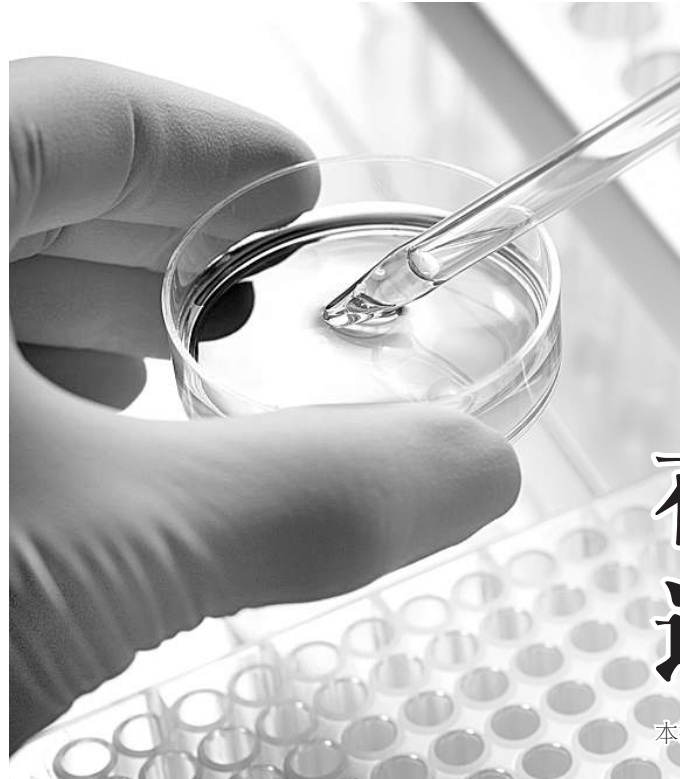


G 学科密码

G 科技前沿



三十年前,人们说:21世纪是生命科学的世纪。这表述中蕴含着一种即将掌握生命密码的豪情,也为人们展开一幅健康可控、寿命延长的美好画卷。然而,生命科学的进展似乎并没有达到人们的预期——并非生命科学进展缓慢,后基因组时代对生命密码的解析、基因编辑手段的应用、基因药物的问世……生命科学正发展成为多学科交叉的前沿科学,现代生物技术、新的基因技术、分子与细胞技术等应运而生,并被广泛地应用于医药卫生、农林畜牧、食品、化工和环境等领域,为解决全球共同面临的人口、健康、环境等诸多问题,增进人类福祉作出了重要贡献。然而,我们却发现,在理解生命这条道路上,还有很长的路要走。对人类来说,“生命”依然充满着迷茫和困惑。但科学家的脚步依然坚定:只有科学,才是解开生命奥秘的钥匙。

## 在理解生命这条路上,还有多长路要走

本报记者 杨 颀

### 探索,让人类健康、快乐且有质量地活着

北京大学未来技术学院教授刘颖说:“我对生命特别感兴趣,很好奇人体内部每时每刻到底在发生着什么,特别想了解清楚我们自己。”

刘颖主要做细胞生物学,研究人的代谢和衰老。“细胞是构成生命最基本的结构和功能单元。生命体太复杂,我们就先把细胞内部的事情研究清楚。研究细胞怎样感知它所处环境的变化,它能利用到的能量或营养物质水平的变化。我们吃饱了饭,蛋白质就会分解成氨基酸,细胞能感受到这些营养物质,然后很聪明地启动合成代谢,把我们过多的、剩余的营养物质储存起来。到了饥饿的时候,细胞感受到外界环境中的能量和物质不足,再把已经储存起来的物质分解掉,去提供能量。”

刘颖认为,研究细胞如何根据周围环境做适应性调整很关键。“很多癌细胞在营养物质有限的环境下,也能够不断地生长,造成细胞过多的增殖和分裂。所以在做相关研究的时候,我们发现一些基因跟癌症的发生密切相关。此外,细胞的代谢状态变化也跟衰老密切相关。人体进行适度的热量限制性进食,可以延缓衰老,这是非常有意思的问题。”

1865年,孟德尔根据豌豆杂交实验,发表遗传定律的假说,遗传学就此诞生。随后,假说和进化论结合,又催生了分子生物学、重组DNA技术。1990年,“人类基因组研究计划”启动。到2003年,科学家完成了对人类基因组30亿对碱基序列的全部测定工作,生命科学被认定将进入后基因组和蛋白质组学的时代。

人类用100多年的时间完成对基因组整体面貌的描述,改变了我们的生命观,让我们可以重新思考生命究竟是什么,看到更加清晰的全新生命图景。但人类对“生命”,在许多情况下仍然无能为力——从细胞研究到整个生命科学领域,对生命的认知,仍然是以有限的遗传信息为基础。科学家们跨越有限基因组的障碍,试图揭示出生生命无限发育和分化行为的原因。

“前段时间我看到一个视频,有人说现在生命科学学科的研究对生命的认知到了90%的程度。”清华大学生命科学学院教授杨茂君对此不敢认同:“依据我个人的理解以及科学研究的规律来看,目前生命科学学科对生命的认知依然处于起步阶段,无法定量。如果非要定量,我估计不到1%。知识就如同一个点,随着知识的积累,当这个点膨大为圆的时候,圆外面接触到的未知知识也就越多,未知的知识是无限的,而我们已知的知识却是有限的,以有限去比无限,我真的无法界定我们目前对生命认知到了什么程度。”

那么,生命科学学科能为我们做什么?杨茂君说,其实这个问题应该问,我们人类想要生命科学学科为我们做什么?当然是健康、快乐且有质量地活着。这些年生命科学的发展有目共睹,就拿我国人口平均寿命来说,1949年的时候是35岁,1957年的时候是57岁,1981年的时候是68岁,2010年的时候是75岁,2019年则上升到77.3岁。以前我们常说,人生七十古来稀,而现在70岁以上的老人比比皆是。这就是这些年我国生命科学飞速发展所带来的最直观的体现。”

本版图片均选自光明图片/视觉中国

### 挑战,对生命有机体运行机制的认识远远不足

科学研究,永远在探索“未知”,但往往在“已知”里面仍有许多“未知”。

在刘颖看来,在生命科学领域研究得越多,就更敬畏生命,“生命真的非常精妙,你想象不到它怎么能这么聪明,这么精妙地去调控每一个步骤。”

但“研究得越多,对生命了解越多,就发现自己越无知”,“一些约定俗成的概念或者现象,近几年在不断地被修正和完善。”

比如瓦伯格效应(Warburg effect),即癌细胞主要是通过糖酵解(分解葡萄糖)的方式来提供能量并产生乳酸。但这种提供能量的方式,产生的能量非常少。因为绝大多数正常细胞90%以上的能量是来自线粒体。

癌细胞为什么要走糖酵解这个途径?刘颖介绍,此前的解读认为癌细胞长成一大团肿瘤的时候,许多细胞挤在一起,尤其是肿瘤内部

中间的细胞,接触不到太多的氧气,所以没办法用线粒体来提供能量。

“随着近几年研究,科学家们逐渐发现,肿瘤内部会促进毛细血管的生成,氧气供应并不少。之所以选择糖酵解的方式,是因为这期间的代谢物会参与到其他生物大分子的合成,合成自己的DNA和蛋白质。有了这些物质,癌细胞才能从一个细胞分裂成两个细胞。”刘颖解释道。

“随着研究的不断深入,许多我们原本认为已经了解的生命科学现象其实在不断改写。未知的内容还有很多。尽管我们研究衰老,但衰老现象本质的原因是什么?我们能用什么样的方式延缓衰老?在人类层面上,这还是一个悬而未决、亟待解决的问题。伴随着人口老龄化的加剧,很多与衰老相关的疾病,像神经退行性疾病等,也是一个较大的社会问题。”刘颖说。

杨茂君坦承,生命科学领域目

前面临的巨大挑战是我们对生命有机体内在运行机制的认识还远远不足,“比如当今世界公认的五大绝症:运动神经元症(渐冻人症)、癌症、艾滋病、白血病、类风湿被世界卫生组织列为世界五大疑难杂症。随着科技的发展,目前白血病可以通过细胞免疫疗法或(和)骨髓移植来治愈,艾滋病也可以得到有效的控制。我相信,随着科学家进一步的研究,其他疾病也可能会在近期内取得重大进展。”

从学科发展来说,生命科学专业还面临着哪些新挑战?杨茂君认为,任何学科发展无外乎是两个维度,一个是往更深入、更细微处研究,比如分子层面、原子层面乃至量子层面的研究;而另一个则是向外延伸,比如生命科学与各学科的综合交叉。“生命科学专业所面临的挑战就是随着我们对未知的探索,我们所面临的未知越来越多。”杨茂君说。

### 开创,生命科学的爆发式突破

从“达尔文进化论”打破生命的神秘,到“细胞学说”揭示生物的共同性,“孟德尔定律”揭示生物遗传变化的规律。再到1953年沃森和克里克提出著名的DNA双螺旋结构模型,确定生命运动的“中心法则”,开辟20世纪分子生物学的新纪元。生命科学领域的每一次重大突破都有着里程碑的意义。

刘颖认为,每一次技术上的突破会相应地伴随着基础科学研究的突破。“几十年前显微镜成像的分辨率很低,顶多只能观察到细胞内部的结构,如果没有荧光的染料,根本看不到细胞内部发生了什么。但随着显微技术不断完善和突破,现在通过高分辨率的显微镜,可以真实地看到细胞内部各种细胞器每时每刻的运动,细胞彼此之间的融合等。每个细胞分裂的时候,甚至可以看到染色体每次复制分裂的整个过程。”

“随着技术上的突破,真正进入到一个多模态的、从微观到宏观到宏观的、对生命现象整体描绘和更深入理解的层面。我觉得21世纪的

生命科学,相比之前说不定会有爆发式或跳跃式的突破。”

“作为科研工作者,我们没办法预估怎样才能产生突破,只能靠自己一点点去摸索。每天一点点地去试验,甚至试错。在此过程中吸取经验教训,掌握新的信息,不断加深对生命科学问题的理解。这是从量变到质变的过程,当积累了一定程度的信息之后,可能就会有一些突破性的发现。”

“科学的发展,人类无法估量,但科学的魅力就在于它的不确定性。”刘颖说。

在杨茂君看来,由于目前自然科学各学科的相互交叉、影响、渗透以及各种新技术的不断涌现,未来生命科学的研究必将会继续在微观与宏观两个层面上同时进行,并且有相互融合的趋势。他说:“未来发展方向无非就是让人类对自身更加了解,比如脑科学;并且让我们具有更强的对抗各种疾病的能力,比如基因治疗、人工智能应用于药物研发等。”

“毫无疑问,生命科学肯定会再

开创一个新的时代,让21世纪真正成为生命科学的世纪。”杨茂君说。“此次夺去了数百万人生命的新冠肺炎疫情全球暴发,无疑会成为把生命科学的研究提高到国家战略层面的契机。我国前期对生命科学研究的巨大投入,比如863计划、国家重点研发计划,以及最近新建的四个国家实验室,都已经对国家未来生命科学的研究做了很好的布局,剩下的就需要坚定不移地继续推进了。”

作为一名科研工作者,杨茂君最直观的感受是,近些年来,尤其是2003年非典以后,国家明显加大了对生命科学的支持力度。“我国生命科学基础研究和应用研究这些年来取得了大量的丰硕成果,所以才能在面对此次新冠肺炎疫情时取得令世界瞩目的成就。从发展的角度来看,未来依然有各式各样、分门别类的疾病会威胁人类的生命健康。解决和突破的关键在于国家积极引导,让广大科技工作者能静下心来,做真正有意义的研究。”杨茂君最后说。

## 人工智能程序预测出98.5%的人类蛋白质结构

□ 张家伟

英国“深度思维”公司研究人员领衔的团队日前在英国《自然》杂志发表报告说,该公司的人工智能程序“阿尔法折叠”(AlphaFold)预测出98.5%的人类蛋白质结构,有助于深入理解一些关键生物学信息,从而更好地开展药物研发。

人类蛋白质组是指人类基因组编码所有蛋白质的集合,考虑到理解人类蛋白质组对健康和医药的重要性,研究人员一直以来付出大量努力来确定其中的蛋白质结构,但用普通实验方法预测蛋白质结构十分耗时。人类基因组中目前只有三分之一的蛋白质3D结构已通过实验确定。

“阿尔法折叠”是“深度思维”公司开发的一款人工智能程序,可用于预测蛋白质结构。该公司研究人员利用“阿尔法折叠”确定了覆盖几乎整个人类蛋白质组(98.5%的所有人类蛋白质)的蛋白质结构,并将这些结构放入公开的数据库免费供全球科研人员使用。

氨基酸是连接起来形成蛋白质的亚单位。研究人员还让“阿尔法折叠”对人类蛋白质组58%的氨基酸结构位置给出可信预测,其中对35.7%的结构位置预测达到很高可信度。

报告作者之一、“深度思维”公司联合创始人德米斯·哈萨比斯在一篇文章中说,了解一部机器的结构之后才能清楚知道它能做什么,因此深入分析蛋白质结构有助于我们理解它的功能。研究人员在寻找疾病治疗方案以及应对包括抗生素耐药性、微塑料污染和气候变化等人类社会面临的重大挑战过程中,也能从中受益。(据新华社)

### 新闻解读

新技术的应用给生命科学带来了新机遇。日前,英国《自然》杂志和美国《科学》杂志同时刊发人工智能软件预测蛋白质结构又快又精确的研究结果——两款新型人工智能系统被宣布能够精准预测蛋白质结构,它们分别是英国深度思维(Deep Mind)公司研发的阿尔法折叠2(Alpha Fold2),以及华盛顿大学蛋白设计研究所David Baker教授课题组的玫瑰折叠(RoseTTAFold)。

这很容易让人联想到阿尔法狗(Alpha Go)。阿尔法狗被人们所熟知,是因为其连续战胜多名世界顶级棋手,收获了公众的关注。也因此,战胜了人类智慧的阿尔法狗被视为人工智能的重要成果。实际上,相较于阿尔法狗,阿尔法折叠2和玫瑰折叠虽不出名,但其存在价值对人类而言更具意义。

阿尔法折叠2和玫瑰折叠作为一个AI软件,其最大作用是既快又准地测定蛋白质的形状,尤其是3D形状。我们知道,生命的本质是蛋白质,蛋白质又由氨基酸按一定顺序结合形成的多肽链组成,而且它们从一维到二维再到三维,以无数方式折叠成各种精致形状,方能完成各种功能和发挥重要作用。

人类蛋白质有成千上万种,其他物种的蛋白质更是多达几十亿种,包括细菌和病毒的蛋白质。过去,人类只能用低温电子显微镜(CryoEM)、核磁共振(MR)和X射线晶体学等手段来测定蛋白质结构,而且需要经过大量试错才能最终确定其结构。即便如此,有些蛋白质结构还测不出来,如在蛋白质数据库(PDB)中,有4种蛋白质无法用MR测定结构,包括牛磺甘氨酸N-乙酰基转移酶、细菌氧化还原酶、细菌表面层蛋白(SLP)和来自真细菌革菌属金孢子菌属的分泌蛋白。

蛋白质结构为何如此重要呢?原因之一是大量疾病与蛋白质的折叠形状有千丝万缕的联系。例如新冠病毒的棘突蛋白(S蛋白)的折叠形式决定了它入侵人体细胞的速度和致病力,普里昂蛋白的折叠形式决定了对包括人在内的哺乳动物的传染性海绵状脑病的致病力和致命性。遗憾的是,虽然人类已经测序人和其他物种的数10亿计的蛋白质的氨基酸序列,但截至目前,只有约10万个蛋白质的结构已经用实验方法得到了解析。

从理论上讲,一个蛋白质从一维到三维有无数种折叠方式。认识和精确测定蛋白质的构型既要耗费大量的时间和精力,同时也不一定测得准,也造成了药物、疫苗研发和疾病治疗的举步维艰。如现在的新冠肺炎,尽管有疫苗手段投入使用,必将极大地推动生命科学的发展。

(作者系科普作家)

## 新技术助力生命科学研究

□ 张田勤

今,还没有有效治疗新冠肺炎的药物,原因之一也是对病毒蛋白质结构认知不清。

既如此,让AI来帮助人们认识和精确测定蛋白质的结构就具有重大意义,并且非常实用。以阿尔法折叠2为例,它测定的大部分蛋白质的结构非常准确,不仅与实验方法测得的蛋白质结构的精确度相同,而且远超解析新蛋白质结构的其他方法。阿尔法折叠2测定的组成蛋白质主链骨架的叠加原子之间的距离中位数(95%的覆盖率)为0.96埃(0.096纳米),而人体细胞的速度和致病力,普里昂蛋白的折叠形式决定了对包括人在内的哺乳动物的传染性海绵状脑病的致病力和致命性。遗憾的是,虽然人类已经测序人和其他物种的数10亿计的蛋白质的氨基酸序列,但截至目前,只有约10万个蛋白质的结构已经用实验方法得到了解析。

从理论上讲,一个蛋白质从一维到三维有无数种折叠方式。认识和精确测定蛋白质的构型既要耗费大量的时间和精力,同时也不一定测得准,也造成了药物、疫苗研发和疾病治疗的举步维艰。如现在的新冠肺炎,尽管有疫苗手段投入使用,必将极大地推动生命科学的发展。

(作者系科普作家)