

不同于液流电池, Form Energy 公司的铁—空气电池是一种静态电池,其基本原理是基于铁的可逆氧化(生锈),可持续多达 10 000 次的循环。相比于铁基液流电池,铁—空气电池的储能容量更大,其可储存电能长达 100 小时(约可为电网提供超过 4 天的电力),这种电池将使具有成本效益的“多日储能”成为可能。上述两种铁基电池在大规模储能方面均具有明显的优势:超长循环寿命、高安全稳定性、可扩展性、低成本和绿色环保,可平衡可再生能源发电的波动式变化,实现低碳长时电网储能。

铁基长时电网储能电池的发展,可以弥补锂离子电池的一些不足,以科技创新的方式将电力系统从化石燃料发电转变为可再生能源发电,有利于在全球范围内减少碳排放,实现低碳电网的发展和碳中和的终极目标。然而,除了长时电网储能电池外,还有一些其他可以提供稳定电力服务的能源组合(核能、化石能源+碳捕捉与封存技术、氢能等)与之竞争,这些技术未来的发展,也会在一定程度上左右长时储能电池在电网中的占比。此外,与其他储能技术的发展一样,长时储能电池从研发、示范、落地到规模化,一路上必将面临产能、供应链、建设、运营等多方面的挑战,必须严格控制每一环的风险,才能实现既定的成本目标。

我国的长时电网储能技术以全钒液流电池为主,其已经经过十多年的示范考核,并且其大规模储能的工程效果已得到了充分的验证,产业配套成熟,可支撑起百兆瓦级储能项目的设计与开发。此外,全钒液流电池系统的单瓦时成本已可控制在 2~3 元的水平,具备了商业化应用的条件。2018 年以来,我国液流电池的装机量呈现爆发式增长。其中,2020 年规划的液流电池装机量超过 6 GW,容量超过 20 GWh。与此同时,单个项目的规模也在不断提升,如 200 MW/800 MWh 的全钒液流电池示范项目。整体而言,我国液流电池的产业研究和技术工艺处于国际领先水平,特别是国内液流电池的龙头企业,大连融科在海外市场的拓展也在如火如荼地进行。然而,全钒液流电池的低能量密度和钒高昂的价格,需要我们开发更具价格和能量密度优势的新型长时电网储能技术。

储能作为“双碳”背景下构建低碳电网的关键组成部分,跨天、跨月乃至跨季节的长时电网储能系统的发展迫在眉睫。目前长时储能技术仍处于百家争鸣的中早期研发示范阶段,孰胜孰劣尚未揭晓。电化学储能由于动力电池产业的推动,不受地理环境

的制约,暂时处于比较有利的竞争地位。未来电网储能系统的发展需要以模型数据开源、学术产业结合等方式集思广益,甄选出最具经济可靠性的电源储能配置方案,形成多能互补的,新能源+储能的电力系统,为实现“双碳”目标提供强有力的支撑。



图 3 廉价、储能持久的铁基电池有望分摊可再生能源的供应压力,并扩大清洁能源的使用范围  
(图片来源:MIT Technology Review 官网)

DOI:10.16262/j.cnki.1000-8217.2022.03.025

#### 4 AI 蛋白质折叠 (Artificial intelligence for protein folding)

作为生命体最重要的功能载体之一,蛋白质在众多生命活动中发挥着关键的作用。蛋白质在行使功能时往往需要折叠成特定的三维结构,因此对蛋白质结构的测定和解析不仅能帮助人们在分子层面上理解大多数生命活动的机理,而且可以有效辅助基于结构的药物开发以及相关疾病的诊治。目前通过实验手段解析蛋白质结构费时费力,远远无法满足现实需求。2020 年底,谷歌重组后的“伞形公司”Alphabet 旗下名为 DeepMind 的人工智能实验室采用多种深度学习技术,开发出了一款名为 AlphaFold2 的软件,能根据蛋白质的氨基酸序列准确预测其三维结构。该软件使用一种称为深度学习的人工智能技术,可以预测蛋白质的形状,甚至精确到原子。由于大多数蛋白质的氨基酸序列已知,该软件可以在数小时内提供目标蛋白质原子分辨率的结构信息,而且其预测的结构模型准确度很高,在很多蛋白上可以与实验解析的真实结果媲美。世界各地的科学团队已经开始使用它来研究癌症、抗生素抗性和新冠病毒。2022 年,该技术被 MIT Technology Review 评选为“全球十大突破技术”之一。

## 专家点评:



龚海鹏 清华大学生命学院副教授, 博士生导师, 生物信息学教育部重点实验室副主任。2009 年加入清华大学生命学院, 主要从事蛋白质结构相关的计算方法研究。近年来的研究兴趣主要集中在结合人工智能技术发展蛋白质结构预测算法和分子模拟的采样方法, 以通讯作者身份在 *Nature Machine Intelligence*、*Advanced*

*Science*、*Bioinformatics*、*PLoS Computational Biology*、*Journal of Chemical Theory and Computation* 等计算生物学主流期刊上发表多篇论文。近 5 年主持国家自然科学基金项目 3 项。

生命体中的主要生命活动都通过蛋白质分子完成, 因此理解单个蛋白质分子的工作机理至关重要。蛋白质的多肽链是由氨基酸顺序连接而成的线性分子, 它往往折叠成特定的三维结构来行使功能。换言之, 蛋白质的序列决定结构, 而结构又决定功能。自上世纪五六十年代起, 蛋白质序列、结构与功能间的关系就一直是生命科学的核心问题。作为这一信息链条的中心点, 蛋白质结构既可以帮助人们理解生命活动的分子机理, 也能有效地辅助蛋白质设计和基于结构的药物设计, 因而结构解析已经成为生物物理领域最重要的研究方向之一。过去二三十年来, 结构生物学取得了长足进展, 包括蛋白质晶体学和冷冻电镜等技术的快速发展, 使得人们可以较为快速地测定生物大分子的三维结构。目前蛋白质结构数据库 (Protein Data Bank, PDB) 中已经积累了超过 18 万个分子的结构。但是, 总体而言, 蛋白质结构的实验测定仍然较为耗时, 往往至少要耗费数月时间。此外, 由于新一代测序技术的发展, 蛋白质序列的积累速度远远大于结构解析的速度。目前蛋白质序列库中的蛋白质数目已经超过结构数据库 3~4 个数量级, 这一差距无法通过实验方法弥补。

根据安芬森法则 (Anfinsen's dogma), 大多数球状蛋白的三维结构由氨基酸序列唯一决定。自 20 世纪八九十年代起, 人们就开始发展计算机算法, 通过研究序列和结构间的关系, 根据氨基酸序列预测蛋白质的三维结构。1994 年, 约翰·莫尔特 (John Moult) 等人组织了第一届国际蛋白质结构预测评估竞赛 (Critical Assessment of protein Structure Prediction, CASP), 用于系统评测各种计算方法的预测准确性。该竞赛每两年举办一次, 组委会收集未发表的结构数据, 对参赛者发布其序列信息, 然后收集其预测结果进行双盲评估。CASP 竞赛极大地促进了蛋白质结构预测领域的发展。在早期 CASP 竞赛中, 发展的基于模板的建模方法 Modeller 以及

基于统计和物理模型的建模方法 Rosetta 和 I-TASSER 等程序, 结合物理知识和对结构数据库的统计分析, 可以对某些特定蛋白提供较为准确的预测模型。但是, 随着实验解析蛋白质数目的快速积累, 这些方法的预测性能并未显示出相应提升, 反而达到了瓶颈。2015 年, 克里斯·桑德斯 (Chris Sanders) 等人提出可以从多重序列比对中获得氨基酸残基间的共进化关系, 从而为结构预测提供额外信息。2016 年的 CASP12 竞赛中, 许锦波等人提出的 RaptorX 程序, 首次使用深度卷积模型, 根据多重序列比对预测氨基酸残基间接触, 再根据预测结果折叠蛋白, 从而显著提升了结构预测的平均准确率。其后, 人工智能方法开始广泛介入蛋白质结构预测领域。2018 年的 CASP13 竞赛中, DeepMind 发展的 AlphaFold 采用了类似的方案预测残基间距离并根据预测距离折叠蛋白。其后的大多数方法也主要沿这一思路进行。2020 年 CASP14 竞赛前, 人们发现这类方法的性能并不能随模型参数量增加而继续提高, 而且这类深度学习模型的预测准确性离现实需求尚有一段难以跨越的距离。但是, 在 2020 年底 CASP14 结果公布时, DeepMind 提出的 AlphaFold2 算法远远超越了其他深度学习模型, 对绝大多数目标蛋白都可以提供高度精准的预测模型。对有些蛋白质而言, AlphaFold2 预测的结果与实验解析的模型高度相似, 甚至仅根据实验数据都很难区分孰优孰劣。这一结果也震惊了整个科学界。

后续的报告和论文显示, DeepMind 在设计 AlphaFold2 时完全摒弃了第一代 AlphaFold 的架构, 规避了残基间接触或距离的预测, 采用了一种全新的端对端模型直接根据序列预测结构。这一设计不仅能加快预测速度, 而且可以有效抑制中间过程中的误差积累。此外, DeepMind 采用了近年来自然语言处理领域较流行的 Transformer 架构。与以前常用的卷积架构不同, 这种基于注意力机制的模型允许所有氨基酸残基在每一步操作中发生信息交互, 能更好地模拟蛋白质折叠过程中的残基间相互作用。最后, AlphaFold2 还使用了重循环和自蒸馏等工程技术, 进一步有效提升了预测准确率。总之, 在 AlphaFold2 中, DeepMind 通过具有高度创新性的深度学习网络架构设计, 初步解决了蛋白质结构预测 (或折叠) 这一困扰人类 50 年之久的科学难题, 也因此入选 MIT Technology Review 评选的 2022 年“全球十大突破性技术”。

AlphaFold2 的提出显然对生命科学有巨大的



促进作用。一方面,它能够快速准确地根据氨基酸序列预测蛋白质的三维结构,因此可以有效弥补现有结构生物学技术的缺陷。另一方面,通过对 AlphaFold2 的进一步研究,人们可以更好地理解蛋白质序列和结构间的映射关系。目前已有许多研究组开始把 AlphaFold2 应用于药物开发和蛋白质设计领域。比如,清华大学的张林琦和彭健合作,通过进一步优化的 AlphaFold2 模型预测抗原和抗体的结合能,并根据预测结果优化抗体设计,最终开发出我国首款新冠特效药。2021 年,DeepMind 与 EBI 合作,建立了基于 AlphaFold2 预测结果的数据库 AlphaFold DB。该数据库中已经储备了近一百万蛋白质的预测结构,为生命科学各个领域的科学家们提供重要的蛋白质结构信息。这一行为很可能会改变很多领域的科研范式,促进分子层面的研究从以序列为基础转变为基于序列和结构的研究,从而加快生命科学量化的步伐。

当然,AlphaFold2 还有一定的局限性。首先,它在很多蛋白上的预测精度还有待进一步提高,目前还不能完全满足药物开发等领域需要的结构精度,特别是对蛋白质复合体的预测精度较低。因此,AlphaFold2 还不能完全替代结构生物学研究。其次,AlphaFold2 模拟的是从多重序列比对到三维结构的映射关系,并没有解决从单一序列到三维结构的映射关系,因此蛋白质折叠问题还没有完美解决。最后,也是最重要的一点,针对一个特定的目标蛋白,AlphaFold2 仅提供有限的结构模型,不能揭示其结构的动态变化。而动态结构才是真正决定功能的基础。

无论如何,AlphaFold2 的提出展示了人工智能技术对生命科学研究的巨大促进作用。在蛋白质结构预测领域,预计人工智能技术将继续引领后续的进展,解决目前 AlphaFold2 的局限性:(1) 提高蛋白质复合体的结构预测精度;(2) 发展根据单一序列预测蛋白质结构的人工智能算法;(3) 根据氨基酸序列预测蛋白质的动态结构。

我国在蛋白质结构预测领域的基础整体上还比较薄弱。近年来虽然有多多个学术研究组在残基间距离预测、能量函数构建和模型质量评估等子领域做出了原创性的工作,但是由于学术研究组的规模和资金有限,还没有形成完整的自主研发的程序算法,能达到与 AlphaFold2 持平的程度。但是,目前已有企业(如深势科技、华深智药、腾讯、百度、华为等)牵头的研发小组以 AlphaFold2 为模板进行二次开发,

并取得了一定的成绩。考虑到人工智能技术的飞速发展,我国在这一领域还有极大的潜力。希望在不久的将来,我国能通过多个学术研究组的联合攻关,或企业和高校的合作研发模式,开发出原创性的、具有完全自主知识产权的蛋白质结构精准预测算法。

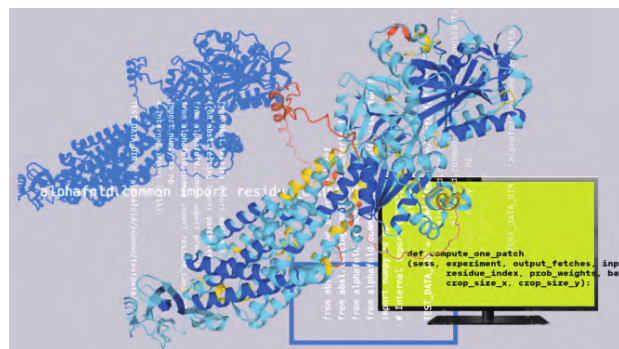


图 4 人工智能实验室 DeepMind 开发的 AlphaFold2 软件初步解决了蛋白质结构预测这一科学难题  
(图片来源:MIT Technology Review 官网)

## 5 疟疾疫苗 (Malaria vaccine)

寄生虫是复杂的多细胞生物,其基因组比大多数病毒和细菌中的基因组大 500~1 000 倍。使其能够通过无数种方式的基因突变来躲避人体免疫系统的监视。疟疾是疟原虫所引起的严重危害人类生命健康的寄生虫病。其主要集中在撒哈拉以南的非洲地区,该地区的病例约占全世界的 95%。每年有 60 多万人死于疟疾,其中大部分是 5 岁以下的儿童。2021 年 10 月,世界卫生组织批准了世界上第一种对抗由蚊子传播的致命疾病的疫苗——疟疾疫苗。然而,这款由葛兰素史克公司研发的疟疾疫苗,RTS,S 或 Mosquirix,被认为不是一种特别有效的疫苗。它需要在 5 至 17 个月大的儿童中接种三剂,并在 12 至 15 个月后再接种第四剂。此外,在肯尼亚、马拉维和加纳的 80 多万名儿童中,这种疫苗在第一年对严重疟疾的有效率约为 50%,而且随着时间的推移,其疗效急剧下降。即便如此,公共卫生官员仍将这种自 1987 年就开始测试的疫苗誉为非洲的“游戏改变者”,主要原因是当其与其他疟疾控制措施(包括驱虫蚊帐和在雨季使用的预防药物)结合使用时,有望将疟疾死亡人数减少多达 70%。Mosquirix 作为第一个被批准用于寄生虫病的疫苗,旨在敲响免疫系统的警钟,保护潜在的宿主免受感染,对鼓励创新以及下一代疟疾疫苗的开发具有重大意义。