



Chen Institute & Science
Prize for
**AI Accelerated
Research**

2024 AI 驱动科学大奖

Dr. Zhuoran Qiao
大奖得主

用“计算显微镜”观察蛋白质

大奖得主 *Zhuoran Qiao* 希望以生成式 AI 彻底革新药物研发

有时候，即便是获奖科学家也会在实验室里栽跟头。

在北京大学读本科时，Zhuoran Qiao 梦想着专攻全合成，利用复杂的实验室方法合成有机化合物。但他很快被现实“叫醒”了。“这就像做饭不好吃一样——我在实验操作上真的不太在行。”Zhuoran Qiao 笑着说，“我总是把实验搞砸，还把自己弄伤。”几次烫伤手指和实验失败后，Zhuoran Qiao 决定转向其他方向。“结果发现，我在用计算机模拟化学反应方面还挺有天赋，”他说，“在计算机上‘做实验’对我来说顺利多了。”

这其实还是谦虚了。今年春天，Zhuoran Qiao 凭借利用生成式 AI 模拟蛋白质结构的研究，成为首届“天桥脑科学研究院与《科学》杂志 AI 驱动科学大奖”的得主。他的突破性成果有望彻底改变制药领域。

“Zhuoran Qiao 的工作令人震撼地展示了 AI 加速科学发现的力量——多亏了他的努力，过去需要耗费数月甚至数年繁重劳动的研究，现在几秒钟就能完成，”天桥脑科学研究院创始人雒芊芊表示，“他让研究人员能够快速设计新分子，为那些曾被认为‘无药可治’的疾病带来了全新的治疗路径。”

科学发现之路

在中国学习计算化学后，Zhuoran Qiao 于 2018 年前往加州理工学院，在 Anima Anandkumar 教授和 Thomas Miller 教授（一位先驱化学家和 AI 创新者）的实验室攻读博士学位。在那里，Zhuoran Qiao 开始利用机器学习，打通量子数据与分子结构高层描述之间的桥梁。“用 AI，我们能获得以往需要极高计算量才能得到的结构洞见，”Zhuoran Qiao 说。

2021 年，正值 Zhuoran Qiao 不断完善自己的模型时，谷歌 DeepMind 发布了 AlphaFold——一个能根据氨基酸序列预测复杂蛋白质结构的 AI 模型，进一步凸显了计算模拟的潜力。AlphaFold 的能力让 Zhuoran Qiao 惊叹不已：蛋白质的折叠方式比宇宙中的原子还多，但 AlphaFold 却能极其准确地快速预测蛋白质结构。“这是个了不起的突破，”Zhuoran Qiao 说，“但我也看到还有两点不足。”

首先，AlphaFold2 只研究孤立的蛋白质，而药物研发需要模拟蛋白质与小分子的相互作用。其次，AlphaFold2 给出的是静态快照，但蛋白质在体内的生化环境中始终处于动态变化之中。“要理解蛋白质的功能，就必须捕捉其全部动态过程，”Zhuoran Qiao 解释道。

就在这时，Zhuoran Qiao 灵机一动：为何不用扩散模型（diffusion model）？这类模型后来也用于流行的生成式 AI 图像工具。正如米开朗基罗雕刻大卫像时，一点点剔除多余的石头，扩散模型则是逐步消除目标之外的噪音。比如，输入“戴高帽的猫”的提示，模型会从一堆随机像素出发，结合关于猫和帽子的训练数据，逐步逼近目标图像。

Zhuoran Qiao 意识到，同样的方法可以用于模拟蛋白质的复杂动态分子景观。“你从所有可能原子位置的随机混合出发，然后逐步去噪，最终得到三维结构，”他解释道。这里不再是文本提示，而是生化数据：蛋白质序列的进化特征、小分子的化学图谱，以及预测小分子与蛋白质如何相互作用的“接触图”。“利用这些，扩散模型可以映射每一个原子，最终还原分子结构，” Zhuoran Qiao 说。

通俗来说，这意味着 Zhuoran Qiao 的预测引擎不仅能预测蛋白质如何折叠，还能预测其与其他分子（包括潜在药物）如何相互作用。

构建“计算显微镜”

在 AI 出现之前，模拟一个蛋白质是个极其庞杂的任务，要么依赖艰苦的晶体学或光谱成像，要么靠逐原子计算机模拟。“过去大家用蛮力解决这个问题，但模拟一个蛋白质要花费数年 GPU 时间，而且还要人工设置参数，”Zhuoran Qiao 解释道。而他的生成式模型只需几秒钟就能准确还原复杂分子体系。“速度提升了一百万倍，结果往往与实验数据非常接近，”他说。

实际上，这个预测引擎为研究者提供了一个“计算显微镜”，让他们无需繁琐计算和实验，就能洞悉复杂生化系统的运作。它甚至可能帮助科学家突破传统成像的极限，探索那些尚未被实验观察到的“暗蛋白”。“我们的模型可以从已知蛋白质中泛化，帮助我们跨越这一障碍，”Zhuoran Qiao 说，“这对新药开发至关重要。”

初步研究会集中在设计针对已知蛋白质的小分子。“如果你已经有关于信号通路的生物学假说，可以用我们的工具找到能触发该效果的分子，”Zhuoran Qiao 说，“这依然不是个简单问题，但我们的引擎显著加速了搜索过程。”更具雄心的是，这个引擎还能用于从零设计蛋白质，比如开发新型蛋白质传感器，甚至创建一连串生物工程蛋白质相互作用，解锁新的生物功能。有前景的应用之一，是利用“分子胶”让致病蛋白与专门降解生化物质的蛋白结合。“通过模拟这些复杂动

态，我们或许能主动清除体内的致病蛋白，”Zhuoran Qiao 说，“这还只是起步阶段，但机会无限。”

最终目标不仅是加速传统研究，更是实现药物发现自动化，让研发走向人类科学家未曾设想的新方向。“我们希望构建能自动生成全新分子的工作流，”Zhuoran Qiao 说，“从长远来看，我们能够替代药物研发中大量传统的人工流程。”

走出学术圈

为了实现这个目标，Zhuoran Qiao 开始将目光投向学术之外。2023 年，他加入了 Iambic Therapeutics，这家初创公司源自他导师在加州理工的实验室；今年，他又以创始科学家的身份创立了 Chai Discovery。“学术界是原型开发的好地方，但要产生真正的现实影响，就需要更大的团队和更多资源，”Zhuoran Qiao 解释道。

团队协作带来了令人瞩目的成果：Iambic 的 NeuralPLexer 模型刷新了结构预测的行业标准，Chai 最新模型能以 16% 的成功率设计新型抗体——这是此前水平的 100 倍提升。Chai Discovery 联合创始人 Joshua Meier 称之为“蛋白质的 Photoshop”，带来了强大的新见解。“数字生物学不再是科幻——它已经成为现实，”他说。

目前，DeepMind 的 AlphaFold 数据库已被超过 200 万名研究者使用，完成了以往需 10 亿年才能完成的分析；2024 年，AlphaFold 的两位创始人获得了诺贝尔化学奖。现在，Zhuoran Qiao 说，研究者们正迈入“后 AlphaFold”时代，利用更强大、更专业的模型加速发现。“我们有机会以前所未有的规模绘制蛋白质相互作用图谱，”他说，“我很幸运能参与这场变革。”

现在，Zhuoran Qiao 表示，是时候抓住这个机遇，打造能带临床变革性成果的实用工具和工作流了。“这对我来说很有情感意义，”Zhuoran Qiao 补充道，“如果我们能做到这一点，整个计算药物研发的方式都将被彻底改变。”