
中国科大等利用人工智能预测蛋白质“光学指纹”

作者：writer 来源：中国科学院

本文原地址：<https://www.iikx.com/news/progress/5505.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

中国科大等利用人工智能预测蛋白质“光学指纹”。蛋白质是生命的基石，生物的功能依赖于既稳定而又灵活可变的蛋白质结构。蛋白质的光谱响应信号，尤其是紫外光谱，可以称之为蛋白质骨架的“指纹”。这个“光学指纹”，经过理论模拟的解读，可以揭示出精确的蛋白质结构，为生命科学和医学诊断提供极其重要的信息。

然而，蛋白质的结构极其复杂多变，需要做大量的高精度的量子化学理论计算。由于计算量太大，即使是最厉害的超级计算机轻易也“吃不消”。所以蛋白质的光谱的理论解读是一个长期的困难与挑战，限制了光谱的准确分析和蛋白质结构的发现。

怎么在光谱理论模拟中避免太昂贵的量子化学计算，解读蛋白质骨架的“光学指纹”，是一个重要的科学课题。而近年来，人工智能技术被广泛应用到各个领域，用于大幅度降低复杂体系的计算量。

最近，中国科学技术大学微尺度物质科学国家研究中心教授江俊，与中国科大教授罗毅和美国加州大学尔湾分校教授Shaul Mukamel合作，通过利用人工智能机器学习中的神经网络技术，模拟了蛋白质肽键结构与性质之间的构效关系，将计算量一下降低了上万倍。最后他们成功地预测了肽键紫外光谱，还用随机森林的办法揭示了具有化学内涵的结构描述子和构效关系。人工智能与量子化学理论计算的结合，为预测蛋白质的光学特性提供了一种高效的工具。相关成果以A Neural Network Protocol for Electronic excitations of N-Methylacetamide 为题发表在《美国国家科学院院刊》(PNAS, DOI:10.1073/pnas.1821044116)。

江俊课题组近些年致力于发展机器学习技术在量化领域的应用，努力探究使其成为解决量化问题的一种重要工具。在该工作中，研究人员首先在300K温度下通过分子动力学模拟以及量子化学计算，得到了五万组不同构型的肽键模型分子。通过机器学习算法筛选出键长、键角、二面角跟电荷信息作为描述符，通过神经网络来构建肽键基态结构与其激发态性质之间的构效关系。基于训练好的机器学习模型，预测出了肽键的基态偶极矩及激发态性质，最后预测出肽键的紫外吸收光谱。为了验证机器学习模型的鲁棒性，研究人员又基于300K的温度下得到的机器学习模型，预测出肽键在200K以及400K温度下的紫外吸收光谱，其结果与时间密度泛函理论计算很好地吻合。

这是人工智能技术首次用于理论计算预测蛋白质的光谱研究。通过理论计算得到大量数据，使用人工智能加以训练构建构效关系，使用最后得到的模型用于预测，为模拟蛋白质的光谱提供了一种新的思路。该项工作确立了机器学习模拟蛋白质肽键骨架紫外吸收光谱的可行性和优势，蛋白质的“光学指纹”解读也将会变得更加轻易和有效。

相关工作得到国家自然科学基金、中科院先导项目的资助，该论文第一作者为博士生叶盛与博士后胡伟、李鑫，江俊与Shaul Mukamel为共同通讯作者。

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发