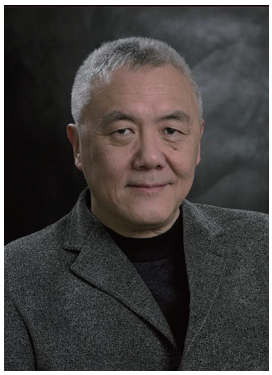
DOI: 10.13376/j.cbls/2025146
文章编号：1004-0374(2025)12-1479-02

·专辑·



赵国屏，分子微生物学家，中国科学院院士，中国医学科学院学术咨询委员会学部委员，发展中国家科学院院士，美国微生物科学院院士。现任国科大杭州高等研究院首席教授，复旦大学微生物组中心主任。兼任中国生物工程学会合成生物学专业委员会名誉主任，上海生物工程学会名誉理事长，上海微生物学会名誉理事长。研究工作涉及微生物生理生化、基因组学、系统与合成生物学以及生物信息学等领域。曾参与启动中国人类基因组计划及相关生命“组学”研究，克隆若干遗传病致病基因；主持若干重要微生物的基因组、功能基因组、比较和进化基因组研究，解析 SARS 冠状病毒分子进化机制。在细菌蛋白质乙酰化组和肠道微生物组等领域作出若干开创性工作。组建并领导中国科学院合成生物学重点实验室，在酵母染色体重构、代谢组与代谢流量组研究、天然化合物细胞工厂制造、基因编辑技术研发等方向上，实现重要突破。近年来，积极参与中国科学院上海营养与健康研究所生物医学大数据中心为建设国家生物医学大数据治理体系所开展的基础性科学工作。

张国庆，研究员，博士生导师。现任中国科学院上海营养与健康研究所生物医学大数据中心执行主任，上海生物医学大数据工程技术研究中心执行主任；中国卫生信息与健康医疗大数据学会多组学生物信息与未来医学工程分会副秘书长，中国生物信息学会 (筹) 生物数据资源专委会副主任，上海生物信息学会生物医学数据专委会主任。近年来，面向生物医学大数据与人工智能，开展生物医学数据科学的关键技术研究及工程平台建设，并在人群队列与专病库、环境与人体微生物组等领域数据的规范化采集、清洗与应用等方向应用，形成了较完整的生物医学大数据集成治理技术体系和示范性应用场景。

序　言

张国庆，赵国屏
(中国科学院上海营养与健康研究所，上海 200031)

基础模型体系正推动生命科学与生物医学进入以高质量数据、模型能力与应用验证协同推进的数据密集型研究新阶段。这里的基础模型既包括面向文本与知识的大语言模型，也包括面向生物序列、结构、单细胞与影像等模态的专用语言模型与跨模态模型。随着多组学、单细胞与空间组学、蛋白质序列与结构、医学影像与连续生理监测等数据快速扩展，研究对象在分子、细胞、组织到人群队列之间形成跨尺度关联，同时也使异构数据对齐、缺失与偏倚处理、跨中心可比性与可复用性成为制约知识产出的关键瓶颈。

在此背景下，模型训练与知识增强正在成为贯通数据与应用的重要抓手。一方面，预训练、指令微调与对齐等训练范式为模型提供通用能力，并通过领域适配提升其在生物医学语境下的可用性与稳健性。另一方面，检索增强生成 (retrieval augmented generation, RAG) 通过将外部证据与模型推理结合，缓解幻觉与可追溯性不足的问题，而融合知识图谱

的 RAG 进一步以结构化关系组织证据链，支持更清晰的推理依据与人机协同决策。与此同时，隐私合规、数据产权边界与真实世界验证仍决定模型能否在科研与临床场景中稳定落地。

本专辑以 AI for Life Science 为主题，面向生命科学与生物医学的共同需求组织稿件，系统呈现垂直领域模型与大语言模型的训练与评估、RAG 与知识图谱增强 RAG 等关键技术路线，并覆盖多组学队列与疾病预测、循证决策与临床转化、蛋白质与细胞基础模型、靶点发现及小分子与核酸药物设计、合成生物学等应用方向。专辑旨在为读者提供贯通数据治理、模型构建、证据推理与转化验证的问题框架，促进高质量数据供给与模型能力协同增值。