

03-作品 Demo

推报学校名称：福建师范大学

揭榜榜题名称：XH-202620 面向一流学科建设的学科垂类大模型与创新应用开发

榜题发榜单位：科大讯飞股份有限公司

申报作品名称：Syn-BioGPT：面向合成生物学的垂类大模型开发及生物学科创新研究支撑

申报者姓名：吴亚祺、曾信伟、徐愿、何宇彬、王晓彤、张俊昊、蔡伟贤

目录

1.Demo 快速入口	3
2.作品定位	4
3.完整 Demo	4
4.三类专业能力的真实执行路线	5
4.1 合成路径设计	5
4.2 蛋白功能与结构	5
4.3 未知序列分析	6
5.技术架构与垂类化实现	7
6.测试问题	8
7.使用说明与注意事项	8

1.Demo 快速入口

体验地址: <https://biogpt.com.cn/>

适配终端: Chrome/Edge 桌面浏览器。打开链接后进入作品门户, 可浏览作品 Demo、作品代码和团队信息; 点击“立即体验”进入智能体对话。



图 1-1 扫码打开 Demo



图 1-2 Syn-BioGPT 作品门户与体验入口



图 1-3 Syn-BioGPT 功能界面

2.作品定位

Syn-BioGPT 面向生命科学与合成生物学科研场景，重点解决“专业数据库分散、科研流程难以衔接、通用大模型证据边界不清”的问题。当前版本不追求让模型直接替代科研人员，而是将合成生物学研究中的关键任务封装为可控、可追溯的专业能力。

轻量垂类化方式：平台托管大语言模型→自研领域 Skill 进行意图路由与规则约束
→只调用白名单学术数据库→结构化证据回答→用户主动选择下一步。

能力模块	功能实现
合成路径设计	目标分子、二维结构、天然来源、KEGG 路径重建、底盘分析、文献增强、优化建议
蛋白功能与结构	已知蛋白功能注释；按需查询 AlphaFold、NCBI、Uniprot；机制/工程问题 按需补论文
未知序列分析	PROSITE 快速识别→CDD 增强→按需 BLAST/UniProt→置信度分层与文献支持

表 2-1 专业能力模块及功能实现

3.完整 Demo

阶段	示例输入	预期展示
01 目标提出	我想在大肠杆菌中合成番茄红素	触发“合成路径设计”：目标产物识别→二维结构→代表性来源植物→路线重建→大肠杆菌底盘→文献增强→路线优化。
02 深入关键酶	CrtI 是什么酶？它在番茄红素合成里具体干什么？	切换到“蛋白功能与结构”：读取已知蛋白注释，不自动跑整条路径，不自动查结构。
03 工程研究	CrtI 有哪些工程改造或者提高催化效率的研究？	在已知蛋白身份基础上补充 PubMed 研究证据，展示“从路径到关键酶”的深入分析。
04 结构层	帮我看一下这个蛋白的 AlphaFold 三维结构：	当用户主动要求时查询 AlphaFold，展示预测结构与置信信息。

表 3-1 完整 Demo

一句话主线：“我想合成什么”→“怎么合成”→“关键酶是什么”→“有什么工程研究”
→“蛋白结构长什么样”。

4.三类专业能力的真实执行路线

4.1 合成路径设计

用户目标→KEGG 目标与反应/模块证据→PubChem 二维结构→代表性来源植物→候选路径重建→指定底盘或默认底盘比较→PubMed 工程证据→定性优选与优化建议→横向 Mermaid 路径图。

主要优化方向

基于近年文献证据，大肠杆菌番茄红素合成的主流优化策略可归纳为4类：

- 前体途径强化：**过表达MEP途径关键限速酶（dxs、idi等），提升IPP/DMAPP池的供给，打通上游萜类合成瓶颈。
- 路径基因表达平衡：**通过启动子工程、质粒拷贝数优化或染色体整合策略，精细调节CrtE/CrtB/CrtI三个基因的表达剂量，避免中间产物积累或酶表达失衡导致的代谢负担。
- 竞争支路敲除：**阻断FPP流向其他内源萜类产物的竞争支路，减少碳流分流，定向引导碳流进入番茄红素合成模块。
- 产物储存与耐受优化：**通过改造细胞膜结构、增加膜表面积，或采用minicell形成突变体，提升细胞对疏水性番茄红素的储存能力和产物耐受性，已有研究报道该类改造可使番茄红素产量提升259%。

图 4-1 合成路径优化建议（局部画面）

4.2 蛋白功能与结构

已知蛋白/基因/Accession→NCBI/Protein 注释→功能/EC/催化信息→用户明确要求结构时查询 AlphaFold→机制、突变、工程改造等问题查询 PubMed。

7. 代表性文献证据汇总

本预研报告所有核心结论均有已发表的权威研究文献支撑，代表性研究汇总如下：

PMID	文献来源	核心贡献
2254247	<i>J Bacteriol</i> , 1990	首次在大肠杆菌中功能解析完整类胡萝卜素合成路径，明确CrtI的四步脱氢功能，是该领域的奠基性工作 (查看文献)
1400305	<i>J Biol Chem</i> , 1992	完成重组CrtI的表达纯化与体外复活，直接验证其FAD依赖的催化活性 (查看文献)
25959020	<i>Metab Eng</i> , 2015	通过定向进化改造CrtI，结合代谢工程优化大幅提升番茄红素合成产量 (查看文献)
40724976	<i>Int J Mol Sci</i> , 2025	通过氧化还原平衡协同工程，在生物反应器中实现3.07 g/L的番茄红素产量，代表当前最高发酵水平

图 4-2 CrtI 功能与工程文献证据（报告节选）



图 4-3 CrtI AlphaFold 结构与置信度结果

4.3 未知序列分析

未知氨基酸序列→ScanProsite 快速家族识别→证据不足时 CDD 增强→用户需要具体同源蛋白时 BLAST→UniProt 注释→置信度分层→中高置信后按需补 PubMed。



图 4-4 未知序列功能判定结果

系统边界：系统不会因为用户问了一条路径，就自动继续执行蛋白结构、未知序列等所有后续任务；每一步由用户主动继续，避免长链串行调用和无关检索。

5. 技术架构与垂类化实现

本项目采用“轻量垂类增强”而非自行训练超大模型权重：垂类能力主要来自任务拆分、专业数据源、证据约束、固定调用规则和领域输出规范。作品门户采用原生HTML、CSS与JavaScript实现，统一提供体验入口、作品 Demo 和作品代码，并通过独立配置文件接入 WebSDK。

层级	实现	作用
交互层	公网对话页面	自然语言输入、多轮上下文与结果展示。
模型层	平台托管的大语言模型	任务理解、信息整合与教学表达；当前版本无独立模型权重文件。
领域能力层	3 个自研 Skill	将合成路径、已知蛋白、未知序列三类意图的执行流程、白名单接口和证据边界固化。
数据层	本地轻量数据库	10 张结构化数据表，支持自然语言查询与计算（NL2SQL）；数据及证明材料按版权授权核验。
知识库层	合成生物学 RAG 向量知识库	汇集领域文献、课程资料与典型案例，通过向量检索和 Rerank 重排召回高相关知识片段，为专业问答、教学表达与报告生成提供上下文支撑。
证据层	KEGG/PubChem/PubMed/NCBI/AlphaFold/PROSITE 等	按照 Skill 中声明的接口进行访问，保证数据可信度。
输出层	结构化 Markdown	路径图、证据表、置信度、文献、下一步入口与安全边界。

表 5-1 技术架构与垂类化实现

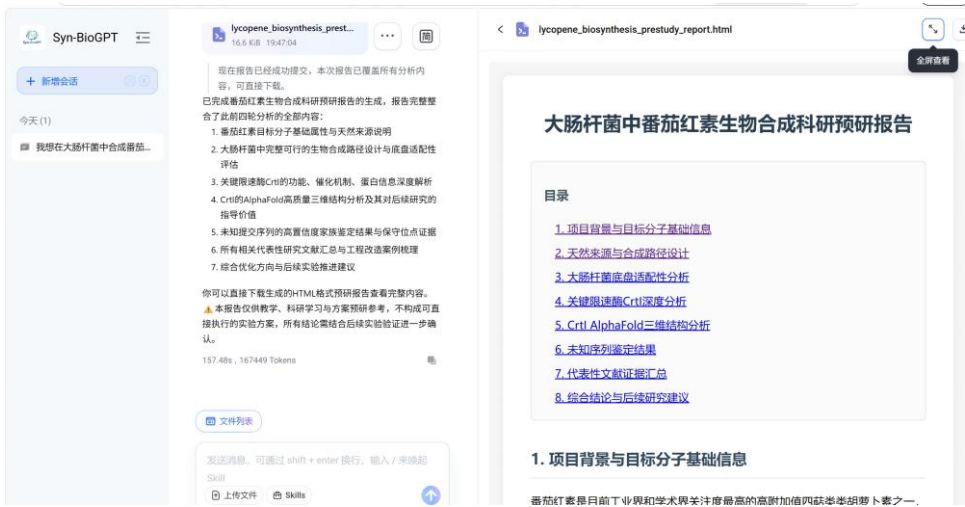
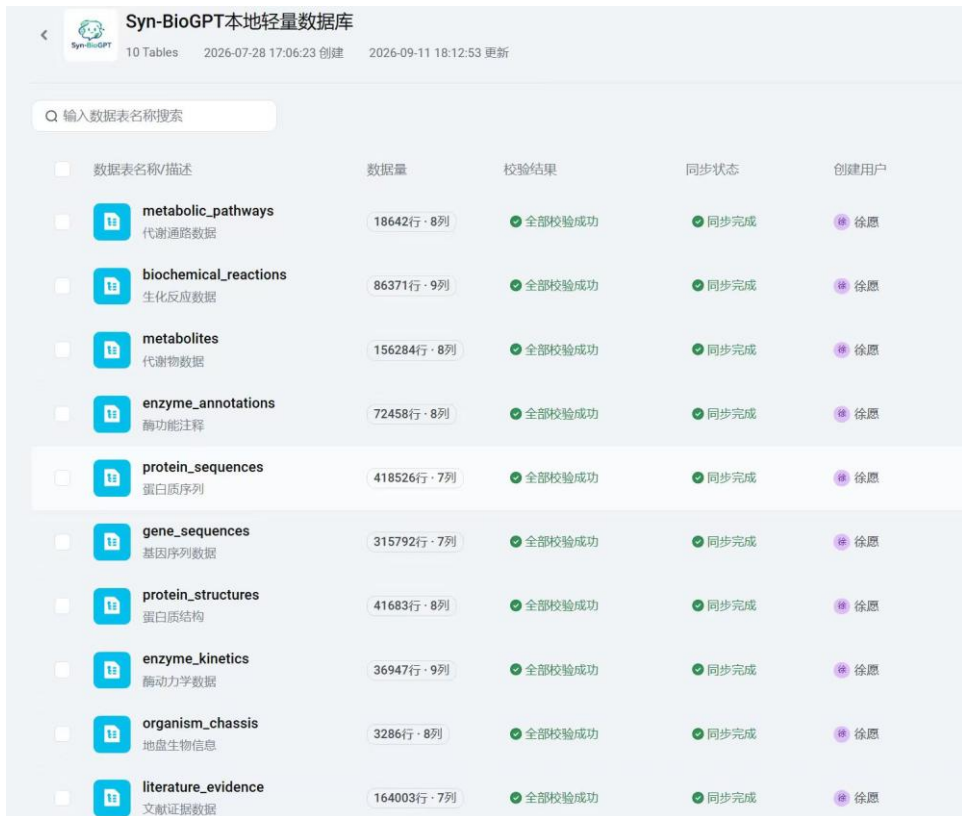


图 5-1 综合预研报告生成与预览



数据表名称/描述	数据量	校验结果	同步状态	创建用户
☐ metabolic_pathways 代谢通路数据	18642行 · 8列	全部校验成功	同步完成	徐恩
☐ biochemical_reactions 生化反应数据	86371行 · 9列	全部校验成功	同步完成	徐恩
☐ metabolites 代谢物数据	156284行 · 8列	全部校验成功	同步完成	徐恩
☐ enzyme_annotations 酶功能注释	72458行 · 8列	全部校验成功	同步完成	徐恩
☐ protein_sequences 蛋白质序列	418526行 · 7列	全部校验成功	同步完成	徐恩
☐ gene_sequences 基因序列数据	315792行 · 7列	全部校验成功	同步完成	徐恩
☐ protein_structures 蛋白质结构	41683行 · 8列	全部校验成功	同步完成	徐恩
☐ enzyme_kinetics 酶动力学数据	36947行 · 9列	全部校验成功	同步完成	徐恩
☐ organism_chassis 底盘生物信息	3286行 · 8列	全部校验成功	同步完成	徐恩
☐ literature_evidence 文献证据数据	164003行 · 7列	全部校验成功	同步完成	徐恩

图 5-2 本地轻量数据库数据表列表

6. 测试问题

能力场景	测试问题	能力验证重点
综合路径设计	我想在大肠杆菌中合成番茄红素	评估系统从目标产物识别、二维结构解析、合成路径重建、底盘适配到文献支撑与优化建议的完整预研能力。
指定起点设计	我想从 GGPP 开始合成番茄红素	评估系统围绕指定起点精准重建后续合成路线，并给出关键反应、所需酶及证据来源。
路径优化决策	这条大肠杆菌番茄红素路线还有哪些值得优先优化的地方？	评估系统结合代谢节点、关键酶和文献证据识别优先优化环节，形成可执行的研究建议。
关键酶解析	CrtI 是什么酶？	评估系统对关键酶身份、催化功能、对应反应及其在目标路径中作用的专业解析能力。
酶工程研究	CrtI 有哪些提高催化效率或工程改造研究？	评估系统整合蛋白功能与前沿文献，归纳酶工程改造方向、研究依据和可借鉴方案。
未知序列解析	粘贴未知氨基酸序列并询问功能	评估系统完成家族特征识别、功能线索整合、置信度分层及后续验证建议的能力。

表 6-1 测试问题

7. 使用说明与注意事项

- 公网体验入口见第 1 节。知识库与数据库因版权保护采用授权核验方式，证明材料请联系 108012025017@student.fjnu.edu.cn。公共数据库繁忙时响应可能波动；结果用于教学与科研预研，需结合专业数据库和实验核验。