arXiv:2603.01464v1 [cs.AI] 2 Mar 2026

ProtRLSearch：通过强化学习训练的大语言模型驱动的多轮多模态蛋白质搜索智能体

Congying Liu[1], Taihao Li[2], Ming Huang[3], Peipei Liu *[1,4],Xingyuan Wei[1,4],Yiqing Shen[5],Yanxu Mao[6],Tiehan Cui[6]1
中国科学院大学，北京，中国[2]中国科学院大学杭州高等研究院，杭州，中国[3]中国科学院深圳先进技术研究院，深圳，中国
[4]中国科学院信息工程研究所，北京，中国[5]约翰斯·霍普金斯大学，美国巴尔的摩[6]河南大学，中国开封

摘要—**医疗保健场景中的蛋白质分析任务通常要求在蛋白质序列约束下进行准确推理，涉及疾病相关变体的功能解读、临床研究中的蛋白质层面分析等任务。为应对此类任务，人们引入搜索智能体来检索蛋白质相关信息，为以蛋白质为中心的推理中的疾病相关变异分析和蛋白质功能推理提供支持。然而，这类搜索智能体大多局限于单轮、纯文本模态搜索，无法将蛋白质序列模态作为多模态输入纳入搜索决策过程。同时，它们依赖的强化学习（RL）监督仅关注最终答案，缺乏对搜索过程的约束，因此难以及时识别并纠正关键词选择和推理方向上的偏差。为克服这些局限，我们提出 ProtRLSearch，这是一种通过多维奖励强化学习训练的多轮蛋白质搜索智能体，可在实时搜索中联合利用蛋白质序列和文本作为多模态输入，生成高质量报告。为评估模型在真实蛋白质查询场景中整合蛋白质序列信息与基于文本的多模态输入的能力，我们构建了 ProtMCQs，这是一个包含 3,000 道选择题、分为三个难度级别的基准测试。该基准测试评估的蛋白质查询任务涵盖了从在序列约束下推理蛋白质功能和表型变化，到结合多维序列特征、信号通路与调控网络进行综合蛋白质推理。实验结果表明，ProtRLSearch 将第 1 级的准确率从 35.7% 提高到 86.9%，将第 2 级的准确率从 30.5% 提高到 77.4%，并将第 3 级的准确率从 26.1% 提高到 72.5%**。

索引词—**多模态搜索智能体、蛋白质、大语言模型、强化学习**

I. 引言

近年来，大语言模型（LLM）已被引入蛋白质分析领域，以支持变异效应解读以及跨蛋白质和信号通路层面的推理等蛋白质相关分析任务，从而助力医疗应用中的疾病相关变异分析和蛋白质推理

*通讯作者：peipeiliu@yeah.net

应用于临床研究 [1]。然而，由于无法直接获取真实的序列信息和相关文献，大语言模型在功能归因和变异解读等任务中往往容易产生幻觉 [2]。因此，要在蛋白质分析中实现可靠推理，就需要整合实时在线搜索，将可验证的蛋白质数据库和文献作为真实的蛋白质搜索结果，为模型的蛋白质多步推理提供支持 [3]。

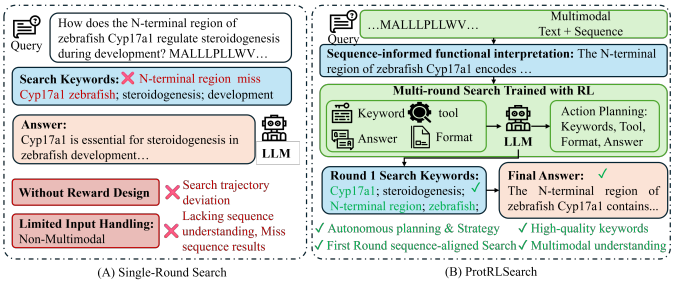

图 1. 单轮搜索方法与所提出方法的对比，展示了两者在搜索过程中的差异。红色标注表示错误情况，例如缺少关键词；绿色标注表示正确结果。

Existing searc h agents suc h as BioMedSearch [4]一个d Mind - 搜索智能体主要基于文本模态运行，并采用单轮搜索：针对给定查询，仅从多个数据库中进行一轮搜索，然后根据筛选后的结果进行蛋白质相关推理。这类搜索智能体高度依赖初始阶段生成的关键词，而蛋白质相关结论往往受到序列层面特征的严格限制，包括蛋白质序列、结构域和突变位点。当关键的序列相关信息无法通过文本关键词充分表达时，模型便难以判断搜索结果是否与给定序列一致，从而限制了序列约束推理的有效性。

为解决这些局限性，研究人员提出了 Search-R1 [6] andR1-

Searcher [7], 等多轮搜索智能体，引入强化学习（RL）和基于最终答案的奖励机制，以规范查询生成和搜索轮次决策，并通过多轮搜索逐步整合额外的蛋白质相关文本信息。然而，这些方法的搜索规划仍完全依赖文本模态，未将蛋白质序列作为并行输入，用于约束搜索计划或指导决策。因此，多轮搜索仍可能偏离关键的序列特征。此外，这些方法主要依赖最终答案奖励，对搜索工具选择和搜索路径规划施加的约束有限，导致早期搜索方向中的错误持续存在，并在后续轮次中不断被放大 [8]。

为解决先前工作的局限性，我们提出 Pro-tRLSearch，这是一种采用强化学习训练、并通过多维奖励信号进行优化的多轮实时搜索智能体，可将蛋白质序列和文本信息作为多模态输入。具体而言，该智能体利用蛋白质基础模型学习氨基酸序列的表征，使模型能够更深入地理解与蛋白质结构和功能相关的内在属性。通过基于关键词、答案、搜索工具和输出格式的多维奖励信号进行训练，搜索组件（包括规划器、检索器和执行器）能够学习自主规划搜索过程并生成答案 [6]。规划器通过关注查询的蛋白质序列所涉及的属性（如相互作用和变体），学习提取有代表性的关键词并分配相应的搜索工具，从而确定后续搜索路径。检索器在多个数据库中执行实时搜索以获取候选结果，并根据多模态查询进一步筛选检索内容，保留符合查询的信息以供下游推理。执行器学习根据搜索结果和多模态输入进行推理并生成答案，同时评估当前答案是否充分满足需求，并据此决定是否触发后续搜索轮次。

现有数据集（如 ProteinLMDataset [9], ）虽包含多模态信息，却缺乏对搜索与推理过程的明确监督，因此在蛋白质搜索任务中的适用性受到限制。为此，我们构建了 3,000 个多模态训练样本，将多模态输入与搜索规划、答案信息及辅助字段明确关联起来。此外，当前基准测试（如 MedMCQA [10] ）并未评估模型能否通过整合序列信息与检索内容来回答蛋白质相关问题 [11]。我们构建了 ProtMCQs，这是一个包含 3,000 道选择题（ MCQ ）的基准测试，涵盖从受序列约束的功能与表型推理，到基于多个序列特征、信号通路及调控网络的综合蛋白质推理等任务。我们的贡献总结如下：

- 我们提出了 ProtRLSearch，这是一种通过强化学习训练的多轮多模态蛋白质搜索智能体。通过将蛋白质序列表示融入搜索过程，ProtRLSearch 能够在多轮搜索中做出序列感知搜索决策，并提升以蛋白质为中心的查询的推理质量。
- 我们在强化学习过程中设计了一种多维奖励方案，综合利用关键词、格式、搜索工具和答案质量等方面的信号。这些奖励组成部分经过高效整合，引导模型自主优化搜索路径。
- 我们推出了 ProtMCQs，这是一个包含 3,000 个多模态蛋白质问题的基准数据集，涵盖三个难度级别和多种关键蛋白质场景，用于评估与序列相关的推理能力。

II. 相关工作

A. 蛋白质查询任务的搜索智能体

大语言模型能够生成蛋白质任务中关于结构域功能和蛋白质相互作用的解释 [1]。然而，其输出往往与经过整理的序列注释和结构知识不一致 [4], ，这表明有必要在推理过程中引入搜索智能体，以提供可靠的蛋白质信息 [12]。现有搜索智能体（如 BioMedSearch [4],、Pasa [13] 和 MindSearch [5], ）通过单轮搜索整合来自多个来源的蛋白质信息；而 Search-R1 [6] 和 R1-Searcher [7] 则采用多轮搜索，逐步纳入搜索结果并优化答案。然而，这些方法通常将蛋白质序列视为文本，未能充分建模序列中蕴含的结构与功能语义，导致搜索路径偏离蛋白质推理所需的逐步结构化流程 [14]。为解决这一局限，ProtRLSearch 将蛋白质序列与文本整合为多模态输入，并在序列相关约束下进行多轮搜索，从而确保搜索信息与蛋白质序列保持一致。

B. 面向蛋白质任务的大语言模型强化学习奖励设计

强化学习已应用于通过反馈信号 [15]优化蛋白质相关任务中的搜索与推理。BioRea- son [16] 和 Search-R1 [6] 在多步推理或多轮搜索中引入基于答案的奖励，从而提升与参考答案的一致性，并支持对结构域功能、蛋白质相互作用及变体效应的分析。然而，这些方法主要依赖针对最终答案定义的奖励，缺乏对蛋白质关键词识别和搜索工具选择的约束，因此限制了在异构数据源中构建有效搜索路径的能力 [17]。相比之下，ProtRLSearch采用多维度奖励设计，引导模型围绕蛋白质序列、结构域和变体信息组织多轮在线搜索 [18]。

III. 方法

A. 基于大语言模型骨干网络的蛋白质序列与文本多模态表示

我们借鉴 ProtLLM [19] 的设计理念，并引入蛋白质序列表示模块，从原始蛋白质序列中提取

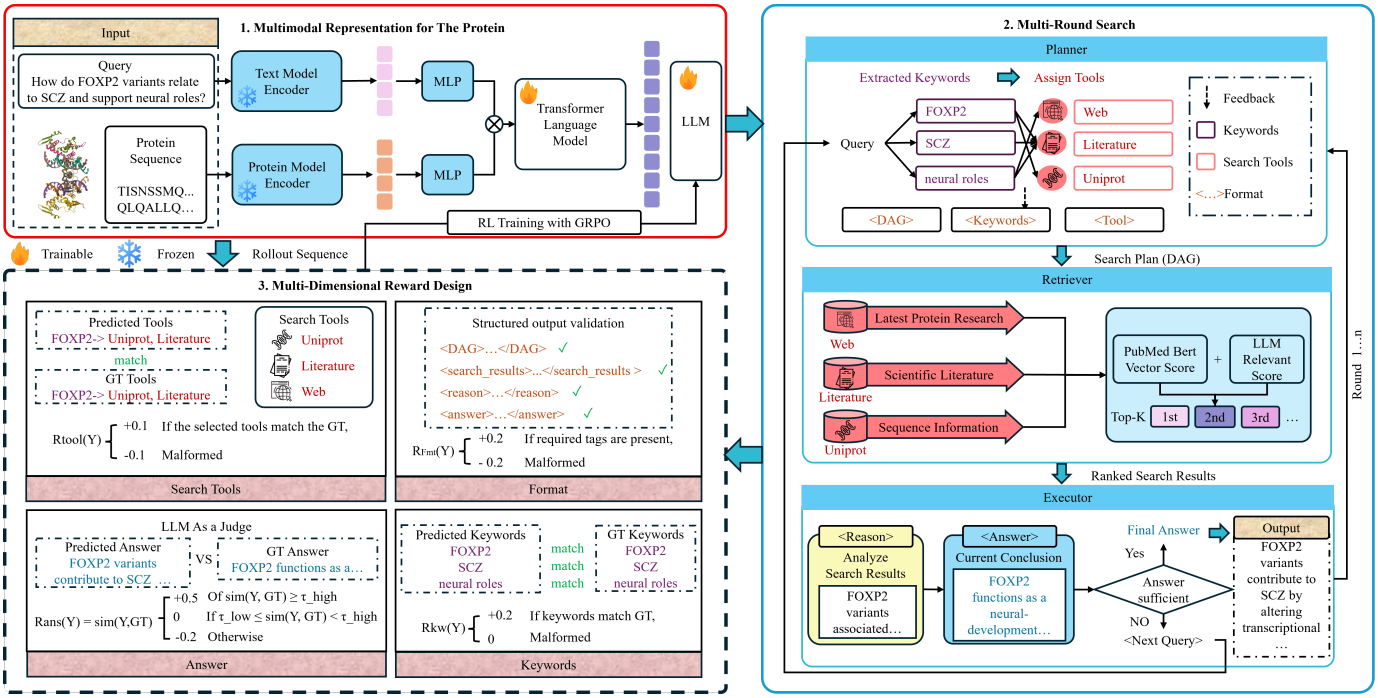

图2. 整体代理涵盖多模态编码、通过强化学习训练的多轮搜索，以及多维奖励的设计

特征表示。模型使用 ESM-2 [20]等冻结的蛋白质预训练模型，获取残基级上下文表示，涵盖结构域边界、保守基序以及与功能相关的序列特征。随后，通过可学习的线性投影，将这些高维蛋白质嵌入映射到大语言模型的输入空间，使蛋白质序列能够在词元级与文本信息一同进入统一的Transformer架构。为提升序列与语义之间的对齐质量，我们采用 ProtLLM [19] 发布的 InterPT数据集作为训练语料，引导模型在蛋白质区域与其对应描述之间建立稳定的映射。投影后的蛋白质嵌入在序列级与查询文本拼接，并共同输入大语言模型，使模型能够在统一的注意力范围内处理两种模态的信息。随着训练的进行，融合层逐渐形成稳定的序列与语义对齐能力。蛋白质嵌入在大语言模型空间中能够更准确地反映残基片段、结构域边界及与功能相关的信号，使大语言模型能够同时关注蛋白质序列中的关键区域和文本中的重要描述。训练完成后，如图2所示，融合表示展现出更强的跨模态关联性和内部一致性，为后续多轮搜索提供连贯且可交互的基础。

B. 具有结构化输出的多轮搜索

我们构建了一个由规划器、检索器和执行器组成的多轮搜索智能体。该智能体将蛋白质序列和查询作为多模态输入，生成搜索计划，执行并行多源搜索，并更新其结论。与仅输出自然语言答案的方法不同，每一轮都会生成结构化搜索输出，包括搜索计划、排序后的搜索结果，以及用于生成下一轮查询的中间结果，从而支持多轮搜索的持续推进。

在每一轮搜索中，规划器模块使用大语言模型将当前查询转换为搜索计划，如图2所示。在此过程中，模型从查询中提取多个以蛋白质为中心的关键词，并为每个关键词分配搜索工具，包括用于网络搜索的 Web、用于文档检索的文献，以及用于蛋白质数据库查询的 UniProt。关键词与工具之间的关联被组织成一个有向无环图（DAG），构成规划器在当前轮次的结构化输出，并作为后续并行搜索的搜索计划。

检索器模块根据有向无环图（DAG）执行多源搜索，并从网页内容、文献来源和蛋白质数据库注释中获取搜索结果。检索结果依据两类相关性信号进行评分。第一类是基于 PubMedBERT [21]，的向量相关性分数，用于衡量查询与文本内容之间的语义相似度。第二类是基于大语言模型的相关性分数，用于评估给定蛋白质查询下每条搜索结果的相关性。结合这两种相关性分数后，检索器会输出一组结构化的 Top-K 搜索结果，并按相关性为当前轮次排序。

执行器模块基于当前轮次的排序后的检索结果进行分析，并输出一组结构化结果字段。具体而言，模型在 <推理> 中分析检索内容，并在 <答案>中生成当前轮次的中间结论。 <判定> 字段表明可用的检索是否

结果足以解答原始查询，而当结果不足时，<下一_查询>字段会提供下一轮搜索所需的查询。这些结构化输出共同构成当前轮次的分析结果，其中 <下一_查询>与蛋白质序列相结合，作为后续轮次中新的多模态输入，以推进下一轮搜索与分析过程。

C. 多维奖励设计

为引导大语言模型在蛋白质相关查询任务中获得稳定的搜索与推理能力，我们将训练目标分解为四个可验证的子目标，并构建由答案、关键词、工具和格式组成的多维奖励函数。每种奖励对应图2中展示的一项独立约束。总体奖励定义为各子奖励的加权和，如下所示：

$$R_{\text{total}} = \lambda_{\text{Ans}} R_{\text{Ans}} + \lambda_{\text{KW}} R_{\text{KW}} + \lambda_{\text{Tool}} R_{\text{Tool}} + \lambda_{\text{Fmt}} R_{\text{Fmt}}.$$

答案奖励用于评估模型生成的答案与真实标注（GT）答案之间的一致性，如图2所示。该奖励由大语言模型评审器计算，用于评估预测答案与GT答案之间的语义相似度。当相似度超过预设阈值时，给予正奖励；否则给予负奖励。此奖励约束模型在多轮搜索后生成符合查询目标的最终答案；关键词奖励约束模型在第一轮搜索规划中提取蛋白质相关关键词的行为。将预测的关键词集合与根据GT得出的关键词集合进行匹配。当预测关键词与GT关键词匹配时，给予正奖励；缺失或不匹配的关键词则不予奖励或给予惩罚。这一设计强调了蛋白质名称、结构域及其他核心关键词在搜索流程中的关键作用；工具奖励引导模型根据提取的关键词选择适当的搜索工具。将预测的关键词与工具分配结果，与GT中定义的工具分配进行比较。当预测工具与GT工具匹配时，给予正奖励；不合理的工具选择或格式错误的输出则会受到惩罚。此奖励确保不同类型的蛋白质相关关键词映射到正确的数据源；格式奖励约束模型输出满足预定义的结构化格式要求。输出必须包含完整且有效的结构化标签，例如 <有向无环图>、 <搜索_结果>、 <推理>和<答案>。当所有必需标签均存在且格式正确时，给予正奖励；否则施加惩罚，以确保多轮搜索结果可解析。

D. 训练

我们通过强化学习在一个包含 3,000 个样本的多模态蛋白质数据集上训练模型，其中多维奖励源自真实蛋白质数据库和文献中可验证的信号，从而在没有对中间搜索步骤进行显式监督的情况下，约束跨模态蛋白质搜索。每个样本均基于可验证的蛋白质–物种配对构建，并通过 UniProt reviewed数据库验证，以确保数据可靠性。该数据集涵盖六大类蛋白质，包括转录因子、蛋白质调节因子、离子通道和转运蛋白、与信号传导和炎症相关的蛋白质、突触和神经发育相关蛋白质，以及线粒体蛋白质。数据集在四种推理任务中保持均衡分布，包括从变异到表型的推理、从结构到功能的推理、跨系统机制推理和跨物种比较。如图3所示，每个样本都通过多阶段流程构建。首先，选取一个经过 UniProt reviewed数据库验证的蛋白质–物种配对，并检索相应的蛋白质序列。接着，从 PubMed 收集与功能相关的文献。最后，大语言模型生成研究查询、由关键词和搜索工具组成的有向无环图（DAG），以及相关的推理标签 <推理>和<答案>。这一过程将蛋白质序列、搜索结果和相关搜索信息整理为结构化训练样本，以支持多轮搜索。

E. 基准设计

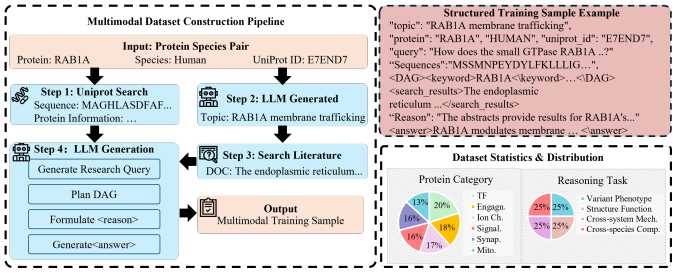

图3. 所构建训练数据集概览 现有生物医学数据集（如 BioMedMCQs [4]）通过模拟真实用户的生物医学查询，在检索与信息整合场景下评估问答表现；而 MedQA [22]和 MedMCQA [10]主要用于衡量医学考试与专业问答场景中的知识理解及基于文本的推理能力。然而，这些数据集并未涵盖以生物序列为核心的蛋白质查询场景，从而限制了对模型蛋白质相关能力的评估 [23]。为弥补这一空白，我们构建了 ProtMCQs，以联合评估模型理解蛋白质序列的能力，以及在搜索过程中对文本信息进行整合推理的能力。其构建遵循与数据生成流程类似的工作流，详细步骤如图3所示。ProtMCQs 按三个逐步递增的难度级别组织，用于评估蛋白质查询任务中的序列约束推理。第 1 级要求模型在相同序列约束下，围绕功能或表型问题整合多项相关检索结果，并对多个序列特征进行联合推理，以支持关于功能变化或表型差异的判断。第 2 级引入基于现有研究生成的多个干扰选项，这些研究涉及序列特征、分子

表 I 不同方法在 BioMedMCQs 和 ProtMCQs 上的性能比较。

数据集	方法	第 1 级			第 2 级			第 3 级		
		Acc%	词元	时间/秒	Acc%	词元	时间/秒	Acc%	词元	时间/秒
BioMedMCQs	基线	45.8±0.6	321±41	6.5±0.7	38.7±1.2	335±58	6.8±1.1	29.5±0.9	379±72	7.6±1.4
	BioMedSearch [4]	73.3±0.8	3029±155	48.1±1.5	69.7±1.4	3348±121	65.1±0.9	62.4±1.0	3987±188	78.9±1.6
	BioReason [16]	63.9±1.7	472±63	11.7±1.3	56.2±0.8	496±71	16.7±0.8	49.3±1.4	528±53	20.5±0.6
	Search-R1 [6]	76.1±0.7	2063±134	19.3±0.4	70.9±1.3	2582±98	27.1±1.2	63.5±1.8	2835±176	32.4±1.9
	ProtLLM [19]	74.8±1.1	409±82	9.6±1.1	60.4±0.9	534±65	14.9±1.3	57.9±0.5	560±103	16.2±1.5
	ProtRLSearch（我们的方法）	**89.2±1.4**	554±117	10.5±0.9	**75.8±0.6**	785±84	13.4±1.6	**71.7±1.9**	873±156	21.3±1.7
ProtMCQs	基线	35.7±0.8	432±74	8.9±0.9	30.5±1.4	468±92	9.6±1.2	26.1±1.1	493±63	12.4±0.6
	BioMedSearch [4]	61.4±0.9	3217±188	53.8±1.3	57.3±1.8	3129±146	58.0±0.7	48.1±0.7	3572±171	69.1±2.5
	BioReason [16]	58.9±1.6	365±55	12.6±1.1	50.6±0.6	503±73	13.3±0.8	47.4±1.3	514±97	15.9±1.4
	Search-R1 [6]	65.6±0.5	2376±162	19.3±2.6	58.1±1.9	2582±183	27.6±1.3	49.2±1.2	2935±115	32.4±1.1
	ProtLLM [19]	72.1±1.4	448±69	10.5±1.5	62.8±0.7	514±82	16.3±1.4	59.1±1.0	546±104	18.2±0.9
	ProtRLSearch（我们的方法）	**86.9±1.3**	632±118	15.9±0.7	**77.4±0.8**	706±91	20.4±1.2	**72.5±1.7**	989±153	24.3±0.4

表 II 在不同预训练配置下 BioMedMCQs 和 ProtMCQs 的性能比较。

数据集	蛋白质	RL	第 1 级			第 2 级			第 3 级		
			Acc%	词元	时间/秒	Acc%	词元	时间/秒	Acc%	词元	时间/秒
BioMedMCQs	✓	✓	**89.2±1.4**	**554±117**	**10.5±0.9**	**75.8±0.6**	**785±84**	**13.4±1.6**	**71.7±1.9**	**873±156**	**21.3±1.7**
	✓	×	76.7±1.9	977±148	15.4±2.3	70.7±0.9	1719±238	20.5±1.7	65.1±1.8	2178±327	35.6±1.3
	×	✓	83.2±0.6	586±103	12.6±1.2	72.9±1.5	874±129	15.3±2.1	69.8±0.7	910±203	22.1±1.8
ProtMCQs	✓	✓	**86.9±1.3**	**632±118**	**15.9±0.7**	**77.4±0.8**	**706±91**	**20.4±1.2**	**72.5±1.7**	**989±153**	**24.3±0.4**
	✓	×	78.6±1.7	1016±175	24.7±1.8	63.7±1.0	1623±136	22.8±1.3	61.2±1.4	2031±164	33.4±2.0
	×	✓	73.5±0.5	795±134	19.2±2.5	61.2±1.9	1072±148	25.7±1.9	55.4±1.1	1518±172	29.7±1.5

特性、生物学背景和物种差异，要求模型判断这些选项是否与给定的序列信息一致。第 3 级进一步提高推理复杂度，要求模型在共享序列背景下，将多维序列特征与来自信号通路和调控网络的信息相融合，从而得出全局一致的蛋白质层面的推理决策。

IV. 实验

A. 实现细节

在 BioMedMCQs [4] 和 ProtMCQs 选择题数据集上，我们记录模型准确率、生成的词元数量和平均推理时间。所有模型均采用统一的输入格式和评估模板，以确保结果具有可比性。此外，我们还在从 MedMCQA [10] 和 MedQA [22]中随机抽取的 1000 道问题上评估对比方法，以衡量其在蛋白质相关任务之外的一般医学问答场景中的泛化能力。所有对比方法均基于 Qwen3-8B [24] ，并采用相同的模型配置。对于所有搜索方法，检索段落数量统一设为 3。ProtRLSearch 中多维奖励信号的权重配置定义如下：

$$\lambda_{\text{Ans}} = 0.5, \quad \lambda_{\text{KW}} = 0.2, \quad \lambda_{\text{Tool}} = 0.2, \quad \lambda_{\text{Fmt}} = 0.1.$$

答案奖励被赋予最大权重，以优先确保与任务目标保持一致；关键词奖励和搜索工具奖励采用相同权重，共同约束搜索方向和工具选择；格式奖励则作为辅助约束，以确保输出规范。

表 III MedMCQA 和 MedQA 上的性能比较（准确率 %）。

方法	MedMCQA	MedQA
BioBERT [25]	36.7	37.1
SciBERT [26]	-	39.2
MedAlpaca-7B [27]	55.2	45.8
MEDAGENTS [28]	64.1	59.3
ProtRLSearch（无强化学习）	85.8	83.9
ProtRLSearch（无蛋白质）	88.5	85.2
ProtRLSearch	**90.4**	**87.4**

B. 主要结果

表 I 展示了 ProtRLSearch 与其他竞争方法在 BioMedMCQs [4] 和 ProtMCQs 上的整体表现。在这两个数据集的所有难度级别中，ProtRLSearch 均取得最高准确率，同时将推理成本控制在合理范围内，某些情况下甚至更低。在 BioMedM- CQs [4],上，ProtRLSearch 在级别 1、2 和 3 的准确率分别达到 89.2%、75.8% 和 71.7%，显著高于 45.8%、38.7% 和 29.5% 的基线结果。同时，其平均推理时间分别为 10.5 秒、13.4 秒和 21.3 秒，仅略高于基线的单步推理时间，且不会随着任务难度增加而呈指数增长，这表明引入多轮搜索并未造成显著的效率损失。相比之下，Search-R1 [6] 在相应级别的准确率分别为 76.1%、70.9% 和 63.5%，但推理时间明显更长。这些结果表明，通过整合多模态信息和多维奖励信号，ProtRLSearch 能够使多轮搜索更早收敛至有效的检索路径，并避免不必要的推理开销。在 ProtMCQs 上，ProtRLSearch 在级别 1–3 的准确率分别达到 86.9%、77.4% 和 72.5%，同样

显著优于 35.7%、30.5% 和 26.1% 的基线结果，凸显了在序列约束下引入搜索与规划的关键作用。在效率方面，相应的推理时间分别为 15.9 秒、20.4 秒和 24.3 秒。尽管高于基线的单步推理时间，但相较于性能提升幅度，这一增幅是可以接受的。ProtLLM [19] 得益于蛋白质序列建模，在该数据集上的表现优于 BioReason；然而，由于缺少搜索规划，在需要结合检索信息进行跨区域推理的高难度查询上，其表现仍明显逊于 ProtRLSearch。总体而言，这些结果进一步表明，基于多模态蛋白质序列编码的多轮搜索规划，能够在提升准确率与控制效率之间实现良好平衡。

表 III 展示了在 MedM- CQA [10]和 MedQA [22] 上的性能比较。ProtRLSearch 在两个数据集上均取得最高准确率，在 MedM- CQA [10] 上达到 90.4%，在 MedQA [22],上达到 87.4%，优于现有的生物医学语言模型和基于智能体的基线。这些结果表明，所提出的方法在一般医学问答场景中具有更强的推理稳定性和泛化能力。

C. 消融研究

为分析 RL 和多模态建模在 ProtRLSearch 中的独立贡献，我们在表 II 所示的 BioMedMCQs [4]和 ProtMCQs 数据集的三个难度级别上开展了消融实验。我们构建了 w/o Protein（移除蛋白质模态）和 w/o RL（移除 RL）两种设置，而完整设置同时包含这两个组件。如表 II 所示，在两个数据集的所有难度级别上，完整设置均取得了最高准确率，同时具有更低的 Token 消耗和检索推理时间。在 BioMedMCQs [4]，中，移除 RL 会使第 1 级准确率从 89.2% 降至 76.7%，并增加 Token 和时间成本；相比之下，移除蛋白质模态仅导致有限的性能下降，这表明在生物医学纯文本任务中，蛋白质序列并不是决定搜索与推理性能的关键因素。在 ProtMCQs 上，移除蛋白质模态会造成最大的性能下降，第 3 级准确率从 72.5% 降至 55.4%，同时 Token 消耗和推理时间均有所增加；相比之下，w/o RL 在所有难度级别上的准确率均低于完整设置，并伴随更高的推理成本。总体而言，完整设置在两个数据集和所有难度级别上均优于所有消融变体，表明引入 RL 和多模态建模有助于获得更加稳定的整体性能。表 III 中的消融结果显示，移除 RL 组件或蛋白质相关建模，都会导致 MedMCQA [10]和 MedQA [22]，上的准确率持续下降；这可能是由于其影响了搜索决策，从而降低了整体答案稳定性。

V. 结论

在蛋白质相关的生成与推理任务中，由于缺乏对蛋白质序列信息的显式建模以及实时检索能力，大语言模型往往会产生不稳定或事实不一致的输出，尤其是在疾病相关变异分析等医疗保健场景中。为解决这一局限性，我们提出了 ProtRLSearch——一种通过 RL 训练的多轮多模态蛋白质搜索智能体。ProtRLSearch 联合建模蛋白质序列表示与文本信息，将多轮搜索与大语言模型相结合，并在由规划器、检索器和执行器组成的闭环框架中，在多维奖励约束下学习稳定的搜索轨迹。实验结果表明，将蛋白质序列明确视为文本之外的互补模态，有助于实现更加稳定且一致的多轮蛋白质搜索与推理。未来的工作将探索自适应蛋白质表示学习和在线搜索监督，以提升模型对未见蛋白质家族及动态演化知识源的泛化能力。

参考文献

[1] Y. Xiao 等，《蛋白质大语言模型：全面综述》，arXiv:2502.17504，2025年。[2] L. Huang 等，《大语言模型幻觉综述：原理、分类、挑战与开放问题》，《ACM 信息系统汇刊》，第43卷第2期，第1–55页，2025年。[3] R. Apweiler 等，《UniProt：通用蛋白质知识库》，《核酸研究》，第51卷，第D523–D531页，2023年。[4] C. Liu 等，《BioMedSearch：基于大语言模型的多源生物医学检索框架》，arXiv:2510.13926，2025年。[5] Z. Chen 等，《MindSearch：模拟人类思维，激发深度人工智能搜索器》，2025年。[6] B. Jin 等，《Search-R1：通过强化学习训练大语言模型进行推理并利用搜索引擎》，arXiv:2503.09516，2025年。[7] H. Song 等，《R1-Searcher：通过强化学习激励大语言模型的搜索能力》，arXiv:2503.05592，2025年。[8] M. Huo 等，《多模态大语言模型实现蛋白质功能预测》，bioRxiv，2024年。[9] Y. Shen 等，《面向蛋白质理解的大语言模型微调数据集与基准测试》，arXiv:2406.05540，2024年。[10] A. Pal 等，《MedMCQA：用于医学问答的大规模多学科选择题数据集》，载于《健康、推理与学习会议论文集》，2022年。[11] S. Niu 等，《重新思考基于文本的蛋白质理解：检索还是大语言模型? 》，arXiv:2505.20354，2025年。[12] Y. Gao 等，《大语言模型的检索增强生成：综述》，arXiv:2312.10997，2024年。[13] Y. He 等，《PASA：用于全面学术论文搜索的大语言模型智能体》，2025年。[14] Z. Zhang 等，《蛋白质语言模型学习相互作用序列基序的进化统计》，《美国国家科学院院刊》，2024年。[15] L. Ouyang 等，《通过人类反馈训练语言模型遵循指令》，NeurIPS 论文集，第35卷，第27 730–27 744页，2022年。[16] A. Fallahpour 等，《BioReason：在 DNA-LLM 模型中激励多模态生物学推理》，arXiv:2505.23579，2025年。[17] H. Orgad 等，《大语言模型知道的比展现出来的更多：关于大语言模型幻觉的内在表示》，arXiv:2410.02707，2024年。[18] R. Ranjan 等，《检索增强生成（RAG）综合综述：演进、现状与未来方向》，arXiv:2410.12837，2024年。[19] L. Zhuo 等，《ProtLLM：采用蛋白质即词预训练的交错式蛋白质语言大语言模型》，arXiv:2403.079205，2024年。[20] Lin 等，《进化尺度上的蛋白质序列语言模型可实现准确的结构预测》，bioRxiv，2022年。

[21]Q. Han 等，《一种采用数据增强策略检测推文中药物提及的 PubMedBERT 分类器》，*arXiv:2112.02998*，2021年。[22] D. Jin 等，《Ffff》，应用科学，2021年。[23] M. Xu 等，《PEER：面向蛋白质序列理解的综合性多任务基准测试》，神经信息处理系统进展，第35卷，第35 156–35 173页，2022年。[24] A. Yang 等，《Qwen3 技术报告》，*arXiv:2505.09388*，2025年。[25] J. Lee 等，《BioBERT：一种预训练的生物医学语言表示模型》，生物信息学，2020年。[26] I. Beltagy 等，《SciBERT：一种面向科学文本的预训练语言模型》，*arXiv:1903.10676*，2019年。[27] T. Han 等，《MedAlpaca——开源医学对话式人工智能模型与训练数据集》，*arXiv:2304.08247*，2023年。[28] X. Tang 等，《MEDAGENTS：作为零样本医学推理协作者的大语言模型》，载于计算语言学协会：*ACL* 会议论文集，2024年。