

---

# X-LoRA：低等级适配器专家混合物，A 大型语言模型的灵活框架 蛋白质力学和分子生物学中的应用 设计

---

Eric L. Buehler\*

Markus J. Buehler†

mbuehler@MIT.EDU

## 摘要

我们报告了一种混合专家策略，利用基于低秩适配（LoRA）的深度层标记级方法创建微调大型语言模型。从一组预先训练好的 LoRA 适应器开始，我们的门控策略利用隐藏状态动态混合适应层，从而使生成的 X-LoRA 模型能够利用不同的能力，创造出前所未有的深度层级组合来解决任务。该设计的灵感来源于生物的普遍性和多样性原则，即神经网络构建模块可在不同层次的表现形式中重复使用。因此，X-LoRA 模型可以轻松地应用于任何现有的大型语言模型（LLM），而无需修改底层结构。我们开发了一个量身定制的 X-LoRA 模型，提供了包括正向/反向分析任务和增强推理能力在内的科学功能，重点关注生物材料分析、蛋白质力学和设计。这项工作的影响包括获得具有强大领域知识和跨领域知识整合能力的可随时扩展和适应的模型。我们邀请了生物学、数学、推理、生物启发材料、力学和材料、化学、蛋白质生物物理学、力学和基于量子力学的分子特性等领域的专家，开展了一系列以物理学为重点的案例研究。我们研究了知识回忆、蛋白质力学正演/反演任务、蛋白质设计、对抗代理建模（包括本体知识图谱构建）以及分子设计。该模型不仅能对蛋白质的纳米力学特性或量子力学分子特性进行定量预测，还能对结果进行推理，并正确预测解释不同分子行为的可能机制。

**关键词** 语言建模 - 科学人工智能 - 生物材料 - 蛋白质 - 逆问题 - 断裂 - 材料科学

## 1 导言

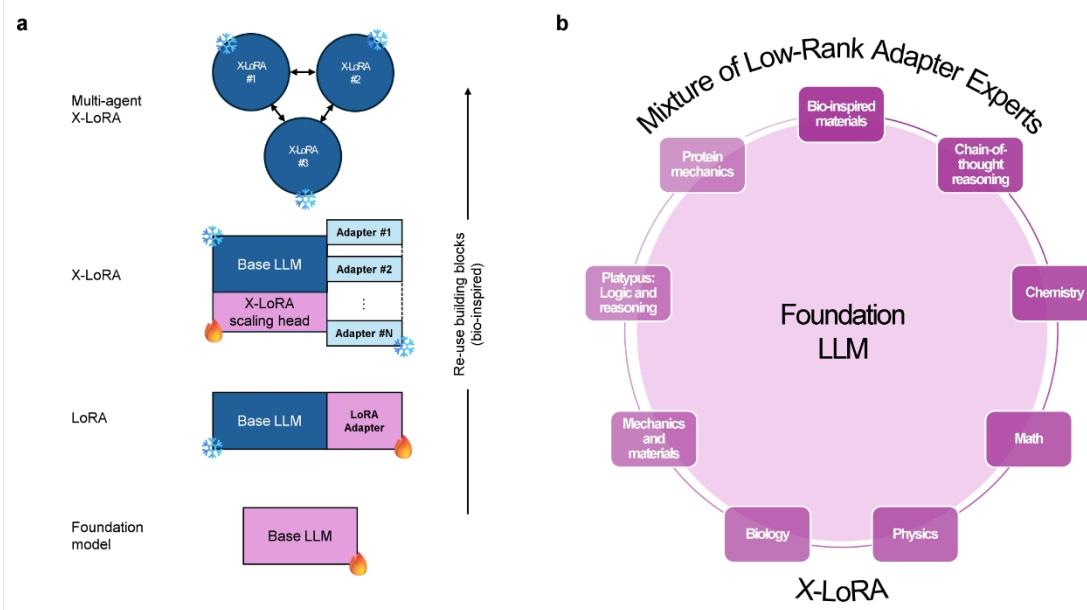
大型语言模型（LLMs）[1, 2, 3, 4, 5, 6]已经获得了极大的普及，包括在以下方面的发展。然而，训练这类模型的成本可能很高，尤其是当需要不同的能力集时。低阶适配器（Low-rank adapters，LoRA）[15] 等方法已被提出作为一种更有效的替代方法，但适配器通常侧重于较窄的知识领域。LoRA 建模的基本概念是使用低秩矩阵添加到原始全矩阵中，并选择这些低秩矩阵作为模型中唯一可训练的部分。由

于只有适配器层是可训练的，这些模型通常保留了基础模型预训练时获取的知识，同时使模型更适用于特定任务，而且在计算需求方面也很高效。由于 LoRA 适配器的训练效率很高，因此有可能利用这种策略开发出大量专用模型。

---

\*Academy for Science and Design (ASD), 9 Townsend W, Nashua, NH 03063, USA

<sup>†</sup>麻省理工学院 (MIT) , 77 Mass.Ave 1-165, Cambridge, MA 02139, USA



**图 1：**适应和代理神经网络架构的多层次设计原则，遵循生物启发范式，其基础是重新利用现有的构建模块。(a)，在示意图中，**蓝色表示冻结（不可训练）组件，红色表示可训练组件**。通过较小的可训练组件，重用和适应性的直观反映清晰可见。尽管这项工作尚未实施，但在多机器人层面也有可能添加可训练组件。作为这项工作的一部分，所进行的训练主要集中在模型的 LoRA 适配器和 X-LoRA 层面，因为预先训练的基础模型被用作训练的基础。(b), 用于构建 X-LoRA 模型的适配器集概览，其中包括生物启发材料、思维链 (CoT) 与推理、化学、数学、物理、力学与材料、逻辑与推理以及蛋白质力学方面的专家。

虽然 LoRA 适配器提供了一种开发专门模型能力的有效方法，但如何将多个此类系统集成到改进的模型中，从而提供一套综合甚至改进的能力，仍然是一个未决问题。事实上，混合 LLM 已成为一个有趣的研究领域[16]。其他实验工作，如使用球面内插模型参数 (SLERP) 和其他概念[17]，都显示出了有趣的结果和混合模型实现新功能的前景。以这些方式创建的混合模型通常是根据特定算法先验计算出来的，虽然它们在一些基准测试中表现出了良好的性能，但仍有必要对混合模型的合理发展以及某些特征是如何出现的进行更多研究。我们假设，参数的动态混合可以是一种更通用的方法，因为它允许混合模型在推理过程中探索新的组合。

相关方法采用了专家混合 (MoE) 策略[18, 19, 20]，包括最近发布的 Mixtral LLM [21]。然而，这些方法的计算成本可能很高，甚至在推理过程中也有很大的需求。在这里，我们提出了一种计算高效的专家混合方法，使用一组不同的 LoRA 适配器，这些适配器可以很容易地进行单独训练。我们的方法实施了一种高粒度和灵活的方法，其灵感来自于一种生物范式，即在构建从基础模型到代理系统 (图 1) 的交互模型层次中重复使用神经网络构建模块。为了从数量上确定其影响，我们的工作特别侧重于将这一策略用于以科学为重点的 LLM，在此特别应用于更广泛的生物启发材料分析和设计领域中的蛋白质力学问题。

我们提出了一种算法，可在标记级状态下动态混合适配器，并通过细粒度实现单个层级的适配器混合。这种方法被称为 X-LoRA，它提供了对多模态语言建模的科学应用至关重要的各种功能[22, 23, 24?]。

### 1.1 X-LoRA 的基本概念

我们首先简要回顾一下开发 LoRA 适配程序时使用的概念。低秩适配（LoRA）[15] 背后的基本策略假定权重更新的 "内在维度" 较低，并利用以下优势

通过冻结原始权重  $W_0 \in \mathbb{R}^{d \times k}$  并将更新限制为低阶分解

$$w_0 + \Delta w_0 = w_0 + ba \quad (1)$$

其中， $B \in \mathbb{R}^{d \times r}$ ， $A \in \mathbb{R}^{r \times d}$ ，秩  $r \ll \min(d, k)$ 。在使用 LoRA [15] 进行训练时，预训练的权重  $W_0$  被冻结，只有  $A$  和  $B$  矩阵保留可训练参数。LoRA 层的前向传递可以是用下式表示

$$h = W_0 x + \Delta W_0 x = W_0 x + BAx \quad (2)$$

我们注意到，在实际应用中，LoRA 适配器有一个固定的标量权重  $\alpha$ 。

$$h = W_0 x + BAx - \alpha \quad (3)$$

在此基础上，X-LoRA 基于使用一组 LoRA 适配器的想法，正如早期以科学为重点的 LLM 工作（如 [10]）所展示的那样，每个适配器都有独特的能力集。与静态混合或集成模型不同的是，我们创建了一种动态门控方法，可根据单个令牌和层的粒度扩展单个 LoRA 适配器，以促进模型内部的深度混合。换句话说，考虑到 LoRA 适配器  $i$ ，原始的  $\alpha_i$  参数乘以一个缩放值（用  $\lambda_i$  表示），以创建适配器  $\alpha^*$  的新缩放值：

$$\alpha_i^* = \alpha_i - \lambda_i \quad (4)$$

缩放值  $\lambda_i$  由 X-LoRA 缩放头预测，该缩放头利用模型的隐藏状态，形成 X-LoRA 模型中的可训练组件。由于缩放头可以利用复杂的编码，因此它可以作为一个或几个层的简单前馈神经网络高效地实现（在 X-LoRA 的实现过程中，用户可以指定构建神经网络的细节）。

缩放计算针对序列中的每个标记进行。因此，由于在 LLM 中特定层运行的每个 LoRA 适配器都会进行缩放，因此提供了较高的粒度。因此，我们将这种对适配器的各个组件进行门控的方法称为缩放。有关该方法的更多详情，请参阅 "材料与方法" 部分。

## 1.2 论文提纲

本文的计划如下。首先，我们将讨论包括相关数据集在内的方法和训练策略，从而开发出一种具有物理科学特殊能力的 X-LoRA 模型，尤其侧重于包括蛋白质在内的生物材料。然后，我们进行了一系列实验，将 X-LoRA 应用于各种任务，如问题解答、对话和代理建模、蛋白质设计和分析等。我们对缩放模式进行了详细分析，并通过与分子建模及其他物理数据和方法的比较对该方法进行了验证。

## 2 结果和讨论

X-LoRA 模型的开发过程包括以下步骤：

1. 培训基础 LLM（已在早期工作中完成）
2. 单独培训一组适应者，以培养他们在一系列领域的专业知识（这些领域反映了不同或重叠的技能和知识领域）
3. 训练综合 X-LoRA 模型

我们的实验从训练一系列 LoRA 适配器开始。我们以建立在 Mistral-7B 模型[5]之上的 Zephyr-7B- $\beta$  模型[25]为基础，开发了一套九个适配器，并根据不同的专长进行了微调：

1. 生物启发材料
2. 思维链（CoT）和推理
3. 化学
4. 数学
5. 物理学
6. 生物学

7. 力学与材料
8. 鸭嘴兽逻辑与推理
9. 蛋白质力学任务（序列到属性的生成和反演能力）

之所以选择这些重点领域和专业领域，是因为我们的最终目标是建立一个以科学、技术、工程和数学为重点、强调材料科学的法学硕士课程。不过，我们认为，我们的方法可以普遍应用于各种领域，特别是创建可以在不同知识领域之间建立联系的模式[8, 13]。

创建 LoRA 适配程序后，我们将对 X-LoRA 模型进行训练。为此，我们使用了用于训练单个 LoRA 适配程序的训练样本集合的子集（详见 "材料与方法"，包括有关数据集、实施等方面的详细说明）。

我们将 X-LoRA 系统作为一个实验实体，评估其在一些不同环境下的性能。我们进行了一系列实验，包括复杂问题解答，使用生成模型进行蛋白质设计、分析和推理，以及使用对抗代理建模创建更复杂的交互策略。在整个分析过程中，我们比较了基础模型和 X-LoRA 模型的性能，并跟踪了缩放头在生成或开发答案时如何利用不同专家的不同层次。

## 2.1 问题解答和观察到的 X-LoRA 层上缩放权重

我们首先研究 X-LoRA 模型与基础模型之间的差异。我们向这两种模型都提出了这样一个问题：

什么材料行为（硬化与软化）会导致模式 II 裂纹产生超音速裂纹？逐步思考。

基本模型的响应是

由于裂纹顶端附近材料的增硬行为，在模式 II 裂纹中会出现超音速裂纹，也称为瑞利波。在模式 II 断裂中，载荷以剪切为主，裂纹垂直于施加应力的方向扩展。下面是逐步解释：

当裂纹尖端接近材料时，附近的应力会迅速增加。这就形成了一个高应力集中区域，称为应力奇点。

在这种应力奇异点中，材料会发生弹性变形，从而导致刚度增加，即出现僵化行为。产生这种僵化效应的原因是，当裂纹接近裂纹尖端时，裂纹尖端附近的材料会变得更加受约束[.....]

利用我们自己的领域知识[26]，这个答案不仅不正确，而且还很啰嗦。相比之下，X-LoRA 模型的回答如下：

在模式 II 裂纹中，导致超音速裂纹的材料行为是硬化。

这个答案正确、清晰、简洁 [26]。

对于专门的蛋白质分析和设计任务来说，基础模型和 X-LoRA 模型之间的区别更加明显。给出的任务是

```
CalculateForceHistory< M S K S N T Y R M L V L E D D T K I N K E D E K F L K G K P  
G K M H E F V D E L I L P F N V D E L D E L N T W F D K F D A E I C I P Y V E I L  
K E S G M K >
```

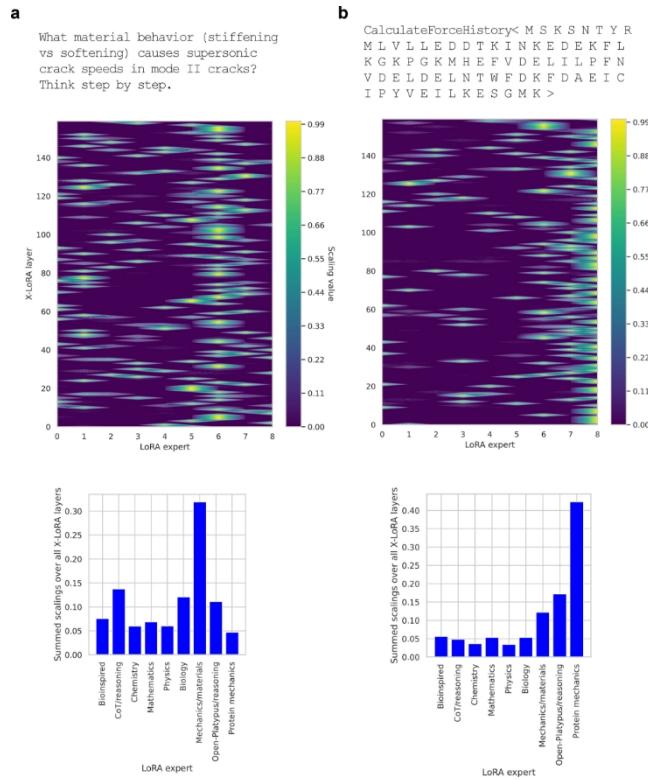
基础模型的响应为

很抱歉，您提供的输入不是有效的数学表达式。它似乎是一串字符，无法构成一个可计算的函数。

X-LoRA：低等级适配器专家混合物

请提供一个有效的数学表达式供我计算。

不过，我可以告诉你，你提供的字符序列实际上是



**图 2：用于比较 X-LoRA 和基础模型的试题的问题回答结果和观察到的 X-LoRA 缩放权重（a 和 b 面板显示了模型解决的两个不同任务的结果）。在每个子面板的顶部，我们显示了所提出的问题，然后分别是 X-LoRA 层和 LoRA 专家的 X-LoRA 缩放权重图。下部的柱状图汇总了所有层的缩放权重，显示了最常用的适配器的整体衡量标准。面板 a 是一个关于动态断裂的问题，大量使用了机械/材料专家。相反，如图 b 所示，蛋白质分析任务则使用了蛋白质力学适配器。据观察，每种情况下都使用了复杂的缩放值模式，这表明 X-LoRA 模型利用了不同适配器在各层间异质混合的优势。由于缩放权重决定了对给定输入使用哪个专家，热图显示了这一决定在模型不同层之间的变化，揭示了巨大的异质性。**

一系列函数和变量名，可能是特定编程语言中更大程序或脚本的一部分。如果没有上下文，就无法确定其计算意图。

如果您能提供更多信息，例如该表达式出现的编程语言和上下文，我或许能更深入地了解其含义和可能的计算方法。

X-LoRA 模型的响应如下，正确完成了生成与该蛋白质相关的力-变形曲线的设计任务：

```
[0.013, 0.066, 0.078, 0.045, 0.072, 0.126, 0.127, 0.107, 0.119, 0.074, 0.106, 0.098, 0.103, 0.128, 0.077, 0.069, 0.131, 0.108, 0.144, 0.174, 0.166, 0.183, 0.185, 0.217, 0.175, 0.230, 0.187, 0.224, 0.267, 0.301, 0.315]
```

图 2 显示了这一问题解答的结果和观察到的 X-LoRA 缩放权重。两个查询，一个侧重于力学和材料（图 2a），另一个侧重于蛋白质力学（图 2b），显示出截然不同的模式。重要的是，我们发现每种情况下都使用了复杂的缩放值模式，这表明 X-LoRA 模型利用了不同适配器跨层异质混合的优势。我们进一步发现，图 2a

## X-LoRA：低等级适配器专家混合物

中的热图在许多专家和层之间具有更广泛的激活范围，其中一些区域的激活程度较高。图 2b 中的热图显示了更大的

我们认为，这可能表明右侧的门控功能对专家的激活更具选择性。我们认为，这可能表明右侧的门控功能在激活专家时更具选择性。

在图 2 中，类似亮线的区域代表高激活路径，这意味着缩放函数对特定层的某些专家激活较多。观察到的模式表明，所有专家和层的重要性并不是均匀分布的；相反，它往往是稀疏和有选择性的。这种稀疏性与在其他专家混合物模型中观察到的情况类似，即对于任何给定的输入，只选择专家的一个子集来专门处理问题空间的不同部分。通过比较不同任务的图谱，稀疏性和模式对于进一步了解模型在处理不同类型的输入或神经网络的不同处理阶段时依赖哪些专家非常重要（我们将在后面讨论图谱在解决复杂任务过程中的动态变化）。

为了进一步分析这种行为，我们现在探讨了更多的问题，以确定 X-LoRA 模型在一组不同的提示下如何使用不同的专家。我们特别想知道 X-LoRA 模型是否在不同技能或领域重叠的问题中混合使用不同的适配器，如果是，又是如何混合使用的。图 3 显示了问题解答的结果和观察到的 X-LoRA 扩展权重，以及对表 2.1 所列问题的回答。在表 2.1 中，我们同时显示了 X-LoRA 的回答和基础模型 *Zephyr-7B-β* 的回答。两个模型之间的差异非常明显。为了清楚直观地显示这一点，我们用红色高亮标记了不正确或有问题的回答。对比结果表明，X-LoRA 的准确度要高得多，而且答案也更加简洁。

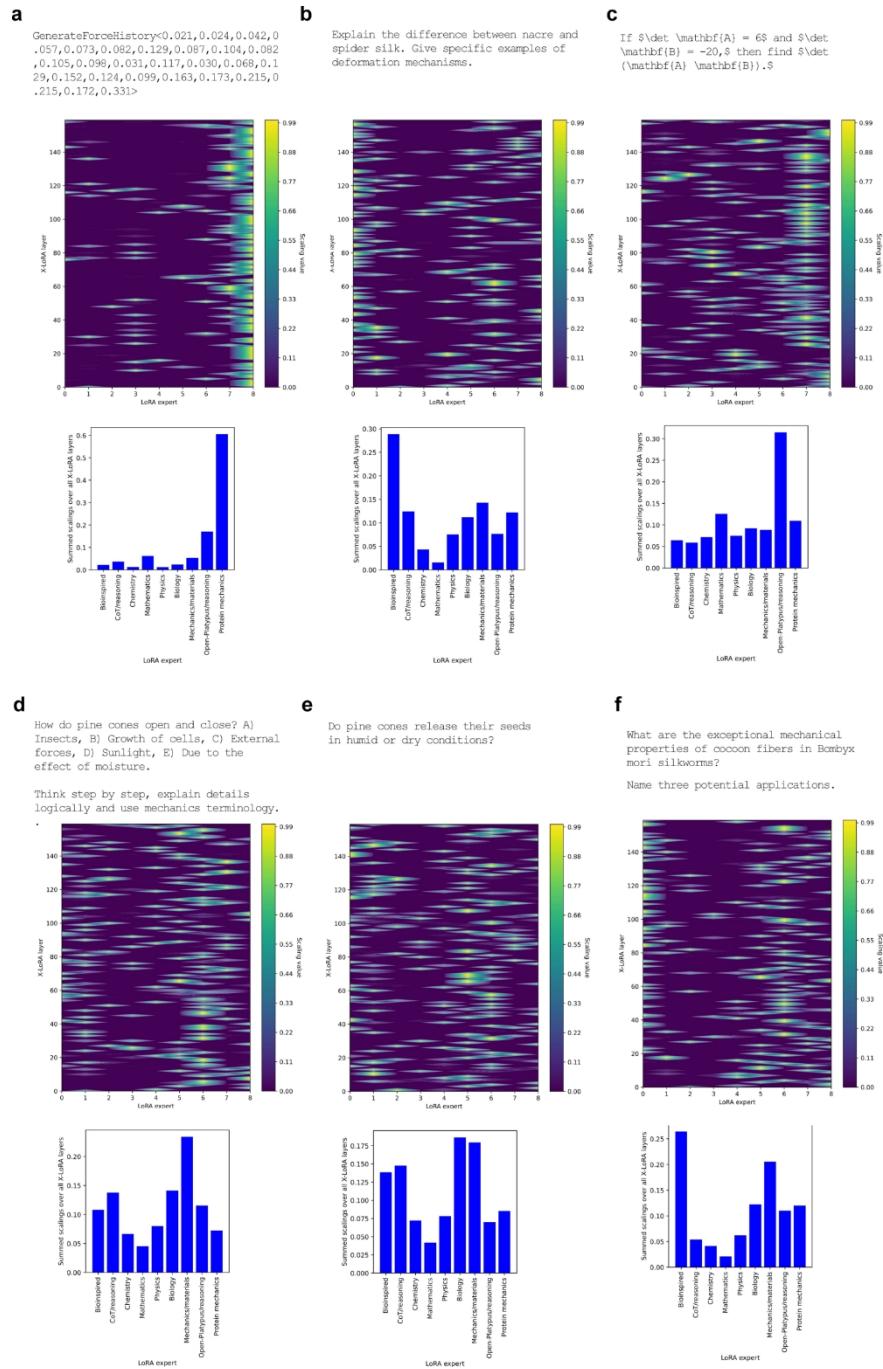
正如我们所假设的那样，我们观察到适配器之间复杂的混合，而且往往会激活几个占主导地位的 LoRA 专家。例如，图 3d 揭示了一个关于松果力学的问题，它导致了力学/材料适配器与 CoT/推理和生物启发适配器的结合使用。虽然 X-LoRA 选出了正确的字母，但答案并不完全正确，因为松果是在干燥而非潮湿的条件下打开的。不过，为了进一步探究松果是在干燥还是潮湿条件下释放，后续问题得出了正确答案。在一个以蛋白质为重点的问题中，图 3f 显示了在回答与桑蚕丝力学有关的问题时适配器的混合情况，激活了生物启发、力学/材料和生物学适配器的组合。对于像蛋白质设计（图 3a）这样专业性很强的任务，我们可以看到其中一个专家基本上是单独激活的。

对图 3a 所示的热图标度进行深入分析后发现，除了蛋白质力学专家的相关性明显较高外，还有几个带状区域的某些专家在多个层级中都具有较高的重要性。这一点在力学/材料和推理专家身上尤为明显。这些条带并不是连续的，而是呈现出不同的模式，这可能暗示了这些专家有效处理的特定条件或特征。事实上，推理专家（紧挨着蛋白质力学专家的左侧）有几个层的激活度最高，尤其是在上层（约 100 到 140）。我们认为，这可能意味着推理专家在模型的最终决策或输出中发挥了重要作用，尽管还需要更多的研究来证实这一点。物理和生物专家在某些层也有更局部但更强的激活，这可能表明该专家专门研究这些区域的特征或表征。

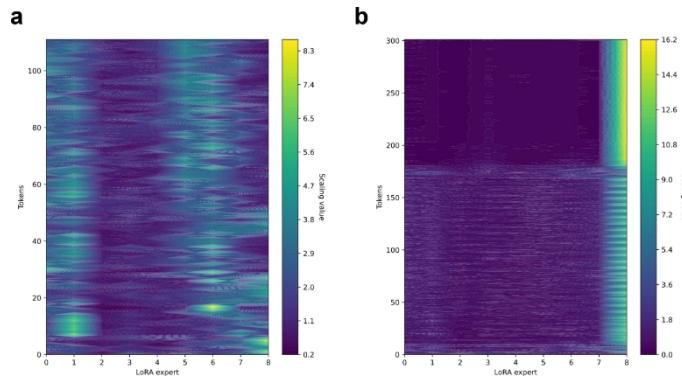
此外，图 4 描述了所有层的总和缩放的标记级历史。这一分析揭示了两个例子的历史，图 4a 是与松果种子释放有关的任务，图 4b 是一系列蛋白质力学任务的结果。蛋白质力学任务涉及一系列提示-回答互动：首先是两个属性计算，最后是设计新蛋白质的生成任务。虽然专家的使用总体上比较稳定，但在这一过程中，LoRA 专家的使用发生了明显的变化。具体来说，蛋白质力学专家在设计新蛋白质的最后生成阶段高度集中。

表 2.1 显示了不同领域的各种用例，以及与基础模型本身性能的比较。我们将对其中几个案例进行更深入的讨论。

例如，我们提出了一个涉及蛋白质科学、生物学和逻辑学的问题，询问蛋白质的表达模式。X-LoRA 得到了正确的答案。在分析每种细胞类型中标记蛋白的表达时，基础模型的回答存在一些不准确和逻辑不一致的地方。基础模型正确指出成纤维细胞表达蛋白 B 而不表达蛋白 A，这是因为蛋白 A 和蛋白 B 具有互斥性。鉴于没有规则禁止成纤维细胞在表达蛋白 B 的同时表达蛋白 C，而且成纤维细胞对蛋白 C 也没有限制，因此它们也会顺理成章地表达蛋白 C。对于神经元来说，由于互斥规则的存在，基础模型一开始就正确地指出神经元表达蛋白 A 而不表达蛋白 B。然而，关于神经元表达蛋白 C 或同时表达蛋白 C 和蛋白 B 的可能性的讨论令人困惑，而且包含一个错误。既然神经元表达蛋白 A 而不表达蛋白 B（如前所述），那么根据规则，唯一的其他选择就是神经元表达蛋白 C。因此，神经元表达的是蛋白 A 和蛋白 C，而不是蛋白 B。



**图 3：**问题解答结果和观察到的 X-LoRA 缩放权重（a-f 面板显示了模型解决不同任务的结果，涵盖了一系列领域和类型）。在每个子面板的顶部，我们显示了所提出的问题，然后分别绘制了 X-LoRA 层和 LoRA 专家的 X-LoRA 缩放权重。下部的柱状图汇总了所有层的缩放权重，显示了最常用的适配器的整体衡量标准。与之前一样，我们可以观察到在每种情况下都使用了复杂的缩放值模式，这表明 X-LoRA 模型利用了不同适配器在各层间异质混合的优势。



**图 4：**作为标记历史函数的所有层的总和缩放。该图显示了两个示例的历史记录，a 面板中是与松果相关的任务（共约 110 个标记），b 面板中是一系列蛋白质力学任务的结果（约 300 个标记；首先是属性计算，然后是设计新蛋白质的生成任务（接近尾声））。虽然专家的使用总体上比较稳定，但在这一过程中，LoRA 专家的使用发生了明显的变化。专家编号（0 至 8）反映了与早期图表中相同的组织结构，例如，请参考图 3 中的标签。

然而，给出的解释错综复杂，其中包括关于蛋白 A 和蛋白 C 相互排斥的错误说法，而这并不是最初观察结果的一部分。简单的推理是，上皮细胞不表达蛋白 A，但必须表达其余两种蛋白，即蛋白 B 和蛋白 C，因为最初的指令只规定蛋白 A 和蛋白 B 不能同时表达。相比之下，X-LoRA 能根据提供的观察结果和逻辑推理，正确识别每种细胞类型表达的标记蛋白。对于成纤维细胞，X-LoRA 正确推断出成纤维细胞不能表达蛋白 A，因为蛋白 A 的存在排除了蛋白 B 的表达。由于成纤维细胞确实表达蛋白 B，而且每种细胞类型都表达三种标记蛋白的独特组合（这是一个添加的不正确假设），因为最初的陈述并没有说明每种细胞类型表达的蛋白数量，而是它们表达的独特组合），因此根据所提供的信息，成纤维细胞也表达蛋白 C 的结论是正确的（除了关于蛋白数量的部分，这似乎是一个误解）。推理合理，指出蛋白 B 不能与蛋白 A 同时表达，既然神经元表达蛋白 A 和另一种标记蛋白（不是蛋白 B），那么它们一定表达蛋白 C。最后，对于上皮细胞，X-LoRA 模型准确地得出了上皮细胞表达蛋白 B 和蛋白 C 的结论。这是因为上皮细胞不表达蛋白 A，而且通过排除法，鉴于上皮细胞表达两种标记蛋白，这两种标记蛋白必须是蛋白 B 和蛋白 C。

在关于潮湿和干燥条件下松果释放量的问题中，基础模型给出了一个与事实不符的复杂答案。X-LoRA 则给出了一个推理清晰的精确答案。还有其他相关的例子，例如在有关蚕丝的特殊机械特性的问题中，基础模型错误地断言蚕丝纤维会导电。该模型在解决其他几个特定领域的任务时也出现失误，例如在有关超弹性在最大断裂速度中的作用的问题中，X-LoRA 提供了简明准确的答案，而基础模型却无法回答。在一个关于壁虎脚的强粘附性的问题中，X-LoRA 正确地指出了刚毛和纳米级效应的重要性，而基础模型却对特殊用途胶蛋白在这一过程中的作用提出了错误的说法。表中和本文其他部分讨论的这些例子和其他例子有力地证明了 X-LoRA 比基础模型更强的能力。

**表 1：**生物材料、材料力学、蛋白质特性、逻辑推理及相关领域的问答对摘要，重点是具有挑战性的问题。

X-LoRA 和 Zephyr-7B- $\beta$  基础模型提出并回答了几个示例问题，以供比较。图 3 显示了一些查询的相关 X-LoRA 缩放值，以说明该模型如何能够根据上下文混合使用不同的适配器和相关能力。我们用 **红色高亮标记**

## X-LoRA：低等级适配器专家混合物

了不正确或有问题的响应。对比结果表明，X-LoRA 的整体准确率要高得多，而且回答也更简洁。

问题	X-LoRA 模型的响应	Zephyr-7B-β 基本模型响应
		下一页

表 1 - 续前页

问题	X-LoRA 合集	Zephyr/Beta 基本型合集
生成力历史记录 <0.021,0.024,0.042,0.057,0.073,0.082,0.129,0.087,0.104,0.082,0.105,0.098,0.031,0.117,0.030,0.068,0.129,0.152,0.124,0.099,0.163,0.173,0.215,0.215,0.172,0.331>	[msksntrymlvilleddtkinkedekllkgpkghmefvdeilpfnvdelntwfdkfdaeicipyveilkesgmk]	<p>所提供的数字序列代表武力的历史测量结果随时间变化。以下是一种可能的解释：</p> <ul style="list-style-type: none"> <li>前七个值 (0.021, 0.024, ..., 0.082) 代表不同时间点的力读数。这些值看起来相对较小，并围绕一个较低的平均值波动。</li> <li>接下来的三个数值 (0.129, 0.087, 0.104) 表明，“力，然后下降，再上升。这可能是表明某种周期性行为或趋势。”</li> <li>在此之后，还会有几次波动，其中一些波动值较大。(0.115, 0.098, ..., 0.173)。目前还不清楚这些数值是否代表测量过程中的重大变化或噪音。</li> <li>最后，最后四个数值 (0.215, 0.215, 0.172, 0.331) 显示了力量急剧增加，可能表明发生了重大事件或异常。然而，仅从这些数据中还看不出是什么原因造成了这些异常现象。这个峰值。更多的背景资料和额外的测量结果将有助于要充分了解这里发生的事情，就需要</li> </ul>
解释珍珠质和蜘蛛丝的区别。举出变形机制的具体例子。	<p>珍珠质又称珍珠母，是一种生物矿化物质，存在于牡蛎和珍珠贝等软体动物的外壳中。它由一层层文石小板组成，由蛋白质和多糖组成的有机基质将其固定在一起。珍珠质的强度和韧性来自于它的分层结构，其中包括纳米级文石晶体与有机夹层砖墙式排列。这种结构可以在变形过程中有效地传递载荷和消散能量。</p> <p>而蛛丝则是蜘蛛为捕捉猎物或结网等各种目的而生产的一种蛋白质纤维。蛛丝由纤维蛋白组成，形成细长的纤维，具有很高的拉伸强度和弹性。蛛丝的机械特性得益于它的半晶体结构，其中包括β片状晶体和无定形区域。蛛丝具有极佳的韧性，在断裂前可发生明显的变形。</p> <p>就变形机制而言，珍珠质和蛛丝都表现出塑性变形，这意味着它们可以永久变形而不会断裂。然而，这种变形背后的机制却各不相同。在珍珠质中，塑性变形是通过文石板块和有机基质之间的界面脱粘以及界面处的裂纹偏转发生的。这些机制可以消散能量和重新分配应力，从而提高韧性。</p> <p>与此相反，蛛丝的塑性变形是通过β-片晶相互滑动实现的，这种滑动可以沿着纤维方向发生，也可以横向发生。这种滑动机制可产生较大的应变而不会断裂，这也是蛛丝具有特殊韧性的原因之一。此外，蛛丝中无定形区域的存在有利于变形过程中的分子滑动，进一步增强了其可塑性。</p> <p>总之，虽然珍珠质和蛛丝都表现出塑性变形，但其基本机制却各不相同，反映了它们不同的结构和组成。</p>	<p>珍珠质又称珍珠母，是一种存在于牡蛎、蛤蜊和贻贝等软体动物外壳中的生物材料。另一方面，蛛丝是一种蛋白质纤维，由蜘蛛生产，用于织网和其他用途。虽然这两种材料都具有出色的机械性能，但它们在结构和变形机制方面却有很大不同。</p> <p>结构：珍珠质由碳酸钙 (CaCO<sub>3</sub>) 制成的薄片（称为片状晶体）组成，层层排列，中间夹着有机物质（主要是蛋白质）。这种分层结构赋予珍珠质特有的彩虹色。相比之下，蛛丝由纤维蛋白组成，这种蛋白质形成的纤维具有独特的层次结构。纤维由纳米级的纤丝组成，这些纤丝又进一步组织成更大的纤维。</p> <p>变形机制：珍珠质和蛛丝的变形行为因其不同的结构而大相径庭。当受到拉力时，珍珠质会表现出独特的脆性断裂行为。随着荷载的增加，片状晶体相互分离，导致裂缝在材料中扩展。然而，板块之间的有机基质阻止了材料的完全破坏，从而产生了相对较高的韧性。珍珠钮扣遇水或受潮后会出现脆性，片状晶体脱层并从钮扣表面脱离，这就是这种行为的一个例子。</p> <p>另一方面，蛛丝在拉力作用下具有高弹性。蛛丝中的纤维蛋白分子形成延伸的贝塔片，为这种材料提供了非凡的强度和刚度。当材料被拉伸时，这些薄片相互滑动，使纤维得以延伸而不会断裂。这种滑动机制会产生很大的应变效率，使蛛丝的强度比同等重量的钢材高出数倍。蜘蛛网在受到冲击时能够伸展和反冲，使蜘蛛能够捕获猎物，这就是这种行为的一个例子。</p> <p>总之，虽然珍珠质和蛛丝都具有重要的机械特性，但它们在结构和变形机制上的差异导致了它们在拉力作用下的独特行为。珍珠质的分层结构使其能够在变形过程中吸收能量，而蛛丝的分层结构使其能够拉伸和反冲而不断裂。这些特性使这两种材料成为开发具有定制机械特性的新型生物启发材料的理想候选材料，可用于各种应用领域。</p>
如果 $\det A=6$ 和 $\det B=-20$ ，求 $\det(AB)$ 。	两个矩阵的乘积的行列式等于它们的行列式的乘积，因此： $\det(AB) = \det A \det B = 6(-20) = -120$ 。	利用乘积的行列式等于行列式的乘积这一性质（前提是矩阵相通），我们可以得出 $\det(AB) = \det(A) \det(B) = 6(-20) = -120$ 。

下一页

表 1 - 续前页

问题	X-LoRA 答案	Zephyr/B-5 基本型号 答案
松果是如何打开和闭合的？ A) 昆虫 B) 细胞的生长 C) 外力 D) 阳光 E) 由于水分的影响  循序渐进地思考，合乎逻辑地解释去尾，并使用术语。	松果的开合受水分的影响。鳞片当湿气进入该区域时，锥体上的微小铰链会使锥体张开。这是一种根据环境条件而发生的生物力学运动。	松果不会因阳光或昆虫的影响而开合。 相反，它们会在水分和外力的作用下开合。这个过程被称为锥体缩放。  具体操作如下 <ol style="list-style-type: none"> <li>1. 闭合球果闭合松果的鳞片紧密重叠，在里面的种子周围形成一个紧凑的结构。这有助于防止水分流失，保护种子不受天敌侵害。</li> <li>2. 湿气会触发鳞片张开：当有足够的水分时，比如降雨后或潮湿的天气，球果上的鳞片就会开始轻微分离。这样可以让空气流通，促进真菌生长，有助于分解种皮，为发芽做好准备。</li> <li>3. 外力有助于打开：随着鳞片的不断分离，重力和风也会在完全打开圆锥体方面发挥作用。这使种子暴露在外界环境中，增加了它们通过风或动物运输散播的机会。</li> <li>4. 闭合：一旦种子散落或掉落到地面上，锥体开始变干，鳞片再次合拢，对剩余的种子形成严密的密封。这有助于防止水分进一步流失并保护种子，直到下一季条件有利于发芽。</li> </ol> 总之，松果的开合是一个由环境因素和机械力驱动的复杂过程。它既能使种子有效传播，又能保护种子不受伤害，直到出现成功发芽的适当条件。
松果是否会在雨季释放种子？中雨还是暴雨？	松果在干燥的条件下释放种子。当周围的空气中雨还是暴雨？	松果一般会在以下时间内释放出种子（也称松子）：I. 对水分和温度的反应。许多松树品种的球果只有在温暖潮湿的条件下才会打开并释放种子。这是因为球果内的种子需要充足的水分才能发芽并长成新的松树。 不过，有些松树品种已经适应了在较干燥的条件下释放种子，尤其是在冬季降雨稀少的地中海气候地区。这些松果可能会一直闭合，直到降雨量较大时才会打开并释放种子。总之，松果释放种子的时间和条件会因松树的具体种类而有很大不同。
有哪些特殊的机械 蚕茧纤维的特性？ 请说出三种可能的应用。	蚕茧纤维的特殊机械特性 蚕茧的纤维具有很高的拉伸强度、弹性和韧性。这些特性得益于纤维的独特结构，它由结晶区和无定形区组成。结晶区提供强度，而无定形区则提供弹性。  这些特殊性能的三个潜在应用领域包括为航空航天和汽车工业开发轻质高强的复合材料。蚕茧纤维的高强度重量比使其成为复合材料的理想候选材料。  为包装和纺织品创造可生物降解的环保材料。茧纤维的天然来源使其成为对环境有害的合成材料的一种有吸引力的替代品。利用蚕茧纤维的独特性能设计新型给药系统。从这些纤维中控制药物的释放可带来更有效、更有针对性的疗法。	蚕茧纤维的特殊机械特性 蚕丝纤维的独特成分和结构决定了它的独特性。蚕丝纤维主要由两种蛋白质组成：纤维蛋白和丝胶蛋白。纤维蛋白是赋予蚕丝强度和韧性的主要成分，而丝胶则是一种粘合剂。  1. 强度：蚕丝纤维的抗拉强度高达 1.3 GPa，高于尼龙和聚酯等许多合成纤维。这种强度得益于纤维蛋白的结晶结构，即蛋白质分子的规则排列。 2. 韧性丝纤维还具有很高的韧性，韧性是衡量一种材料在断裂前吸收能量的能力。丝纤维在断裂前可吸收原长度的 35%，高于许多合成纤维。 3. 弹性：蚕丝纤维还具有高弹性，这意味着它们可以被拉伸，然后恢复到原来的形状而不会断裂。这种弹性是由于纤维蛋白中存在无定形区域，使纤维具有一定的柔韧性。  基于丝纤维的生物启发材料有三种潜在应用： <ol style="list-style-type: none"> <li>1. 生物医学应用：蚕丝纤维的生物相容性和生物降解能力使其成为生物医学应用中极具吸引力的材料，如伤口愈合敷料、缝合线和给药系统。</li> <li>2. 纺织品：丝纤维的高强度和韧性使其成为纺织品中合成纤维的潜在替代品，特别是在运动服和防护装备等高性能应用领域。</li> <li>3. 电子学：蚕丝纤维独特的电气特性（如导电能力）使其成为传感器和柔性显示器等电子设备的潜在材料。</li> </ol>

表 1 - 续前页

问题	X-LoRA 合集	Zephyr/B-5 基本型号 合集
超弹性对脆性材料的动态断裂，特别是最大裂纹速度。	<p>超弹性对脆性材料动态断裂的作用是 对最大裂纹速度的影响很大。在线性弹性材料中，最大裂纹速度由材料中的声速决定。然而，当存在超弹性时，最大裂纹速度会明显高于音速，接近超音速。这是由于超弹性材料在大应变时会产生刚化效应，从而使裂纹达到更高的速度。动态裂纹速度由裂纹驱动力和材料刚度之间的平衡决定，而超弹性对这两个因素都有很大影响。</p>	<p>超弹性，即材料的非线性弹性行为，会对脆性材料的动态断裂产生重大影响，尤其是最大裂纹速度。在动态断裂中，由于弹性能量的释放，裂纹会以很高的速度扩展。<b>最大裂纹速度受限于材料通过塑性变形或断裂消散能量的能力。</b></p> <p>在脆性材料中，由于不存在塑性变形，最大裂纹速度通常低于韧性材料。然而，超弹性可在断裂前产生较大变形，从而提高最大裂纹速度。这是因为与线性弹性材料相比，超弹性材料在单位体积内可以储存和释放更多的能量，从而可以延迟断裂的发生。</p> <p>超弹性对动态断裂的影响可以用裂纹速度-裂纹张开位移 (COD) 曲线来表示。COD 是裂纹扩展时裂纹表面之间的位移。该曲线显示了裂纹速度与断裂所需的 COD 之间的关系。在超弹性材料中，断裂所需的 COD 要高于线性弹性材料，这就导致了更高的最大裂纹速度。</p> <p>总之，超弹性可使脆性材料在断裂前产生更大的变形，从而提高其最大裂纹速度，这就推迟了断裂的开始时间，并增加了断裂所需的能量。这对于理解脆性材料在动态加载条件下的行为非常重要。</p>
为什么原始纤维素纳米晶 (CNC) 薄膜是否与骨骼或螳螂虾俱乐部等坚韧的生物材料相媲美？  A) 缺乏增塑剂 B) 脆性过大 C) 拉伸模量高  循序渐进地思考。	<p>原始纤维素纳米晶 (CNC) 薄膜的特性并不是由于其脆性过大，无法与骨骼或螳螂虾俱乐部等坚韧的生物复合材料相媲美。CNC 薄膜具有较高的拉伸模量，但由于缺乏塑性，在应力作用下容易破裂。为了改善 CNC 薄膜的机械性能，通常会添加增塑剂来降低其脆性并提高其韧性。</p>	<p>虽然原始纤维素纳米晶 (CNC) 薄膜表现出很高的拉伸强度，但由于两个主要原因，它们的性能无法与骨骼或螳螂虾俱乐部等坚韧的生物复合材料相提并论：缺乏增塑剂：骨和螳螂虾等生物复合材料含有称为增塑剂的有机分子，有助于减少纳米级构件之间的分子间相互作用，从而增加分子的流动性，实现塑性变形。相比之下，数控薄膜缺乏这种增塑剂，因此脆性过大，缺乏塑性变形。</p> <p>脆性过大：CNC 薄膜的高拉伸模量是由于纳米晶体之间强烈的分子间相互作用。然而，这也会导致脆性过大，因为这种材料缺乏在负载下发生塑性变形的能力。另一方面，骨骼和螳螂虾俱乐部等生物复合材料则结合了强键和弱键，从而实现了刚度和延展性之间的平衡。</p> <p>总之，棱纹 CNC 薄膜缺乏增塑剂和脆性过大是其性能无法与骨骼或螳螂虾俱乐部等坚韧的生物复合材料相媲美的主要原因。</p>
一位科学家正在研究三种不同的细胞类型：成纤维细胞、神经元和上皮细胞。每种细胞类型都表达三种标记蛋白的独特组合：根据以下观察结果，确定每种细胞类型表达的标记蛋白。  1. 当蛋白质 A 存在时，蛋白质 B 无法表达。 2. 成纤维细胞表达蛋白质 B。 3. 神经元表达蛋白 A 和另一种标记蛋白，但不表达蛋白 B。 4. 上皮细胞不表达 Protein A，但表达另外两种标记蛋白。  确定成纤维细胞、神经元和上皮细胞表达的标记蛋白，并为您的确定提供合理解释。	<p>成纤维细胞不能表达蛋白 A，因为当蛋白 A 存在时，蛋白 B 无法表达。因此不能表达蛋白 B。然而，成纤维细胞能表达蛋白 B，所以它们也一定能表达蛋白 C，因为每种细胞类型都能表达三种标记蛋白的独特组合。</p> <p>神经元表达蛋白 A 和另一种标记蛋白，但不表达蛋白 B。由于蛋白 A 存在时蛋白 B 无法表达，因此神经元必须表达蛋白 C。</p> <p>上皮细胞不表达蛋白 A，但表达另外两种标记蛋白。由于每种细胞类型都表达三种标记蛋白的独特组合，上皮细胞必须表达蛋白 B 和蛋白 C。</p> <p>总之，成纤维细胞表达蛋白 B 和蛋白 C，神经元表达蛋白 A 和蛋白 C，上皮细胞表达蛋白 B 和蛋白 C。</p>	<p>如观察结果 2 所述，成纤维细胞会表达蛋白 B。由于蛋白 A 和蛋白 B 不能同时表达（观察结果 1），成纤维细胞不表达蛋白 A。</p> <p>如观察结果 3 所述，神经元表达蛋白 A。由于神经元不表达蛋白 B（如蛋白 A 和蛋白 B 不能一起表达），它们必须表达蛋白 C。</p> <p>然而，观察结果 4 指出，上皮细胞同时表达蛋白 C 和蛋白 B，但不表达蛋白 A。既然神经元不表达蛋白 A，那么它们也不可能表达蛋白 B，因为这与上皮细胞的观察结果相冲突。因此，神经元表达蛋白 C，但不表达蛋白 B。</p> <p>根据观察结果 4，上皮细胞不表达蛋白 A。由于上皮细胞同时表达蛋白 C 和蛋白 B，这意味着蛋白 A 和蛋白 C 不可能同时表达（否则，上皮细胞就会同时表达蛋白 A 和蛋白 C，但它们并不表达）。因此，蛋白 A 和蛋白 C 是这些细胞类型的互斥标记。这也意味着上皮细胞不能同时表达蛋白 A 和蛋白 B，因为这与成纤维细胞不能同时表达蛋白 B 和蛋白 A 的观察结果相冲突。因此，上皮细胞表达蛋白 B，而不表达蛋白 A。</p> <p>总之，成纤维细胞表达蛋白 B，神经元表达蛋白 C，而上皮细胞表达蛋白 B，但不表达蛋白 A。这些发现的逻辑解释基于在特定条件下观察到的某些标记蛋白的互斥性。</p>



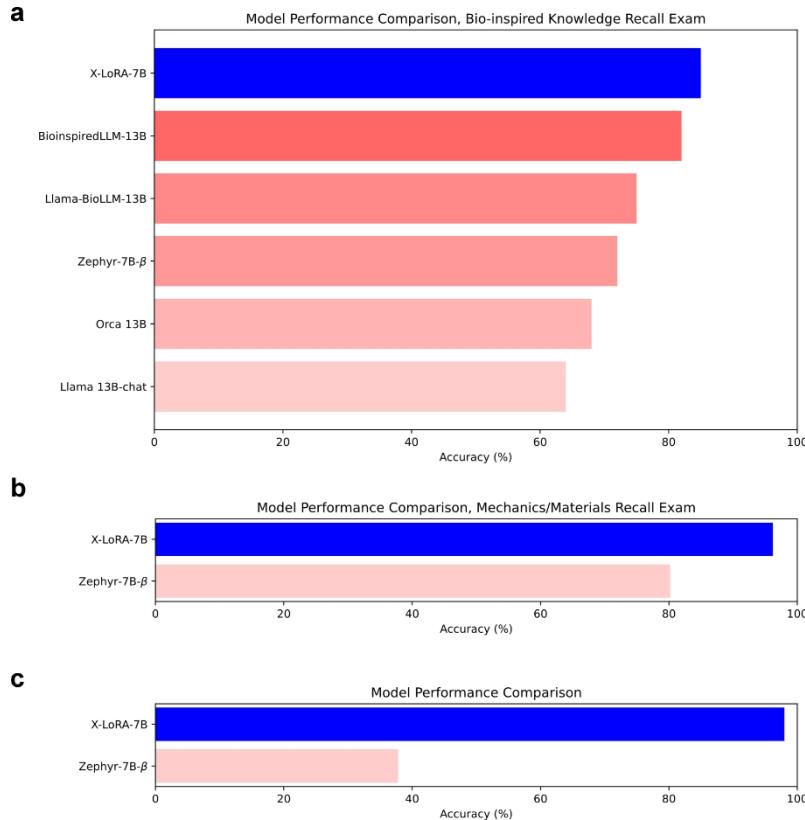
表 1 - 续前页

问题	X-LoRA 答案	Zephyr/B-β 基本型 答案
讨论结构和功能 蛋白质胶原的多样性，强调其在结缔组织细胞外基质 (ECM) 中的作用。  讨论内容包括胶原蛋白的生物合成、其独特的三重螺旋结构，以及翻译后修饰如何影响其机械性能。  请举例说明一种与胶原蛋白有关的疾病，这种疾病具有很强的力学特征。	<p>胶原蛋白是一种蛋白质家族，在以下方面发挥着重要作用。它是结缔组织的细胞外基质 (ECM)。它是人体中含量最高的蛋白质，为各种组织提供结构支撑和机械强度。胶原蛋白由三条多肽链组成，排列成独特的三螺旋结构。这种结构由氢键稳定，为胶原蛋白提供了特殊的机械性能。</p> <p>胶原蛋白的生物合成始于成纤维细胞中胶原蛋白基因的转录。产生的 mRNA 随后被翻译成原胶原蛋白，原胶原蛋白在内质网和高尔基体中进行翻译后修饰。原胶原蛋白会发生羟基化、糖基化和其他修饰，从而影响其机械性能。成熟的胶原蛋白分子随后由成纤维细胞分泌并组装成纤维，从而形成结缔组织的 ECM。</p> <p>胶原纤维是一种高度组织化的结构，可为组织提供机械强度和稳定性。它们由多个平行排列的胶原分子组成。胶原蛋白的三重螺旋结构使其能够承受较大的拉伸负荷，是承重组织的理想材料。</p> <p>成骨不全症 (OI) 是胶原蛋白相关疾病的一个例子，这种疾病也被称为脆骨病。成骨不全症是一种遗传性疾病，会影响胶原蛋白的合成和加工，导致骨骼脆弱。OI 患者经常会发生骨折、关节脱位和骨骼畸形。OI 患者体内胶原蛋白的机械性能发生改变，导致骨强度降低，更容易骨折。</p> <p>总之，胶原蛋白是一种高度多样化的功能性蛋白，在结缔组织的 ECM 中发挥着至关重要的作用。其独特的三重螺旋结构和翻译后修饰使其具有卓越的机械性能。与胶原蛋白相关的疾病（如 OI）凸显了胶原蛋白在维持组织完整性和机械强度方面的重要性。</p>	<p>胶原蛋白是一个结构和功能多样化的蛋白质家族。胶原蛋白在结缔组织的细胞外基质 (ECM) 中发挥着至关重要的作用。胶原蛋白是哺乳动物体内含量最丰富的蛋白质，约占人体蛋白质总含量的 25%。胶原蛋白由成纤维细胞、软骨细胞和成骨细胞等细胞合成，其生物合成涉及转录、翻译、翻译后修饰和分泌等复杂过程。</p> <p>胶原蛋白合成为一种叫做原胶原蛋白的前体蛋白，经过翻译后修饰形成成熟的胶原纤维。胶原蛋白分子由三条多肽链组成，每条多肽链长约 1000 个氨基酸，通过链间二硫键连接在一起。胶原蛋白的三螺旋结构非常独特，是由三条多肽链以左旋超螺旋排列而成。这种结构使胶原蛋白具有特有的机械特性，包括强度、刚度和弹性。</p> <p>胶原蛋白在结缔组织的 ECM 中发挥着关键作用，它提供结构支撑，促进细胞粘附和迁移，并调节细胞增殖、分化和凋亡等过程。胶原蛋白还参与伤口愈合、组织修复和重塑等各种生理过程。</p> <p>成骨不全症 (OI) 是胶原蛋白相关疾病的一个例子，这种疾病也被称为脆骨病。OI 是一种遗传性疾病，会影响胶原蛋白的合成，导致骨骼脆弱和畸形。引起 OI 的胶原基因突变可导致胶原纤维组装异常、胶原交联改变或胶原合成减少，从而导致骨强度降低和更易骨折。OI 患者体内胶原蛋白的机械特性发生改变，导致骨硬度降低和骨脆性增加。这凸显了胶原蛋白结构和功能在维持结缔组织机械完整性方面的重要性，以及基于胶原蛋白的疗法在治疗胶原蛋白相关疾病方面的潜在作用。</p>
选择最可能的答案 (A、B 或 C)：较高的湿度条件如何影响鸟类羽毛的硬度？  A) 湿度越高，硬度越低 B) 湿度越高，硬度越高 C) 湿度越高，硬度不变	<p>鸟类羽毛的层次结构包括<math>\alpha</math>-角蛋白纤维、倒钩和钩刺。这些结构的硬度受湿度、温度和相对湿度等因素的影响。</p> <p>较高的湿度条件会使羽毛的角蛋白纤维吸收水分，导致膨胀加剧，并可能降低硬度（软化）。因此，最可能的答案是 A。</p>	<ol style="list-style-type: none"> <li>1. 羽毛的分层结构使其具有独特的机械特性，例如硬度。羽毛的硬度由多个因素决定，包括羽毛钩刺内微纤维的排列和方向。</li> <li>2. 湿度可通过改变材料的含水量来影响其行为。高湿度会使材料吸收更多水分，从而导致其物理性质发生变化。</li> <li>3. 据报道，就鸟类羽毛而言，高湿度实际上会增加其硬度。这是因为空气中水分含量的增加会导致羽毛轻微膨胀，从而压缩内部结构并增加硬度。不过，这种影响只是暂时的，而且是可逆的，因为当湿度降低时，羽毛又会恢复到原来的硬度。</li> </ol> <p>因此，最可能的答案是选项 B：湿度越高，硬度越大。</p> <p>注：在依赖任何具体信息之前，仔细核对信息来源并确保其可靠始终是非常重要的。在这种情况下，我的答案是基于已发表的研究报告，这些报告调查了湿度对羽毛硬度的影响。</p>
这些概念中哪一个是壁虎脚的强大附着力的原因。  (A)，动物使用超级胶水。(B)，尺寸效应，纳米级纤维产生强力和耐缺陷的粘附力。(C)，壁虎的高效设计使其重量轻。(D)，壁虎具有从丝胶蛋白中提取的特殊用途蛋白质，可确保其粘附在许多表面上。	<p>(B) 是解释壁虎脚具有强大粘性的关键。壁虎的脚上覆盖着细小的毛发状结构，称为刚毛，刚毛又分裂成更小的纤维，称为 spatulae。这些结构使壁虎能够利用范德华力附着在表面上，而这种强大的附着力是由尺寸效应造成的，纳米级的纤维能够产生强大的、不受缺陷影响的附着力。</p>	<p>(D)，壁虎体内有从丝氨酸中提取的特殊用途蛋白质，这种蛋白质使壁虎的脚具有很强的粘附力，其关键概念是确保它能粘附在许多表面上。壁虎的脚上覆盖着数百万根被称为刚毛的细小绒毛，这些绒毛上包裹着一种被称为壁虎丝胶蛋白的蛋白质。这种蛋白质能在壁虎脚和接触表面之间产生范德华力，从而在不使用超级胶水或任何其他粘合剂的情况下实现强力粘合。壁虎的体型效应和高效设计可能有助于提高其整体粘附力，但正是丝胶蛋白的独特特性使壁虎具备了攀爬垂直表面和在天花板上倒立行走的非凡能力。</p>

图 5 显示了使用 [11] 中介绍的生物启发知识测试集进行的知识召回评估实验对比。我们可以看到，尽管 X-

## X-LoRA：低等级适配器专家混合物

LoRA 模型比 BioinspiredLLM 模型或 Llama-BioLLM 模型小得多（X-LoRA 的参数为 7B 而其他模型的参数为 13B），但它的性能却最好。图中还包括与最近发表的其他 LLM 的比较，特别是 Orca-13B 和 Llama-13b-chat。图中还包括 X-LoRA 与基础模型 *Zephyr-7B-β* 模型的比较，结果表明 X-LoRA 模型的性能显著提高。这些比较表明，即使使用较小的基础模型，X-LoRA 也能实现很强的性能。



**图 5：**知识召回评估实验结果。(a), 使用 [11] 中介绍的生物启发知识召回考试的结果。尽管 X-LoRA 模型比 BioinspiredLLM 模型小得多（7B 个参数对 13B 个参数），但 X-LoRA 的性能最好。图中还包括 X-LoRA 与基础模型 Zephyr-7B- $\beta$  模型的比较。值得注意的是，尽管 X-LoRA 模型要小得多，尽管它具有许多不同的功能，但与所有其他模型相比，X-LoRA 性能更优。(其他模型的性能数据摘自文献[11]) (b), 文献[8]中机械/材料知识召回基准的结果。(c) 表 2.1 中提出的问题的结果。

图 6 显示了图像合成中的一个应用，我们使用 X-LoRA 生成一个提示，供其他模型使用，例如这里的 DALL-E 3。

创建一个用于生成图像的提示，以准确描述分层复合材料的微观结构。

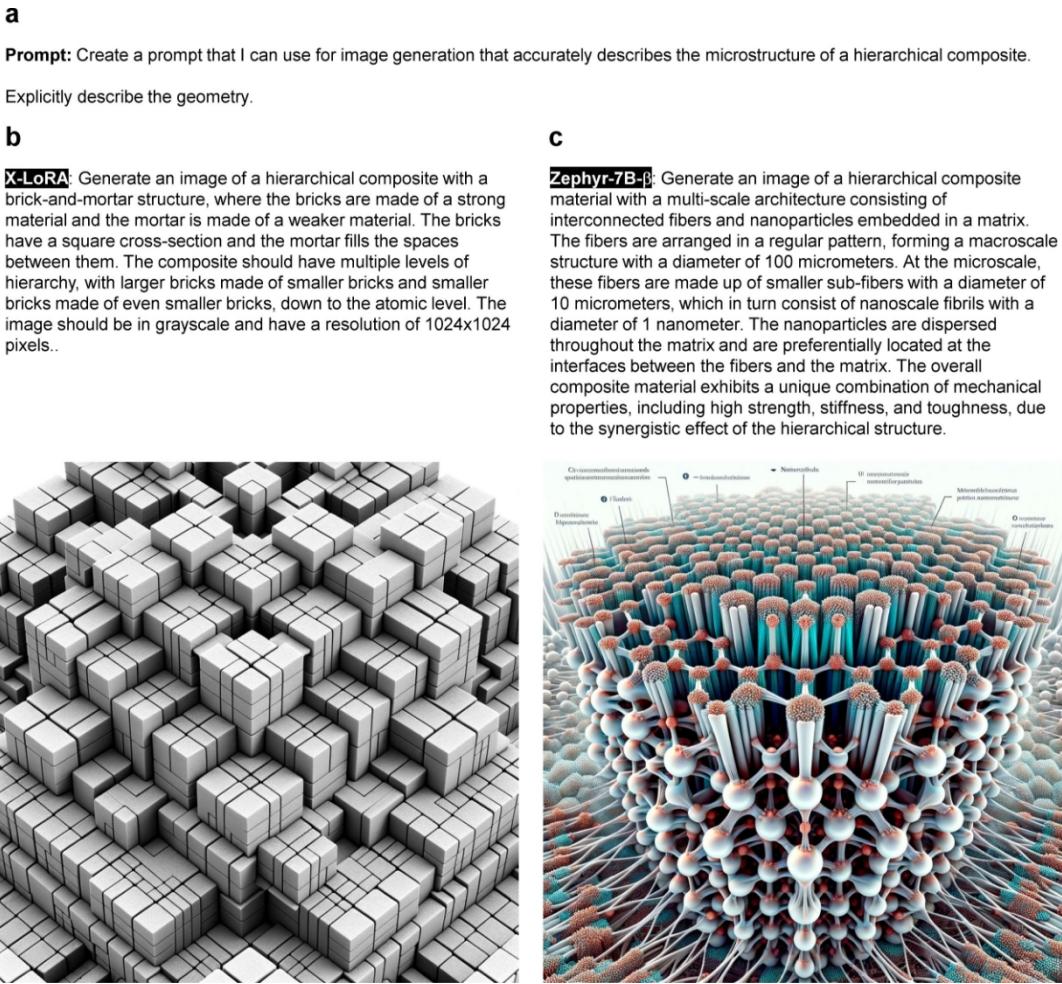
明确描述几何图形。

对比结果（图 6(b)-(c)）清楚地表明，DALL-E 3 的提示更加简洁，因此生成的图像也更加精确[27]。值得注意的是，Zephyr-7B- $\beta$  生成的提示并没有准确反映所提供的指令，而是引入了各种其他设计成分，如纤维和纳米颗粒。而在最初的提示中并没有提到这些内容，因为最初的提示特别关注分层复合材料的微观结构。X-LoRA 提示更简短，重点更突出，因此结果更好。

## 2.2 蛋白质分析的生成解决方案

## X-LoRA：低等级适配器专家混合物

接下来，我们将讨论具体的蛋白质生成和分析任务。这些能力在基于蛋白质力学适配器的 X-LoRA 模型中显现出来，并通过与模型中的生物学知识、生物启发材料和推理以及逻辑推理相结合而得到增强。本节将专门讨论对蛋白质任务本身的分析，评估这种能力的定量表现。



**图 6：**图像合成应用在这里，我们要求模型生成图像生成提示

(a)图中显示了 X-LoRA (b) 和 Zephyr-7B- $\beta$  (c) 的结果。对比结果清楚地表明，X-LoRA 的提示更为简洁，因此 DALL-E 3 所生成的图像也更为精确。

图 7 显示了基于蛋白质任务从序列预测力变形行为的结果。图 15 显示了根据序列预测解折力和解折能的整体性能，并将地面实况与预测结果进行了比较（我们发现  $R_2 = 0.85$ ）。总体而言，该模型的性能非常出色，正如下一节将探讨的那样，它在设计和周期一致性评估的反向任务中也非常有效[28]。然而，与早期使用高度专业化的 GPT 模型解决此类任务的工作不同，我们的 X-LoRA 模型将这些特殊用途能力与通过各种微调适应性赋予的其他深度专业化能力结合在一起。

### 2.3 蛋白质设计与分析

现在，我们将测试案例扩展到涉及多个能力领域的综合性应用。具体来说，本节报告的实验展示了 X-LoRA 模型如何利用一系列独特的技能，以及代理建模如何带来更具挑战性和更详尽的响应。我们以蛋白质设计任务开始讨论，要求模型设计一种新型蛋白质，以满足一定的目标力-变形响应。这些力-形变响应已在早期

## X-LoRA：低等级适配器专家混合物

的工作中进行过研究[29]，反映了蛋白质两端受拉时力与形变的测量结果（反映了经典的蛋白质展开实验[30]）。

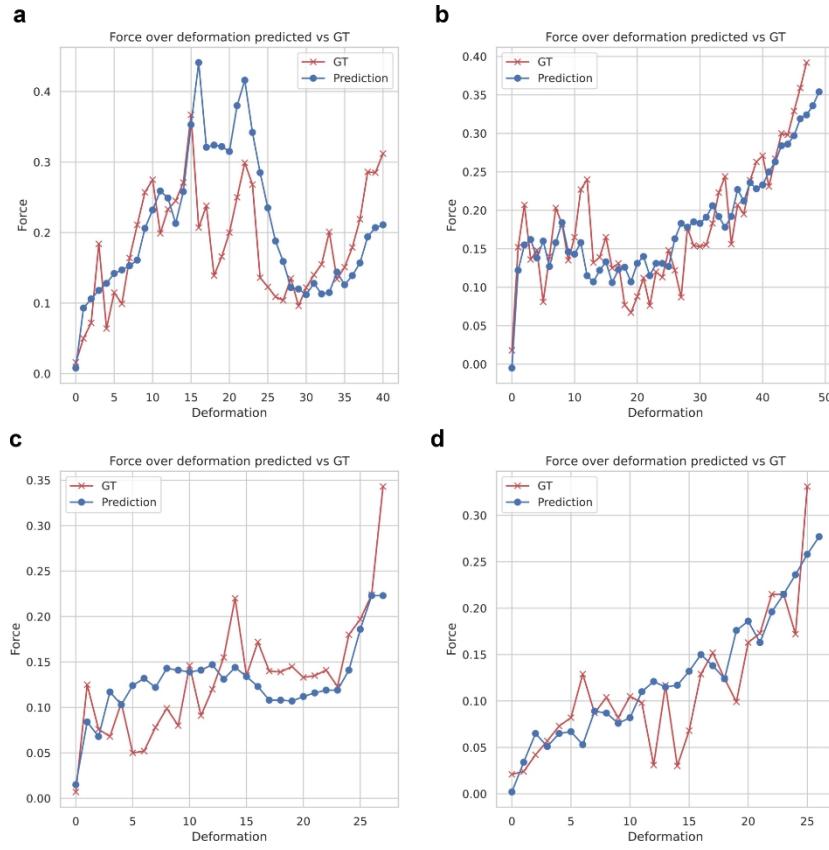


图 7：使用 X-LoRA 中的蛋白质任务预测氨基酸序列的受力变形行为（如图 a 至 d 所示）。可以看出，该模型具有出色的前向能力，可以很好地预测非线性力学行为。

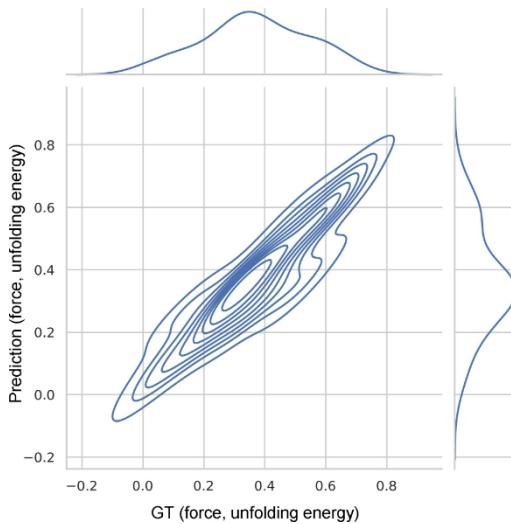
图 9 展示了我们如何使用生成蛋白质任务。我们设计一种具有所需受力变形行为的蛋白质（原始分子动力学（MD）模拟中使用的蛋白质牵拉训练数据和边界条件，见[29]），然后测试预测序列，检查其是否符合所需目标特性（图 9a）。这种周期一致性分析对于检验逆向能力是否与正向预测相匹配非常重要。图 9b 显示了预测序列 MSKSNTYRMLVLLEDDTKINKEDEKFLKGKPGKMHE- FVDELILPFNVDELDNTWFDKFDAEICIPYVEILKESGMK 的结果。此外，图 9c 显示了该蛋白质的折叠模型（使用 AlphaFold 2 [31]）。为评估所设计蛋白质的关系，图 9d 通过基本局部比对搜索工具（BLAST）树[32]检验了所设计蛋白质与其他已知序列的关系。有趣的是，所设计的蛋白质与早期工作中探索过但尚未确定或解决结构问题的假定蛋白质有某些关系。

通过代理建模，并将多个 X-LoRA 模型作为多代理模型中的元素，我们利用 X-LoRA 模型进一步分析了预测序列。这就利用了我们模型的增强功能，不仅能理解蛋白质力学，还能结合其他领域的知识加以利用。在这里，我们将上述预测结果作为输入，要求对序列的变体进行更深入的分析并提出设计建议。我们从这项任务开始：

我已经确定了这一氨基酸序列：msksntyrmvlleddtkinkedekflkgkpgkmhefvdel ilpfnvdelelntwfdkfdaeicipyveilksgmk.

如何使用？

提供合理的步骤，具体讨论问题的工程设计。



**图 8：**X-LoRA 从序列中预测展开力和展开能的总体性能，比较地面实况与预测结果 ( $R_2 = 0.85$ )。此处显示的数据汇总了测试集的展开力和展开能。

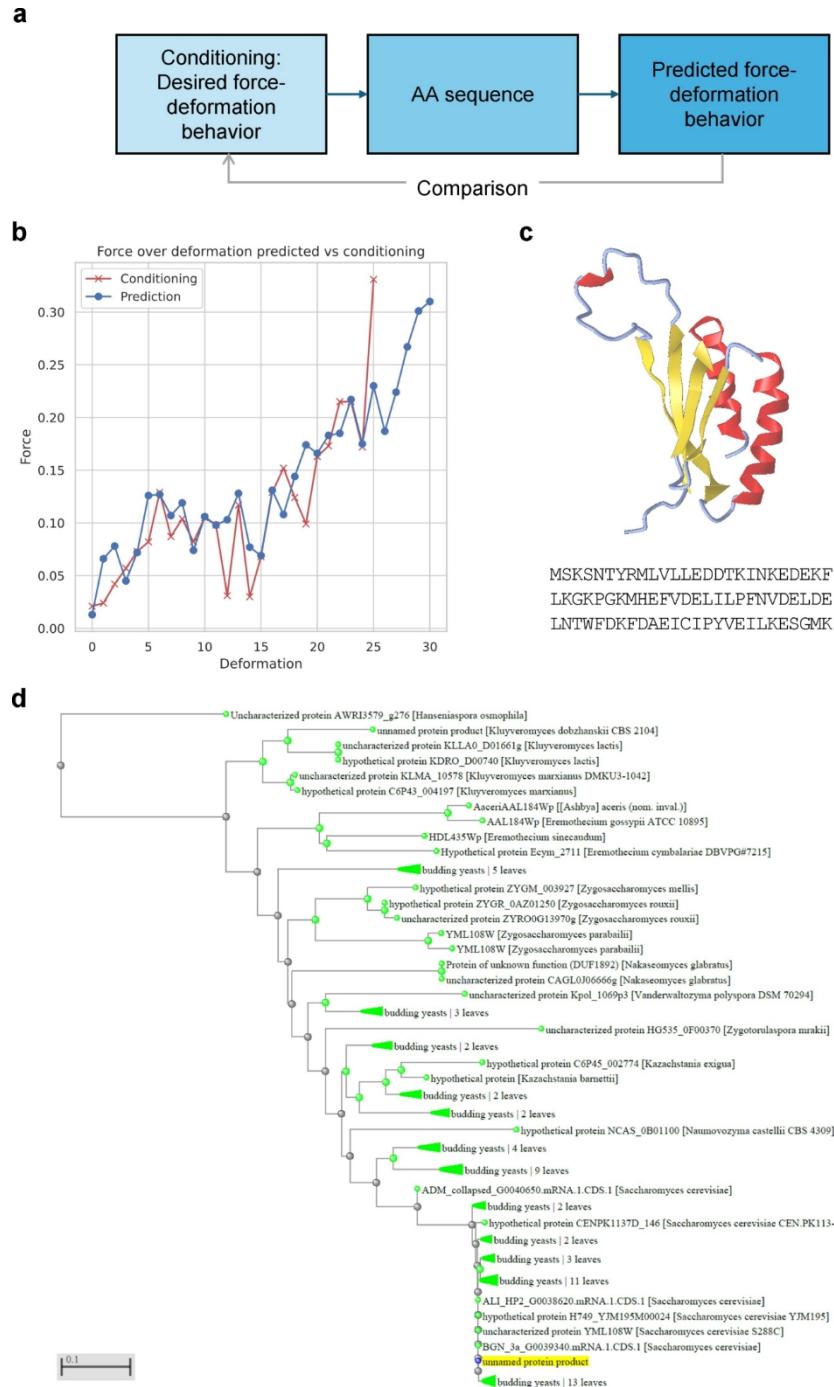
我们使用多重代理互动来开发解决方案，特别是使用逆向策略（详见材料与方法），推动模型探索所讨论概念的更深层次和更复杂的方面。根据 X-LoRA 模型的识别，对话摘要如下（完整对话见补充材料，SI Conv 示例 1）：

对话讨论了特定氨基酸序列在蛋白质工程中的潜在应用。该序列包含疏水和亲水残基的混合体，表明它可以形成稳定的结构并与其他分子相互作用。要利用该序列设计蛋白质，可以使用计算方法预测二级和三级结构，但局限性包括蛋白质折叠的复杂性和计算成本。改进方法可以通过开发更高效的算法和纳入新的实验数据，计算模型还将不断改进。计算模型的验证对于确保准确性和实际应用的实用性至关重要。交叉验证、正则化、早期停止、集合方法和迁移学习等技术可用于防止过拟合和欠拟合。对模型的解释和说明也有助于为实际应用提供信息。

我们还要求该模型编制一份要点清单，结果如下

- 氨基酸序列 MSKSNTYRMLVLLEDDTKINKEDEKFLKGKPGKMHEFVDELI LPFNVDDELNTWFDKFDAEICIPYVEILKESGMK 由于混合了疏水和亲水残基，在蛋白质工程中具有潜在的应用价值。计算方法可用于预测蛋白质的二级和三级结构，但其局限性包括蛋白质折叠的复杂性和计算成本。
- 计算模型的验证对于确保准确性和实际应用的实用性至关重要。交叉验证、正则化、早期停止、集合方法和迁移学习等技术可用于防止过拟合和欠拟合。
- 对模型的解释和说明也有助于为实际应用提供信息。

最后，关键的启示是



**图 9：**利用生成任务，我们设计了一种具有所需受力变形行为的蛋白质，然后对预测序列进行了测试，检验其是否符合所需目标特性（a 面板）。面板 b 显示了预测序列 MSKSNTYRMLVLLEDDTKINKEDEKF LKGKPGKMHEFVDELILPFNVDELNTWFDKF- DAEICIPYVEILKESGMK 的结果，比较了调节（目标响应）与模型预测的实际响应。请注意，我们还确认，使用计算力 [...] 直接计算峰值力的结果为 0.305，与力拳理指令的预测结果十分吻合。面板 c 显示了蛋白质的折叠模型（使用 AlphaFold 2 获得），面板 d 通过 BLAST 树[32]检验了设计的蛋白质与其他已知序列的关系。

最重要的一点是，氨基酸序列

MSKSNTYRMLVLLEDDTKINKEDEKFLKGKPGKMHEFVDELILPFNVDELDNTWFDKFDAEICIPYVEILKESGMK

由于它混合了疏水和亲水残基，因此在蛋白质工程中具有潜在的应用价值。要设计具有这种序列的蛋白质，可以使用计算方法来预测二级和三级结构，但其局限性包括蛋白质折叠的复杂性和计算成本。

验证计算模型对于确保准确性和实用性至关重要

在实际应用中，可以使用交叉验证、正则化、早期停止、集合方法和迁移学习等技术来防止过拟合和欠拟合。对模型的解释和说明也有助于为实际应用提供信息。

我们还可以更深入地探索具体的序列工程步骤。为此，我们要求模型提出更具体的设计建议，以提高蛋白质的稳定性：

如何改变以下蛋白质序列使其更加稳定？

msksntyrmvlvleddtkinkedekf1kgkpgkmhefvdelilpfnvdeleldntwfdkfdaeicipyveilksgmk?

互动摘要如下（完整对话见补充材料，SI Conv 示例 2）：

对话讨论了如何通过引入疏水氨基酸和二硫键来提高蛋白质序列的稳定性。生物学家认为，疏水氨基酸往往能提高蛋白质的稳定性，而二硫键也能提高稳定性。工程师要求提供疏水氨基酸的具体例子，以及应该加入多少个二硫键。生物学家提供了利用这些修饰成功稳定蛋白质的例子，包括胰岛素、牛胰蛋白酶抑制剂和绿色荧光蛋白。对话中还谈到了引入疏水氨基酸或二硫键的潜在限制或缺点，如蛋白质功能的变化、复杂性的增加以及聚集的可能性。这位生物学家建议，可以通过仔细选择氨基酸、计算分析和实验验证来解决这些局限性。

要点清单：

- 为了使蛋白质序列更加稳定，可以引入疏水氨基酸和二硫键。

疏水氨基酸往往能提高蛋白质的稳定性，而二硫键也能提高稳定性。

利用这些修饰成功稳定蛋白质的例子包括胰岛素、牛胰蛋白酶抑制剂和绿色荧光蛋白。

- 引入疏水氨基酸或二硫键的潜在限制或缺点包括蛋白质功能的改变、复杂性的增加以及聚集的可能性。
- 这些限制可以通过精心选择氨基酸、计算分析和实验验证来解决。

最后，关键的启示是

对话中最重要的一点是，要使蛋白质序列

MSKSNTYRMLVLLEDDTKINKEDEKFLKGKPGKMHEFVDELILPFNVDELDNTWFDKFDAEICIPYVEILKESGMK

可以引入更稳定的疏水氨基酸和二硫键。这直接回答了最初的问题。

从蛋白质工程的角度来看，这些建议是有意义的，因为它们确实能带来更稳定的设计。我们提出了具体的例子和策略，研究人员可以进一步探讨。

在另一个例子中，我们考虑这三个序列。我们首先要求模型计算每种序列的强度，然后让它对结果进行推理：找出最强的候选序列，并解释为什么这种蛋白质可能是最强的。此外，我们还要求模型解释在实验室中制造它的步骤。这三个序列（通过手动操作测试集中的一个序列创建）分别是

- LNTWFDKFDAEIPEYVEILKSGMK
- aitwfdkfdaeicipyvealkiaamk
- Aaaaaaaakfdaaeiciiaaaaaaaakiaamk

本示例测试的是该模型是否能够结合不同的功能，并随着对话的进展动态地实现这一点。也就是说，在蛋白质力学、理性分析、生成新假设等不同能力之间动态切换。

图 10 总结了对话过程，并描述了对话过程中缩放比例的变化情况（分析显示了深层适配器缩放比例以及对缩放比例的汇总分析，以衡量各种专家的有效使用情况）。结果清楚地显示了蛋白质任务最初的高度相关性是如何逐渐转向生物启发适配器和生物学适配器的更突出表现的。在对话的最后，生物适配器是最突出的适配器。

为了分析预测结果以评估响应质量，图 11 展示了本任务中考虑的三种蛋白质（使用 Alpha Fold 2 [31]，通过 ColabFold [33] 折叠的蛋白质）。我们发现，中间最稳定的结构具有最有组织的二级组织，这与图 10 中 X-LoRA 模型预测的产生最高展开力的观点一致。其他两个结构的组织性较差；第一个结构基本上是螺旋形的，有一些转折，最后一个结构部分是非结构化的，几何形状上有一个扭结。这些特征解释了为什么观察到的展开力与模型预测的最低值相同。

通过分析折叠结构的视觉表现，我们证实中间最稳定的结构具有最有组织的二级组织，这与模型预测的它能产生最高的展开力这一概念是一致的。我们发现，其他两个模型的组织程度较低。第一个模型在很大程度上呈螺旋状，但有一些转折；最后一个模型部分结构不完整，几何形状上有一个扭结。我们认为，这些特征很可能是模型预测的最低解折力的原因，这与 X-LoRA 模型的预测一致。

在第二个例子中，我们将重点放在蛋白质的能量特性上，考虑蛋白质从初始状态到完全展开几何形状的端对端拉伸过程中的展开能[29]。我们再次考虑三个序列（前两个序列与上述序列完全相同，但第三个序列由一个较长的集合体组成，代表位于起始/末端的第二个序列和位于中心的第一个序列的混合体）：

- LNTWFDKFDAEIPEYVEILKSGMK
- aitwfdkfdaeicipyvealkiaamk
- aitwfdkfdaeicipyvealkiaamklntwfdkfdaeicipyveilksgmkaitwfdkfdaeicipyvealkiaamk

我们首先要求模型计算每种蛋白质的展开能，然后对结果进行推理：根据氨基酸序列，我们要求解释为什么这种蛋白质可能是最稳定的。最后，我们询问最稳定的蛋白质可能具有的功能。

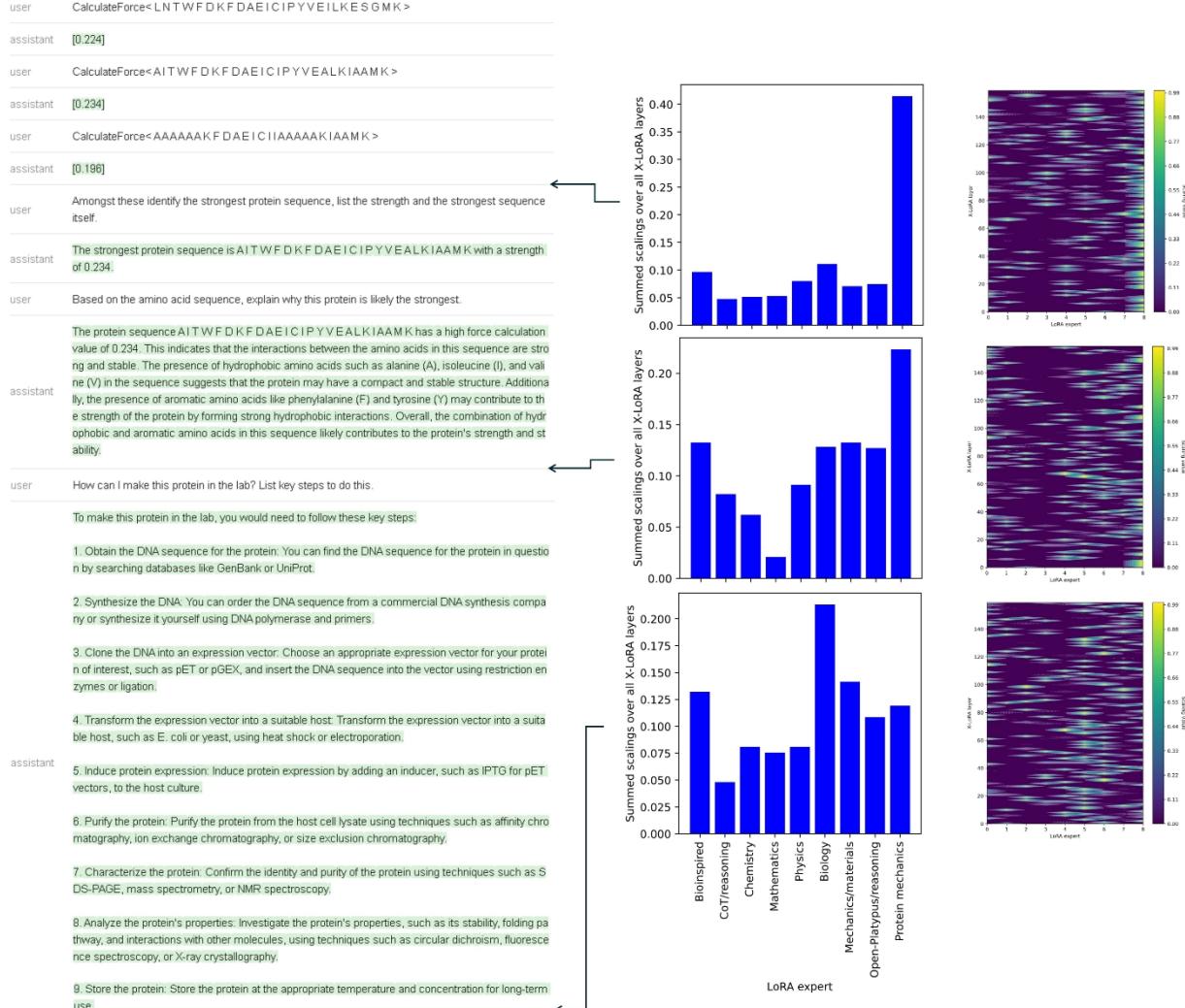
图 12 显示了一个对话记录，该对话调用了不同专家的组合，并整合了这些能力的知识，在此应用于三种蛋白质的展开能。在对话的最后，CoT/推理适配器是最突出的一个，因为 X-LoRA 模型响应查询，对结果进行推理，并预测可能的蛋白质结构特征以及蛋白质的潜在功能。具体来说，关于蛋白质稳定性起源的机理问题包括非极性氨基酸之间疏水相互作用以及极性氨基酸之间氢键作用的出现。该模型预测，有几个疏水氨基酸（I、F、W、Y）和氢键氨基酸（D、E、K）有助于提高蛋白质的稳定性。该模型还预测，序列中半胱氨酸（C）残基的存在将导致二硫键的形成，从而进一步稳定蛋白质结构。通过图 13 所示的结构分析可以验证，关于结构特征的关键预测被正确识别，从而解释了高稳定性的原因，这一点非常重要。图中的分析证实了所有这些特征的存在。另外两个蛋白质的螺旋结构域更短，稳定性也不尽相同，但它们的解折能

却低得多。

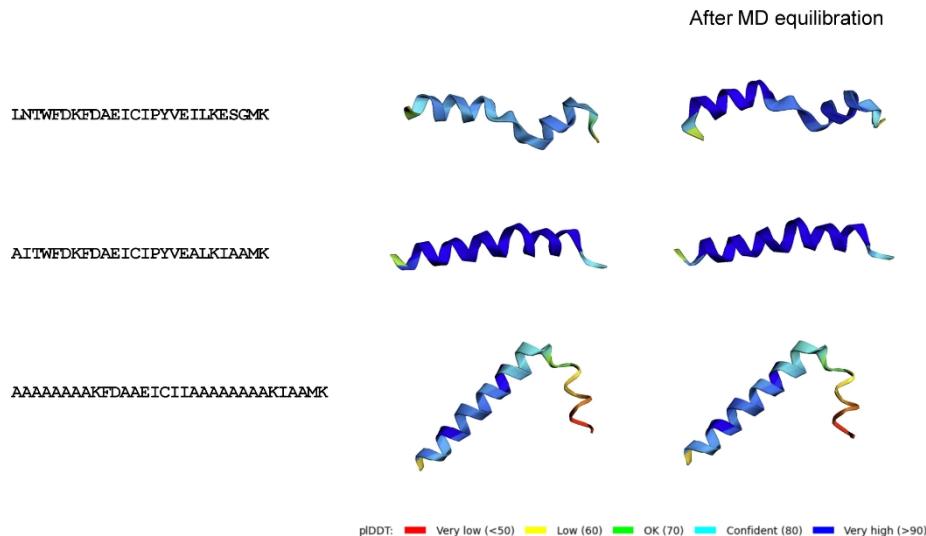
#### 2.4 通过对抗代理建模连接不同的学术学科和知识，生成本体论知识图谱

由于我们的 X-LoRA 模型拥有跨领域的深厚知识，因此我们可以利用它来探究不同想法、知识库和专业领域之间的联系。在一个实验中，我们使用该模型提出了两个查询，每个查询的内容都是

## X-LoRA：低等级适配器专家混合物



**图 10：**调用不同专家组并整合这些能力知识的复杂对话。右侧面板显示了对话过程中标度的变化情况。结果清楚地显示了蛋白质力学任务最初的高度相关性，然后逐渐转向生物启发适配器和生物学适配器的更突出表现。在对话结束时，生物学适配器的表现最为突出，这反映了模型在响应蛋白质的生物学制造步骤时，对生物学关键要素的关注。



**图 11：**图 10 中折叠蛋白质的直观图，显示预测结果（左列）和 MD 平衡结果（右列）。中间最稳定的结构（AITWFDKFDAAEICIPYVEALKIAAMK）具有最有组织的二级组织，这与它产生最大解折力的结果一致。其他两个组织较少；第一个主要是螺旋状，有一些转折，最后一个部分无组织，几何形状上有一个扭结。这些特征很可能是模型预测的最低解折力的原因。

类似，但使用的重点不同。然后，我们生成三元组，生成本体论知识图谱，将答案提炼为更有条理的输出 [34, 35]。由此产生的图谱可以综合理解所产生的见解，并以可解释和机制化的方式直观显示概念之间的联系。

我们使用两个对抗代理来回答这个问题。代理 1 被指定为 "批判物理学家"，代理 2 被指定为 "哲学家"。代理 1 提出最初的问题，并负责根据代理 2 的每一个回答进行探究和提出更多问题。随着对话的展开，对话深入主题，揭示出各种不同的层面。提出的问题有意涵盖多个学科知识领域，以考察模型整合思想和概念的能力。

问题是

将古典音乐作品抽象地表示为由相互作用的音符、和弦、旋律、节奏等组成的图形。

如果我们对这一抽象的音乐概念施加压力，直至失败，会发生什么呢？

提供逻辑步骤和数学推理，并具体讨论问题的物理原理。

对话全文见补充材料（见 SI Conv 示例 3）。为简明扼要起见，我们在此仅报告摘要和主要启示，其中包含了基本特征：

对话讨论了对古典音乐作品的抽象表现形式施加压力的问题，将其视为一个具有相互作用的音符、和弦、旋律和节奏的物理系统。压力可以用应力来量化，当应力达到临界值时，就确定了失效临界点。为了检验数学模型的预测结果，可以使用抽象音乐作品的物理模型，并将实验观察结果与预测结果进行比较。所提数学模型的局限性包括对相互作用的简化表示、静态表示、均匀分布以及应力和压力之间的线性关系。为了提高模型的准确性和有效性，需要建立更详细、更逼真的模型，并考

## X-LoRA：低等级适配器专家混合物

虑动态效应、非均匀分布和非线性因素。