

评估维度

1、正确性

评估答案在事实层面的准确性，包括是否基于可靠的事实、书籍、研究成果，所提及的物理、化学概念、参数和数值是否符合领域内的公认标准。若涉及计算，计算的精确性，以及单位的使用是否得当。

评分标准：

10 分：答案完全准确，无任何错误。所有事实、数据和计算均正确，单位使用恰当且精确。

4-9 分：答案基本准确，存在少量事实性错误（不超过 2 处），根据错误数量和严重程度扣分。

1-3 分：答案有一定量事实性错误（超过 2 处，但低于 4 处），根据错误数量和严重程度扣分。

0 分：答案有较多事实性错误（超过 4 处）。

2、逻辑性

评估答案在逻辑层面的合理性，包括整体结构的严谨性、内容之间的逻辑衔接以及是否存在自相矛盾或前后不一致的情况。回答的每一部分是否紧密围绕主题展开。若有推测性信息，表述是否严谨合理。

评分标准：

10 分：逻辑严谨，内容衔接自然，无任何矛盾或不一致，整体结构紧密。

4-9 分：逻辑基本合理，仅存在少量矛盾或不一致（不超过 2 处），但不影响整体理解。

1-3 分：逻辑存在明显问题，存在较多矛盾或不一致（超过 2 处，但低于 4 处），影响整体连贯性。

0 分：逻辑极其不合理，内容之间缺乏衔接，存在重大矛盾（超过 4 处）。

1.3 清晰度

评估答案语言描述是否清晰，是否存在冗余，回答是否针对性地回应问题，是否表达清晰、简洁、易懂，使用户能够轻松理解所提供的信息，是否存在冗余描述，若回答中有冗余描述，需扣分。

评分标准：

10 分：表达极为清晰、简洁，无冗余。

4-9 分：表达较为清晰，存在较少（不超过论述篇幅的 20%）冗余。

1-3 分：表达模糊，或存在较多冗余（超过论述篇幅的 20%，但低于 50%）。

0 分：表达极不清晰，语言混乱，或存在大量冗余（超过论述的篇幅的 50%）。

1.4 完备性【总分：10 分】

评估答案是否全面完整地回答了问题，是否完全回应了用户的核心问题，是否提供了足够细节和相关知识，是否给出必要的相关理论、研究进展、实验数据或技术应用等内容。若回答缺少遗漏了重要的相关信息，导致对问题的回应不够充分，将扣分。

评分标准：

10 分：答案全面完整，包含所有要点。

4-9 分：答案较完整，相较参考答案，存在较少遗漏（要点超过参考答案要点的 50%）。

1-3 分：答案不完整，相较参考答案，存在较多遗漏（要点超过参考答案要点的 30%，但低于 50%）。

0 分：答案极不完整（要点低于参考答案要点的 30%）

1.5 理论深度【总分：10 分】

评估维度描述：评估答案是否能够运用普适性的化学基本原理（如热力学、动力学、量子化学、非共价相互作用理论等），对具体问题进行深入、具体、特异化的分析和阐释。该维度不关注知识点的罗列广度（例如，在回答奎宁时是否提及其历史、药理等），而关注对核心化学问题进行纵向挖掘的深度。一个有深度的回答，应能清晰地展示出从普遍原理到具体现象的逻辑推演路径以及微观结构到宏观现象的机制链条。

例如，在解释一个主客体结合现象时，一个有深度的回答不应止步于描述“发生了氢键和疏水作用”，而应进一步：

1. 分析相互作用的来源：具体是哪个官能团的哪个原子参与了氢键？分子的哪部分构象使其能有效进入疏水空腔？
2. 评估相互作用的相对贡献：在该体系中，是焓驱动（强的氢键）还是熵驱动（疏水效应释放水分子）更多？为什么？
3. 关联到宏观结果：这种微观相互作用如何最终决定了宏观可测量的热力学结果（例如，解释为何焓变有利而熵变不利，但最终 ΔG 为负，导致高结合常数 K_a 。这个结合常数还可能继续影响光谱、电流等可测量性质）。

评分标准:

10 分: 能够系统性地运用一或多个基本化学原理, 对具体问题进行深入、多层次的特异性分析。清晰地揭示了现象背后的内在机制, 成功地将回答从“是什么”的层面深入到了“为什么”和“怎么样”的层面。

4-9 分: 能够提及相关的基本原理来解释现象, 但分析不够深入或具体, 在原理和实例之间的逻辑连接不够紧密或存在跳跃。展示了一定的理论思考, 但未能完全特异化。

1-3 分: 基本停留在对事实或现象的描述上, 很少或未能运用基本原理解释其内在机制。提及了理论词汇, 但没有进行有效分析。

0 分: 完全没有理论分析, 纯粹的事实罗列。

1.6 论述严谨性与信息密度 【总分：10 分】

评估维度描述: 评估答案在科学论述上的严谨程度和信息传递的效率。考察内容包括: **1) 术语精确性:** 是否始终使用领域内最精确、公认的专业术语, 术语是否精准的表达概念意图; **2) 信息密度:** 回答是否凝练, 有效信息占比是否高, 是否存在为通俗易懂而牺牲信息量的冗余解释; **3) 论证合理性:** 是否存在不严谨的概括或可能引起误解的简化。

评分标准:

10 分: 术语使用极其精准; 信息密度极高, 几乎无任何冗余; 论证过程严密, 无任何不严谨的推论或简化。

4-9 分: 术语使用较为准确, 但偶有不够精确之处; 信息密度较高, 但存在少量 (低于篇幅 **15%**) 的铺垫或简化性描述; 论证基本严谨。

1-3 分: 使用了较多非专业或过于宽泛的术语; 信息密度低, 存在较多 (超过篇幅 **15%**) 冗余或过度简化的内容; 论证不够严谨。

0 分: 术语使用混乱; 信息冗余严重; 论证缺乏严谨性。