Mn(II) 溴化物发光材料：数据需求与研究计划 v2

致：化学合成团队、实验表征团队

来自：数据科学团队

日期：2026-10-13

版本：v2（新增①全库数据面覆盖审计 ②各项预估工时与排期）

---

0. 一句话总结

请优先做两件**最便宜、最缺**的事：① **每个化合物的吸收光谱（→ 10Dq 与 Racah B，我们完全缺失的核心物理变量）；② 「寿命 + PLQY」同批配对（→ k_r / k_nr，现在寿命数据为 0）**。

并请**优先扩充低配位（Mn 配位数 ≤3）与「纯溴配位」两个我们已证明可预测的窗口**。

时间上：P0 两项约需 2–3 周即可交付第一批数据，是整个项目的最短见效路径。

---

1. 我们的现状（诚实数字，不是最好的一次）

按化学式分组、10 个折分种子重复的交叉验证：

目标	诚实表现	判定
精确发射波长	R^2 = **+0.120 ± 0.197**	基本不可预测
精确 PLQY 数值	R^2 = **+0.028 ± 0.293**	基本不可预测
发射颜色（蓝/绿/红）	平衡准确率 **0.48**	等于随机
PLQY「亮/暗」（≥50%）分类	平衡准确率 **0.754 ± 0.084**	✅ 可靠可用

根因不是算法，是数据面有巨大空洞（见第 2 节）。

---

2. 数据面覆盖审计：我们只看到了整张地图的一小块

我们对全库 **796 个结构**做了逐维度覆盖统计，发现**建模所需的很多维度几乎是空的**：

2.1 卤素家族 —— 我们只做了 Br

卤素	结构数	占比
Br	645	82.6%
mixed（混合卤）	134	17.2%
Cl	**1**	0.1%
I	**1**	0.1%
F	**0**	0%

> **Cl / I / F 几乎完全空白。** 我们此前所有结论（cn≤3 可预测、cn≥4 含 N/O 是黑洞）**仅在 Br 上验证过**，是否可推广到 I（预期红移）/Cl（预期蓝移）完全未知——这是最便宜就能大幅扩展覆盖面的一步。

2.2 配位数 —— 主体区域（6 配位）几乎是空的

配位数	全库结构数	已标记	标记率
cn2	41	4	9.8%
cn3	65	10	15.4%
cn4	121	34	**28.1%（最厚）**
cn5	86	**1**	**1.2%** ⚠️
cn6	**235**	**3**	**1.3%** ⚠️ **最大空洞**
cn7–9	76	4	5.3%

> **6 配位（最经典、最商用的八面体 Mn(II) 磷光体区域，如 $CsMnBr_3$ 类）有 235 个结构，却只有 3 个标记。** 这是全库最大的一块空白。

2.3 配体家族

配体家族	全库	已标记	标记率
all-Br（纯卤素）	101	36	**35.6%（最好）**
Br+N/O	271	19	7.0% ⚠️
noBr（Br 作抗衡离子）	219	2	**0.9%** ⚠️
Br+C（有机）	46	**0**	**0%** ⚠️

2.4 性质字段覆盖率（764 条主记录）

字段	覆盖	严重程度
发射波长	57 / 645 Br	—
PLQY	54 / 645 Br	—
寿命 lifetime	**0 / 764**	🔴 **完全为零**
带隙 band gap	**0 / 764**	🔴 **完全为零**
吸收峰	80 / 764	🟡 仅 10%
CIE 色坐标	95 / 764	🟡 仅 12%
LED 波长	75 / 764	🟡 仅 10%
激发光谱	极少	🟡
温度依赖（τ(T)/PLQY(T)）	**0**	🔴 完全为零

2.5 标签在"有价值的区域"分布极不均

发射颜色	数量		PLQY 段	数量
蓝光 <500nm	10		暗 0–20%	12
绿光 500–570nm	**41（72%）**		中 20–50%	10
橙/红 ≥570nm	**6** ⚠️		亮 50–80%	19
			很亮 >80%	13

> **红光（≥570nm，商业上最想要的长波长）只有 6 个标记。** 我们的模型天然偏向绿光，对红光几乎无训练数据。

---

3. 数据需求（含预估工时）

🔴 P0-A：吸收光谱 → 10Dq 与 Racah B 【最短见效路径】

项	内容
测量什么	**UV-Vis 吸收/漫反射光谱，250–800 nm**
为什么	Mn(II) 第一吸收带 ≈ **10Dq**；多带拟合 Tanabe–Sugano 图得 **Racah B/C**。这是我们模型完全缺失、且**扫描 100 篇文献 0 篇报告**的变量
样本量建议	**先做 50 个现有样品**（立刻见效），再随新化合物滚动补充
预估工时	仪器 2–3 个工作日（50 个样品，UV-Vis 很快）+ 拟合分析 1–2 天 → **约 1 周**
交付	原始光谱文件（保留，我们自行重拟合）+ 10Dq(cm^{-1})、B(cm^{-1})、各带峰位

> **对 6 配位尤其关键**：八面体 d^5 的 Tanabe–Sugano 分析方法最成熟，所以 10Dq 提取在 cn6 区域最可靠——正好补上最大的数据空洞。

🔴 P0-B：寿命 + PLQY 同批配对 → k_r / k_nr 【补上完全为 0 的字段】

项	内容
测量什么	**荧光寿命 τ（ns）+ PLQY，必须同一样品同一批次**
为什么	k_r = Φ/τ、**k_nr = (1−Φ)/τ**。k_nr 才是决定"亮不亮"的物理量，与配体刚性/共价性直接相关
样本量建议	与 P0-A 同一批 50 个
预估工时	若有专用 TCSPC/磷光寿命仪：约 **2–3 周**（含排队与拟合）；**若需外借机时，时间翻倍，请提前告知**
交付	τ 及拟合方式（多指数给主分量）、PLQY（含标准曲线与不确定度）

🟠 P1-A：扩充「cn≤3」窗口（我们已证明可预测）

目标窗口	当前标记	建议新增	预估工时
cn≤3（低配位，胺/N 给体）	15	**+90~130**	**4–6 周**（含合成+表征）
全 Br 任意配位（PLQY 最佳窗口）	32	**+50~80**	**4–6 周**（可与上项并行）

建议的实验设计（最重要的一条）：固定 Mn:Br:配体 比例、溶剂、温度、时间、反离子等全部条件，**只改变胺类配体 L**（同一系列 R–NH_2，R 逐步变化）。这样得到**干净的单变量系列**——我们分析发现当前最大问题是标签噪声，单变量

系列能真正建立"结构→性质"的因果关系。

高通量提示：若做 96 孔板组合库，一个批次可覆盖 96 个化合物，**约 1–2 周/批**（合成）+ 1 周（表征）。

> 📋 **我们已备好候选清单**：随本计划附 `SYNTHESIS_SHORTLIST.csv`，列出落在这两个窗口内、**库中已有结构但尚未完整表征**的 **106 个化合物**（窗口 A 低配位 79 个 / 窗口 B 全溴 27 个），含 COD 编号、分子式、Mn 配位数、Mn 核数，并标注了「多核 Mn」「纯无机（最快上手）」「cn≤1 描述符可能失真需核对」等提示。

> 已自动排除 122 个含异金属（Fe/Ni/Cr/Re/Ge 等）的杂金属配合物——它们不是简单的 Mn(II) 溴化物磷光体。

> 其中 `Br2Mn`、`Br3Mn·Rb`、`Br3CsMn` 等**纯无机体系合成最简单，可作为最快上手的起点**。

🟠 P1-B：补齐 6 配位（最大空洞）

目标	当前标记	建议新增	预估工时
cn6（八面体）	3	**+30~50**	**4–6 周**

> 建议与 P0-A 绑定：先对 30 个八面体化合物测吸收，验证 10Dq 是否真能解释其发射；若成立，这是打开商用磷光体市场最关键的一块。

🟡 P2-A：拓展到 Cl / I 卤素（目前几乎空白）

目标	现状	建议新增	预估工时
Mn(II) 氯化物 / 碘化物	1 / 1	**各 +15~20**	**6–8 周**（需换前驱体）

> 价值：验证我们的机制结论是否可推广；碘化物天然红移，是获取**红光**（当前仅 6 个标记）的捷径。

🟡 P2-B：温度依赖 + 晶体结构

项	内容	预估工时
温度依赖	τ(T)、PLQY(T)、发射(T) 77–300 K，5–6 点	**约 1 周**（样品就绪后）
晶体结构	新化合物尽量给 CIF	含在合成流程内
CIE / 带隙 / 激发光谱	随常规表征一并记录	几乎零额外成本

---

4. 请不要在这些地方投入（黑洞区）

❌ **Mn 配位数 ≥4 且含 N/O 给体**：该区域发射误差 **80.6 nm**，比不做任何限定**更差**，属物理+数据双重死局。**注意**：该区域恰有 271 个结构（占全库 35%），所以这对我们是**战略性放弃**而非小损失——但请先不要往这里投数据。

（例外：若 P0-A 的 10Dq/B 在此区域证明有效，可重启——这正是我们要先测吸收的原因。）

---

5. 测量规范（降低标签噪声——最容易被忽略、最关键）

我们在自己数据里发现过同系列化合物被错标（多个结构被标成同一 365 nm）、未区分多晶/溶剂化物的问题。请务必：

同一样品、同一批次同时测 PLQY 与寿命。

注明测量条件：固体（粉末/薄膜/单晶）或溶液；溶液注明溶剂；注明测试温度。

给出不确定度：PLQY 报 ±，寿命报 ±，光谱报重复次数或信噪比。

同步记录 CIE 色坐标、激发光谱、带隙（目前这三项覆盖都极低，几乎零成本）。

多晶/溶剂化物请标注，避免与主相混淆。

重复测量请给**重复值**而非单点。

合成条件元数据（前驱体比例、溶剂、温度、时间、反离子）请一并记录——我们要用它做"按合成路线限定"的预测。

---

6. 数据交付格式（可填写模板见附随的 `DATA_REQUEST_TEMPLATE.csv`）

每行一个化合物，至少包含：

```
sample_id, formula_with_ligand, cn_target, ligand_family, cn_measured,
emission_lambda_nm, em_fwhm_nm, cie_x, cie_y, excitation_lambda_nm,
plqy_pct, plqy_uncertainty, lifetime_ns, lifetime_fit_type, bandgap_eV,
abs_lambda_nm, Dq_cm-1, Racah_B_cm-1,
synthesis_route, precursor_ratio, solvent, temp_C, time_h, counterion,
measurement_state, temperature_dependence, cif_id, notes
```

若有**原始光谱文件**（吸收、发射、寿命衰减），请一并保留——10Dq/B 的数值依赖拟合方法，原始文件让我们能自行重拟合。

---

7. 预估排期（总览）

阶段	内容	预估时间	累计
阶段 1	P0-A 吸收光谱（50 个现有样品）	**1 周**	1 周
阶段 1	P0-B 寿命+PLQY（同批 50 个）	**2–3 周**	3 周
阶段 1	顺带补 CIE/带隙/激发（零成本）	并行	—
阶段 2	P1-A cn≤3 + 全Br 高通量合成与表征	**4–6 周**（可并行开工）	6–9 周
阶段 2	P1-B cn6 八面体系列（30–50 个）	**4–6 周**	6–9 周
阶段 3	P2-A Cl / I 拓展	**6–8 周**	12–17 周
阶段 3	P2-B 温度依赖	**1 周**	—

关键路径提示：**阶段 1（吸收+寿命，约 3 周）是最短见效路径**，建议无论其他如何都优先启动；阶段 2 的合成可与之**并行**，不必等阶段 1 完成。

若资源紧张，请只做阶段 1——它成本最低，且直接补上模型完全缺失的物理变量。

---

8. 我们承诺回馈

每收一批数据就回一次结论：哪些合成路线/配体值得继续、哪些是黑洞，避免你们在无效区域浪费试剂与机时。

3D 结构浏览器与筛选器（已有 viewer），帮助挑选下一批目标。

可复现的建模脚本与评估报告（含泄漏控制与多种子验证）。

「亮/暗」候选排序（当前平衡准确率约 75%，可作粗筛）。

联合申请/经费建议：若需要特定表征机时预算（尤其 TCSPC），我们可以提供数据支撑材料。

---

9. 预期效果与不确定性（诚实版）

不承诺数据一到就能把发射预测做到很高。保守预期：

补 **10Dq/B** 后，发射预测有望从"接近随机"提升到**中等水平**——最有把握的一项。

补 **k_nr** 后，PLQY 相关预测有望显著改善（目前 PLQY 数值预测几乎为零）。

cn≤3 / 全Br 窗口做深后，该窗口内可做到很准（发射 ±6~34 nm，PLQY ±17.7%），但**不外推到中高配位区**。

打开 cn6（最大空洞，235 结构） 依赖 10Dq/B 是否有效——这是本项目最大的机会点，也是不确定性最大的一点。

核心判断：当前是**方差受限**（52 个标记，折分随机性就能让 R^2 在 −0.32~+0.38 摆动）。**数据量从 52 增到 200+ 是唯一可靠的提升路径**，请优先支持。

10. 一页速查（给合成团队）

优先级	做什么	样本量	工时	为什么
🔴 1	**吸收光谱**	50 现有样品	**1 周**	最便宜，补全缺失的 10Dq/B
🔴 2	**寿命+PLQY 同批**	同批 50 个	**2–3 周**	补上完全为 0 的字段，拆出 k_nr
🟠 3	cn≤3 低配位高通量	+90~130	4–6 周	唯一能靠数据稳住预测的窗口
🟠 4	全 Br 窗口加深	+50~80	4–6 周	PLQY 最佳窗口（±17.7%）
🟠 5	cn6 八面体（最大空洞）	+30~50	4–6 周	235 结构仅 3 标记，商用磷光体
🟡 6	Cl / I 拓展	各 +15~20	6–8 周	目前几乎空白；碘化物=红光捷径
🟡 7	温度依赖 + CIE/带隙	顺带	~1 周	零成本补字段
❌	**cn≥4 含 N/O**	**别做**	—	黑洞（误差 80.6nm，比不限定更差）

顺带：CIE、带隙、激发光谱请顺手记录——目前覆盖仅 10–12%，几乎零成本。

有任何字段拿不到、测量有困难或机时排不开，**请先告诉我们**，我们愿意为可获得的测量重新排布优先级。