

CRO客户报告撰写指南

核心理念

面向CRO客户的报告 = 结果导向 + 应用价值 + 专业呈现

- **突出结果和用途**: 客户关心"是什么"和"怎么用", 而非"怎么算"
- **不透露核心算法**: 保护知识产权, 避免技术细节泄露
- **专业且可执行**: 报告应可直接用于客户会议和决策

一、报告结构 (推荐)

1. 执行摘要 (Executive Summary) - 1-2页

目的: 让客户在5分钟内了解核心发现和建议

包含内容:

- **项目目标** (一句话)
- **主要发现** (3-5个要点)
- **推荐方案** (Top 1-2)
- **关键风险提示**
- **算法名称、数据库名称、技术细节**

示例:

```
## 执行摘要
```

****项目目标****: 为抗EGFR VHH 7D12设计人源化方案, 降低免疫原性风险, 保持抗原结合能力。

****主要发现****:

1. 7D12的CDR构型为H2-10-1类别, 在人类种系中罕见, 建议采用表面改造(SR)策略
2. SR变体实现了低免疫原性 (预测MHC-II结合负荷=6, 处于临床SR分子的第29百分位)
3. 结构验证显示, 67%的SR突变位点表面暴露, 支持策略合理性
4. 可开发性评分78分, 处于可接受范围, 但疏水斑块风险略高 ($hp_{max9}=0.889$)

****推荐方案****:

- ****首选****: SR变体 (6个FR突变, CDR完全保留)
- ****可选优化****: IMGT101 V→S突变, 将疏水斑块风险降至临床中位数水平

****关键风险****:

- 基于计算预测, 需实验验证抗原结合亲和力
- 疏水斑块风险可通过制剂优化管理

2. 输入序列分析 (Input Sequence Analysis) - 1页

目的：展示对客户序列的理解

包含内容：

- 序列基本信息（长度、CDR序列、长度）
- 关键特征识别（VHH标志性位置、CDR构型类别）
- 质量检查结果（完整性、异常检测）
- ANARCI版本、IMGT映射细节、内部编号规则

示例：

```
## 输入序列分析

**序列信息**:
- 总长度: 127个氨基酸
- CDR1: GRTSRSYGMGW (12 aa)
- CDR2: SWRGDSTGYA (10 aa)
- CDR3: TLYEYDYWGQGT (13 aa)

**关键特征**:
- CDR构型类别: H2-10-1 ("稳定盆"构型, CDR2长度8)
- VHH标志性位置: 位置37为F (典型VHH特征)
- 结构来源: PDB 4KRL (实验解析结构)

**质量检查**:
-  序列完整性: 通过
-  CDR识别: 通过
-  异常检测: 无异常半胱氨酸、无终止密码子
```

3. 策略推荐与理由 (Strategy Recommendation) - 2-3页

目的：解释为什么推荐某个策略

包含内容：

- 策略选择依据 (CDR构型、临床先例、风险评估)
- 与临床分子的对比 (同类策略的成功案例)
- 风险-收益分析
- 决策树算法、熵计算、位置集定义细节

示例：

```
## 策略推荐: 表面改造(SR)策略

**推荐理由**:
1. **CDR构型保护**: 7D12的H2-10-1构型在人类种系中罕见 (<5%)。如果采用框架移植(BM)策略, 强制移植到常见人类框架可能导致CDR构型改变, 影响抗原结合。
```

2. **临床先例支持**: 已批准的H2-10-1 VHH (如Ozoralizumab、Caplacizumab) 均采用SR策略, 证明了这一策略的可行性。

3. **免疫原性优势**: SR变体保持了Native的低免疫原性 (预测MHC-II结合负荷=6), 而BM变体显著增加 (16, 是SR的2.7倍)。

4. **结构验证**: 基于PDB 4KRL结构分析, 67%的SR突变位点在结构中表面暴露, 支持策略合理性。

风险提示:

- SR策略的人类序列相似性相对较低 (78.6% vs BM的84.7%), 但临床数据表明这一差异不影响免疫原性表现

- 疏水斑块风险略高, 但可通过制剂优化管理

4. 人源化结果总览 (Humanization Results) - 2-3页

目的: 展示所有变体的对比

包含内容:

- 变体对比表 (Native/SR/BM)
- 关键指标对比 (免疫原性、可开发性、突变数)
- 与临床基准的对比 (百分位排名)
- 突变规则、评分算法、内部计算流程

示例:

```
## 人源化结果总览
```

变体	人类身份	免疫原性*	可开发性评分	疏水斑块风险	突变数	推荐等级
Native	72.4%	6 (低)	80	0.889 (高)	0	对照
SR	78.6%	6 (低)	78	0.889 (高)	6	★★★ 推荐
BM	84.7%	16 (中)	78	0.889 (高)	12	★ 对比

*免疫原性: 预测的MHC-II强结合体数量 (排名≤1%)

与临床基准对比:

- SR变体的免疫原性处于临床SR分子的第29百分位 (中位数13, 范围0-19)
- SR变体的可开发性评分处于第57百分位 (中位数77, 范围67-83)
- 疏水斑块风险处于第100百分位 (最高值), 但Gefurulimab和Tarpcerprumig等临床分子也达到此水平

5. 突变详情 (Mutation Details) - 2-3页

目的: 提供可执行的突变列表

包含内容:

- 突变位置和残基变化 (IMGT位置、区域、Native→SR)
- 突变目的 (降低免疫原性、改善可开发性)
- 结构验证结果 (表面暴露/掩埋)
- 位置集定义、突变规则、评分细节

示例：

```
## SR变体突变详情
```

IMGT位置	区域	Native	SR	目的	结构状态
12	FR1	S	L	人源化	表面暴露
40	FR2	G	S	人源化	掩埋
42	FR2	F	V	人源化	掩埋
83	FR3	A	S	人源化	表面暴露
96	FR3	P	A	人源化	表面暴露
101	FR3	I	V	人源化	表面暴露

****关键保护**：**

- 所有CDR区域 (CDR1、CDR2、CDR3) 完全保留
- VHH标志性位置 (37、44、45、47) 保留
- 结构锚点 (23、41、104) 保留
- Vernier区域 (28、29、94) 保留

****结构验证**：**

基于PDB 4KRL结构分析，6个突变位点中有4个 (67%) 表面暴露，支持SR策略的合理性。

6. 可选优化方案 (Optional Optimizations) - 1页

目的： 提供进一步改进的建议

包含内容：

- 具体优化建议 (单点突变、组合突变)
- 预期效果 (指标改善)
- 风险评估 (对功能的影响)
- 优化算法、搜索策略

示例：

```
## 可选优化方案
```

****方案1：降低疏水斑块风险****

- ****突变**：** IMGT101 V→S
- ****效果**：** hp_max9从0.889降至0.778 (达到临床SR中位数)
- ****风险**：** 低 (表面暴露位置，不涉及CDR)
- ****免疫原性**：** 无变化 (B_total_1pct保持为6)

****方案2：进一步提升人类身份****

- **突变**: IMGT5 E→L, IMGT15 T→P (BM变体中的额外突变)
- **效果**: 人类身份从78.6%提升至84.7%
- **风险**: 中等 (可能影响CDR构型, 需实验验证)
- **建议**: 仅在SR变体功能验证后进行

推荐顺序:

1. 优先验证SR变体的抗原结合亲和力
2. 如功能保持, 考虑方案1 (低风险优化)
3. 方案2需谨慎, 建议先进行结构建模验证

7. 最终建议 (Final Recommendations) - 1页

目的: 提供清晰的行动建议

包含内容:

- 推荐序列 (可直接用于合成)
- 实验验证建议 (优先级排序)
- 风险缓解措施
- 内部工作流程、算法参数

示例:

```
## 最终建议
```

**推荐序列 (SR变体) **:

```
QVKLEESGGGLVQTGGSLRLTCAASGRTSRSYGMWSVRQAPGKEREFVSGISWRGDSTGYADSVKGRFTISRDNSKNTVDL  
QMNSLKAEDTAVYYCAAAAGSAWYGTLYEYDYWGQGTQVTVSSALE
```

**实验验证建议 (优先级排序) **:

1. **高优先级**:
 - 抗原结合亲和力 (SPR/BLI) : 验证CDR保留后功能是否保持
 - 表达和纯化: 验证可制造性
2. **中优先级**:
 - 热稳定性 (DSC) : 评估T_m值
 - 聚集倾向 (SEC) : 评估单体比例
3. **低优先级**:
 - 免疫原性实验 (PBMC回忆检测) : 验证计算预测
 - 长期稳定性: 评估制剂兼容性

风险缓解措施:

- 疏水斑块风险: 通过制剂优化 (表面活性剂、pH调节) 管理

- 免疫原性：如需要，可考虑HSA融合或表位屏蔽策略
- CDR构型：建议进行结构建模（AlphaFold2）验证

8. 术语解释 (Glossary) - 1页

目的：帮助客户理解专业术语

包含内容：

- 术语定义（通俗易懂）
- 应用场景
- 算法细节、技术实现

示例：

术语解释

表面改造(SR)策略：仅对预测为溶剂暴露的框架位置进行人源化，保留内部核心位置的骆驼序列。适用于CDR构型独特或罕见的VHH。

框架移植(BM)策略：将CDR序列移植到人类框架模板上，随后对关键位置进行选择性回变。适用于CDR构型与常见人类种系匹配的VHH。

免疫原性预测：基于MHC-II结合预测，评估序列中可能引发T细胞反应的区域。数值越低，风险越低。

可开发性评分：综合评估序列的CMC（化学、制造和控制）风险，包括聚集倾向、稳定性等。评分越高，可开发性越好。

疏水斑块：序列中连续的疏水残基区域，可能导致蛋白质聚集。通过9-mer窗口扫描识别最差区域。

二、撰写原则

应该做的

1. 结果导向

- 先说“是什么”，再说“为什么”
- 用数据说话（百分位、对比值）
- 提供可执行的建议

2. 应用价值

- 强调临床相关性（与已批准分子对比）
- 提供风险评估和缓解措施
- 给出明确的下一步行动

3. 专业呈现

- 使用清晰的表格和图表
- 结构化的章节组织
- 一致的术语和格式

✗ 不应该做的

1. 暴露算法细节

- ✗ "使用Shannon熵计算表面可塑性位置"
- ☑ "基于人类种系序列变异分析识别可人源化位置"

2. 暴露内部工具

- ✗ "使用ANARCI v1.4进行IMGT编号"
- ☑ "采用标准IMGT编号系统进行序列分析"

3. 暴露数据库名称

- ✗ "从Thera-SAbDab数据库获取序列"
- ☑ "基于公开的临床序列数据"

4. 暴露评分算法

- ✗ "使用公式 $H(X) = -\sum p_i \log_2(p_i)$ 计算熵"
- ☑ "通过序列变异分析评估位置可塑性"

5. 暴露内部规则

- ✗ "根据surface_plasticity_positions_v1_strict集合选择突变位点"
 - ☑ "基于结构分析和临床先例选择表面暴露位置"
-

三、语言风格

推荐表达

- ☑ "基于结构分析" (而非"使用Shrake-Rupley算法计算SASA")
- ☑ "与临床基准对比" (而非"使用ECDF函数计算百分位")
- ☑ "预测的免疫原性" (而非"IEDB MHC-II API预测结果")
- ☑ "序列变异分析" (而非"Shannon熵计算")
- ☑ "结构引导的设计" (而非"基于relSASA阈值的过滤")

避免表达

- ✗ 算法名称 (ANARCI、IEDB、AlphaFold2等)
 - ✗ 数据库名称 (Thera-SAbDab、IMGT等)
 - ✗ 技术细节 (API调用、版本号、参数设置)
 - ✗ 内部代码 (函数名、变量名、文件路径)
-

四、报告长度建议

- **执行摘要**: 1-2页
- **输入序列分析**: 1页
- **策略推荐**: 2-3页
- **结果总览**: 2-3页
- **突变详情**: 2-3页
- **可选优化**: 1页
- **最终建议**: 1页
- **术语解释**: 1页

总计: 12-16页 (不含附录)

五、质量检查清单

在提交报告前，检查以下项目：

- 是否包含执行摘要？
 - 是否提供了推荐方案和理由？
 - 是否包含与临床基准的对比？
 - 是否提供了可执行的突变列表？
 - 是否包含风险评估和缓解措施？
 - 是否避免了算法名称和数据库名称？
 - 是否避免了技术细节和内部代码？
 - 术语是否通俗易懂？
 - 表格和图表是否清晰？
 - 建议是否可执行？
-

六、示例对比

✗ 不好的写法 (暴露算法)

```
## 方法
```

本研究使用ANARCI v1.4对输入序列进行IMGT编号，然后从Thera-SAbDab数据库获取19个临床VHH序列。通过计算Shannon熵 $H(X) = -\sum p_i \log_2(p_i)$ ，识别出surface_plasticity_positions_v1_strict集合中的22个位置。使用IEedb MHC-II API（默认11个HLA-DRB1等位基因）进行免疫原性预测，调用格式为POST /tools_api/mhcii/。

✓ 好的写法 (结果导向)

```
## 分析结果
```

基于标准IMGT编号系统，对输入序列和19个临床VHH进行了系统分析。通过序列变异分析和结构验证，

识别出22个适合人源化的表面位置。免疫原性评估显示，SR变体的预测风险处于临床SR分子的第29百分位，属于低风险水平。

七、总结

CRO客户报告的核心：

1. **结果第一**: 突出发现和建议
2. **应用导向**: 强调临床价值和可执行性
3. **专业呈现**: 结构清晰、数据充分
4. **保护IP**: 不暴露算法和技术细节

记住: 客户关心"是什么"和"怎么用"，而非"怎么算"。