

# 分子优化-Chem-CoT 数据集质量评测

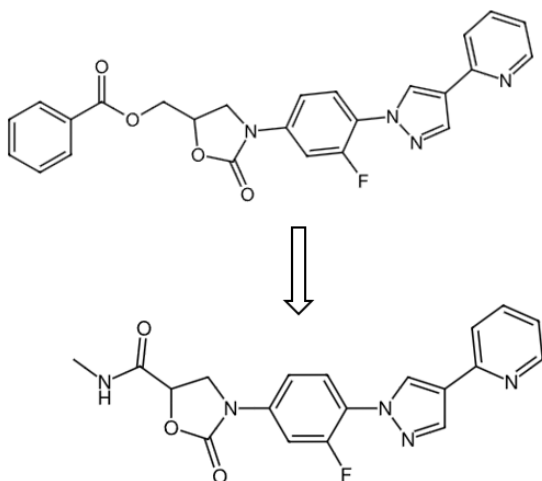
**目标：**评测单目标分子优化任务的Benchmark对应的sample，Deepseek生成的思维链质量如何，并修改错误位置

下面是一个其中的例子：

1. 我们可视化了QED属性对应的large分子对以及他的思维链，发现在对源分子进行结构分析时（思维链第一段）会存在一些结构上的识别错误（主要是环结构有问题）
2. 为此，我们需要：
  - a. 记录整个思维链中针对**分子片段(苯甲基 / 吡啶 ...)**的描述一共有多少个，记录在total\_entity中
  - b. 查看这些**分子片段是否错误**，记录错误数量在error\_entity中
  - c. 修改这些错误的分子片段

## • Chain-of-Thoughts Distillation Samples: QED-large

```
"src_smiles": "O=C(OCC1CN(C2CCC(-N3CC(-C4CCCCN4)CN3)C(F)C2)C(=O)O1)C1CCCCC1",  
"tgt_smiles": "CNC(=O)C1CN(C2CCC(-N3CC(-C4CCCCN4)CN3)C(F)C2)C(=O)O1",
```



1. 分子A的分析：
- 结构概要：中心吗啉衍生环，带有氟苯基-三唑-吡啶取代基、苯甲酸酯和环状酯基。
  - QED的局限性：芳环数量较多（4个），分子量较大，以及由于疏水基团导致的潜在logP值。
2. QED分析：
- QED的主要特征：芳环过多（4个，而理想 $\alpha$ 值较低），氢键受体较多（酯类、吡啶类、三唑类），以及柔性链上的可旋转键。
  - A结构中的缺口：芳环过载（苯甲酯、氟苯基、三唑类、吡啶类），分子量较高，以及酯类主导的极性降低了药物相似性。
3. 分子A的优化：
- 用甲基酰胺取代苯甲酸酯，以消除一个芳环并降低分子量/logP。
  - 将环状酯转化为氨基甲酸酯，以保持环的稳定性，同时降低氢键受体。
  - 保留氟苯基-三唑-吡啶用于靶标结合，但简化取代基以平衡疏水性。
4. 最终分子B: CNC(=O)C1CN(C2CCC(-N3CC(-C4CCCCN4)CN3)C(F)C2)C(=O)O1

我们会提供一个 json文件，包含20个large-molecule的CoT标注，改哪些位置如下（即 cot\_result, total\_entity, error\_entity）

```
{
  "src": "O=C(CCN1CCN(CCOC(c2ccccc2)c2ccccc2)CC1)c1ccco1",
  "tgt": "O=C(CCN1CCN(CCOC(c2ccccc2)c2ccccc2)CC1)c1ccc(Br)cc1",
  "src_logd": 3.4487500985063706,
  "tgt_logd": 4.3170710274715365,
  "src_solubility": 1.9096424292618417,
  "tgt_solubility": 0.1979279118522929,
  "src_clint": 2.6662705233451,
  "tgt_clint": 2.204315680331008,
  "src_qed": 0.46014746523748207,
  "tgt_qed": 0.3421730579143774,
  "src_drd2": 0.22388610909267082,
  "tgt_drd2": 0.6673706773943849,
  "src_jnk3": 0.0,
  "tgt_jnk3": 0.0,
  "src_gsk": 0.01,
  "tgt_gsk": 0.0,
  "type": "large",
  "raw_cot": "",
  "cot_result": {
    "\n  \"Structural Analysis of Source Molecule\": \"The source molecule contains a central piperazine ring linked to a diphenylethoxy group and a ketone-substituted furan ring. The furan ring (c1ccco1) is unsubstituted and lacks electron-withdrawing/halogen substituents, which are common in DRD2-active compounds. The diphenylethoxy group provides lipophilicity but may reduce selectivity.\",\n  \"Property Analysis\": \"DRD2 activity requires optimal aromatic/hydrophobic interactions and balanced polarity. Unsubstituted furan rings often exhibit weaker receptor binding due to insufficient electron density modulation and limited hydrophobic surface area.\",\n  \"Limitation in Source Molecule for Property\": \"The unsubstituted furan ring (c1ccco1) reduces binding affinity by limiting hydrophobic interactions and electron density effects critical for DRD2 engagement. The absence of halogen substituents further diminishes potential halogen-bonding interactions.\",\n  \"Optimization for Source Molecule\": \"Introduce a bromine atom at the para position of the furan ring to enhance hydrophobic interactions, electron-withdrawing effects, and potential halogen bonding. This modification retains the core scaffold while improving DRD2 binding.\",\n  \"Final Target Molecule\": \"O=C(CCN1CCN(CCOC(c2ccccc2)c2ccccc2)CC1)c1ccc(Br)cc1\"
  },
  "total_entity": 0,
  "error_entity": 0
},
```

分子

看这

实体