

DENV2 vs DENV3 结构分析使用指南

快速开始

1. 安装依赖



```
pip install biopython requests numpy matplotlib
```

2. 运行第一阶段（序列分析）



```
python denv_structure_analysis.py
```

这会自动完成：

- ✓ 从UniProt下载DENV2和DENV3序列
- ✓ 序列比对和活性位点保守性分析
- ✓ 生成AlphaFold输入文件

3. AlphaFold预测（选择一种方式）

方式A：使用ColabFold（推荐，免费）

1. 访问 <https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>
2. 上传生成的 sequences/DENV2_NS3_protease.fasta
3. 设置参数：
 - num_models = 1 (只需要最佳模型)
 - use_templates = False (从头预测)
4. 运行预测（约5-10分钟）
5. 下载结果PDB文件，重命名为 DENV2_predicted.pdb

方式B：使用本地ColabFold（如已安装）



```
bash
```

```
colabfold_batch sequences/DENV2_NS3_protease.fasta structures/ --num-models 1
```

方式C：使用AlphaFold Server（最简单）

1. 访问 <https://alphafoldserver.com/>
2. 上传序列文件
3. 下载预测结果

4. 准备DENV3实验结构

从PDB数据库下载：



bash

```
# 示例: DENV3 NS2B-NS3蛋白酶复合物
```

```
wget https://files.rcsb.org/download/3U1I.pdb -O structures/DENV3_experimental.pdb
```

常见DENV3 NS2B-NS3结构PDB ID：

- 3U1I - 闭合构象，与抑制剂复合
- 2F0M - 开放构象
- 3U1J - 另一个抑制剂复合物

选择建议：优先使用与抑制剂复合的闭合构象结构

5. 运行第二阶段（结构比对）

确保文件结构如下：



```
structures/
├── DENV2_predicted.pdb
└── DENV3_experimental.pdb
```

然后运行：



bash

```
python denv_structure_analysis.py --continue
```

输出结果解读

1. 序列分析报告 (results/sequence_analysis.json)

- **序列同源性 >85%** → 两个血清型非常相似，可安全使用DENV3结构
- **序列同源性 70-85%** → 中等相似度，建议双轨验证
- **序列同源性 <70%** → 差异较大，优先用DENV2预测结构

催化三联体检查：

- His51, Asp75, Ser135 必须完全保守
-  如有突变，说明酶活机制可能不同

底物结合位点：

- 85% 保守 → 抑制剂结合模式可能相似
- <70% 保守 → 血清型特异性明显

2. 结构质量评估 (results/plddt_scores.png)

AlphaFold pLDDT评分标准：

- >90 - 高置信度，可与实验结构媲美
- 70-90 - 良好质量，主链位置可靠
- 50-70 - 低置信度，可能需要优化
- <50 - 无序或错误，不可用

重点关注活性位点区域的pLDDT值！

3. 结构叠合结果 (results/DENV2_aligned_to_DENV3.pdb)

RMSD (均方根偏差) 判断标准：

- 整体RMSD <2Å → 结构高度相似
- 整体RMSD 2-3Å → 中等相似度
- 整体RMSD >3Å → 结构差异显著
- 活性位点RMSD <1.5Å → 结合口袋几乎相同
- 活性位点RMSD 1.5-2.5Å → 有差异但可接受
- 活性位点RMSD >2.5Å → 结合模式可能不同

4. 决策建议报告 (results/decision_report.txt)

脚本会给出三种建议之一：

建议A：使用DENV3实验结构 ✓

触发条件：

- 序列同源性 >85%
- 催化三联体完全保守
- 活性位点保守性 >85%
- RMSD <2Å

实施方案：



toml

```
# 在REINVENT4配置中
[[scoring.component]]
[scoring.component.Maize]
[[scoring.component.Maize.endpoint]]
name = "DENV3 Docking"
weight = 0.5
params.workflow = "docking_DENV3.yaml" # 使用DENV3结构
```

建议B：双轨验证 ⚖

触发条件：

- 序列同源性 70-85%
- 活性位点部分保守
- 结构相似但有差异

实施方案：



toml

```
# 同时使用两个对接评分
[[scoring.component]]
[scoring.component.Maize]
[[scoring.component.Maize.endpoint]]
name = "DENV2 Docking"
weight = 0.3
params.workflow = "docking_DENV2.yaml"

[[scoring.component]]
[scoring.component.Maize]
[[scoring.component.Maize.endpoint]]
name = "DENV3 Docking"
weight = 0.2
params.workflow = "docking_DENV3.yaml"

# QSAR模型权重更高 (因为基于真实DENV2数据)
[[scoring.component]]
[scoring.component.QSARScorer]
[[scoring.component.QSARScorer.endpoint]]
name = "DENV2 pIC50"
weight = 0.5
params.model_path = "qsar_model.pkl"
```

建议C：优先DENV2预测结构

触发条件：

- 序列同源性 <70%
- 活性位点差异显著
- 催化三联体有突变 (严重情况)

实施方案：



toml

```
# 主要使用QSAR，对接仅作参考
[[scoring.component]]
[scoring.component.QSARScorer]
[[scoring.component.QSARScorer.endpoint]]
name = "DENV2 pIC50"
weight = 0.7
params.model_path = "qsar_model.pkl"

[[scoring.component]]
[scoring.component.Maize]
[[scoring.component.Maize.endpoint]]
name = "DENV2 Docking"
weight = 0.3
params.workflow = "docking_DENV2.yaml"
```

🎯 实战示例

场景1：高同源性（最理想）



序列同源性: 92.5%

催化三联体: ✓✓✓ 完全保守

活性位点保守性: 89.3%

整体RMSD: 1.42Å

活性位点RMSD: 0.87Å

pLDDT: 94.2

→ 推荐：直接使用DENV3实验结构

→ 理由：实验结构更可靠，保守性极高

场景2：中等相似度（需权衡）



序列同源性: 78.3%
催化三联体: ✓✓✓ 完全保守
活性位点保守性: 71.4%
整体RMSD: 2.31Å
活性位点RMSD: 1.89Å
pLDDT: 87.6

- 推荐: 双轨验证
- 先用DENV3快速筛选, 再用DENV2精确优化

场景3: 显著差异 (谨慎处理)



序列同源性: 64.7%
催化三联体: ✓✗✓ Asp75突变!
活性位点保守性: 58.9%
整体RMSD: 3.87Å
活性位点RMSD: 3.12Å
pLDDT: 91.3

- 推荐: 仅使用DENV2预测结构
- ⚠️ 催化机制可能不同, 需实验验证

🔧 高级技巧

1. 如果AlphaFold预测质量不佳



bash

```
# 使用模板建模 (以DENV3为模板)
# 在ColabFold中设置:
use_templates = True
# 或手动准备模板文件
```

2. 优化活性位点构象



python

```
# 使用分子动力学优化预测结构
# 示例：使用GROMACS进行短时MD
gmx pdb2gmx -f DENV2_predicted.pdb -o processed.gro
gmx editconf -f processed.gro -o boxed.gro -d 1.0
gmx solvate -cp boxed.gro -o solvated.gro
# ... 能量最小化和短时MD
```

3. 验证对接可靠性



python

```
# 用已知抑制剂验证对接方案
known_inhibitors = [
    "CC1=C(C(=NO1)C2=CC=CC=C2Cl)C(=O)NC3CCC(CC3)C(=O)O", # DENV抑制剂1
    # ... 更多已知活性化合物
]

# 检查对接能否重现实验结构
# 期望：已知抑制剂的对接分数与实验pIC50相关性 >0.6
```

4. 处理缺失loop区域



python

```
# 如果AlphaFold在loop区预测差 (pLDDT<50)
# 使用Modeller重建loop
from modeller import *
from modeller.automodel import *

env = Environ()
a = LoopModel(env, alnfile='alignment.ali',
              knowns='DENV3_template', sequence='DENV2_target',
              loop_assess_methods=assess.DOPE)
a.starting_model = 1
a.ending_model = 5
a.loop.starting_model = 1
a.loop.ending_model = 10
a.make()
```

完整文件结构



```
denv_analysis/
├── denv_structure_analysis.py    # 主脚本
├── sequences/
│   ├── DENV2_NS3.fasta          # 自动下载
│   ├── DENV3_NS3.fasta          # 自动下载
│   └── DENV2_NS3_protease.fasta # AlphaFold输入
└── structures/
    ├── DENV2_predicted.pdb      # 需手动放置
    ├── DENV3_experimental.pdb   # 需手动下载
    └── DENV2_aligned_to_DENV3.pdb # 自动生成
└── results/
    ├── sequence_analysis.json   # 序列比对结果
    ├── plddt_scores.png         # 质量评估图
    ├── decision_report.txt      # 最终建议
    └── final_analysis.json      # 完整数据
```

⚠ 常见问题

Q1: AlphaFold预测的是什么构象？

A: 通常是最稳定的构象（可能是开放态）。如果需要闭合态：

- 使用DENV3闭合态作为模板建模
- 或从开放态进行分子动力学模拟诱导闭合

Q2: pLDDT很高但RMSD很大？

A: 说明预测质量好，但两个血清型真实结构确实不同。这种情况应该：

- 使用DENV2预测结构（匹配你的活性数据）
- 不要强行用DENV3结构

Q3: 催化三联体有突变怎么办？

A: 这是严重问题！建议：

1. 仔细检查序列比对是否正确
2. 查文献确认该突变是否真实存在
3. 如果确实突变，考虑两个血清型的抑制机制可能不同
4. 必须使用DENV2特异性的结构和数据

Q4: 如何在REINVENT4中集成?

A: 将推荐的结构配置到对接工作流:



```
# MAIZE docking workflow (docking_workflow.yaml)
graph:
- docking:
  receptor: structures/DENV2_predicted.pdb # 或 DENV3_experimental.pdb
  box_center: [15.2, 23.7, 10.5] # 活性位点坐标
  box_size: [20, 20, 20]
  exhaustiveness: 8
```

然后在REINVENT4中引用:



```
[[scoring.component]]
[scoring.component.Maize]
[[scoring.component.Maize.endpoint]]
name = "DENV Protease Docking"
weight = 0.4
params.executable = "maize"
params.workflow = "docking_workflow.yaml"
transform.type = "reverse_sigmoid"
transform.high = -6.0
transform.low = -10.0
transform.k = 0.3
```

🚀 下一步行动

根据脚本输出的建议:

如果推荐使用DENV3:

1. 下载最佳的DENV3-抑制剂复合物结构
2. 准备对接配置 (定义结合口袋)
3. 用10-20个已知抑制剂验证对接
4. 整合到REINVENT4配置

如果推荐双轨验证：

1. 同时准备两个对接配置
2. 先用DENV3快速生成候选（计算快）
3. 再用DENV2精确评估（更准确）
4. 对比两种方案的生成结果

如果推荐仅用DENV2：

1. 检查预测结构的活性位点质量
2. 如需要，进行MD优化
3. 考虑增加QSAR权重 (>0.6)
4. 对接仅作为几何筛选

需要更多帮助？

-  检查 `results/decision_report.txt` 查看详细建议
-  查看 `results/final_analysis.json` 获取所有数值
-  查看 `results/plddt_scores.png` 评估预测质量

祝分子设计顺利！  