

# REINVENT4 COMPREHENSIVE ANALYSIS REPORT

*De Novo Drug Design for DENV NS2B-NS3 Protease Inhibitors*

---

---

## EXECUTIVE SUMMARY

---

Report Generated: 2025-10-18 11:47:33

Total Runs Analyzed: 3

Runs: archive, run3, run4

Total Gold Candidates: 0

Total High Quality: 0

Total Good Quality: 0

---

---

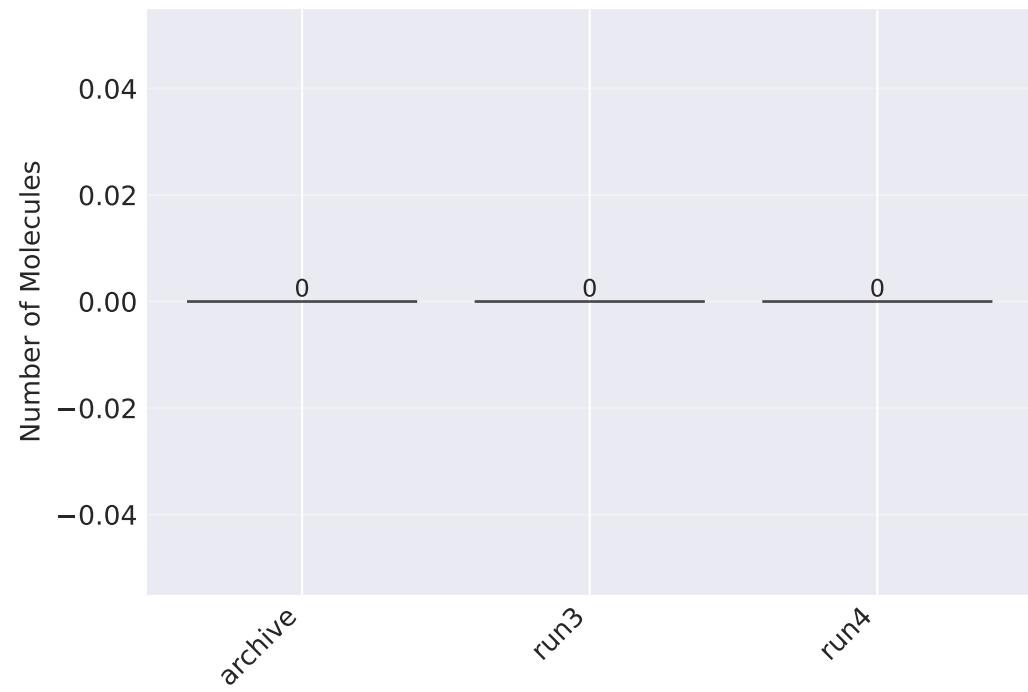
## TABLE OF CONTENTS

---

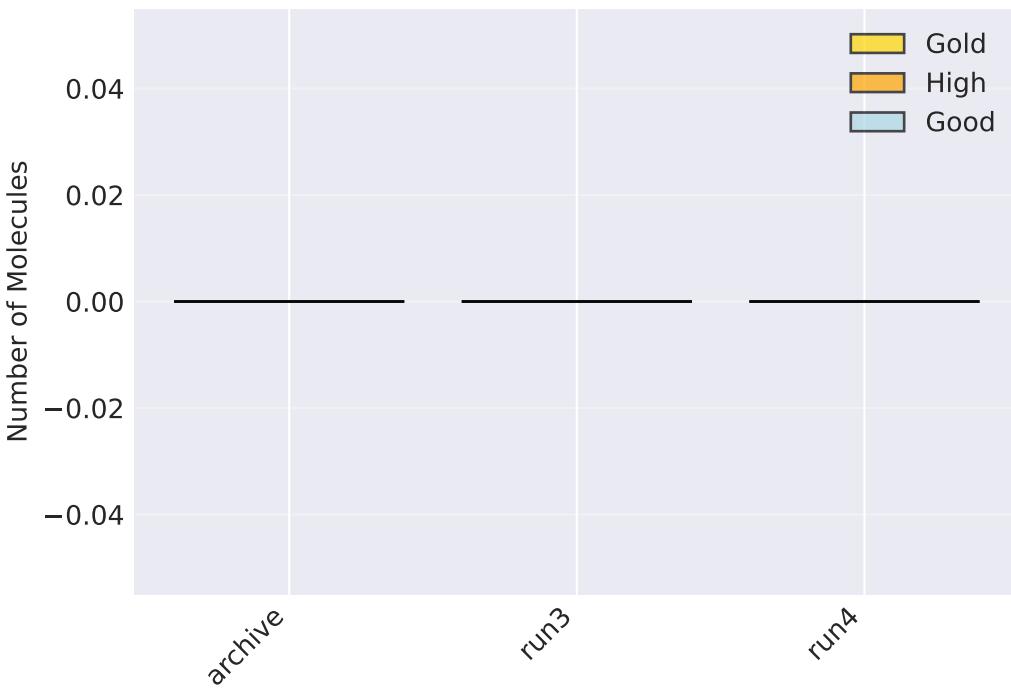
1. Gold Standard Molecules Comparison
2. Molecular Structure Comparison
3. Drug-likeness Metrics Comparison
4. Activity & QSAR Prediction Comparison
5. Molecular Properties Comparison
6. Model Optimization Trends
7. Configuration Differences Analysis
8. Detailed Individual Run Reports

# Gold Standard Molecules Comparison

Gold Candidates Count  
(Score  $\geq$  threshold & Pass filters)



Candidate Quality Distribution



Gold Candidate Success Rate



## GOLD STANDARD CRITERIA:

- ✓ Total Score  $\geq$  threshold
- ✓ QSAR predicted pIC50  $\geq 6.0$
- ✓ QED  $\geq 0.5$  (drug-likeness)
- ✓ Molecular Weight: 200-500 Da
- ✓ Pass structural alerts
- ✓ Lipinski's Rule of 5 compliant
- ✓ Synthetic Accessibility  $< 6$

## HIGH QUALITY CRITERIA:

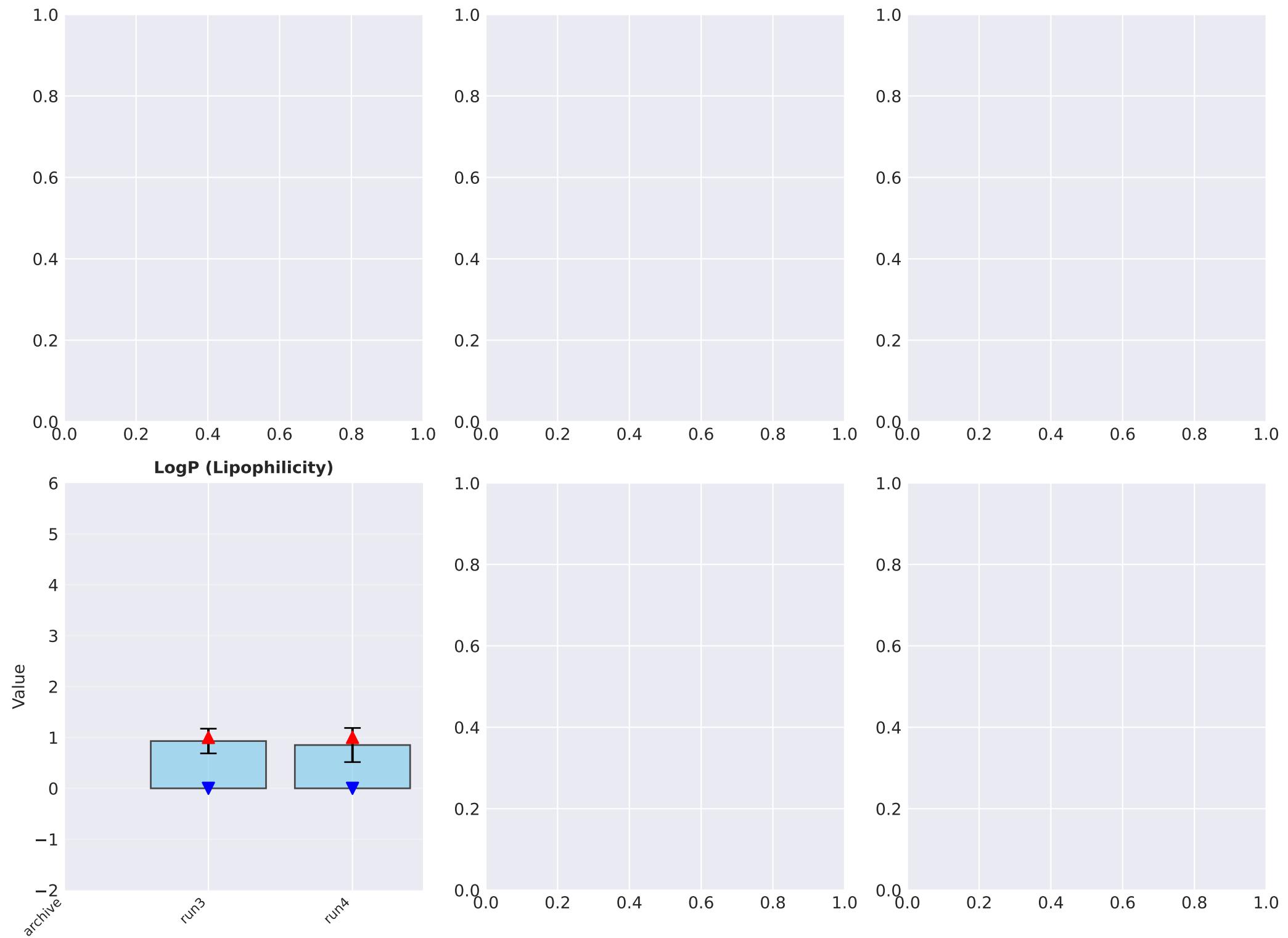
- ✓ Total Score  $\geq 0.7$
- ✓ QSAR pIC50  $\geq 5.5$
- ✓ QED  $\geq 0.4$

## GOOD QUALITY CRITERIA:

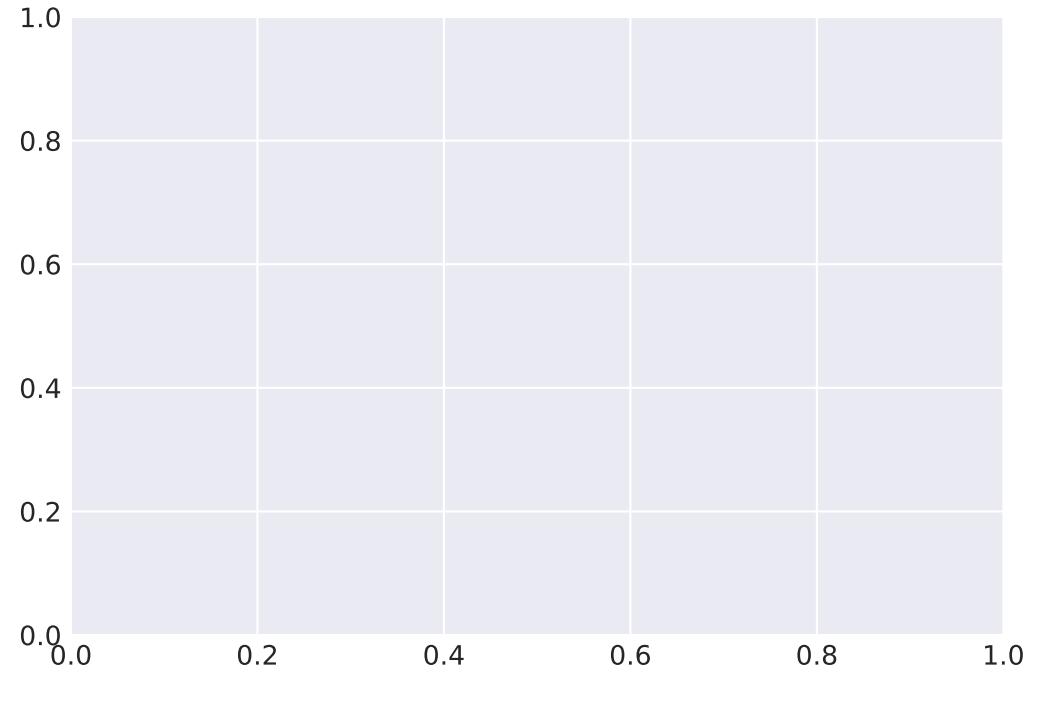
- ✓ Total Score  $\geq 0.5$
- ✓ Pass basic filters

# Molecular Structures Comparison Across Run

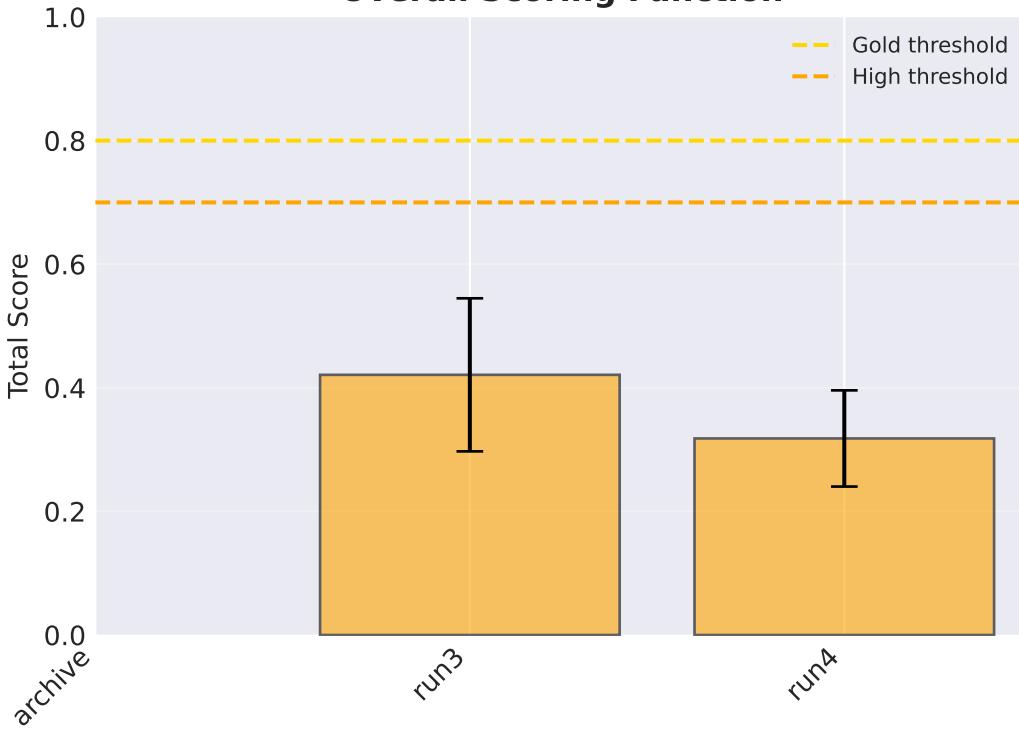
# Drug-likeness Metrics Comparison



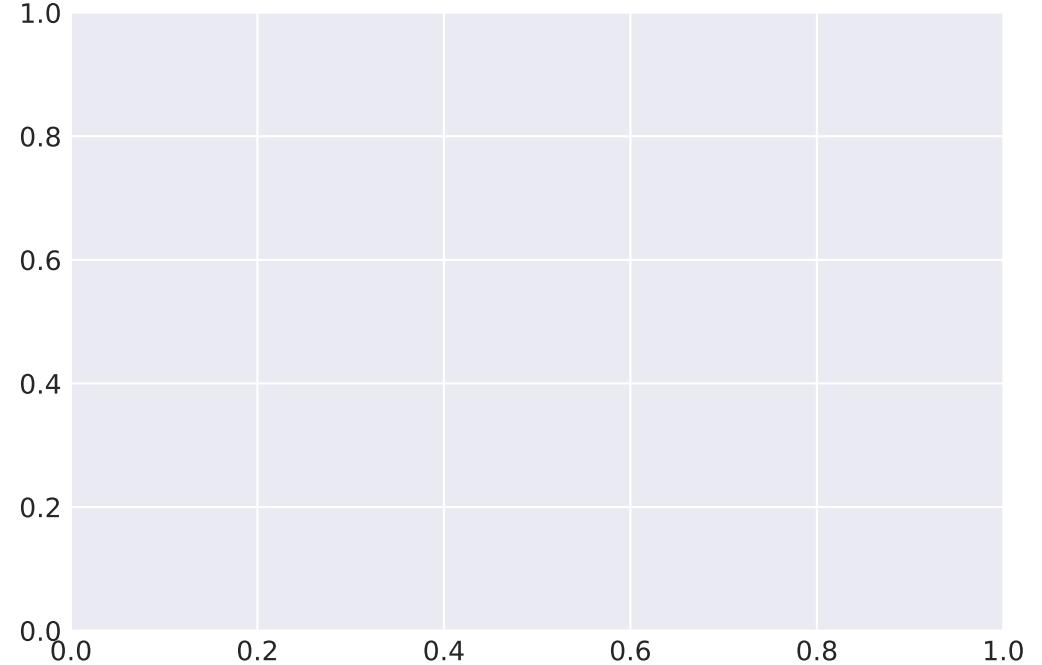
# Activity & QSAR Prediction Comparison



## Overall Scoring Function

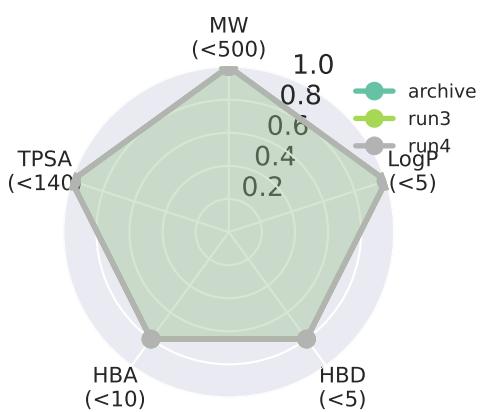
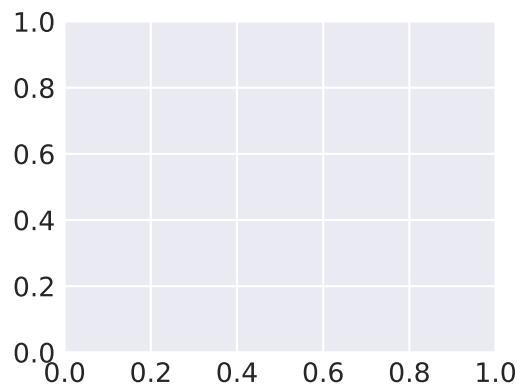
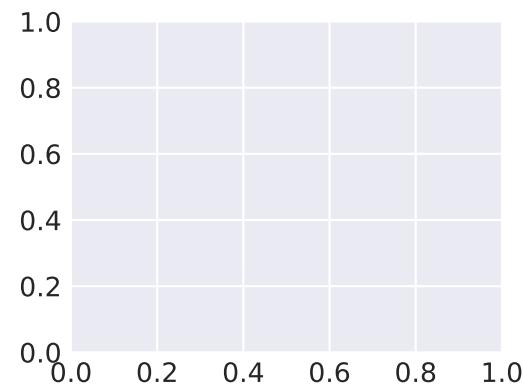
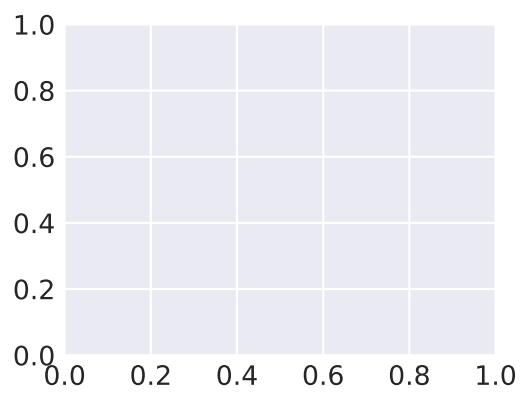
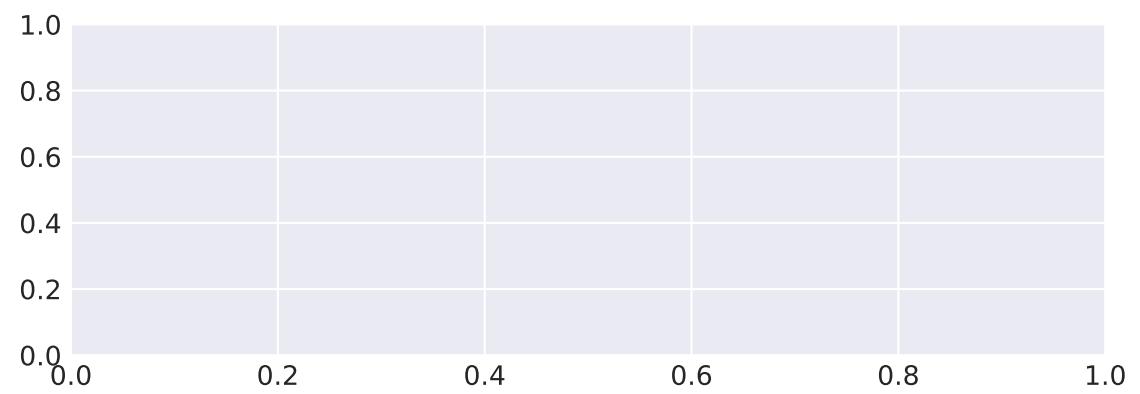


## Activity Statistics Summary



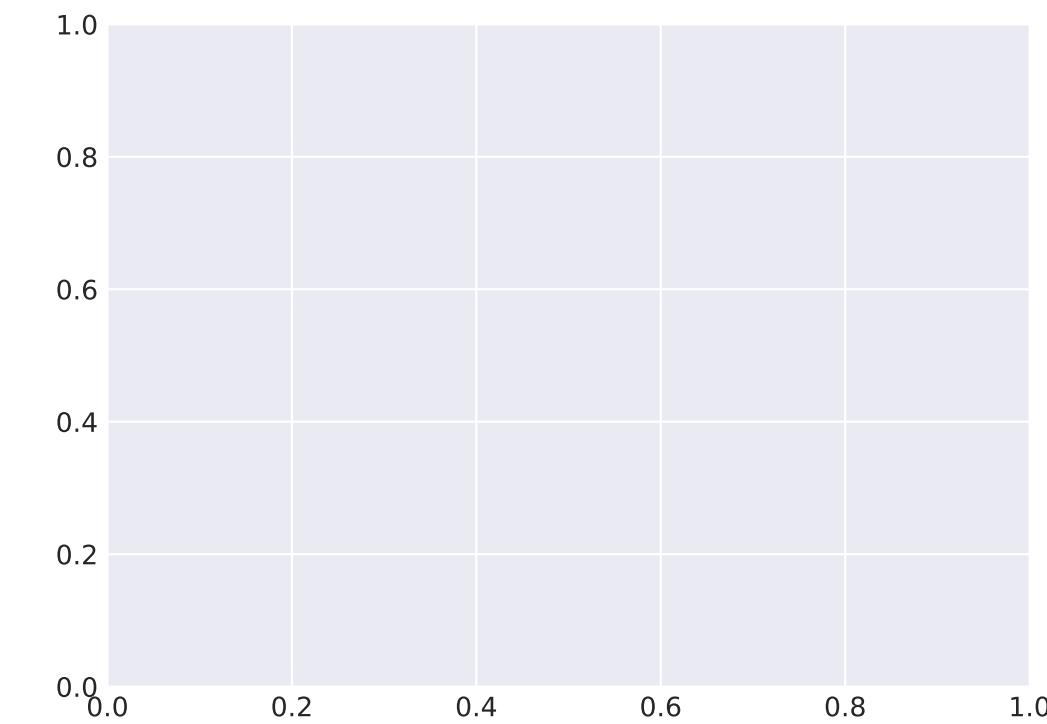
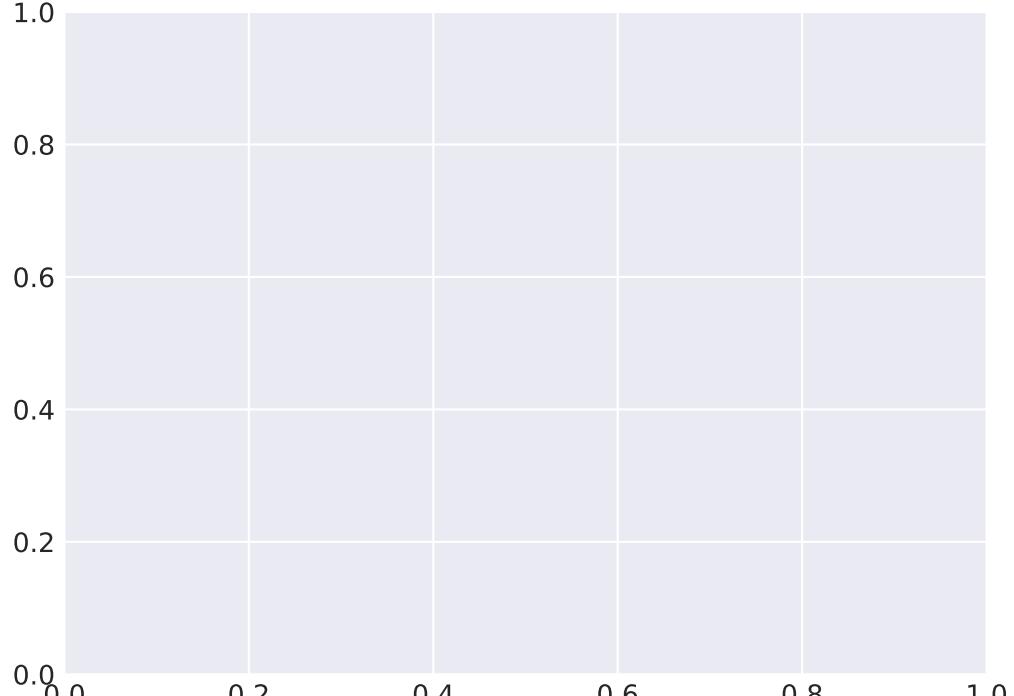
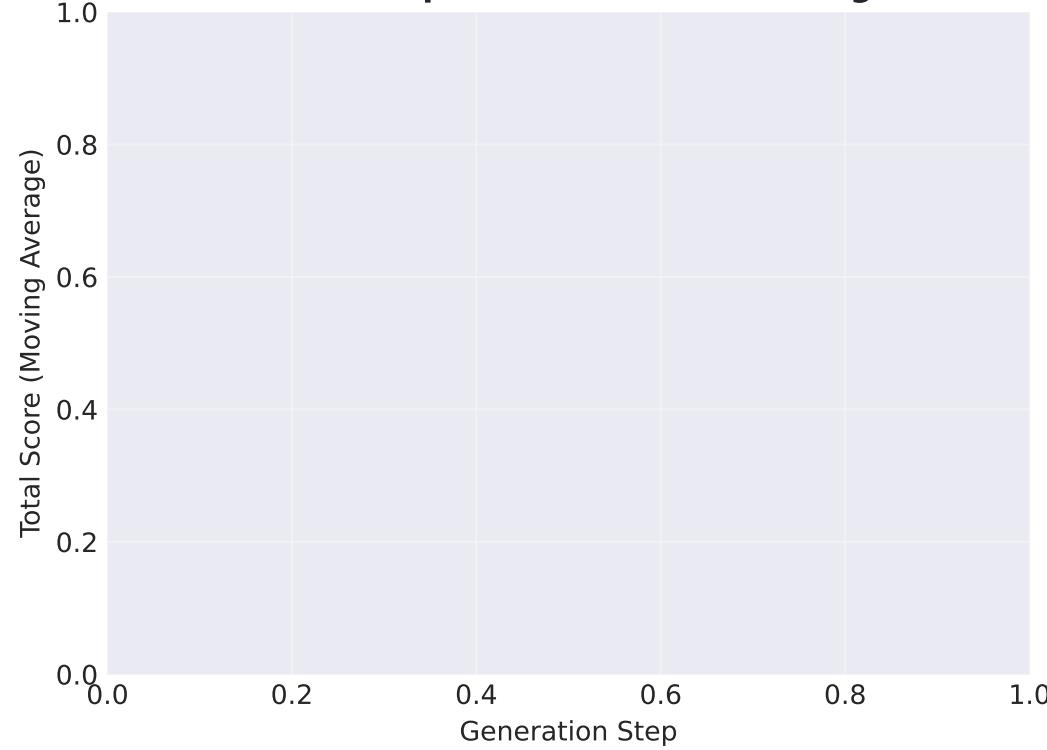
| Run     | pIC50<br>(Mean±SD) | Max<br>pIC50 | Mean<br>Score | N<br>Gold |
|---------|--------------------|--------------|---------------|-----------|
| archive | 0.00±0.00          | 0.00         | nan           | nan       |
| run3    | 0.00±0.00          | 0.00         | 0.42          | 0         |
| run4    | 0.00±0.00          | 0.00         | 0.32          | 0         |

# Molecular Properties Detailed Comparison



# Model Optimization & Convergence Analysis

## Score Improvement Over Training



### OPTIMIZATION QUALITY METRICS:

#### CONVERGENCE CRITERIA

- ✓ Fast: < 500 steps to 80% max
- ✓ Stable: CV < 5% in final phase
- ✓ Improved: Final > Initial by 20%

### MODEL FITTING INDICATORS:

- Low CV% → Good model stability
- High final score → Effective optimization
- Smooth curves → Proper learning rate
- Early plateau → May need longer training

### COMPARISON INSIGHTS:

# Configuration Differences & Optimization Strategy

## Configuration Parameters Comparison (Yellow = Different values)

| Run     | Run Type        | Device | Scoring | LR     | Sigma  | Batch |
|---------|-----------------|--------|---------|--------|--------|-------|
| archive | nan             | nan    | nan     | nan    | nan    | nan   |
| run3    | staged_learning | cuda:0 | nan     | 0.0001 | 120.00 | 64.0  |
| run4    | staged_learning | cuda:0 | nan     | 0.0001 | 120.00 | 64.0  |

## Scoring Components Matrix (✓ = Component used)

| Component   | Run 1 | Run 2 | Run 3 | Run 4 |
|-------------|-------|-------|-------|-------|
| Component A | ✓     |       |       |       |
| Component B |       | ✓     |       |       |

### KEY CONFIGURATION DIFFERENCES:

#### OPTIMIZATION PARAMETERS

#### SCORING STRATEGY

### OPTIMIZATION RECOMMENDATIONS:

- LEARNING RATE:
  - Too high → Unstable training
  - Too low → Slow convergence
  - Optimal: 0.0001 - 0.001
- SIGMA (Reward shaping):
  - Controls exploration vs exploitation
  - Higher  $\sigma$  → More exploration
  - Typical range: 60 - 120
- BATCH SIZE:
  - Larger → More stable gradients
  - Smaller → Faster updates
  - Balance: 50 - 200
- SCORING COMPONENTS:
  - More components → Better filtering
  - But may reduce diversity
  - Balance quality vs quantity
- INSIGHTS FROM COMPARISON:

Best performing: run3  
Gold candidates: 0  
→ Used LR: 0.0001  
→ Used  $\sigma$ : 120.0

# Detailed Analysis: archive

RUN: archive

Directory: experiments/runs/archive

---

## SUMMARY STATISTICS

---

Total Molecules Generated: N/A

Gold Candidates: 0

High Quality: 0

Good Quality: 0

---

## MOLECULAR PROPERTIES (Mean $\pm$ Std)

---

QSAR pIC50:  $0.00 \pm 0.00$   
Range: [0.00, 0.00]

QED Score:  $0.000 \pm 0.000$

Molecular Wt:  $0.0 \pm 0.0$  Da

LogP:  $0.00 \pm 0.00$

SA Score:  $0.00 \pm 0.00$

Total Score:  $0.000 \pm 0.000$

---

## CONFIGURATION

---

## Detailed Analysis: run3

RUN: run3  
Directory: experiments/runs/run3

## SUMMARY STATISTICS

Total Molecules Generated: 122551  
Gold Candidates: 0  
High Quality: 0  
Good Quality: 0

### MOLECULAR PROPERTIES (Mean $\pm$ Std)

QSAR pIC50:       $0.00 \pm 0.00$   
                   Range: [0.00, 0.00]  
  
 QED Score:         $0.000 \pm 0.000$   
  
 Molecular Wt:      $0.0 \pm 0.0$  Da  
  
 LogP:               $0.93 \pm 0.24$   
  
 SA Score:           $0.00 \pm 0.00$   
  
 Total Score:        $0.421 \pm 0.124$

## CONFIGURATION

```
Run Type:  
  staged_learning  
  
Device:  
  cuda:0  
  
Learning Rate:  
  0.0001
```

## Detailed Analysis: run4

RUN: run4  
Directory: experiments/runs/run4

## SUMMARY STATISTICS

Total Molecules Generated: 125965  
Gold Candidates: 0  
High Quality: 0  
Good Quality: 0

## MOLECULAR PROPERTIES (Mean $\pm$ Std)

|               |                     |
|---------------|---------------------|
| QSAR pIC50:   | $0.00 \pm 0.00$     |
|               | Range: [0.00, 0.00] |
| QED Score:    | $0.000 \pm 0.000$   |
| Molecular Wt: | $0.0 \pm 0.0$ Da    |
| LogP:         | $0.85 \pm 0.34$     |
| SA Score:     | $0.00 \pm 0.00$     |
| Total Score:  | $0.318 \pm 0.078$   |

## CONFIGURATION

```
Run Type:  
    staged_learning  
  
Device:  
    cuda:0  
  
Learning Rate:  
    0.0001  
  
Sigma:  
    120.0
```