

# Results Overview

- **Designs generated: 1**

- Designs passing all filters: 0

- **MMCIF files of final designs are in:**

output/final\_ranked\_designs/final\_1\_designs

- **Metrics and sequences of 1 final designs are in:**

output/final\_ranked\_designs/final\_designs\_metrics\_1.csv

- **Metrics of all designs are in:**

output/final\_ranked\_designs/all\_designs\_metrics.csv

**You can rerun filtering (very quick), using this command with changed parameters:**

```
-- boltzgen run input_spec.yaml --steps filtering --config filtering budget=60 alpha=0.05
```

You can also rerun filtering in a jupyter notebook if you want using `filter.ipynb`

## **What was run to produce this in the Filter task:**

1. Filtering: each design is evaluated against mandatory thresholds.
2. Ranking: for every metric we compute its rank, then scale it by the metric's inverse-importance weight.

Designs with fewer passed filters are automatically penalised because the ranking key is the pair (num\_filters\_passed, metric). The overall quality score is the worst (maximum) of these scaled ranks.

The 10 best designs form the Top set.

3. Diversity: a lazy-greedy algorithm selects 1 designs that jointly maximise quality and minimise sequence similarity (sequence-identity distance). The trade-off is controlled by  $\alpha = 0.001$ :

- $\alpha = 0 \rightarrow 100\%$  quality focus (same as Top set)
- $\alpha = 1 \rightarrow 100\%$  diversity focus (ignores quality)
- Quality – composite of metrics such as iPTM, salt-bridges,  $\Delta$ SASA, etc. Each metric has an "inverse importance" weight (see "Sorting Criteria" table). A larger weight divides the rank by a bigger number and therefore down-weights that metric.
- Diversity – 1 – sequence identity between designs.

We use  $\alpha = 0.001$ , meaning 100 % emphasis on quality, 0 % on diversity.

# **CSV Column Reference**

Column	Description
id	filename to retrieve the file
design_sequence	designed amino acids (may be subset of chain)
designed_chain_sequence	full sequence of the chain containing designed residues (recommended for synthesis)
num_design	number of designed residues
secondary_rank	intermediate rank from the sorting procedure
design_ptm	predicted TM score for intra-design contacts (higher = better)
design_ip tm	predicted TM score for design-target contacts (higher = better)
design_to_target_ip tm	same as design_ip tm but for multi-chain designs
min_design_to_target_pae	minimum PAE between design & target (lower = better)
plip_saltbridge	number of salt-bridge interactions
plip_hbonds	number of hydrogen-bond interactions
plip_hydrophobic	number of hydrophobic interactions
delta_sasa_original	$\Delta$ SASA when binder present vs absent
delta_sasa_refolded	same as above but on the refolded structure

# Filtering & Sorting Criteria

This section first lists threshold filters that every design must pass (Filtering Criteria table) then explains how designs are ranked (Sorting Criteria table) weighted by their inverse importance.

# Filtering Criteria

feature	lower_is_better	threshold	Pass
has_x	True	0.0	1
filter_rmsd	True	2.5	0
filter_rmsd_design	True	2.5	1
CYS_fraction	True	0.0	1
ALA_fraction	True	0.3	1
GLY_fraction	True	0.2	1
GLU_fraction	True	0.2	1
LEU_fraction	True	0.3	1
VAL_fraction	True	0.2	1

## Sorting Criteria

Metric	Inverse Importance
design_to_target_ip tm	1
design_ptm	1
neg_min_design_to_target_pae	1
plip_hbonds_refolded	2
plip_saltbridge_refolded	2
delta_sasa_refolded	2

# **Results Summary**

## *Aggregate statistics*

Quick numeric overview comparing (i) ALL incoming designs, (ii) the top-10 highest-quality designs, and (iii) the 1 quality+diversity designs produced by the lazy-greedy selection.

Metric	Mean	Mean top 10	Mean top 1 diverse
Num designs	1	-	-
num_design	35.000	35.000	35.000
filter_rmsd	21.339	21.339	21.339
design_ip tm	0.127	0.127	0.127
design_to_target_ip tm	0.138	0.138	0.138
min_design_to_target_pae	21.684	21.684	21.684
delta_sasa_refolded	1771.332	1771.332	1771.332
plip_saltbridge_refolded	3.000	3.000	3.000
plip_hbonds_refolded	13.000	13.000	13.000



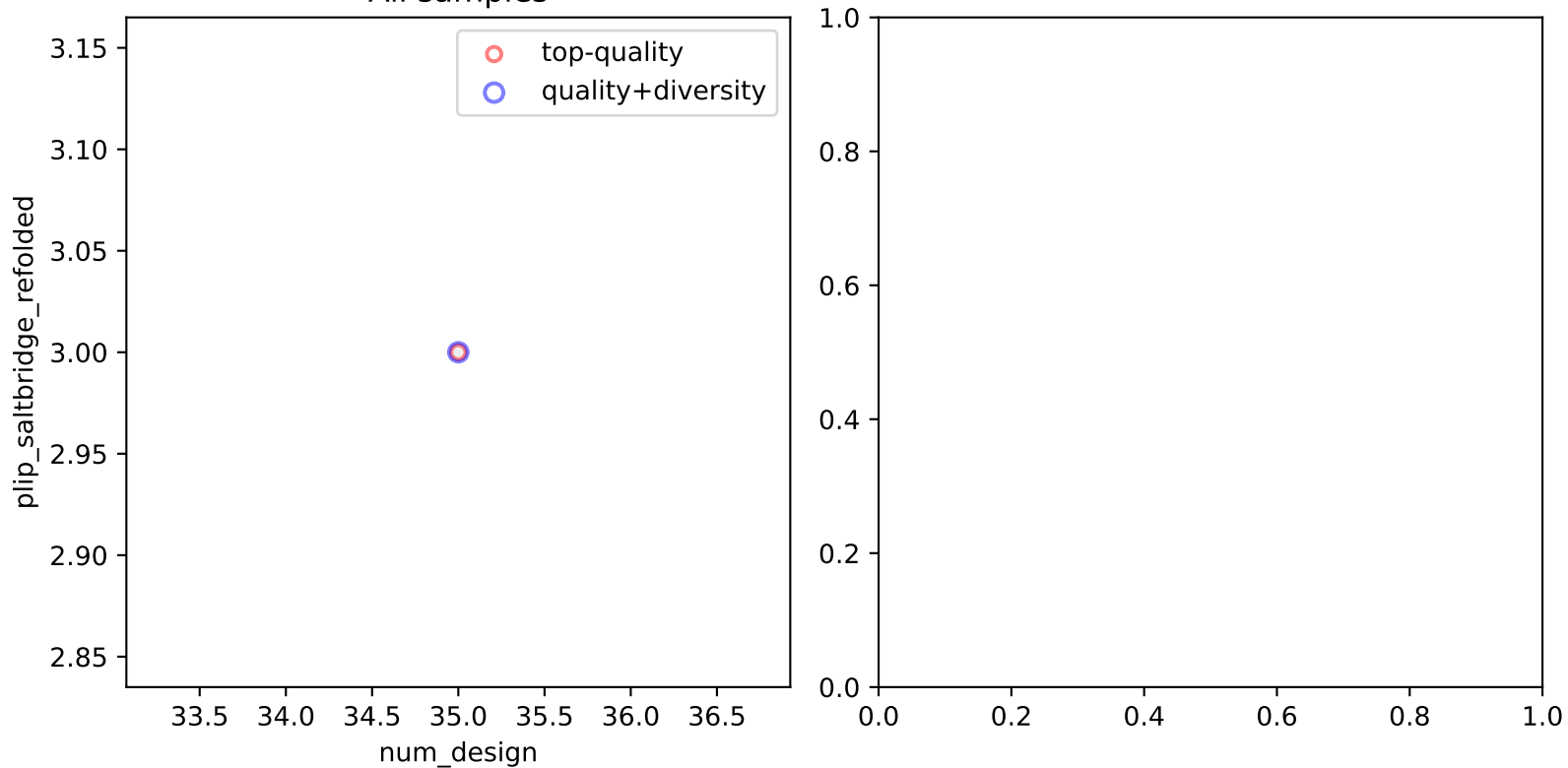
# Scatter Plots

## *Metric relationships*

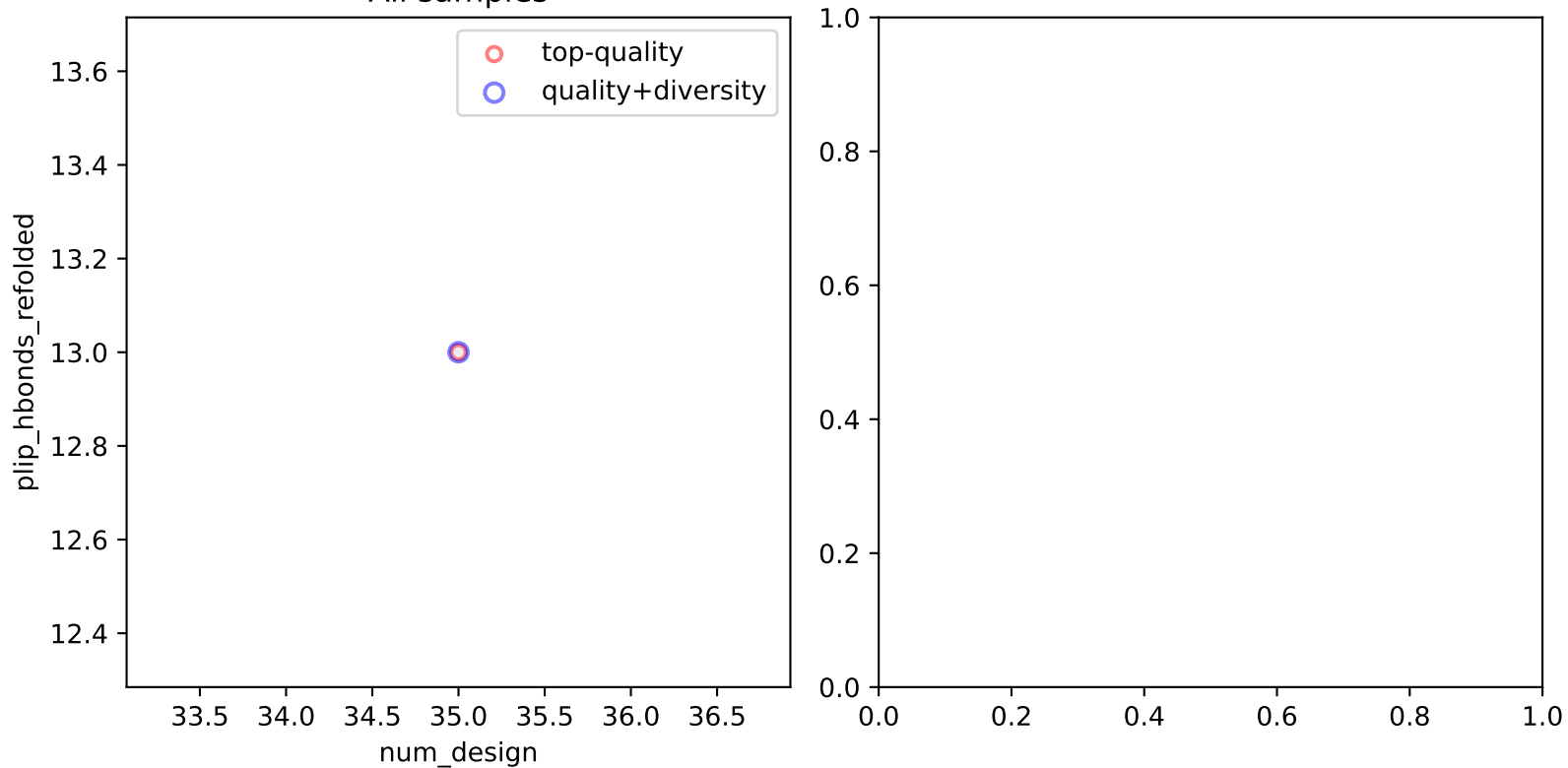
Each scatter page contains two panels:

- Left – all designs (grey) with overlays of Top (red) and Diverse (blue).
- Right – same but limited to designs passing the RMSD filter.

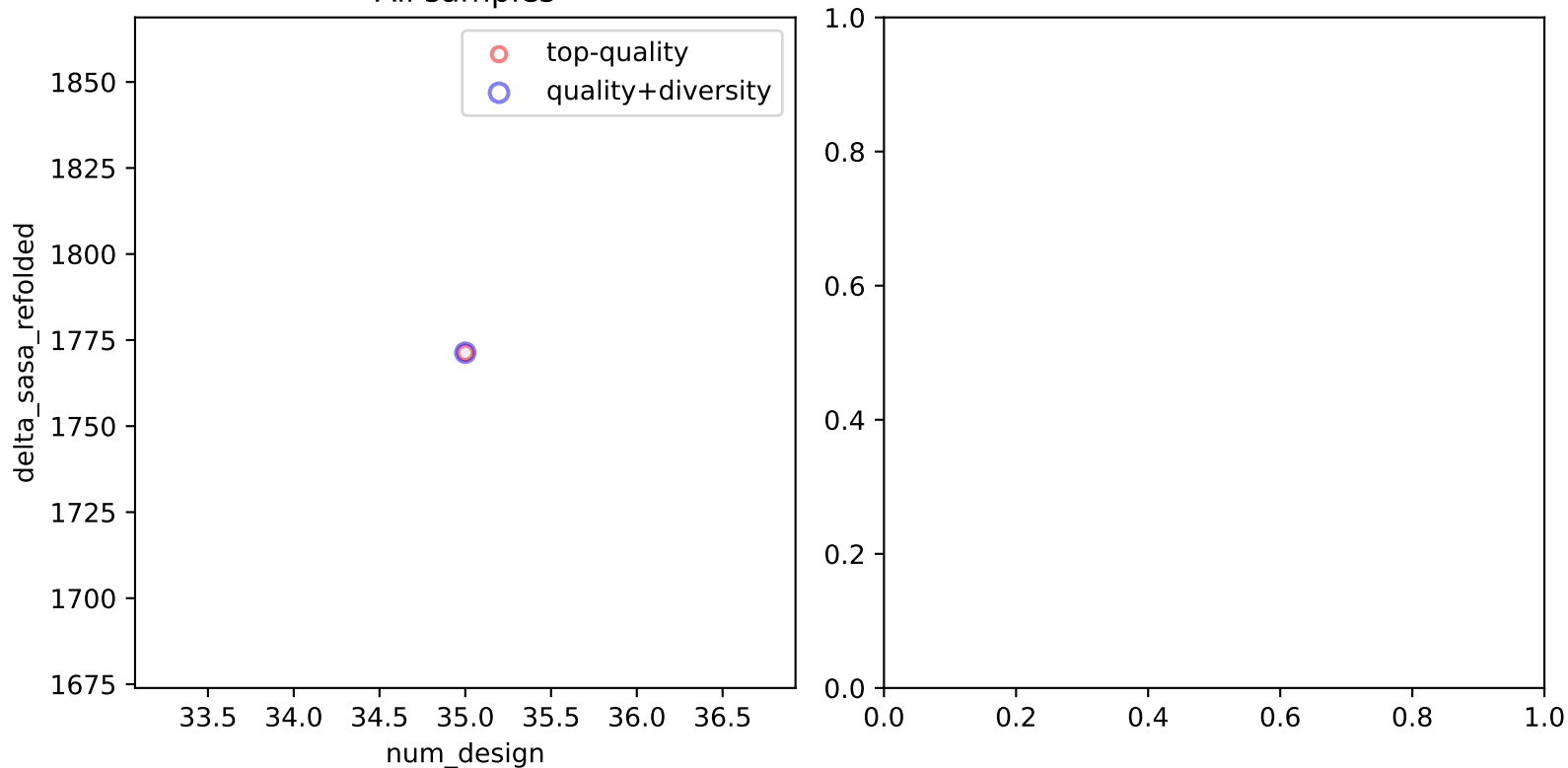
All samples



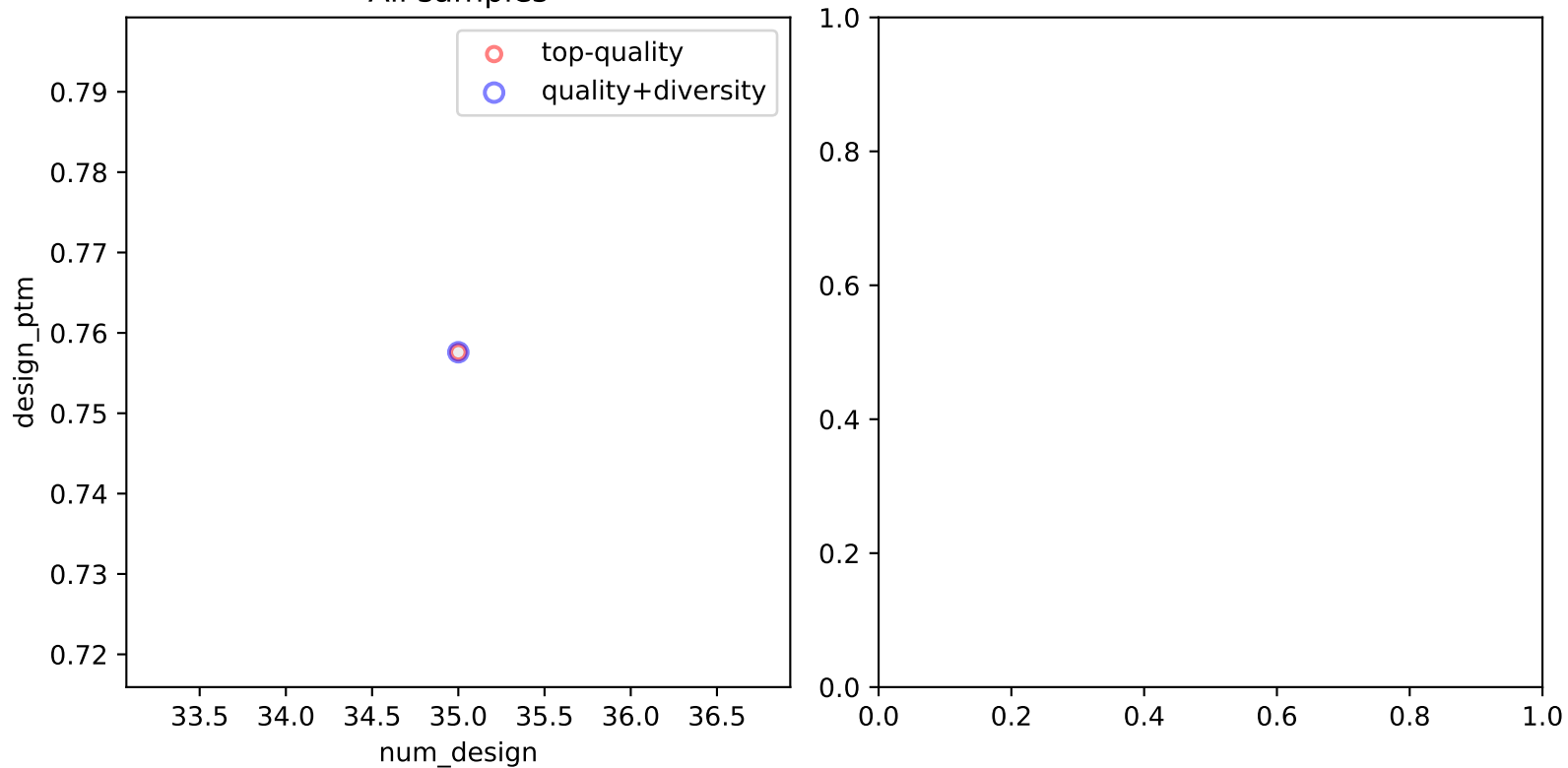
All samples



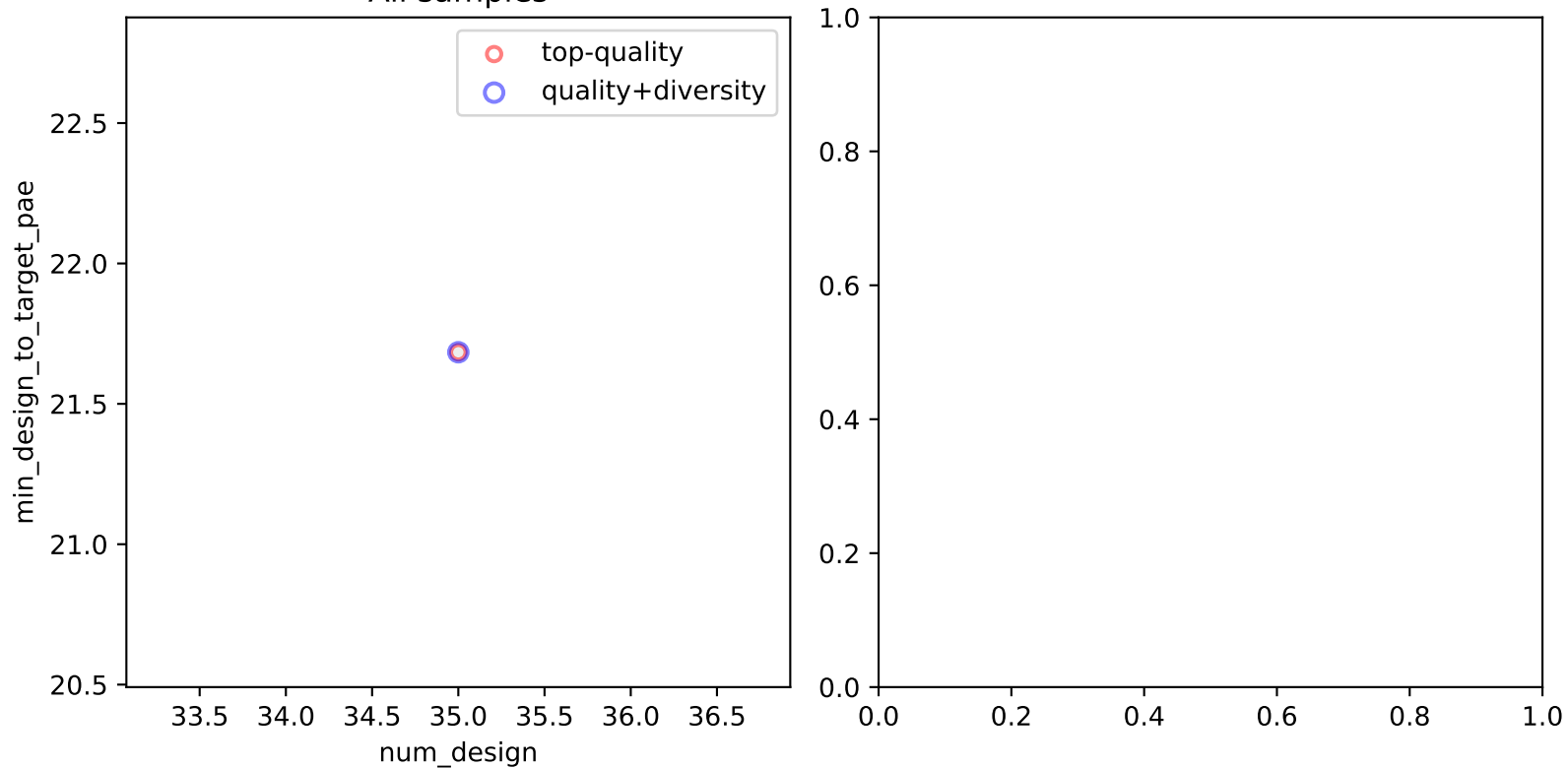
All samples



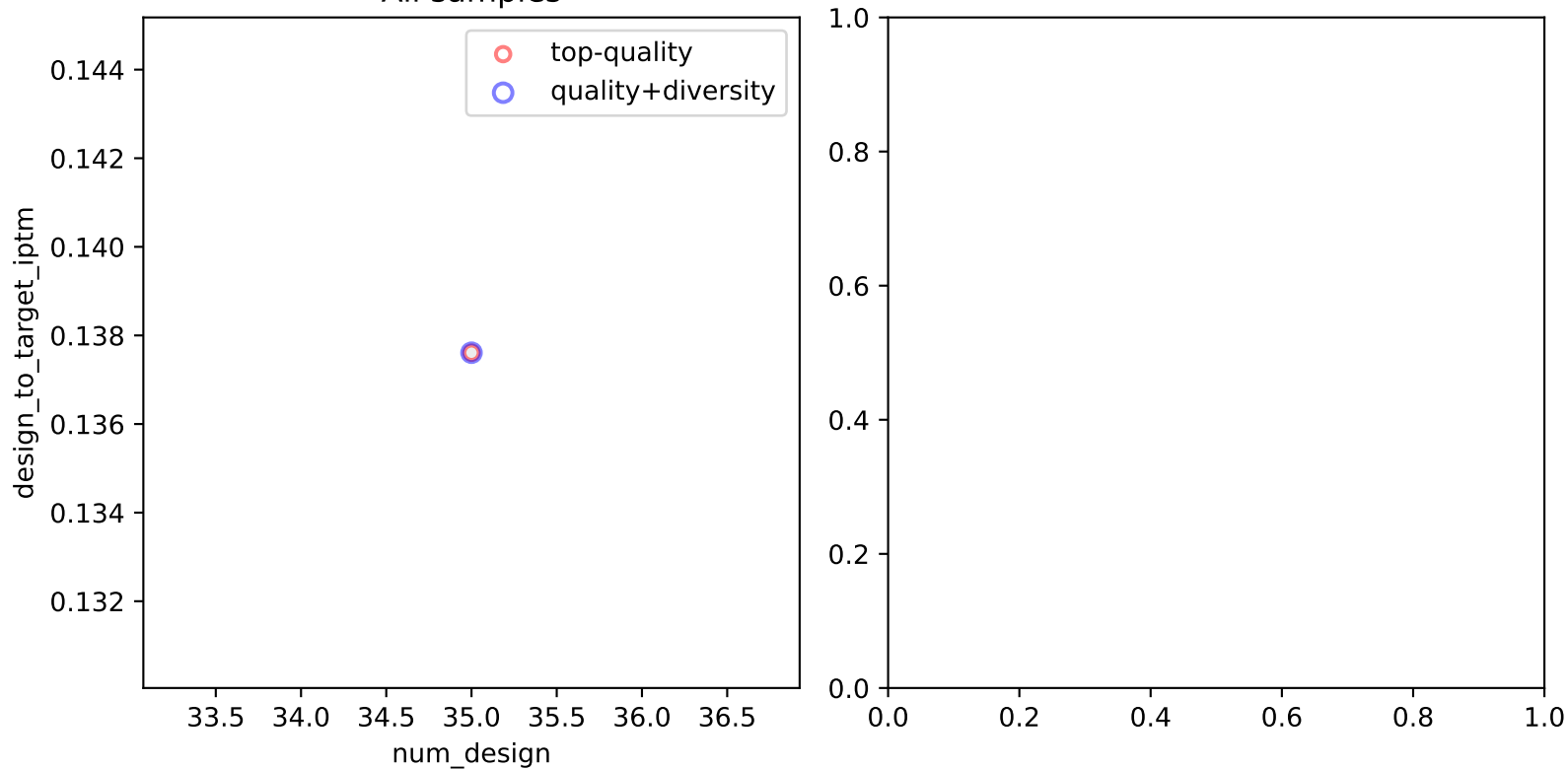
All samples



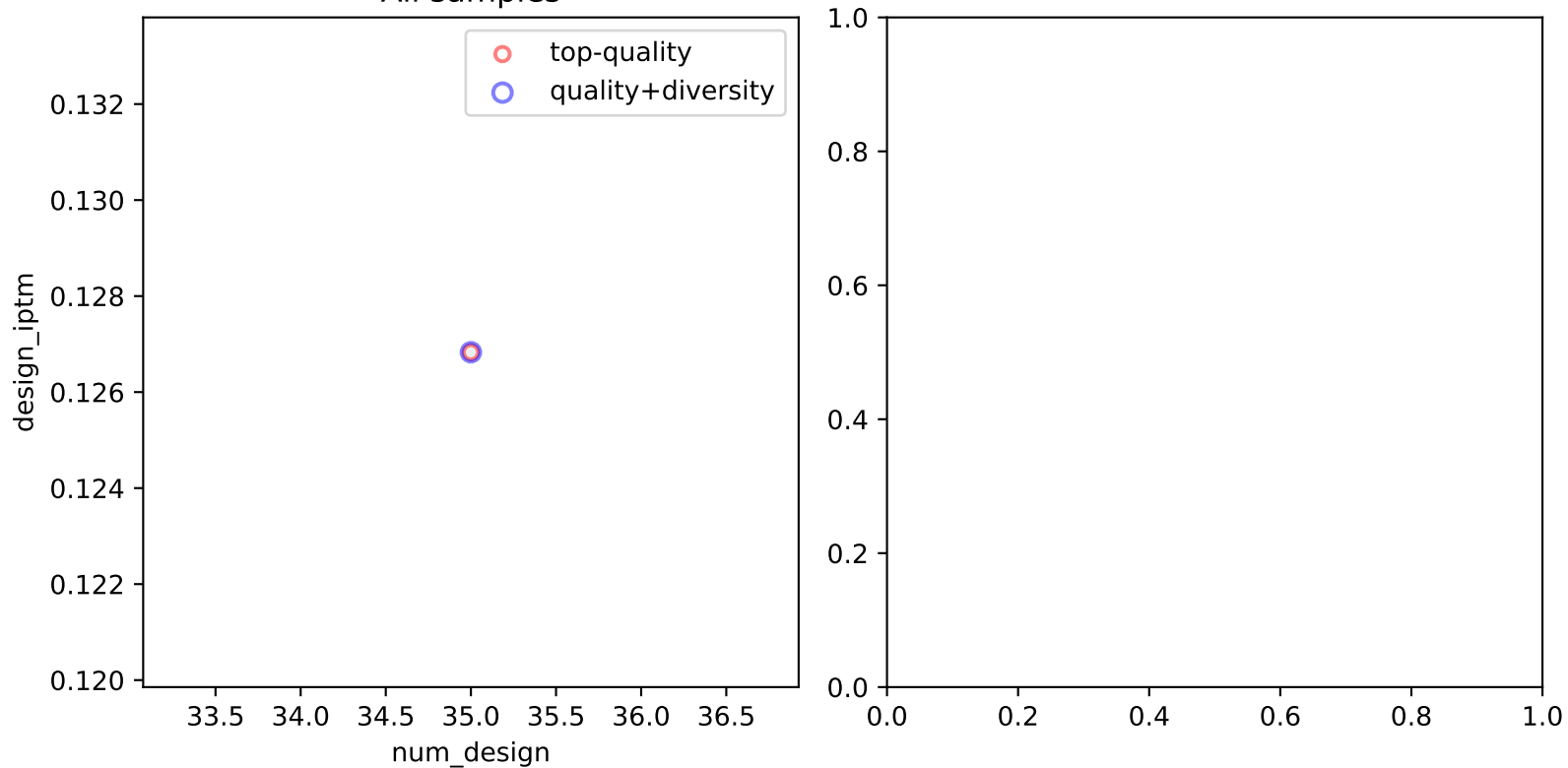
All samples



All samples

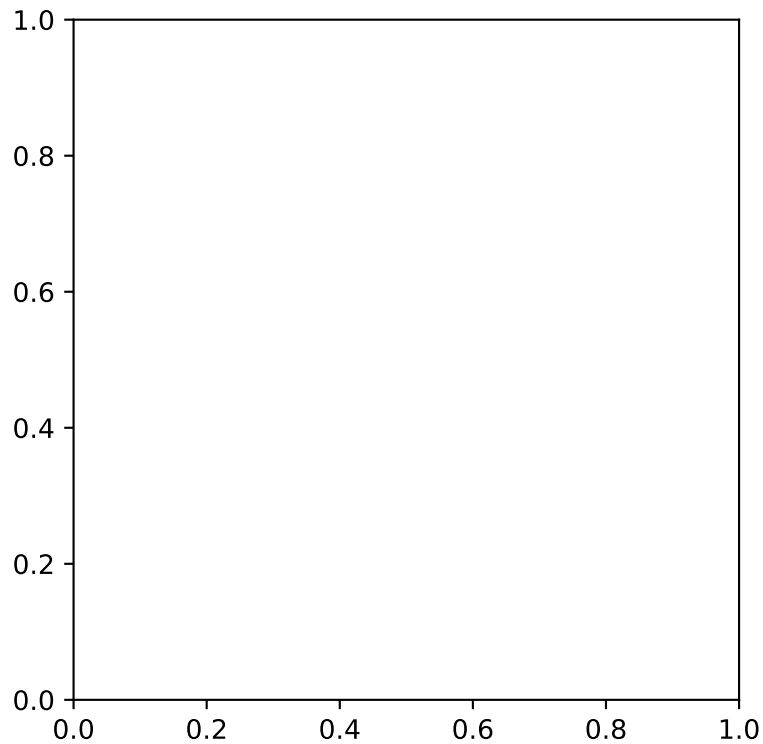
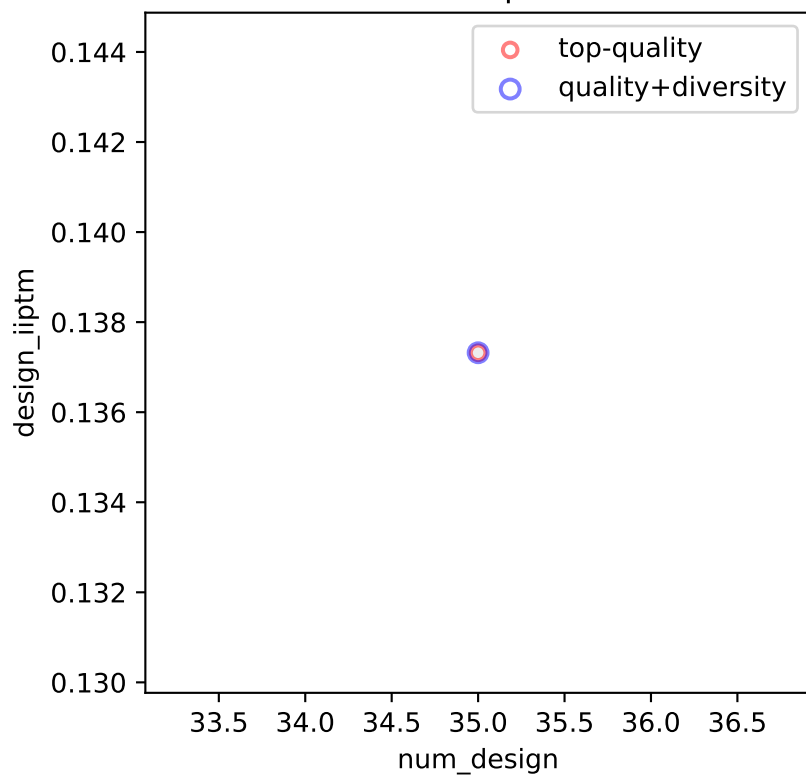


All samples

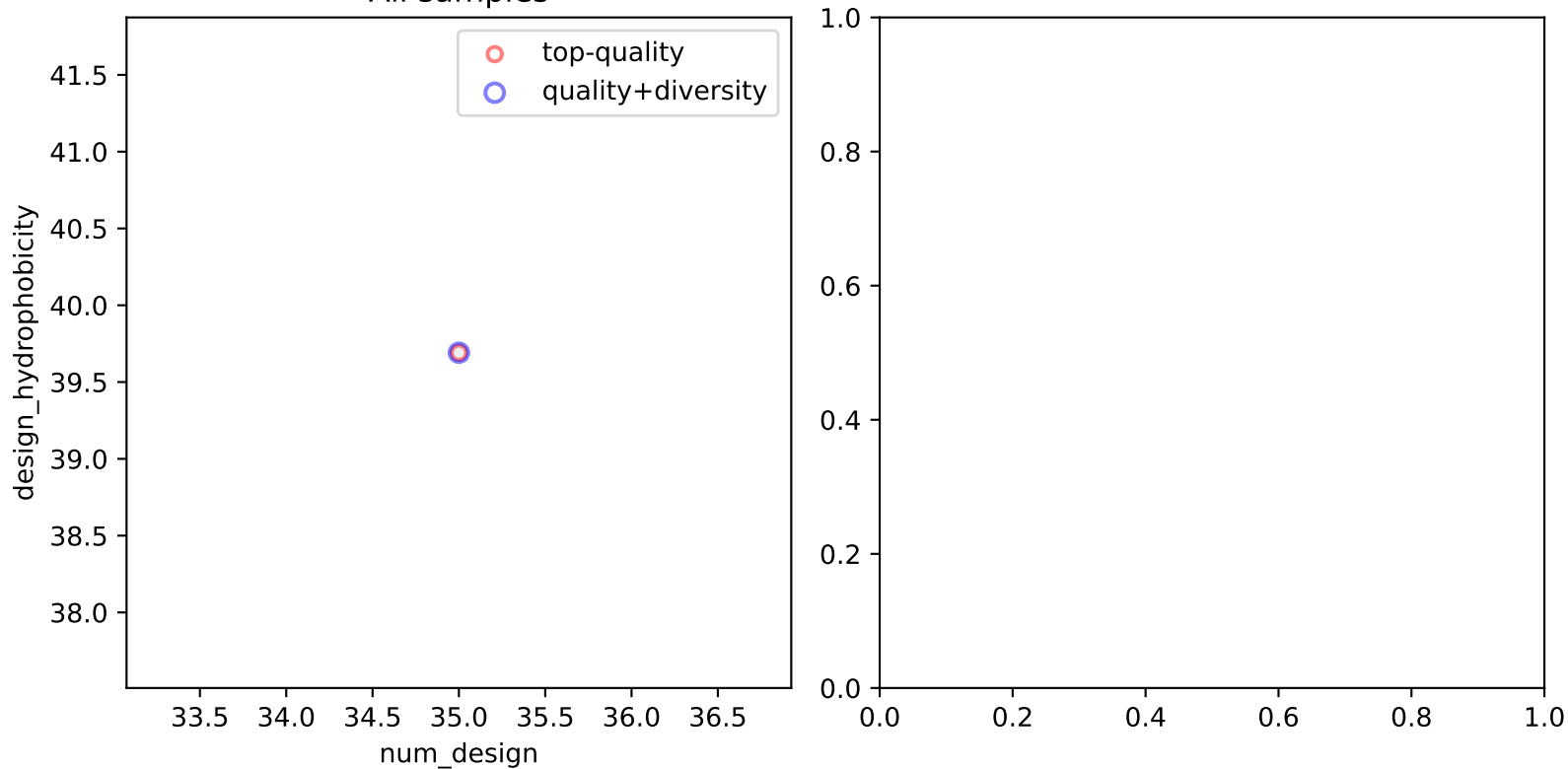




All samples



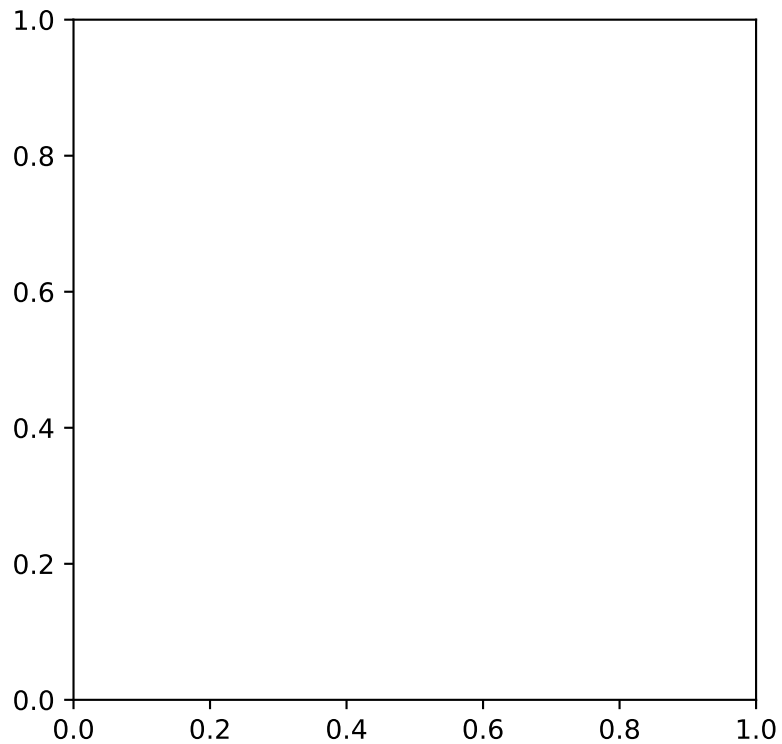
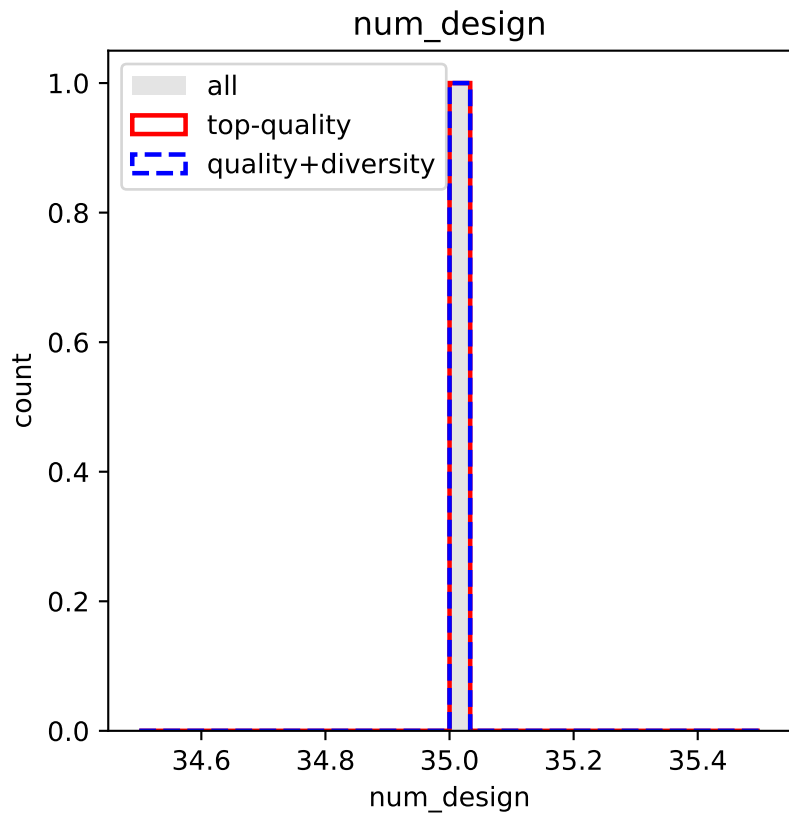
All samples

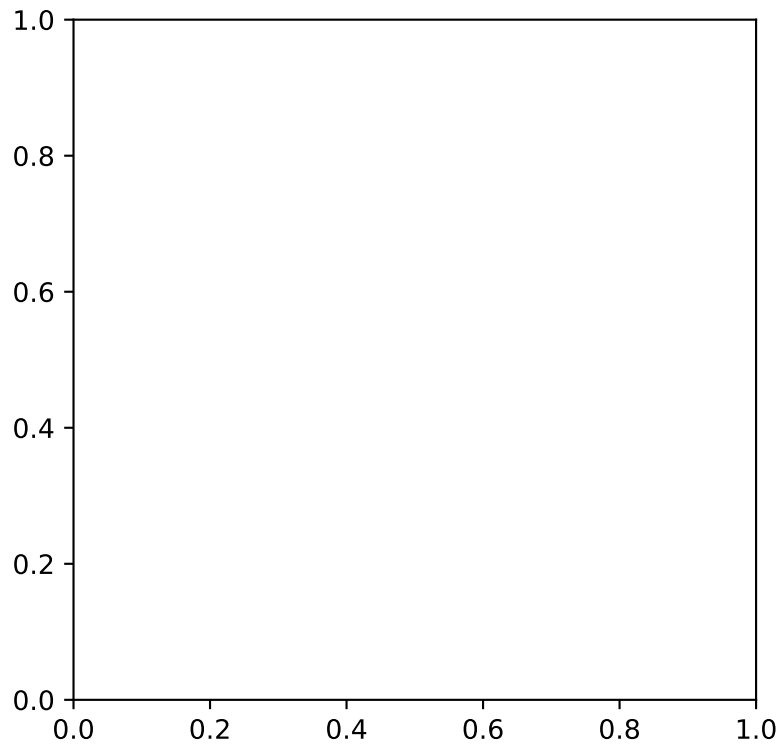
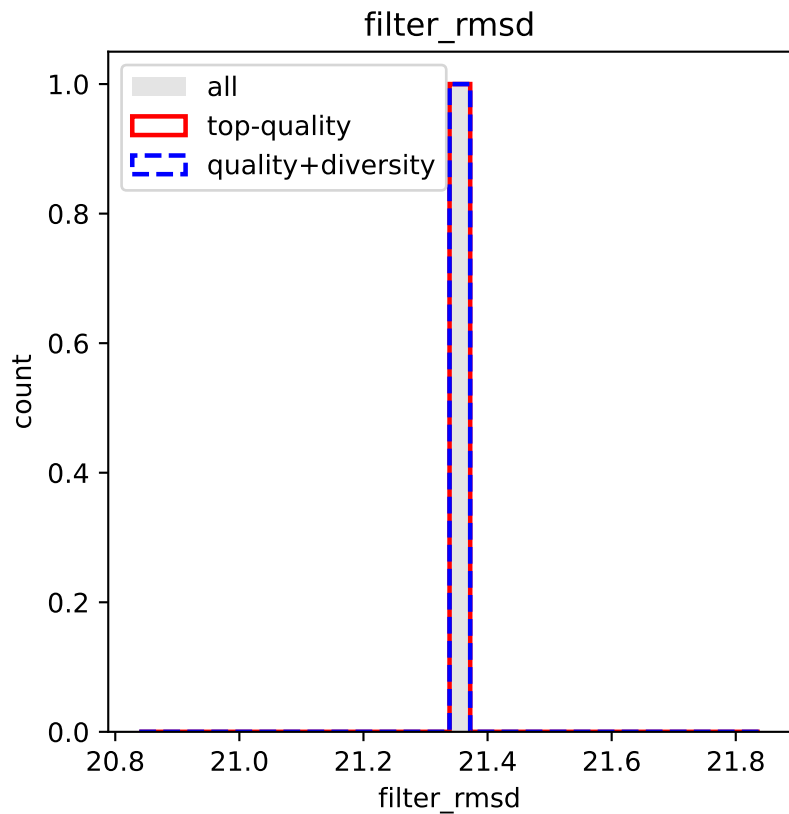


# **Metric Distributions**

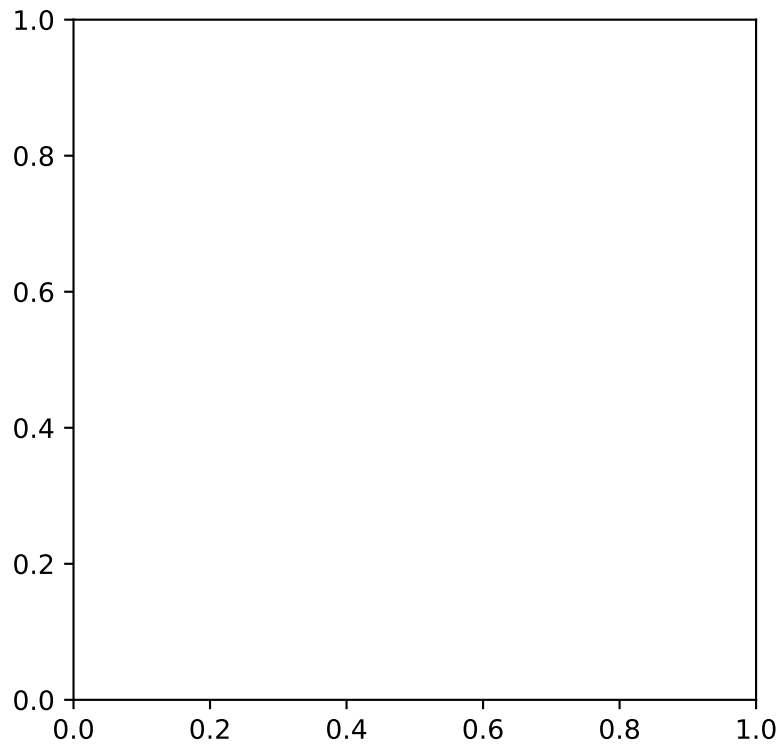
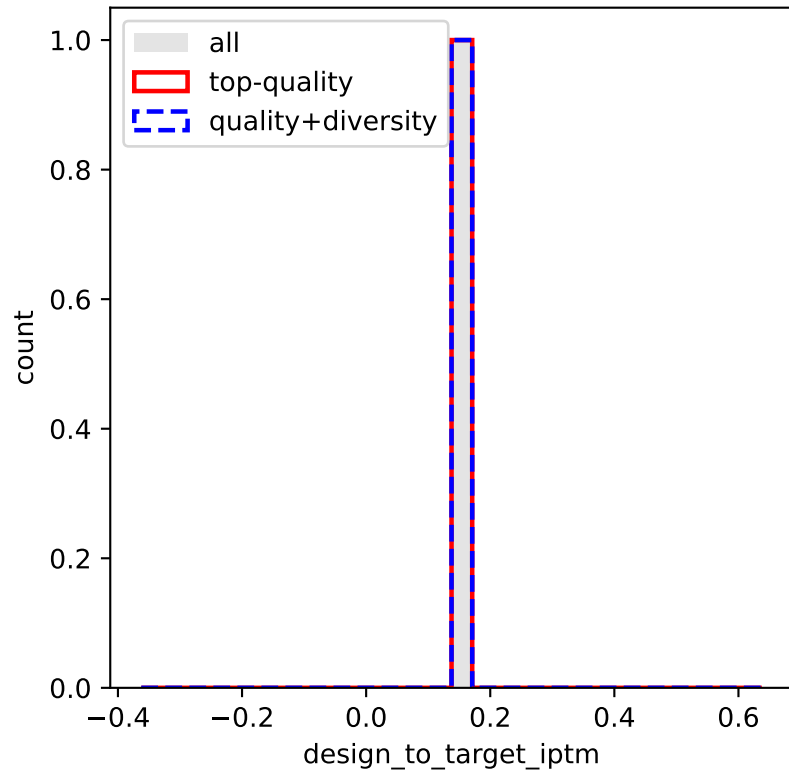
## *Histograms*

Distribution of each metric across all designs (grey) with overlays for Top-quality (red outline) and Diversity-optimised (blue dashed) subsets. The right panel repeats the histogram but only for designs that pass the RMSD threshold.

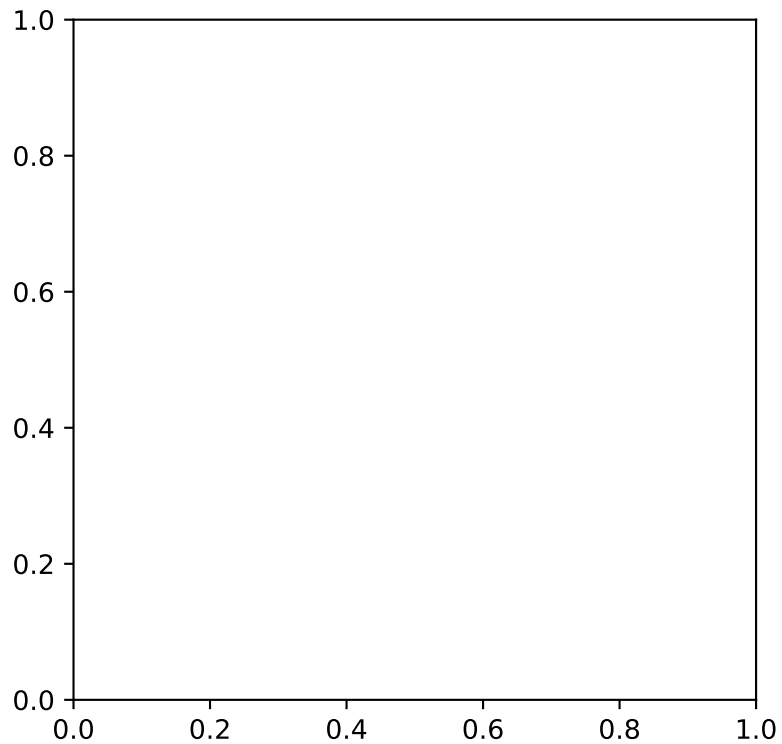
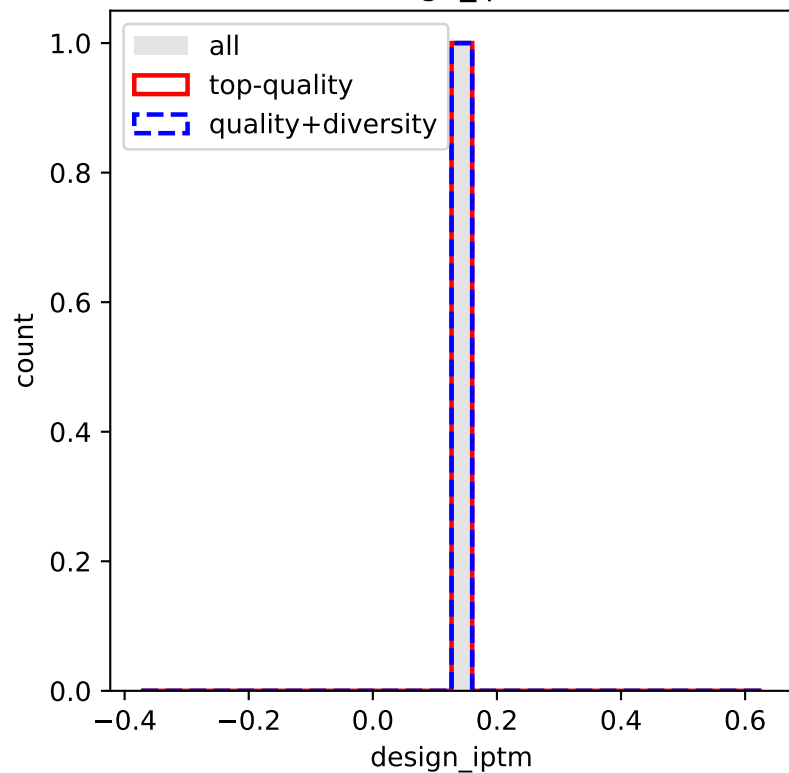




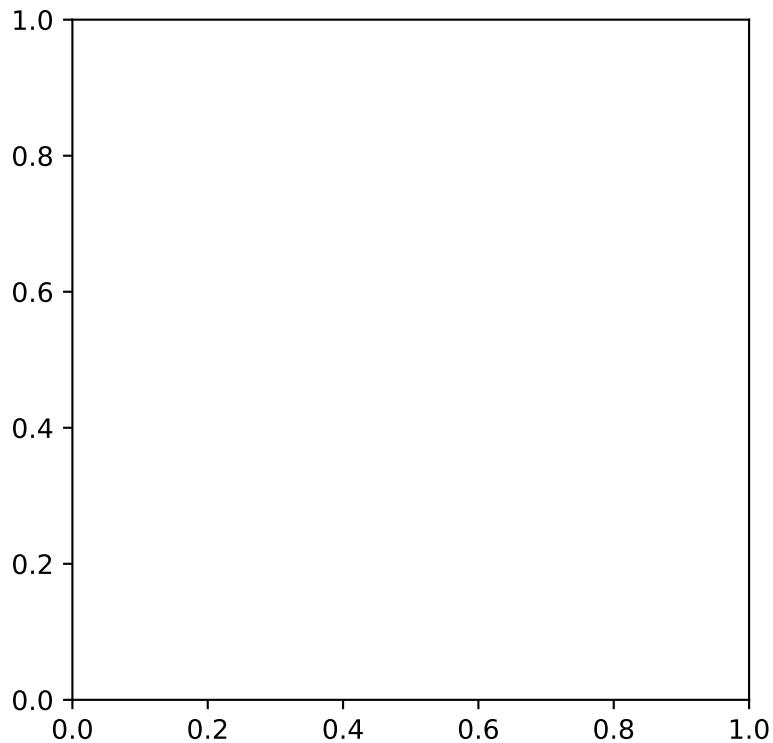
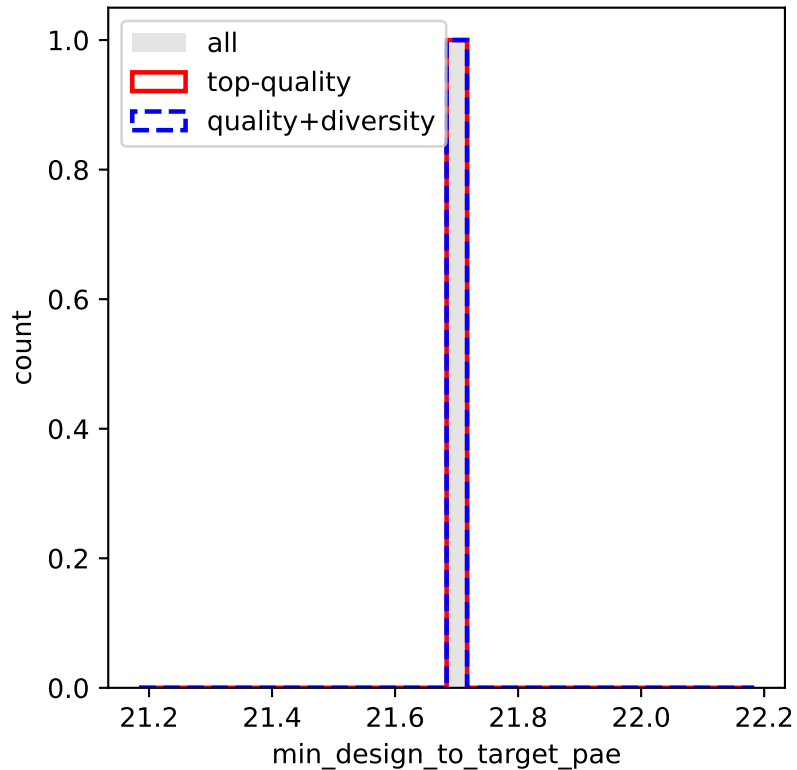
design\_to\_target\_ip\_tm



design\_iptm

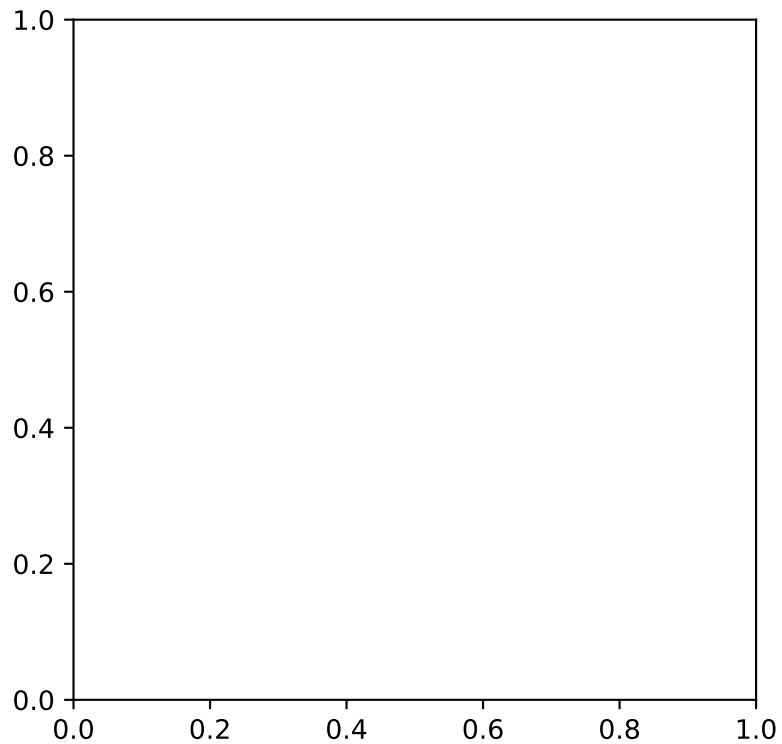
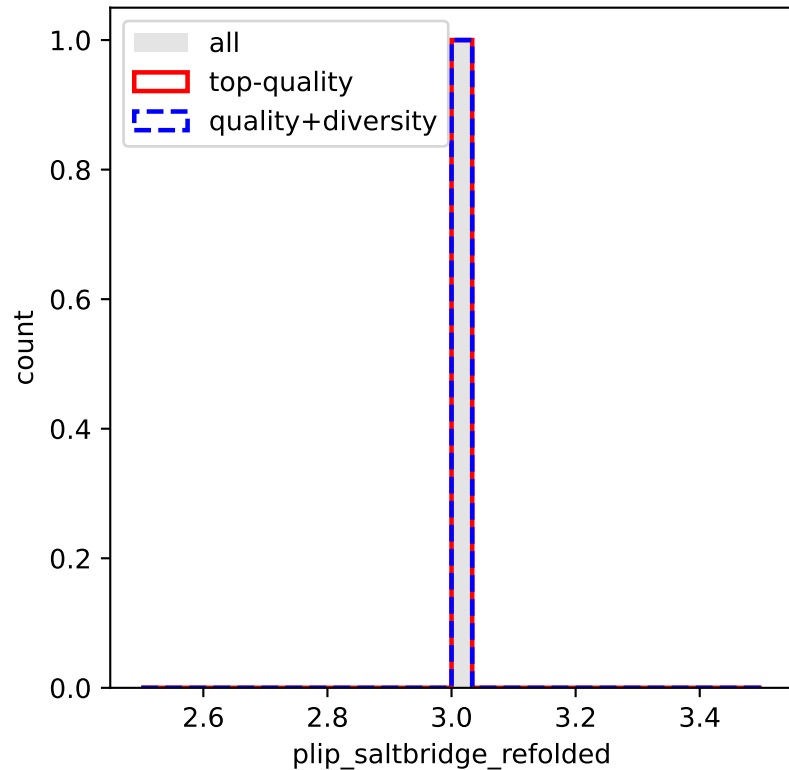


min\_design\_to\_target\_pae

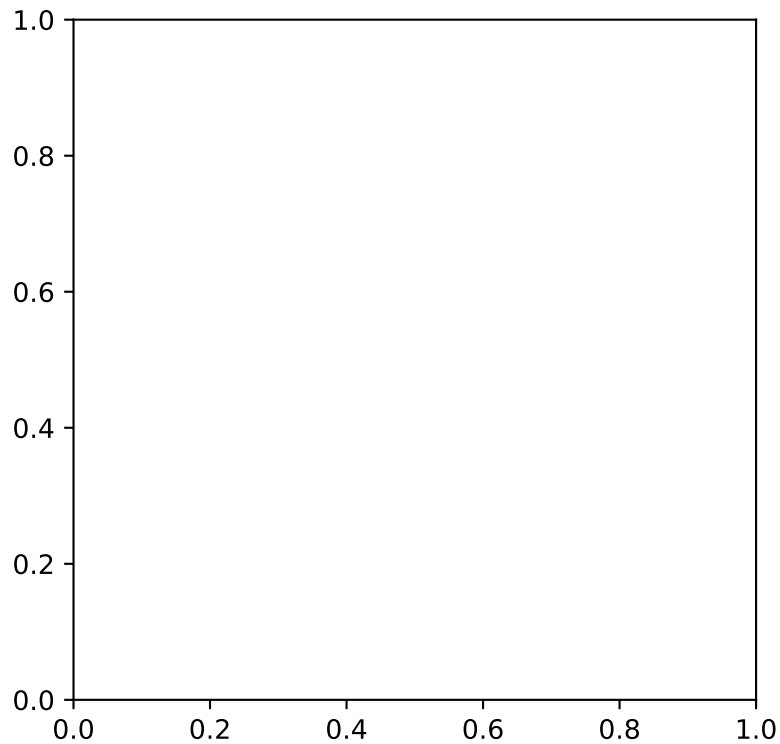
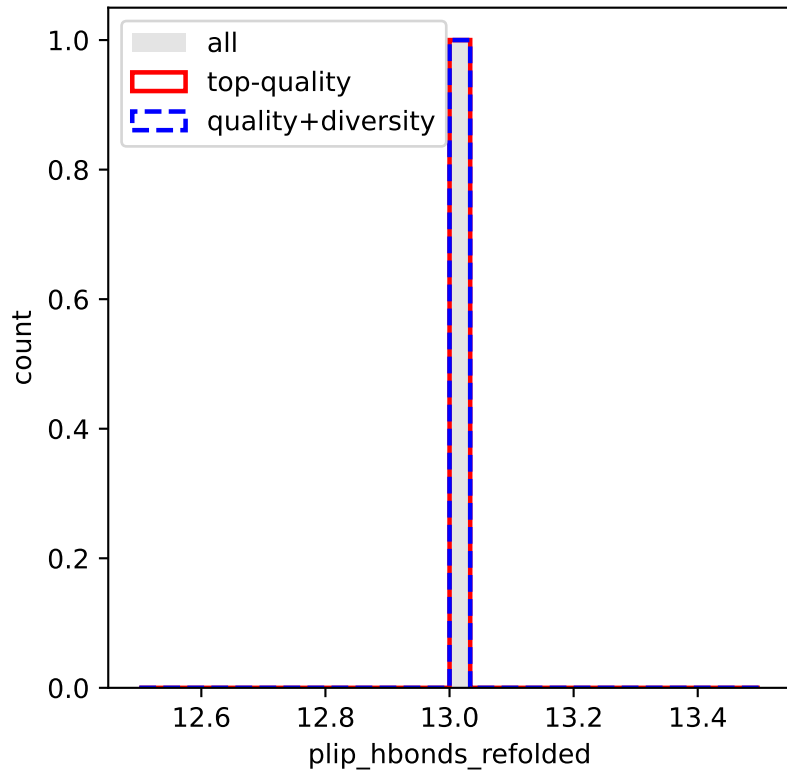




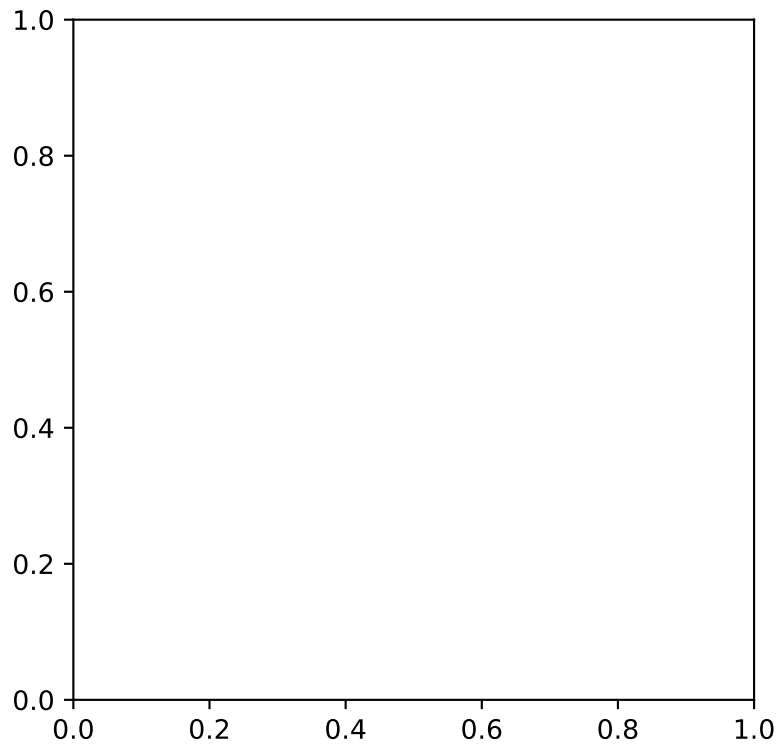
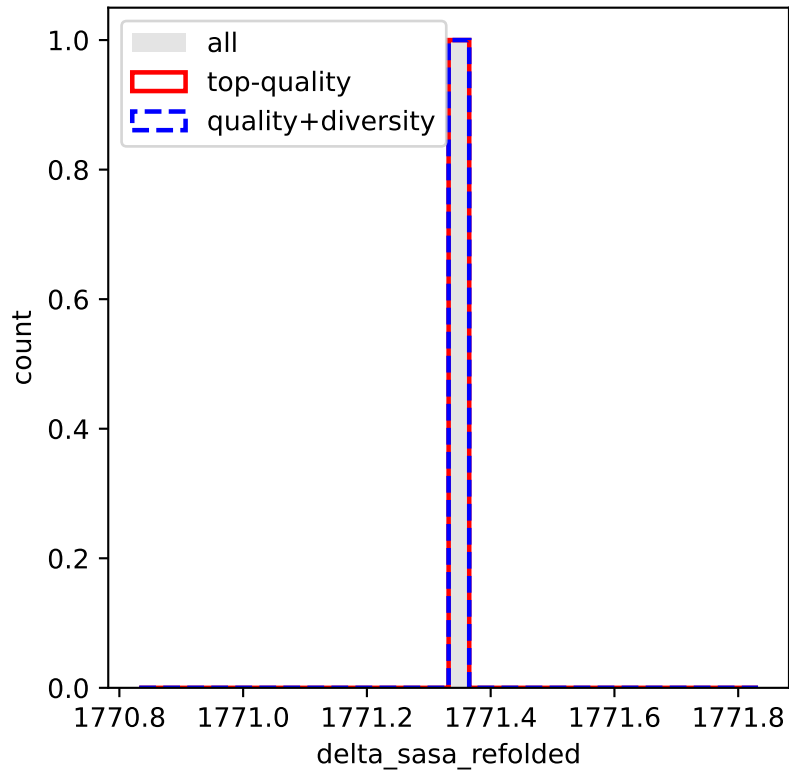
plip\_saltbridge\_refolded



plip\_hbonds\_refolded



delta\_sasa\_refolded



design\_hydrophobicity

