

Results Overview

- **Designs generated: 10**

- Designs passing all filters: 1

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

/home/ubuntu/out/final_ranked_designs/final_2_designs

- **Metrics and sequences of 2 final designs are in:**

/home/ubuntu/out/final_ranked_designs/final_designs_metrics_2.csv

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

/home/ubuntu/out/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 2 designs that jointly maximise quality and minimise sequence similarity (sequence-identity distance). The trade-off is controlled by $\alpha = 0.001$:
 - $\alpha = 0 \rightarrow 100\% \text{ quality focus}$ (same as Top set)
 - $\alpha = 1 \rightarrow 100\% \text{ diversity focus}$ (ignores quality)
 - Quality – composite of metrics such as iPTM, salt-bridges, Δ 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_iptm	predicted TM score for design-target contacts (higher = better)
design_to_target_iptm	same as design_iptm 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	Δ 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	10
filter_rmsd	True	2.5	1
filter_rmsd_design	True	2.5	3
CYS_fraction	True	0.0	10
ALA_fraction	True	0.2	10
GLY_fraction	True	0.2	5
GLU_fraction	True	0.2	10
LEU_fraction	True	0.2	10
VAL_fraction	True	0.2	9

Sorting Criteria

Metric	Inverse Importance
design_iiptm	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 2 quality+diversity designs produced by the lazy-greedy selection.

Metric	Mean	Mean top 10	Mean top 2 diverse
Num designs	10	-	-
num_design	114.200	114.200	118.500
filter_rmsd	9.885	9.885	3.363
design_iptm	0.237	0.237	0.350
design_to_target_iptm	0.237	0.237	0.350
min_design_to_target_pae	16.084	16.084	10.867
delta_sasa_refolded	1119.714	1119.714	1254.795
plip_saltbridge_refolded	1.500	1.500	1.000
plip_hbonds_refolded	4.300	4.300	6.500

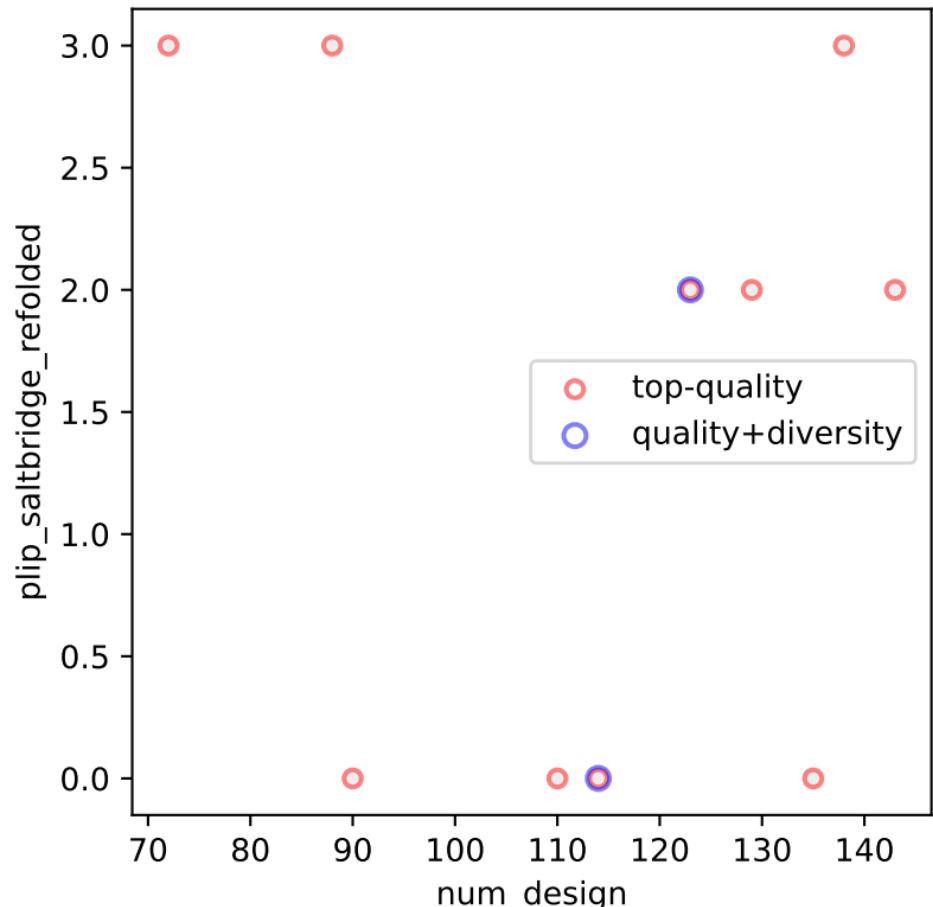
Scatter Plots

Metric relationships

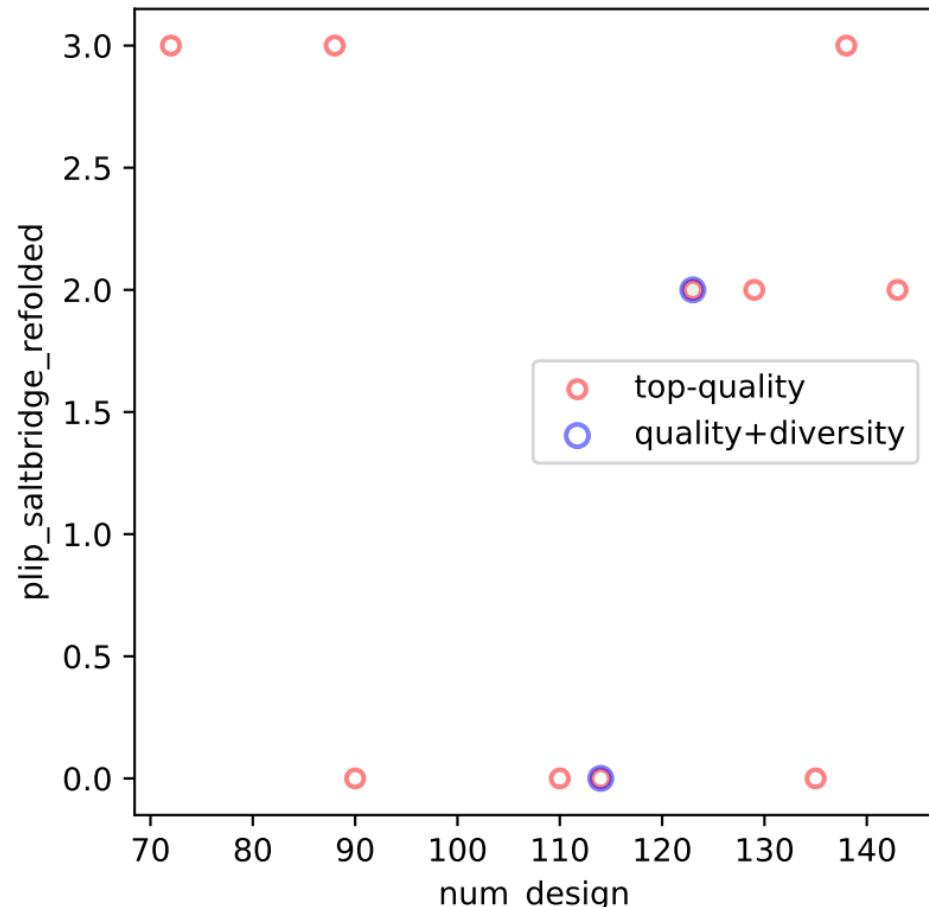
Each scatter page contains two panels:

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

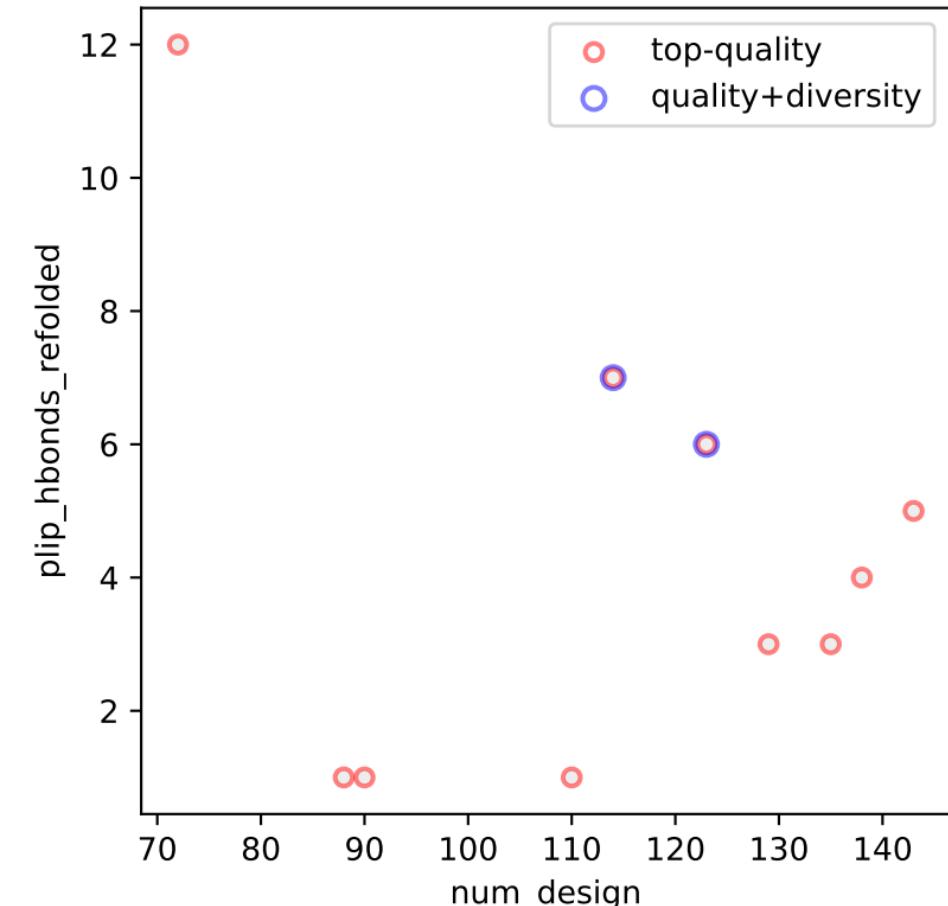
All samples



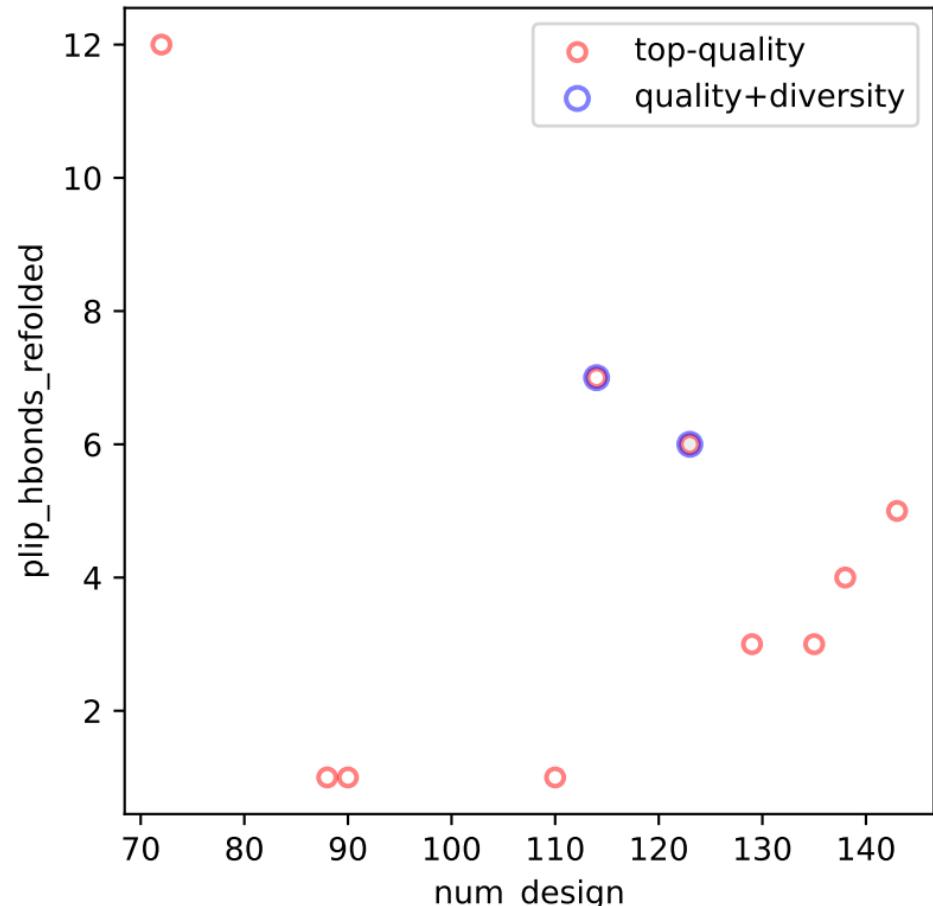
(Designs passing RMSD threshold)



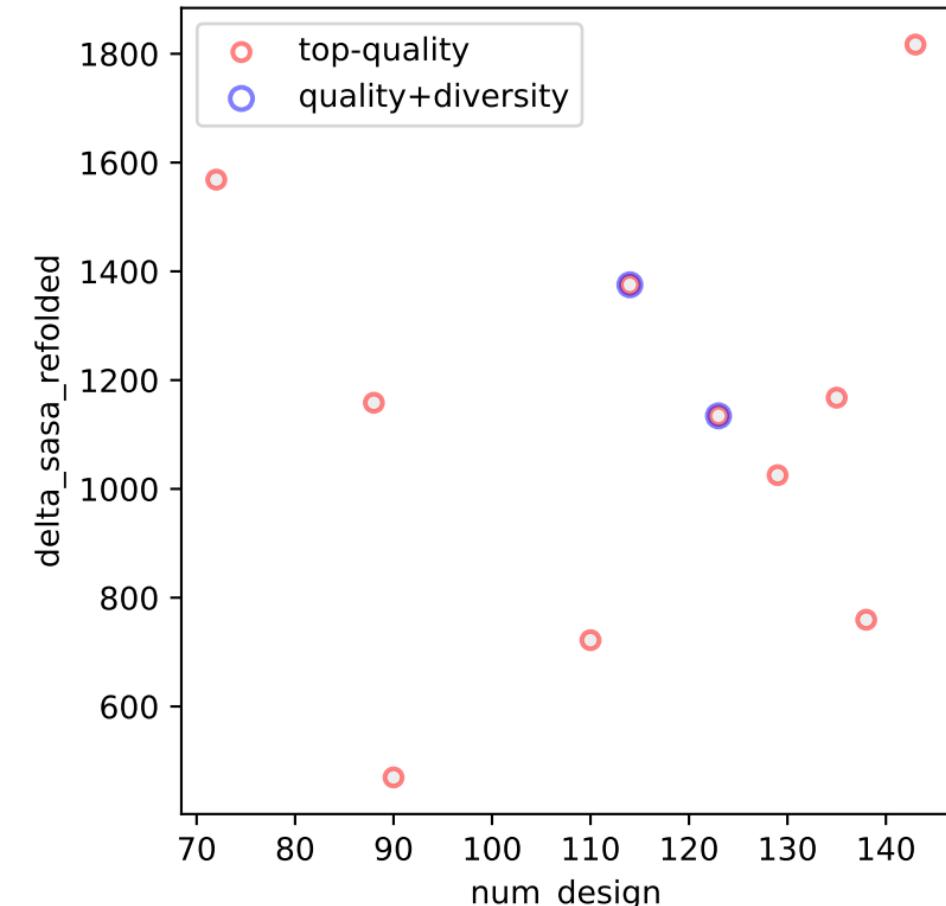
All samples



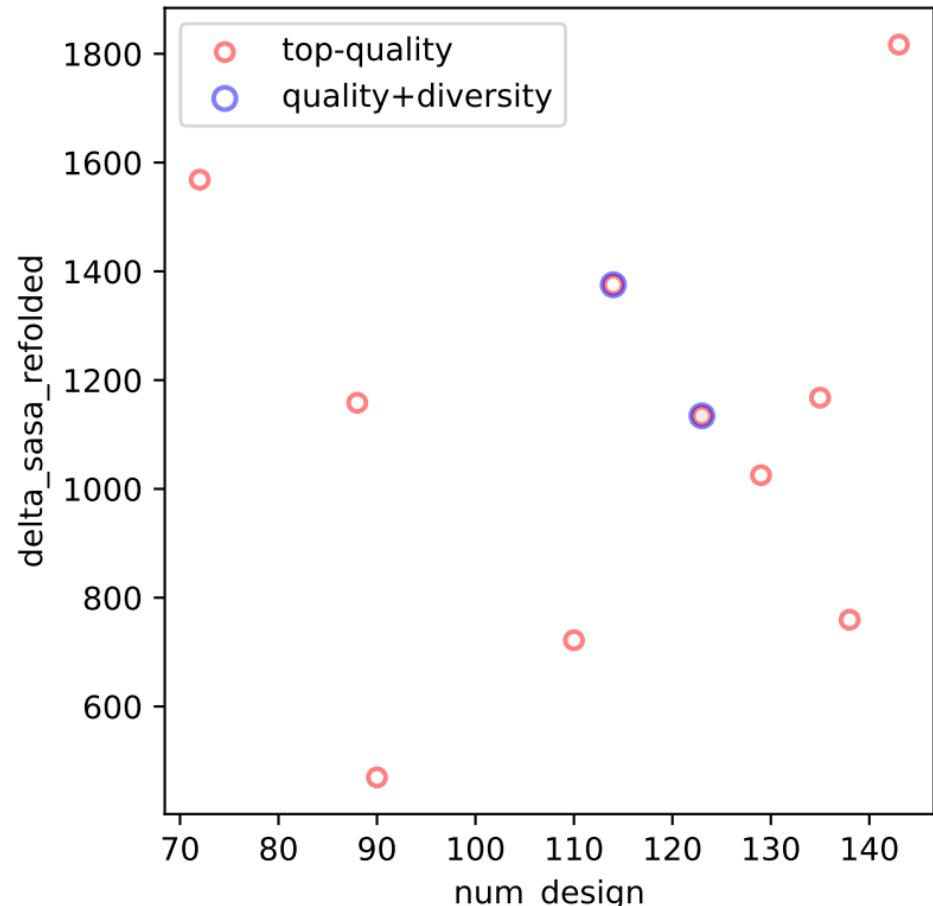
(Designs passing RMSD threshold)



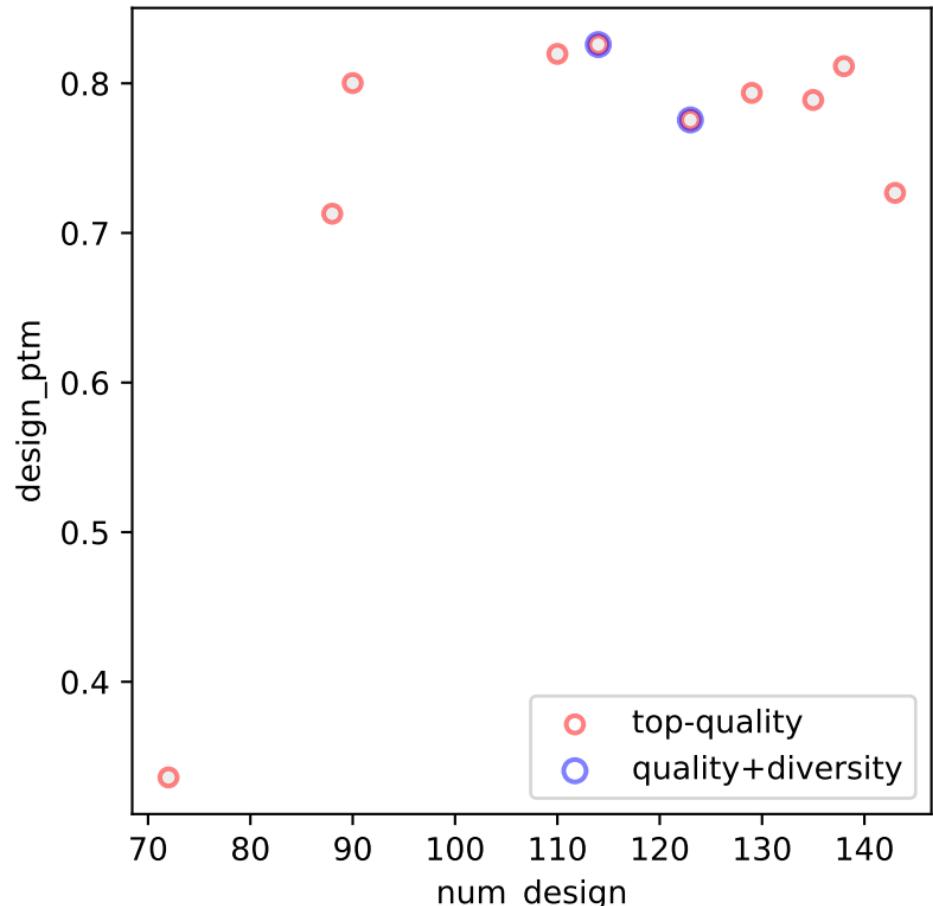
All samples



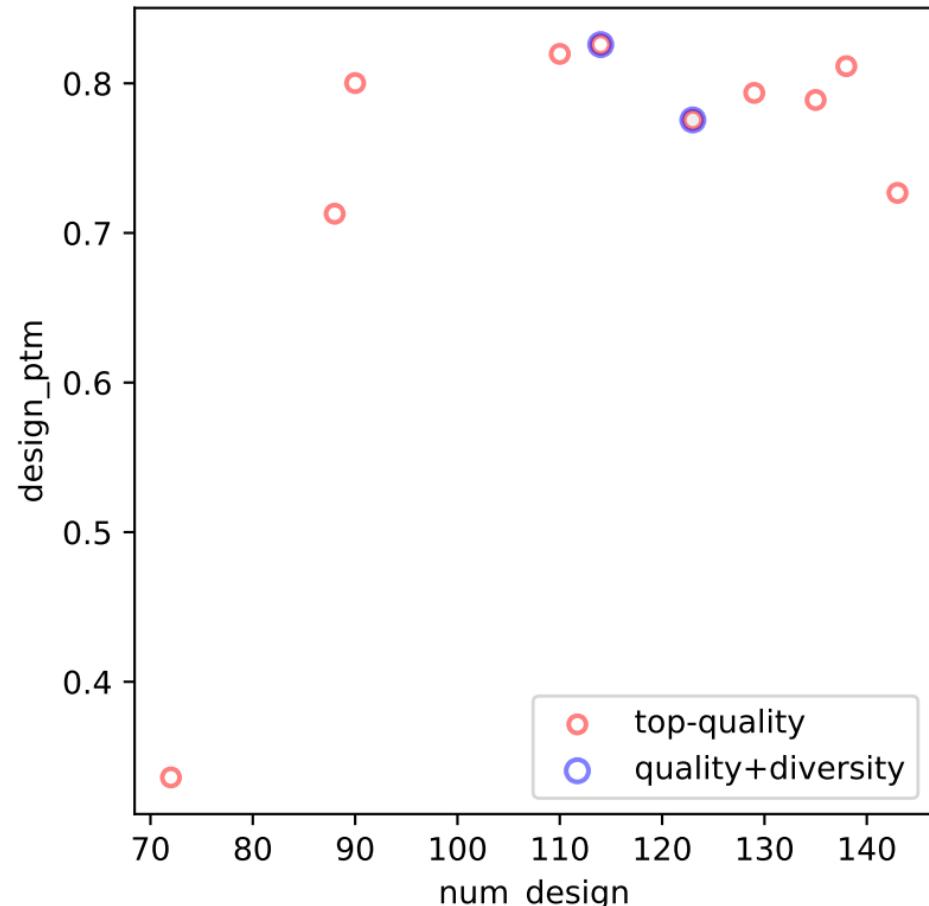
(Designs passing RMSD threshold)



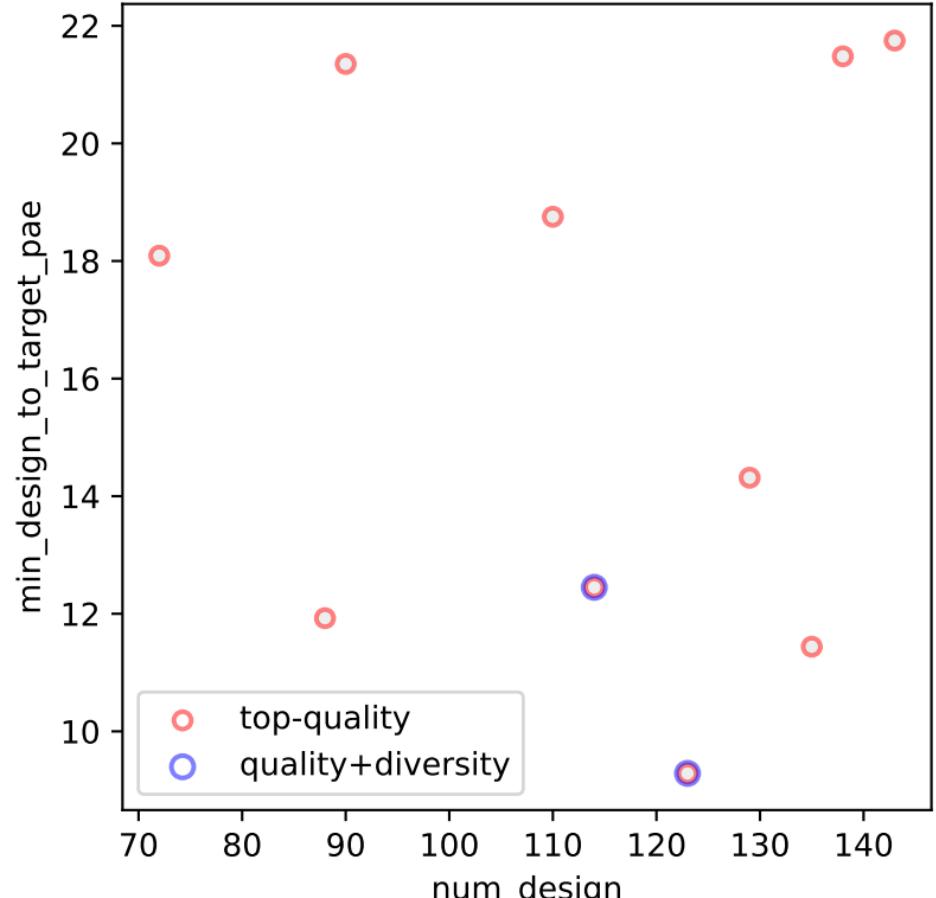
All samples



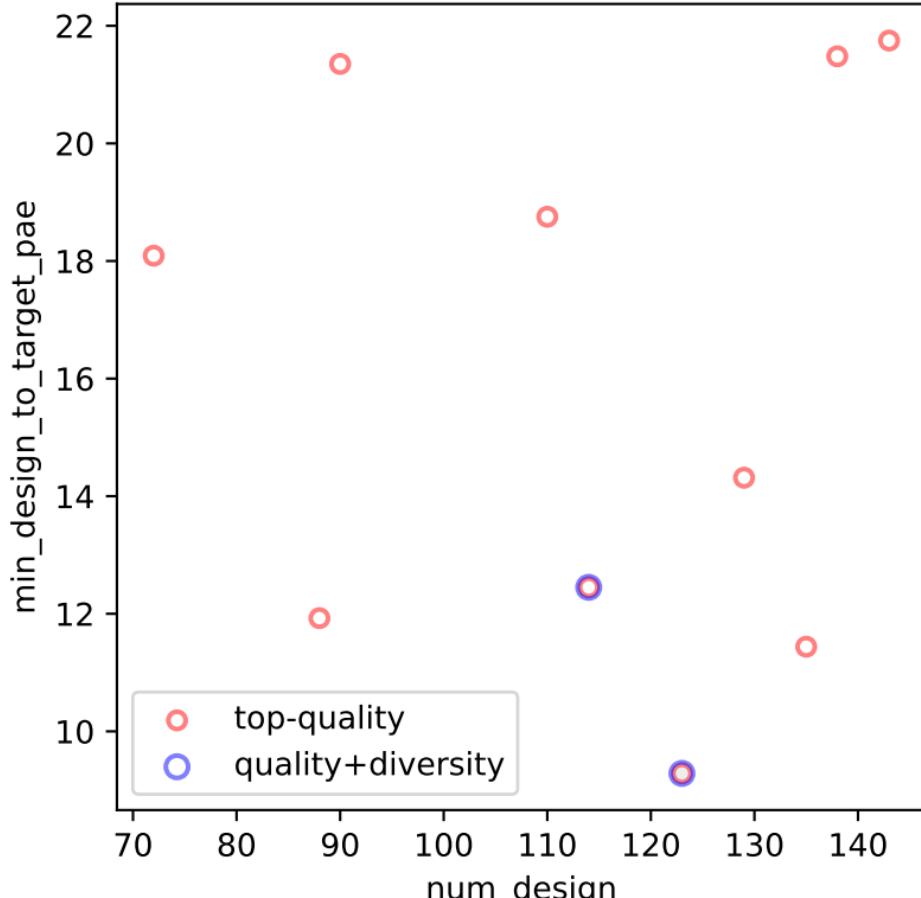
(Designs passing RMSD threshold)



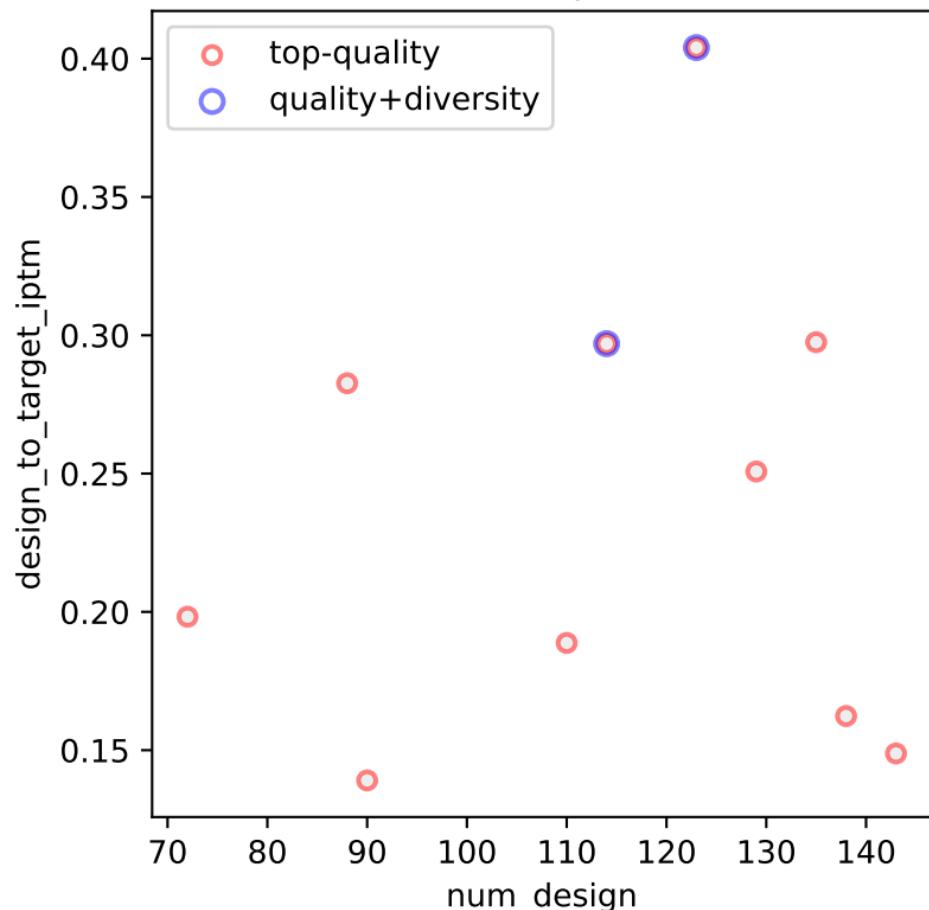
All samples



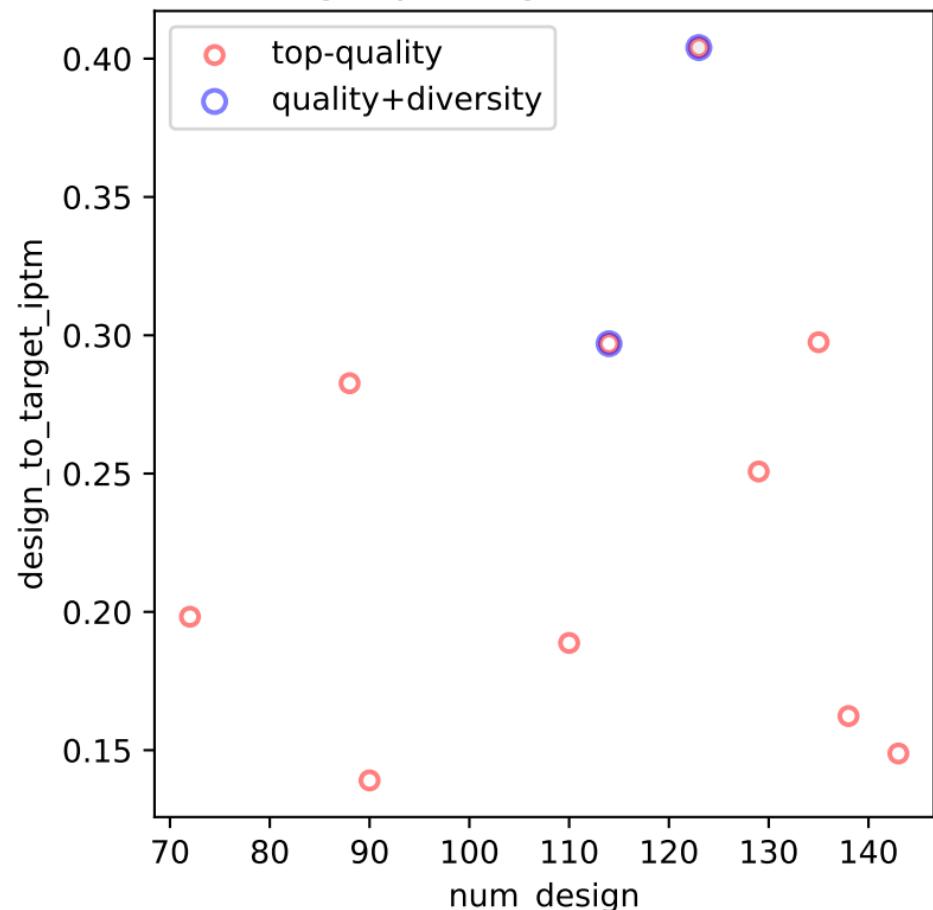
(Designs passing RMSD threshold)



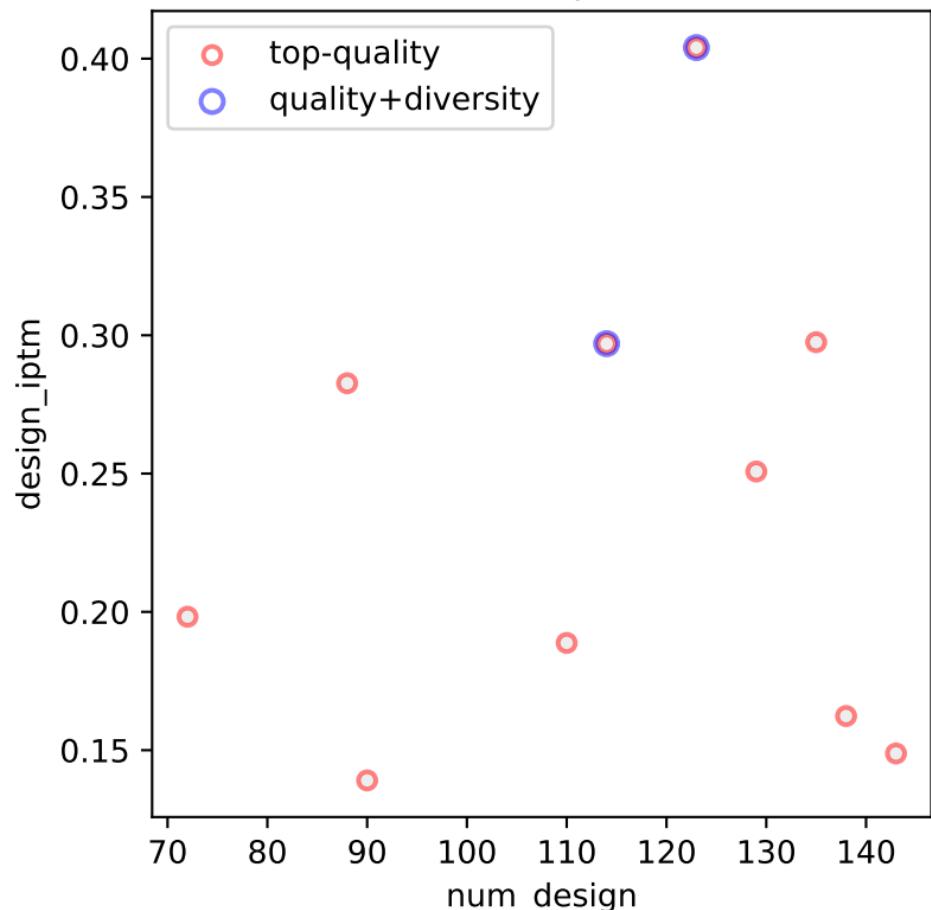
All samples



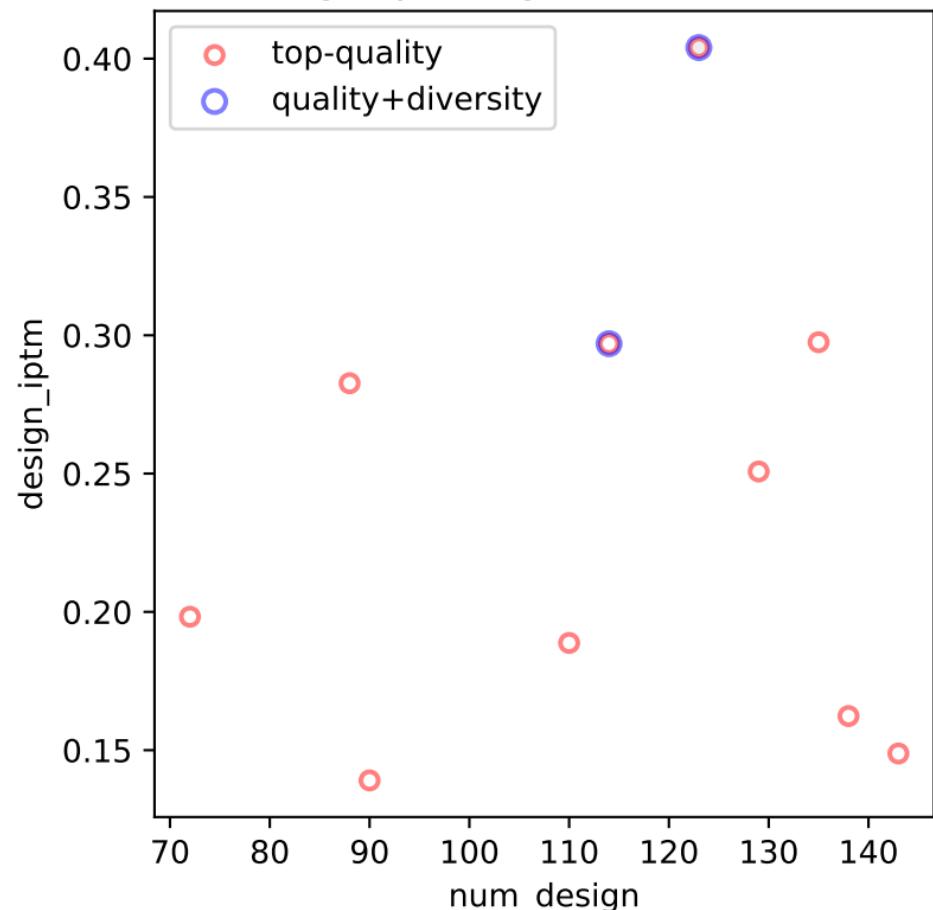
(Designs passing RMSD threshold)



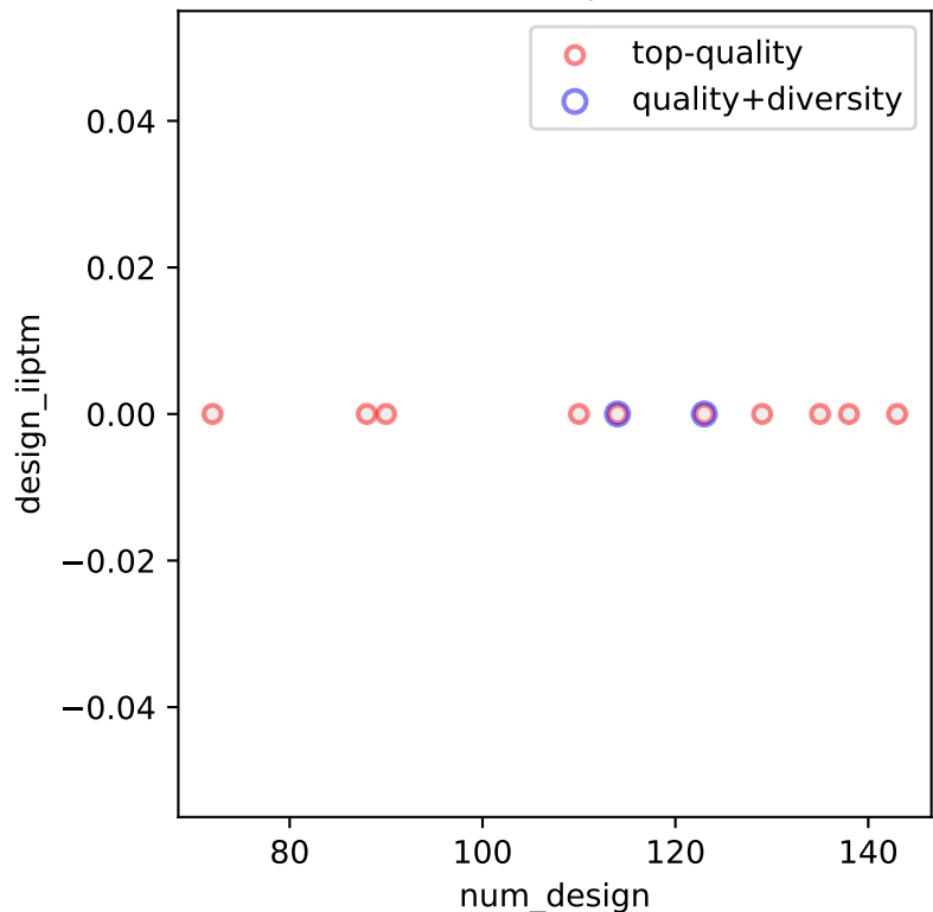
All samples



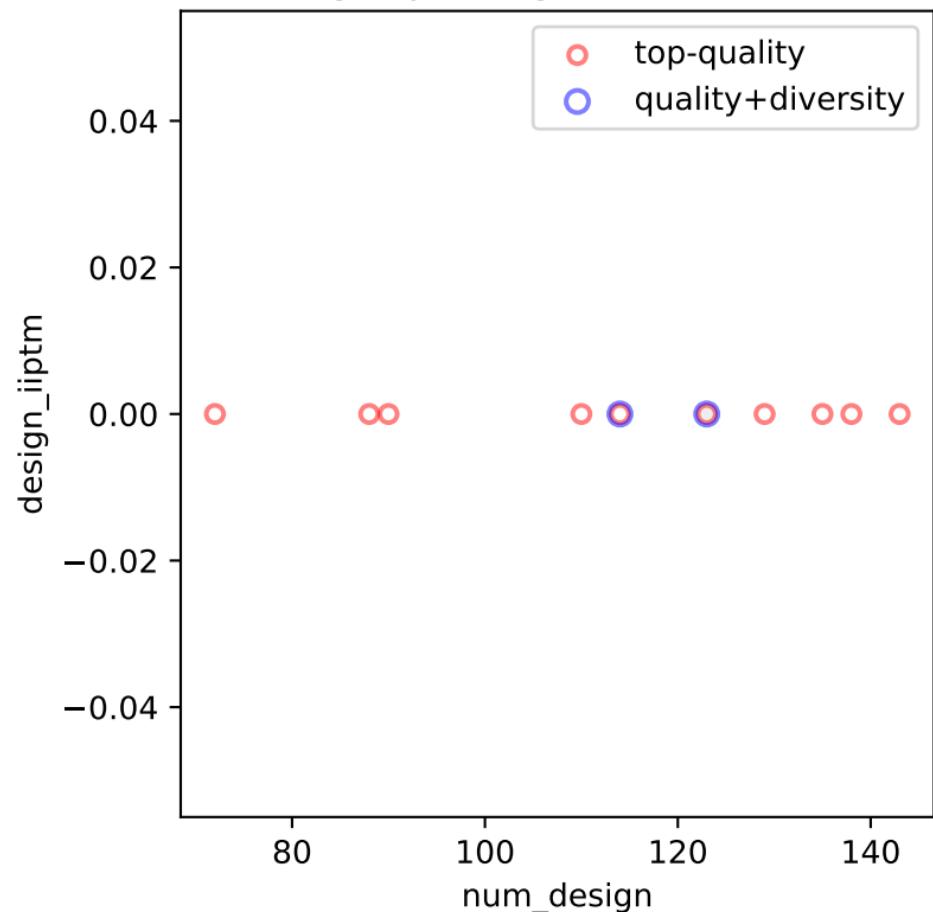
(Designs passing RMSD threshold)



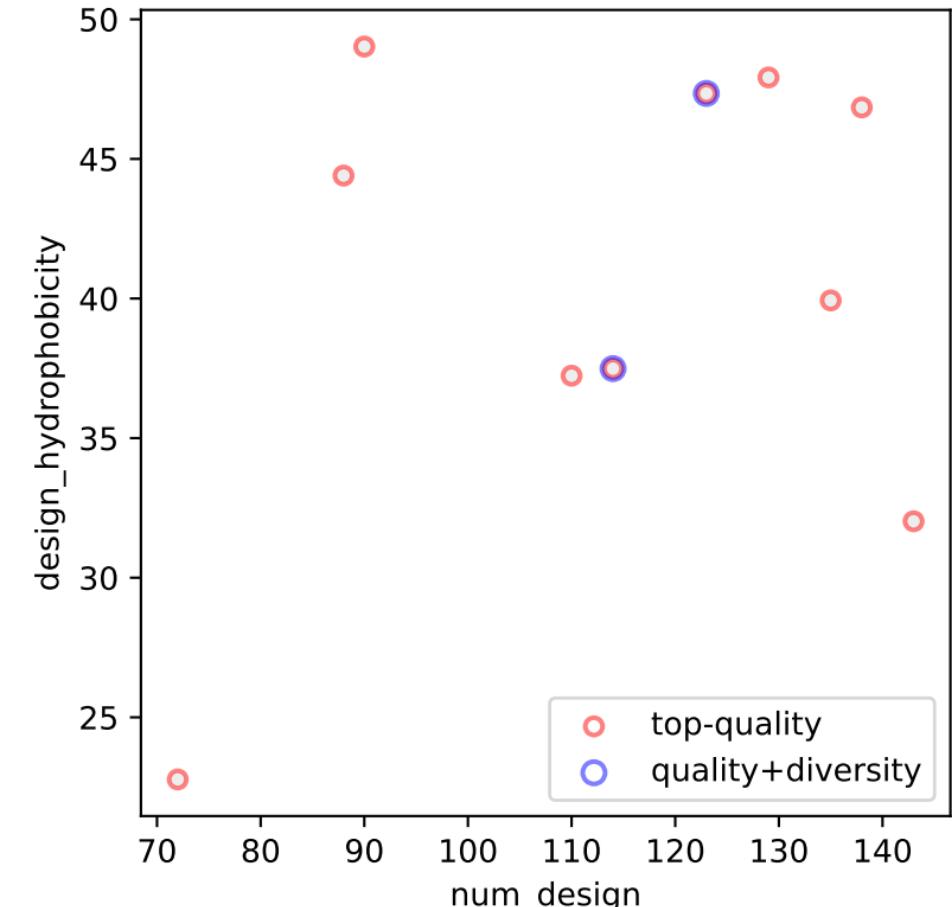
All samples



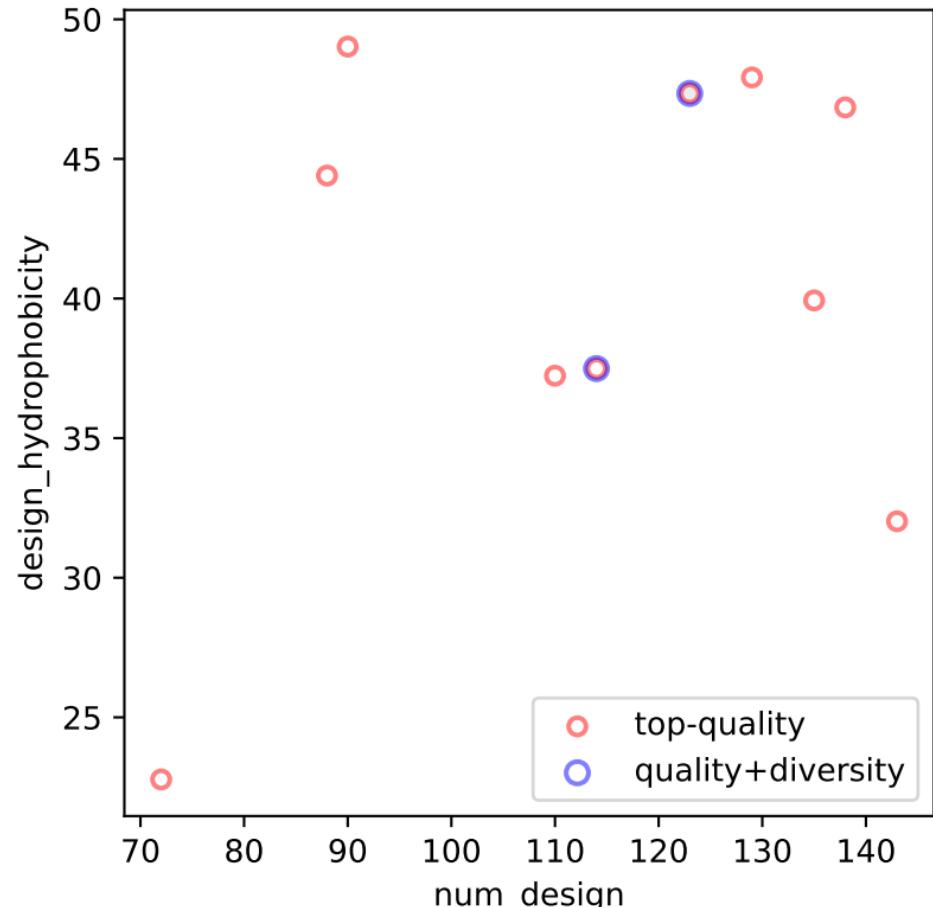
(Designs passing RMSD threshold)



All samples



(Designs passing RMSD threshold)

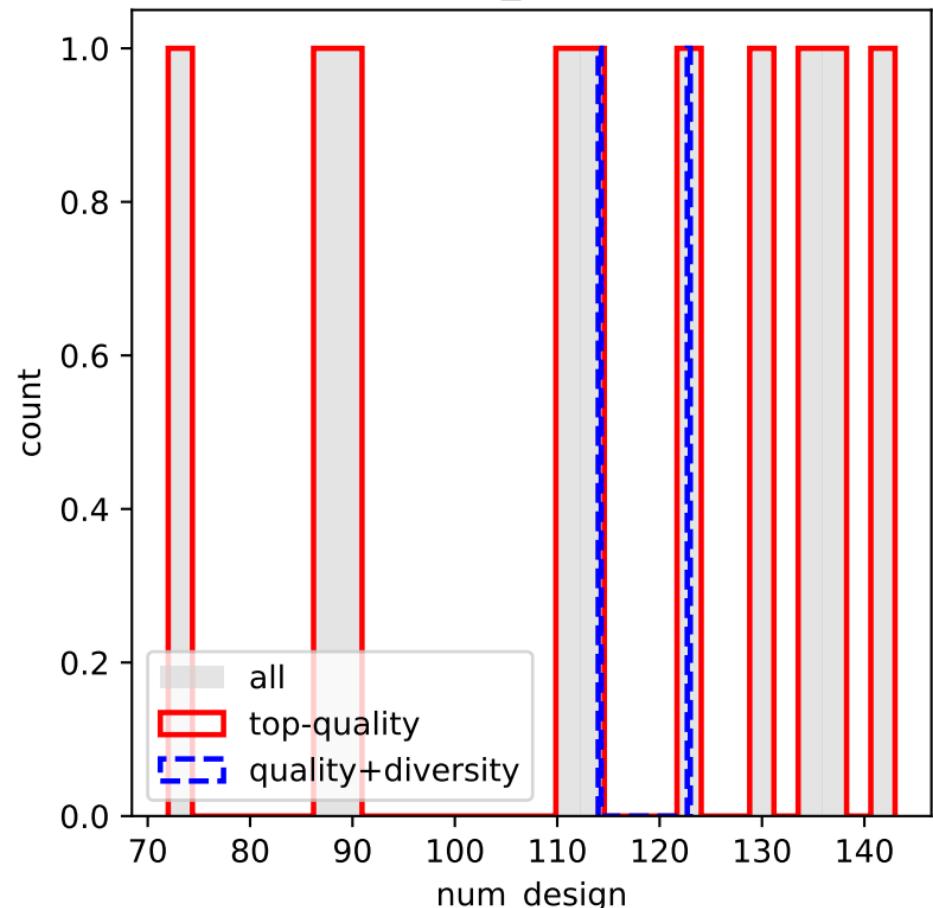


Metric Distributions

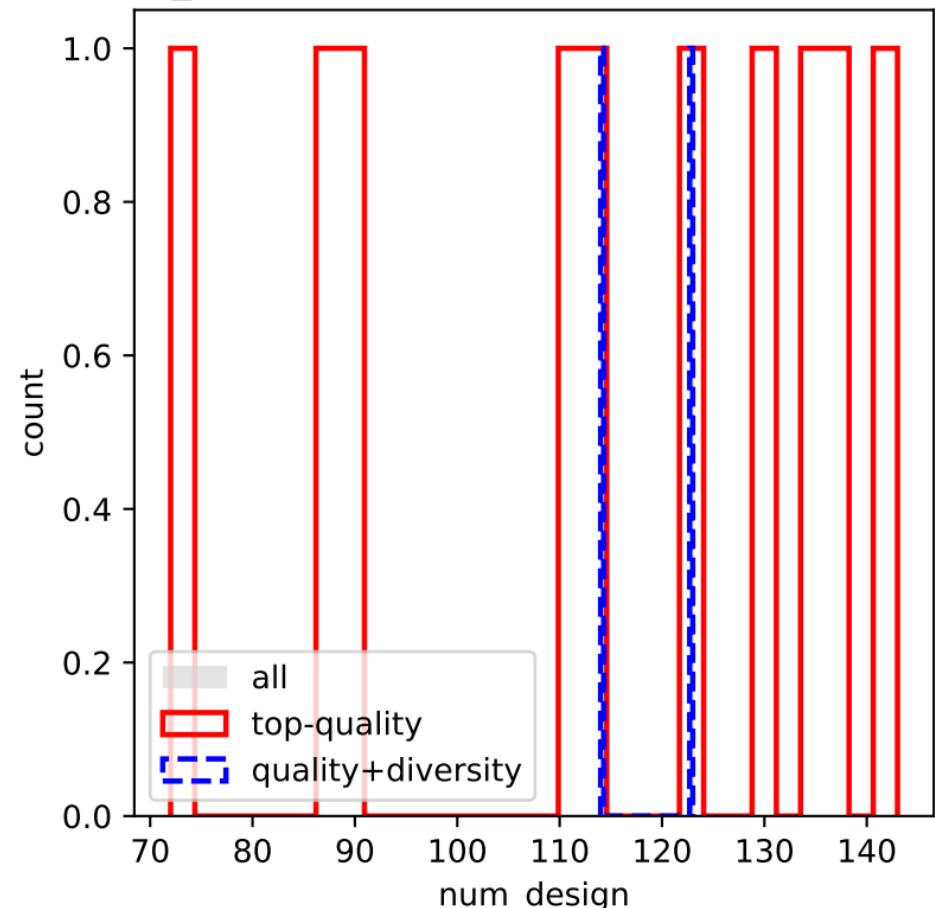
Histograms

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.

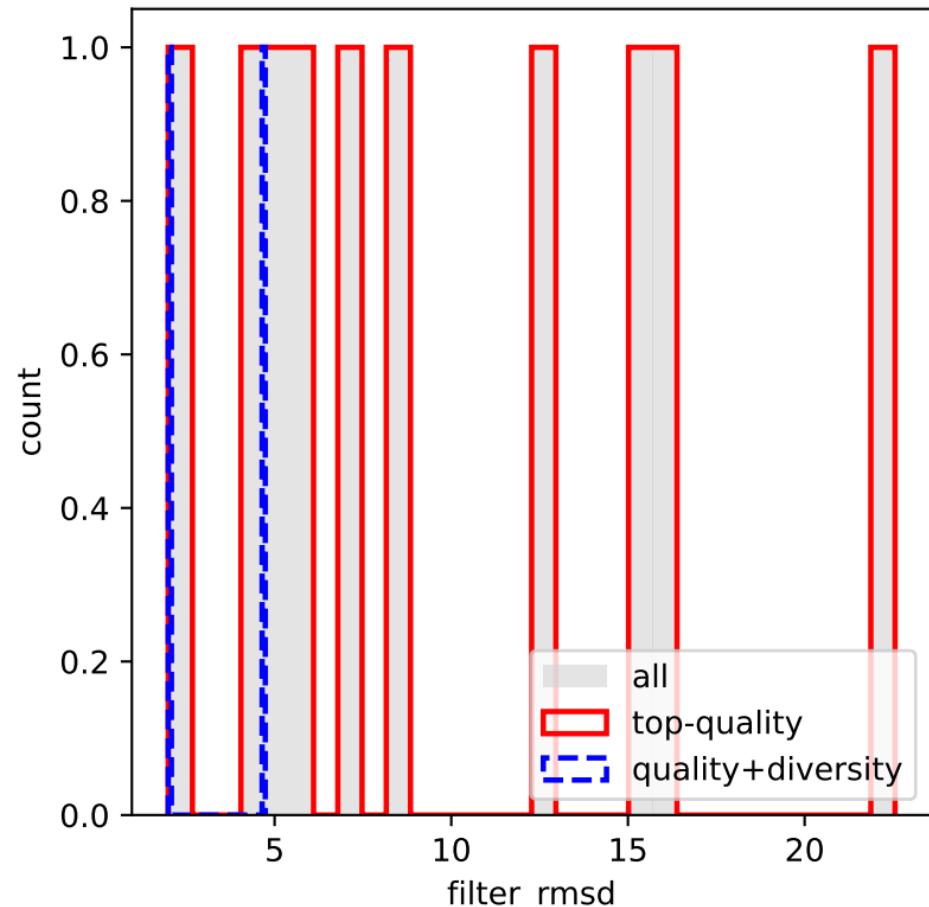
num_design



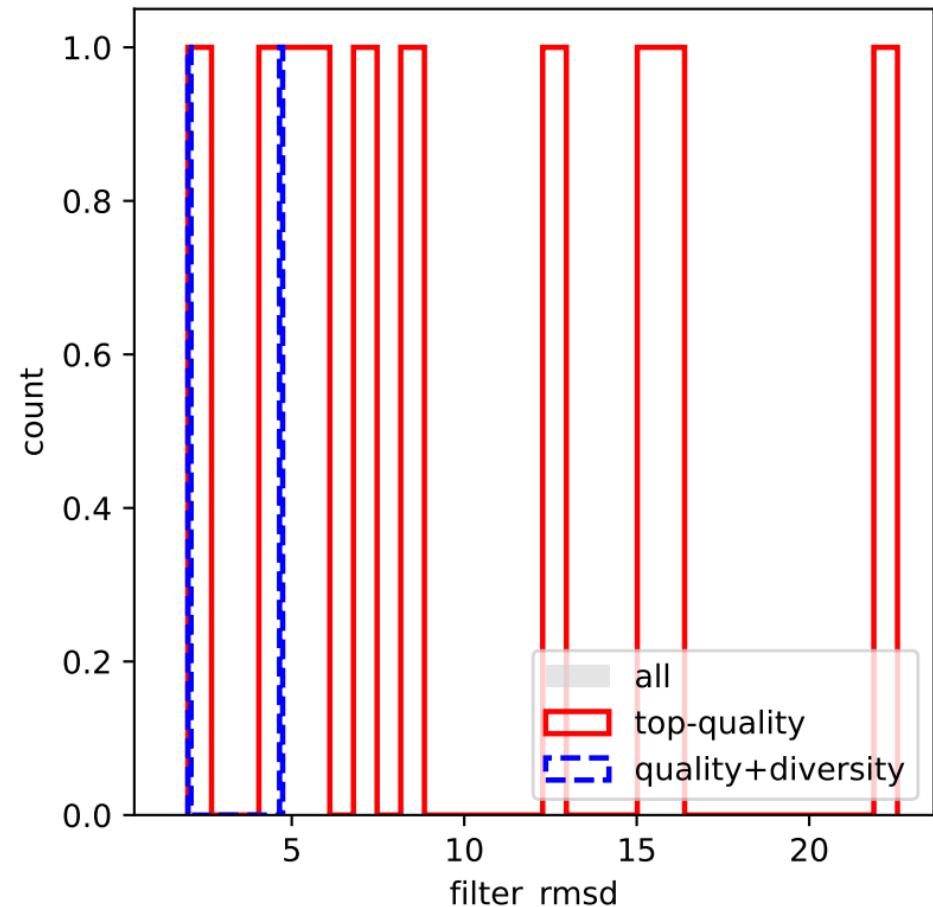
num_design (Designs passing RMSD threshold)



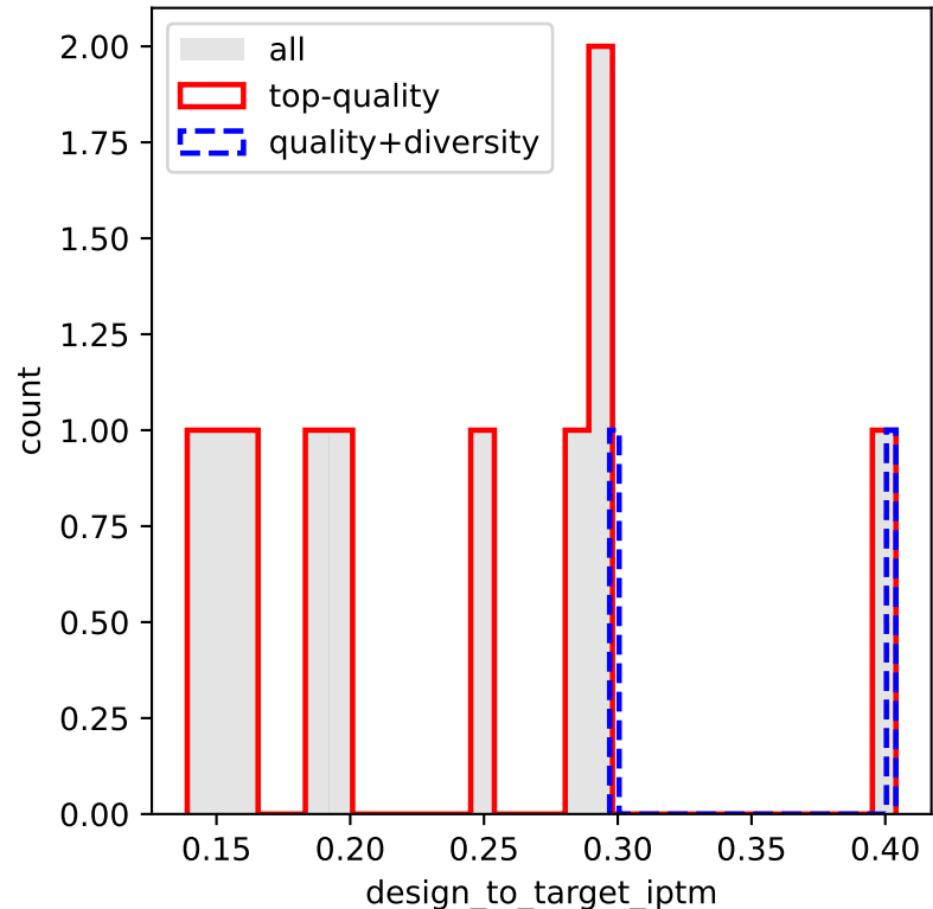
filter_rmsd



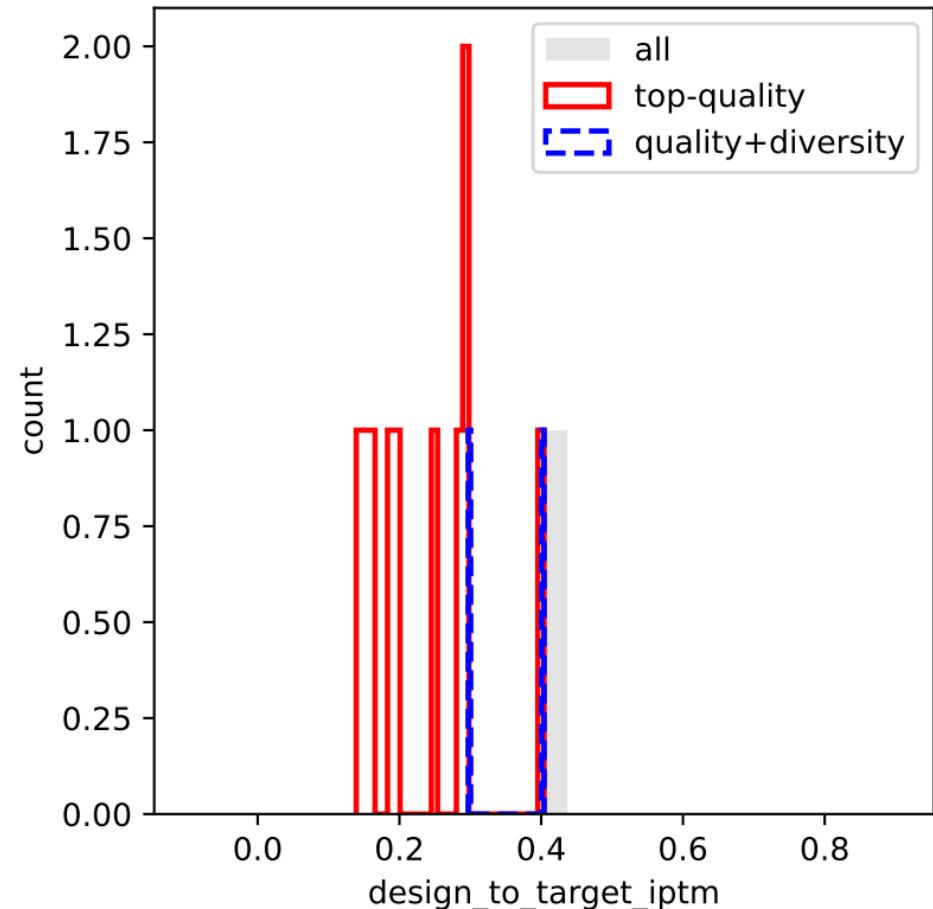
filter_rmsd (Designs passing RMSD threshold)



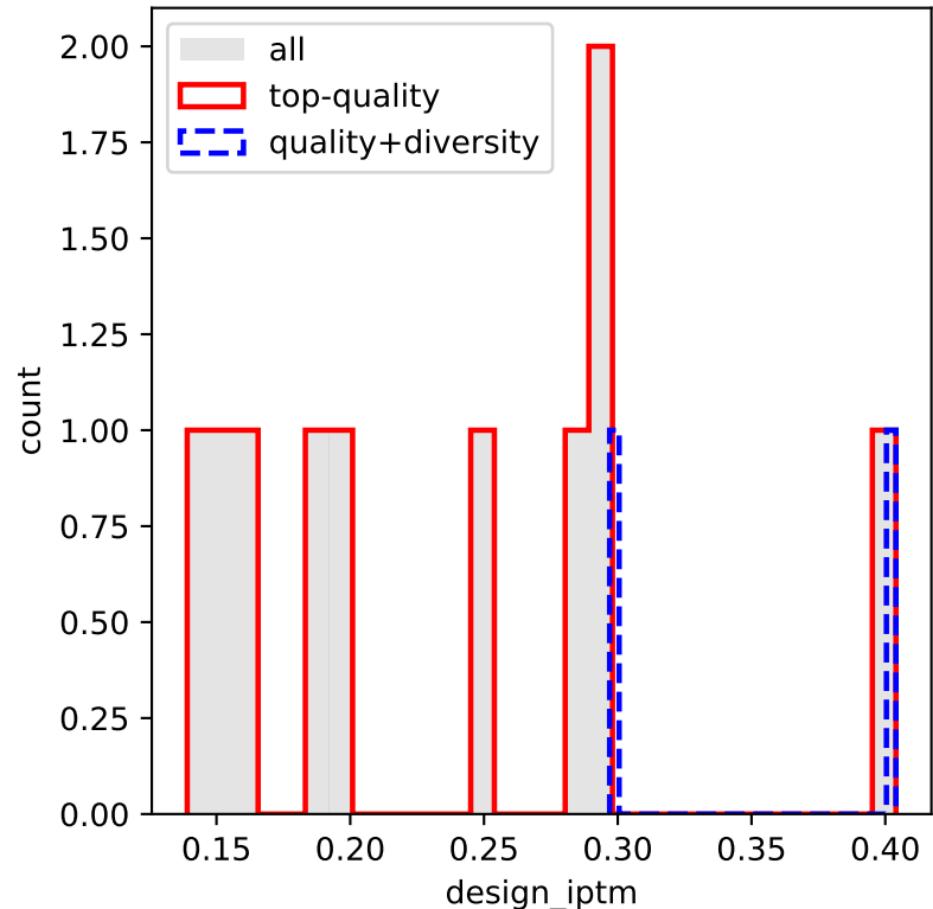
design_to_target_iptm



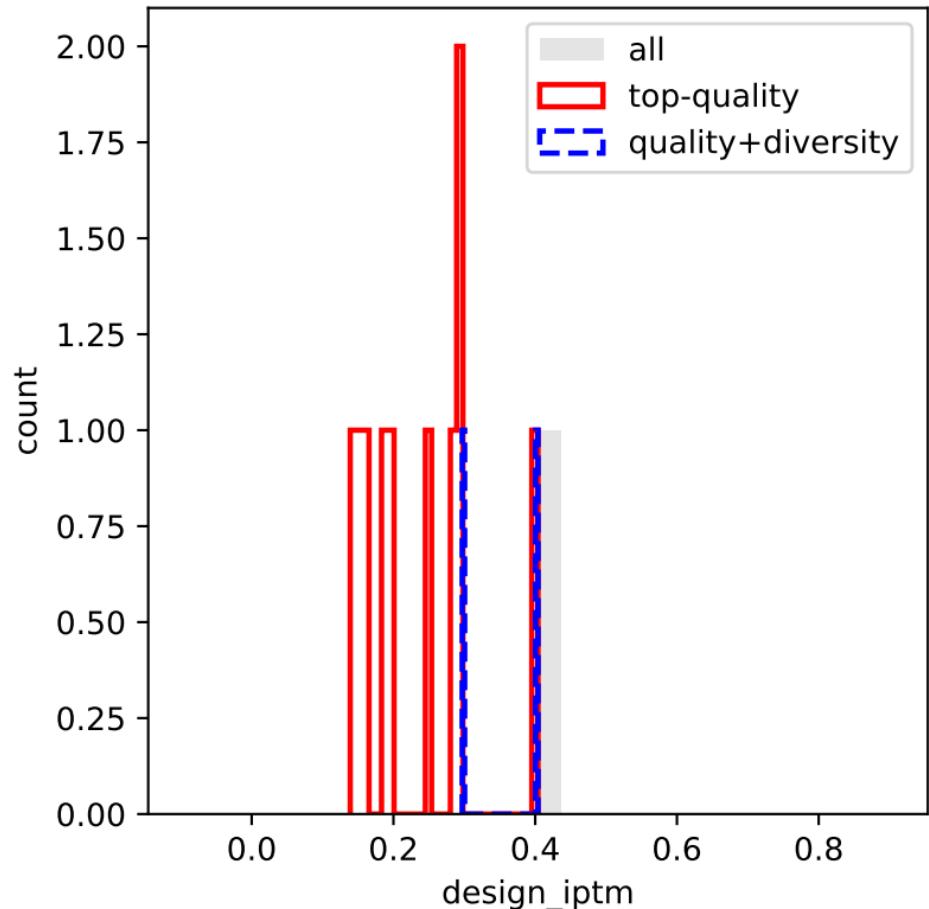
design_to_target_iptm (Designs passing RMSD thre



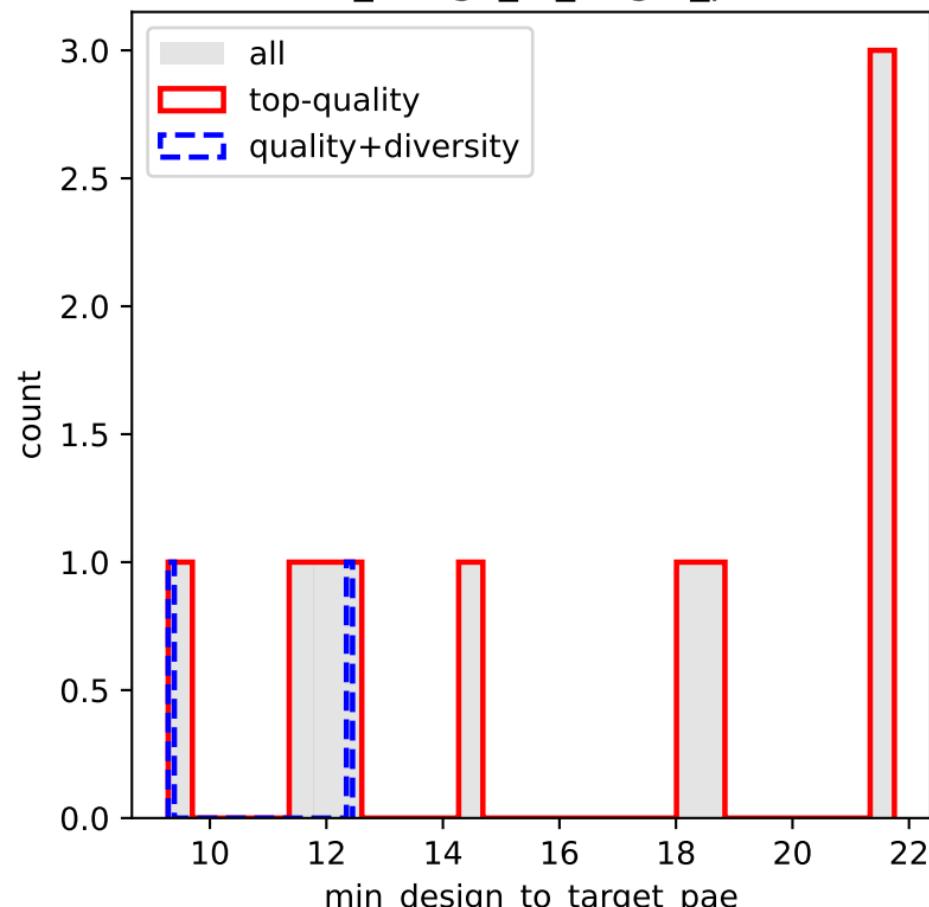
design_iptm



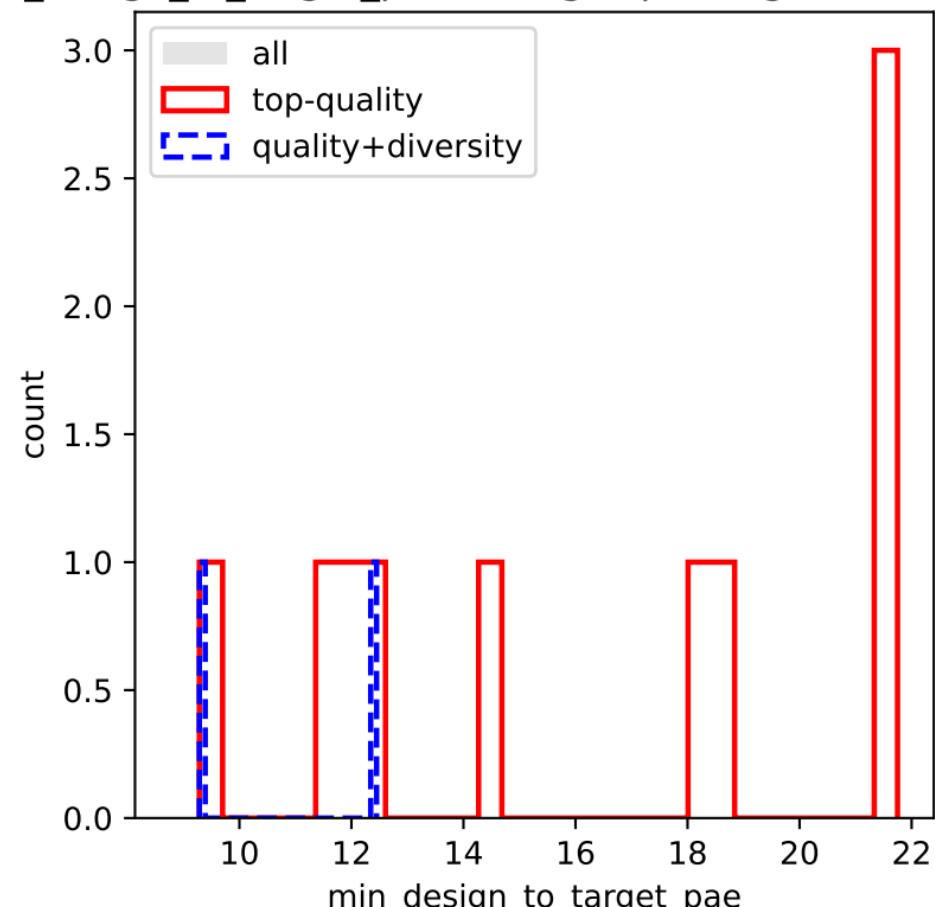
design_iptm (Designs passing RMSD threshold)



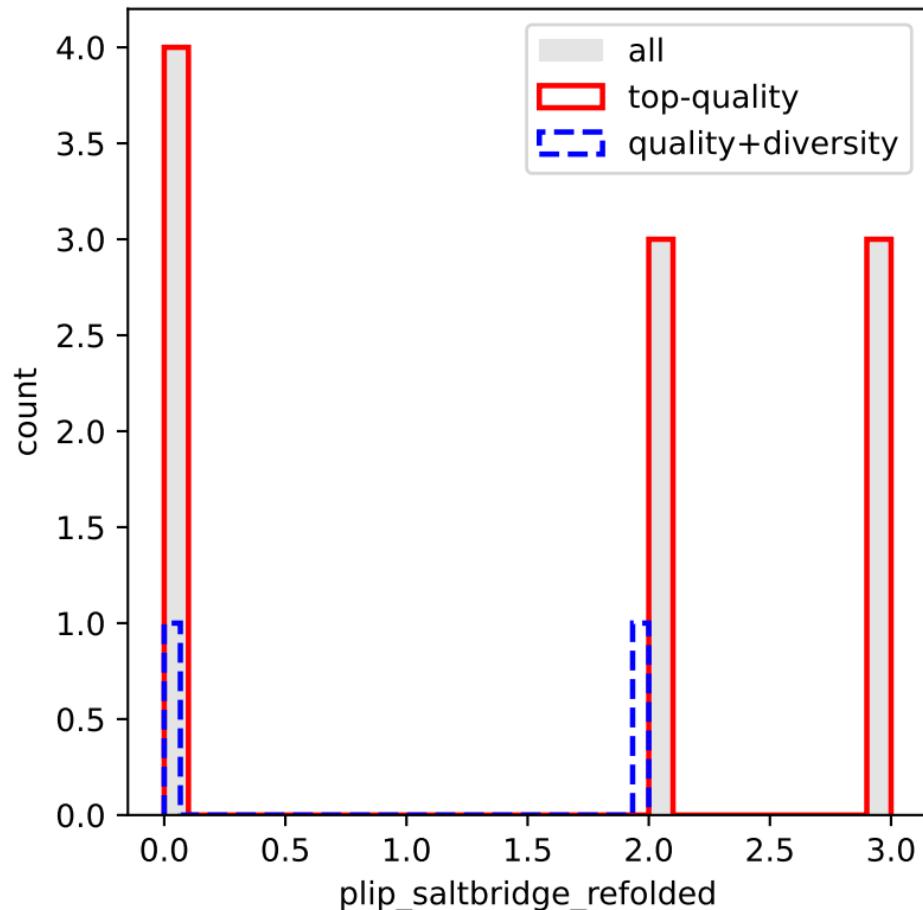
min_design_to_target_pae



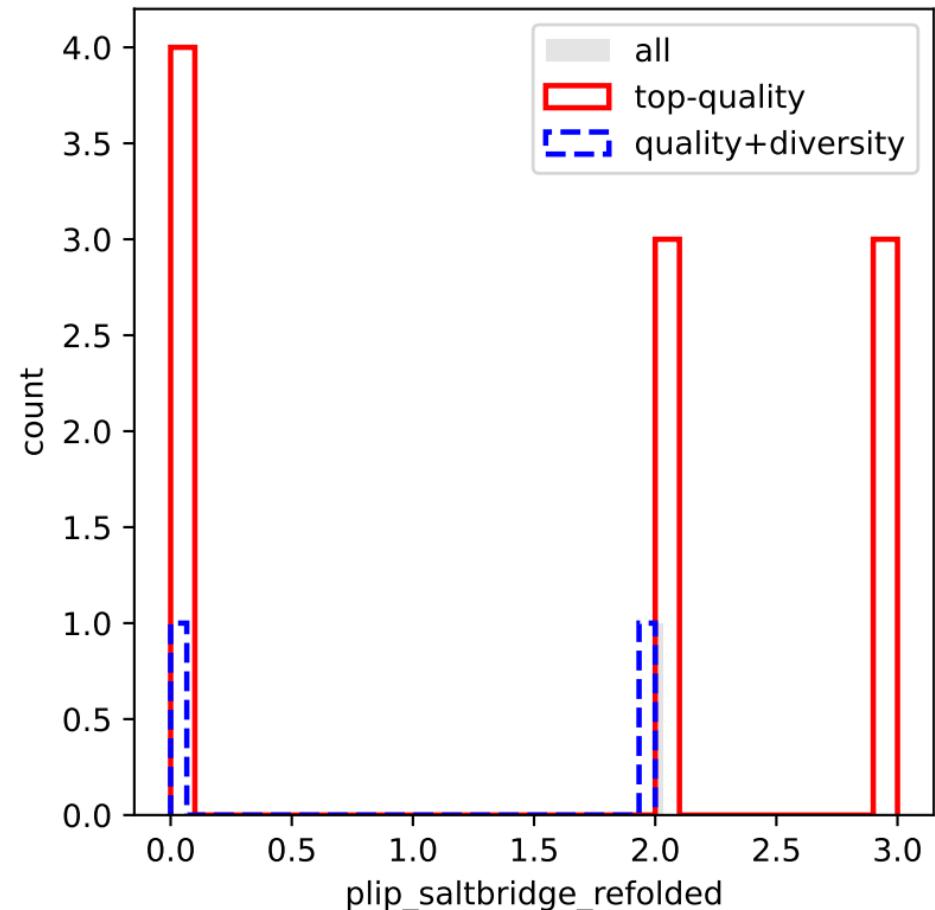
min_design_to_target_pae (Designs passing RMSD threshold)



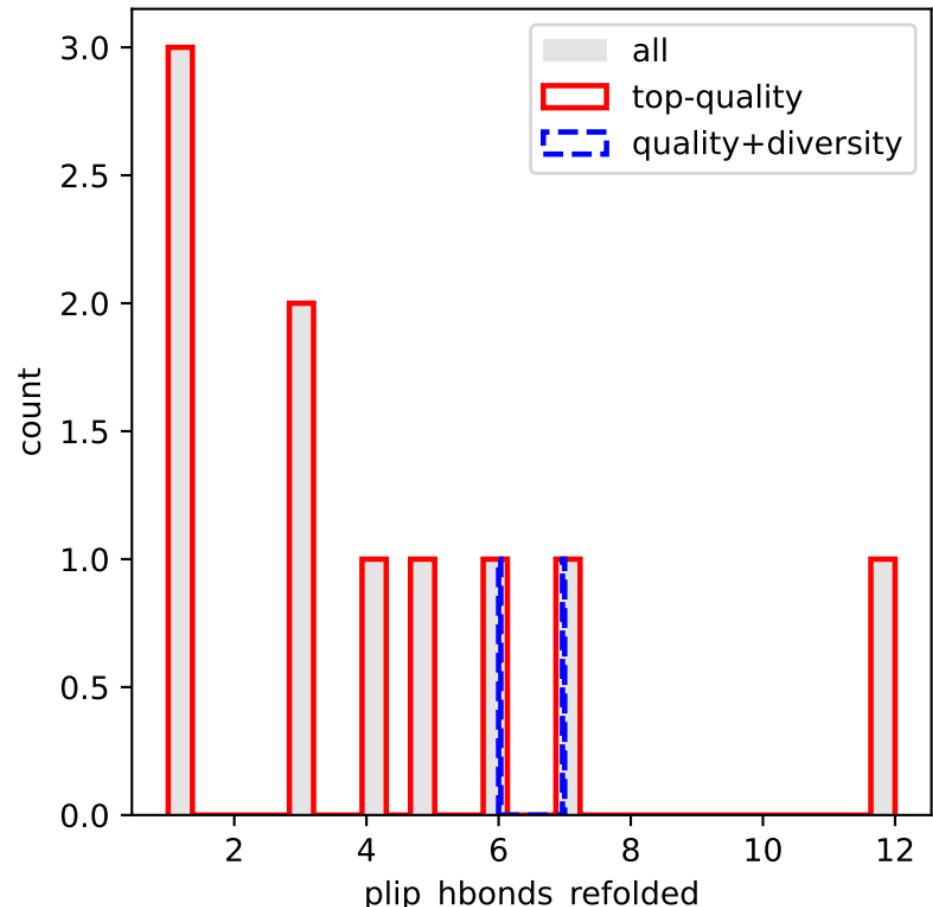
plip_saltbridge_refolded



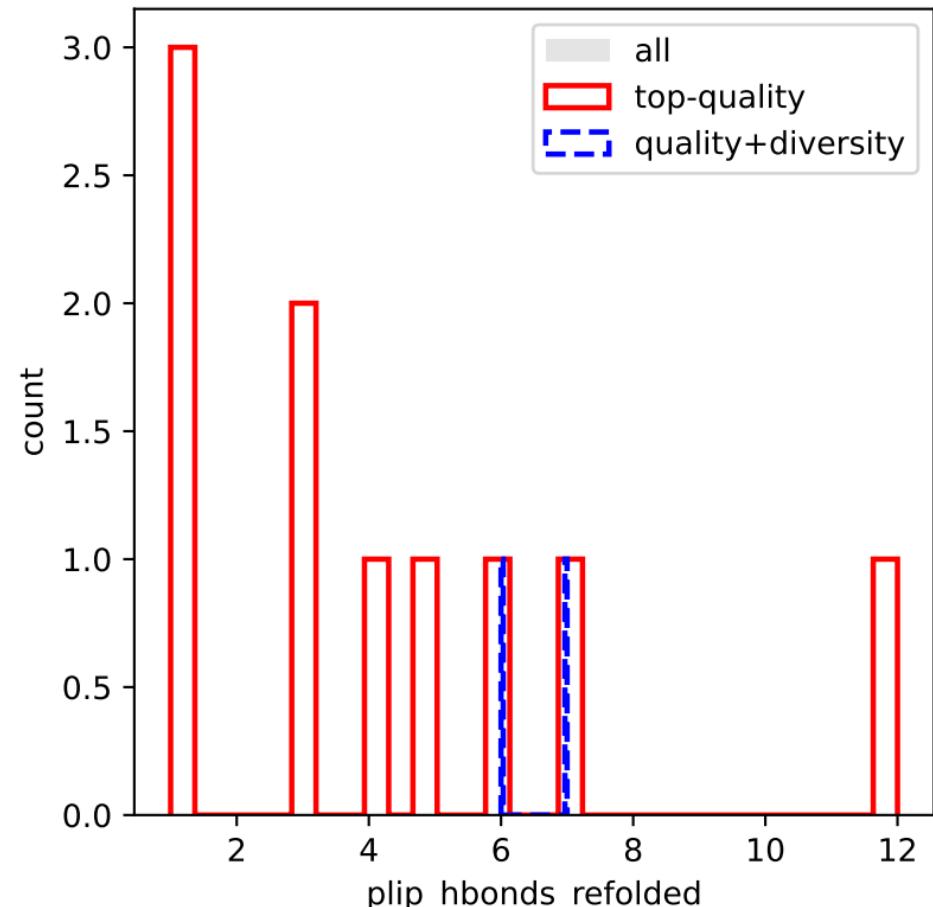
plip_saltbridge_refolded (Designs passing RMSD threshold)



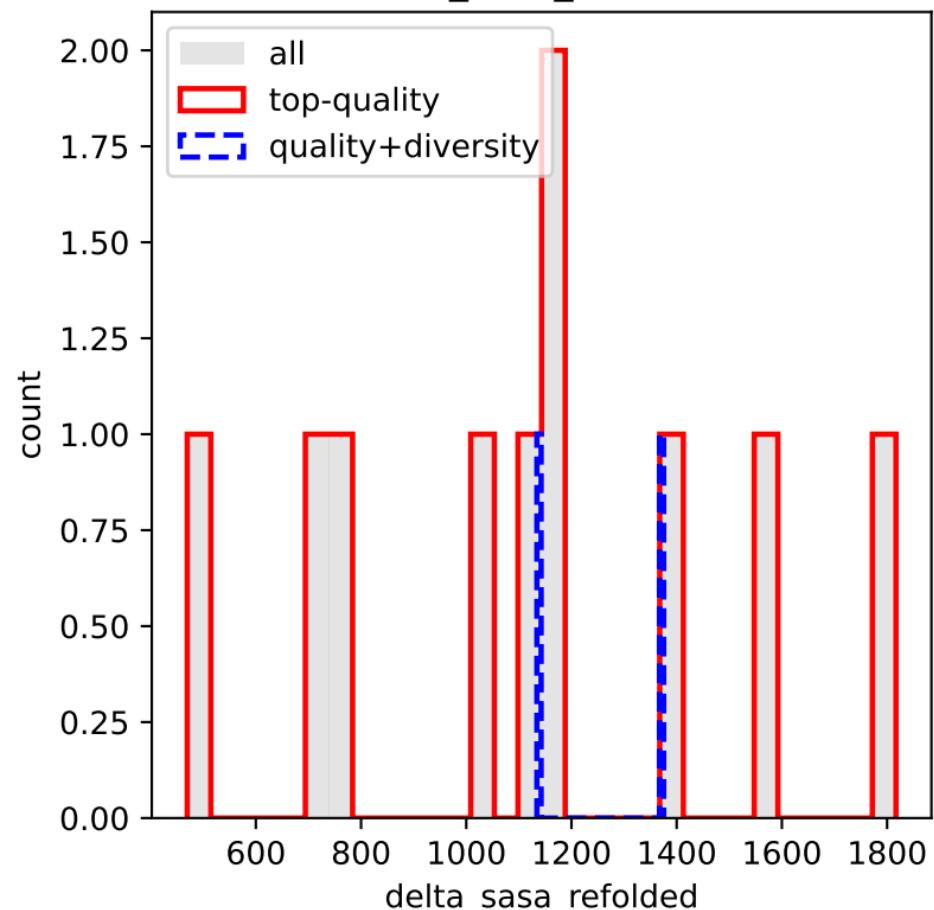
plip_hbonds_refolded



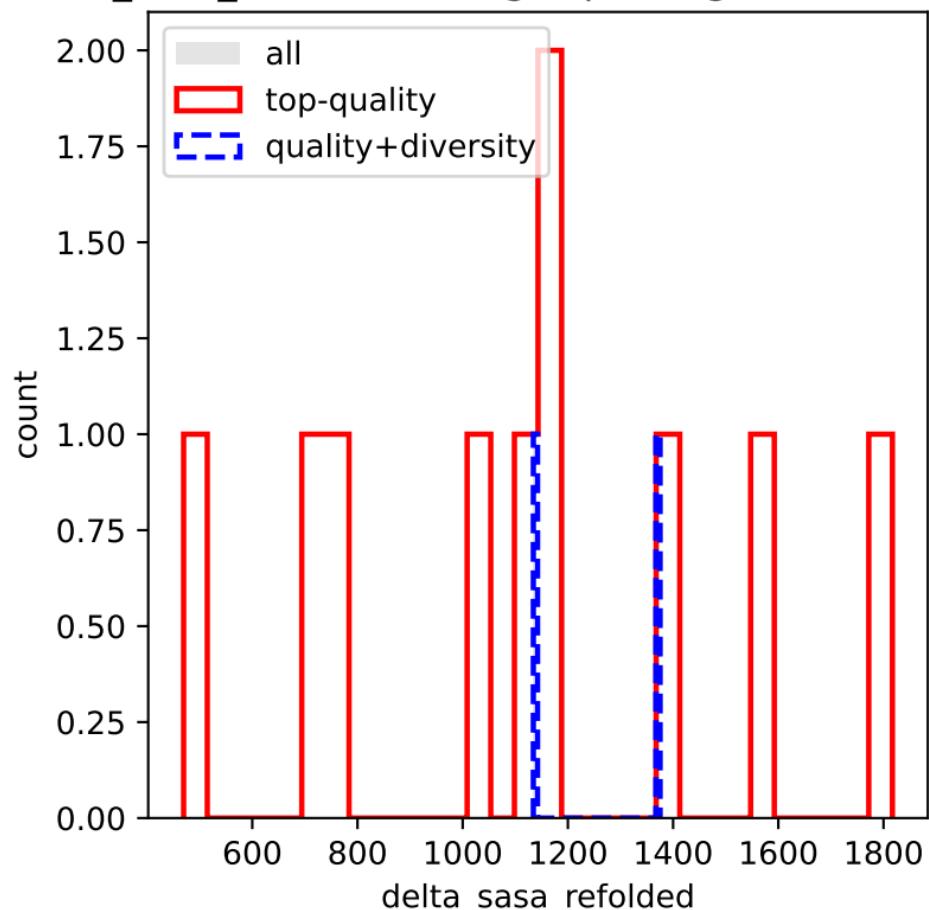
plip_hbonds_refolded (Designs passing RMSD thres)



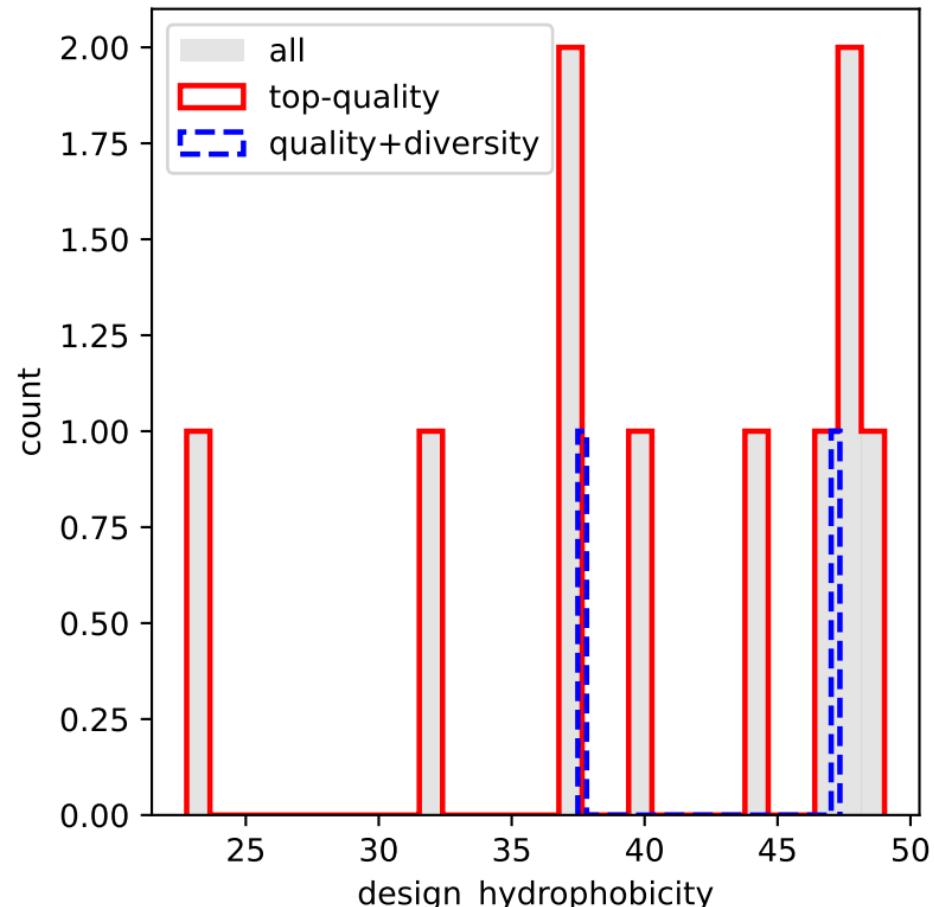
delta_sasa_refolded



delta_sasa_refolded (Designs passing RMSD thres)



design_hydrophobicity



design_hydrophobicity (Designs passing RMSD thre

