

Supporting Information:

Pareto Optimization to Accelerate Multi-Objective Virtual Screening

Jenna C. Fromer,[†] David E. Graff,^{†,‡} and Connor W. Coley^{*,†,¶}

[†]*Department of Chemical Engineering, MIT, Cambridge, MA 02139*

[‡]*Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA
02138*

[¶]*Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA 02139*

E-mail: ccoley@mit.edu

S1 Model-guided Optimization Algorithms

Algorithm S1: Multi-objective Bayesian optimization using scalarization

Input: Objective function $\mathbf{f} : x \mapsto \mathbb{R}^N$, weight vector $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_N]$, surrogate model \hat{f} , candidate set \mathcal{X} , acquisition function $\alpha : x \mapsto \mathbb{R}$, initial observation size b_0 , batch size b

- 1 Select random subset of design space: $\mathcal{X}_0 \subset \mathcal{X} : |\mathcal{X}_0| = b_0$
- 2 Initialize dataset: $\mathcal{D}_0 \leftarrow \{(x, \boldsymbol{\lambda} \cdot \mathbf{f}(x)) : x \in \mathcal{X}_0\}$
- 3 **for** $t \leftarrow 1 \dots T$ **do**
- 4 Train \hat{f} on \mathcal{D}_{t-1}
- 5 Select new batch: $\mathcal{X}_t \leftarrow \arg \max_{\mathcal{X}_t \subset \mathcal{X} : |\mathcal{X}_t|=b} \sum_{x \in \mathcal{X}_t} \alpha(x; \hat{f}, \mathcal{D}_{t-1})$
- 6 Update dataset: $\mathcal{D}_t \leftarrow \mathcal{D}_{t-1} \cup \{(x, \boldsymbol{\lambda} \cdot \mathbf{f}(x)) : x \in \mathcal{X}_t\}$
- 7 **end**

Output: $\arg \max_{x \in \mathcal{D}_t} \boldsymbol{\lambda} \cdot \mathbf{f}(x)$

Algorithm S2: Multi-objective Bayesian optimization using Pareto optimization

Input: Objective function $\mathbf{f} : x \mapsto \mathbb{R}^N$, surrogate models $\{\hat{f}^{(n)}\}_{n=1}^N$, candidate set \mathcal{X} , acquisition function $\alpha : x \mapsto \mathbb{R}$, initial observation size b_0 , batch size b

- 1 Select random subset of design space: $\mathcal{X}_0 \subset \mathcal{X} : |\mathcal{X}_0| = b_0$
- 2 Initialize dataset: $\mathcal{D}_0 \leftarrow \{(x, \mathbf{f}(x)) : x \in \mathcal{X}_0\}$
- 3 Calculate Pareto front: $\mathcal{P}_0 \leftarrow \text{pareto_front}(\mathcal{D}_0)$
- 4 **for** $t \leftarrow 1 \dots T$ **do**
- 5 **for** $n \leftarrow 1 \dots N$ **do**
- 6 Train $\hat{f}^{(n)}$ on $\{(x, f_n) : x, \mathbf{f}(x) \in \mathcal{D}_{t-1}\}$
- 7 **end**
- 8 Select new batch: $\mathcal{X}_t \leftarrow \arg \max_{\mathcal{X}_t \subset \mathcal{X} : |\mathcal{X}_t|=b} \sum_{x \in \mathcal{X}_t} \alpha(x; \{\hat{f}^{(n)}\}, \mathcal{D}_{t-1}, \mathcal{P}_{t-1})$
- 9 Update dataset: $\mathcal{D}_t \leftarrow \mathcal{D}_{t-1} \cup \{(x, \mathbf{f}(x)) : x \in \mathcal{X}_t\}$
- 10 Update Pareto front: $\mathcal{P}_t \leftarrow \text{pareto_front}(\mathcal{D}_t)$
- 11 **end**

Output: \mathcal{P}_T

S2 Performance Metrics

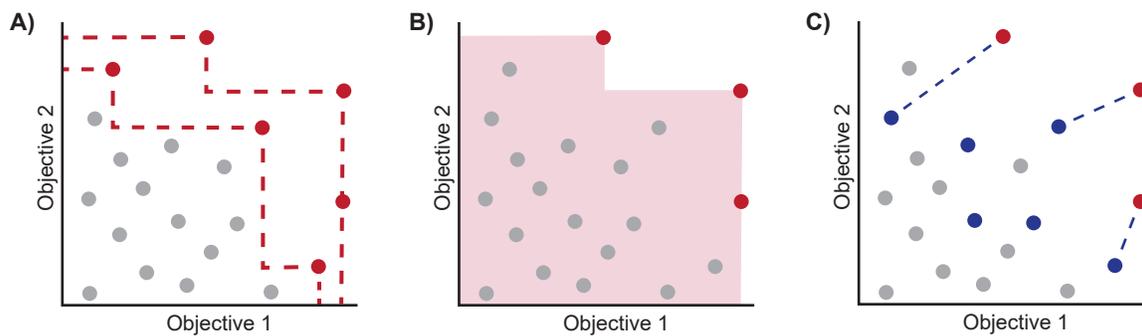


Figure S1: Illustration of the Pareto optimization evaluation metrics considered in this work. (A) Definition of true top- k molecules through non-dominated sorting. Here, the top 30% are shown in red. (B) Hypervolume metric. (C) Inverted generation distance, which averages the shortest distance between points on the true Pareto front (red) and acquired points (blue).

S3 Additional Metrics to Compare Acquisition Functions

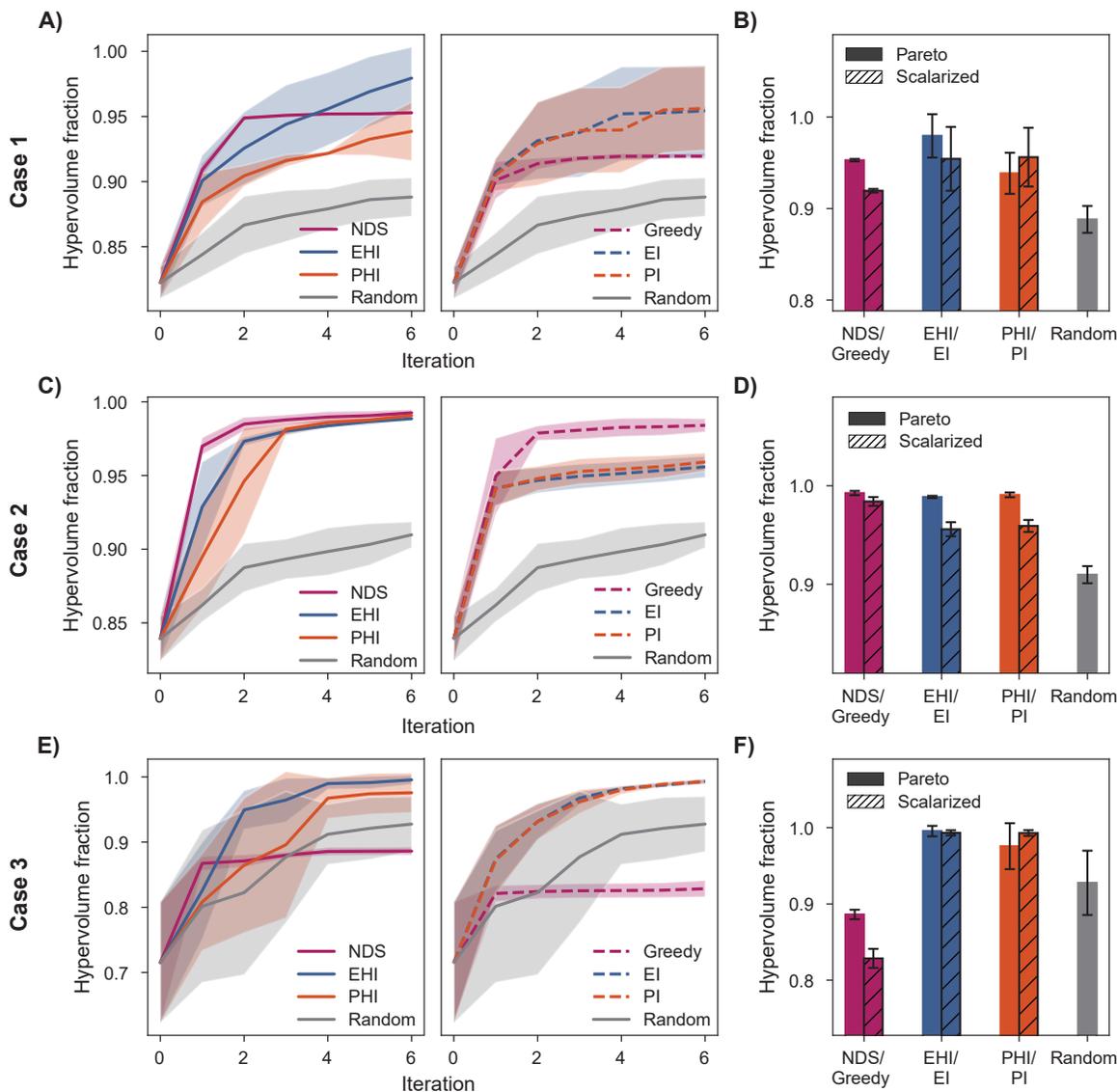


Figure S2: Fraction of hypervolume acquired across iterations (A, C, E) and after six iterations (B, D, F) for three case studies. All cases minimized docking scores to one target and maximized docking scores to one off-target. Cases 1, 2, and 3 aimed to identify putative selective inhibitors of DRD3 over DRD2, JAK2 over LCK, and IGF1R over CYP3A4, respectively. Docking scores to all targets were extracted from the published DOCKSTRING dataset of 260k molecules.^{S1} 1% of the virtual library was acquired at each iteration using top-k batching. Error bars (B, D, F) and shaded regions (A, C, E) denote ± 1 standard deviation across five runs.

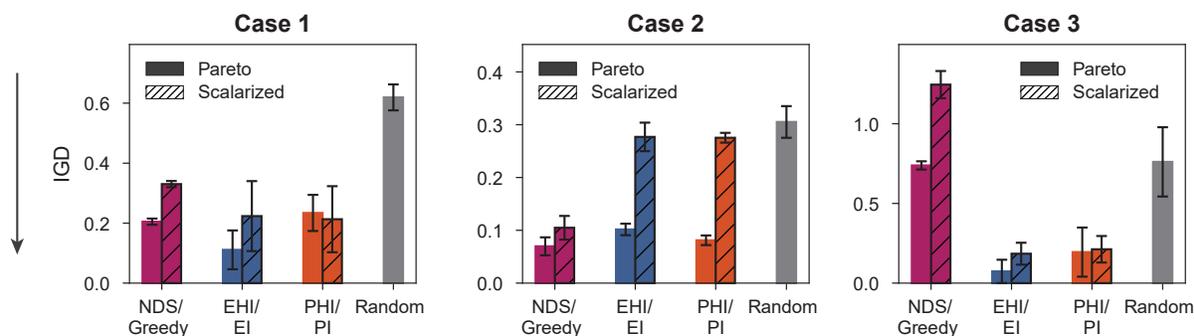


Figure S3: Inverted generational distance^{S2,S3} (IGD) after six iterations for three case studies, calculated with the python package pymoo.^{S4} IGD quantifies the distance between the true Pareto front and the solution set. While IGD measures the quality of solution sets and captures the uniformity of solution set distributions, the metric can be misleading because is not Pareto compliant and requires a uniformly distributed Pareto front.^{S5} All cases minimized docking scores to one target and maximized docking scores to one off-target. Cases 1, 2, and 3 aimed to identify putative selective inhibitors of DRD3 over DRD2, JAK2 over LCK, and IGF1R over CYP3A4, respectively. Docking scores to all targets were extracted from from the published DOCKSTRING dataset of 260k molecules.^{S1} 1% of the virtual library was acquired at each iteration using top-k batching. Error bars denote \pm one standard deviation across five runs.

S4 Molecular Diversity Visualization

2-dimensional projections of molecular fingerprints can illustrate molecular diversity in a qualitative sense. We use UMAP projections^{S6} to visualize the improvement in molecular diversity of acquired points with diversity-enhanced acquisition strategies. Figure S4 shows UMAP projections of acquired points at iterations 1, 3, and 5 for single experiments using different diversity-enhancing acquisition strategies. These experiments were for the identification of putative IGF1R inhibitors with selectivity over CYP3A4. UMAP embeddings were trained on the entire searched library, shown as blue density plots. Diversity is compared for acquisition that implements no clustering, clustering in the feature space, clustering in the objective space, and clustering in both spaces. The acquired molecules (red points) are qualitatively more dispersed across the chemical space spanned by the library when compared to the points acquired without clustering. While the visualization of molecular diversity through dimensional reduction is qualitative in nature, the difference in the chemical space acquired in the two runs suggests that diversity-enhanced acquisition improves the structural diversity of acquired points.

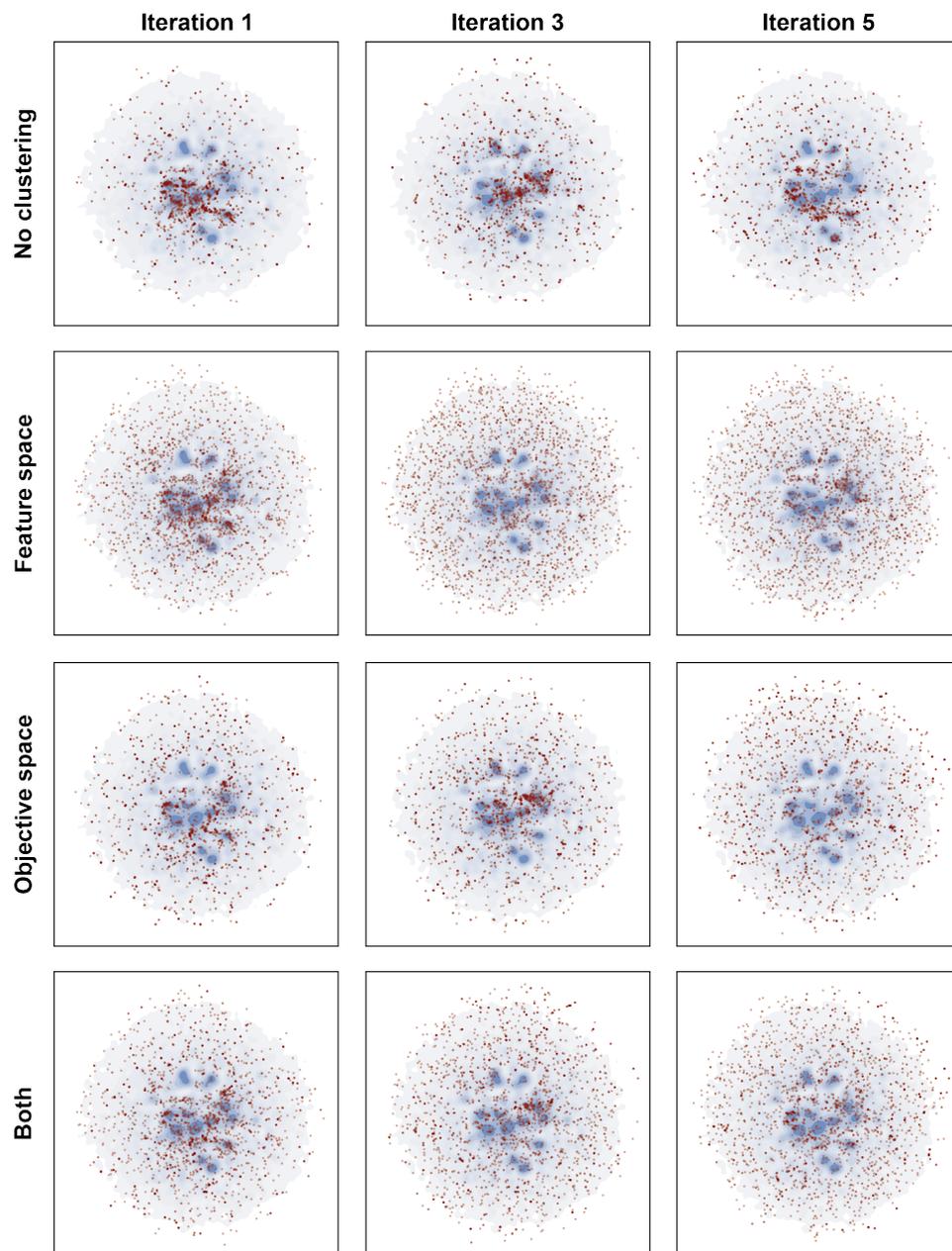


Figure S4: UMAP projections demonstrating molecular diversity of acquired points using standard and diversity-enhanced acquisition. Each row is a single run using the specified diversity-enhanced acquisition strategy with PHI. The runs corresponding to each row were initialized with the same random set of acquired points at iteration 0 and the same model seed. Docking scores computed with DOCKSTRING^{S1} to IGF1R and CYP3A4 were minimized and maximized, respectively, to identify putative selective inhibitors of IGF1R. Points acquired at iterations 1, 3, and 5 are shown for the four acquisition strategies tested in red. UMAP projections were trained on the entire virtual library (shown as a blue density behind acquired points).

S5 Large-Scale Multi-Objective Screen Results

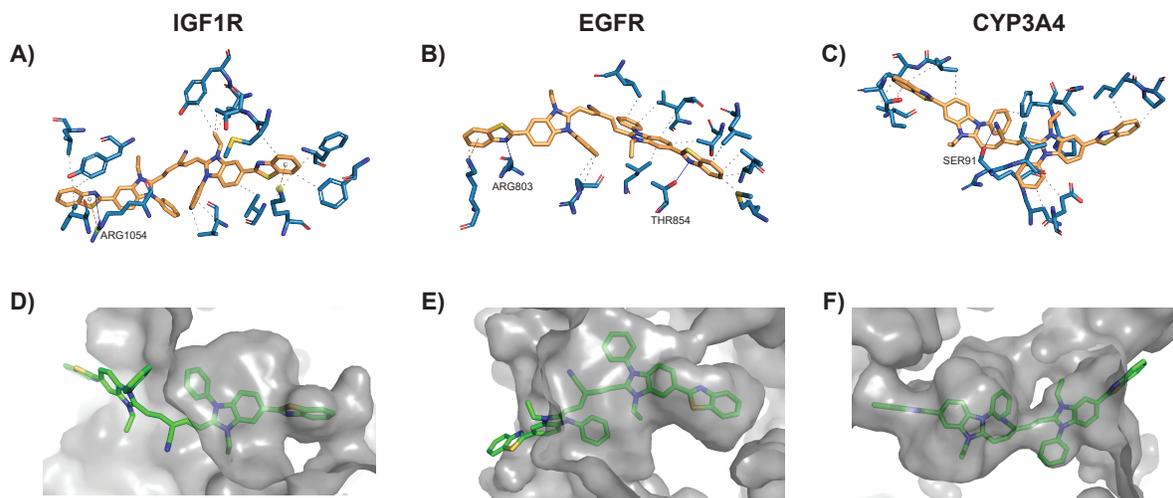


Figure S5: Docking poses of one non-dominated molecule (M4, Figure S6) predicted to selectively bind IGF1R and EGFR over CYP3A4. Docking poses were computed with DOCK-STRING,^{S1} an AutoDock Vina wrapper with prepared docking settings for IGF1R, EGFR, and CYP3A4. The Enamine screening database^{S7} of over 4M molecules was used as the virtual library. (A-C) Protein-ligand interactions of M4 with IGF1R, EGFR, and CYP3A4 prepared with PLIP.^{S8} (D-F) Space-filling visualization of M4 in the binding pockets of IGF1R, EGFR, and CYP3A4. The docking-based optimization for predicted selectivity favors bulky molecules like M4 that can fit in the binding pockets of targets IGF1R and EGFR but form unresolvable steric clashes with residues that form the pocket of off-target CYP3A4.

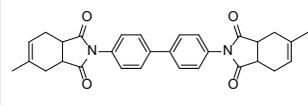
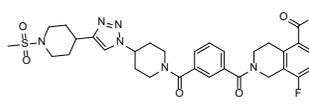
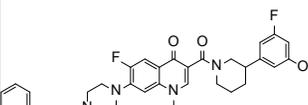
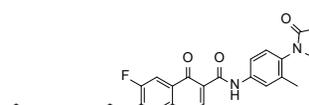
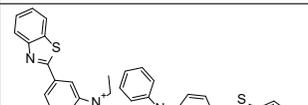
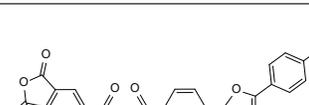
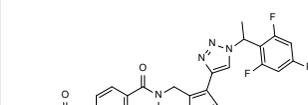
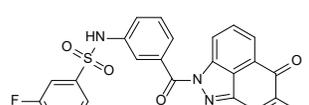
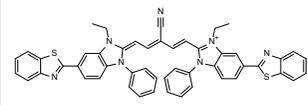
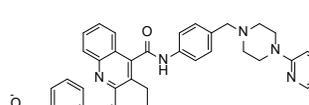
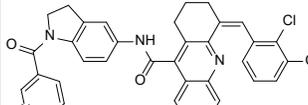
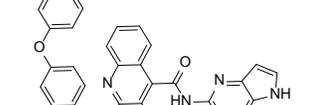
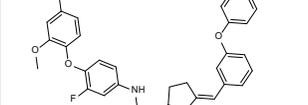
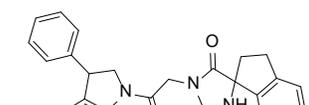
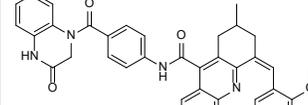
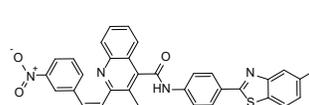
ID	Structure	Scores	ID	Structure	Scores
M0		IGF1R: -8.4 EGFR: -12.7 CYP3A4: -8.3	M8		IGF1R: -10.2 EGFR: -13.5 CYP3A4: -11.8
M1		IGF1R: -9.2 EGFR: -12.5 CYP3A4: -8.0	M9		IGF1R: -10.4 EGFR: -12.2 CYP3A4: -9.1
M2		IGF1R: -9.3 EGFR: -11.9 CYP3A4: -2.8	M10		IGF1R: -10.5 EGFR: -13.4 CYP3A4: -11.2
M3		IGF1R: -9.5 EGFR: -14.1 CYP3A4: -12.1	M11		IGF1R: -10.5 EGFR: -13.7 CYP3A4: -12.1
M4		IGF1R: -9.5 EGFR: -11.6 CYP3A4: 0	M12		IGF1R: -10.9 EGFR: -13.3 CYP3A4: -9.6
M5		IGF1R: -9.7 EGFR: -12.1 CYP3A4: -7.3	M13		IGF1R: -11.1 EGFR: -14.2 CYP3A4: -12.7
M6		IGF1R: -10 EGFR: -13.8 CYP3A4: -11.9	M14		IGF1R: -11.1 EGFR: -13 CYP3A4: -11.9
M7		IGF1R: -10.2 EGFR: -13.2 CYP3A4: -8.5	M15		IGF1R: -11.1 EGFR: -12.5 CYP3A4: -9.3

Figure S6: Molecules 0-15 of the 39 non-dominated molecules for an exemplary 3-objective optimization aiming to identify binders of IGF1R and EGFR with selectivity over CYP3A4 from the Enamine screening library of over 4M molecules.^{S7} Docking scores were computed using DOCKSTRING.^{S1} Docking scores to IGF1R and EGFR were minimized, and scores to the off-target CYP3A4 were maximized.

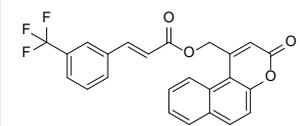
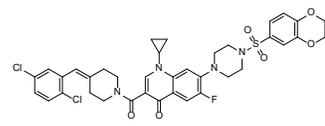
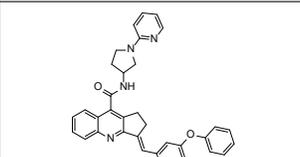
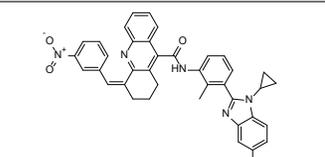
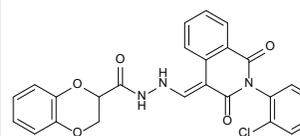
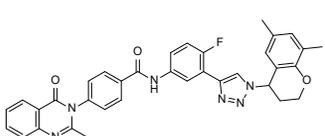
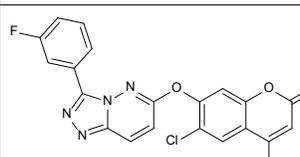
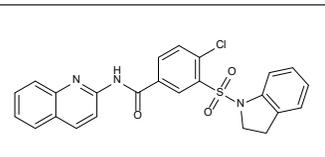
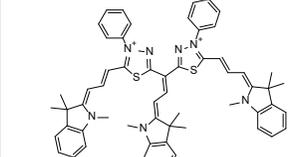
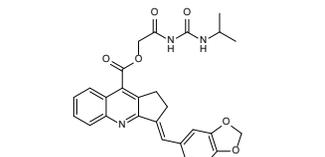
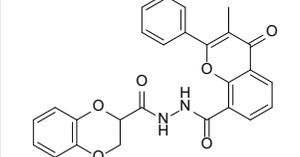
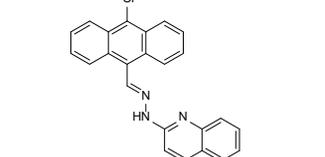
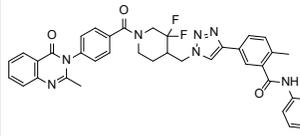
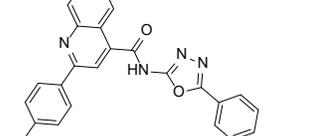
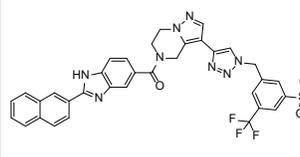
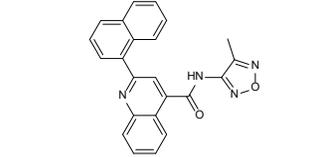
ID	Structure	Scores	ID	Structure	Scores
M16		IGF1R: -11.2 EGFR: -12.2 CYP3A4: -9.9	M24		IGF1R: -11.6 EGFR: -11.8 CYP3A4: -9.1
M17		IGF1R: -11.2 EGFR: -12.2 CYP3A4: -9.9	M25		IGF1R: -11.6 EGFR: -13.3 CYP3A4: -12.2
M18		IGF1R: -11.2 EGFR: -12.6 CYP3A4: -10.3	M26		IGF1R: -11.6 EGFR: -12.5 CYP3A4: -11.3
M19		IGF1R: -11.3 EGFR: -12.2 CYP3A4: -11.1	M27		IGF1R: -11.8 EGFR: -12.1 CYP3A4: -11.7
M20		IGF1R: -11.3 EGFR: -12 CYP3A4: -4.3	M28		IGF1R: -11.9 EGFR: -11.3 CYP3A4: -10.5
M21		IGF1R: -11.4 EGFR: -12.9 CYP3A4: -11.6	M29		IGF1R: -11.9 EGFR: -11.9 CYP3A4: -11
M22		IGF1R: -11.4 EGFR: -13.8 CYP3A4: -13.2	M30		IGF1R: -11.9 EGFR: -12.3 CYP3A4: -12.1
M23		IGF1R: -11.6 EGFR: -12.9 CYP3A4: -12	M31		IGF1R: -12 EGFR: -11.1 CYP3A4: -11.3

Figure S7: Molecules 16-31 of the 39 non-dominated molecules in the searched library for an exemplary 3-objective optimization aiming to identify binders of IGF1R and EGFR with selectivity over CYP3A4 from the Enamine screening library of over 4M molecules.^{S7} Docking scores were computed using DOCKSTRING.^{S1} Docking scores to IGF1R and EGFR were minimized, and scores to the off-target CYP3A4 were maximized.

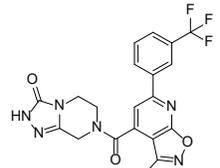
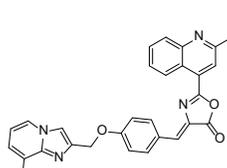
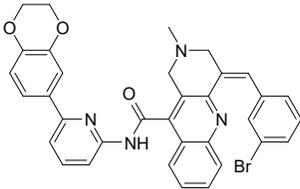
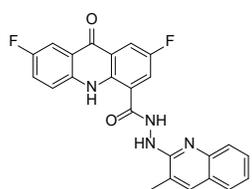
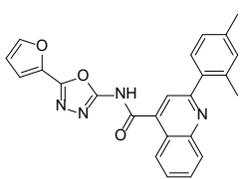
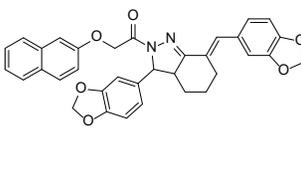
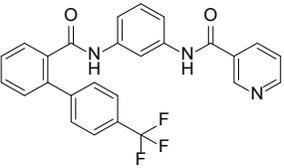
ID	Structure	Scores	ID	Structure	Scores
M32		IGF1R: -12.1 EGFR: -11.4 CYP3A4: -11.4	M36		IGF1R: -12.5 EGFR: -12.9 CYP3A4: -13.1
M33		IGF1R: -12.1 EGFR: -13.6 CYP3A4: -12.6	M37		IGF1R: -12.5 EGFR: -12.7 CYP3A4: -12.2
M34		IGF1R: -12.1 EGFR: -11.5 CYP3A4: -11.6	M38		IGF1R: -13 EGFR: -12 CYP3A4: -12.6
M35		IGF1R: -12.2 EGFR: -10.6 CYP3A4: -11.3			

Figure S8: Molecules 32-38 of the 39 non-dominated molecules in the searched library for an exemplary 3-objective optimization aiming to identify binders of IGF1R and EGFR with selectivity over CYP3A4 from the Enamine screening library of over 4M molecules.^{S7} Docking scores were computed using DOCKSTRING.^{S1} Docking scores to IGF1R and EGFR were minimized, and scores to the off-target CYP3A4 were maximized.

S6 Tables

Table S1: Fraction of top-1% acquired across iterations for retrospective multi-objective virtual screening experiments in Section 3.2, shown in Figure 4. A comparison with MO-MEMES^{S9} is also included. All cases minimized docking scores to one target and maximized docking scores to one off-target. Cases 1, 2, and 3 aimed to identify putative selective inhibitors of DRD3 over DRD2, JAK2 over LCK, and IGF1R over CYP3A4, respectively. Docking scores to all targets were extracted from the published DOCKSTRING dataset of 260k molecules.^{S1} 1% of the virtual library was acquired at each iteration using top-k batching. Means \pm one standard deviation across 5 trials are shown.

Case	Acquisition Function	Iteration						
		0	1	2	3	4	5	6
1	EHI	0.01 \pm 0.00	0.15 \pm 0.01	0.21 \pm 0.03	0.25 \pm 0.03	0.30 \pm 0.02	0.33 \pm 0.02	0.36 \pm 0.02
1	EI	0.01 \pm 0.00	0.14 \pm 0.02	0.17 \pm 0.02	0.20 \pm 0.02	0.23 \pm 0.02	0.25 \pm 0.02	0.27 \pm 0.02
1	Greedy	0.01 \pm 0.00	0.17 \pm 0.01	0.25 \pm 0.01	0.31 \pm 0.02	0.36 \pm 0.01	0.41 \pm 0.01	0.45 \pm 0.01
1	MO-MEMES	0.01 \pm 0.00	0.09 \pm 0.02	0.11 \pm 0.02	0.13 \pm 0.02	0.15 \pm 0.02	0.18 \pm 0.02	0.19 \pm 0.01
1	NDS	0.01 \pm 0.00	0.16 \pm 0.02	0.25 \pm 0.01	0.30 \pm 0.01	0.34 \pm 0.01	0.37 \pm 0.01	0.40 \pm 0.01
1	PHI	0.01 \pm 0.00	0.17 \pm 0.01	0.25 \pm 0.02	0.31 \pm 0.02	0.36 \pm 0.02	0.40 \pm 0.01	0.43 \pm 0.01
1	PI	0.01 \pm 0.00	0.14 \pm 0.02	0.18 \pm 0.02	0.21 \pm 0.01	0.24 \pm 0.02	0.27 \pm 0.02	0.29 \pm 0.02
1	Random	0.01 \pm 0.00	0.02 \pm 0.00	0.03 \pm 0.00	0.04 \pm 0.00	0.05 \pm 0.01	0.06 \pm 0.01	0.07 \pm 0.01
2	EHI	0.01 \pm 0.00	0.13 \pm 0.04	0.15 \pm 0.04	0.17 \pm 0.03	0.20 \pm 0.03	0.23 \pm 0.04	0.25 \pm 0.04
2	EI	0.01 \pm 0.00	0.03 \pm 0.00	0.03 \pm 0.00	0.04 \pm 0.00	0.05 \pm 0.00	0.07 \pm 0.00	0.08 \pm 0.00
2	Greedy	0.01 \pm 0.00	0.09 \pm 0.03	0.13 \pm 0.03	0.16 \pm 0.04	0.20 \pm 0.04	0.22 \pm 0.04	0.25 \pm 0.04
2	MO-MEMES	0.01 \pm 0.00	0.02 \pm 0.00	0.03 \pm 0.00	0.05 \pm 0.00	0.06 \pm 0.00	0.07 \pm 0.00	0.07 \pm 0.01
2	NDS	0.01 \pm 0.00	0.13 \pm 0.01	0.21 \pm 0.01	0.26 \pm 0.01	0.29 \pm 0.01	0.32 \pm 0.01	0.34 \pm 0.01
2	PHI	0.01 \pm 0.00	0.18 \pm 0.02	0.21 \pm 0.02	0.25 \pm 0.02	0.28 \pm 0.01	0.31 \pm 0.00	0.35 \pm 0.02
2	PI	0.01 \pm 0.00	0.03 \pm 0.00	0.04 \pm 0.00	0.05 \pm 0.00	0.06 \pm 0.00	0.07 \pm 0.01	0.08 \pm 0.00
2	Random	0.01 \pm 0.00	0.02 \pm 0.00	0.03 \pm 0.00	0.04 \pm 0.00	0.05 \pm 0.00	0.06 \pm 0.00	0.07 \pm 0.00
3	EHI	0.01 \pm 0.00	0.04 \pm 0.01	0.06 \pm 0.01	0.11 \pm 0.01	0.17 \pm 0.02	0.23 \pm 0.01	0.26 \pm 0.01
3	EI	0.01 \pm 0.00	0.03 \pm 0.01	0.04 \pm 0.01	0.05 \pm 0.01	0.06 \pm 0.01	0.08 \pm 0.01	0.09 \pm 0.01
3	Greedy	0.01 \pm 0.00	0.03 \pm 0.00	0.06 \pm 0.00	0.08 \pm 0.00	0.11 \pm 0.01	0.13 \pm 0.01	0.15 \pm 0.01
3	MO-MEMES	0.01 \pm 0.00	0.02 \pm 0.00	0.03 \pm 0.00	0.05 \pm 0.00	0.06 \pm 0.00	0.08 \pm 0.00	0.10 \pm 0.00
3	NDS	0.01 \pm 0.00	0.05 \pm 0.01	0.08 \pm 0.00	0.12 \pm 0.01	0.15 \pm 0.01	0.17 \pm 0.01	0.19 \pm 0.01
3	PHI	0.01 \pm 0.00	0.03 \pm 0.01	0.07 \pm 0.02	0.11 \pm 0.03	0.17 \pm 0.04	0.24 \pm 0.04	0.28 \pm 0.03
3	PI	0.01 \pm 0.00	0.03 \pm 0.01	0.04 \pm 0.01	0.05 \pm 0.01	0.06 \pm 0.01	0.07 \pm 0.00	0.09 \pm 0.00
3	Random	0.01 \pm 0.00	0.02 \pm 0.00	0.03 \pm 0.00	0.04 \pm 0.00	0.05 \pm 0.00	0.06 \pm 0.00	0.07 \pm 0.00

Table S2: Hypervolume profiles for retrospective multi-objective virtual screening experiments in Section 3.2, shown in Figure S2. A comparison with MO-MEMES^{S9} is also included. 1% of the virtual library was acquired at each iteration using top-k batching. Means \pm one standard deviation across 5 trials are shown.

Case	Acquisition Function	0	1	2	Iteration 3	4	5	6
1	EHI	0.82 \pm 0.01	0.90 \pm 0.02	0.93 \pm 0.03	0.94 \pm 0.03	0.96 \pm 0.03	0.97 \pm 0.03	0.98 \pm 0.02
1	EI	0.82 \pm 0.01	0.91 \pm 0.01	0.93 \pm 0.03	0.94 \pm 0.03	0.95 \pm 0.04	0.95 \pm 0.04	0.95 \pm 0.03
1	Greedy	0.82 \pm 0.01	0.90 \pm 0.01	0.91 \pm 0.00	0.92 \pm 0.00	0.92 \pm 0.00	0.92 \pm 0.00	0.92 \pm 0.00
1	MO-MEMES	0.82 \pm 0.01	0.89 \pm 0.02	0.90 \pm 0.02	0.91 \pm 0.01	0.93 \pm 0.01	0.95 \pm 0.03	0.96 \pm 0.03
1	NDS	0.82 \pm 0.01	0.91 \pm 0.00	0.95 \pm 0.00				
1	PHI	0.82 \pm 0.01	0.88 \pm 0.02	0.90 \pm 0.01	0.92 \pm 0.00	0.92 \pm 0.00	0.93 \pm 0.01	0.94 \pm 0.02
1	PI	0.82 \pm 0.01	0.91 \pm 0.01	0.93 \pm 0.03	0.94 \pm 0.03	0.94 \pm 0.03	0.95 \pm 0.03	0.96 \pm 0.03
1	Random	0.82 \pm 0.01	0.84 \pm 0.02	0.87 \pm 0.02	0.87 \pm 0.02	0.88 \pm 0.02	0.89 \pm 0.02	0.89 \pm 0.01
2	EHI	0.84 \pm 0.01	0.93 \pm 0.03	0.97 \pm 0.00	0.98 \pm 0.00	0.98 \pm 0.00	0.99 \pm 0.00	0.99 \pm 0.00
2	EI	0.84 \pm 0.01	0.94 \pm 0.01	0.95 \pm 0.01	0.95 \pm 0.01	0.95 \pm 0.01	0.95 \pm 0.01	0.96 \pm 0.01
2	Greedy	0.84 \pm 0.01	0.95 \pm 0.03	0.98 \pm 0.00	0.98 \pm 0.01	0.98 \pm 0.01	0.98 \pm 0.01	0.98 \pm 0.00
2	MO-MEMES	0.84 \pm 0.02	0.88 \pm 0.01	0.91 \pm 0.02	0.95 \pm 0.01	0.96 \pm 0.00	0.96 \pm 0.00	0.96 \pm 0.00
2	NDS	0.84 \pm 0.01	0.97 \pm 0.01	0.98 \pm 0.00	0.99 \pm 0.00	0.99 \pm 0.00	0.99 \pm 0.00	0.99 \pm 0.00
2	PHI	0.84 \pm 0.01	0.89 \pm 0.03	0.95 \pm 0.04	0.98 \pm 0.00	0.99 \pm 0.00	0.99 \pm 0.00	0.99 \pm 0.00
2	PI	0.84 \pm 0.01	0.94 \pm 0.01	0.95 \pm 0.01	0.95 \pm 0.01	0.95 \pm 0.01	0.96 \pm 0.01	0.96 \pm 0.01
2	Random	0.84 \pm 0.01	0.86 \pm 0.01	0.89 \pm 0.02	0.89 \pm 0.01	0.90 \pm 0.02	0.90 \pm 0.01	0.91 \pm 0.01
3	EHI	0.72 \pm 0.09	0.83 \pm 0.07	0.95 \pm 0.03	0.96 \pm 0.03	0.99 \pm 0.01	0.99 \pm 0.01	1.00 \pm 0.01
3	EI	0.72 \pm 0.09	0.87 \pm 0.05	0.93 \pm 0.03	0.97 \pm 0.01	0.98 \pm 0.00	0.99 \pm 0.00	0.99 \pm 0.00
3	Greedy	0.72 \pm 0.09	0.82 \pm 0.01	0.82 \pm 0.01	0.83 \pm 0.01	0.83 \pm 0.01	0.83 \pm 0.01	0.83 \pm 0.01
3	MO-MEMES	0.72 \pm 0.09	0.91 \pm 0.12	0.98 \pm 0.00	0.98 \pm 0.00	0.99 \pm 0.00	0.99 \pm 0.00	1.00 \pm 0.00
3	NDS	0.72 \pm 0.09	0.87 \pm 0.01	0.87 \pm 0.01	0.88 \pm 0.00	0.89 \pm 0.01	0.89 \pm 0.01	0.89 \pm 0.01
3	PHI	0.72 \pm 0.09	0.81 \pm 0.07	0.86 \pm 0.10	0.90 \pm 0.11	0.97 \pm 0.03	0.97 \pm 0.03	0.98 \pm 0.03
3	PI	0.72 \pm 0.09	0.87 \pm 0.05	0.93 \pm 0.03	0.96 \pm 0.02	0.98 \pm 0.01	0.99 \pm 0.00	0.99 \pm 0.00
3	Random	0.72 \pm 0.09	0.80 \pm 0.12	0.82 \pm 0.13	0.88 \pm 0.10	0.91 \pm 0.05	0.92 \pm 0.05	0.93 \pm 0.04

Table S3: Comparison of acquisition functions (including against MO-MEMES^{S9}) using all four evaluation metrics after a fixed exploration budget of 6 iterations. Top-1%, hypervolume (HV), inverted generational distance (IGD), and fraction of the true Pareto front are shown. Values are plotted in Figures 4, 5, and S2. All cases minimized docking scores to one target and maximized docking scores to one off-target. Cases 1, 2, and 3 aimed to identify putative selective inhibitors of DRD3 over DRD2, JAK2 over LCK, and IGF1R over CYP3A4, respectively. Docking scores to all targets were extracted from the published DOCKSTRING dataset of 260k molecules.^{S1} 1% of the virtual library was acquired at each iteration using top-k batching. Means \pm one standard deviation across 5 trials are shown.

Case	Acquisition Function	Top 1%	HV	IGD	Fraction of True Front
1	EHI	0.36 \pm 0.02	0.98 \pm 0.02	0.11 \pm 0.06	0.68 \pm 0.05
1	EI	0.27 \pm 0.02	0.95 \pm 0.03	0.22 \pm 0.12	0.52 \pm 0.06
1	Greedy	0.45 \pm 0.01	0.92 \pm 0.00	0.33 \pm 0.01	0.51 \pm 0.02
1	MO-MEMES	0.19 \pm 0.01	0.96 \pm 0.03	0.28 \pm 0.12	0.40 \pm 0.08
1	NDS	0.40 \pm 0.01	0.95 \pm 0.00	0.20 \pm 0.01	0.49 \pm 0.04
1	PHI	0.43 \pm 0.01	0.94 \pm 0.02	0.23 \pm 0.06	0.61 \pm 0.03
1	PI	0.29 \pm 0.02	0.96 \pm 0.03	0.21 \pm 0.11	0.53 \pm 0.07
1	Random	0.07 \pm 0.01	0.89 \pm 0.01	0.62 \pm 0.04	0.09 \pm 0.04
2	EHI	0.25 \pm 0.04	0.99 \pm 0.00	0.10 \pm 0.01	0.41 \pm 0.03
2	EI	0.08 \pm 0.00	0.96 \pm 0.01	0.28 \pm 0.03	0.20 \pm 0.05
2	Greedy	0.25 \pm 0.04	0.98 \pm 0.00	0.10 \pm 0.02	0.38 \pm 0.05
2	MO-MEMES	0.07 \pm 0.01	0.96 \pm 0.00	0.26 \pm 0.02	0.19 \pm 0.05
2	NDS	0.34 \pm 0.01	0.99 \pm 0.00	0.07 \pm 0.02	0.48 \pm 0.05
2	PHI	0.35 \pm 0.02	0.99 \pm 0.00	0.08 \pm 0.01	0.47 \pm 0.03
2	PI	0.08 \pm 0.00	0.96 \pm 0.01	0.28 \pm 0.01	0.21 \pm 0.03
2	Random	0.07 \pm 0.00	0.91 \pm 0.01	0.31 \pm 0.03	0.03 \pm 0.03
3	EHI	0.26 \pm 0.01	1.00 \pm 0.01	0.07 \pm 0.07	0.82 \pm 0.10
3	EI	0.09 \pm 0.01	0.99 \pm 0.00	0.18 \pm 0.07	0.62 \pm 0.10
3	Greedy	0.15 \pm 0.01	0.83 \pm 0.01	1.25 \pm 0.09	0.02 \pm 0.04
3	MO-MEMES	0.10 \pm 0.00	1.00 \pm 0.00	0.20 \pm 0.04	0.64 \pm 0.05
3	NDS	0.19 \pm 0.01	0.89 \pm 0.01	0.74 \pm 0.03	0.24 \pm 0.08
3	PHI	0.28 \pm 0.03	0.98 \pm 0.03	0.19 \pm 0.15	0.72 \pm 0.15
3	PI	0.09 \pm 0.00	0.99 \pm 0.00	0.21 \pm 0.08	0.60 \pm 0.11
3	Random	0.07 \pm 0.00	0.93 \pm 0.04	0.76 \pm 0.22	0.16 \pm 0.12

Table S4: Top-1% and hypervolume profiles for Case 3 experiments comparing diversity-enhancing acquisition strategies (Section 3.3, Figure 6). All runs used PHI for acquisition. Results are shown for top- k batching without clustering and three diversity-enhanced acquisition strategies that apply clustering. Docking scores to IGF1R and CYP3A4 were minimized and maximized, respectively. The virtual library and docking scores were use as published in DOCKSTRING.^{S1} 1% of the virtual library was acquired at each iteration using top- k batching. Means \pm one standard deviation across 5 trials are shown.

Top-1%							
Cluster Type	0	1	2	Iteration 3	4	5	6
Feature	0.01 \pm 0.00	0.04 \pm 0.01	0.08 \pm 0.02	0.12 \pm 0.02	0.16 \pm 0.02	0.19 \pm 0.02	0.23 \pm 0.02
Feature + Obj	0.01 \pm 0.00	0.04 \pm 0.01	0.07 \pm 0.01	0.12 \pm 0.02	0.16 \pm 0.02	0.20 \pm 0.02	0.24 \pm 0.02
No clustering	0.01 \pm 0.00	0.03 \pm 0.01	0.07 \pm 0.02	0.11 \pm 0.03	0.17 \pm 0.04	0.24 \pm 0.04	0.28 \pm 0.03
Obj	0.01 \pm 0.00	0.03 \pm 0.00	0.05 \pm 0.01	0.09 \pm 0.01	0.13 \pm 0.01	0.17 \pm 0.01	0.20 \pm 0.01
Hypervolume							
Cluster Type	0	1	2	Iteration 3	4	5	6
Feature	0.72 \pm 0.09	0.86 \pm 0.07	0.97 \pm 0.01	0.97 \pm 0.01	0.98 \pm 0.01	0.98 \pm 0.01	0.99 \pm 0.01
Feature + Obj	0.72 \pm 0.09	0.87 \pm 0.04	0.93 \pm 0.04	0.96 \pm 0.04	0.98 \pm 0.02	1.00 \pm 0.00	1.00 \pm 0.00
No clustering	0.72 \pm 0.09	0.81 \pm 0.07	0.86 \pm 0.10	0.90 \pm 0.11	0.97 \pm 0.03	0.97 \pm 0.03	0.98 \pm 0.03
Obj	0.72 \pm 0.09	0.85 \pm 0.06	0.94 \pm 0.03	0.96 \pm 0.04	0.98 \pm 0.01	0.99 \pm 0.01	0.99 \pm 0.01

Table S5: Comparison of diversity-enhancing acquisition strategies using all four evaluation metrics after a fixed exploration budget of 6 iterations (Section 3.3, Figure 6). All runs used PHI for acquisition. Results are shown for top- k batching without clustering and three diversity-enhanced acquisition strategies that apply clustering. Docking scores to IGF1R and CYP3A4 were minimized and maximized, respectively. The virtual library and docking scores were use as published in DOCKSTRING.^{S1} 1% of the virtual library was acquired at each iteration using top- k batching. Means \pm one standard deviation across 5 trials are shown.

Cluster Type	Top 1%	Hypervolume	IGD	Fraction of True Front	Number of Scaffolds
Feature	0.23 \pm 0.02	0.99 \pm 0.01	0.26 \pm 0.08	0.58 \pm 0.07	10605 \pm 146
Feature + Obj	0.24 \pm 0.02	1.00 \pm 0.00	0.09 \pm 0.05	0.84 \pm 0.05	9851 \pm 49
No clustering	0.28 \pm 0.03	0.98 \pm 0.03	0.19 \pm 0.15	0.72 \pm 0.15	7946 \pm 186
Obj	0.20 \pm 0.01	0.99 \pm 0.01	0.25 \pm 0.08	0.64 \pm 0.16	9036 \pm 198

References

- (S1) García-Ortegón, M.; Simm, G. N. C.; Tripp, A. J.; Hernández-Lobato, J. M.; Bender, A.; Bacallado, S. DOCKSTRING: Easy Molecular Docking Yields Better Benchmarks for Ligand Design. *Journal of Chemical Information and Modeling* **2022**, *62*, 3486–3502, downloaded June 2023.
- (S2) Tanabe, R.; Ishibuchi, H. An Analysis of Quality Indicators Using Approximated Optimal Distributions in a 3-D Objective Space. *IEEE Transactions on Evolutionary Computation* **2020**, *24*, 853–867.
- (S3) Bosman, P.; Thierens, D. The Balance Between Proximity and Diversity in Multiobjective Evolutionary Algorithms. *IEEE Transactions on Evolutionary Computation* **2003**, *7*, 174–188.
- (S4) Blank, J.; Deb, K. Pymoo: Multi-Objective Optimization in Python (version 0.6.0.1). *IEEE Access* **2020**, *8*, 89497–89509.
- (S5) Li, M.; Chen, T.; Yao, X. How to Evaluate Solutions in Pareto-Based Search-Based Software Engineering: A Critical Review and Methodological Guidance. *IEEE Transactions on Software Engineering* **2022**, *48*, 1771–1799.
- (S6) McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection (version 0.5.4). *Journal of Open Source Software* **2018**, *3*, 861.
- (S7) Enamine Screening Collections. <https://enamine.net/compound-collections/screening-collection>, downloaded May 2023.
- (S8) Adasme, M. F.; Linnemann, K. L.; Bolz, S. N.; Kaiser, F.; Salentin, S.; Haupt, V.; Schroeder, M. PLIP 2021: Expanding the Scope of the Protein–Ligand Interaction Profiler to DNA and RNA. *Nucleic Acids Research* **2021**, *49*, W530–W534.

- (S9) Mehta, S.; Goel, M.; Priyakumar, U. D. MO-MEMES: A Method for Accelerating Virtual Screening Using Multi-Objective Bayesian Optimization. *Frontiers in Medicine* **2022**, *9*.