Design-Bench:

Benchmarks for Data-Driven Offline Model-Based Optimization

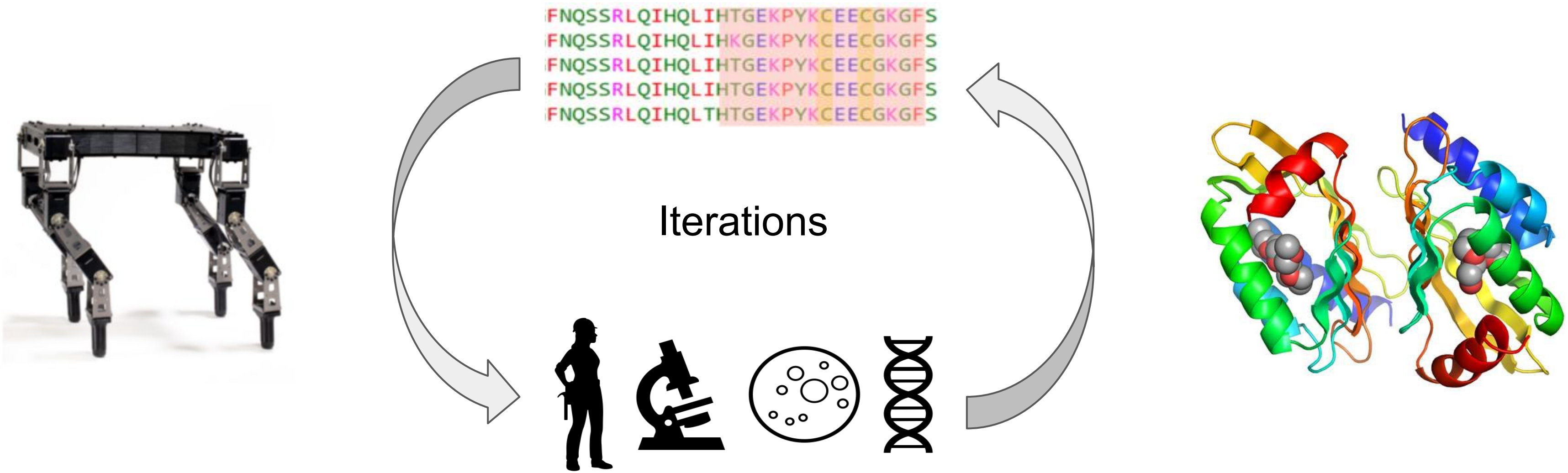
Brandon Trabucco*, Xinyang Geng*, Aviral Kumar, Sergey Levine

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Berkeley
UNIVERSITY OF CALIFORNIA

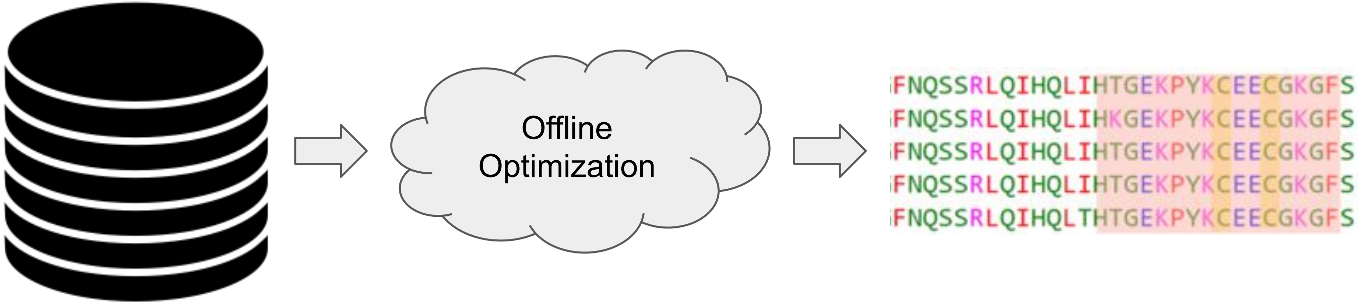
Optimization Based Design

- Iterative process of generating the design and evaluation.
- For problems like protein design, the wet lab evaluations can take **months.**

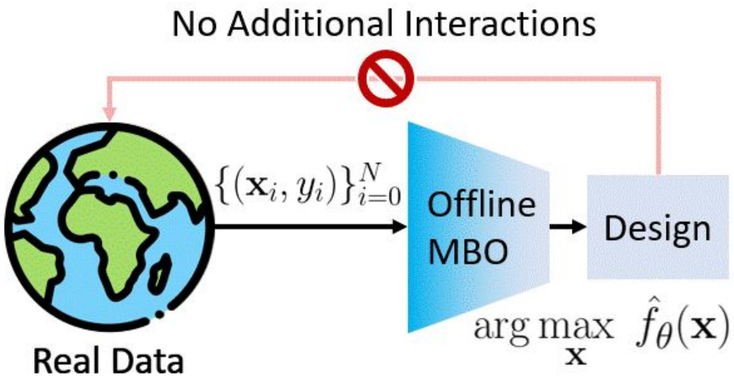


Optimization Based Design

- Iterative process of generating the design and evaluation.
- For problems like protein design, the wet lab evaluations can take **months.**
- **How can we leverage existing data?**

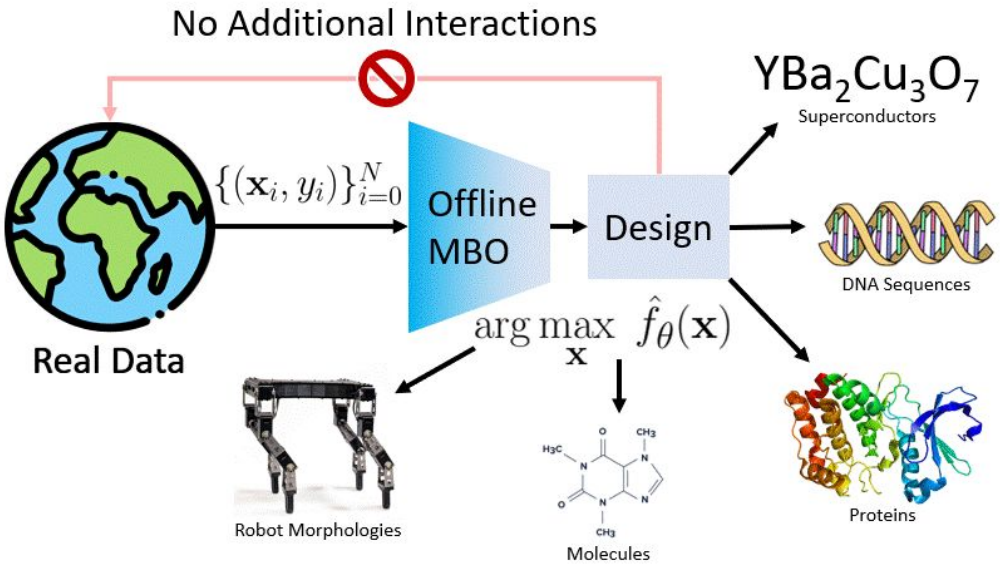


Offline Model-Based Optimization (MBO)



- **Offline MBO**: given a dataset $\mathcal{D} = \{(\mathbf{x}_0, y_0), (\mathbf{x}_1, y_1), \ldots (\mathbf{x}_N, y_N)\}$ maximize $f(\mathbf{x})$ **without** new function evaluations.

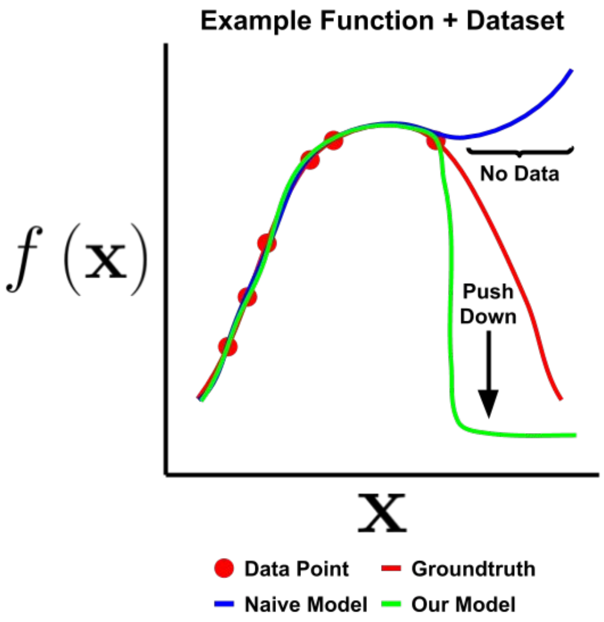
Offline Model-Based Optimization (MBO)



- **Offline MBO**: given a dataset $\mathcal{D} = \{(\mathbf{x}_0, y_0), (\mathbf{x}_1, y_1), \ldots (\mathbf{x}_N, y_N)\}$ maximize $f(\mathbf{x})$ **without** new function evaluations.

What Makes Offline MBO Uniquely Challenging?

- We can only estimate the objective function accurately **within distribution.**
- However, optimization inevitably requires going to **out-of-distribution** regions.
- Successful offline MBO algorithms find high objective points at the edge of data distribution.



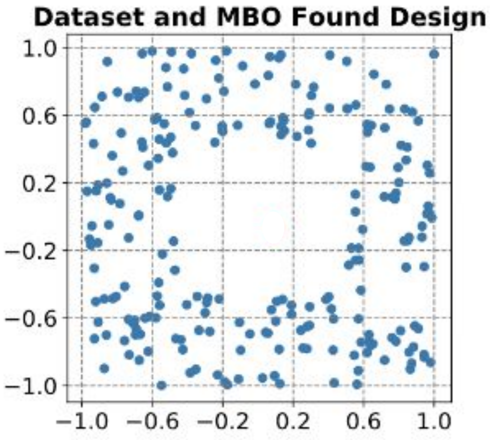
Why Is Offline MBO Tractable?

- What prevents offline MBO from simply copying over the best design in the dataset?
- Many problems have **compositionality**:
    - The dataset contains optimal choices for all component, but not in combination.

Why Is Offline MBO Tractable?

- What prevents offline MBO from simply copying over the best design in the dataset?
- Many problems have **compositionality**:
  - The dataset contains optimal choices for all component, but not in combination.

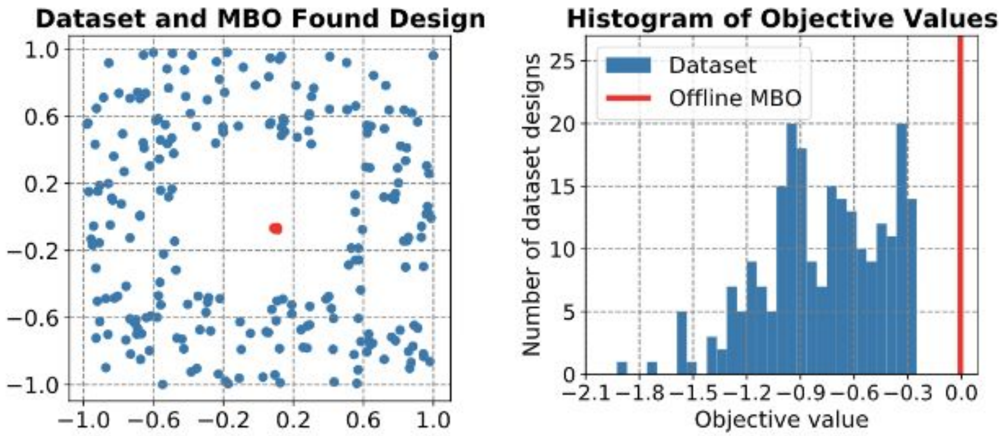
Example: dataset contains optimal choice for x and y, but not together.



$$f(x) = -x^2 - y^2$$

Why Is Offline MBO Tractable?

- What prevents offline MBO from simply copying over the best design in the dataset?
- Many problems have **compositionality**:
  - The dataset contains optimal choices for all component, but not in combination.
- Good algorithms learn to **combine the optimal components.**



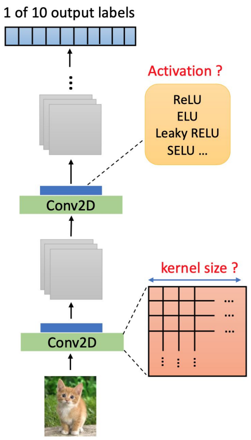
$$f(x) = -x^2 - y^2$$

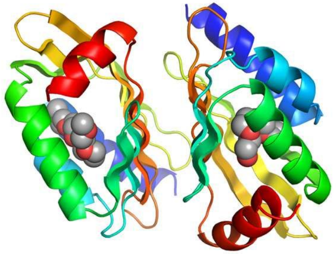
Offline MBO Methods We Benchmark

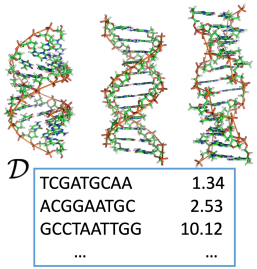
- Search on learned objective
    - Gradient descent, REINFORCE, CMA-ES
- Bayesian optimization
    - BO-qEI
- Generative modeling
    - CbAS, Autofocused CbAS, MINs
- Conservative objectives
    - COMs

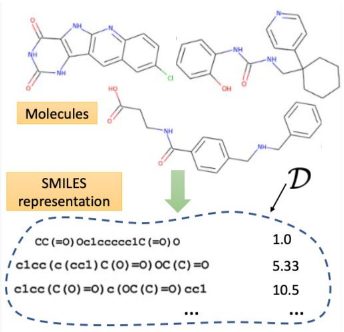
Benchmark Tasks for Offline MBO

- **Realistic**: represent important applications
- Captures the **core challenges** of offline MBO
  - High dimensionality, sensitive objectives and heavy-tailed data
- **Diverse** in problems and domains



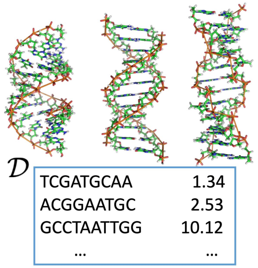


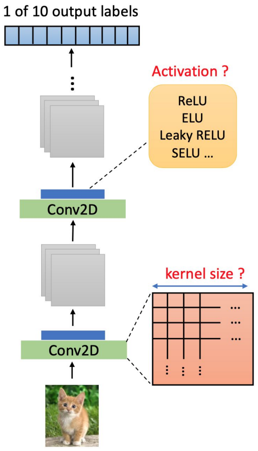


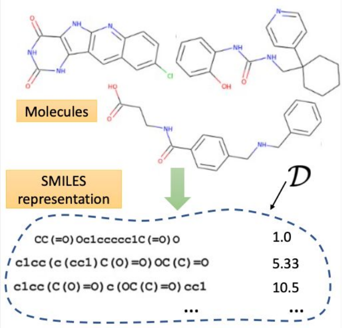


Benchmark Tasks for Offline MBO

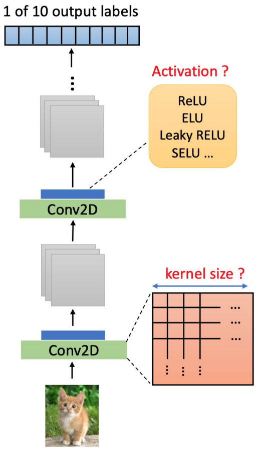
Dataset Name	Size	Dimensions	Categories	Type	Oracle
TF Bind 8	32898	8	4	Discrete	Exact
TF Bind 10	50000	10	4	Discrete	Exact
NAS	1771	64	5	Discrete	Exact
ChEMBL	1093	31	591	Discrete	Random Forest
Superconductor	21263	86	N/A	Continuous	Random Forest
Ant Morphology	25009	60	N/A	Continuous	Exact
D'Kitty Morphology	25009	56	N/A	Continuous	Exact
Hopper Controller	3200	5126	N/A	Continuous	Exact



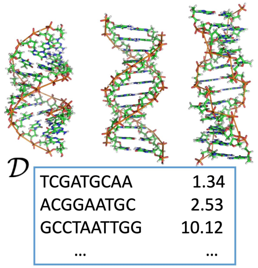




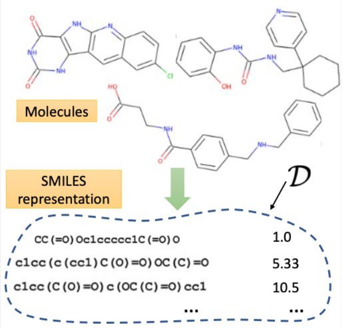
Benchmark Tasks for Offline MBO



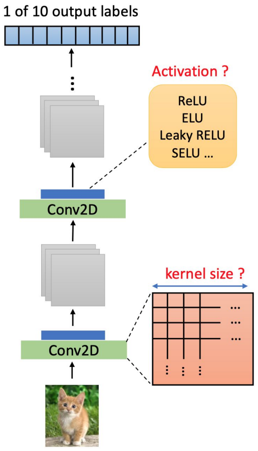
Dataset Name	Size	Dimensions	Categories	Type	Oracle
TF Bind 8	32898	8	4	Discrete	Exact
TF Bind 10	50000	10	4	Discrete	Exact
NAS	1771	64	5	Discrete	Exact
ChEMBL	1093	31	591	Discrete	Random Forest
Superconductor	21263	86	N/A	Continuous	Random Forest
Ant Morphology	25009	60	N/A	Continuous	Exact
D'Kitty Morphology	25009	56	N/A	Continuous	Exact
Hopper Controller	3200	5126	N/A	Continuous	Exact



- 8 Tasks from **Biology**, **Chemistry**, **Robotics** and **Deep Learning**
- Tasks with **High-Dimensional**, Discrete and Continuous input space
- **Sensitive** and **Non-Linear** Objective Functions

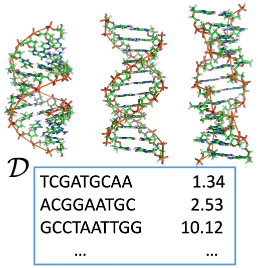


Benchmark Tasks for Offline MBO

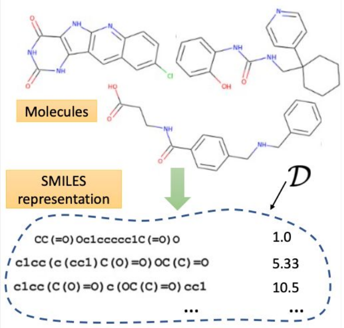


Dataset Name	Size	Dimensions	Categories	Type	Oracle
TF Bind 8	32898	8	4	Discrete	Exact
TF Bind 10	50000	10	4	Discrete	Exact
NAS	1771	64	5	Discrete	Exact
ChEMBL	1093	31	591	Discrete	Random Forest
Superconductor	21263	86	N/A	Continuous	Random Forest
Ant Morphology	25009	60	N/A	Continuous	Exact
D'Kitty Morphology	25009	56	N/A	Continuous	Exact
Hopper Controller	3200	5126	N/A	Continuous	Exact

Significantly High-Dimensional!



- 8 Tasks from **Biology**, **Chemistry**, **Robotics** and **Deep Learning**
- Tasks with **High-Dimensional**, Discrete and Continuous input space
- **Sensitive** and **Non-Linear** Objective Functions



Performance Analysis

	TF Bind 8	TF Bind 10	ChEMBL	NAS	Superconductor	Ant Morph.	DKitty Morph.	Hopper
$\mathcal{D}$ **(best)**	0.439	0.467	0.605	0.436	0.400	0.565	0.884	1.0
Auto. CbAS	0.910 ± 0.044	0.630 ± 0.045	0.249 ± 0.305	0.506 ± 0.074	0.421 ± 0.045	0.882 ± 0.045	0.906 ± 0.006	0.137 ± 0.005
CbAS	0.927 ± 0.051	0.651 ± 0.060	0.473 ± 0.264	0.683 ± 0.079	0.503 ± 0.069	0.876 ± 0.031	0.892 ± 0.008	0.141 ± 0.012
BO-qEI	0.798 ± 0.083	0.652 ± 0.038	0.596 ± 0.226	1.079 ± 0.059	0.402 ± 0.034	0.819 ± 0.000	0.896 ± 0.000	0.550 ± 0.118
CMA-ES	0.953 ± 0.022	0.670 ± 0.023	0.085 ± 0.225	0.985 ± 0.079	0.465 ± 0.024	1.214 ± 0.732	0.724 ± 0.001	0.604 ± 0.215
Grad.	0.977 ± 0.025	0.657 ± 0.039	0.307 ± 0.308	0.433 ± 0.000	0.518 ± 0.024	0.293 ± 0.023	0.874 ± 0.022	1.035 ± 0.482
Grad. Min	0.984 ± 0.012	0.649 ± 0.032	0.653 ± 0.024	0.433 ± 0.000	0.506 ± 0.009	0.479 ± 0.064	0.889 ± 0.011	1.391 ± 0.589
Grad. Mean	0.986 ± 0.012	0.645 ± 0.018	0.652 ± 0.005	0.433 ± 0.000	0.499 ± 0.017	0.445 ± 0.080	0.892 ± 0.011	1.586 ± 0.454
REINFORCE	0.948 ± 0.028	0.663 ± 0.034	0.164 ± 0.285	-1.895 ± 0.000	0.481 ± 0.013	0.266 ± 0.032	0.562 ± 0.196	-0.020 ± 0.067
MINs	0.905 ± 0.052	0.616 ± 0.021	0.000 ± 0.000	0.717 ± 0.046	0.499 ± 0.017	0.445 ± 0.080	0.892 ± 0.011	0.424 ± 0.166
COMs	0.973 ± 0.016	0.730 ± 0.136	0.633 ± 0.000	0.459 ± 0.139	0.439 ± 0.033	0.944 ± 0.016	0.949 ± 0.015	2.056 ± 0.314

- Gradient-based MBO is **surprisingly** effective

Performance Analysis

	TF Bind 8	TF Bind 10	ChEMBL	NAS	Superconductor	Ant Morph.	DKitty Morph.	Hopper
$\mathcal{D}$ (best)	0.439	0.467	0.605	0.436	0.400	0.565	0.884	1.0
Auto. CbAS	0.910 ± 0.044	0.630 ± 0.045	0.249 ± 0.305	0.506 ± 0.074	0.421 ± 0.045	0.882 ± 0.045	0.906 ± 0.006	0.137 ± 0.005
CbAS	0.927 ± 0.051	0.651 ± 0.060	0.473 ± 0.264	0.683 ± 0.079	0.503 ± 0.069	0.876 ± 0.031	0.892 ± 0.008	0.141 ± 0.012
BO-qEI	0.798 ± 0.083	0.652 ± 0.038	0.596 ± 0.226	1.079 ± 0.059	0.402 ± 0.034	0.819 ± 0.000	0.896 ± 0.000	0.550 ± 0.118
CMA-ES	0.953 ± 0.022	0.670 ± 0.023	0.085 ± 0.225	0.985 ± 0.079	0.465 ± 0.024	1.214 ± 0.732	0.724 ± 0.001	0.604 ± 0.215
Grad.	0.977 ± 0.025	0.657 ± 0.039	0.307 ± 0.308	0.433 ± 0.000	0.518 ± 0.024	0.293 ± 0.023	0.874 ± 0.022	1.035 ± 0.482
Grad. Min	0.984 ± 0.012	0.649 ± 0.032	0.653 ± 0.024	0.433 ± 0.000	0.506 ± 0.009	0.479 ± 0.064	0.889 ± 0.011	1.391 ± 0.589
Grad. Mean	0.986 ± 0.012	0.645 ± 0.018	0.652 ± 0.005	0.433 ± 0.000	0.499 ± 0.017	0.445 ± 0.080	0.892 ± 0.011	1.586 ± 0.454
REINFORCE	0.948 ± 0.028	0.663 ± 0.034	0.164 ± 0.285	-1.895 ± 0.000	0.481 ± 0.013	0.266 ± 0.032	0.562 ± 0.196	-0.020 ± 0.067
MINs	0.905 ± 0.052	0.616 ± 0.021	0.000 ± 0.000	0.717 ± 0.046	0.499 ± 0.017	0.445 ± 0.080	0.892 ± 0.011	0.424 ± 0.166
COMs	0.973 ± 0.016	0.730 ± 0.136	0.633 ± 0.000	0.459 ± 0.139	0.439 ± 0.033	0.944 ± 0.016	0.949 ± 0.015	2.056 ± 0.314

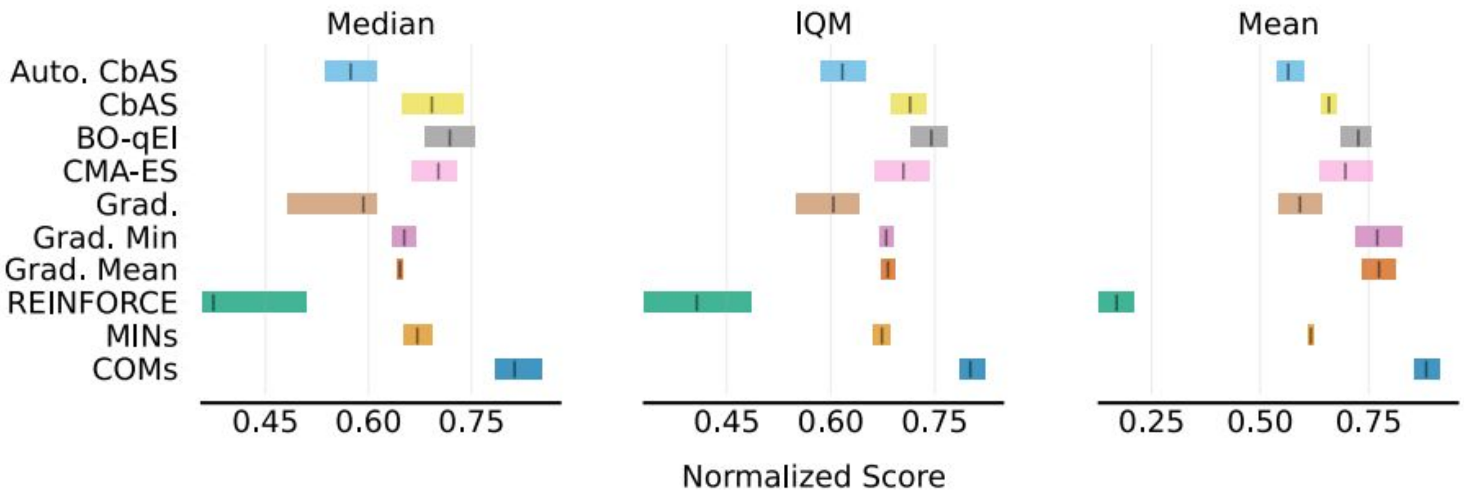
- Gradient-based MBO is **surprisingly** effective
- No **single method** works in all domains (yet)

Performance Analysis

	TF Bind 8	TF Bind 10	ChEMBL	NAS	Superconductor	Ant Morph.	DKitty Morph.	Hopper
$\mathcal{D}$ (best)	0.439	0.467	0.605	0.436	0.400	0.565	0.884	1.0
Auto. CbAS	0.910 ± 0.044	0.630 ± 0.045	0.249 ± 0.305	0.506 ± 0.074	0.421 ± 0.045	0.882 ± 0.045	0.906 ± 0.006	0.137 ± 0.005
CbAS	0.927 ± 0.051	0.651 ± 0.060	0.473 ± 0.264	0.683 ± 0.079	0.503 ± 0.069	0.876 ± 0.031	0.892 ± 0.008	0.141 ± 0.012
BO-qEI	0.798 ± 0.083	0.652 ± 0.038	0.596 ± 0.226	1.079 ± 0.059	0.402 ± 0.034	0.819 ± 0.000	0.896 ± 0.000	0.550 ± 0.118
CMA-ES	0.953 ± 0.022	0.670 ± 0.023	0.085 ± 0.225	0.985 ± 0.079	0.465 ± 0.024	1.214 ± 0.732	0.724 ± 0.001	0.604 ± 0.215
Grad.	0.977 ± 0.025	0.657 ± 0.039	0.307 ± 0.308	0.433 ± 0.000	0.518 ± 0.024	0.293 ± 0.023	0.874 ± 0.022	1.035 ± 0.482
Grad. Min	0.984 ± 0.012	0.649 ± 0.032	0.653 ± 0.024	0.433 ± 0.000	0.506 ± 0.009	0.479 ± 0.064	0.889 ± 0.011	1.391 ± 0.589
Grad. Mean	0.986 ± 0.012	0.645 ± 0.018	0.652 ± 0.005	0.433 ± 0.000	0.499 ± 0.017	0.445 ± 0.080	0.892 ± 0.011	1.586 ± 0.454
REINFORCE	0.948 ± 0.028	0.663 ± 0.034	0.164 ± 0.285	-1.895 ± 0.000	0.481 ± 0.013	0.266 ± 0.032	0.562 ± 0.196	-0.020 ± 0.067
MINs	0.905 ± 0.052	0.616 ± 0.021	0.000 ± 0.000	0.717 ± 0.046	0.499 ± 0.017	0.445 ± 0.080	0.892 ± 0.011	0.424 ± 0.166
COMs	0.973 ± 0.016	0.730 ± 0.136	0.633 ± 0.000	0.459 ± 0.139	0.439 ± 0.033	0.944 ± 0.016	0.949 ± 0.015	2.056 ± 0.314

- Gradient-based MBO is **surprisingly** effective
- No **single method** works in all domains (yet)
- Nearly all methods fail in **high-dimensional** tasks

Performance Analysis: Aggregated



- Gradient-based MBO is **surprisingly** effective
- No **single method** works in all domains (yet)
- Nearly all methods fail in **high-dimensional** tasks

Thank you and looking forward to your algorithms on our benchmark!

Benchmark

Paper