Table S1. Detailed collection list of peer-reviewed papers regarding offline BBO.

Paper title	**Venue**	**Year**
Conditioning by adaptive sampling for robust design.	ICML	2019
Autofocused oracles for model-based design.	NeurIPS	2020
Model inversion networks for model-based optimization.	NeurIPS	2020
Offline model-based optimization via normalized maximum likelihood estimation.	ICLR	2021
Conservative objective models for effective offline model-based optimization.	ICML	2021
RoMA: Robust model adaptation for offline model-based optimization.	NeurIPS	2021
Data-driven offline optimization for architecting hardware accelerators.	ICLR	2022
Design-Bench: Benchmarks for data-driven offline model-based optimization.	ICML	2022
Data-driven offline decision-making via invariant representation learning.	NeurIPS	2022
Bidirectional learning for offline infinite-width model-based optimization.	NeurIPS	2022
Generative pretraining for black-box optimization.	ICML	2023
Diffusion models for black-box optimization.	ICML	2023
Bidirectional learning for offline model-based biological sequence design.	ICML	2023
Parallel-mentoring for offline model-based optimization.	NeurIPS	2023
Importance-aware co-teaching for offline model-based optimization.	NeurIPS	2023
ExPT: Synthetic pretraining for few-shot experimental design.	NeurIPS	2023
Bootstrapped training of score-conditioned generator for offline design of biological sequences.	NeurIPS	2023
Degradation-resistant offline optimization via accumulative risk control.	ECAI	2023
Offline model-based optimization via policy-guided gradient search.	AAAI	2024
Functional graphical models: Structure enables offline data-driven optimization.	AISTATS	2024
Learning surrogates for offline black-box optimization via gradient matching.	ICML	2024
Boosting offline optimizers with surrogate sensitivity.	ICML	2024
Offline multi-objective optimization.	ICML	2024
Incorporating surrogate gradient norm to improve offline optimization techniques.	NeurIPS	2024
Generative adversarial model-based optimization via source critic regularization.	NeurIPS	2024
Guided trajectory generation with diffusion models for offline model-based optimization.	NeurIPS	2024
Exploring validation metrics for offline model-based optimisation with diffusion models.	TMLR	2024
Robust guided diffusion for offline black-box optimization.	TMLR	2024
Offline model-based optimization by learning to rank.	ICLR	2025
SOO-Bench: Benchmarks for evaluating the stability of offline black-box optimization.	ICLR	2025
ParetoFlow: Guided flows in multi-objective optimization.	ICLR	2025

Table S2. Comparison of different model-inner BBO optimizers (BO-qEI, CMA-ES, and EA) in improved UniSO-T. As CMA-ES cannot operate in categorical space, we conduct experimental comparison on continuous tasks from SOO-Bench, where the best and runner-up results on each task are **Blue** and **Violet**. $\mathcal{D}$(best) denotes the best score in the offline dataset. Here all optimizers are allowed for an evaluation budget of 1000 (i.e., the optimizer can access the model's outputs for upmost 1000 times) for fair comparison.

Task	$\mathcal{D}$(best)	BO-qEI	CMA-ES	EA
GTOPX 2	-195.586	**-80.220 ± 11.852**	-138.716 ± 45.523	**-99.778 ± 18.512**
GTOPX 3	-151.190	**-48.493 ± 3.745**	-81.524 ± 19.054	**-69.278 ± 20.009**
GTOPX 4	-215.716	**-80.232 ± 13.582**	-177.559 ± 29.664	**-132.543 ± 38.303**
GTOPX 6	-112.599	**-73.306 ± 12.892**	-101.525 ± 16.986	**-62.024 ± 23.828**
Avg. Rank	/	**1.250 ± 0.433**	3.000 ± 0.000	**1.750 ± 0.433**

Table S3. Ablation studies on different metadata components (name, description, and objective) on various tasks. All experiments are conducted based on improved UniSO-T. The first part includes tasks appear in training dataset, i.e., unconstrained tasks from Design-Bench and SOO-Bench. The remaining two parts contain unseen tasks during training, where we compare them under both zero-shot and few-shot settings. The best and runner-up results on each task are **Blue** and **Violet**. $\mathcal{D}$(best) denotes the best score in the offline dataset.

Task	$\mathcal{D}$(best)	UniSO-T	w/o name	w/o desc.	w/o obj.	w/o metadata
Ant	165.326	**374.665 ± 56.057**	345.369 ± 40.675	362.842 ± 33.008	**363.812 ± 59.380**	358.379 ± 64.211
D'Kitty	199.363	225.752 ± 8.521	**235.715 ± 11.617**	226.479 ± 15.311	**245.135 ± 15.571**	227.169 ± 12.278
Superconductor	74.000	92.200 ± 15.209	85.207 ± 6.261	**99.113 ± 12.742**	**97.610 ± 10.297**	90.871 ± 10.611
TF Bind 8	0.439	0.903 ± 0.041	**0.954 ± 0.025**	0.935 ± 0.041	0.937 ± 0.012	**0.950 ± 0.025**
TF Bind 10	0.005	**0.823 ± 0.542**	**0.696 ± 0.126**	0.664 ± 0.141	0.596 ± 0.148	0.651 ± 0.121
GTOPX 2	-195.586	**-72.848 ± 9.576**	-97.806 ± 40.646	**-63.670 ± 20.381**	-80.789 ± 8.908	-79.864 ± 13.338
GTOPX 3	-151.190	**-45.602 ± 8.433**	-50.788 ± 8.706	**-45.981 ± 4.211**	-50.660 ± 10.713	-48.178 ± 12.638
GTOPX 4	-215.716	-84.271 ± 8.307	-84.962 ± 11.300	-92.163 ± 9.529	**-75.233 ± 5.734**	**-79.887 ± 14.729**
GTOPX 6	-112.599	-47.794 ± 11.943	**-42.181 ± 11.671**	**-45.591 ± 12.310**	-48.050 ± 13.901	-45.764 ± 7.685
RobotPush (zero-shot)	0.102	**3.171 ± 0.984**	2.747 ± 1.455	**3.416 ± 1.455**	2.634 ± 0.953	2.517 ± 1.640
Rover (zero-shot)	-16.148	**-8.888 ± 2.119**	-11.009 ± 0.598	-10.854 ± 0.859	**-9.099 ± 2.202**	-9.089 ± 3.070
LunarLander (zero-shot)	7.038	**31.186 ± 27.971**	30.105 ± 57.577	30.892 ± 54.657	**52.108 ± 47.941**	6.251 ± 53.042
RobotPush (few-shot)	0.102	**7.067 ± 0.169**	7.026 ± 0.219	**7.129 ± 0.486**	6.310 ± 1.677	6.155 ± 1.495
Rover (few-shot)	-16.148	**-8.239 ± 1.270**	-8.850 ± 0.703	**-8.084 ± 0.569**	-8.342 ± 1.573	-10.511 ± 2.070
LunarLander (few-shot)	7.038	**248.573 ± 45.386**	226.726 ± 65.244	**244.252 ± 38.329**	233.169 ± 51.037	233.919 ± 60.467
Avg. Rank	/	**2.333 ± 1.350**	3.600 ± 1.451	**2.467 ± 1.310**	3.067 ± 1.340	3.533 ± 1.087

Table S4. References of response to Reviewer 4Ajw.

No.	Paper title	Venue	Year
1	Position: Leverage foundational models for black-box optimization.	ICML	2024
2	A systematic survey on large language models for algorithm design.	arXiv	2024
3	LLM4AD: A platform for algorithm design with large language model.	arXiv	2024
4	Mathematical discoveries from program search with large language models.	Nature	2024
5	Evolution of heuristics: Towards efficient automatic algorithm design using large language model.	ICML	2024
6	Can GPT-4 perform neural architecture search?	arXiv	2023
7	OmniPred: Language models as universal regressors.	TMLR	2024
8	Understanding LLM embeddings for regression.	TMLR	2025
9	Predicting from strings: Language model embeddings for Bayesian optimization.	arXiv	2024
10	Towards learning universal hyperparameter optimizers with Transformers.	NeurIPS	2022
11	LICO: Large language models for in-context molecular optimization.	ICLR	2025
12	Understanding the behaviour of contrastive loss.	CVPR	2021
13	Advancing Bayesian optimization via learning correlated latent space.	NeurIPS	2023
14	pymoo: Multi-objective optimization in python.	IEEE Access	2020
15	Offline multi-objective optimization.	ICML	2024