Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# RepMOBO: Representation-Guided Multi-Objective Bayesian Optimization for Pareto-Optimal MOF Discovery

**Suihua He**

suihua.he@bit.edu.cn

Beijing Institute of Technology, Zhuhai https://orcid.org/0000-0002-5054-6496

**Zhongyuan Ming**

Beijing Institute of Technology, Zhuhai

**Xing Liu**

University of Science and Technology of China

**Yihe Zhang**

Beijing Institute of Technology, Zhuhai

**Hao Ren**

State Key Laboratory of NBC Protection for Civilian

**Yudie Liu**

Beijing Institute of Technology, Zhuhai

**Xinbo Wang**

State Key Laboratory of NBC Protection for Civilian

**Yue Zhao**

State Key Laboratory of NBC Protection for Civilian

**Article**

**Keywords:** Metal–organic frameworks, Multi-objective Bayesian optimization, Active learning, Representation learning, Toxic gas adsorption, Materials discovery

**Posted Date:** November 3rd, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11033745/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** There is **NO** Competing Interest.

# Abstract

Screening metal–organic frameworks (MOFs) across chemical space by simulation or experiment is costly, while trade-offs among adsorption objectives hinder Pareto-optimal discovery. Material representation learning offers a promising route by encoding crystal structures into compact feature spaces. However, prior work has been confined to offline prediction, leaving unclear how representation choice shapes search behaviour and Pareto discovery in multi-objective Bayesian optimization. Here we propose RepMOBO, a representation-guided multi-objective Bayesian optimization framework, using multi-vapour MOF adsorption to examine how handcrafted physicochemical descriptors and pretrained structural embeddings shape low-budget Pareto discovery. At a 10% query budget, the hybrid-representation configuration yielded 4.5-fold more configuration-specific strict Pareto-front discoveries than the handcrafted-descriptor configuration. RepMOBO-prioritized candidates guided experimental screening, yielding synthesizable MOFs with target-dependent, state-of-the-art adsorption performance. These results reveal a prediction–discovery mismatch: strong offline prediction need not translate into efficient low-budget Pareto discovery. RepMOBO therefore elevates representation from static encoding to an active search variable and motivates transferable representation–surrogate–policy co-design for costly multi-objective discovery in numerically representable candidate spaces.


# Introduction

The release of toxic vapours poses immediate threats to human health and industrial safety, driving an urgent need for high-performance adsorbent materials capable of effective capture[1]. Metal–organic frameworks (MOFs), assembled from metal nodes and organic linkers, have emerged as important candidates for toxic-vapour capture owing to their high surface areas, tunable pore environments and designable functional sites[2,3]. For practical protection scenarios, adsorbent design needs to move beyond optimizing performance against a single threat and instead address multiple hazardous vapours simultaneously. However, different vapours vary substantially in their molecular physicochemical properties and host-guest interactions, making it unlikely for a single MOF to achieve optimal adsorption performance across all targets. Multi-vapour adsorbent screening therefore constitutes an intrinsically multi-objective materials-discovery problem, where the aim is no longer to identify a single best material, but rather to discover Pareto-optimal MOF candidates and the corresponding Pareto front[4–6].

Across large MOF spaces, exhaustively evaluating multiple adsorption properties by grand canonical Monte Carlo (GCMC) simulation or experiment is prohibitively expensive, making Pareto-optimal MOFs difficult to identify through exhaustive screening. To reduce expensive property evaluations while efficiently locating high-value candidates, active learning (AL) and Bayesian optimization (BO) have emerged as key frameworks for sample-efficient materials discovery and closed-loop optimization, applied across electrolytes, alloys, phase-change materials, metamaterials and perovskites[7–13]. In MOF gas adsorption and separation, AL and BO use surrogate-guided sequential querying to prioritize promising candidates, thereby cutting labelling costs compared with full-library evaluation[4,6,14]. Their

effectiveness depends on the ability of the surrogate model to learn generalizable structure–property relationships from a limited number of high-fidelity labels and to provide predictions and uncertainty estimates for unevaluated candidates[15,16]. For highly tunable MOFs, representation learning encodes complex crystal structures into computable and comparable representations that retain adsorption-relevant information[17,18]. In recent years, MOF representations have progressed from handcrafted physicochemical descriptors of pore geometry, composition and topology to crystal graphs, energy-based representations, and multimodal or three-dimensional pretrained models such as MOFTransformer and Uni-MOF, enabling increasingly complex structural information to be captured[19–21].

However, most existing studies of MOF representation learning still evaluate representations primarily through offline property prediction on fixed labelled datasets, using metrics such as mean absolute error (MAE), root mean square error (RMSE) and the coefficient of determination ($R^2$) to quantify average predictive performance, with much less attention to their practical utility for discovering rare, high-value candidates under limited budgets[22,23]. More broadly, studies of AI-guided discovery in areas including perovskites and drug discovery have shown that task-relevant structural or phenotypic representations, high-fidelity automated data and experimental feedback can improve candidate ranking and discovery efficiency in low-data regimes[11,24,25]. For MOFs, representations lacking task-relevant information can weaken the ability of single-objective BO to identify high-performing candidates[25], while descriptor choice and the coverage of labelled samples in descriptor and target spaces can also affect surrogate generalization and labelling efficiency[14]. For Gaussian-process-based BO, the representation further defines distances, similarities and local neighbourhoods among candidates, and can therefore alter posterior prediction, uncertainty estimation and acquisition priorities. Yet existing studies remain focused largely on static property prediction or single-objective optimization. It remains unclear whether strong offline predictive performance translates into effective Pareto discovery under limited budgets, or how representation, GP kernel parameterization and selection policy jointly shape tail enrichment, search trajectories and strict Pareto-front recovery.

To address this gap, we introduce RepMOBO, a representation-guided multi-objective Bayesian optimization (MOBO) framework, and establish an integrated workflow spanning representation benchmarking, low-budget Pareto discovery and experimental validation (Fig. 1). Specifically, a quality-controlled MOF benchmark library was constructed with complete GCMC adsorption labels for CEES, DMMP and EtCl as a testbed for multi-vapour, multi-objective discovery. We first compared the predictive and discovery utility of handcrafted physicochemical descriptors (Manual) with pretrained structural embeddings, and then, under a unified GP-based MOBO protocol, examined how representation space, kernel parameterization and selection policy jointly influence low-budget Pareto discovery. Our results demonstrate that representation choice shapes not only offline predictive performance but also the interaction between GP surrogate parameterization and selection policy during sequential discovery. In particular, the low-dimensional Manual descriptor space and the Hybrid space incorporating pretrained structural embeddings exhibit complementary Pareto-discovery behaviours, ultimately leading to synthesizable MOFs with target-dependent adsorption advantages. Overall, RepMOBO treats

representation not merely as a static encoding for offline prediction, but as a search variable in sequential discovery. Representation quality should therefore be assessed not only by predictive accuracy, but also by how it structures the candidate space, alters query priorities and ultimately shapes Pareto discovery.

## Results

# 2.1 Multi-vapour adsorption defines a sparse and heterogeneous Pareto landscape

To quantitatively resolve the compatibility and potential trade-offs among multiple vapour-adsorption objectives, we constructed a fixed, label-complete MOF oracle library comprising 7,532 MOFs with grand canonical Monte Carlo (GCMC) adsorption labels for 2-chloroethyl ethyl sulfide (CEES), dimethyl methylphosphonate (DMMP) and ethyl chloride (EtCl) (Supplementary Notes 1 and 2). These three vapours represent a sulfur-mustard simulant, a highly polar organophosphorus simulant and a small, highly volatile chlorinated hydrocarbon, respectively, and differ substantially in molecular size, polarity and volatility[2,26–28]. We first asked whether the high-adsorption tails for individual vapours were dominated by the same materials, and thus whether multi-vapour discovery could be approximated by single-objective ranking. As shown in Fig. 2a, the three single-vapour Top-1% sets, each containing 76 MOFs, were nearly disjoint. The CEES–DMMP and DMMP–EtCl sets shared only 3 and 7 candidates, respectively, whereas both the CEES–EtCl overlap and the three-way intersection contained no candidates. Single-vapour ranking therefore predominantly enriched target-specific candidates but failed to capture the full three-objective non-dominated set, motivating multi-objective search rather than independent single-objective screening.

The full-library adsorption distributions further showed that high-performing candidates were concentrated in rare tails for each objective (Fig. 2b). Adsorption uptakes for CEES, DMMP and EtCl were all strongly right-skewed, with most MOFs occupying the low-to-intermediate adsorption regime and Top-1% cutoffs of 9.75, 6.39 and 4.01 mol $kg^{-1}$, respectively. To determine whether the three vapours shared a common material ranking, we calculated pairwise Spearman rank correlations, where ρ approaching 1 indicates increasingly concordant rankings and ρ near 0 indicates little monotonic association (Fig. 2c). Only weak-to-moderate ranking consistency was observed across the three objectives. CEES–DMMP showed the strongest correlation (ρ = 0.47), followed by DMMP–EtCl (ρ = 0.29), whereas CEES–EtCl exhibited almost no monotonic relationship (ρ = 0.02). These results indicate that high-adsorption candidates are both rare and target specific, and that the three adsorption responses cannot be represented by a common ranking axis. Neither single-objective ranking nor scalarization based on a single prespecified weighting scheme can preserve the full three-objective trade-off structure.

To characterize these three-objective trade-offs explicitly, we mapped the full MOF library into the three-dimensional CEES–DMMP–EtCl adsorption space and defined the strict Pareto front and Pareto layers

using the full-library oracle (Supplementary Note 3)[29]. Among the 7,532 MOFs, the strict oracle Pareto front contained only 15 materials (0.20%); even when expanded to the first three layers obtained by iterative non-dominated sorting, the Pareto-relevant set comprised only 113 MOFs (1.50%) (Fig. 2d). The strict front and Top-3 Pareto set were subsequently used as full-library references for strict-front recovery and Top-3 Pareto-layer capture, respectively, allowing the discovery performance of different low-budget search configurations to be compared against a common ground truth. These Pareto-relevant candidates occupied the sparse non-dominated boundary at the periphery of the material distribution rather than its densely populated interior. Two-dimensional projections further showed that strict-front and near-front candidates populated different boundary regions across the three objective planes, reflecting distinct performance trade-offs (Fig. 2e). Some three-dimensional strict-front candidates did not appear in the upper-right corner of every two-dimensional projection because their non-dominated status depended on the third objective omitted from that projection. Overall, the strict oracle front was markedly asymmetric across objectives: most front members were concentrated in the high-CEES and CEES–DMMP trade-off regions, whereas candidates extending the front towards high EtCl adsorption were comparatively scarce. This pattern reflects an intrinsic imbalance in the joint three-objective response space of the current candidate library.

Finally, we examined whether these Pareto-relevant candidates could be identified directly from simple structural or compositional features. Comparison of the normalized distributions of 14 handcrafted physicochemical descriptors revealed shifts for strict-front and near-front MOFs relative to the remaining materials, most prominently for Void Fraction and VSA (Fig. 2f and Supplementary Note 1). However, Pareto-relevant and background materials still overlapped extensively in pore dimensions, framework density, elemental composition, electronegativity statistics and local metal coordination. None of the individual descriptors examined, nor any simple threshold on them, clearly separated Pareto-relevant candidates from the remainder of the library. Pareto relevance therefore does not arise from a single isolated feature, but is distributed across multiple structural and compositional dimensions. Taken together, this multi-vapour adsorption landscape combines target-specific long tails, weak cross-objective ranking consistency, a sparse and asymmetric non-dominated boundary, and distributed structural signals, providing a controlled benchmark for testing how different representations organize and search these high-value regions.

## 2.2 Representation benchmarking reveals a prediction–discovery mismatch

Given that Pareto-relevant candidates are rare, distributed across multiple objective trade-off regions and not clearly separable by any single descriptor examined above, we next asked whether richer material representations could improve both average adsorption prediction and identification of the rare high-performance tail. In Fig. 2f, the 14 handcrafted physicochemical descriptors are hereafter referred to as the Manual representation and serve as a low-dimensional, physically interpretable baseline. Five different representation spaces were compared using the same dataset: Manual (14D), MOFTransformer

(768D), Uni-MOF (512D), and two hybrid representations created by concatenating Manual with the corresponding pretrained embeddings, Manual + MOFTransformer (782D) and Manual + Uni-MOF (526D). All representations were evaluated using identical data splits and model-selection procedures, with adsorption prediction assessed using nonlinear LightGBM and linear Ridge regression[30,31]. In addition to held-out test-set $R^2$, we compared Spearman rank correlation, Top-1% hit rate and cumulative recall to distinguish average regression performance, global ranking consistency and the ability to identify the high-performance tail (Supplementary Notes 4 and 5).

Under nonlinear LightGBM, the Manual representation provided a strong predictive baseline (Fig. 3a), achieving the highest test-set $R^2$ for DMMP and EtCl (0.662 and 0.667, respectively), while its CEES $R^2$ of 0.715 was only slightly below the 0.723 obtained with Manual + MOFTransformer. Neither MOFTransformer nor Uni-MOF embeddings alone outperformed Manual, and the hybrid representations did not yield consistent gains across objectives. Among them, Manual + Uni-MOF maintained the most balanced predictive performance, indicating that the contribution of pretrained structural information was strongly target dependent. The ranking of representations also changed with the downstream model. Under linear Ridge regression, Manual performance decreased substantially, whereas the hybrid representations became more competitive: Manual + MOFTransformer performed best for CEES ($R^2$ = 0.544), while Manual + Uni-MOF was best for DMMP and EtCl ($R^2$ = 0.451 and 0.431, respectively). Thus, the relative utility of a representation depends on how effectively its information can be decoded by the downstream model, rather than reflecting a fixed, model-independent hierarchy among representation spaces. Materials discovery, however, depends not only on overall regression accuracy but also on ranking the full library and identifying candidates in the extreme high-performance tail. Based on LightGBM predictions, Manual achieved the highest Top-1% hit rate for all three objectives, reaching 0.625 for CEES, 0.688 for DMMP and 0.375 for EtCl (Fig. 3b), followed closely by Manual + Uni-MOF, which was the most balanced among hybrids. Global Spearman correlations showed a similar, although not identical, ordering, with Manual performing best overall and Manual + Uni-MOF following closely (Fig. 3c). Notably, EtCl exhibited the highest Spearman correlation under most representations but the lowest Top-1% hit rate among the three objectives. This contrast shows that global ranking consistency and extreme-tail recovery capture distinct aspects of model performance: a high overall rank correlation does not guarantee effective identification of rare, strongly adsorbing candidates. To visualize this distinction, a prediction–discovery alignment map was constructed (Fig. 3d) by using an equally weighted combination of normalized $R^2$ and Spearman correlation to represent overall modelling quality and the Top-1% hit rate to quantify tail-screening utility. Manual occupied the high-performing region for both dimensions across all objectives, while Manual + Uni-MOF was the most balanced among the hybrid representations. In contrast, Manual + MOFTransformer slightly improved CEES $R^2$, but this gain did not translate into improved Top-1% recovery.

Because Manual + Uni-MOF showed the most balanced prediction-discovery profile among the hybrid representations, we retained it for subsequent MOBO analysis. We therefore used it as a representative hybrid case to examine whether aggregate metrics concealed systematic prediction errors across

different regions of the adsorption distributions (Fig. 3e; parity plots for all representations are shown in Supplementary Fig. 1). For all three vapours, predictions were broadly distributed along the identity line, but the predicted ranges were noticeably narrower than the real adsorption ranges. Specifically, low-adsorption candidates were slightly overestimated, whereas for higher-adsorption candidates, an increasing fraction of candidates fell below the identity line and outside the ± 15% relative-error band, revealing a clear regression-to-the-mean behaviour. Thus, even a representation that balances overall prediction, ranking and Top-1% hit rate can still systematically underestimate candidates in the rare high-adsorption regime. Nevertheless, cumulative-recall curves derived from the same static prediction rankings showed that all representations outperformed the random baseline across most of the screening range, indicating that they retained useful relative-ranking information for candidate enrichment (Fig. 3f). The relative advantage of each representation varied with both objective and screening stage: Manual + MOFTransformer led during early CEES screening, whereas Manual + Uni-MOF achieved higher recall during the early-to-intermediate stages for DMMP and EtCl.

Taken together, the results across average prediction, global ranking and high-performance-tail screening show that no single representation consistently dominates across all objectives, downstream models and screening stages. Manual provides the strongest overall low-dimensional baseline in the current benchmark while retaining direct physical interpretability. Manual + Uni-MOF does not surpass Manual on every offline metric, but offers the most balanced prediction–discovery profile among the hybrid representations and introduces pretrained structural information not explicitly encoded by the Manual descriptors. Considering this complementarity, we retained Manual as the physical baseline and selected Manual + Uni-MOF as the structure-enhanced representation for subsequent controlled MOBO comparisons. This approach enables the assessment of whether distinct offline characteristics translate into different low-budget Pareto-discovery outcomes.

# 2.3 Representations steer low-budget Pareto discovery

# 2.3.1 Representation–kernel coupling defines low-budget discovery performance

Building on the representation benchmark above, we retained the 14-dimensional Manual descriptors as a low-dimensional, physically interpretable baseline and selected the 526-dimensional Manual + Uni-MOF representation as a hybrid representation incorporating pretrained structural information. We used these two representations to examine the joint effects of representation space, selection rule and GP kernel parameterization on low-budget Pareto discovery[32]. To accommodate GP-based MOBO under limited labelled data, we further compressed the Hybrid representation by principal component analysis (PCA) into P80–P95 latent spaces (retaining 80–95% cumulative variance). These latent spaces, together with the Manual space, were used as GP inputs in a factorial comparison of qEHVI versus Greedy selection and ARD on versus off under matched initial designs, batch settings and 10% cumulative query budget (Methods 4.6–4.8 and Supplementary Notes 6–8). qEHVI evaluates expected hypervolume improvement

under the joint GP posterior, whereas Greedy selects candidates solely according to a posterior-mean composite score across the three objectives; ARD fits an independent length scale for each input dimension when on, and a shared scale within each objective when off[33,34].

At the 10% budget, strict-front recovery and Top-3 Pareto-layer capture revealed a clear representation–ARD crossover (Fig. 4a,b). In Manual space, ARD on outperformed ARD off under both selection policies; in all P80–P95 hybrid latent spaces, ARD off was superior, with qEHVI–P85 (ARD off) performing best overall, achieving a strict-front recovery of 0.587 ± 0.056 and a Top-3 capture of 0.559 ± 0.026. In comparison, P95 performed consistently poorly across all configurations, indicating that retaining more cumulative variance and increasing dimensionality did not improve low-budget discovery. Additionally, matched factor contrasts (ARD off minus ARD on; qEHVI minus Greedy) were calculated, where positive values therefore indicate an advantage for ARD off or qEHVI (Fig. 4c; Top-3 contrasts are shown in Supplementary Fig. 3). This analysis confirmed the reversal of disabling ARD, which reduced recovery in the Manual descriptor space under both selection rules, but improved recovery across all hybrid latent spaces; under ARD off, qEHVI outperformed Greedy in Manual, P80, P85 and P90, whereas the difference was negligible in P95. This pattern reflects input dimensionality: the compact Manual space benefits from dimension-specific length scales that capture distinct correlation scales, whereas higher-dimensional hybrid spaces incur a larger hyperparameter-estimation burden, favouring a shared length scale[32,33]. Accordingly, the Manual paired most effectively with ARD on, whereas the hybrid latent spaces paired most effectively with ARD off, under which qEHVI generally outperformed Greedy.

Objective-specific enrichment further revealed distinct but complementary behaviours for Manual descriptor space and P85 hybrid latent space (Fig. 4d). Top-1% recovery measures the fraction of full-library top performers within the cumulative queried set. Correspondingly, qEHVI generally outperformed Greedy for Manual, P80 and P85 under both ARD settings, but this advantage faded for P90 and P95. The two leading pairings were qEHVI–Manual (ARD on) and qEHVI–P85 (ARD off), which achieved mean Top-1% recoveries of 54% and 56%, respectively, placing both among the leading configurations across the 20 tested settings. qEHVI–Manual (ARD on) excelled for DMMP, reaching the highest recovery (87%) observed across all configurations. In contrast, qEHVI–P85 (ARD off) maintained high DMMP recovery but broader CEES and EtCl coverage, resulting in a more balanced cross-objective enrichment profile. Thus, both configurations serve complementary roles: Manual favours DMMP-led enrichment, whereas P85 balances all three objectives and recovers a broader set of Pareto-relevant candidates.

Material-level analysis further showed that configurations with similar aggregate performance did not necessarily recover the same Pareto-optimal materials, indicating that representation, kernel parameterization and selection policy also shape the composition of the discovered front (Fig. 4e). Integrating normalized hypervolume[35], strict-front recovery, objective-specific enrichment and material-level coverage, we identified two representative configurations with complementary discovery roles (Fig. 4f): qEHVI–Manual (ARD on) achieved the highest normalized hypervolume, whereas qEHVI–P85 (ARD off) achieved the highest strict-front recovery and broader material-level coverage. We therefore retained their matched Greedy counterparts under the same representation and ARD settings, yielding

two matched qEHVI–Greedy comparison pairs: Manual (ARD on) and P85 (ARD off), for subsequent analyses of budget-dependent trajectories and search behaviour. The four resulting configurations are hereafter denoted qE–Manual, Gr–Manual, qE–Hybrid and Gr–Hybrid, with each label referring to a complete input-space–kernel–policy configuration.

## 2.3.2 Complementary configurations produce distinct Pareto discovery trajectories

Across query budgets, the selected configurations exhibited distinct discovery trajectories (Fig. 5a–c). qE–Manual most rapidly improved normalized hypervolume, reaching 0.997 at the 10% budget, while the Manual configurations also enriched high-performing candidates earlier in the search. By contrast, qE–Hybrid achieved the highest strict-front recovery at the primary 10% operating point (58.7% ± 5.6%) with relatively small seed-to-seed variation, consistent with uncertainty-aware acquisition facilitating exploration of Pareto-relevant regions under limited labelled data[32,34]. This advantage was not universal: Gr–Manual remained competitive across much of the early-to-intermediate budget range[36], and Gr–Hybrid eventually surpassed qE–Hybrid as the budget increased. Random sampling consistently lagged behind the model-guided configurations. Together, these trajectories show that the relative advantage of qEHVI over posterior-mean selection depends on both representation space and query budget, with qE–Manual favouring rapid Pareto approximation and qE–Hybrid providing stronger exact-front recovery at the principal low-budget operating point.

At 10% query budget, the preceding trajectories showed that qE–Manual and qE–Hybrid achieved nearly identical normalized hypervolume (0.997 and 0.994, respectively) and mean Top-1% recovery (~ 54% and ~ 56%), yet differed substantially in strict-front recovery (40.0% versus 58.7%). We therefore examined their selection distributions and material-level recovery compositions in greater detail (Fig. 5d–f). When the cumulative queries from all five seeds were projected onto the common PC1–PC2 plane of the fixed P85 hybrid latent space, the two configurations showed broadly overlapping regions, with qE–Hybrid exhibiting only slightly broader two-dimensional coverage at later stages. A clearer distinction emerged in the timing of front-member discovery: qE–Manual concentrated Top-3 hits at early budgets, whereas qE–Hybrid continued to acquire additional strict-front members between 8% and 10%, with individual runs recovering 8–10 members compared with 4–8 for qE–Manual (Fig. 5d). Shared–specific decomposition at the 10% budget showed an average of 5.2 shared strict-front members, whereas Hybrid-only recovery (3.6) was 4.5-fold higher than Manual-only recovery (0.8). This asymmetric incremental coverage persisted at 15% and 20% budgets, where Hybrid-only counts were 3.0- and 6.5-fold higher, respectively (Fig. 5e). Thus, the additional exact-front recovery of qE–Hybrid primarily arose from expanding beyond a shared recovery core, rather than replacement of commonly discovered materials by a different set. Similar near-front capture and two-dimensional selection coverage can therefore coexist with markedly different strict-front recovery.

To identify the trade-offs underlying these incremental discoveries, we classified the 15 strict-front members by recovery counts across the five seeds into six shared-stable, four Hybrid-enriched, one Manual-enriched and four rare/missed members (Fig. 5f). Members recovered by at least three seeds under both configurations were deemed shared-stable; and those recovered by neither as rare/missed; all remaining members were assigned as Hybrid-enriched or Manual-enriched according to the configuration that recovered them more frequently. Figure 5f shows material identities (left), three-objective performance using true uptakes (centre), and per-configuration recovery counts across the five seeds (right). The recovered set of materials included both single objective extremes and members with competitive trade-offs across multiple objectives, indicating that both configurations could access distinct classes of high-value solutions. Notably, every strict-front member recovered by qE–Manual was also recovered by qE–Hybrid, while qE–Hybrid additionally recovered three front members that qE–Manual missed entirely and increased the recovery frequency of another. The Hybrid advantage therefore stemmed from extending beyond the shared recovery core to Pareto-optimal materials that Manual failed to access, yielding clear incremental front coverage.

Taken together, Manual (ARD on) and P85 (ARD off) exhibit complementary low-budget discovery modes. The former supports faster hypervolume growth and earlier enrichment of high-performing candidates, while allowing Greedy selection to remain competitive across a broad early-to-intermediate budget range. The latter enables qEHVI to achieve higher and more stable strict-front recovery at 10% budget, while extending front coverage beyond the shared recovery core; Gr–Hybrid then progressively catches up and becomes advantageous as the budget increases. Representation choice therefore affects not only overall search performance, but also the discovery priorities and the relative merits of GP kernel parameterization and selection policy. RepMOBO accordingly does not seek a single globally optimal strategy; instead, it identifies representation–kernel–policy configurations with complementary discovery capabilities and consolidates the high-value candidates jointly enriched by these configurations into an experimentally oriented candidate pool, providing the computational basis for subsequent material prioritization and experimental translation.

## 2.4 RepMOBO-prioritized candidates translate into experimentally validated MOF adsorbents

Having established that representation choice reshapes low-budget Pareto discovery, we integrated the candidates across ten qE–Manual and qE–Hybrid runs and progressively filtered them by cross-run stability, Pareto relevance, and overall adsorption performance to obtain an experimental shortlist from the 7,532 GCMC-labelled MOFs (Fig. 6a). At the 10% query budget, the ten runs collectively queried 2,385 unique MOFs, and retaining materials appearing in at least six runs reduced the pool to 428 MOFs. Further intersection of the first three Pareto layers with Top-1% composite-score candidates yielded 30 Pareto-enriched candidates, of which eight ranked among the top ten materials by composite score and constituted the high-priority shortlist.

From these eight high-priority candidates, five were selected for synthesis—OLOGAY, KAZXEQ, OYEWEW, QAJJOD01 and IXEJOM, identified by their CSD refcodes (Fig. 6b). Four of them (OLOGAY, OYEWEW, QAJJOD01 and IXEJOM) yielded activated samples consistent with the target structures, corresponding to NJU-Bai1[37], MOF-515[38], UiO-67-Me[39] and PCN-700-$(CF_3)_2$[40], respectively, and were characterized by PXRD, $N_2$ sorption and BET surface-area analysis, FTIR, TGA and SEM (Figure S7-S12 in Supplementary Notes9). KAZXEQ, by contrast, did not yield a product consistent with the target structure under the tested conditions. These results distinguish computational prioritization from experimental realizability and establish four materials for subsequent vapour-adsorption measurements.

Single-component CEES, DMMP and EtCl adsorption isotherms were measured for the four successfully synthesized MOFs with uptakes compared at $P/P_0 = 0.1$ (CEES and DMMP) and $V_i/V = 0.1$ (EtCl), as well as their adsorption capacities at the highest tested condition (Fig. 6c,d). As shown in Fig. 6c, the performance ranking was target-dependent: for CEES and DMMP, the order was UiO-67-Me > MOF-515 > PCN-700-$(CF_3)_2$ > NJU-Bai1; whereas for EtCl, MOF-515 surpassed UiO-67-Me to rank first, while the remaining ordering was unchanged. At the low-exposure test point, UiO-67-Me reached uptakes of 585 and 493 mg g-1 for CEES and DMMP, respectively, while MOF-515 reached 551 mg $g^{-1}$ for EtCl. At the highest tested condition, the leading material for each target remained unchanged, with the corresponding capacities rising to 805, 834 and 695 mg $g^{-1}$, respectively (Fig. 6d). These results demonstrated that RepMOBO-prioritized candidates can translate into experimentally realizable MOFs with complementary, target-dependent adsorption advantages, validating the practical value of representation-guided multi-objective BO for narrowing large MOF spaces and guiding material selection.

# Discussion

Most existing material-representation benchmarks rely on offline prediction metrics such as $R^2$, RMSE, and rank correlation. This work, however, demonstrates that these metrics do not fully capture the practical utility of a representation for materials discovery under limited evaluation budgets. High-value candidates often occupy sparse, target-specific tails of the data distribution, whereas global regression and ranking metrics are dominated more readily by the much larger population of low- to intermediate-performing samples, thereby attenuating local prediction errors in the tails. In the present multi-vapour benchmark, highly adsorbing candidates for CEES, DMMP and EtCl were all rare, and the corresponding high-performance regions differed across objectives. Consequently, a higher $R^2$ or Spearman correlation did not necessarily translate into stronger tail recovery. Representation evaluation for materials discovery should therefore extend beyond aggregate offline metrics to explicitly assess the identification of rare, high-value candidates under limited budgets, including tail recovery and downstream Pareto-discovery performance.

Motivated by this prediction–discovery mismatch, we developed RepMOBO as a representation-guided multi-objective BO framework. Its central premise is that a representation does more than provide inputs

to a surrogate model: it defines distances, neighbourhoods and the overall geometry of the candidate space, and thereby interacts with GP kernel parameterization and candidate-selection policy to shape both the search process and its discovery outcomes. Systematic comparisons showed that the relative utility of surrogate parameterization and selection policy depends on the representation under a given query budget, precluding a single universally optimal configuration independent of the input space. RepMOBO thus reframes the task from selecting an individual model or acquisition strategy to jointly designing the representation, surrogate, and search policy, with low-budget downstream Pareto-discovery outcomes serving as the ultimate evaluation criterion. This framework thereby connects representation learning with low-budget sequential discovery and lays a foundation for future prospective validation and generalization.

More broadly, these results show that representation space determines not only what information is available to the surrogate, but also which surrogate parameterization and search behaviour are most effective under limited budgets. The representation–ARD crossover and the lack of consistent gains from higher-dimensional Hybrid spaces indicate that retaining more structural variation does not necessarily preserve more information relevant to Pareto discovery. Instead, different representations support distinct discovery functions: Manual-based configurations favour rapid improvement of Pareto approximation and early enrichment of high-performing candidates, whereas the Hybrid configuration provides broader and more stable recovery of exact Pareto-front members. The continued increase in strict-front recovery after hypervolume saturation further shows that objective-space coverage and recovery of individual Pareto-optimal materials are not equivalent. These findings motivate evaluating and designing representation–surrogate–policy configurations according to their downstream discovery roles, rather than seeking a single universally optimal representation or optimization setting.

Although this study is limited to retrospective evaluation on a fixed candidate library with reference points calibrated from full-range responses, RepMOBO reframes representation as a search variable that shapes the candidate space and Pareto discovery, offering a foundation for representation-guided sequential discovery under more realistic constraints. Future investigation could derive reference point from domain constraints, pilot observations or currently observed responses, while integrating online computation, synthesizability constraints and experimental feedback directly into the sequential loop to examine whether these design principles remain robust in prospective closed-loop discovery. Additionally, the same concepts could be explored in Bayesian optimization and active-learning workflows for candidate spaces that admit task-relevant numerical representations but require costly response evaluation.

## Methods

# 4.1 MOF library curation and dataset assembly

The initial structure library was assembled from computation-ready entries in the CoRE MOF 2019 database[41]. A predefined, label-independent curation workflow removed structures with unusable

geometric quantities, failed structure processing, crystallographic or compositional inconsistencies, physically extreme pore-size, density or metal-coordination values, or incomplete values among the selected structure-derived descriptors. A unique MOF identifier was retained throughout structure processing, descriptor construction and adsorption-label alignment.

The final curated library contained 7,532 MOFs with complete values for the 14 Manual descriptors. No descriptor imputation or representation- or search-policy-specific filtering was applied after the candidate library had been fixed. The CEES, DMMP and EtCl GCMC adsorption labels generated as described below were aligned to this library by MOF identifier, producing the complete label table used as the full-library oracle. Descriptor definitions, curation criteria and dataset attrition are provided in Supplementary Note 1.

## 4.2 GCMC simulation and adsorption-label extraction

Adsorption labels were generated by grand canonical Monte Carlo simulations using RASPA 2.0.47[42]. For each target vapour, a separate air-like ternary mixture containing $N_2$, $O_2$ and either CEES, DMMP or EtCl was simulated at 298 K and a total pressure of 101.325 kPa. The corresponding gas-phase mole fractions were 0.789, 0.210 and 0.001, respectively, giving a target-vapour concentration of 1,000 ppm. MOF frameworks were treated as rigid, with all framework atoms fixed at their crystallographic positions.

Framework Lennard–Jones parameters were assigned from the Universal Force Field[43], and framework partial charges were assigned using the EQeq method[44]. $N_2$, $O_2$, CEES, DMMP and EtCl were represented using fixed interaction-site models with TraPPE-type Lennard–Jones parameters and fixed partial charges[45–49]. Cross Lennard–Jones parameters were obtained using the Lorentz–Berthelot mixing rules. Lennard–Jones interactions were truncated at 12 Å without long-range tail corrections, and electrostatic interactions were evaluated using Ewald summation with a 12 Å real-space cutoff.

Each simulation comprised 50,000 equilibration cycles followed by 50,000 production cycles. The production-averaged absolute loading of the target vapour was extracted and reported in mol $kg^{-1}$. These values served as the CEES, DMMP and EtCl adsorption labels; the $N_2$ and $O_2$ loadings were not used as optimization objectives. Simulation outputs were checked for normal termination and successful extraction of the target-vapour loading before alignment with the curated MOF library. Complete force-field definitions, simulation parameters and label-extraction procedures are provided in Supplementary Note 2.

## 4.3 Dataset-level adsorption and Pareto landscape analysis

Dataset-level analyses used the original, untransformed GCMC uptake values of all 7,532 MOFs. For each vapour, the 76 highest-uptake MOFs were defined as the single-vapour Top-1% set, corresponding to ceiling rounding of 1% of the library. The lowest uptake within this fixed set was used as the target-

specific Top-1% cutoff. Overlaps among the target-specific sets were calculated by set intersection, and pairwise Spearman rank correlations were calculated across the complete library.

Pareto membership was determined jointly from CEES, DMMP and EtCl uptake, with all three objectives maximized. A MOF was defined to dominate another when it was no worse in all three objectives and strictly better in at least one. The strict oracle Pareto set comprised the full-library non-dominated materials, and their corresponding objective vectors formed the strict oracle Pareto front. Subsequent Pareto layers were generated by iterative non-dominated sorting, and the Top-3 Pareto set was defined as the union of the first three layers[29]. Full mathematical definitions of the single-vapour statistics and oracle Pareto construction are provided in Supplementary Note 3.

For Fig. 2f, each Manual descriptor was independently min–max normalized using its full-library range. Descriptor distributions were compared among the strict oracle Pareto set, Pareto layers 2–3 and the remaining MOFs. The median normalized value of each descriptor within the strict oracle Pareto set was overlaid as a connected reference profile. Manual-descriptor definitions and calculation procedures are provided in Supplementary Note 1.

## 4.4 Material representation construction

Material representations were constructed for the final 7,532-MOF library. The Manual representation comprised the 14 handcrafted physicochemical descriptors defined in Supplementary Note 1. Pretrained structural representations were extracted from MOFTransformer and Uni-MOF using frozen pretrained encoders, without fine-tuning or adsorption-label supervision[20,21].

For MOFTransformer, the final-layer [CLS] token representation from the frozen pretrained encoder was used as a 768-dimensional structural embedding. For Uni-MOF, the first-token representation from the final encoder layer was extracted as a 512-dimensional structural embedding. Hybrid representations were generated by concatenating the Manual descriptors with each pretrained embedding after MOF-ID alignment.

All representation matrices were indexed by MOF identifier and aligned with the same adsorption-label table. Representation comparisons were therefore performed on the identical set of 7,532 MOFs. Detailed preprocessing procedures, embedding definitions and alignment checks are provided in Supplementary Note 4.

## 4.5 Representation benchmark protocol

Representation performance was evaluated using supervised regression and discovery-oriented ranking analyses for the CEES, DMMP and EtCl adsorption targets. Five feature spaces were compared: Manual, MOFTransformer, Uni-MOF, Manual + MOFTransformer and Manual + Uni-MOF.

Each adsorption target was modelled independently as a single-task regression problem. For each target, all representations and model families used the same fixed training set of 6,025 MOFs and held-

out test set of 1,507 MOFs. Model selection was performed exclusively within the training set using five-fold cross-validation. Feature standardization was implemented within the modelling pipeline, with the scaler fitted separately within each training fold, to prevent information from the validation or test data entering model fitting.

LightGBM was used as the primary nonlinear regression model and Ridge regression as a linear reference[30,31]. Hyperparameters were optimized independently for each target–representation pair using Optuna[50], and the selected pipeline was refitted on the complete training set before a single held-out test evaluation. Performance was assessed using $R^2$, RMSE, MAE and Spearman rank correlation. High-uptake screening performance was evaluated using the Top-1% hit rate and cumulative recall of the true Top-1% materials. The discovery-oriented analyses in Fig. 3b–f were calculated from the held-out LightGBM predictions. For Fig. 3d, target-wise min–max-normalized $R^2$ and Spearman correlation were combined with equal weights to define the overall modelling-quality score. Full benchmark settings, metric definitions, selected hyperparameters and prediction results are provided in Supplementary Note 5.

## 4.6 Retrospective pool-based multi-objective Bayesian optimization benchmark

We evaluated multi-objective discovery using a retrospective, pool-based Bayesian optimization benchmark over the fixed library of 7,532 MOFs. CEES, DMMP and EtCl uptake were treated as simultaneous maximization objectives. The complete GCMC-labelled library served as the retrospective oracle, but the objective values of a material were made available to a search policy only after that material had been queried.

Five matched seeds (42, 123, 456, 789 and 1024) were evaluated. For a given seed, all configurations used the same 100-MOF MaxMin initial design[51], candidate pool, oracle labels, batch schedule and cumulative query budgets. The initial designs were constructed in the standardized Manual descriptor space without using adsorption labels and were included in the cumulative budget. The principal low-budget operating point was 753 evaluated MOFs, corresponding to 10% of the library, and representative-policy trajectories were followed to 1,506 evaluated MOFs, corresponding to 20%. The nominal post-initial batch size was ten; batches were shortened when required to reach a predefined budget milestone exactly.

Performance was calculated from the revealed GCMC responses using normalized hypervolume, strict-front recovery, Top-3 Pareto-layer capture and target-specific Top-1% recovery[52]. Material-level analyses additionally used strict-front recovery frequencies, shared and configuration-specific recovery counts, and nearest-neighbour diagnostics in a fixed shared representation space. Results were summarized as the arithmetic mean ± sample s.d. across the five seeds. Full definitions and implementation details are provided in Supplementary Note 6.

## 4.7 Bayesian-optimization input spaces and shared-space diagnostics

The Bayesian-optimization benchmark used five fixed input spaces. The Manual space comprised the standardized 14-dimensional Manual descriptor vector. The Hybrid representation was constructed by concatenating the 14 Manual descriptors with the frozen 512-dimensional Uni-MOF embedding after MOF-ID alignment. The resulting 526-dimensional matrix was standardized feature-wise, compressed by principal-component analysis and standardized again component-wise before GP fitting.

Four cumulative explained-variance thresholds were evaluated: P80, P85, P90 and P95, retaining 41, 58, 86 and 145 principal components, respectively. Scaling and PCA used representation values only and did not use adsorption labels, target rankings or Pareto assignments. Each model-guided configuration used its assigned Manual or Hybrid PCA space for surrogate fitting and candidate selection. Following the complete configuration screen, P85 was retained as the representative Hybrid space for the detailed analyses in Fig. 5.

Cross-configuration spatial analyses used the same fixed, post-PCA-standardized 58-dimensional P85 coordinate matrix for every policy and seed. PC1 and PC2 were used only for the selection maps, whereas new-to-history and cumulative nearest-neighbour distances were calculated from all 58 coordinates. These shared coordinates were not refitted or re-standardized for individual configurations and did not replace their assigned Bayesian-optimization input spaces.

## 4.8 Gaussian-process surrogates and candidate-selection policies

Model-guided configurations used three independent single-output Gaussian-process surrogates, one for each adsorption objective. The objective values observed in the current queried set were standardized before fitting. Each surrogate used a constant mean, a scaled Matérn-5/2 covariance kernel and a learned homoskedastic noise term[33]. The three models were refitted after each completed batch. Automatic relevance determination was evaluated as an explicit model factor: ARD-on models fitted an independent length scale for every input coordinate, whereas ARD-off models used one shared length scale for each objective-specific GP.

Two model-guided candidate-selection rules were compared. qEHVI combined the three GP posteriors to prioritize candidates according to expected hypervolume improvement and therefore incorporated both posterior means and uncertainty[34,53]. For qEHVI, a single benchmark-wide reference point was defined once from the full-library objective ranges and held fixed across all qEHVI configurations, seeds and query budgets. This shared calibration defined a common hypervolume origin; no unqueried candidate-specific responses, target ranks or Pareto assignments were supplied to the GP surrogates. Posterior-mean Greedy selection independently min–max normalized the predicted means for CEES, DMMP and EtCl and ranked candidates by their equally weighted average; it did not use posterior variance or

expected hypervolume improvement. Random sampling selected unqueried materials uniformly without fitting a surrogate.

The five input spaces, two selection rules and two ARD states defined 20 model-guided configurations; Random sampling was retained as a separate unguided reference. qEHVI batches were assembled sequentially, with candidates already selected within the current batch treated as pending points. Their oracle labels were revealed only after the complete batch had been formed, and the GP models were not refitted between within-batch selections. For memory efficiency, full-pool posterior prediction was performed in chunks, and qEHVI evaluation used a temporary posterior-based candidate-screening step followed by chunked acquisition evaluation.

Following the 10% configuration screen, four model-guided configurations were retained for Fig. 5: qEHVI–Manual and Greedy–Manual with ARD enabled, and qEHVI–P85 and Greedy–P85 with ARD disabled. Random sampling was added as the fifth representative policy. Comparisons between qEHVI–Manual and qEHVI–P85 therefore represent comparisons between two selected complete configurations rather than a representation-only ablation. Full GP, acquisition, screening and configuration settings are reported in Supplementary Notes 6 and 7.

## 4.9 Diversity- and feasibility-aware experimental candidate selection

Candidate prioritization began from the materials queried by the two representative qEHVI configurations retained after the configuration benchmark: RepMOBO-Manual, corresponding to qEHVI–Manual with ARD on, and RepMOBO-Hybrid, corresponding to qEHVI–P85 with ARD off. For each configuration, five matched-seed runs were considered at the 10% cumulative query budget, corresponding to 753 evaluated MOFs per run including the common 100-MOF initial design. Pooling materials queried in at least one of the ten runs produced 2,385 unique candidates.

For each material, cross-run occurrence was defined as the number of the ten RepMOBO runs in which it had been queried by the 10% budget. Materials with a total occurrence of at least 6/10 were retained, reducing the candidate pool to 428 MOFs. The recurrent candidates were then intersected with two fixed full-library performance sets: the Top-3 Pareto set and the Top-1% set ranked by a composite adsorption score. The composite score was calculated as the arithmetic mean of the target-wise min–max-normalized CEES, DMMP and EtCl GCMC uptakes. This intersection yielded 30 Pareto-enriched candidates, of which eight were ranked among the top ten materials in the complete library by composite score.

Five candidates were advanced to synthesis on the basis of their computational priority, CSD traceability, available synthesis procedures and experimental feasibility: OLOGAY, KAZXEQ, OYEWEW, QAJJOD01 and IXEJOM. Four candidates produced activated samples consistent with the target structures and were used for adsorption measurements: OLOGAY (NJU-Bai1), OYEWEW (MOF-515), QAJJOD01 (UiO-67-Me) and IXEJOM (PCN-700-$(CF_3)_2$). The attempted synthesis of KAZXEQ did not produce the target material

under the conditions tested. The composite-score definition, candidate provenance and experimental outcomes are provided in Supplementary Note 8.

## 4.10 MOF synthesis, characterization and vapour adsorption measurements

NJU-Bai1, MOF-515, UiO-67-Me and PCN-700-$(CF_3)_2$ were synthesized using solvothermal procedures adapted from the corresponding reported preparations. The attempted synthesis of KAZXEQ was performed separately. Reagent quantities, solvent compositions, reaction vessels, temperatures, durations, washing, solvent-exchange and activation procedures are provided in Supplementary Note 9.

The synthesized materials were characterized by powder X-ray diffraction, field-emission scanning electron microscopy, thermogravimetric analysis, Fourier-transform infrared spectroscopy and $N_2$ adsorption–desorption measurements at 77 K. BET surface areas were calculated from the $N_2$ sorption data, and pore-size distributions were derived using the pore-kernel method[54]. Instrumentation, measurement conditions and data-processing procedures are detailed in Supplementary Note 9.

Single-component gravimetric adsorption isotherms for CEES, DMMP and EtCl were measured at 298 K using activated samples. CEES and DMMP adsorption was evaluated as a function of relative pressure ( $P/P_0$), whereas EtCl adsorption was evaluated as a function of inlet volume fraction ( $V_i/V$); uptake was normalized to the activated sample mass. Detailed sample activation, vapour generation, equilibration criteria, measurement ranges and data-processing procedures are provided in Supplementary Note 9.

## Declarations

# Competing interests

The authors declare no competing interests.

# Author contributions

**Zhongyuan Ming** conceived the study, developed and implemented the RepMOBO framework, led the curation of the MOF candidate library and GCMC-labelled adsorption dataset, implemented the representation-learning and multi-objective Bayesian optimization workflows, performed the computational analyses, prepared the figures and drafted the manuscript. **Xing Liu** contributed to data curation, adsorption-property analysis and validation of the computational workflow. **Yihe Zhang** contributed to the design of the Bayesian optimization protocol, analysis of representation-dependent search behaviour and interpretation of the computational results. **Hao Ren** contributed to data curation and validation of the computational results. **Xinbo Wang** contributed to experimental data analysis and

characterization of the MOF samples. **Heguo Li** performed MOF synthesis, structural characterization and vapour adsorption measurements, and analysed the experimental results. **Yue Zhao** and **Suihua He** supervised the project, acquired funding, contributed to the conceptual development of the study and reviewed and edited the manuscript. All authors discussed the results and approved the final manuscript.

# Acknowledgements

This work was supported by the Shandong Provincial Natural Science Foundation (grant no. ZR2026QC0095), the General Program of the China Postdoctoral Science Foundation, the State Key Laboratory of Chemistry for NBC Hazards Protection Research Fund (grant no. SKLNBC20250110), and the Guangdong Association for Science and Technology Young Talents Cultivation Program (grant no. SKXRC2025280).

## Data availability

The data supporting the findings of this study, including the curated MOF identifiers, handcrafted physicochemical descriptors, CEES, DMMP and EtCl GCMC adsorption labels, structural representations, PCA coordinates, RepMOBO query records, configuration-level performance metrics, material-level recovery records and the source data underlying the main and supplementary figures, are available in Zenodo at [ZENODO DATA DOI OR URL]. Raw and processed experimental characterization and vapour-adsorption data are available in the same repository. The starting crystal structures were obtained from the CoRE MOF 2019 database, as described and cited in the Methods. Source data are provided with this paper.

## Code availability

All custom code used for MOF-library curation, descriptor and representation alignment, PCA construction, representation benchmarking, Gaussian-process modelling, qEHVI and Greedy RepMOBO searches, Pareto-recovery analysis and figure generation is available on GitHub at [GITHUB REPOSITORY URL]. The version used in this study has been archived in Zenodo at [ZENODO CODE DOI OR URL]. The archived release includes environment specifications, configuration files, random seeds and instructions for reproducing the principal analyses and figures. Third-party software and pretrained models used in this study are identified and cited in the Methods.

## References

1. Bobbitt NS et al (2017) Metal–organic frameworks for the removal of toxic industrial chemicals and chemical warfare agents. Chem Soc Rev 46:3357–3385

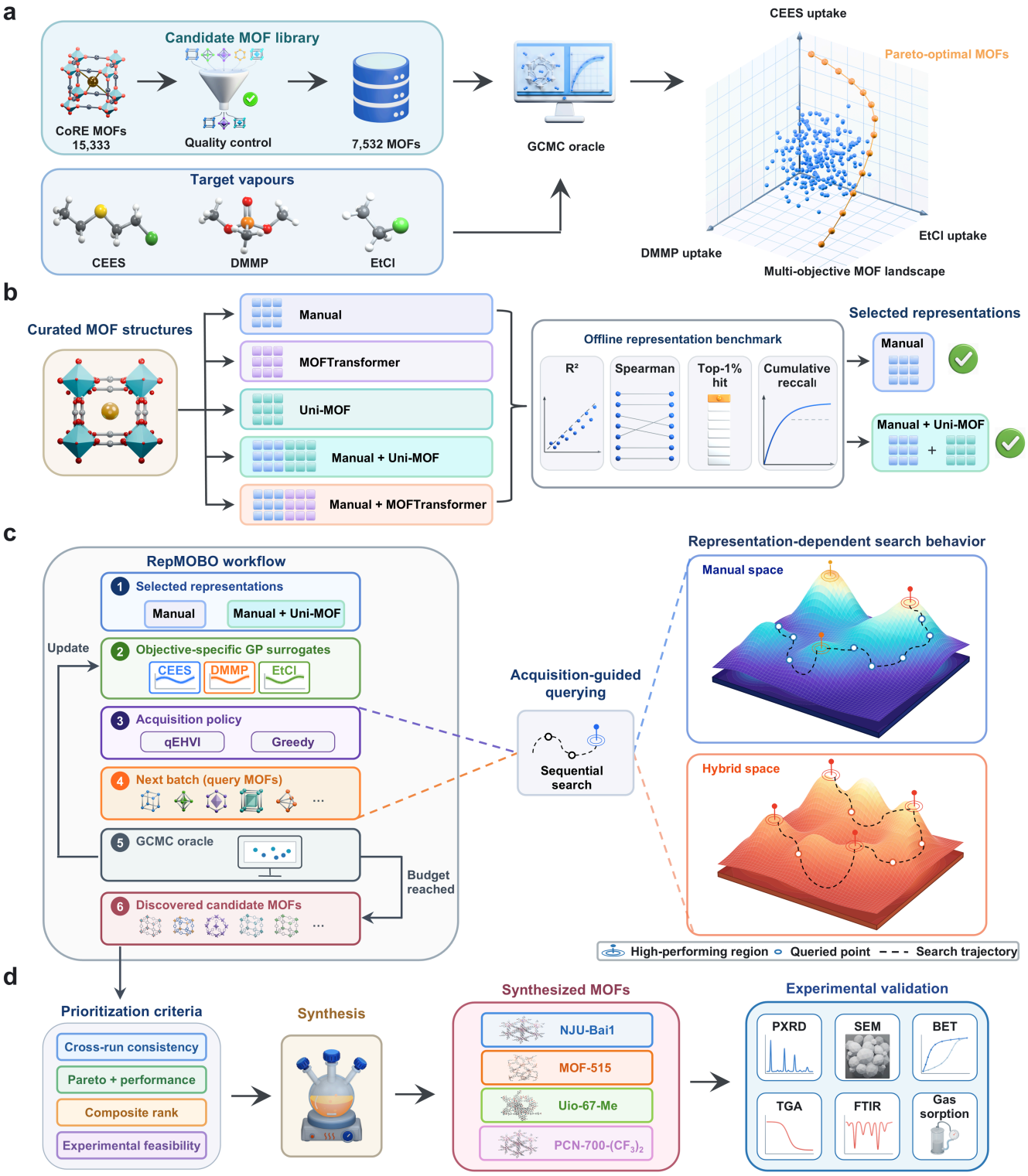
2. Islamoglu T et al (2020) Metal–organic frameworks against toxic chemicals. Chem Rev 120:8130–8160
3. Furukawa H, Cordova KE, O'Keeffe M, Yaghi OM (2013) The chemistry and applications of metal-organic frameworks. Science 341:1230444
4. Comlek Y, Pham TD, Snurr RQ, Chen W (2023) Rapid design of top-performing metal-organic frameworks with qualitative representations of building blocks. npj Comput Mater 9:171–183
5. Jablonka KM, Jothiappan GM, Wang S, Smit B, Yoo B (2021) Bias free multiobjective active learning for materials design and discovery. Nat Commun 12:2312
6. Schultz LE, Gantzler N, Bobbitt NS, Gallis S, D. F., Dingreville R (2026) Graph neural network-based multi-objective bayesian optimization for enhanced screening of metal–organic frameworks with optimal separation performance. J Mater Chem A 14:10836–10853
7. Hu S et al (2026) Experimental mechanician for plate lattice metamaterial discovery. Nat Commun 17:3933
8. Rao Z et al (2022) Machine learning–enabled high-entropy alloy discovery. Science 378:78–85
9. Kusne AG et al (2020) On-the-fly closed-loop materials discovery via bayesian active learning. Nat Commun 11:5966
10. Lee JA et al (2025) Active learning framework to optimize process parameters for additive-manufactured ti-6Al-4V with high strength and ductility. Nat Commun 16:931
11. Gao D et al (2026) Autonomous closed-loop framework for reproducible perovskite solar cells. Nature 653:707–714
12. Hong X et al (2026) Deep active learning and knowledge transfer for rapid discovery of lithium metal battery electrolytes. Nat Commun 17:5146
13. Jing L et al (2026) Enhancing mechanical properties of al-li alloys via multi-objective active learning optimization. npj Comput Mater 1–42. 10.1038/s41524-026-02188-6
14. Jose A, Devijver E, Jakse N, Poloni R (2024) Informative training data for efficient property prediction in metal–organic frameworks by active learning. J Am Chem Soc 146:6134–6144
15. Shahriari B, Swersky K, Wang Z, Adams RP (2016) & De Freitas, N. Taking the human out of the loop: A review of bayesian optimization. *Proc. IEEE* 104, 148–175
16. Settles B (2012) Active Learning. Springer International Publishing, Cham. 10.1007/978-3-031-01560-1
17. Damewood J et al (2023) Representations of materials for machine learning. Annu Rev Mater Res 53:399–426
18. Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A (2018) Machine learning for molecular and materials science. Nature 559:547–555
19. Jablonka KM, Ongari D, Moosavi SM, Smit B (2020) Big-Data Science in Porous Materials: Materials Genomics and Machine Learning. Chem Rev 120:8066–8129

20. Wang J et al (2024) A comprehensive transformer-based approach for high-accuracy gas adsorption predictions in metal-organic frameworks. Nat Commun 15:1904
21. Kang Y, Park H, Smit B, Kim J (2023) A multi-modal pre-training transformer for universal transfer learning in metal–organic frameworks. Nat Mach Intell 5:309–318
22. Trabucco B, Kumar A, Geng X, Levine S (2021) Conservative Objective Models for Effective Offline Model-Based Optimization. in *Proceedings of the 38th International Conference on Machine Learning* vol. 139 10358–10368PMLR
23. Zuluaga M, Krause A, Puschel M (2016) ε-PAL: An Active Learning Approach to the Multi-Objective Optimization Problem. J Mach Learn Res 17:1–32
24. Moon J et al (2024) Active learning guides discovery of a champion four-metal perovskite oxide for oxygen evolution electrocatalysis. Nat Mater 23:108–115
25. Rajabi-Kochi M, Mahboubi N, Gill APS, Moosavi SM (2025) Adaptive representation of molecules and materials in bayesian optimization. Chem Sci 16:5464–5474
26. Wang Y, LeVan MD (2008) Nanopore Diffusion Rates for Adsorption Determined by Pressure-Swing and Concentration-Swing Frequency Response and Comparison with Darken's Equation. Ind Eng Chem Res 47:3121–3128
27. Kim H-H et al (2021) Decomposition of the simulant 2-chloroethyl ethyl sulfide blister agent under ambient conditions using metal–organic frameworks. ACS Appl Mater Interfaces 13:3782–3792
28. Ma F-J et al (2011) A sodalite-type porous metal–organic framework with polyoxometalate templates: Adsorption and decomposition of dimethyl methylphosphonate. J Am Chem Soc 133:4178–4181
29. Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 6:182–197
30. Ke G et al (2017) LightGBM: A highly efficient gradient boosting decision tree. In: Guyon I et al (eds) Advances in Neural Information Processing Systems, vol 30. Curran Associates, Inc., pp 3146–3154
31. Hoerl AE, Kennard RW (1970) Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12:55–67
32. Chen G, Fleck M, Stuyver T (2026) Leveraging chemical hidden-space representations effectively in bayesian optimization for experiment design through dimension-aware hyperpriors. J Chem Theory Comput 22:5594–5608
33. Rasmussen CE, Williams CK, I (2006) Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA
34. Daulton S, Balandat M, Bakshy E (2020) Differentiable expected hypervolume improvement for parallel multi-objective bayesian optimization. in *Advances in Neural Information Processing Systems* vol. 33 9851–9864Curran Associates, Inc
35. Guerreiro AP, Fonseca CM, Paquete L (2022) The hypervolume indicator: Computational problems and algorithms. ACM Comput Surv 54:1–42

36. De Ath G, Everson RM, Rahat AAM, Fieldsend JE (2021) Greed is good: Exploration and exploitation trade-offs in bayesian optimisation. ACM Trans Evol Learn Optim 1:1–22
37. Duan J, Bai J, Zheng B, Li Y, Ren W (2011) Controlling the shifting degree of interpenetrated metal–organic frameworks by modulator and temperature and their hydrogen adsorption properties. Chem Commun 47:2556
38. Liu C-S et al (2017) Highly stable aluminum-based metal-organic frameworks as biosensing platforms for assessment of food safety. Biosens Bioelectron 91:804–810
39. Øien-Ødegaard S et al (2016) UiO-67-type metal–organic frameworks with enhanced water stability and methane adsorption capacity. Inorg Chem 55:1986–1991
40. Yuan S et al (2016) Flexible zirconium metal-organic frameworks as bioinspired switchable catalysts. Angew Chem Int Ed 55:10776–10780
41. Chung YG et al (2019) Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal–Organic Framework Database: CoRE MOF 2019. J Chem Eng Data 64:5985–5998
42. Dubbeldam D, Calero S, Ellis DE, Snurr RQ (2016) RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. Mol Simulat 42:81–101
43. Rappe AK, Casewit CJ, Colwell KS, Goddard WA, Skiff WM (1992) UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. J Am Chem Soc 114:10024–10035
44. Wilmer CE, Kim KC, Snurr RQ (2012) An Extended Charge Equilibration Method. J Phys Chem Lett 3:2506–2511
45. Potoff JJ, Siepmann JI (2001) Vapor–liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen. AIChE J 47:1676–1682
46. Hansen N, Agbor FAB, Keil FJ (2007) New force fields for nitrous oxide and oxygen and their application to phase equilibria simulations. Fluid Phase Equilib 259:181–188
47. Sokkalingam N, Kamath G, Coscione M, Potoff JJ (2009) Extension of the transferable potentials for phase equilibria force field to dimethylmethyl phosphonate, sarin, and soman. J Phys Chem B 113:10292–10297
48. Herdes C et al (2007) Selective adsorption of volatile organic compounds in micropore aluminum methylphosphonate-α: A combined molecular simulation–experimental approach. Langmuir 23:7299–7305
49. Wang W et al (2023) Machine learning assisted high-throughput computational screening of MOFs for the capture of chemical warfare agents from the air. Sep Purif Technol 325:124546
50. Akiba T, Sano S, Yanase T, Ohta T, Koyama M, Optuna (2019) A next-generation hyperparameter optimization framework. in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* 2623–2631 (ACM, Anchorage AK USA, 10.1145/3292500.3330701
51. Kennard RW, Stone LA (1969) Computer aided design of experiments. Technometrics 11:137–148

52. Zitzler E, Thiele L (1999) Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach. IEEE Trans Evol Comput 3:257–271
53. Balandat M et al (2020) Curran Associates, Inc.,. BoTorch: A framework for efficient Monte-Carlo Bayesian optimization. in *Advances in Neural Information Processing Systems* vol. 33 21524–21538
54. Gómez-Gualdrón DA, Moghadam PZ, Hupp JT, Farha OK, Snurr RQ (2016) Application of consistency criteria to calculate BET areas of micro- and mesoporous metal–organic frameworks. J Am Chem Soc 138:215–224

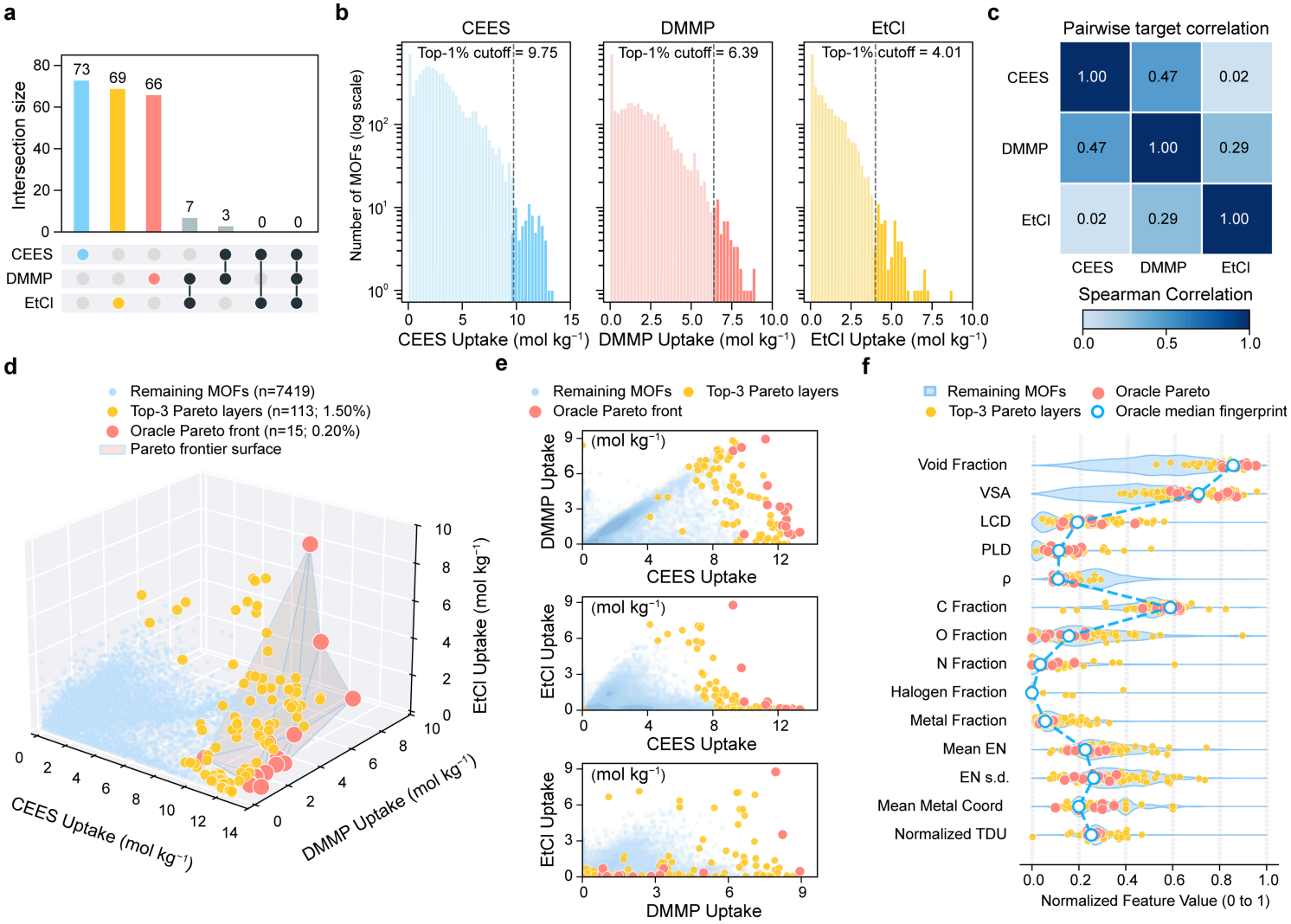
# Figures




## Figure 1

**Multi-vapour MOF discovery by RepMOBO and downstream experimental validation. a**, Construction of the multi-vapour benchmark from 15,333 CoRE MOFs, quality control to 7,532 structures, GCMC evaluation of CEES, DMMP and EtCl uptake, and the resulting multi-objective landscape with Pareto-optimal MOFs. **b**, Offline benchmarking of five structure representations using $R^2$, Spearman correlation, top-1% hit rate and cumulative recall, yielding Manual and Manual + Uni-MOF as the selected

representations. **c**, RepMOBO combines objective-specific Gaussian-process surrogates, qEHVI or Greedy acquisition, batch querying and iterative GCMC-oracle updates; the central schematic highlights acquisition-guided querying, while the landscapes illustrate representation-dependent search behaviour. **d**, Downstream experimental translation of discovered candidates through prioritization and synthesis, yielding four MOFs for PXRD, SEM, $N_2$ sorption/BET, TGA, FTIR and vapour adsorption measurements. In the vapour models, C, H, O, S, P and Cl are shown in grey, white, red, yellow, orange and green, respectively.




## Figure 2

**Multi-vapour adsorption defines a sparse and heterogeneous Pareto landscape across the MOF library.** **a**, UpSet plot showing overlaps among the single-vapour top-1% MOF sets for CEES, DMMP and EtCl. **b**, Full-library uptake distributions, with dashed lines indicating the target-specific Top-1% cutoffs. **c**, Pairwise Spearman correlation matrix for the three adsorption targets. **d**, Three-dimensional CEES–DMMP–EtCl adsorption landscape of 7,419 non-Top-3 MOFs and 113 Top-3 Pareto-layer MOFs, including the 15-member strict oracle front. The shaded surface connects the strict-front MOFs as a visual guide. **e,** Pairwise projections of the adsorption landscape onto the CEES–DMMP, CEES–EtCl and DMMP–EtCl

planes. **f**, Normalized distributions of 14 handcrafted physicochemical descriptors for all other MOFs, the top three Pareto layers and the strict oracle Pareto front. Open circles connected by a dashed line denote the median descriptor profile of the strict oracle Pareto front.

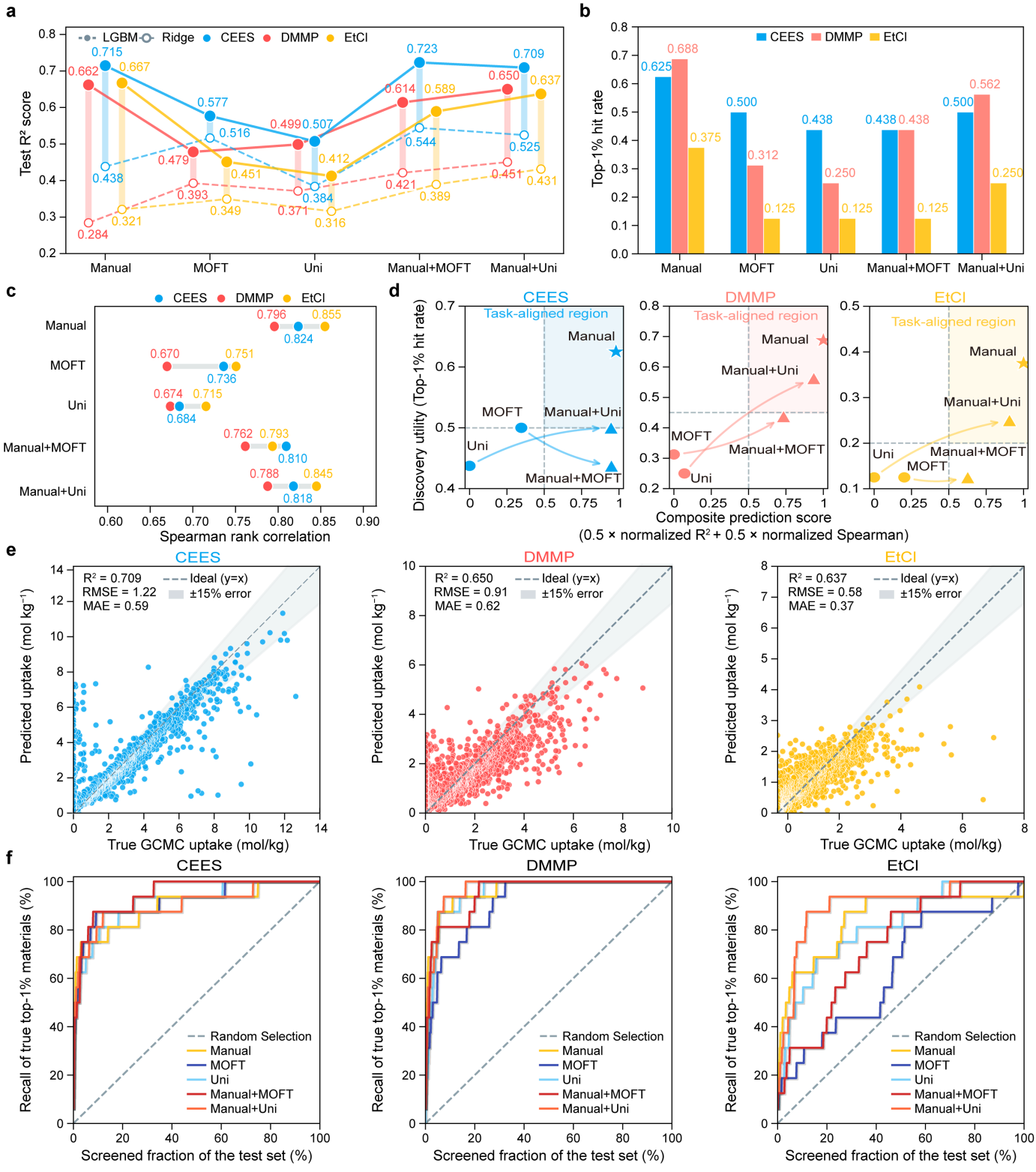
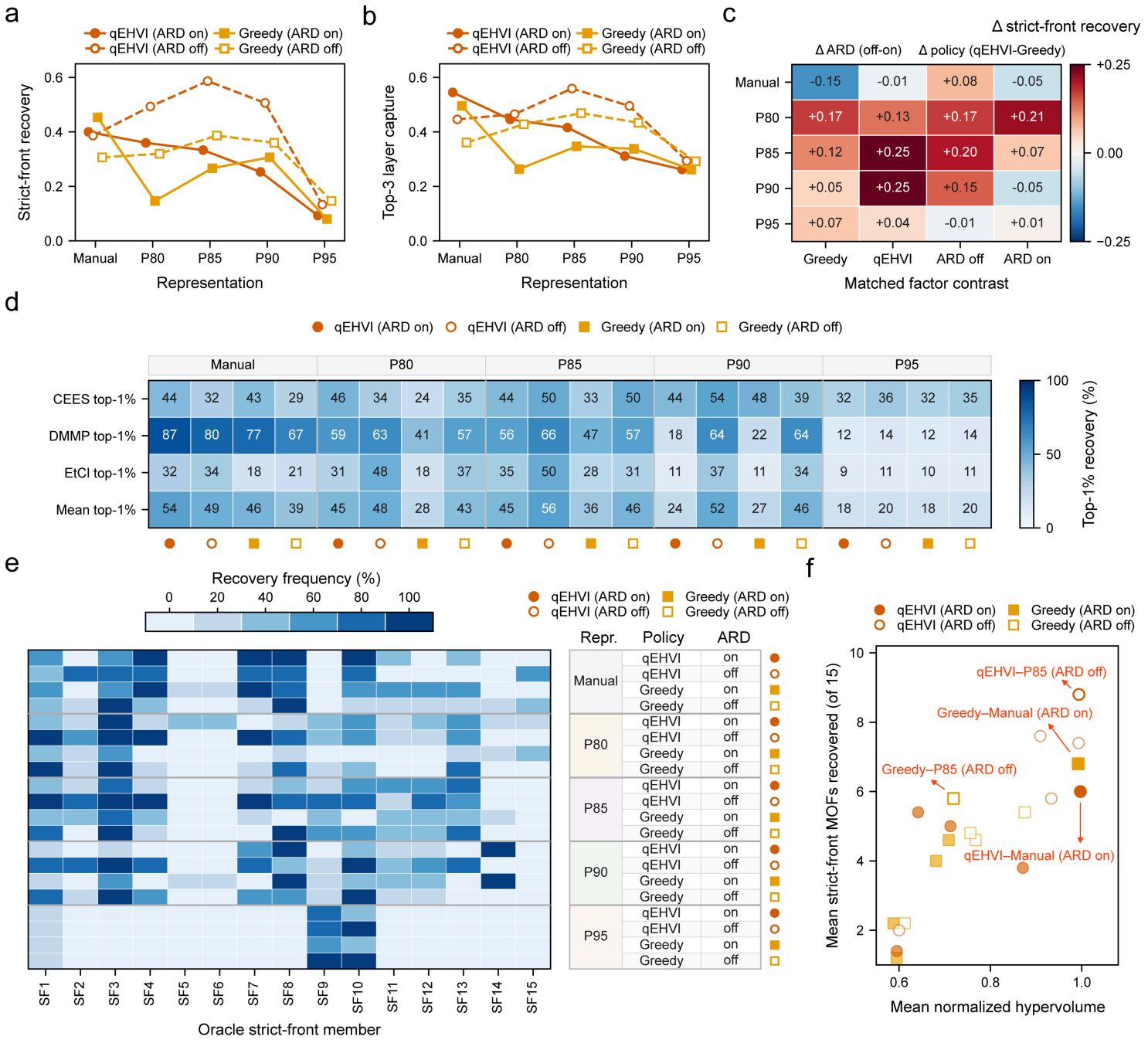



**Figure 3**

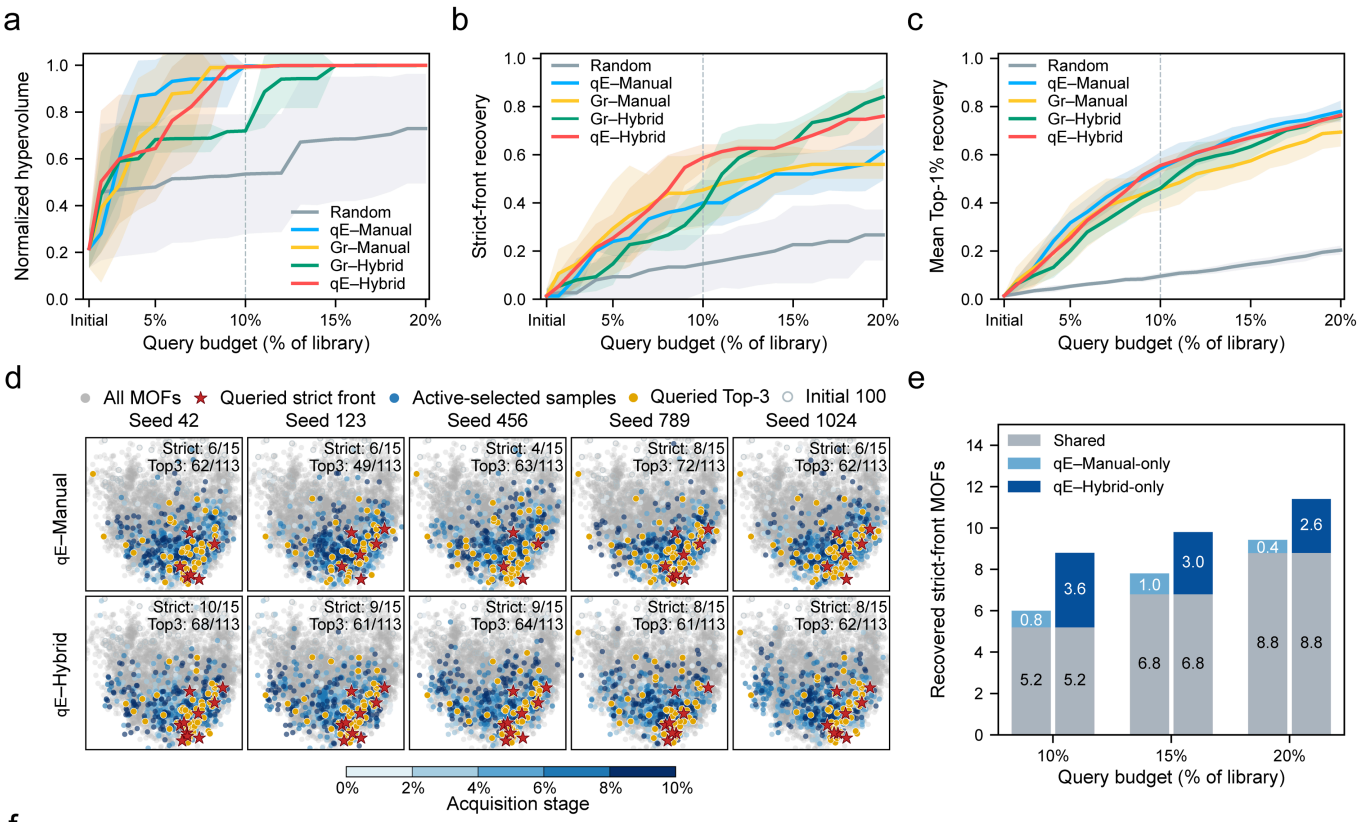
**Representation benchmark across handcrafted and pretrained MOF feature spaces. a**, Held-out test-set $R^2$ values for LightGBM and Ridge regression across five representation spaces and three vapour targets. **b**, Top-1% hit rates calculated from the corresponding held-out LightGBM predictions. **c**, Spearman rank correlations for the same predictions. **d**, Prediction–discovery alignment map. The composite prediction score is calculated as 0.5 × normalized $R^2$ + 0.5 × normalized Spearman correlation, and tail-screening utility is measured by the Top-1% hit rate. The shaded upper-right region is included as a visual guide to representations with jointly high composite prediction scores and Top-1% hit rates. **e**, Parity plots for the representative Manual + Uni-MOF hybrid representation. Diagonal lines indicate ideal agreement, and grey shaded regions denote the ±15% relative-error band. **f**, Cumulative recall of the true Top-1% MOFs as a function of the screened test-set fraction. The dashed diagonal line denotes the expected recall under random screening. In all panels, Manual denotes the 14 handcrafted physicochemical descriptors; MOFT denotes the MOFTransformer embedding; Uni-MOF denotes the Uni-MOF embedding; Manual + MOFT and Manual + Uni-MOF denote the corresponding concatenated hybrid representations.




## Figure 4

**Configuration-level benchmarking and representative-configuration selection at the 10% cumulative query budget. a**, Recovery of the 15-member oracle strict Pareto front across 20 configurations combining five GP input spaces, two selection rules and ARD on/off. **b,** Capture of the 113 MOFs in the Top-3 oracle Pareto layers across the same configurations. In a,b, points and error bars show the mean ± sample s.d. across five matched seeds. **c,** Matched factor contrasts in strict-front recovery. ARD contrasts are calculated as ARD off minus ARD on within each selection rule, whereas policy contrasts are calculated as qEHVI minus Greedy within each ARD setting. Positive values favour ARD off or qEHVI, respectively. **d,** Five-seed mean recovery of the CEES, DMMP and EtCl single-objective Top-1% sets and their arithmetic mean. Configurations are grouped by GP input space, with annotation strips indicating

selection rule, input space and ARD setting. **e,** Five-seed recovery frequencies of the 15 oracle strict-front MOFs. Each cell indicates the percentage of seeds in which the corresponding material was recovered, ranging from 0% to 100% in increments of 20%. Rows are ordered by average-linkage clustering of weighted-Jaccard distances between the 15-dimensional recovery fingerprints. **f,** Five-seed mean normalized hypervolume versus mean strict-front recovery. The four configurations retained for subsequent analysis are annotated.




f

Objective-space profiles of 15 strict-front MOFs at 10% budget across five seeds

| Group (count) | Class | SF ID | CSD code | Normalized true uptake (0–1) DMMP | CEES | EtCl | Recovered in (out of 5 seeds) qE–Manual | qE–Hybrid |
|---|---|---|---|---|---|---|---|---|
| Differential (5) | | SF2 | IBAPIM | 0.090 | 0.965 | 0.012 | 0 | 4 |
| | | SF9 | RAYMIP | 0.316 | 0.930 | 0.006 | 0 | 4 |
| | | SF12 | UMODEH01 | 0.356 | 0.909 | 0.015 | 0 | 4 |
| | | SF13 | UMOVEZ04 | 0.168 | 0.936 | 0.010 | 2 | 3 |
| | | SF11 | UMOBEF02 | 0.107 | 0.946 | 0.009 | 2 | 1 |
| Shared-stable (6) | | SF3 | LAPZEK | 0.373 | 0.850 | 0.079 | 3 | 5 |
| | | SF4 | MOYZUW | 1.000 | 0.842 | 0.052 | 5 | 5 |
| | | SF7 | OLOGAY | 0.889 | 0.692 | 1.000 | 5 | 5 |
| | | SF8 | OYEWEW | 0.920 | 0.731 | 0.403 | 5 | 4 |
| | | SF10 | SUKYON | 0.177 | 0.919 | 0.016 | 5 | 4 |
| | | SF1 | FOPFAS | 0.557 | 0.851 | 0.039 | 3 | 5 |
| Rare/missed (4) | | SF5 | NILGIA | 0.232 | 0.938 | 0.003 | 0 | 0 |
| | | SF6 | NILGOG | 0.351 | 0.944 | 0.002 | 0 | 0 |
| | | SF14 | XICKOL | 0.094 | 0.744 | 0.080 | 0 | 0 |
| | | SF15 | YUGZIK | 0.113 | 1.000 | 0.004 | 0 | 0 |

Discovery class: Hybrid-enriched, Manual-enriched, Shared-stable, Rare/missed

Normalized true uptake (0–1)

## Figure 5

**Budget-dependent performance and material-level discovery of representative RepMOBO configurations.** qE–Manual and Gr–Manual denote qEHVI and posterior-mean Greedy selection in the Manual descriptor space with ARD on, whereas qE–Hybrid and Gr–Hybrid denote the corresponding configurations in the P85 hybrid space with ARD off. **a**, Normalized hypervolume versus cumulative query budget. **b,** Recovery of the 15 oracle strict-front MOFs. **c,** Mean Top-1% recovery across CEES, DMMP and EtCl. In a–c, curves and shaded bands show the mean ± sample s.d. across five matched seeds, and the vertical dashed line marks the 10% cumulative query budget. **d,** Selection maps for qE–Manual and qE–Hybrid at the 10% budget across five seeds, projected onto the fixed fixed P85 hybrid PC1–PC2 space. Grey points denote the full MOF library, open circles the common initial design, blue points post-initial selections coloured by acquisition stage, yellow points queried Top-3 Pareto-layer materials and red stars queried strict-front materials. **e,** Seed-averaged decomposition of strict-front recovery at the 10%, 15% and 20% budgets into shared, Manual-only and Hybrid-only members. **f,** Objective profiles and five-seed recovery counts of the 15 oracle strict-front MOFs at the 10% budget. Rows are grouped by recovery class, the heatmap shows normalized true uptake for DMMP, CEES and EtCl, and the side bars report recovery counts for qE–Manual and qE–Hybrid. SF identifiers correspond to the recovery fingerprints in Fig. 4e, and CSD codes identify the individual MOFs.




**Figure 6**

**Experimental prioritization and adsorption validation of computationally prioritized MOF candidates. a**, Funnel from the full GCMC-labelled MOF library to the five candidates selected for synthesis. **b**, Cross-run occurrence of the eight highest-priority candidates across five RepMOBO-Manual and five RepMOBO-Hybrid runs at the 10% query budget, together with their experimental status; candidates are identified by CSD refcodes. **c**, Single-vapour adsorption isotherms of the four successfully synthesized MOFs for CEES, DMMP and EtCl. **d**, Uptake at the 0.1 reference condition ($P/P_0$ = 0.1 for CEES and DMMP and $V_i/V$ = 0.1 for EtCl) and at the highest tested condition.

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx