

# Chemical Science

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: S. Bonakala, M. Wahiduzzaman, T. Watanabe, K. Hamzaoui and G. Maurin, *Chem. Sci.*, 2026, DOI: 10.1039/D6SC00831C.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

## ARTICLE

# Towards Accurate and Scalable High-throughput MOF Adsorption Screening: Merging Classical Force Fields and Universal Machine Learned Interatomic Potentials

Satyanarayana Bonakala<sup>a</sup>, Mohammad Wahiduzzaman<sup>a</sup>, Taku Watanabe<sup>b</sup>, Karim Hamzaoui<sup>b</sup>, and Guillaume Maurin<sup>a,c\*</sup>Received 00th January 20xx,  
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

High-throughput computational screening (HTCS) of gas adsorption in metal–organic frameworks (MOFs) typically relies on classical generic force fields which are computationally efficient but often fail to capture complex host–guest interactions. Universal machine learned interatomic potentials (u-MLIPs) offer near-quantum accuracy at far lower cost than density-functional theory (DFT), yet their large-scale application in adsorption screening remains limited. Here, we introduce an efficient hybrid screening workflow that merges classical generic force fields and u-MLIPs within a Monte Carlo scheme to accurately assess the adsorption performance of a large MOF structure database. As a proof-of-concept, this HTCS is applied to identify top performing MOFs for selective ethylene capture under humid conditions, highly relevant to food-preservation packaging technologies. Our workflow demonstrates that accurate treatment of both host–guest energetics and framework flexibility, enabled by u-MLIPs, is essential to achieve reliable adsorption-performance rankings.

## Introduction

Metal–Organic Frameworks (MOFs), a class of hybrid porous materials with remarkable chemical and structural diversity, hold immense promise across a wide range of applications. Yet, their full potential, particularly for sorption-based processes, remains largely untapped. One key reason is that only a small fraction of the 128,000+ MOF structures reported in the Cambridge Structural Database (CSD)<sup>1</sup> have been explored for specific applications. This limited exploration is primarily due to the impracticality of synthesizing and experimentally testing thousands of candidates, given the substantial investment of time, resources and specialized infrastructure required. To overcome these limitations, computational approaches have emerged as indispensable tools. In particular, high-throughput computational screening (HTCS) has revolutionized the evaluation of MOFs and related porous materials properties across large structural databases. This strategy enables the rapid identification of promising sorbent candidates for a wide array of adsorption/separation applications, including carbon capture, H<sub>2</sub> storage, natural gas purification, among others.<sup>2–4</sup> Traditionally, these computational screenings have relied on classical interatomic potentials, commonly referred to as classical universal interatomic potentials (u-IPs) to describe the MOF/guest interactions. The most widely used classical u-IPs, e.g., UFF<sup>5</sup> employ transferable Lennard-Jones parameters to account for van der Waals interactions, complemented by

electrostatic terms to describe interatomic interactions within the system.

While this modelling approach enables an effective qualitative assessment of thousands of MOF sorbents, it inherently lacks the precision required to capture complex host-guest interactions, especially between guest molecules and functionalized linkers as well as open metal sites (OMS) and polar groups within the inorganic nodes, where polarization effect may play a dominant role.<sup>6</sup> For example, UFF significantly underestimates CO<sub>2</sub> interaction energies with the OMSs of Mg-MOF-74 by 60–65% compared to density-functional theory (DFT) calculations,<sup>7</sup> which can potentially lead to the mis-ranking of this sub-class of MOFs. To overcome these shortcomings, case-specific FF parametrization has been developed using exhaustive DFT-derived interaction energies for individual MOF-guest pairs, e.g., CO<sub>2</sub>@Mg-MOF-74,<sup>8</sup> C<sub>2</sub>H<sub>2</sub>@Cu-BTC,<sup>9</sup> N<sub>2</sub>@MIL-100(Cr)<sup>10</sup> and SO<sub>2</sub>@Mg<sub>2</sub>(dobpdc).<sup>11</sup> More recently, Machine-learned potentials (MLPs), trained on high-quality quantum mechanical datasets, have proven effective in capturing intricate host/guest interactions across a range of MOFs, achieving with near quantum mechanical-accuracy at significantly lower computational cost.<sup>12–14</sup> For instance, this strategy has been successfully applied to accurately predict the H<sub>2</sub> adsorption isotherm in the low-pressure domain of Al-*soc*-MOF-1d containing OMSs.<sup>15</sup> Despite their improved accuracy in targeted studies for a few MOFs, both case-specific FFs and trained MLPs remain limited in transferability and are computationally impractical for large-scale screenings. In this context, the on-going development of universal machine learned interatomic potentials (u-MLIPs) offers a potentially transformative advance. Foundation models such as the Multi-Atomic Cluster Expansion (MACE),<sup>16</sup> Materials

<sup>a</sup> ICGM, Univ. Montpellier, CNRS, ENSCM, 34293 Montpellier, France<sup>b</sup> Matlantis Corporation, Otemachi, Chiyoda-ku, 100-0004 Tokyo, Japan<sup>c</sup> Institut Universitaire de France, France

Graph Network with three-body interactions (M3GNet)<sup>17</sup> and Crystal Hamiltonian Graph Neural Network (CHGNet)<sup>18</sup> architectures have demonstrated near-DFT accuracy for predicting MOF framework properties offering a promising pathway towards MLP-based generalizable MOF screening platforms. Notably, Meta's Equiformer V2 trained on the Open DAC datasets, ODAC23 and ODAC25 have shown superior performance over the UFF in predicting CO<sub>2</sub> and H<sub>2</sub>O adsorption energies for a few MOFs, especially those featuring OMSs.<sup>19</sup> A few recent benchmark studies extended this comparison across a small sub-set of MOFs evaluating a range of u-MLIPs including EqV2-ODAC,<sup>20</sup> MACE, M3GNet, CHGNet, and the Equivariant Smooth Energy Network (eSEN)<sup>21</sup> against DFT and UFF calculations for computing their CO<sub>2</sub> and H<sub>2</sub>O adsorption energies.<sup>22</sup>

However, despite these promising developments, u-MLIPs have not yet been systematically applied to predict adsorption properties across a large and diverse set of MOFs. Their integration into HTCS pipelines is still not feasible due to computational costs, which, while significantly lower compared to DFT, are still substantially higher than those associated with classical FFs. Herein, we advance a hybrid HTCS workflow for MOFs that bridges classical u-IPs and u-MLIPs, using ethylene (C<sub>2</sub>H<sub>4</sub>) capture under humid conditions as a test case. This adsorption scenario poses a fundamental challenge: weakly physisorbed ethylene must compete with strongly interacting polar water molecules, requiring accurate modelling of the MOF/guests interactions. Ethylene was selected due to its high-impact relevance in food preservation technologies, particularly in regulating the ripening of fruits and vegetables.<sup>23</sup> Its uncontrolled accumulation in postharvest storage and food packaging, environments typically characterized by high humidity,<sup>24</sup> can accelerate spoilage, contributing significantly to food loss.<sup>25</sup> A few technologies have been proposed to mitigate ethylene accumulation, including chemical inhibitors such as 1-methylcyclopropene (1-MCP),<sup>26</sup> oxidative scavenging agents like potassium permanganate (KMnO<sub>4</sub>)<sup>27</sup> and TiO<sub>2</sub>-based photocatalysts.<sup>28</sup> A truly viable sorbent-based solution has yet to be realized, as conventional porous materials, such as zeolites and activated carbons, exhibit inherently low C<sub>2</sub>H<sub>4</sub> affinity and/or inadequate C<sub>2</sub>H<sub>4</sub>/H<sub>2</sub>O selectivity.<sup>29,30</sup> Recently, a few biodegradable and non-toxic MOFs have been explored as alternative sorbents for food-related applications.<sup>31,32</sup> For example, cyclodextrin-based MOFs (CD-MOF-Na/K),<sup>32</sup> synthesized from bio-derived  $\alpha$ -cyclodextrins and alkali metal ions, have been proposed as effective C<sub>2</sub>H<sub>4</sub> sorbents for delaying the ripening of fruits under ambient conditions. These preliminary findings highlight the potential of MOFs in selective C<sub>2</sub>H<sub>4</sub> capture and pave the way for the identification of other existing MOFs with optimized C<sub>2</sub>H<sub>4</sub> sorption properties.

Our proposed computational workflow begins with a HTCS of a curated database of sustainable MOFs selected from the 128,799 experimentally synthesized MOFs in the latest release of the Cambridge Structure Database.<sup>1</sup> Following a rigorous MOF structure curation and geometry optimization pipeline, key adsorption descriptors, Henry's law constants ( $K_H$ ) and enthalpy of adsorption at infinite dilution ( $\Delta H_{0,ads}$ ) are

computed via Widom insertion method at room temperature for both C<sub>2</sub>H<sub>4</sub> and H<sub>2</sub>O as well as the resulting ideal selectivity ( $S_{C_2H_4/H_2O} = K_{H,C_2H_4}/K_{H,H_2O}$ ) using the UFF. This first conventional stage enables the identification of hydrophobic MOF candidates capable of selectively capturing ethylene over water. The Preferred Potential (PFP) u-MLIP, an equivariant graph neural network AI-model trained on a comprehensive dataset of 42 million DFT-calculated structures (including molecules, crystals, slabs, clusters and disordered systems),<sup>33</sup> is subsequently employed to refine the performance predictions of the 88 experimentally viable ethylene selective MOFs identified in the initial screening. Following its rigorous validation against DFT-calculated adsorption energies for water and ethylene across these 88 MOFs, the PFP u-MLIP is systematically applied to compute  $\Delta H_{0,ads}$  and  $K_H$  for both guest molecules and the corresponding  $S(C_2H_4/H_2O)$ . Notably, under the commonly adopted rigid-MOF framework approximation in HTCS workflows, the UFF approach demonstrates reliable performance in capturing the hydrophobic behaviour of most MOFs examined in this study, and in yielding ethylene adsorption enthalpies that are comparable to those predicted by the PFP u-MLIP model with a mean absolute deviation (MAD) of 5 kJ mol<sup>-1</sup>. The use of PFP u-MLIP becomes more critical for a small number of outlier MOFs featuring small pore sizes or polar functional groups, where UFF proves less effective at accurately capturing C<sub>2</sub>H<sub>4</sub> and H<sub>2</sub>O adsorption energetics, particularly under high confinement or when hydrogen-bonding interactions play a significant role. While UFF performs adequately for most cases, these outliers highlight the limitations of classical generic FFs and the advantages of higher-fidelity models like PFP u-MLIP. Notably, PFP u-MLIP results show that 99% of the MOFs exhibit ideal selectivity,  $S(C_2H_4/H_2O) > 1$ , confirming their strong ethylene preference and suitability for separation under humid conditions.

Although guest-triggered flexibility in MOFs is well documented and has been shown to significantly affect host-guest interactions in specific cases,<sup>34</sup> incorporating such structural responsiveness into HTCS workflows remains computationally prohibitive. Assessing whether the rigid-framework approximation introduces systematic deviations or remains within acceptable error margins is critical to ensuring the reliability of HTCS-derived adsorption properties. To this end, the PFP u-MLIP model was also employed to refine the predictions of  $\Delta H_{0,ads}(C_2H_4)$  for the 88 top MOF candidates, explicitly accounting for adsorption-induced framework flexibility, including atomic displacements alone or in combination with unit cell relaxations. It was found that for most MOFs exhibiting guest-induced unit cell volume changes below 10% led to only moderate deviations in  $\Delta H_{0,ads}(C_2H_4)$  relative to the rigid-framework approximation (mean absolute deviation (MAD) of 1.5 kJ mol<sup>-1</sup>). This refined step enables to exclude a few otherwise attractive MOFs that undergo significantly larger guest-induced unit cell expansion, accompanied by a notable decrease in C<sub>2</sub>H<sub>4</sub> affinity.

Beyond introducing a robust data-driven strategy for identifying MOFs suitable for ethylene removal under humid food storage conditions as a test case, this work proposes a generalized

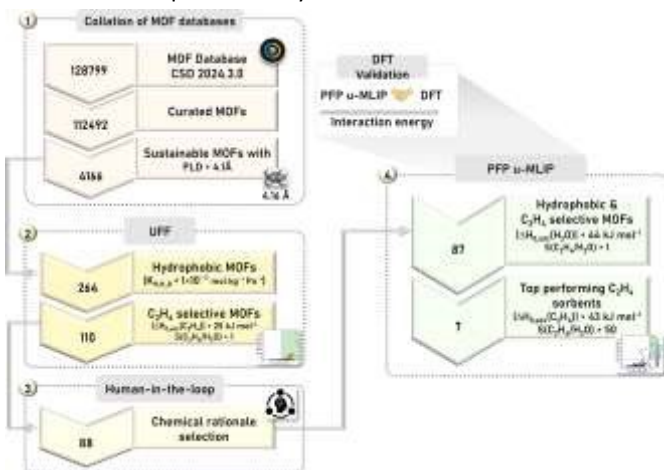


hybrid HTCS workflow that balances computational efficiency with predictive accuracy. We show that classical generic FF, UFF combined with rigid framework approximations provide a reliable and scalable first-pass filter to identify top-performing MOF candidates for sorption-based applications. For the top-ranked MOFs, we then leverage PFP u-MLIP to refine UFF-derived performance predictions, capturing host/guest interactions and structural flexibility more accurately. Our findings provide valuable insights into the transformative potential of u-MLIPs in HTCS, offering a powerful strategy to refine classical u-IPs predictions and to enable more accurate and resource-efficient screening pipelines for MOF-based sorption properties.

## Computational methodology

### Collation and curation of a sustainable MOFs structure database

A high-quality and structurally consistent MOFs database is a prerequisite for reliable HTCS of MOF properties.<sup>35</sup> As such, a systematic curation pipeline was implemented starting from the CSD MOF database (version 2024.3.0)<sup>1</sup> as outlined in Fig. 1. We cross-referenced the initial CSD entries with two well-established molecular simulations-ready MOFs: QMOF<sup>36</sup> and MOSAEC-DB.<sup>37</sup> To avoid redundant, time- and computation-intensive curation steps, we excluded CSD MOF entries that overlapped with QMOF (15,966) and/or MOSAEC-DB (71,138), retaining a total of 25,388 MOFs from the non-disordered subset of the CSD MOF database. These MOFs were subsequently curated using SAMOSA (Structural Activation via Metal Oxidation State Analysis) tool,<sup>38</sup> which removes free solvent molecules while preserving chemically relevant monodentate species. This yielded a combined baseline dataset



**Fig. 1** Schematic overview of the HTCS workflow for identifying hydrophobic and ethylene-selective MOF candidates from the CSD MOF database (version 2024.3.0). The numbered labels in the top-left corners of each panel indicate the sequential steps in the screening process: **(1)** collation and filtering of MOF structures (PLD > 4.1 Å); **(2)** Evaluation of hydrophobicity and  $\text{C}_2\text{H}_4/\text{H}_2\text{O}$  ideal selectivity using UFF-based Widom insertion Monte Carlo simulations; **(3)** Human-in-the-loop chemical rationale selection, yielding 88 promising MOFs; **(4)** PFP u-MLIP screening, identifying 87 hydrophobic MOFs [ $-\Delta H_{0,ads}(\text{H}_2\text{O}) < 44 \text{ kJ mol}^{-1}$ ] and confirming 7 top-performing sorbents with both high affinity [ $-\Delta H_{0,ads}(\text{C}_2\text{H}_4) > 43 \text{ kJ mol}^{-1}$ ] and high selectivity  $S(\text{C}_2\text{H}_4/\text{H}_2\text{O}) > 50$ ]. u-MLIP predictions were benchmarked against DFT adsorption energies as a supporting validation step to confirm model fidelity.

of 112,492 unique MOF structures for our HTCS pipeline. To ensure the accessibility of ethylene (kinetic diameter of 4.163 Å), Zeo++ software was used to filter MOFs with a pore-limiting diameter (PLD) > 4.1 Å. Additionally, MOFs were pre-selected based on their environmentally benign metal ions (Al, Zn, Hf, Fe or Ti) to align with sustainability and health safety standards relevant to food-related applications. Combining these geometric and sustainability criteria yielded a refined subset of 4,166 MOFs.

We then implemented a rigorous refinement process to ensure highest possible chemical correctness and to eliminate structure redundancy. First, we applied the Metal Oxidation State Automated Error Checker (MOSAEC)<sup>39</sup> to identify chemically inconsistent structures by analysing metal oxidation states. MOFs flagged as “BAD” by MOSAEC, typically due to unbalanced charges or chemically implausible valence assignments, were excluded. Next, we addressed structural redundancy by evaluating both topological and geometric similarity using the CoRE-MOF-Tools package.<sup>35</sup> Specifically, we employed the StructureMatcher module to identify duplicate or near-duplicate geometries. MOFs with charged frameworks were finally excluded, as these require additional considerations, such as counterion placement, that fall beyond the scope of this study. This multi-stage refinement process yielded a final dataset of 1,881 neutral and computation-ready MOFs that were geometry optimized using the EqV2-ODAC u-MLIP via BFGS optimization in Atomic Simulation Environment (ASE), with a force convergence threshold of 0.05 eV/Å.<sup>40</sup>

### Monte Carlo simulations

#### Classical generic FF Widom’s test particle insertion method

$K_H$  and  $\Delta H_{0,ads}$  for both water and ethylene were computed for all EqV2-ODAC u-MLIP optimized MOFs, employing Widom’s test particle insertion method at 298 K.<sup>41</sup> Ideal selectivity  $S(\text{C}_2\text{H}_4/\text{H}_2\text{O})$  was calculated as the ratio of  $K_H$  for ethylene to that for water. Each of these simulations consisted of  $2 \times 10^6$  production steps following  $1 \times 10^5$  equilibration steps, ensuring statistical convergence. In this calculation, the MOF frameworks were treated as rigid. MOF-guest interactions were modelled as the sum of van der Waals (Lennard-Jones-LJ) and Coulombic contributions. Long-range electrostatic interactions were treated via Ewald summation,<sup>42</sup> and a 12.0 Å cut-off was applied to the van der Waals terms. The UFF parameters<sup>5</sup> were used to define non-bonded terms of the MOF atoms. Atomic partial charges for MOF atoms were assigned using the MEPO-ML (MOF Electrostatic POTential–Machine Learned) model.<sup>43</sup> Ethylene and water molecules were described by a united atom potential model developed by Fisher et al.,<sup>44</sup> and the four center TIP4P-Ew model,<sup>45</sup> respectively. Site-site LJ interactions between unlike force field centers were treated using the Lorentz–Berthelot mixing rules. All these simulations were performed using the Complex Adsorption and Diffusion Simulation Suite (CADSS).<sup>46</sup>

#### PFP u-MLIP Widom’s insertion simulations

The PFP u-MLIP, version 6.0.0<sup>33</sup> was implemented in the Widom’s test particle insertion method to assess  $K_H$  and  $\Delta H_{0,ads}$



for both water and ethylene in the 88 top performing MOFs identified by UFF-based HTCS. To accurately account for dispersion interactions, the PBE-D3 correction<sup>47,48</sup> was applied via `CRYSTAL_U0_PLUS_D3` module, improving the description of the van der Waals interactions between MOFs and the guests. Each simulation was carried out for 50,000 MC cycles. To further investigate the impact of MOF framework flexibility on  $\Delta H_{0,ads}$  of ethylene, simulations were carried out using three distinct structural configurations for each of the 88 MOFs: (i)  $MOF_{reference}$  - structures optimized (atomic positions relaxed) in the absence of guest molecules using the EqV2-ODAC u-MLIP model; (ii)  $MOF_{geometry-opt}$  - structures with atomic positions relaxed in the presence of a single ethylene molecule using the PFP u-MLIP; (iii)  $MOF_{cell-opt}$  - structures with both atomic positions and unit cell parameters relaxed in the presence of a single ethylene molecule using the PFP u-MLIP. The initial structures for (ii) and (iii) were derived from the minimum energy ethylene adsorption configurations identified during the PFP u-MLIP based Widom insertion simulations. These geometry optimizations were performed using the BFGS optimizer<sup>49</sup> with a force convergence threshold of  $0.005 \text{ eV} \cdot \text{\AA}^{-1}$ . All Widom insertion and geometry optimization calculations using the PFP u-MLIP were performed via the MATLANTIS platform,<sup>50</sup> under an academic license. In this framework, the PFP u-MLIP is implemented through the `Estimator` module interfaced with the `ASECalculator` enabling seamless integration with the Atomic Simulation Environment (ASE) for energy and force evaluations.

#### PFP u-MLIP bulk modulus calculations

To evaluate the mechanical properties of the 88 MOFs, we computed their elastic response across an isotropic volumetric strain of  $\pm 8\%$  over 10 equidistant steps, utilizing the PFP u-MLIP. At each step, the lattice vectors were scaled to the target volume, and internal atomic positions were relaxed via the BFGS algorithm until the interatomic forces converged below  $0.01 \text{ eV}/\text{\AA}$ . Finally, the bulk modulus was computed by fitting the resulting energy-volume (E-V) data to the Vinet equation of state (EOS). To assess the reliability of the u-MLIP in describing framework mechanical stiffness, A520 was considered as a test-case, with calculated bulk modulus (4.9 GPa) in excellent agreement with available experimental data (5 GPa)<sup>51</sup>.

#### PFP u-MLIP benchmarking against DFT calculations

To validate the accuracy of the PFP u-MLIP used in this study for modelling MOF-guest interactions, specifically in the case of ethylene and water, the current release of the PFP u-MLIP model (version 6.0.0<sup>33</sup>) was rigorously benchmarked against 0 K DFT calculations. The analysis focused on ethylene and water adsorption in 88 top-performing MOFs (average system size of 366 atoms; **Fig. S2**), with guest positions taken from the lowest-energy adsorption configurations identified via PFP-based Widom insertion MC simulations on  $MOF_{reference}$ . Each resulting MOF+guest configuration, containing a single guest molecule per simulation box, was then subjected to PFP u-MLIP and DFT geometry optimization, with only the atomic positions relaxed. The MOF-guest interaction energies ( $\Delta E$ ) for ethylene and water were then evaluated using

the following expression:  $\Delta E = E(MOF + guest) - E(MOF) - E(guest)$

where,  $E(MOF + guest)$ , is the total energy of the geometry optimized MOF-guest configurations, while  $E(MOF)$  and  $E(guest)$  are the single-point energies of the MOF framework for which the guest molecule is removed and of the isolated guest molecule (ethylene or water), respectively. This approach was applied consistently across both PFP u-MLIP and DFT calculations to ensure a fair comparison.

DFT calculations were carried out using the Quickstep module in the CP2K package.<sup>52</sup> We used the Perdew-Burke-Ernzerhof (PBE)<sup>53</sup> exchange-correlation functional within the Gaussian plane-wave (GPW) formalism. Double-zeta valence polarized MOLOPT basis sets (DZVP-MOLOPT)<sup>54</sup> and Goedecker-Teter-Hutter (GTH) pseudopotentials<sup>55</sup> were used. Long-range dispersion interactions were included via Grimme's D3 correction<sup>47</sup>, consistent with the treatment employed in the PFP u-MLIP framework. A density cutoff of 450 Ry was applied and the geometry optimizations were performed until maximum residual forces were less than  $2 \times 10^{-4} \text{ au} \cdot \text{\AA}^{-1}$ .

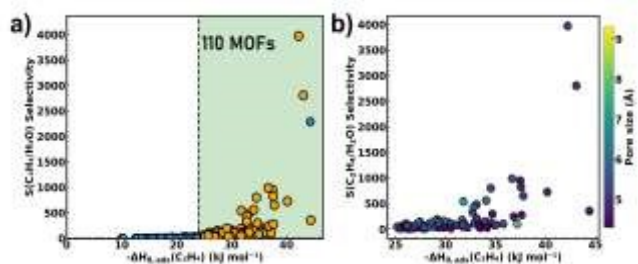
## Results and Discussion

### UFF-guided identification of MOFs for ethylene capture under humidity

Our multi-stage screening pipeline started with 1,881 charge-neutral, computation-ready MOFs, validated via MOSAEC and CoRE MOF protocols and geometry-optimized at the u-MLIP (EqV2-ODAC) level. The hydrophobic character of these MOFs was first assessed by computing their  $K_{H,H_2O}$  using Widom's test particle insertion MC simulations with UFF. Using ZIF-8 (a well-established benchmark highly hydrophobic MOF with  $K_{H,H_2O} \approx 2.5 \times 10^{-6} \text{ mol kg}^{-1} \text{ Pa}^{-1}$ ,<sup>56</sup>) as a reference, we applied a conservative threshold for  $K_{H,H_2O} = 1 \times 10^{-5} \text{ mol kg}^{-1} \text{ Pa}^{-1}$  to ensure the selection of a broader range of hydrophobic MOFs. This yielded to a subset of 264 hydrophobic MOFs suitable for ethylene capture in moisture-rich environments.

The ethylene/water selective adsorption of this subset of hydrophobic MOFs was then assessed by calculating their ideal selectivity values  $S(C_2H_4/H_2O) (=K_{H,C_2H_4}/K_{H,H_2O})$ . MOFs with  $S(C_2H_4/H_2O) > 1$  were considered as promising sorbents for ethylene adsorption over water. Additionally, to ensure practical relevance, we imposed a thermodynamic constraint, requiring the adsorption enthalpy of ethylene ( $\Delta H_{0,ads}(C_2H_4)$ ) to exceed  $25 \text{ kJ mol}^{-1}$  (in absolute value) significantly higher than ethylene's vaporization enthalpy ( $14 \text{ kJ mol}^{-1}$ ).<sup>57</sup> This criterion ensured strong host-guest interactions, critical for trace ethylene capture. These combined filtering, hydrophobicity,  $S(C_2H_4/H_2O) > 1$  and high  $C_2H_4$  affinity, narrowed the list of potential candidates to 110 high-performance MOFs. These materials, which collectively satisfy all three criteria, are highlighted in the shaded green region of **Fig. 2a**. To ensure practical applicability, we rigorously evaluated the synthetic feasibility of linkers in the 110 candidate MOFs using the MOFid toolkit<sup>58</sup>. Through systematic filtering and manual inspection





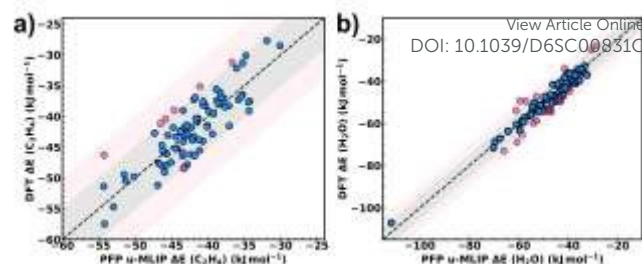
**Fig. 2** (a)  $S(\text{C}_2\text{H}_4/\text{H}_2\text{O})$ -negative adsorption enthalpy of ethylene ( $-\Delta H_{0,\text{ads}}(\text{C}_2\text{H}_4)$ ) plot across the dataset of 264 hydrophobic MOFs. The shaded green region highlights 110 MOFs that exhibit both high ethylene affinity ( $-\Delta H_{0,\text{ads}}(\text{C}_2\text{H}_4) > 25 \text{ kJ mol}^{-1}$ ) and  $S(\text{C}_2\text{H}_4/\text{H}_2\text{O}) > 1$ , the 88 identified viable MOFs being represented in orange colour; (b) Same plot for this selected sub-set of 88 MOFs alongside their pore sizes.

(**Fig. S1**), we excluded MOFs containing excessively long linkers (backbone lengths  $>3 \text{ nm}$ ), highly complex architectures including polycyclic or sterically hindered structures, and non-commercially available linkers. This comprehensive screening process eliminated 22 impractical candidates, yielding a final set of 88 synthetically viable MOFs highlighted in orange colour in **Fig. 2a**. **Fig. 2b** shows that these selected MOFs predominantly feature moderate pore sizes ranging from 4.5 to 6.0 Å underscoring the importance of pore confinement in promoting selective ethylene adsorption over water. To validate the reliability of this first-pass filter, UFF-predicted ethylene adsorption enthalpies were benchmarked against available experimental isosteric heats of adsorption from the literature, showing overall good agreement (**Table S1**). In addition, comparison of the UFF-based workflow with the DREIDING force field (**Fig. S3**) confirms that the relative screening trends are robust, with strong correlations observed for both ethylene adsorption enthalpy ( $R^2 = 0.89$ ) and ideal selectivity ( $R^2 = 0.96$ ), indicating very limited sensitivity to the specific choice of generic non-bonded LJ parameters.

Beyond the LJ parameters, we also assessed the sensitivity of the screening results to the electrostatic description of the MOF/guest interactions by comparing three machine-learning-based charge assignment schemes for MOF atoms: MEPO-ML, PACMAN,<sup>59</sup> and PACMOF.<sup>60</sup> As shown in **Fig. S4**, when combined with UFF, the three charge models lead to predicted  $S(\text{C}_2\text{H}_4/\text{H}_2\text{O})$  consistently yielding high pairwise correlation coefficients ( $R^2 > 0.94$ ), thereby demonstrating the robustness of our electrostatic description for performance ranking.

#### Benchmarking u-MLIP against DFT calculations

We first validated the applicability of PFP u-MLIP to our test case by benchmarking its predicted interaction energies ( $\Delta E$ ) for both  $\text{C}_2\text{H}_4$  and  $\text{H}_2\text{O}$  against the values obtained by DFT calculations across the representative set of 88 MOFs. **Fig. 3a** and **3b** emphasize an excellent agreement for both ethylene and water with very low MAD of  $2.4 \text{ kJ mol}^{-1}$  and  $3.0 \text{ kJ mol}^{-1}$  respectively. Notably, 91% of the investigated MOFs exhibit deviations within  $\pm 5 \text{ kJ mol}^{-1}$  for ethylene, while 78% MOFs fall within the same range for water. The remaining MOFs show moderate deviations of 5–10  $\text{kJ mol}^{-1}$ . This benchmarking also confirms PFP u-MLIP's quantitative reliability in modelling MOF–guest interactions for both polar ( $\text{H}_2\text{O}$ ) and non-polar ( $\text{C}_2\text{H}_4$ ) guests across diverse MOF frameworks.



**Fig. 3** Comparison between interaction energies of (a) ethylene and (b) water for a dataset of 88 MOFs. The dashed black line indicates the ideal parity line. Blue data points within the shaded grey region denote a mean absolute deviation of  $\pm 5 \text{ kJ mol}^{-1}$ , while pink data points within the light pink region indicate deviations between  $\pm 5$  and  $\pm 10 \text{ kJ mol}^{-1}$ .

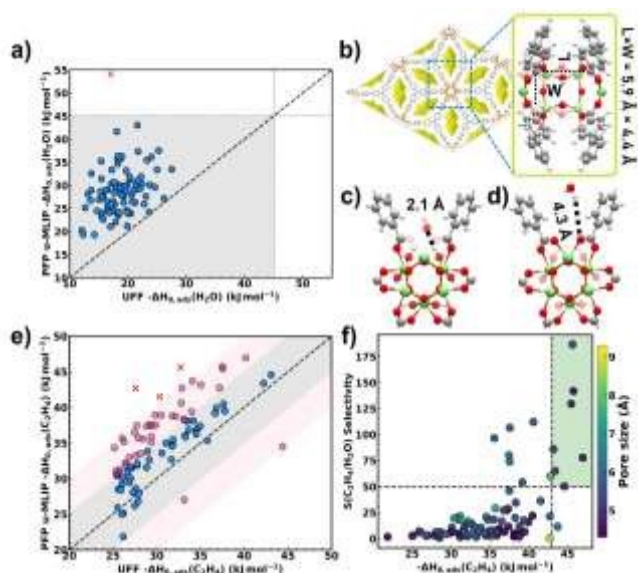
#### Comparison of classical generic FF and PFP u-MLIP adsorption predictions

To evaluate the reliability of the conclusions drawn from our initial UFF-based HTCS, we conducted a systematic comparison of adsorption energetics predicted by UFF and those calculated by PFP u-MLIP across the final set of 88 MOF structures. As a critical first validation step, we examined whether PFP u-MLIP confirms the hydrophobic character of these MOFs by computing their  $\Delta H_{0,\text{ads}}(\text{H}_2\text{O})$  under the same rigid MOF framework approximation applied in the initial UFF calculations. As shown in **Fig. 4a**, u-MLIP predicts  $\Delta H_{0,\text{ads}}(\text{H}_2\text{O})$  values below the water vaporization enthalpy threshold ( $\Delta H_{\text{vap}}(\text{H}_2\text{O}) \approx 44 \text{ kJ mol}^{-1}$ ) for 87 out of 88 MOFs, indicating weak water-MOF affinity and thereby supporting the hydrophobic classification originally established using UFF, which predicted all values below  $30 \text{ kJ mol}^{-1}$ . Although UFF does not quantitatively agree with the adsorption enthalpy values computed with PFP u-MLIP, it effectively captures the qualitative trend, demonstrating its utility in discriminating hydrophobic MOFs in large-scale screening efforts.

The single MOF outlier ZSTU-3<sup>61</sup> (CSD refcode TOYHOH; marked with a red cross in **Fig. 4a**) leads to a PFP u-MLIP calculated  $-\Delta H_{0,\text{ads}}(\text{H}_2\text{O})$  of  $\sim 54 \text{ kJ mol}^{-1}$ , which exceeds  $\Delta H_{\text{vap}}(\text{H}_2\text{O})$ , indicating significantly stronger water-MOF framework interactions compared to UFF predictions. The pore architecture of this MOF is shown in **Fig. 4b**, where polar  $-\text{OH}$  moieties are pointing towards a narrow pore space of  $5.9 \text{ \AA} \times 4.4 \text{ \AA}$ . Typically, this type of pore confinement plays a dominant role in enhancing water/polar functions interactions, an effect UFF fails to accurately capture, and would require tailored LJ parameters,<sup>62</sup> which is impractical in a HTCS workflow. Analysis of the lowest-energy adsorption configurations deduced from Widom insertion simulations confirms that PFP u-MLIP captures this confinement effect, placing water molecules as close as  $2.1 \text{ \AA}$  to framework atoms near the  $-\text{OH}$  moieties (**Fig. 4c**). In contrast, UFF predicts the nearest interaction at  $4.3 \text{ \AA}$ , occurring at carboxylate oxygen atoms (**Fig. 4d**).

Following the confirmation of the overall hydrophobic ranking of the MOF candidates, we evaluated their ethylene affinity using both UFF- and PFP u-MLIP-derived  $\Delta H_{0,\text{ads}}(\text{C}_2\text{H}_4)$ . **Fig. 4e** reveals a very good quantitative agreement between both values across all 88 MOFs, with a MAD of  $5.1 \text{ kJ mol}^{-1}$ . This comparative analysis highlights that 52% of the MOFs (blue





**Fig. 4** (a) Correlation between  $\Delta H_{0,\text{ads}}(\text{H}_2\text{O})$  values calculated using UFF and PFP u-MLIP. MOFs in the grey region exhibit hydrophobic behaviour with  $-\Delta H_{0,\text{ads}}(\text{H}_2\text{O}) < 44$   $\text{kJ mol}^{-1}$ ; (b) Crystal structure of the outlier MOF, ZSTU-3, with its pore architecture shown as a yellow surface. The inset shows a top view of the confined pore region where  $-\text{OH}$  groups are enclosed within a narrow pore window of dimensions  $L$  (length) = 5.9 Å and  $W$  (width) = 4.4 Å; (c, d) Water adsorption configurations in ZSTU-3 predicted by (c) u-MLIP, showing tight confinement (closest contact distance 2.1 Å), and (d) UFF, showing weaker interactions (about 4.3 Å); (e) Correlation between  $\Delta H_{0,\text{ads}}(\text{C}_2\text{H}_4)$  values from UFF and u-MLIP. Grey and pink shaded regions indicate deviations of  $< 5$  and 5–10  $\text{kJ mol}^{-1}$ , respectively; red crosses mark deviations  $> 10$   $\text{kJ mol}^{-1}$ ; (f) Correlation between  $S(\text{C}_2\text{H}_4/\text{H}_2\text{O})$  and  $-\Delta H_{0,\text{ads}}(\text{C}_2\text{H}_4)$  deduced from u-MLIP predictions. Data points are color-coded by pore size.

points) show an excellent correlation between u-MLIP and UFF calculations, with deviations below 5  $\text{kJ mol}^{-1}$ . Another 44% (pink points), show moderate deviations in the range of 5–10  $\text{kJ mol}^{-1}$ , while only three MOFs (3%, marked with red crosses), exhibit significant deviations greater than 10  $\text{kJ mol}^{-1}$ . A closer inspection of these three outlier MOFs: ZTUS-3,<sup>61, 63</sup> and Fe-CFA-6<sup>64</sup> reveals they all contain bridging  $\mu\text{-OH}$  sites confined within angular ('V-shaped') pockets defined by two adjacent linkers (Fig. S5). In the PFP u-MLIP simulations, ethylene molecules are positioned in such a way to favour short  $\pi(\text{C}_2\text{H}_4)\cdots\text{OH}$  contacts ( $\sim 2.5$  Å). In contrast, UFF does not capture this interaction mode, instead predicting different ethylene orientations and associated longer host–guest separating distances (3.0–4.1 Å). These differences in predicted geometries illustrate how the choice of potential model can substantially influence the representation of adsorption environments and associated interaction energies in confined pores. Similar to the adsorbed water geometries discussed above, analysis of the ethylene adsorption modes also revealed noticeable differences in the orientation of ethylene molecules between PFP u-MLIP and UFF, underscoring that classical generic force fields may overlook key directional interactions in spatially constrained environments.

We further evaluated the ethylene/water ideal selectivity,  $S(\text{C}_2\text{H}_4/\text{H}_2\text{O})$ , using  $K_{\text{H}_2\text{C}_2\text{H}_4}$  and  $K_{\text{H}_2\text{O}}$  values obtained from the PFP u-MLIP model (Fig. S6). The ranking of the top 99% ethylene selective MOFs remains largely consistent with those derived from UFF predictions, with 87 out of the 88 investigated MOFs

exhibiting  $S(\text{C}_2\text{H}_4/\text{H}_2\text{O}) > 1$ . ZSTU-3 was the sole exception, to exhibit a selectivity value of 0.2 arising from a rather hydrophilic behaviour predicted by PFP u-MLIP simulations as discussed above. In general, the absolute values of  $S(\text{C}_2\text{H}_4/\text{H}_2\text{O})$  predicted by the PFP u-MLIP (Fig. 4f) are consistently lower than those obtained using UFF (Fig. 2b) mostly due to a more accurate assessment of  $K_{\text{H}_2\text{O}}$  with PFP u-MLIP paving the way towards a more realistic estimation of the selectivity.

Further analysis of Fig. 4f enables to identify among these 87 MOFs, seven top performing candidates combining high  $S(\text{C}_2\text{H}_4/\text{H}_2\text{O}) > 50$  and strong ethylene affinity ( $-\Delta H_{0,\text{ads}}(\text{C}_2\text{H}_4) > 43$   $\text{kJ mol}^{-1}$ ). These MOFs are JNU-90 (CSD refcode: NOPGUY<sup>65</sup>), ZnPF-1 (CSD refcode: XEGXUF<sup>66</sup>), EVONOS<sup>67</sup>, A520 (CSD refcode: DOYBEA<sup>68</sup>), CYCU-5 (CSD refcode: HONCEU<sup>69</sup>), ATULIM<sup>70</sup>, and PERROW<sup>71</sup> (Fig. S7). Structurally, these MOFs typically possess microporous cavities in the range of 4.5–5.5 Å, as indicated by the pore size colour gradient represented in Fig. 4f. The combination of high ethylene/water selectivity and strong ethylene affinity positions these materials as promising candidates for targeted ethylene capture, particularly in low-trace applications such as food preservation.

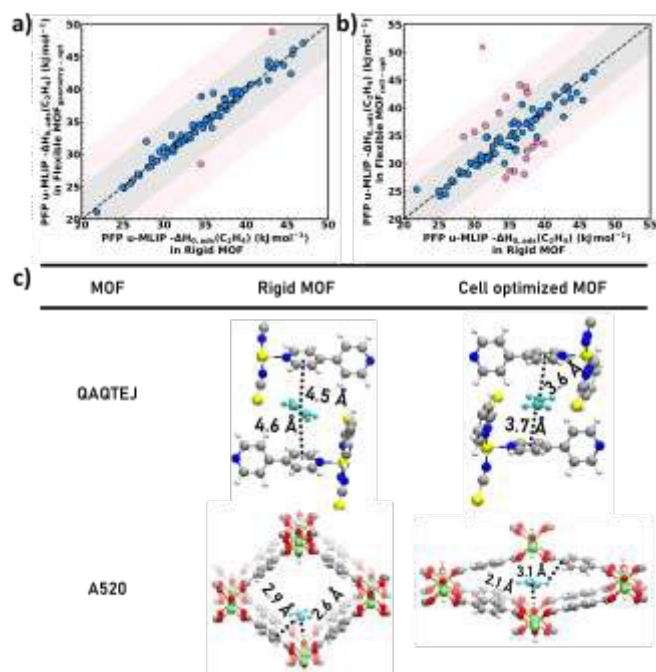
#### MOF framework flexibility and its ethylene affinity

Most HTCS workflows inherently assume rigid MOF frameworks, prompting a critical question: to what extent explicit consideration of framework flexibility is essential to refine the prediction of the adsorption properties? To address this question, we equally simulated  $\Delta H_{0,\text{ads}}(\text{C}_2\text{H}_4)$  for our shortlisted 88 MOFs via PFP u-MLIP based Widom insertion MC method, applied to geometry optimized MOFs, relaxed either by atomic positions alone or by both atomic positions and unit cell parameters.

The comprehensive analyses presented in Fig. 5a reveal that, for majority of the MOFs, the relaxation of their atomic positions exerts only a limited effect on ethylene adsorption energetics, largely preserving the overall ranking as predicted using the rigid MOF framework approximation. Specifically, when only atomic positions are relaxed, 98% of the MOFs (86 out of 88) exhibit  $\Delta H_{0,\text{ads}}(\text{C}_2\text{H}_4)$  variations within 5  $\text{kJ mol}^{-1}$ , with a very low MAD of 0.72  $\text{kJ mol}^{-1}$ . Only two structures show a more substantial deviation that remains however moderate in the range of 5–10  $\text{kJ mol}^{-1}$ . This supports those tiny localized atomic rearrangements have negligible influence on ethylene affinity for this MOF set.

In the case of full structure optimization, which includes both atomic positions and unit cell parameters relaxation, our u-MLIP calculations revealed that a significant fraction of the 88 MOFs shows a substantial unit cell volume change upon  $\text{C}_2\text{H}_4$  adsorption. To better understand the origin of these adsorption-induced structural responses, we evaluated the intrinsic mechanical stiffness of the 88 MOFs through bulk modulus calculations at the u-MLIP level (Fig. S8). The results show that larger adsorption-induced volume changes tend to occur preferentially in MOF frameworks with low to moderate bulk moduli as shown in Fig. 6, consistent with a higher ability for geometrical deformation in mechanically compliant MOFs.

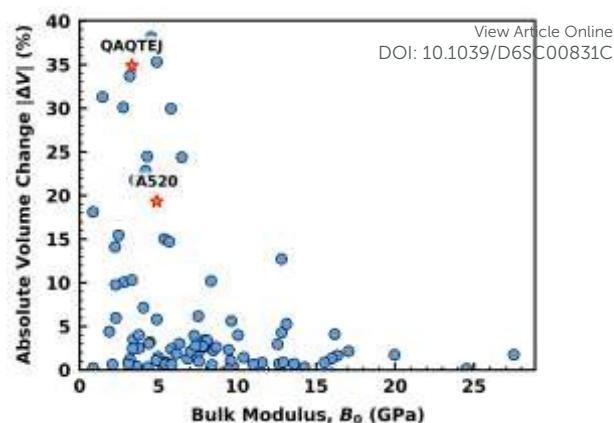




**Fig. 5** (a) Comparison of  $-\Delta H_{0,\text{ads}}(\text{C}_2\text{H}_4)$  computed with PFP u-MLIP for rigid versus geometry-optimized (atomic positions only) MOFs. The grey region denotes deviations within  $\pm 5 \text{ kJ mol}^{-1}$ , while the pink region corresponds to deviations between 5 and 10  $\text{kJ mol}^{-1}$ . (b) Comparison of  $-\Delta H_{0,\text{ads}}(\text{C}_2\text{H}_4)$  between rigid and fully optimized MOFs (both atomic positions and lattice parameters). The grey and pink regions represent deviations of  $\leq 5 \text{ kJ mol}^{-1}$  and 5–10  $\text{kJ mol}^{-1}$ , respectively. Red cross indicates sole outlier MOF with a deviation exceeding 10  $\text{kJ mol}^{-1}$ . (c) Ethylene adsorption snapshots of QAQTEJ and A520 in both rigid and fully unit cell-optimized configurations. Distances indicate the shortest framework–ethylene contacts.

**Fig. 5b** further shows that 80% of MOFs (70 out of 88) still show moderate  $\Delta H_{0,\text{ads}}(\text{C}_2\text{H}_4)$  changes within 5  $\text{kJ mol}^{-1}$ . 17 MOFs exhibit deviations of 5–10  $\text{kJ mol}^{-1}$ , and just one structure shows a deviation exceeding 10  $\text{kJ mol}^{-1}$ . Correspondingly, the MAD increases to 2.4  $\text{kJ mol}^{-1}$ , which however remains modest overall but clearly reflects the larger variability introduced by lattice flexibility for certain MOF frameworks. These results highlight the critical role of lattice flexibility in modulating host–guest interactions, particularly for MOFs whose structural adaptability significantly influences adsorption behaviour. Microscopic insight into these effects is provided in **Fig. 5c**, which compares rigid and fully cell-optimized configurations for the outlier MOF QAQTEJ<sup>72</sup> (red cross in **Fig. 5b**), a layered structure which displays a distinct flexibility mode: while its internal square nets remain largely intact, unit cell optimization triggers a lateral shift of the layers, reducing the interlayer spacing from 9 to 6 Å. This structural rearrangement results in 35% reduction in unit cell volume and increases  $-\Delta H_{0,\text{ads}}(\text{C}_2\text{H}_4)$  from 31 to 51  $\text{kJ mol}^{-1}$ . The enhancement in adsorption enthalpy can be attributed to increased confinement and stronger  $\pi\cdots\pi$  interactions between ethylene and the pyridine rings of the linkers, with contact distances shortening from 4.5 Å to 3.6 Å.

Notably, among the seven identified top-performing MOFs, six of them display small volume changes of 1–4% leading to a tiny variation of their adsorption enthalpy of ethylene (0.4–4  $\text{kJ mol}^{-1}$ ) (**Table S2**). In addition, while A520 (CSD refcode: DOYBEA<sup>68</sup>) undergoes a substantial relative volume variation of



**Fig. 6** Correlation between  $\text{C}_2\text{H}_4$ -induced absolute unit cell volume change ( $|\Delta V|$ %) and bulk modulus ( $B_0$ ) predicted for the 88 MOFs by PFP u-MLIP.

19%, yet the change in adsorption enthalpy is only 0.39  $\text{kJ mol}^{-1}$ . In the relaxed state, the framework retains nearly identical host–guest geometries, with  $-\text{OH}\cdots\pi$  interactions (2.6  $\rightarrow$  2.1 Å) and  $\text{C}-\text{H}\cdots\pi$  contacts (2.9  $\rightarrow$  3.1 Å) showing only marginal adjustments, as illustrated in **Fig. 5c**.

## Conclusions

This work presents a scalable, multi-scale HTCS framework that couples the speed of classical force fields with the quantum-level accuracy of universal machine learning potentials, validated against DFT, for the identification of high-performance MOF adsorbents. This strategy uses ethylene capture under humid conditions as a challenging test case, which is highly relevant to food-preservation packaging technologies. We demonstrate how classical u-IPs and u-MLIPs can be synergistically integrated to achieve both computational efficiency and accurate predictions for a large library of MOFs. Specifically, our results show that classical u-IP methods like UFF-based HTCS remain a robust first-pass filter, correctly capturing adsorption trends in  $\sim 97$ – $99\%$  of cases under the rigid-MOF framework approximation. u-MLIPs are essential for further refining the adsorption predictions for the top-performing candidates, especially for MOFs involving strong hydrogen bonding, directional  $\pi\cdots\text{H}/\pi\cdots\pi$  interactions, or guest-induced framework flexibility, where deviations in  $\Delta H_{0,\text{ads}}(\text{C}_2\text{H}_4)$  can reach  $\sim 20 \text{ kJ mol}^{-1}$ . This workflow enables the identification of top MOF performers combining microporous confinement (4.5–6.0 Å), strong ethylene affinity ( $\Delta H_{0,\text{ads}}(\text{C}_2\text{H}_4) > 43 \text{ kJ mol}^{-1}$ ), and high selectivity ( $S(\text{C}_2\text{H}_4/\text{H}_2\text{O}) > 50$ ). These computational findings are expected to prioritize MOFs for experimental validation and accelerate their integration into scalable postharvest ethylene management technologies. Beyond ethylene capture, the proposed hybrid classical generic FF/u-MLIP strategy offers a transferable computational workflow for other adsorption-driven separations, enabling the screening of large structural databases while retaining the resolution needed to treat complex host–guest interactions and framework flexibility.



## Author contributions

G.M. conceived the idea and supervised the project. S.B. performed and interpreted the computational calculations. M.W. developed the u-MLIP-based Widom simulation code and contributed to the analysis of the simulated data. S.B. wrote the original draft with input from M.W. and G.M. T.W. and K.H. provided support with the MATLANTIS framework and contributed to manuscript improvement.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

The code supporting the PFP u-MLIP-based Widom simulations and the data (Excel file) generated in this study are available on GitHub (<https://github.com/gmaurin-group/MLP-WIDOM-SIM>).

## Acknowledgements

The computational work was performed using HPC resources from GENCI-CINES (Grant A0180907613) and MATLANTIS. G.M. thanks Institut Universitaire de France for the Senior Chair.

## Notes and references

- 1 P. Z. Moghadam, A. Li, S. B. Wiggin, A. Tao, A. G. P. Maloney, P. A. Wood, S. C. Ward and D. Fairen-Jimenez, *Chemistry of Materials*, 2017, **29**, 2618–2625.
- 2 H. Daglar and S. Keskin, *Coord. Chem. Rev.*, 2020, **422**, 213470.
- 3 Y. Qiu, L. Wang, C. Liu, X. Zhang, Y. Tian and Z. Zhou, *Materials Today*, DOI: <https://doi.org/10.1016/j.mattod.2025.10.017>.
- 4 S. Ga, N. An, G. Y. Lee, C. Joo and J. Kim, *Renewable and Sustainable Energy Reviews*, 2024, **192**, 114275.
- 5 T. R. Cundary and M. S. Gordon, *UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations*, 1992, vol. 114.
- 6 M. Fischer, J. R. B. Gomes, M. Fröba and M. Jorge, *Langmuir*, 2012, **28**, 8537–8549.
- 7 L.-C. Lin, K. Lee, L. Gagliardi, J. B. Neaton and B. Smit, *J. Chem. Theory Comput.*, 2014, **10**, 1477–1488.
- 8 A. L. Dzubak, L.-C. Lin, J. Kim, J. A. Swisher, R. Poloni, S. N. Maximoff, B. Smit and L. Gagliardi, *Nat. Chem.*, 2012, **4**, 810–816.

- 9 A. R. Kulkarni and D. S. Sholl, *The Journal of Physical Chemistry C*, 2016, **120**, 23044–23054. View Article Online DOI: 10.1039/D6SC00831C
- 10 J. W. Yoon, H. Chang, S.-J. Lee, Y. K. Hwang, D.-Y. Hong, S.-K. Lee, J. S. Lee, S. Jang, T.-U. Yoon, K. Kwac, Y. Jung, R. S. Pillai, F. Faucher, A. Vimont, M. Daturi, G. Férey, C. Serre, G. Maurin, Y.-S. Bae and J.-S. Chang, *Nat. Mater.*, 2017, **16**, 526–531.
- 11 E. Martínez-Ahumada, D. won Kim, M. Wahiduzzaman, P. Carmona-Monroy, A. López-Olvera, D. R. Williams, V. Martis, H. A. Lara-García, S. López-Morales, D. Solis-Ibarra, G. Maurin, I. A. Ibarra and C. S. Hong, *J. Mater. Chem. A Mater.*, 2022, **10**, 18636–18643.
- 12 H. Kraß, J. Huang and S. M. Moosavi, *MOFSimBench: Evaluating Universal Machine Learning Interatomic Potentials In Metal--Organic Framework Molecular Modeling*, 2025.
- 13 R. Goeminne and V. Van Speybroeck, *J. Am. Chem. Soc.*, 2025, **147**, 3615–3630.
- 14 D. Fan, F. L. Oliveira, M. Wahiduzzaman and G. Maurin, *Decoding local framework dynamics in the ultra-small pore MOF MIL-120(AI) CO<sub>2</sub> sorbent with Machine Learned Potentials*, 2025.
- 15 S. Liu, R. Dupuis, D. Fan, S. Benzaria, M. Bonneau, P. Bhatt, M. Eddaoudi and G. Maurin, *Chem. Sci.*, 2024, **15**, 5294–5302.
- 16 I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, M. Avaylon, W. J. Baldwin, F. Berger, N. Bernstein, A. Bhowmik, F. Bigi, S. M. Blau, V. Cărare, M. Ceriotti, S. Chong, J. P. Darby, S. De, F. Della Pia, V. L. Deringer, R. Elijošius, Z. El-Machachi, E. Fako, F. Falcioni, A. C. Ferrari, J. L. A. Gardner, M. J. Gawkowski, A. Genreith-Schriever, J. George, R. E. A. Goodall, J. Grandel, C. P. Grey, P. Grigorev, S. Han, W. Handley, H. H. Heenen, K. Hermansson, C. H. Ho, S. Hofmann, C. Holm, J. Jaafar, K. S. Jakob, H. Jung, V. Kapil, A. D. Kaplan, N. Karimitari, J. R. Kermode, P. Kourtis, N. Kroupa, J. Kullgren, M. C. Kuner, D. Kuryla, G. Liepuoniute, C. Lin, J. T. Margraf, I.-B. Magdău, A. Michaelides, J. H. Moore, A. A. Naik, S. P. Niblett, S. W. Norwood, N. O'Neill, C. Ortner, K. A. Persson, K. Reuter, A. S. Rosen, L. A. M. Rosset, L. L. Schaaf, C. Schran, B. X. Shi, E. Sivonxay, T. K. Stenczel, C. Sutton, V. Svahn, T. D. Swinburne, J. Tilly, C. van der Oord, S. Vargas, E. Varga-Umbrich, T. Vegge, M. Vondrák, Y. Wang, W. C. Witt, T. Wolf, F. Zills and G. Csányi, *J. Chem. Phys.*, 2025, **163**, 184110.
- 17 C. Chen and S. P. Ong, *Nat. Comput. Sci.*, 2022, **2**, 718–728.



## Journal Name

## ARTICLE

- 18 B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel and G. Ceder, *Nat. Mach. Intell.*, 2023, **5**, 1031–1041. Charoenphakdee and T. Ibuka, *Nat. Commun.*, 2022, **13**, 2991. View Article Online  
DOI: 10.1039/D6SC00831C
- 19 A. Sriram, L. M. Brabson, X. Yu, S. Choi, K. Abdelmaqsoud, E. Moubarak, P. de Haan, S. Löwe, J. Brehmer, J. R. Kitchin, M. Welling, C. L. Zitnick, Z. Ulissi, A. J. Medford and D. S. Sholl, 2025, preprint, <https://arxiv.org/abs/2508.03162>. 34 L. M. Brabson, A. J. Medford and D. S. Sholl, *The Journal of Physical Chemistry C*, 2025, **129**, 16811–16825.
- 20 A. Sriram, S. Choi, X. Yu, L. M. Brabson, A. Das, Z. Ulissi, M. Uyttendaele, A. J. Medford and D. S. Sholl, *ACS Cent. Sci.*, 2024, **10**, 923–941. 35 X. Jin, K. M. Jablonka, E. Moubarak, Y. Li and B. Smit, *Digital Discovery*, 2025, **4**, 1560–1569.
- 21 X. Fu, B. M. Wood, L. Barroso-Luque, D. S. Levine, M. Gao, M. Dzamba and C. L. Zitnick, DOI:10.48550/arxiv.2502.12147. 36 A. S. Rosen, S. M. Iyer, D. Ray, Z. Yao, A. Aspuru-Guzik, L. Gagliardi, J. M. Notestein and R. Q. Snurr, *Matter*, 2021, **4**, 1578–1597.
- 22 Y. Lim, H. Park, A. Walsh and J. Kim, *Matter*, 2025, **8**, 102203. 37 M. Gibaldi, A. Kapeliukha, A. White, J. Luo, R. A. Mayo, J. Burner and T. K. Woo, *Chem. Sci.*, 2025, **16**, 4085–4100.
- 23 C. S. Barry and J. J. Giovannoni, *J. Plant Growth Regul.*, 2007, **26**, 143–159. 38 M. Gibaldi, A. Kapeliukha, A. White and T. K. Woo, *J. Chem. Inf. Model.*, 2025, **65**, 275–287.
- 24 N. Y. State, C. B. Watkins, J. F. Nock and J. Reid, *Production Guide for 2012 Production Guide for Storage of Organic Fruits and Vegetables Funded through a grant from the New York State Department of Agriculture and Markets*, . 39 M. Gibaldi, A. Kapeliukha, A. White, J. Luo, R. A. Mayo, J. Burner and T. K. Woo, *Chem. Sci.*, 2025, **16**, 4085–4100.
- 25 J. Gustavsson, C. Cederberg, U. Sonesson, R. van Otterdijk and A. Meybeck, *Food and Agriculture Organization of the United Nations*, 2011, preprint. 40 A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Duřak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, *Journal of Physics: Condensed Matter*, 2017, **29**, 273002.
- 26 L.-Y. Chang and J. K. Brecht, *Sci. Hortic.*, 2023, **309**, 111636. 41 D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press, San Diego, 2nd edn., 2002, vol. 1.
- 27 M. H. Álvarez-Hernández, G. B. Martínez-Hernández, F. Avalos-Belmontes, F. D. Miranda-Molina and F. Artés-Hernández, *Postharvest Biol. Technol.*, 2020, **160**, 111061. 42 P. P. Ewald, *Ann. Phys.*, 1921, **369**, 253–287.
- 28 H. Li, F. Li, L. Wang, J. Sheng, Z. Xin, L. Zhao, H. Xiao, Y. Zheng and Q. Hu, *Food Chem.*, 2009, **114**, 547–552. 43 J. Luo, O. Ben Said, P. Xie, M. Gibaldi, J. Burner, C. Pereira and T. K. Woo, *NPJ Comput. Mater.*, 2024, **10**, 224.
- 29 W. You, Y. Liu, J. D. Howe and D. S. Sholl, *Journal of Physical Chemistry C*, 2018, **122**, 8960–8966. 44 M. Fischer, F. Hoffmann and M. Fröba, *ChemPhysChem*, 2010, **11**, 2220–2229.
- 30 Y. Qi, H. Yang, C. Li and H. Li, *Sep. Purif. Technol.*, 2024, **347**, 127634. 45 C. Vega, J. L. F. Abascal and I. Nezbeda, *J. Chem. Phys.*, 2006, **125**, 034503.
- 31 Z. Zhang, R. Zhou, L. Ke, J. Li, H. Jayan, H. R. El-Seedi, X. Zou and Z. Guo, *Trends Food Sci. Technol.*, 2024, **154**, 104771. 46 Q. Yang and C. Zhong, *J. Phys. Chem. B*, 2006, **110**, 17776–17783.
- 32 H. Li, L. Shi, C. Li, X. Fu, Q. Huang and B. Zhang, *ACS Appl. Mater. Interfaces*, 2020, **12**, 34095–34104. 47 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 33 S. Takamoto, C. Shinagawa, D. Motoki, K. Nakago, W. Li, I. Kurata, T. Watanabe, Y. Yayama, H. Iriguchi, Y. Asano, T. Onodera, T. Ishii, T. Kudo, H. Ono, R. Sawada, R. Ishitani, M. Ong, T. Yamaguchi, T. Kataoka, A. Hayashi, N. 48 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.



## ARTICLE

## Journal Name

- 49 R. Battiti and F. Masulli, in *International Neural Network Conference: July 9–13, 1990 Palais Des Congres — Paris — France*, Springer Netherlands, Dordrecht, 1990, pp. 757–760.
- 50 Preferred Computational Chemistry Inc., preprint, <https://matlantis.com/>.
- 51 P. Ramaswamy, J. Wieme, E. Alvarez, L. Vanduyfhuys, J.-P. Itié, P. Fabry, V. Van Speybroeck, C. Serre, Pascal. G. Yot and G. Maurin, *J. Mater. Chem. A Mater.*, 2017, **5**, 11047–11054.
- 52 J. Hutter, M. Iannuzzi, F. Schiffmann and J. VandeVondele, *WIREs Computational Molecular Science*, 2014, **4**, 15–25.
- 53 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
- 54 J. VandeVondele and J. Hutter, *J. Chem. Phys.*, 2007, **127**, 114105.
- 55 S. Goedecker, M. Teter and J. Hutter, *Phys. Rev. B*, 1996, **54**, 1703–1710.
- 56 G. Ercakir, G. O. Aksu, C. Altintas and S. Keskin, *ACS Engineering Au*, 2023, **3**, 488–497.
- 57 R. M. Stephenson and S. Malanowski, *Handbook of the Thermodynamics of Organic Compounds*, Springer Dordrecht, 1st edn., 1987.
- 58 B. J. Bucior, A. S. Rosen, M. Haranczyk, Z. Yao, M. E. Ziebel, O. K. Farha, J. T. Hupp, J. I. Siepmann, A. Aspuru-Guzik and R. Q. Snurr, *Cryst. Growth Des.*, 2019, **19**, 6682–6697.
- 59 G. Zhao and Y. G. Chung, *J. Chem. Theory Comput.*, 2024, **20**, 5368–5380.
- 60 S. Kancharlapalli, A. Gopalan, M. Haranczyk and R. Q. Snurr, *J. Chem. Theory Comput.*, 2021, **17**, 3052–3064.
- 61 C. Li, H. Xu, J. Gao, W. Du, L. Shangguan, X. Zhang, R.-B. Lin, H. Wu, W. Zhou, X. Liu, J. Yao and B. Chen, *J. Mater. Chem. A Mater.*, 2019, **7**, 11928–11933.
- 62 J. Shen, A. Kumar, M. Wahiduzzaman, D. Barpaga, G. Maurin and R. K. Motkuri, *Chem. Rev.*, 2024, **124**, 7619–7673.
- 63 E. Alvarez, N. Guillou, C. Martineau, B. Bueken, B. Van de Voorde, C. Le Guillouzer, P. Fabry, F. Nouar, F. Taulelle, D. de Vos, J.-S. Chang, K. H. Cho, N. Ramsahye, T. Devic, M. Daturi, G. Maurin and C. Serre, *Angew. Chem. Int. Ed.*, 2015, **54**, 3664–3668.
- 64 S. Spirk, M. Grzywa, C. S. Zehe, J. Senker, S. Demeshko, F. Meyer, S. Riegg and D. Volkmer, *CrystEngComm*, 2019, **19**, 313–322.
- 65 Y. Wang, Z.-J. Jiang, D.-R. Wang, W. Lu and D. Li, *J. Am. Chem. Soc.*, 2024, **146**, 6955–6961.
- 66 C. Castillo-Blas, V. A. de la Peña-O’Shea, I. Puente-Orench, J. R. de Paz, R. Sáez-Puche, E. Gutiérrez-Puebla, F. Gándara and Á. Monge, *Sci. Adv.*, 2025, **3**, e1700773.
- 67 Y. Xu, R.-H. Wang, B.-Y. Lou, L. Han and M.-C. Hong, *Acta Crystallographica Section C*, 2004, **60**, m296–m298.
- 68 E. Alvarez, N. Guillou, C. Martineau, B. Bueken, B. Van de Voorde, C. Le Guillouzer, P. Fabry, F. Nouar, F. Taulelle, D. de Vos, J.-S. Chang, K. H. Cho, N. Ramsahye, T. Devic, M. Daturi, G. Maurin and C. Serre, *Angew. Chem. Int. Ed.*, 2015, **54**, 3664–3668.
- 69 C.-Y. Wu, D. S. Raja, C.-C. Yang, C.-T. Yeh, Y.-R. Chen, C.-Y. Li, B.-T. Ko and C.-H. Lin, *CrystEngComm*, 2014, **16**, 9308–9319.
- 70 M. Armaghan, X.-J. Shang, Y.-Q. Yuan, D. J. Young, W.-H. Zhang, T. S. A. Hor and J.-P. Lang, *Chempluschem*, 2015, **80**, 1231–1234.
- 71 X.-G. Liu, C.-L. Tao, H.-Q. Yu, B. Chen, Z. Liu, G.-P. Zhu, Z. Zhao, L. Shen and B. Z. Tang, *J. Mater. Chem. C Mater.*, 2018, **6**, 2983–2988.
- 72 C. J. Adams, M. C. Muñoz, R. E. Waddington and J. A. Real, *Inorg. Chem.*, 2011, **50**, 10633–10642.



### Data availability

The code supporting the PFP u-MLIP-based Widom simulations and the data (Excel file) generated in this study are available on GitHub (<https://github.com/gmaurin-group/MLP-WIDOM-SIM>).

