

PCCP

Physical Chemistry Chemical Physics

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: J. Avelar, M. Rivera-Almazo, R. Vargas and J. Garza, *Phys. Chem. Chem. Phys.*, 2026, DOI: 10.1039/D6CP00919K.



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.

Cite this: DOI: 00.0000/xxxxxxxxxx

Systematic exploration of host–guest potential energy surfaces in metal–organic frameworks[†]

Joana Avelar^{*a}, Marcos Rivera-Almazo^b, Rubicelia Vargas^a and Jorge Garza^{*a}Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

We present a computational framework for the systematic exploration of high-dimensional potential energy surfaces associated with host–guest interactions in porous materials. As case studies, the adsorption of benzyl alcohol, benzene, benzaldehyde, CO, CO₂, H₂S, and methanol in the metal–organic frameworks MFM-300(Sc) and MFM-300(In) was investigated. A stochastic sampling strategy generated thousands of candidate configurations, which were first screened at the semiempirical PM7 level and subsequently refined using dispersion-corrected density functional theory. To identify representative low-energy structures while preserving configurational diversity, a data-driven selection procedure based on outlier detection and unsupervised clustering of energetic and geometric descriptors was implemented. The exploration of approximately 39 000 configurations reveals that the adsorption landscapes are characterized by multiple low-lying minima within narrow energy windows, particularly for weakly interacting guest molecules. This behavior reflects the complex and highly corrugated potential energy surfaces associated with confinement in MOF pores and highlights the limitations of approaches based on chemically intuitive initial guesses or single optimized structures. The multilevel protocol provides a statistically robust and computationally efficient strategy for identifying adsorption sites and representative configurations in porous materials, and offers a transferable framework for exploring configurational landscapes in host–guest systems, heterogeneous interfaces, and other complex condensed-phase environments.

Introduction

Porous chemical systems have gained significant attention due to their central role in gas storage, molecular separations, catalysis, and energy-related applications. A wide range of materials exhibit permanent porosity, including zeolites,^{1,2} mesoporous silicas such as MCM-41,^{3,4} porous carbons,^{5,6} metal–organic frameworks (MOFs),^{7,8} and polymers of intrinsic microporosity (PIMs).^{9,10} These platforms differ in composition, pore topology, surface chemistry, and structural flexibility, which determine their adsorption capacity and selectivity toward specific guest molecules. Among these porous platforms, Metal–Organic Frameworks (MOFs) have attracted particular interest due to their exceptional structural diversity and synthetic modularity^{11,12}.

Metal–organic frameworks are a class of hybrid crystalline porous materials composed of inorganic metal ions or clusters interconnected by rigid organic linkers bearing multiple coordination sites. The resulting frameworks are stabilized through co-

ordination bonds between the metal nodes and organic ligands, often reinforced by Coulombic interactions^{8,11–14}. Owing to their high crystallinity, exceptional porosity, and structural tunability, MOFs enable precise control over both physical and chemical properties. This remarkable versatility has driven extensive research into their use across a broad range of applications, including catalysis, gas adsorption and separation, energy harvesting and storage, luminescence, sensing, drug delivery, biomedicine, magnetism, and light-emitting diodes (LEDs)^{15,16}.

Due to their astonishing surface area, consequence of their porous nature, the role of MOFs in gas and molecular capture and storage has been extensively studied. Furthermore, the strategic design and functionalization of MOFs allow for the selective and efficient adsorption of hazardous and industrially relevant gases, thereby contributing to environmental sustainability and process efficiency. For example, CO₂ capture is critical for mitigating greenhouse gas emissions, and frameworks such as MUF-16 have demonstrated exceptional CO₂ selectivity over hydrocarbons, making them suitable for natural-gas purification and feedstock refinement¹⁷. Likewise, the removal of toxic gases such as H₂S is essential to prevent corrosion and catalyst poisoning in industrial settings. MOFs including Mg-MOF-74 and MIL-101(Cr) have shown notable selectivity for H₂S over CO₂, exhibiting reversible adsorption behavior and excellent structural stabil-

^a Departamento de Química, División de Ciencias Básicas e Ingeniería, Universidad Autónoma Metropolitana-Iztapalapa, San Rafael Atlixco 186, Col. Vicentina, Iztapalapa, C.P. 09340, Ciudad de México, México. E-mail: joar@xanum.uam.mx; jgo@xanum.uam.mx

^b Instituto de Física, Universidad Nacional Autónoma de México, Coyoacán, C.P. 01000, Ciudad de México, México.



ity^{18–21}. The π - π stacking interactions and hydrophobic pore environments within MOFs promote the selective uptake of aromatic pollutants such as benzene and benzaldehyde, compounds associated with carcinogenicity and respiratory health risks²². In parallel, the presence of polar functional groups in certain MOFs facilitates the adsorption of polar volatile organic compounds (VOCs) such as methanol and benzyl alcohol, thereby mitigating occupational exposure in industrial environments¹⁵. In this context, research on the capture of these molecules is essential for advancing environmental protection, ensuring industrial safety, and optimizing chemical processing.

MFM-300(Sc) and MFM-300(In) were selected as model systems because they are isostructural frameworks that provide closely related pore environments while differing in the nature of the metal center. Both materials contain μ_2 -OH groups lining the pore walls, which act as well-defined interaction sites capable of establishing hydrogen-bonding and electrostatic contacts with guest molecules. In addition, the MFM-300 family has been reported to exhibit dynamic metal–ligand behavior, in which hemilabile coordination between the metal centers and carboxylate ligands may promote the reversible formation of semi-open metal sites upon guest adsorption^{23,24}. These structural features make MFM-300(Sc) and MFM-300(In) suitable platforms for studying adsorption modes governed by different types of host–guest interactions, ranging from weak dispersive contacts in small gases such as CO and CO₂, to hydrogen-bonding and electrostatic interactions in polar molecules such as methanol and benzyl alcohol, and π -related interactions in aromatic guests such as benzene and benzaldehyde. Moreover, the experimental characterization available for these frameworks and related MFM-300 materials provides a useful reference for benchmarking computational methodologies. For these reasons, MFM-300(Sc) and MFM-300(In) constitute appropriate model systems for assessing a systematic protocol aimed at exploring complex MOF–guest potential energy surfaces.

The efficiency of gas capture in MOFs is primarily determined by the strength and nature of host–guest interactions within their porous frameworks.^{25,26} Computational modeling plays a central role in elucidating the structural features of these complexes and quantifying interaction energies with sufficient accuracy to rationalize experimental observations.^{27,28}

A central challenge in the computational investigation of MOF–guest systems is the reliable identification of low-energy adsorption configurations on highly multidimensional potential energy surfaces. Chemically intuitive initial guesses are frequently employed^{29–31}; however, such approaches do not guarantee identification of the global minimum or a representative ensemble of metastable states. Consequently, systematic exploration of the configurational landscape is required to obtain a statistically meaningful description of host–guest interactions.

Given the structural diversity and size of MOFs, simulations must address multiple length and time scales. For large systems, force-field-based atomistic simulations provide computational efficiency,³² albeit with limited accuracy in describing detailed chemical interactions. Conversely, electronic-structure methods offer greater fidelity in locating adsorption sites and characteriz-

ing host–guest interactions, though at significantly higher computational cost and typically under low guest loadings.³³ These approaches can be combined within multiscale frameworks. In parallel, the integration of machine-learning techniques with high-throughput data generation has enabled rapid screening of experimental and hypothetical MOFs, using descriptors such as adsorption energies, geometric parameters, electrostatic potentials, polarizability, and electronic features to predict adsorption performance and reactivity.^{27,34–37}

In this work, we introduce a computational framework for the unbiased exploration of complex potential energy surfaces and the systematic identification of low-energy configurations in high-dimensional configurational spaces. Rather than relying on chemically intuitive initial guesses, the protocol combines stochastic configurational sampling, descriptor-driven machine-learning selection, and hierarchical density functional theory (DFT) refinement to efficiently navigate large structural ensembles while preserving both energetic and geometric diversity.

Although the framework is demonstrated here through its application to MFM-300(Sc) and MFM-300(In) interacting with representative guest molecules—including alcohols, benzene, benzaldehyde, CO, CO₂, and H₂S—the conceptual foundation of the methodology extends well beyond MOF–guest systems. Its design is inherently general and can be readily adapted to heterogeneous interfaces, catalytic active sites, supramolecular assemblies, molecular clusters, and condensed-phase environments where a statistically robust and unbiased exploration of the configurational landscape is essential.

More broadly, the present work establishes a reproducible, descriptor-driven, and statistically informed strategy for systematic potential energy surface exploration. By combining extensive sampling with hierarchical electronic-structure refinement, the protocol provides a transferable framework capable of supporting data-driven materials discovery, adsorption studies, and mechanistic investigations across a wide spectrum of problems in computational chemistry and materials science.

Methodology

The primary objective of this study is to identify the most favorable initial geometries for the interactions between the MOFs MFM-300(Sc) and MFM-300(In) and a set of guest molecules—benzyl alcohol, benzene, benzaldehyde, CO, CO₂, H₂S, and methanol.

To accomplish this, we explored the potential energy surface (PES) of each MOF–guest system using a stochastic sampling strategy. In particular, the Restricted Isomers Searching by Simulated Annealing (RISSA) code^{38,39} was employed to efficiently sample the configurational space and generate thousands of candidate MOF–guest configurations. Within RISSA, the relative energy of each generated configuration is evaluated using the self-consistent field (SCF) energy obtained at the semiempirical PM7 level of theory, providing a computationally efficient estimate of the energetics during the sampling process⁴⁰.

Although RISSA has been previously used for molecular and cluster-like systems, the code also includes different spatial constraints for configurational sampling, namely confinement within



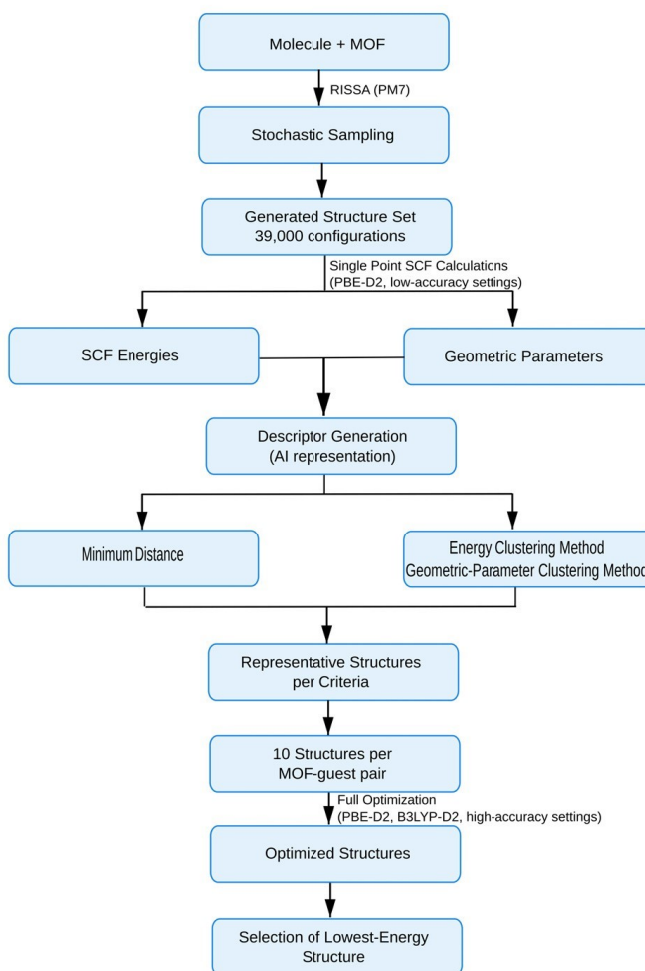


Fig. 1 Schematic overview of the computational workflow employed in this study. The protocol starts with stochastic generation of MOF–guest configurations using RISSA at the PM7 level, followed by single-point PBE-D2 calculations with low-accuracy settings to obtain energetic descriptors. SCF energies and geometric parameters are then used to construct descriptors for three selection criteria: minimum-distance filtering, energy-based clustering, and geometric-parameter clustering. The selected representative structures are subsequently fully optimized using periodic DFT calculations with high-accuracy settings, leading to the identification of the lowest-energy MOF–guest configurations.

a sphere, three-dimensional periodic boundary conditions (PBC), and two-dimensional periodic boundary conditions. In the present work, the MOF–guest configurations were generated using the three-dimensional PBC option, with the guest molecule sampled inside the periodic unit cell of the corresponding MF300 framework. Therefore, the stochastic exploration was carried out directly on the periodic MOF–guest systems, and no finite cluster models were constructed or used in the sampling stage.

For each system, RISSA generated a variable number of distinct structures, ranging from 2000 to 3470, depending on the configurations identified during the stochastic sampling process. Previous studies have demonstrated the effectiveness of RISSA in identifying reliable candidates for global minima on complex potential energy surfaces^{41–45}.

In the next stage of the protocol, single-point energy calculations were carried out at the DFT level to im-

prove accuracy and obtain reliable electronic properties. The Perdew–Burke–Ernzerhof (PBE) exchange–correlation functional⁴⁶ was employed together with Grimme’s D2 dispersion correction scheme⁴⁷. Calculations were performed using plane waves and pseudopotentials with an energy cutoff of 400 eV to achieve a balanced compromise between accuracy and computational efficiency. This stage DFT computations were conducted with the Vienna Ab initio Simulation Package (VASP)⁴⁸.

In a subsequent step, a machine learning–based selection strategy was employed to systematically identify a reduced yet representative set of configurations for further high-level calculations. This step aims to mitigate the large structural redundancy inherent to stochastic sampling methods while preserving both energetic and geometric diversity across the explored configurational space. Prior to model construction, total energies were normalized to the [0, 1] interval, and anomalous configurations were



removed using the Isolation Forest algorithm. Three complementary selection models were then developed, each designed to capture different aspects of the MOF-guest interaction landscape: (a) Energy-based clustering, in which configurations were grouped solely according to their total energy, enabling the identification of energetically representative structures without bias toward a single low-energy region; (b) Multivariable clustering, combining total energy with four distances between the most electronegative atom of the guest and selected metal centers in the MOF, thus accounting simultaneously for energetic and geometric similarity and allowing distinct adsorption motifs with comparable energies to be retained; and (c) Minimum-distance selection, based exclusively on the shortest distance between the most electronegative atom of the guest and the nearest MOF metal center, targeting configurations with potentially strong host-guest interactions. For models (a) and (b), unsupervised clustering was performed using the K-means algorithm^{49,50}. The optimal number of clusters was determined to be three using the elbow method, and the configuration closest to each cluster centroid was selected as the representative structure. For model (c), the four configurations exhibiting the smallest guest-metal distances were retained.

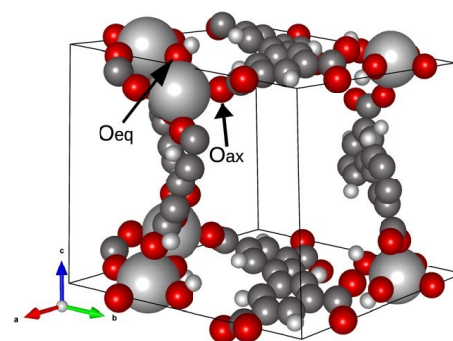
It is important to emphasize that, at this stage, all configurations correspond to unrelaxed geometries generated directly from the stochastic sampling procedure and evaluated only through single-point semiempirical or low-cost DFT calculations. Therefore, the corresponding energies should be regarded as screening descriptors rather than final adsorption energies. Consequently, the configurations exhibiting the lowest initial energies are not necessarily those that converge to the lowest-energy structures after full geometry optimization at a higher level of theory. This is particularly relevant for MOF-guest systems, where geometrically distinct adsorption motifs may have similar preliminary energies but relax toward substantially different local minima due to changes in guest orientation, host-guest contacts, and framework relaxation. Since full periodic DFT optimization of all sampled configurations would be computationally prohibitive, selecting structures exclusively according to their initial energy ranking could introduce a strong bias toward a narrow region of the configurational space and potentially overlook relevant adsorption modes. To address this issue, the present protocol incorporates a descriptor-driven unsupervised selection strategy that considers both energetic and geometric variability prior to the final DFT optimizations. This approach increases the likelihood of retaining representative configurations associated with distinct host-guest interaction mechanisms and enables a broader and more statistically robust exploration of the potential energy surface.

In the final stage of the protocol, both the cell parameters and atomic positions of all selected structures were relaxed using DFT with high-accuracy settings. These calculations were performed with VASP employing the PBE-D2 method in a similar fashion to the second stage. For this optimization section, in contrast, the interaction between core and valence electrons via the Projector Augmented-Wave (PAW) method^{51,52}, an energy cutoff of 520 eV, as well as spin polarization were considered. Geometry optimizations were conducted with a plane-wave energy cutoff of 520 eV and Brillouin-zone sampling using a $1 \times 1 \times 1$ k-point mesh. The

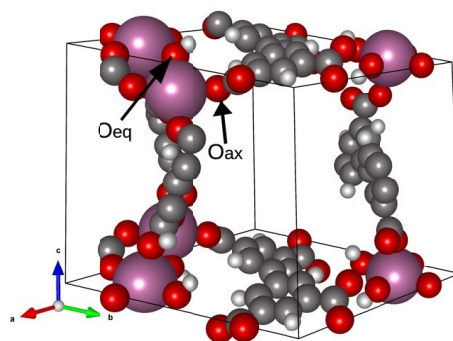
electronic energy and force convergence thresholds were set to 1×10^{-6} eV and $0.01 \text{ eV } \text{\AA}^{-1}$, respectively.

Additionally, complementary DFT calculations employing Gaussian-type basis sets were carried out with the CRYSTAL23 software package⁵³. These calculations used the hybrid B3LYP exchange-correlation functional⁵⁴ with Grimme's D2 dispersion correction and the POB-TZVP basis set⁵⁵ for all atoms, except for the In(III) ion, where a pseudopotential compatible with the POB-TZVP basis was employed. A k-points sampling of 2×2 was set together with the electronic energy and force convergence thresholds of 1×10^{-7} a.u. and $0.0003 \text{ a.u. } \text{\AA}^{-1}$, respectively. Although more recent dispersion schemes such as D3 are available in both VASP and CRYSTAL23, the D2 correction was deliberately retained in this work to ensure direct methodological consistency with the previous study by Rivera-Almazo et al. on MFM-300(Sc) and MFM-300(In)³⁰. This choice allows the differences observed in the adsorption configurations and relative energies to be attributed primarily to the configurational-search protocol rather than to changes in the dispersion model. Therefore, D2 was adopted as a common reference level for both the large-scale screening and the subsequent comparison with previously reported structures, rather than as an assertion of its superiority over more recent dispersion corrections. The overall computational workflow is summarized in Figure 1. All data preprocessing and machine-learning analyses were implemented in Python using the scikit-learn library^{49,50}.

1 Results



(a) MFM-300(Sc)



(b) MFM-300(In)

Fig. 2 Unit cells of the MOFs (a) MFM-300(Sc) and (b) MFM-300(In) obtained using the PBE-D2 exchange-correlation functional.



Table 1 Cell parameters of MFM-300(Sc) and MFM-300(In) predicted using PBE-D2/plane-waves and B3LYP-D2/POB-TZVP methods. Distances (Å) and angles (°) are reported, and mean absolute percentage errors (MAPE) are given in % relative to experimental data.

Parameter	B3LYP-D2	PBE-D2	Exp ^{56,57}	MAPE (B3LYP-D2)	MAPE (PBE-D2)
MFM-300(Sc)					
a	12.56	12.61	12.49	0.54	0.93
b	12.56	12.61	12.49	0.54	0.93
c	12.56	12.61	12.49	0.54	0.93
α	104.28	104.26	104.22	0.06	0.03
β	104.28	104.26	104.22	0.06	0.03
γ	120.45	120.50	120.58	0.11	0.07
Volume	1483.23	1500.30	1458.46	1.70	2.87
d_{M-Oeq}	2.11	2.11	2.12	0.48	0.34
d_{M-Oax}	2.12	2.12	2.11	0.41	0.52
d_{M-OH}	2.09	2.09	2.05	2.21	1.96
MFM-300(In)					
a	12.64	12.77	12.55	0.7	1.74
b	12.64	12.77	12.55	0.7	1.74
c	12.64	12.77	12.55	0.7	1.74
α	103.99	103.91	103.97	0.02	0.06
β	103.99	103.91	103.97	0.02	0.06
γ	121.10	121.27	121.14	0.04	0.11
Volume	1506.35	1552.03	1474.82	2.14	5.24
d_{M-Oeq}	2.14	2.18	2.13	0.59	2.48
d_{M-Oax}	2.19	2.24	2.16	1.55	3.53
d_{M-OH}	2.12	2.15	2.11	0.65	1.82

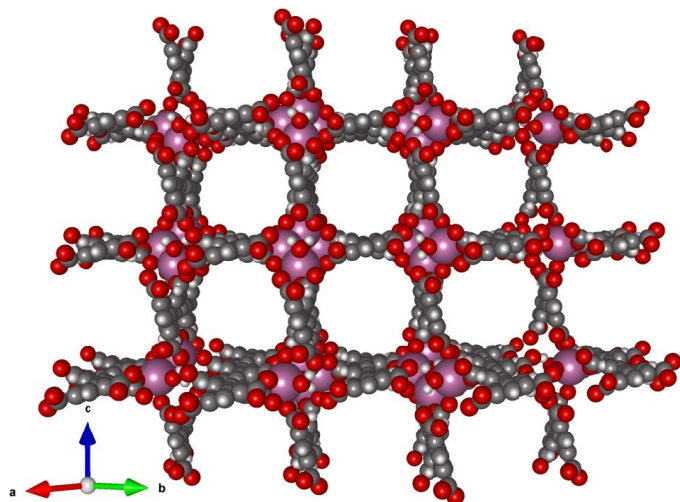


Fig. 3 Pore network of MFM-300(In) constructed through unit cell replication, highlighting the formation of extended channels.

MFM-300(Sc) and MFM-300(In) structures

Before exploring the PES of the MOF-guest systems, a structural analysis of the MOFs MFM-300(Sc) and MFM-300(In) was conducted using the methodologies described in this work. Complete geometry optimizations were performed with both the PBE-D2/plane-wave and B3LYP-D2/POB-TZVP approaches, and the resulting structural parameters were compared with the corresponding experimental data (Table 1). The initial structures were based on those reported by Rivera-Almazo et al.³⁰

Each MOF unit cell comprises 72 atoms, including four

metal centers that exhibit an octahedral coordination environment formed by the carboxylate groups of the biphenyl-3,3',5,5'-tetracarboxylate ligand and the bridging μ_2 -OH group. In addition, Table 1 reports the optimized unit-cell parameters together with selected metal-oxygen bond lengths of the framework. The metal centers exhibit a coordination environment defined by equatorial and axial oxygen atoms, including contributions from hydroxyl (OH) groups (Figure 2). Given their relevance to the structural robustness and chemical stability of the materials, representative metal-oxygen distances were analyzed and are summarized in Table 1. In this context, d_{M-Oeq} denotes the distance between the metal center and the equatorial oxygen atom associated with the hydroxyl group, d_{M-Oax} corresponds to the metal-oxygen separation along the axial direction, perpendicular to the hydroxyl plane, as shown in Figure 2, and d_{M-OH} refers to the bond length between the metal center and the oxygen atom directly bonded to the OH ligand.

The porosity of these materials is consistent with that reported for the isostructural MFM-300 family. Although the framework forms a fully connected three-dimensional crystalline network, the accessible pore space is best described as a system of one-dimensional (1D) channels running along the crystallographic *c*-axis, as illustrated in Fig. 3. These channels display an approximately cylindrical shape, with a slightly corrugated internal surface arising from the periodic arrangement of μ_2 -OH groups lining the pore walls. The effective channel diameters are similar in both materials, with values of approximately 13.6 Å for MFM-300(Sc) and 13.8 Å for MFM-300(In). The slightly larger pore diameter in MFM-300(In) is consistent with the expansion of the framework associated with the larger ionic radius of In compared



with Sc. Overall, both materials provide well-defined 1D adsorption environments embedded within a robust three-dimensional framework, making them suitable platforms for assessing how guest position and orientation affect the host-guest potential energy surface.

For all these metal-ligand distances, both methods slightly overestimate the experimental values for MFM-300(Sc) and MFM-300(In), although the deviations remain within acceptable limits. Both computational methods accurately reproduced the experimental lattice parameters (a , b , and c) and cell angles (α , β , and γ) for MFM-300(Sc), with mean absolute percentage errors (MAPEs) below 1% for PBE-D2/plane-wave calculations. Slightly larger deviations were observed for MFM-300(In), approaching 2%. The B3LYP-D2/POB-TZVP method yielded approximately half the error of PBE-D2/plane-wave for these parameters. Regarding the unit-cell volume, both methods slightly overestimated the experimental values, with PBE-D2/plane-wave showing a larger deviation. The highest MAPEs were 3% for MFM-300(Sc) and 5.2% for MFM-300(In) using PBE-D2/plane-wave. In contrast, B3LYP-D2/POB-TZVP provided smaller deviations of 1.7% and 2.1%, respectively, demonstrating its superior accuracy in reproducing the experimental structural features. Although slight discrepancies are observed for certain properties, both computational approaches show good overall agreement with the experimental data, with all deviations remaining below 6%. These results confirm that both the B3LYP-D2/POB-TZVP and PBE-D2/plane-wave methods are suitable for accurately reproducing the structural parameters of the studied MOFs, providing reliable results within acceptable error margins.

Potential energy surface exploration of MOF-guest systems

Approximately 39 000 configurations were generated across all MOF-guest systems, providing extensive sampling of the host-guest interaction landscape. For MFM-300(Sc), the number of configurations per adsorbate ranged from 2 082 to 3 250, while for MFM-300(In) it ranged from 2 183 to 3 471. Each configuration produced by RISSA was evaluated through single-point PBE-D2 plane-wave calculations. Although this stage required significant computational resources, the use of a 400 eV plane-wave cutoff enabled efficient large-scale screening within practical computational times.

Figure 4 summarizes the distributions of relative energies ($\Delta E = E_{\text{conf}} - E_{\text{min}}$) for the configurational ensembles of the different adsorbates confined in MFM-300(Sc) and MFM-300(In). In all cases, the lowest-energy configuration within each ensemble was taken as the zero-energy reference, allowing a direct comparison of the shape and width of the energy distributions independently of the absolute interaction energies. The number of configurations retained after outlier removal is indicated in the figure, and the corresponding statistical descriptors are summarized in Table 2. The statistical descriptors support the trends observed in the violin plots. For both frameworks, larger organic adsorbates such as benzyl alcohol and benzaldehyde exhibit broader ΔE distributions, larger standard deviations, and higher median energies than the smaller gaseous molecules. This behavior in-

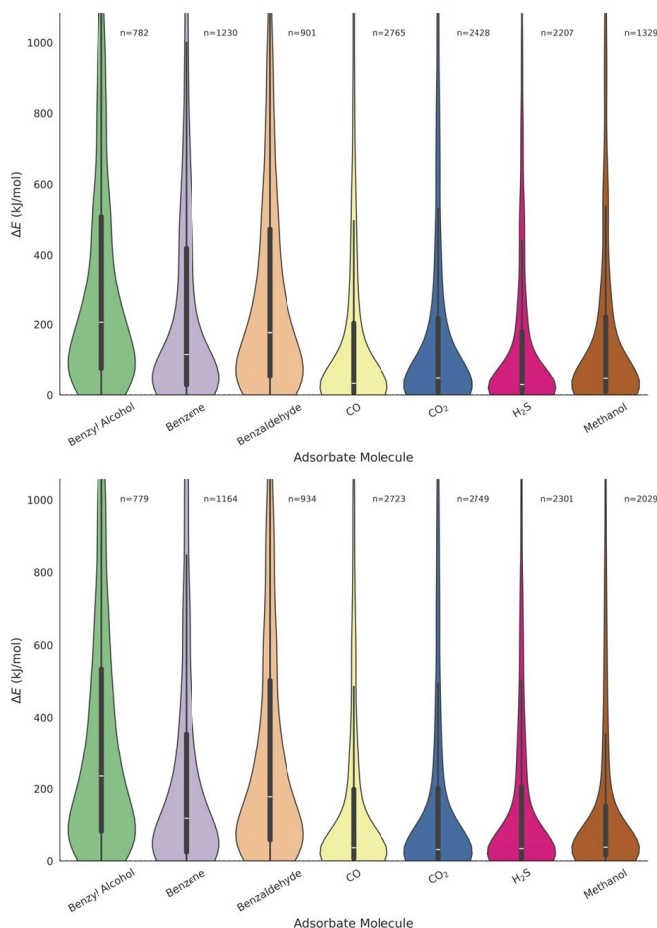


Fig. 4 Distribution of relative energies ($\Delta E = E_{\text{conf}} - E_{\text{min}}$) for the configurational ensembles of different guest molecules adsorbed in MFM-300(Sc) and MFM-300(In). The violin plots represent the probability density of the sampled configurations, while the embedded boxplots indicate the median and interquartile range. Individual sampled points were omitted to improve readability. The number of configurations retained after data cleaning with the Isolation Forest algorithm is indicated for each adsorbate. The x-axis corresponds to the adsorbate molecule and the y-axis reports relative energies in kJ mol^{-1} . Quantitative statistical descriptors of the distributions are reported in Table 2.

dicates a more corrugated configurational landscape, consistent with the larger number of possible orientations, adsorption motifs, and host-guest contacts available to these molecules inside the pore. Benzyl alcohol displays the broadest distribution in both materials, which can be associated with its conformational flexibility and with the coexistence of dispersive and hydrogen-bond-type interactions. Benzaldehyde shows a similar, although slightly less pronounced, energetic dispersion. In contrast, CO and CO₂ exhibit lower median ΔE values and lower energetic dispersion, as reflected by their statistical descriptors. These quantitative descriptors make the differences among CO, CO₂, and H₂S clearer than visual inspection of the violin plots alone. The skewness coefficients reported in Table 2 provide a quantitative measure of the asymmetry of the distributions. Positive skewness values indicate that most configurations are concentrated in a relatively low-energy region, while a smaller number of configurations extend toward high relative energies. This behavior is consistent



Table 2 Descriptive statistical parameters of the relative-energy distributions (ΔE) obtained for the adsorbed configurations of the studied molecules in MFM-300(Sc) and MFM-300(In). Reported values include the mean, median, standard deviation, and skewness in kJ mol^{-1} . The statistics were computed from the datasets retained after Isolation Forest outlier removal and further restricted to the 95th percentile to reduce the influence of extreme high-energy configurations.

Molecule	Mean	Median	Std Deviation	Skewness
MFM-300(Sc)				
Benzyl Alcohol	360.96	206.44	396.36	1.69
Benzene	283.13	116.07	373.95	1.94
Benzaldehyde	337.03	177.15	399.98	1.86
CO	183.50	32.69	328.43	2.80
CO ₂	202.46	48.02	344.35	2.52
H ₂ S	156.61	29.86	278.45	2.97
Methanol	199.59	47.38	335.34	2.59
MFM-300(In)				
Benzyl Alcohol	372.03	235.15	390.99	1.61
Benzene	271.70	119.46	373.54	2.09
Benzaldehyde	340.91	177.33	393.01	1.67
CO	174.37	36.62	301.85	2.73
CO ₂	191.66	31.39	342.57	2.56
H ₂ S	171.75	34.44	300.25	2.85
Methanol	169.90	38.33	304.83	2.95

with the extended high-energy tails observed in Fig. 4. In particular, the smaller gaseous adsorbates show pronounced positive skewness, suggesting that although many sampled configurations remain close to the low-energy region, a limited number of orientations lead to strongly destabilized host–guest arrangements. Overall, the analysis indicates that molecular size, adsorption-site diversity, and interaction complexity determine the ruggedness of the configurational energy landscape in both frameworks. The similar statistical trends observed for MFM-300(Sc) and MFM-300(In) suggest that pore topology primarily governs the qualitative features of the potential energy surface, whereas the identity of the metal center modulates the energetic dispersion more subtly. These results support the need for a selection protocol that preserves both energetic and geometric diversity before final DFT optimization.

Relative energies

Among the ten representative structures selected for each MOF–guest pair, those with relative energies lower than 10 kcal mol^{-1} are reported in Table 3. The table compares the results obtained from the three selection criteria: energy clustering (EC), geometric-parameter clustering (GP), and metal–ligand distance (MLD). The screening stage is performed on non-optimized configurations, whereas the energies reported in Table 3 correspond to fully optimized structures. Since the initial sampling generated approximately 39,000 configurations, full periodic DFT optimization of the complete ensemble would be computationally prohibitive, particularly because the final relaxation includes both guest and framework degrees of freedom. Table 3 should be interpreted by identifying which selection criterion leads to the lowest-energy optimized configurations. If total energy alone were sufficient at the screening stage, the lowest-energy structures would be expected to arise systematically from the EC column. Instead,

several minima are obtained from the GP or MLD criteria, indicating that geometrical and chemically motivated descriptors retain adsorption motifs that may be missed by energy ranking alone. Therefore, combining energy-, geometry-, and distance-based criteria provides a more reliable strategy for identifying representative low-energy structures on the MOF–guest PES. The coexistence of several low-lying structures within narrow energy windows reflects the corrugated nature of the PES in confined MOF environments. The purpose of the present protocol is therefore not only to locate a single minimum, but also to identify representative adsorption motifs and their relative energies. Direct comparison with experiment is limited by the availability of complete guest positions; nevertheless, for benzene@MFM-300, comparison with previously proposed structures derived from partial experimental information shows that the present protocol identifies lower-energy adsorption geometries, as discussed below. Fully optimized structures are provided in the Electronic Supporting Information (ESI). The ESI contains detailed structural data that may be of interest to researchers seeking to perform further analyses or comparative studies on these MOF–guest systems.

The energies and selected geometrical parameters of the MOF–benzene systems obtained using our protocol were compared with those reported previously by Rivera-Almazo et al.³⁰. To ensure a direct and consistent comparison, and to avoid discrepancies arising from software version differences, single-point energy calculations were performed with CRYSTAL23 on the structures reported by Rivera-Almazo et al.³⁰.

A comparison of the empty MOF cells, i.e., in the absence of benzene, revealed only negligible differences between both studies, with the largest energy deviation being 0.003 Hartree . This indicates that the isolated frameworks are essentially equivalent at the level of theory used for the comparison. In contrast, when benzene adsorption is considered, the structures reported by Rivera-Almazo et al. are substantially higher in energy than



Table 3 Relative energies (kcal/mol) for MFM-300(Sc)–guest and MFM-300(In)–guest structures determined through energy clustering (EC), geometric parameters clustering (GP) and metal–ligand distance (MLD) methods, employing the PBE-D2/plane-wave and B3LYP-D2/POB-TZVP methods.

MFM-300(Sc)						MFM-300(In)					
PBE-D2			B3LYP-D2			PBE-D2			B3LYP-D2		
EC	GP	MLD	EC	GP	MLD	EC	GP	MLD	EC	GP	MLD
Benzyl Alcohol											
1.97	3.52	0.00	–	0.00	6.68	7.10	0.00	–	0.00	–	3.16
6.24	4.74	2.33	–	2.53	–	–	1.15	–	–	–	–
–	4.89	4.75	–	7.39	–	–	1.47	–	–	–	–
Benzene											
0.21	0.00	–	0.01	0.02	0.00	0.55	0.00	5.81	0.00	3.47	3.50
0.22	0.05	–	3.80	3.37	–	3.69	0.47	–	–	4.13	4.14
–	0.41	–	5.12	4.57	–	–	0.50	–	–	4.76	5.08
Benzaldehyde											
4.11	1.21	0.00	0.32	0.90	0.00	1.89	6.20	0.00	0.25	8.08	0.00
4.90	5.08	0.17	–	6.98	0.36	5.11	6.76	–	0.82	9.15	–
–	5.55	1.21	–	7.27	0.51	5.89	8.17	–	1.29	–	–
–	–	5.41	–	–	9.73	–	–	–	–	–	–
CO											
0.11	0.00	–	1.18	0.00	–	0.00	0.09	–	0.00	0.04	0.03
0.43	0.16	–	1.21	0.01	–	0.41	0.41	–	2.32	2.37	–
0.17	9.97	–	1.33	1.14	–	0.65	2.99	–	2.61	4.33	–
CO ₂											
0.26	0.00	0.51	0.53	0.03	0.00	0.66	0.00	–	0.00	0.05	–
0.51	0.31	–	0.71	0.03	–	3.02	2.25	–	2.37	2.07	–
0.73	0.57	–	3.45	0.43	–	–	2.78	–	4.44	2.23	–
H ₂ S											
0.72	0.00	–	0.05	0.00	–	0.50	0.80	0.00	0.00	0.82	0.78
1.16	0.44	–	2.51	0.07	–	0.67	0.95	–	0.74	1.15	–
–	1.09	–	–	0.61	–	0.88	2.98	–	4.17	1.83	–
Methanol											
0.00	5.64	4.83	0.00	0.00	8.42	4.48	3.02	0.00	5.22	0.00	0.01
6.33	6.01	5.01	–	0.08	9.21	7.03	7.00	–	–	0.18	–
–	6.03	–	–	0.11	–	9.89	7.55	–	–	–	–

those identified in the present work, by 106.68 kcal mol⁻¹ for MFM-300(Sc) and 130.44 kcal mol⁻¹ for MFM-300(In). Therefore, these differences arise mainly from the MOF–benzene arrangement and the subsequent host–guest relaxation, rather than from intrinsic differences in the empty MOF frameworks.

The optimized cell parameters and selected metal–oxygen distances remain very similar in both sets of structures. In particular, the parameters a , b , c , γ , $d_{M-O_{eq}}$, $d_{M-O_{ax}}$, and d_{M-OH} show only minor variations. Nevertheless, the superposition of the structures, shown in Fig. 5, reveals subtle but relevant atomic rearrangements of the framework upon benzene adsorption. These displacements are more pronounced for MFM-300(Sc), with deviations in the range of 0.50–0.80 Å, than for MFM-300(In), where deviations of 0.10–0.40 Å are observed. The most important differences, however, are associated with the adsorption site and orientation of the benzene molecule inside the pore.

These rearrangements are also reflected in the optimized unit-cell volumes. Relative to the structures reported by Rivera-Almazo et al., the configurations obtained in this work exhibit volume reductions of -2.30 \AA^3 for MFM-300(Sc) and -9.62 \AA^3 for MFM-300(In). These changes suggest a more compact host–guest arrangement, which may enhance cooperative framework–guest contacts and contribute to the additional stabilization. However, the energetic stabilization cannot be rationalized solely in terms of cell-volume changes or the shortest metal–guest distance. For MFM-300(Sc), the shortest C–M distance is 4.41 Å in the structure identified in this work, compared with 4.61 Å in the previously

reported configuration. For MFM-300(In), the corresponding distances are 4.54 and 4.50 Å, respectively. Thus, the lower energies obtained here result from the combined effect of benzene reorientation, changes in the adsorption site, and cooperative relaxation of the framework. This analysis demonstrates that the systematic exploration of the PES is able to identify adsorption motifs that are not evident from chemically intuitive initial guesses or partially defined experimental structural information.

It is important to note that the structures reported by Rivera-Almazo et al. were derived from partial experimental information, whereas the configurations presented here result from a systematic and statistically guided exploration of the potential energy surface. Although the distance between the metal center and the nearest carbon atom of benzene is similar in both approaches, the corresponding geometries and energies differ significantly. The stochastic sampling performed with RISSA generates a broad set of initial MOF–guest arrangements, which are subsequently relaxed through full DFT optimization. This finding highlights the robustness of the protocol proposed in this study, which provides a systematic and unbiased framework for exploring MOF–guest configurations when the atomic arrangement of the guest molecule is not experimentally well defined. Overall, our protocol constitutes a useful tool for identifying MOF–guest binding sites and, more generally, for investigating pore–guest interactions.



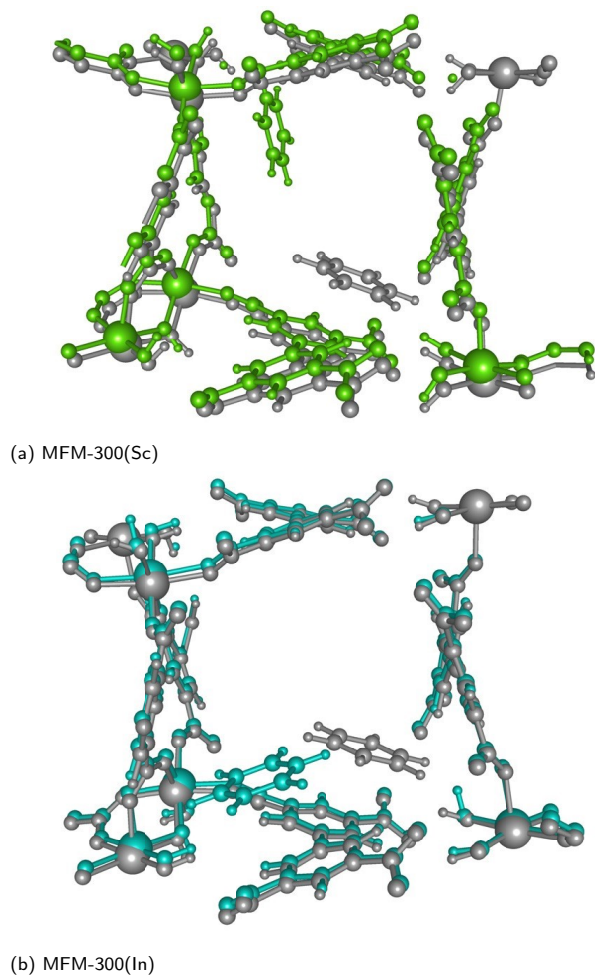


Fig. 5 Superposition of the optimized benzene@MFM-300(Sc) and benzene@MFM-300(In) structures obtained in this work with those reported by Rivera-Almazo et al. The previously reported structures are shown in gray in both panels. The structures obtained in the present study are shown in green for MFM-300(Sc) and in cyan for MFM-300(In).

Conclusions

In this work, we presented a multilevel computational protocol designed to systematically explore the potential energy surface (PES) of MOF-guest systems. By combining stochastic sampling, semiempirical prescreening, density functional theory (DFT) refinement, and descriptor-driven unsupervised selection, our approach enables the efficient identification of low-energy configurations while avoiding the biases inherent in chemically intuitive initial guesses. The protocol was applied to the adsorption of benzyl alcohol, benzene, benzaldehyde, CO, CO₂, H₂S, and methanol in the MOFs MFM-300(Sc) and MFM-300(In). For each system, thousands of candidate structures were generated using stochastic sampling and subsequently filtered and clustered using complementary energy-, geometry-, and distance-based descriptors. A set of ten representative configurations per MOF-guest pair was then reoptimized with high-level DFT methods, providing a detailed picture of the accessible low-energy region of the PES. Our results show that relying exclusively on total energy during the screening stage is insufficient to reliably identify low-energy con-

figurations. Based on the present analysis, we recommend that initial energetic rankings be complemented with geometrical and chemically motivated descriptors. In practical terms, the selection should include: (i) energy-based clustering to retain representative configurations from different energetic regions; (ii) geometry-based or multivariable clustering to preserve distinct adsorption motifs with similar preliminary energies; and (iii) distance- or contact-based criteria to retain structures involving potentially relevant host-guest interactions. This multidescrptor strategy reduces the risk of discarding configurations that are not the lowest in energy before relaxation but may converge to competitive minima after full DFT optimization. The application of this protocol reveals that several low-lying MOF-guest structures can coexist within relatively narrow energy windows, particularly for weakly interacting adsorbates. This structural diversity reflects the corrugated nature of the PES in confined environments and emphasizes the importance of preserving configurational diversity during the screening stage. Comparison with previously reported benzene-MFM-300 structures further demonstrates that systematic PES exploration can identify adsorption motifs with noticeably lower energies and distinct optimized geometries, especially when the experimental position and orientation of the guest molecule are not fully defined. Overall, the protocol proposed in this study offers a generalizable, computationally efficient, and unbiased framework for exploring MOF-guest configurations. Its modular design enables straightforward extension to other porous materials, guest molecules, descriptors, and sampling strategies. We anticipate that this methodology will assist future studies of adsorption, catalysis, and host-guest recognition in crystalline porous materials by providing practical guidelines for selecting representative configurations prior to high-level electronic-structure optimization.

Author contributions

All authors contributed equally to this work and have approved the final version of the manuscript.

Conflicts of interest

There are no conflicts to declare.

Data availability

All data supporting this study are openly available at GitHub (<https://github.com/joana20-source/MOFs.git>). The repository contains the scripts, structure files in the *.34 format obtained with CRYSTAL23, structure files in the *.vasp format generated with VASP, and the associated analysis files required to reproduce the results.

Acknowledgements

We thank the Laboratorio de Supercómputo y Visualización en Paralelo for providing access to its computational facilities. J.A. and M.R. acknowledge financial support from SECIHTI through a postdoctoral fellowship.



References

- 1 D. W. Breck, W. G. Eversole, R. M. Milton, T. B. Reed and T. L. Thomas, *J. Am. Chem. Soc.*, 1956, **78**, 5963–5972.
- 2 F. Bahmanzadegan and A. Ghaemi, *J. Hazard. Mater. Adv.*, 2025, **17**, 100617.
- 3 C. T. Kresge, M. E. Leonowicz, W. J. Roth, J. C. Vartuli and J. S. Beck, *Nature*, 1992, **359**, 710–712.
- 4 R. Narayan, U. Y. Nayak, A. M. Raichur and S. Garg, *Pharmaceuticals*, 2018, **10**, 3.
- 5 J. Wang and S. Kaskel, *J. Mater. Chem.*, 2009, **19**, 8404–8413.
- 6 D. Kobina Sam, H. Li, Y.-T. Xu and Y. Cao, *J. Ind. Eng. Chem.*, 2024, **135**, 17–42.
- 7 H. Li, M. Eddaoudi, M. O’Keeffe and O. M. Yaghi, *Nature*, 1999, **402**, 276–279.
- 8 S. Kitagawa *et al.*, *Chem. Soc. Rev.*, 2014, **43**, 5415–5418.
- 9 P. M. Budd, N. B. McKeown and D. Fritsch, *J. Mater. Chem.*, 2004, **14**, 1975–1986.
- 10 N. B. McKeown, *Polymer*, 2020, **202**, 122736.
- 11 B. F. Hoskins and R. Robson, *J. Am. Chem. Soc.*, 1989, **111**, 5962–5964.
- 12 H. Furukawa, K. E. Cordova, M. O’Keeffe and O. M. Yaghi, *Science*, 2013, **341**, 1230444.
- 13 L. Chen and Q. Xu, *Matter*, 2019, **1**, 57–89.
- 14 W. Luo, Z. Zhang, G. Zhu, X. Zhang, G. Huang, T. Zhou and X. Lu, *J. Environ. Chem. Eng.*, 2025, **13**, 118229.
- 15 D. Li, A. Yadav, H. Zhou, K. Roy, P. Thanasekaran and C. Lee, *Glob. Chall.*, 2024, **8**, 2300244.
- 16 V. F. Yusuf, N. I. Malek and S. K. Kailasa, *ACS Omega*, 2022, **7**, 44507–44531.
- 17 O. T. Qazvini and S. G. Telfer, *ACS Appl. Mater. Interfaces.*, 2021, **13**, 12141–12148.
- 18 Z. Gao, L. Liang, X. Zhang, P. Xu and J. Sun, *ACS Appl. Mater. Interfaces.*, 2021, **13**, 61334–61345.
- 19 J. Liu, Y. Wei, P. Li, Y. Zhao and R. Zou, *J. Phys. Chem. C.*, 2017, **121**, 13249–13255.
- 20 F. Yulia, A. Zulys, R. Ruliandini *et al.*, *J. Adv. Res. Fluid Mech. Therm. Sci.*, 2019, **57**, 158–174.
- 21 P. B. S. Rallapalli, K. Cho, S. H. Kim, J.-N. Kim and H. C. Yoon, *Fuel*, 2020, **281**, 118985.
- 22 N. A. Khan, Z. Hasan and S. H. Jhung, *J. Hazard. Mater.*, 2013, **244–245**, 444–456.
- 23 R. A. Peralta and I. A. Ibarra, *CrystEngComm*, 2024, **26**, 6100–6107.
- 24 V. B. López-Cervantes, H. Alhashimi, C. A. Celaya, M. Solórzano, M. L. Martínez, Y. A. Amador-Sánchez, E. Castaldelli, E. Lester, R. A. Peralta, E. Lima, D. Solis-Ibarra, S. Yang, I. A. Ibarra and A. Laybourn, *Small*, 2025, **21**, e07448.
- 25 Y. Chen, W. Lu, M. Schröder and S. Yang, *Acc. Chem. Res.*, 2023, **56**, 2569–2581.
- 26 Y. Zhao, L. Wu, K. Wu, R.-J. Wei, H. Zeng, H. Pang, W. Lu and D. Li, *Coord. Chem. Rev.*, 2025, **524**, 216302.
- 27 M. Ernst, J. D. Evans and G. Grynóva, *Chem. Phys. Rev.*, 2023, **4**, 041303.
- 28 M. L. Díaz-Ramírez, S. H. Park, M. Rivera-Almazo, E. Medel, R. A. Peralta, I. A. Ibarra, R. Vargas, J. Garza and N. C. Jeong, *Chem. Sci.*, 2025, **16**, 2581–2588.
- 29 P. Horcajada, C. Serre, G. Maurin, N. A. Ramsahye, F. Balas, M. Vallet-Regí, M. Sebban, F. Taulelle and G. Férey, *J. Am. Chem. Soc.*, 2008, **130**, 6774–6780.
- 30 M. Rivera-Almazo, E. Perez-Sanchez, E. Martínez-Ahumada, A. Martínez, J. Garza, I. A. Ibarra and R. Vargas, *J. Phys. Chem. C.*, 2022, **126**, 6465–6471.
- 31 J. G. Flores, J. A. Zárate-Colín, E. Sánchez-González, J. R. Valenzuela, A. Gutiérrez-Alejandre, J. Ramírez, V. Jancik, J. Aguilar-Pliego, M. C. Zorrilla, H. A. Lara-García, E. González-Zamora, G. Guzmán-González, I. González, G. Maurin and I. A. Ibarra, *ACS Appl. Mater. Interfaces.*, 2020, **12**, 18885–18892.
- 32 Z. Deng and L. Sarkisov, *Chem. Mater.*, 2024, **36**, 9806–9821.
- 33 Y. G. Chung, D. A. Gómez-Gualdrón, P. Li, K. T. Leperi, P. Deria, H. Zhang, N. A. Vermeulen, J. F. Stoddart, F. You, J. T. Hupp, O. K. Farha and R. Q. Snurr, *Sci. Adv.*, 2016, **2**, e1600909.
- 34 Y. Wu, H. Duan and H. Xi, *Chem. Mater.*, 2020, **32**, 2986–2997.
- 35 I. B. Orhan, H. Daglar, S. Keskin, T. C. Le and R. Babarao, *ACS Appl. Mater. Interfaces.*, 2022, **14**, 736–749.
- 36 X. Deng, W. Yang, S. Li, H. Liang, Z. Shi and Z. Qiao, *Appl. Sci.*, 2020, **10**, 569.
- 37 I. B. Orhan, T. C. Le, R. Babarao and A. W. Thornton, *Commun. Chem.*, 2023, **6**, 214.
- 38 J.-J. García, R. Hernández-Esparza, R. Vargas, W. Tiznado and J. Garza, *New J. Chem.*, 2019, **43**, 4342–4348.
- 39 J. J. García Miranda, *Master’s thesis*, Universidad Autónoma Metropolitana, Unidad Iztapalapa, Mexico City, Mexico, 2018.
- 40 J. J. P. Stewart, *J. Mol. Model.*, 2013, **19**, 1–32.
- 41 E. García-Hernández, A. M. García-Crisóstomo, L. Palomino-Asencio, C. Cuautli, S. M. Mejía, E. Shakerzadeh and R. Catarino-Centeno, *Comput. Theor. Chem.*, 2024, **1239**, 114806.
- 42 M. Rivera-Almazo, M. L. Díaz-Ramírez, R. Hernández-Esparza, R. Vargas, A. Martínez, V. Martis, P. A. Sáenz-Cavazos, D. Williams, E. Lima, I. A. Ibarra and J. Garza, *Phys. Chem. Chem. Phys.*, 2021, **23**, 1454–1463.
- 43 E. Medel, J. L. Obeso, C. Serrano-Fuentes, J. Garza, I. A. Ibarra, C. Leyva, A. K. Inge, A. Martínez and R. Vargas, *Chem. Commun.*, 2023, **59**, 8684–8687.
- 44 L. Palomino-Asencio, E. Chigo-Anota and E. García-Hernández, *ChemPhysChem*, 2022, **23**, e202200310.
- 45 E. Medel, J. Garza, I. A. Ibarra, A. Martínez and R. Vargas, *Comput. Theor. Chem.*, 2023, **1228**, 114265.
- 46 J. P. Perdew and W. Yue, *Phys. Rev. B*, 1986, **33**, 8800–8802.
- 47 S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787–1799.
- 48 G. Kresse and J. Furthmüller, *Phys. Rev. B*, 1996, **54**, 11169–11186.
- 49 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel,



- B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.
- 50 L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt and G. Varoquaux, ECML PKDD Workshop: Languages for Data Mining and Machine Learning, 2013, pp. 108–122.
- 51 P. E. Blöchl, *Phys. Rev. B*, 1994, **50**, 17953–17979.
- 52 G. Kresse and D. Joubert, *Phys. Rev. B*, 1999, **59**, 1758–1775.
- 53 A. Erba, J. K. Desmarais, S. Casassa, B. Civalleri, L. Donà, I. J. Bush, B. Searle, L. Maschio, L. Edith-Daga, A. Cossard, C. Ribaldone, E. Ascrizzi, N. L. Marana, J.-P. Flament and B. Kirtman, *J. Chem. Theory Comput.*, 2023, **19**, 6891–6932.
- 54 B. Civalleri, C. M. Zicovich-Wilson, L. Valenzano and P. Ugliengo, *CrystEngComm*, 2008, **10**, 405–410.
- 55 J. Laun, D. Vilela Oliveira and T. Bredow, *J. Comput. Chem.*, 2018, **39**, 1285–1290.
- 56 I. A. Ibarra, S. Yang, X. Lin, A. J. Blake, P. J. Rizkallah, H. Nowell, D. R. Allan, N. R. Champness, P. Hubberstey and M. Schröder, *Chem. Commun.*, 2011, **47**, 8304–8306.
- 57 J. Qian, F. Jiang, D. Yuan, M. Wu, S. Zhang, L. Zhang and M. Hong, *Chem. Commun.*, 2012, **48**, 9696–9698.



Cite this: DOI: 00.0000/xxxxxxxxxx

Systematic exploration of host–guest potential energy surfaces in metal–organic frameworks[†]

Joana Avelar^{*a}, Marcos Rivera-Almazo^b, Rubicelia Vargas^a and Jorge Garza^{*a}

Data availability

All data supporting this study are openly available at GitHub (<https://github.com/joana20-source/MOFs.git>). The repository contains the scripts, structure files in the *.34 format obtained with CRYSTAL23, structure files in the *.vasp format generated with VASP, and the associated analysis files required to reproduce the results.

^a Departamento de Química, División de Ciencias Básicas e Ingeniería, Universidad Autónoma Metropolitana-Iztapalapa, San Rafael Atlixco 186, Col. Vicentina, Iztapalapa, C.P. 09340, Ciudad de México, México. E-mail: joar@xanum.uam.mx; jgo@xanum.uam.mx

^b Instituto de Física, Universidad Nacional Autónoma de México, Coyoacán, C.P. 01000, Ciudad de México, México.

