Benchmarking binding energy calculations for organic structure-directing agents in pure-silica zeolites

Daniel Schwalbe-Koda and Rafael Gómez-Bombarelli[a)]

Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

(Dated: 20 January 2021)

Molecular modeling plays an important role in the discovery of organic structure-directing agents (OSDAs) for zeolites. By quantifying the intensity of host-guest interactions, it is possible to select cost-effective molecules that maximize binding towards a given zeolite framework. Over the last decades, a variety of methods and levels of theory have been used to calculate these binding energies. Nevertheless, there is no consensus on the best calculation strategy for high-throughput virtual screening undertakings. In this work, we compare binding affinities from density functional theory (DFT) and force field calculations for 272 zeolite-OSDA pairs obtained from static and time-averaged simulations. Enabled by automation software, we show that binding energies from the frozen pose method correlate best with DFT time-averaged energies. They are also less sensitive to the choice of initial lattice parameters and optimization algorithms, as well as less computationally expensive. Furthermore, we demonstrate that a broader exploration of the conformation space from molecular dynamics simulations does not provide significant improvements in binding energy trends over single-point calculations. The code and benchmark data are open-sourced and provide robust and computationally-efficient guidelines to calculating binding energies in zeolite-OSDA pairs.

I. INTRODUCTION

Zeolites are nanoporous materials widely used in catalysis, separation, sorption and many other industrially-relevant applications.[1] These metastable polymorphs are typically crystallized by adding inorganic cations and organic molecules to amorphous precursor gels in hydrothermal conditions,[2–4] although organic-free approaches are also possible.[5–12] In particular, a combination of electrostatic effects and dispersion interactions allows the molecules to act as organic structure-directing agents (OSDAs) that template the pore structure of the zeolite and determine its topology.[4,13,14] By exploring a variety of OSDAs under different synthesis conditions, several new zeolites have been discovered over the last decades.[15] Nevertheless, only around 250 different topologies have been experimentally identified,[16,17] a majority of which rely on OSDAs to be realized. Accessing known and new zeolites with desired pore structures and compositions through OSDA design is still an open-ended problem, often relying on trial and error.[18]

Computational simulation of zeolite-OSDA interactions has been an important resource to rationalize the design of templates, explain synthesis outcomes, or locate preferential placements for OSDAs.[14,19] Early molecular modeling works relied on shape-matching methods[20–23] to assess the goodness-of-fit between an OSDA and a zeolite. Later, it was shown that host-guest interactions computed from atomistic simulations are good predictors of synthesis outcomes for zeolites.[24–30] This descriptor has enabled theory-driven discovery of many zeolites over the last years.[15,31–36]

Yet, the interaction energy between guests and hosts is strongly dependent on the level of theory and computational pipeline employed. Methods based on quantum mechanics, such as density functional theory (DFT) calculations, offer mostly parameter-free descriptions of the total energy once a suitable exchange-correlation functional is chosen. However, the large size of typical zeolite-OSDA systems demands considerable computational resources, preventing the use of DFT for OSDA screening purposes. A more cost-effective approach is to use force fields (FF) to compute interatomic interactions. Several parametrizations have been employed to model host-guest interactions in zeolites, including CVFF,[25,28,37–39] Dreiding,[40–43] COMPASS,[44] UFF,[45,46] and others.[24,47–51]

Even after a level of theory is appropriately chosen, different ways to calculate interaction energies between zeolites and OSDAs remain. As an example, one could calculate binding affinities by computing energies by performing structural optimizations,[24,44,51] molecular dynamics (MD) simulations,[40,41,43,47,49,50,52] or calculating van der Waals (vdW) contributions by freezing host and guest structures.[25,39,45,46,53,54] Furthermore, optimizations and dynamics can be performed at constant pressure[50,51] or volume.[25,39,43,45,46,52] In screening approaches, calibrating strategies to obtain consistent, computationally-efficient results often make up for biases in simulation methods, matching experimental results better than higher levels of theory with less systematic errors.[55,56] This is especially important in a diverse, combinatorial chemical space such as the one provided by pairings of OSDAs and zeolites.

In this work, we compare methods to calculate binding energies of OSDAs in zeolites and propose guidelines to compute these interactions in high-throughput virtual screening approaches. In particular, the following contri-

[a)]Electronic mail: rafagb@mit.edu

butions are put forward:

1. Different simulation constraints, structural optimization algorithms and initial conditions of the hosts significantly change the trends in binding energies for all simulation methods;

2. Binding energies of zeolite-OSDA pairs from the Dreiding FF are best calculated by constant volume simulations with the frozen pose method, as opposed to structural optimizations or MD simulations;

3. A reference dataset of 272 OSDA-zeolite pairs calculated with DFT and FF approaches can enable further benchmarks with other FF parametrizations, software packages or simulation pipelines;

4. An open-source Python interface to the General Utility Lattice Program (GULP), GULPy, to enable faster generation of input files, parsing of outputs, structure manipulation, and execution of FF calculations.

II.   METHODS

A.   Simulation Details

DFT calculations were performed using the Vienna Ab-initio Simulation Package (VASP),[57,58] version 5.4.4, within the projector-augmented wave (PAW) method.[59,60] The Perdew–Burke–Ernzerhof (PBE) functional within the generalized gradient approximation (GGA)[61] was used as the exchange-correlation functional. vdW interactions were taken into account through Grimme's D3 corrections.[62,63] Several benchmarks have shown that including dispersion corrections in DFT calculations is imperative to accurately predict trends in binding energies between zeolites and guests.[53,54,64–72] Although D3 is known to overbind the guest species to zeolites,[54,70–72] it provides an excellent balance between cost and accuracy for binding energies compared to higher levels of theory such as random phase approximation, Møller-Plesset perturbation theory, or many-body dispersion. The kinetic energy cutoff for plane waves was restricted to 520 eV. Integrations over the Brillouin zone were performed using Monkhorst-Pack k-point meshes[73] (Γ-centered for hexagonal unit cells) with a uniform density of 64 k-points/Å^{-3} (see Table S1 for complete k-point meshes for each zeolite). For isolated molecules, a vacuum of 15 Å thickness was employed in all directions to avoid unphysical interactions between periodic images. A stopping criterion of 10^{-6} eV was adopted for the self-consistent field (SCF) cycle energy convergence. Relaxation of unit cell parameters and atomic positions was performed until the Hellmann–Feynman forces on atoms were smaller than 10 meV/Å. *Ab initio* MD (AIMD) simulations were performed in the NPT ensemble with

Langevin dynamics[74,75] within the Parrinello-Rahman method[76,77], and 0.5 fs timesteps. The sampling temperature was fixed at 400 K to simulate typical hydrothermal conditions in zeolite synthesis.[3] The fictitious lattice mass was set to 1,000 atomic mass units, and all Langevin friction coefficients were set to 1 ps^{-1}. The external pressure has been set to 0 kbar to allow comparisons with similar works where this value is implicit.[50,51] Ground-state geometries were thermalized by randomly displacing the atoms by up to 0.02 Å in each Cartesian coordinate before being used as initial configurations for AIMD calculations. AIMD simulations were performed for 500 fs, with only the last 200 fs considered for production. Despite the short time lengths, this is enough to have well-equilibrated trajectories (see Fig. S1).

FF simulations were performed using the General Utility Lattice Program (GULP), version 5.1.1,[78,79] through the new GULPy package.[80] MD simulations were performed in the NVT and NPT ensembles with modified Nosé-Hoover dynamics[81] using the Leapfrog Verlet integrator, 0.5 fs timesteps, temperature of 400 K, and 0 kbar of external pressure. Fully optimized geometries were used as initial configurations for MD calculations. All MD trajectories consisted of a production run of 5 ps preceded by a 5 ps equilibration run. The Dreiding force field[82] was used to model interactions between the zeolite and the OSDA. It has been widely used for OSDA screening[41–43] and some of its predictions have been experimentally verified[33,83]. Initial FF optimization of unloaded zeolites was performed with the Sanders-Leslie-Catlow (SLC) parametrization.[84] Whereas several other parametrizations have been proposed for pure-silica zeolites,[85–89] SLC is still widely used in the field due to the good correlation of predicted energies with experimental enthalpies of formation.[90]

Initial zeolite structures were downloaded from the International Zeolite Association (IZA) database and pre-optimized using either DFT or SLC, as described above. Conformers for OSDAs were generated using RDKit[91] with the MMFF94 force field,[92,93] and further optimized using the BP86-D3/def2-SVP[94–96] level of theory as implemented in ORCA.[97,98] Subsequent optimizations of these geometries with VASP did not change the final structure of the isolated molecule. Nonetheless, the energies derived from these calculations were necessary to adopt the same reference method (PBE-D3/PAW) throughout the work. Generation of OSDA-zeolite poses was performed using the VOID package.[99] A complete description of the docking strategy, software and parameters is found in Ref. 99.

B.   Calculation of binding energies

Typically, host-guest interactions are described in terms of binding energies, since free energies of binding are typically unavailable from simple simulations. A general expression for the binding energy (E_b) of a molecule
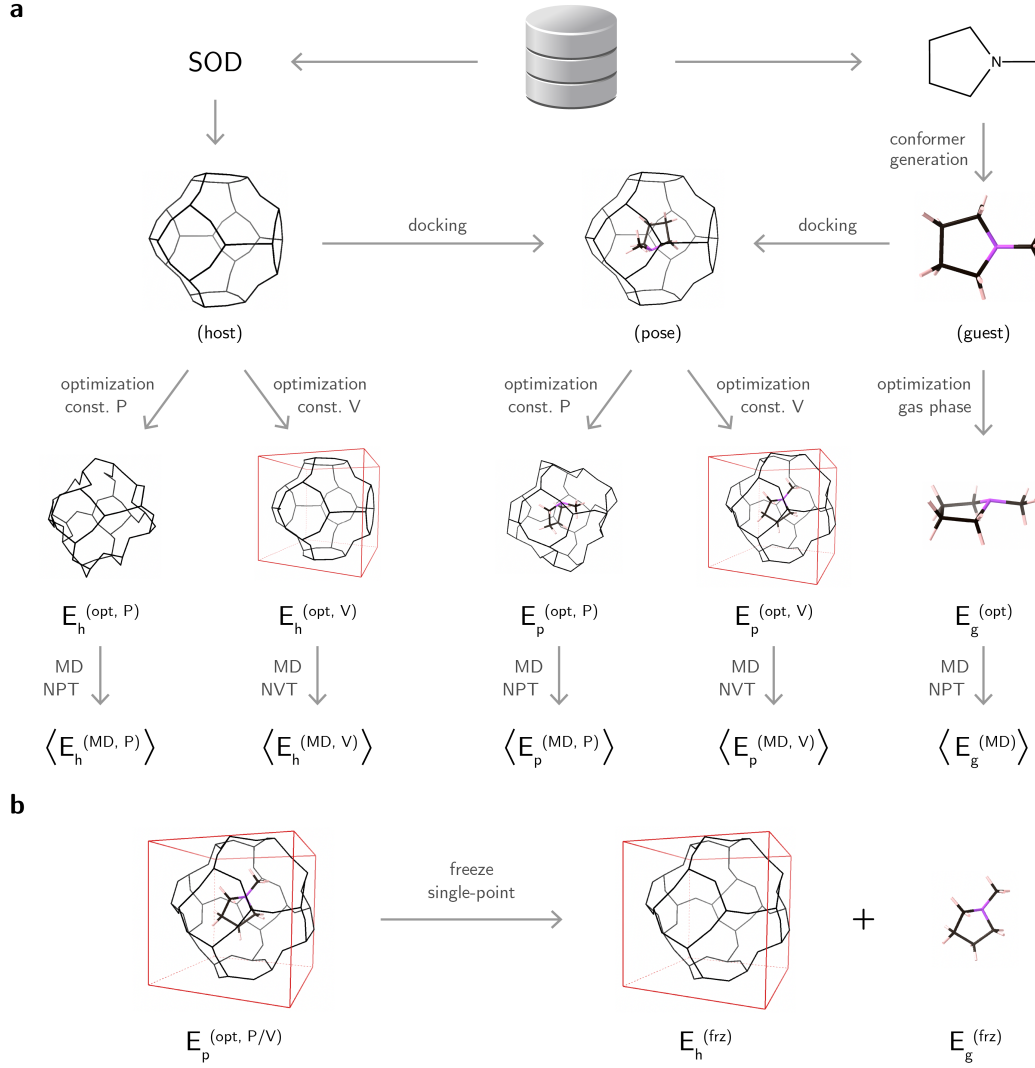
FIG. 1. **a**, Workflow of calculations performed to obtain different binding energies for each of the OSDA-zeolite pairs in the dataset. Resulting energies for poses (E_p), zeolite hosts (E_h) and OSDA guests (E_g) are shown in the figure. **b**, Schematic on the calculation of $E_h^{\text{(frz)}}$ and $E_g^{\text{(frz)}}$ energies within the frozen pose scheme. The host and guest geometries are kept fixed at the relaxed positions obtained in the pose optimization and a single point calculation is carried for each of the isolated systems.

in a host is given by

$$E_b = E_p - E_h - nE_g, \qquad (1)$$

in which E_p is the energy of the zeolite-OSDA pose, E_h is the energy of the pure-silica, unloaded zeolite host, E_g is the energy of the guest template, and n is the number of guests per pose. Whereas the interaction energy between docked guests could be subtracted from Eq. (1), it is often useful to include this term when computing the stabilization provided by the OSDA towards the zeolite. It is well-known that guest-guest interactions influence the product zeolite, particularly when more than one molecule is packed into the same cavity.[39,100] Therefore, removing these contributions from the final binding

energy expression may not be representative of the actual stabilization obtained in experimental settings.

The literature contains multiple examples on how to calculate each of the terms in Eq. (1). For example, the energies can be obtained by time-averaging the host-guest energies along an MD simulation,[40,41,43,47,49,50,52]

$$\left\langle E_b^{\text{(MD)}} \right\rangle = \left\langle E_p^{\text{(MD)}} \right\rangle - \left\langle E_h^{\text{(MD)}} \right\rangle - n \left\langle E_g^{\text{(MD)}} \right\rangle, \quad (2)$$

or by subtracting energies resulting from structural optimizations (opt),[24,44,51]

$$E_b^{\text{(opt)}} = E_p^{\text{(opt)}} - E_h^{\text{(opt)}} - nE_g^{\text{(opt)}}. \qquad (3)$$

Moreover, the resulting geometries from structural optimizations can be frozen in their relaxed positions, and the energy components calculated for the isolated host and guests without further relaxation,

$$E_b^{(\text{frz})} = E_p^{(\text{opt})} - E_h^{(\text{frz})} - \sum_i E_{g,i}^{(\text{frz})}, \qquad (4)$$

where the summation is performed for all guests docked in the host. In this scenario, there are different values of E_h and E_g for each pose, since the final atomic structure is dependent on the host-guest interactions. Here, we refer to this strategy, widely used in the zeolite literature,[25,39,45,46,53,54] as the "frozen pose" method. Finally, structural relaxations and MD simulations can be performed at constant pressure or volume, adding an additional degree of freedom for each of these simulations. Fig. 1 schematizes different pathways for obtaining all these energies.

III. RESULTS

A. Correlations between binding energy calculations

To compare the different methods of calculating binding energies, we created a dataset of 272 zeolite-OSDA poses from 164 unique complexes, which cover 60 neutral OSDAs and 55 zeolite frameworks (Fig. S2 and Tables S2-3). Different loadings were considered for pairs in which the molecule was small compared to the zeolite cavity, following the method described in Ref. 99. Only neutral OSDAs were simulated due to ambiguities on how to calculate DFT energies for charged systems without considering the presence of heteroatoms. It is unclear whether the charge-compensating background potential added in the DFT calculation may affect the lattice parameters and energy references for the pose. Furthermore, typical approaches employ FFs without explicitly considering charges, even for cationic molecules. Therefore, this dataset of neutral molecules docked in puresilica zeolites allows exploring variations of binding energies with a focus on OSDA/zeolite shapes and sizes.

DFT-optimized structures of zeolites and molecules were used as inputs for the docking scheme. We later repeated this docking step for some SLC-optimized zeolites as substrates to analyze the effects of different initial host lattices on the binding energies (Sec. III C). For each of the 272 poses created from DFT substrates, we carried structural optimizations at constant pressure using DFT, and calculated binding energies using Eq. 3. Nonetheless, the physical reality is best described by a dynamic simulation at the NPT ensemble, with the temperature set to ranges of typical hydrothermal conditions. To obtain such binding energies from *ab initio* MD without incurring into excessive computational cost, we performed AIMD simulations for 40 different complexes whose poses contained less than 80 atoms. All other energies listed in

Fig. 1 are calculated for all 272 poses. A summary of the calculation tree is shown in Fig. 2a (see Fig. S3 for a complete description).

After performing all calculations, binding energies obtained for the same initial docked structures were compared. If the property of interest in the benchmark were the mean absolute error (MAE) between the methods, systematic shifts due to parametrization inaccuracies or energy rescaling due to dynamic effects (e.g. sampling temperature in MD simulations) would largely influence the final results. Instead, a theoretical screening method should quantify trends of host-guest binding energies to inform OSDA selection, as in many other computational screening approaches.[41,43,101] To quantify the correlation in the ordering between two different methods, the Spearman's rank correlation coefficient (ρ) is employed as a figure of merit. It correlates two variables according to the rank of the data points, and is invariant to translations and rescalings of the sets under comparison.[102] An increasing, monotonic relationship between host-guest interactions from two different methods has $\rho = 1$. Since DFT was chosen as the reference method, a higher correlation with DFT suggests that the method is better in capturing trends in binding energies.

Fig. 2b summarizes all pairwise correlation coefficients. First, we observe that binding energies from DFT optimizations at constant pressure, henceforth denoted DFT (opt, P), correlate well with those from DFT (MD, P, $N = 40$) simulations ($\rho = 0.82$). The best FF strategy to calculate binding energies is the frozen pose method at constant volume, whose correlation coefficient with DFT (MD, P) is 0.78. MD-derived energies from simulations with the NVT ensemble have a similar correlation with DFT (MD, P) ($\rho = 0.77$). These values are in excellent agreement with the baseline of $\rho = 0.82$ from the two DFT methods. The same trend is observed if DFT (opt, P, $N = 272$) energies are adopted as reference, with an even higher statistical power due to the larger number and diversity of poses. FF (frz, V) outperforms other methods by achieving a correlation of $\rho = 0.68$ with DFT (opt, P) binding affinities, followed by FF (MD, V) ($\rho = 0.55$). Although MD simulations in principle allow sampling a larger fraction of molecular conformations within the guest, average binding energies derived from constant volume MD simulations are extremely correlated with their frozen pose counterparts ($\rho = 0.88$, see also Fig. S4). This suggests that further exploring the phase space beyond the local minimum does not significantly change the trends in binding energies for most cases. To ensure this subsampling effect was not due to short trajectories, we increased the total time of the FF-based MD simulation to 30 ps, 5 ps of which were dedicated to an initial equilibration run. Nevertheless, we did not find significant changes in the average energies obtained from longer trajectories. This result has important consequences. It demonstrates that frozen pose calculations are slightly better predictors of reference binding energies than MD simulations while also being orders
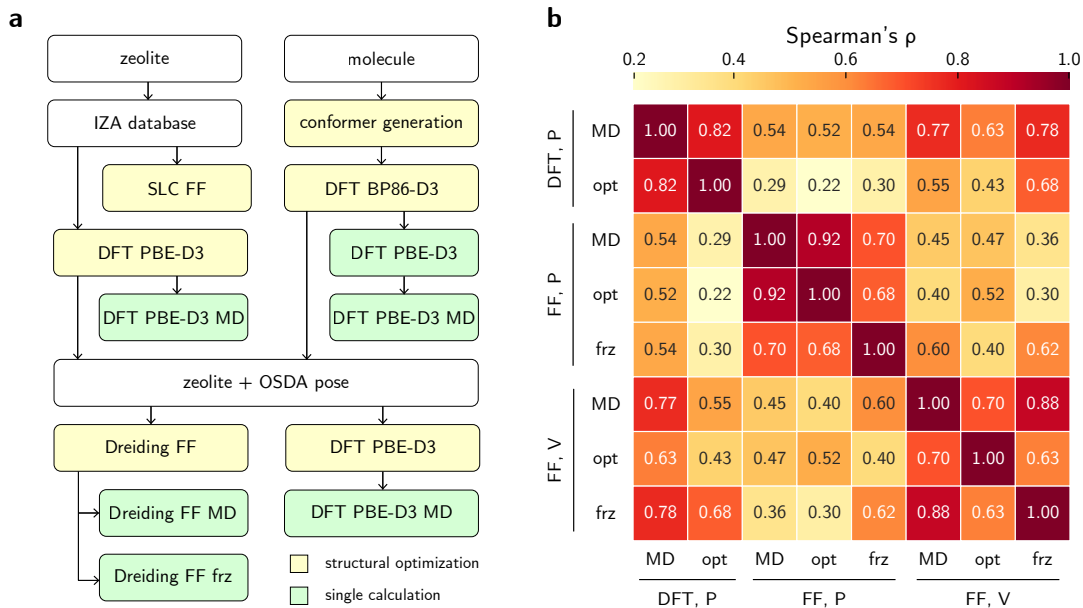
FIG. 2. **a**, Dependency tree of the atomistic simulations performed in this work. The docking algorithm generates poses from host structures optimized using PBE-D3 or SLC and ligands optimized with BP86-D3. Then, subsequent calculations with Dreiding and PBE-D3 are performed for each pose. **b**, Correlation matrix of binding energies calculated with different methods. The value of Spearman's correlation coefficient is reported in each element of the matrix.

of magnitude faster to compute.

Additionally, we also observe in Fig. 2b that the correlation between FF binding energies from constant pressure calculations and DFT is much worse than the one obtained by their constant volume counterparts. Although the physical reality is in principle better described by a constant pressure constraint, the Dreiding FF does not correctly capture the behavior of the isolated silicate frameworks, often leading to unphysical distortions in the zeolite structure. As such, the configuration space sampled by FF calculations at constant pressure is further from the ground truth than that from constant volume calculations, leading to much poorer predictions of binding energy trends. Indeed, an analysis of the density of pure-silica zeolites shows that constant pressure FF optimizations lead to structures which are 45% denser, on average, than their experimental counterparts (see Fig. S5).

B. Stability of structural optimizers for reference hosts

Even at constant volume calculations, structural optimizations of unloaded zeolites affect the binding energies of zeolite-OSDA pairs by changing the host reference energy E_h in Eq. (1). If the minimization algorithms employed in the atomic relaxation are unable to find the global energy minimum for a given structure, then all binding energies for that zeolite will be lower than the ground truth, as $E_h > E_h^{(\text{global})}$. While the stability of geometry optimization algorithms has

been studied before,[103,104] their effects are investigated here to quantify the errors in binding energies of zeolites. We compared the Dreiding energies of pure-silica, unloaded zeolites optimized at constant volume through four algorithms: the conjugate gradient (CG) method, rational function optimization (RFO), Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS), and symmetry-lowering method (Lower), all as implemented in GULP. To accelerate the convergence of the structural relaxation, we started the optimization with the CG, BFGS or Lower methods, and switched to the RFO method when the norm of the gradient ($|G|$) was smaller than a threshold of choice. In principle, if the algorithms were equally effective in finding the global energy minimum, all structures would converge to the same ground state. However, the results indicate that the algorithms often disagree on the equilibrium energies and geometries. Fig. 3 shows the distribution of energies for zeolites with respect to the minimum energy configuration for the same framework found among all four optimization runs. No optimization scheme outperforms the others across all zeolites. Whereas at least 75% of the energy differences tend to be smaller than 1 kJ/mol SiO_2, some algorithms tend to overestimate the energy of the ground state of a zeolite by up to 100 kJ/mol SiO_2. Changing the $|G|$ threshold also leads to different energy minima. If we use BFGS as initial optimizer, but switch to RFO at different values of $|G|$, the outcomes of the simulation are different. We did not find a threshold for $|G|$ that outperforms others, nor did we find significant advantage in not using RFO altogether. In fact, avoiding switching to RFO often increases the number of steps necessary to reach a

local energy minimum.

While optimization energies $E_h^{(\text{opt},V)}$ can get trapped in local minima, MD simulations allow atoms to move and avoid these higher energy traps. Thus, reference energies $\left\langle E_h^{(\text{MD},V)} \right\rangle$ should be more robust to the optimization methods used to generate the initial structures. Fig. 3 compares $\left\langle E_h^{(\text{MD},V)} \right\rangle$ differences according to the algorithm that generated the structure used as input for the MD simulation. MD energies lower the differences between the optimizers, but the discrepancies are still as high as 10 kJ/mol SiO_2 for some systems. This suggests that structural relaxations can affect even time-averaged energy references. Therefore, finding the global ground state energy of different hosts is unlikely without a thorough combination of different minimizers. These systematic shifts explain the existence of binding energies of large absolute value, such as the ones observed in ACO zeolite (see Table S3). Rather than exposing a flaw in the calculation, they suggest that even after multiple optimization attempts, the isolated host has not converged to its ground state geometry. As a consequence, all binding energies for that particular host are biased if Eq. (3) is used to compute the binding affinity of an OSDA towards a zeolite. When OSDAs are ranked across a single zeolite framework according to their binding energy,[99] this is often not a problem. Since the energy reference is shifted by the same amount for all poses, the trends of binding energies along a single host are preserved, but comparison for different hosts is hindered. Nevertheless, non-systematic errors are undesirable in high-throughput discovery of materials and should be avoided. The role of energy references in binding energy calculations does not alter the conclusions drawn from Fig. 2b. If we analyze the correlations between binding energies on a single framework (see Fig. S6 for an analysis on SOD zeolite), the same trends with respect to DFT energies are still observed. The only difference is that a larger correlation between FF binding energies at constant pressure and volume is found, and also between FF and DFT binding energies at constant pressure.

C. Variability of binding energies according to the initial substrate parameters

Another constraint required in constant volume simulations is fixing the host lattice parameters prior to docking. Since methods such as DFT and SLC lead to different equilibrium lattice constants, often with SLC predicting denser zeolites (see Fig. S5b), host-guest interactions are also affected by the choices of unit cells. To compare the influence of the initial host parameters in the final binding energy, we performed the docking procedure a second time for 81 different complexes using either the DFT or SLC geometry of the host (Fig. S3d) as an input. The VOID package generated an average of 14 poses for each host geometry and guest con-
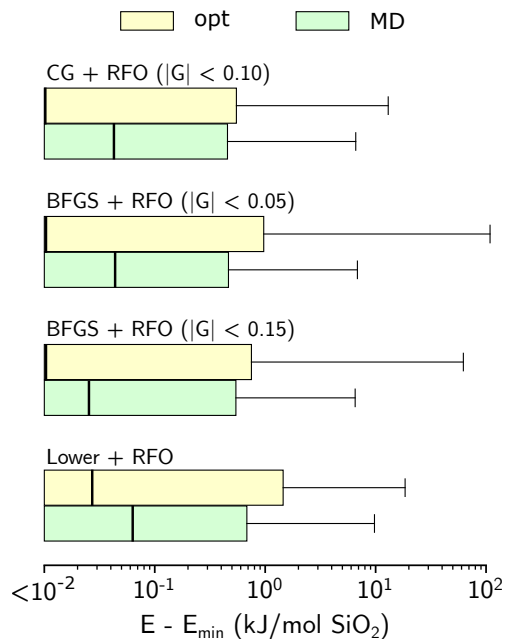

FIG. 3. Distribution of Dreiding FF energies of unloaded zeolite frameworks (E) with respect to the minimum energy obtained among all optimization algorithms for the same framework (E_{min}). The shaded box represents the interquartile range, the vertical line is the median, and the whiskers span the range of the distribution. DFT-optimized frameworks were used as inputs for the Dreiding FF optimization.

former, which has been shown to generate binding energies that qualitatively correlate with typical experimental outcomes.[99] Host energies were obtained by relaxing each unloaded zeolite with the four optimization schemes shown in Fig. 3 and selecting the resulting energy minimum as the reference value. Unit cells from DFT- and SLC-optimized frameworks were assigned different reference energies, since all relaxations were performed at constant volume. Then, we selected the strongest binding affinities among all poses created with the given guests, hosts and their lattice parameters.[99] Fig. 4 compares the best binding energies for each complex according to the starting host. Ideally, small changes in unit cell parameters should not significantly affect the ground state host-guest interactions, as long as the configuration space is thoroughly explored. However, we observe that optimization and MD binding energies vary significantly as a function of the initial substrate. The MAE between the best binding energies is 52.7 and 50.3 kJ/mol OSDA for opt and MD methods, respectively. In contrast, the best binding energies from the frozen pose method are consistent across different initial zeolite structures, with a MAE of 10.1 kJ/mol OSDA, five times better than the other methods. This suggests that binding energies from the frozen pose method are more robust to variations of unit cell parameters, rendering them a desirable choice for high-throughput computational workflows.
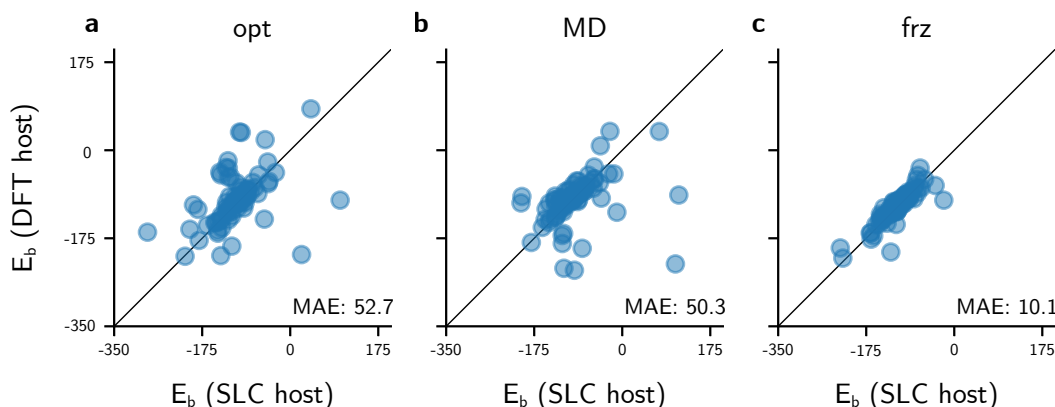
FIG. 4. Correlation between the minimum FF-based binding energy (E_b) for complexes with different initial host lattices, as calculated by the **a**, optimization, **b**, MD, and **c**, frozen pose methods. The unloaded zeolites were independently optimized using DFT and FF-SLC before the docking (see Fig. 2a). The MAE indicates that the frozen pose method is more robust to variations of the initial conditions. All values are given in kJ/mol OSDA.

IV. DISCUSSION

For decades, host-guest interactions have been modeled using a variety of methods. Yet, high-throughput screening methods typically require selecting parameters that yield computationally-efficient results and can be deployed robustly without manual supervision. While constant pressure simulations better represent the synthesis conditions of zeolites, Dreiding FF binding energies at the NPT ensemble do not correlate well with their DFT counterparts. Rather, FF simulations at constant volume show good correlation with DFT optimizations and MD simulations at constant pressure. This might be a limitation on the Dreiding FF, which is unable to correctly describe the unloaded zeolite framework when volume relaxation is allowed. Several other general-purpose force fields and parametrizations specific to zeolites have been proposed and could be benchmarked according to the guidelines discussed here.

Even when binding energies are compared across different simulation pathways and optimization methods within a single FF parametrization, results vary drastically. We have shown that energies from structure optimizations and MD simulations are more susceptible to initialization issues than frozen pose methods. The higher correlation between the latter and the DFT binding energies is also supportive of this robustness. Furthermore, contrary to intuition, a larger sampling of the configuration space through MD simulations does not necessarily lead to significant changes in trends of binding energies when compared to the frozen pose method. In practice, this conclusion opens an opportunity for simulating zeolite-OSDA pairs in larger scales. One of the major bottlenecks of zeolite-OSDA simulations is to simulate long MD trajectories of guests docked inside the host for a variety of loadings and initial configurations, as has been typically done in screening works the field.[41–43]

We propose to replace MD simulations by frozen pose methods within FF calculations, drastically reducing the time necessary to perform computations while increasing the robustness of the final binding energy with respect to the choice of optimization algorithms and unloaded zeolite geometry. Even if FF (MD, V) binding energies are better predictors of experimental outcomes than DFT (MD, P), which has yet to be verified, the use of the frozen pose method is still justified by its correlation with the former. Moreover, we show that the absolute values of the binding energy tend to be more transferable across different initial configurations through this method, suggesting it can also be used to compare the influence of each molecule in stabilizing different zeolite frameworks.[99]

It is important to note that the chemical space used in this analysis is comprised of neutral molecules. However, most known OSDAs are positively charged and direct the formation of zeolites with heteroatoms in their backbone. Theoretical studies on how OSDAs affect the position of the heteroatoms have been developed,[105–110] although at a high computational cost. Nevertheless, we suggest that the current analysis should be transferable to cationic OSDAs as well. OSDAs are often modeled without charges in FFs, and vdW interactions tend to dominate the templating effects.[2,13,24,33,41–43,45,83] It is also assumed that trends in binding energy in pure-silica frameworks hold with changes in zeolite composition. More rigorous analyses would be necessary to simulate charged OSDAs in zeolites through DFT calculations. Typical methods of charge-compensating background potentials may shift energy differences depending on the system, and combinatorial studies on heteroatom distribution are prone to be computationally expensive.

Finally, benchmarks enable the development of high-throughput computation infrastructures by providing clear guidelines for simulating materials in large scales.

To support further dissemination of these ideas, we are releasing the Python interface to the GULP code used to perform these calculations as the package GULPy, as well as the data generated in this article.[111] They lay down standards to test and automate calculation workflows of OSDAs and zeolites with increased reliability.

V. CONCLUSIONS

In summary, we benchmarked different methods to calculate binding energies of OSDAs in zeolites by performing DFT and FF calculations for 272 zeolite-OSDA pairs. We showed that Dreiding FF binding energies calculated with the frozen pose method correlate best with DFT energies. This method offers additional robustness to the binding energy with respect to the choice of geometry optimization algorithms and initial docking conditions. On the other hand, a larger sampling of the phase space from MD simulations does not provide significant benefits, since MD-based binding energies correlate very well with those from the frozen pose method. Remarkably, simulations at constant volume significantly outperform those at constant pressure within the Dreiding FF. This might result from the inability of this parametrization to correctly model the unloaded, pure-silica zeolite structure. These results provide reliable parameters for high-throughput computation of binding energies for zeolites and OSDAs. This benchmark, code and data aims to enable robust, large-scale screenings of OSDAs with significantly less computational overhead.

VI. ACKNOWLEDGEMENTS

REFERENCES

[1] M. E. Davis, "Ordered porous materials for emerging applications," Nature **417**, 813–821 (2002).

[2] R. F. Lobo, S. I. Zones, and M. E. Davis, "Structure-direction in zeolite synthesis," Journal of Inclusion Phenomena and Molecular Recognition in Chemistry **21**, 47–78 (1995).

[3] C. S. Cundy and P. A. Cox, "The hydrothermal synthesis of zeolites: History and development from the earliest days to the present time," Chemical Reviews **103**, 663–701 (2003).

[4] C. S. Cundy and P. A. Cox, "The hydrothermal synthesis of zeolites: Precursors, intermediates and reaction mechanism," Microporous and Mesoporous Materials **82**, 1–78 (2005).

[5] B. Xie, J. Song, L. Ren, Y. Ji, J. Li, and F.-S. Xiao, "Organotemplate-Free and Fast Route for Synthesizing Beta Zeolite," Chemistry of Materials **20**, 4533–4535 (2008).

[6] B. Marler and H. Gies, "Hydrous layer silicates as precursors for zeolites obtained through topotactic condensation: a review," European Journal of Mineralogy **24**, 405–428 (2012).

[7] M. Maldonado, M. D. Oleksiak, S. Chinta, and J. D. Rimer, "Controlling crystal polymorphism in organic-free synthesis of Na-zeolites," Journal of the American Chemical Society **135**, 2641–2652 (2013).

[8] P. Eliášová, M. Opanasenko, P. S. Wheatley, M. Shamzhy, M. Mazur, P. Nachtigall, W. J. Roth, R. E. Morris, and J. Čejka, "The ADOR mechanism for the synthesis of new zeolites," Chemical Society Reviews **44**, 7177–7206 (2015).

[9] S. Goel, S. I. Zones, and E. Iglesia, "Synthesis of Zeolites via Interzeolite Transformations without Organic Structure-Directing Agents," Chemistry of Materials **27**, 2056–2066 (2015).

[10] K. Itabashi, Y. Kamimura, K. Iyoki, A. Shimojima, and T. Okubo, "A Working Hypothesis for Broadening Framework Types of Zeolites in Seed-Assisted Synthesis without Organic Structure-Directing Agent," Journal of the American Chemical Society **134**, 11542–11549 (2012).

[11] L. Van Tendeloo, E. Gobechiya, E. Breynaert, J. A. Martens, and C. E. A. Kirschhock, "Alkaline cations directing the transformation of FAU zeolites into five different framework types," Chemical Communications **49**, 11737–11739 (2013).

[12] D. Schwalbe-Koda, Z. Jensen, E. Olivetti, and R. Gómez-Bombarelli, "Graph similarity drives zeolite diffusionless transformations and intergrowth," Nature Materials **18**, 1177–1181 (2019).

[13] A. W. Burton, S. I. Zones, and S. Elomari, "The chemistry of phase selectivity in the synthesis of high-silica zeolites," Current Opinion in Colloid and Interface Science **10**, 211–219 (2005).

[14] M. Moliner, F. Rey, and A. Corma, "Towards the Rational Design of Efficient Organic Structure-Directing Agents for Zeolite Synthesis," Angewandte Chemie International Edition **52**, 13880–13889 (2013).

[15] J. Li, A. Corma, and J. Yu, "Synthesis of new zeolite structures," Chemical Society Reviews **44**, 7112–7127 (2015).

[16] C. Baerlocher, L. B. McCusker, and D. H. Olson, *Atlas of Zeolite Framework Types*, 6th ed. (Elsevier Science, Amsterdam, 2007) p. 404.

[17] Ch. Baerlocher and L.B. McCusker, "Database of Zeolite Structures," http://www.iza-structure.org/databases/ (2020).

[18] E. M. Gallego, M. T. Portilla, C. Paris, A. León-Escamilla, M. Boronat, M. Moliner, and A. Corma, '"Ab initio" synthesis of zeolites for preestablished catalytic reactions," Science **355**, 1051–1054 (2017).

[19] M. Dusselier and M. E. Davis, "Small-Pore Zeolites: Synthesis and Catalysis," Chemical Reviews **118**, 5265–5329 (2018).

[20] B. M. Lok, T. R. Cannan, and C. A. Messina, "The role of organic molecules in molecular sieve synthesis," Zeolites **3**, 282–291 (1983).

[21] H. Gies and B. Marker, "The structure-controlling role of organic templates for the synthesis of porosils in the systems SiO2/template/H2O," Zeolites **12**, 42–49 (1992).

[22] R. E. Boyett, "A quantitative shape analysis of organic templates employed in zeolite synthesis," Zeolites **17**, 508–512 (1996).

[23] D. W. Lewis, D. J. Willock, C. R. A. Catlow, J. M. Thomas, and G. J. Hutchings, "De novo design of structure-directing agents for the synthesis of microporous solids," Nature **382**, 604–606 (1996).

[24] D. W. Lewis, C. M. Freeman, and C. R. Catlow, "Predicting the templating ability of organic additives for the synthesis of microporous materials," Journal of Physical Chemistry **99**, 11194–11202 (1995).

[25] S. I. Zones, Y. Nakagawa, L. T. Yuen, and T. V. Harris, "Guest/host interactions in high silica zeolite synthesis: [5.2.1.02.6]tricyclodecanes as template molecule," Journal of the

American Chemical Society **118**, 7558–7567 (1996).

[26] D. J. Willock, D. W. Lewis, C. A. Catlow, G. J. Hutchings, and J. M. Thomas, "Designing templates for the synthesis of microporous solids using de novo molecular design methods," Journal of Molecular Catalysis A: Chemical **119**, 415–424 (1997).

[27] C. R. A. Catlow, D. S. Coombes, D. W. Lewis, and J. C. G. Pereira, "Computer Modeling of Nucleation, Growth, and Templating in Hydrothermal Synthesis," Chemistry of Materials **10**, 3249–3265 (1998).

[28] Y. Nakagawa, G. S. Lee, T. V. Harris, L. T. Yuen, and S. I. Zones, "Guest/host relationships in zeolite synthesis: Ring-substituted piperidines and the remarkable adamantane mimicry by 1-azonio spiro [5.5] undecanes," Microporous and Mesoporous Materials **22**, 69–85 (1998).

[29] P. Wagner, Y. Nakagawa, G. S. Lee, M. E. Davis, S. Elomari, R. C. Medrud, and S. I. Zones, "Guest/host relationships in the synthesis of the novel cage-based zeolites SSZ-35, SSZ-36, and SSZ-39," Journal of the American Chemical Society **122**, 263–273 (2000).

[30] A. Burton and S. Elomari, "SSZ-60: a new large-pore zeolite related to ZSM-23," Chemical Communications **22**, 2618 (2004).

[31] R. Simancas, D. Dari, N. Velamazán, M. T. Navarro, A. Cantín, J. L. Jordá, G. Sastre, A. Corma, and F. Rey, "Modular organic structure-directing agents for the synthesis of zeolites," Science **330**, 1219–1222 (2010).

[32] J. E. Schmidt, M. W. Deem, and M. E. Davis, "Synthesis of a Specified, Silica Molecular Sieve by Using Computationally Predicted Organic Structure-Directing Agents," Angewandte Chemie International Edition **53**, 8372–8374 (2014).

[33] T. M. Davis, A. T. Liu, C. M. Lew, D. Xie, A. I. Benin, S. Elomari, S. I. Zones, and M. W. Deem, "Computationally Guided Synthesis of SSZ-52: A Zeolite for Engine Exhaust Clean-up," Chemistry of Materials **28**, 708–711 (2016).

[34] S. K. Brand, J. E. Schmidt, M. W. Deem, F. Daeyaert, Y. Ma, O. Terasaki, M. Orazov, and M. E. Davis, "Enantiomerically enriched, polycrystalline molecular sieves," Proceedings of the National Academy of Sciences of the United States of America **114**, 5101–5106 (2017).

[35] M. Kumar, Z. J. Berkson, R. J. Clark, Y. Shen, N. A. Prisco, Q. Zheng, Z. Zeng, H. Zheng, L. B. McCusker, J. C. Palmer, B. F. Chmelka, and J. D. Rimer, "Crystallization of Mordenite Platelets using Cooperative Organic Structure-Directing Agents," Journal of the American Chemical Society **141**, 20155–20165 (2019).

[36] D. Jo and S. B. Hong, "Targeted Synthesis of a Zeolite with Pre-established Framework Topology," Angewandte Chemie International Edition **58**, 13845–13848 (2019).

[37] R. Millini, F. Frigerio, G. Bellussi, G. Pazzuconi, C. Perego, P. Pollesel, and U. Romano, "A priori selection of shape-selective zeolite catalysts for the synthesis of 2,6-dimethylnaphthalene," Journal of Catalysis **217**, 298–309 (2003).

[38] S. B. Hong, E. G. Lecar, P. A. Wright, W. Zhou, P. A. Cox, C. H. Shin, J. H. Park, and I. S. Nam, "Synthesis, Structure Solution, Characterization, and Catalytic Properties of TNU-10: A High-Silica Zeolite with the STI Topology," Journal of the American Chemical Society **126**, 5817–5826 (2004).

[39] L. Gómez-Hortigüela, F. Corà, C. R. A. Catlow, and J. Pérez-Pariente, "Computational study of the structure-directing effect of benzylpyrrolidine and its fluorinated derivatives in the synthesis of the aluminophosphate AlPO-5," Journal of the American Chemical Society **126**, 12097–12102 (2004).

[40] L. Shi, J. Li, F. Duan, J. Yu, Y. Li, and R. Xu, "[C3N2H12]·[MnAl3P 4O17] · [H3O]: A manganese (II)-substituted aluminophosphate with zeotype AFN topology," Microporous and Mesoporous Materials **85**, 252–259 (2005).

[41] R. Pophale, F. Daeyaert, and M. W. Deem, "Computational prediction of chemically synthesizable organic structure directing agents for zeolites," Journal of Materials Chemistry A **1**, 6750–6760 (2013).

[42] F. Daeyaert and M. W. Deem, "Design of Organic Structure-Directing Agents for the Controlled Synthesis of Zeolites for Use in Carbon Dioxide/Methane Membrane Separations," ChemPlusChem **84**, 1–9 (2019).

[43] K. Muraoka, W. Chaikittisilp, and T. Okubo, "Multi-objective de novo molecular design of organic structure-directing agents for zeolites using nature-inspired ant colony optimization," Chemical Science **11**, 8214–8223 (2020).

[44] R. Millini, L. C. Carluccio, A. Carati, G. Bellussi, C. Perego, G. Cruciani, and S. Zanardi, "ERS-12: A new layered tetramethylammonium silicate composed by ferrierite layers," Microporous and Mesoporous Materials **74**, 59–71 (2004).

[45] A. W. Burton, G. S. Lee, and S. I. Zones, "Phase selectivity in the syntheses of cage-based zeolite structures: An investigation of thermodynamic interactions between zeolite hosts and structure directing agents by molecular modeling," Microporous and Mesoporous Materials **90**, 129–144 (2006).

[46] S. I. Zones, A. W. Burton, G. S. Lee, and M. M. Olmstead, "A study of piperidinium structure-directing agents in the synthesis of silica molecular sieves under fluoride-based conditions," Journal of the American Chemical Society **129**, 9066–9079 (2007).

[47] E. Jaramillo, C. P. Grey, and S. M. Auerbach, "Molecular dynamics studies of hydrofluorocarbons in faujasite-type zeolites: Modeling guest-induced cation migration in dry zeolites," Journal of Physical Chemistry B **105**, 12319–12329 (2001).

[48] G. Sastre, S. Leiva, M. J. Sabater, I. Gimenez, F. Rey, S. Valencia, and A. Corma, "Computational and Experimental Approach to the Role of Structure-Directing Agents in the Synthesis of Zeolites: The Case of Cyclohexyl Alkyl Pyrrolidinium Salts in the Synthesis of β, EU-1, ZSM-11, and ZSM-12 Zeolites," The Journal of Physical Chemistry B **107**, 5432–5440 (2003).

[49] G. Sastre, A. Cantin, M. J. Diaz-Cabañas, and A. Corma, "Searching Organic Structure Directing Agents for the Synthesis of Specific Zeolitic Structures: An Experimentally Tested Computational Study," Chemistry of Materials **17**, 545–552 (2005).

[50] A. Chawla, R. Li, R. Jain, R. J. Clark, J. G. Sutjianto, J. C. Palmer, and J. D. Rimer, "Cooperative effects of inorganic and organic structure-directing agents in ZSM-5 crystallization," Molecular Systems Design and Engineering **3**, 159–170 (2018).

[51] M. Gálvez-Llompart, A. Cantín, F. Rey, and G. Sastre, "Computational screening of structure directing agents for the synthesis of zeolites. A simplified model," Zeitschrift fur Kristallographie - Crystalline Materials **234**, 451–460 (2019).

[52] L. Gómez-Hortigüela, S. Hamad, F. López-Arbeloa, A. B. Pinar, J. Perez-Pariente, and F. Corà, "Molecular insights into the self-aggregation of aromatic molecules in the synthesis of nanoporous aluminophosphates: A multilevel approach," Journal of the American Chemical Society **131**, 16509–16524 (2009).

[53] H. Fang, P. Kamakoti, J. Zang, S. Cundy, C. Paur, P. I. Ravikovitch, and D. S. Sholl, "Prediction of CO 2 adsorption properties in zeolites using force fields derived from periodic dispersion-corrected DFT calculations," Journal of Physical Chemistry C **116**, 10692–10701 (2012).

[54] H. Fang, R. Awati, S. E. Boulfelfel, P. I. Ravikovitch, and D. S. Sholl, "First-Principles-Derived Force Fields for CH4 Adsorption and Diffusion in Siliceous Zeolites," Journal of Physical Chemistry C **122**, 12880–12891 (2018).

[55] E. O. Pyzer-Knapp, C. Suh, R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, A. A. A. Aspuru-Guzik, and D. R. Clarke, "What Is High-Throughput Virtual Screening? A Perspective from Organic Materials Discovery," Annual Review of Materials Research **45**, 195–216 (2015).

[56] R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, T. D. Hirzel, D. K. Duvenaud, D. Maclaurin, M. A. Blood-Forsythe, H. S. Chae, M. Einzinger, D.-G. G. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. Kim, W. Huang, S. I. Hong, M. Baldo, R. P. Adams, and A. Aspuru-Guzik, "Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach,"

Nature Materials **15**, 1120–1127 (2016).

[57] G. Kresse and J. Furthmüller, "Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set," Computational Materials Science **6**, 15–50 (1996).

[58] G. Kresse and J. Furthmüller, "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set," Physical Review B **54**, 11169–11186 (1996).

[59] P. E. Blöchl, "Projector augmented-wave method," Physical Review B **50**, 17953–17979 (1994).

[60] G. Kresse and D. Joubert, "From ultrasoft pseudopotentials to the projector augmented-wave method," Physical Review B **59**, 1758–1775 (1999).

[61] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple," Physical Review Letters **77**, 3865–3868 (1996).

[62] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," The Journal of Chemical Physics **132**, 154104 (2010).

[63] S. Grimme, S. Ehrlich, and L. Goerigk, "Effect of the damping function in dispersion corrected density functional theory," Journal of Computational Chemistry **32**, 1456–1465 (2011).

[64] S. Svelle, C. Tuma, X. Rozanska, T. Kerber, and J. Sauer, "Quantum Chemical Modeling of Zeolite-Catalyzed Methylation Reactions: Toward Chemical Accuracy for Barriers," Journal of the American Chemical Society **131**, 816–825 (2009).

[65] F. Göltl, A. Grüneis, T. Bučko, and J. Hafner, "Van der Waals interactions between hydrocarbon molecules and zeolites: Periodic calculations at different levels of theory, from density functional theory to the random phase approximation and Møller-Plesset perturbation theory," The Journal of Chemical Physics **137**, 114111 (2012).

[66] F. Göltl and J. Hafner, "Modelling the adsorption of short alkanes in protonated chabazite: The impact of dispersion forces and temperature," Microporous and Mesoporous Materials **166**, 176–184 (2013).

[67] F. Göltl and P. Sautet, "Modeling the adsorption of short alkanes in the zeolite SSZ-13 using "van der Waals" DFT exchange correlation functionals: Understanding the advantages and limitations of such functionals," The Journal of Chemical Physics **140**, 154105 (2014).

[68] J. Shang, G. Li, R. Singh, P. Xiao, D. Danaci, J. Z. Liu, and P. A. Webley, "Adsorption of CO2, N2, and CH4 in Cs-exchanged chabazite: A combination of van der Waals density functional theory calculations and experiment study," The Journal of Chemical Physics **140**, 84705 (2014).

[69] G. Piccini, M. Alessio, J. Sauer, Y. Zhi, Y. Liu, R. Kolvenbach, A. Jentys, and J. A. Lercher, "Accurate Adsorption Thermodynamics of Small Alkanes in Zeolites. Ab initio Theory and Experiment for H-Chabazite," The Journal of Physical Chemistry C **119**, 6128–6137 (2015).

[70] Y. Zhang, J. Yu, Y.-H. Yeh, R. J. Gorte, S. Rangarajan, and M. Mavrikakis, "An Adsorption Study of CH4 on ZSM-5, MOR, and ZSM-12 Zeolites," The Journal of Physical Chemistry C **119**, 28970–28978 (2015).

[71] J. Sauer, "Ab Initio Calculations for Molecule-Surface Interactions with Chemical Accuracy," Accounts of Chemical Research **52**, 3502–3510 (2019).

[72] F. R. Rehak, G. Piccini, M. Alessio, and J. Sauer, "Including dispersion in density functional theory for adsorption on flat oxide surfaces, in metal-organic frameworks and in acidic zeolites," Physical Chemistry Chemical Physics **22**, 7577–7585 (2020).

[73] H. J. Monkhorst and J. D. Pack, "Special points for Brillouin-zone integrations," Physical Review B **13**, 5188–5192 (1976).

[74] W. G. Hoover, A. J. C. Ladd, and B. Moran, "High-Strain-Rate Plastic Flow Studied via Nonequilibrium Molecular Dynamics," Physical Review Letters **48**, 1818–1820 (1982).

[75] D. J. Evans, "Computer "experiment" for nonlinear thermodynamics of Couette flow," The Journal of Chemical Physics **78**, 3297–3302 (1983).

[76] M. Parrinello and A. Rahman, "Crystal Structure and Pair Potentials: A Molecular-Dynamics Study," Physical Review Letters **45**, 1196–1199 (1980).

[77] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," Journal of Applied Physics **52**, 7182–7190 (1981).

[78] J. D. Gale, "GULP: A computer program for the symmetry-adapted simulation of solids," Journal of the Chemical Society-Faraday Transactions **93**, 629–637 (1997).

[79] J. D. Gale and A. L. Rohl, "The General Utility Lattice Program (GULP)," Molecular Simulation **29**, 291–341 (2003).

[80] Available under the MIT License at `https://github.com/learningmatter-mit/gulpy`.

[81] S. Melchionna, G. Ciccotti, and B. L. Holian, "Hoover NPT dynamics for systems varying in shape and size," Molecular Physics **78**, 533–544 (1993).

[82] S. L. Mayo, B. D. Olafson, and W. A. Goddard, "DREIDING: A generic force field for molecular simulations," Journal of Physical Chemistry **94**, 8897–8909 (1990).

[83] J. E. Schmidt, D. Xie, and M. E. Davis, "High-silica, heulandite-type zeolites prepared by direct synthesis and topotactic condensation," J. Mater. Chem. A **3**, 12890–12897 (2015).

[84] M. J. Sanders, M. Leslie, and C. R. Catlow, "Interatomic potentials for SiO2," Journal of the Chemical Society, Chemical Communications , 1271–1273 (1984).

[85] J. R. Hill and J. Sauer, "Molecular mechanics potential for silica and zeolite catalysts based on ab initio calculations. 1. Dense and microporous silica," The Journal of Physical Chemistry **98**, 1238–1244 (1994).

[86] J.-R. Hill and J. Sauer, "Molecular Mechanics Potential for Silica and Zeolite Catalysts Based on ab Initio Calculations. 2. Aluminosilicates," The Journal of Physical Chemistry **99**, 9536–9550 (1995).

[87] A. F. Combariza, D. A. Gomez, and G. Sastre, "Simulating the properties of small pore silicazeolites using interatomic potentials," Chemical Society Reviews **42**, 114–127 (2013).

[88] A. Ghysels, S. L. C. Moors, K. Hemelsoet, K. De Wispelaere, M. Waroquier, G. Sastre, and V. Van Speybroeck, "Shape-Selective Diffusion of Olefins in 8-Ring Solid Acid Microporous Zeolites," The Journal of Physical Chemistry C **119**, 23721–23734 (2015).

[89] S. E. Boulfelfel, P. I. Ravikovitch, L. Koziol, and D. S. Sholl, "Improved Hill-Sauer Force Field for Accurate Description of Pores in 8-Ring Zeolites," Journal of Physical Chemistry C **120**, 14140–14148 (2016).

[90] V. Van Speybroeck, K. Hemelsoet, L. Joos, M. Waroquier, R. G. Bell, and C. R. A. Catlow, "Advances in theory and their application within the field of zeolite chemistry," Chemical Society Reviews **44**, 7044–7111 (2015).

[91] G. Landrum, "RDKit: Open-source cheminformatics," (2006).

[92] T. A. Halgren, "Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94," Journal of Computational Chemistry **17**, 490–519 (1996).

[93] P. Tosco, N. Stiefl, and G. Landrum, "Bringing the MMFF force field to the RDKit: implementation and validation," Journal of Cheminformatics **6** (2014), 10.1186/s13321-014-0037-3.

[94] Becke and A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," Physical Review A **38**, 3098–3100 (1988).

[95] J. P. Perdew, "Density-functional approximation for the correlation energy of the inhomogeneous electron gas," Physical Review B **33**, 8822–8824 (1986).

[96] F. Weigend and R. Ahlrichs, "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy," Physical Chemistry Chemical Physics **7**, 3297–3305 (2005).

[97] F. Neese, "The ORCA program system," Wiley Interdisciplinary Reviews: Computational Molecular Science **2**, 73–78 (2012).

[98] F. Neese, "Software update: the ORCA program system, version 4.0," WIREs Computational Molecular Science **8**, e1327 (2018).

[99] D. Schwalbe-Koda and R. Gomez-Bombarelli, "Supramolecular Recognition in Crystalline Nanocavities Through Monte Carlo and Voronoi Network Algorithms," ChemRxiv , 13211012 (2020).

[100] A. Corma, F. Rey, J. Rius, M. J. Sabater, and S. Valencia, "Supramolecular self-assembled molecules as organic directing agent for synthesis of zeolites," Nature **431**, 287–290 (2004).

[101] F. Daeyaert, F. Ye, and M. W. Deem, "Machine-learning approach to the design of OSDAs for zeolite beta," Proceedings of the National Academy of Sciences of the United States of America **116**, 3413–3418 (2019).

[102] G. W. Corder and D. I. Foreman, *Nonparametric statistics : a step-by-step approach*, 2nd ed. (John Wiley & Sons, Hoboken, New Jersey, 2014).

[103] B. J. Berne, G. Ciccotti, and D. F. Coker, *Classical and Quantum Dynamics in Condensed Phase Simulations* (WORLD SCIENTIFIC, 1998) p. 880.

[104] A. R. Leach, *Molecular Modelling: Principles and Applications* (Prentice Hall, 2001).

[105] M. J. Sabater and G. Sastre, "A computational study on the templating ability of the trispyrrolidinium cation in the synthesis of ZSM-18 zeolite," Chemistry of Materials **13**, 4520–4526 (2001).

[106] A. J. Jones and E. Iglesia, "The Strength of Brønsted Acid Sites in Microporous Aluminosilicates," ACS Catalysis **5**, 5741–5755 (2015).

[107] J. E. Schmidt, D. Fu, M. W. Deem, and B. M. Weckhuysen, "Template–Framework Interactions in Tetraethylammonium-Directed Zeolite Synthesis," Angewandte Chemie International Edition **55**, 16044–16048 (2016).

[108] B. C. Knott, C. T. Nimlos, D. J. Robichaud, M. R. Nimlos, S. Kim, and R. Gounder, "Consideration of the Aluminum Distribution in Zeolites in Theoretical and Experimental Catalysis Research," ACS Catalysis **8**, 770–784 (2018).

[109] K. Muraoka, W. Chaikittisilp, Y. Yanaba, T. Yoshikawa, and T. Okubo, "Directing Aluminum Atoms into Energetically Favorable Tetrahedral Sites in a Zeolite Framework by Using Organic Structure-Directing Agents," Angewandte Chemie International Edition **57**, 3742–3746 (2018).

[110] J. R. Di Iorio, S. Li, C. B. Jones, C. T. Nimlos, Y. Wang, E. Kunkes, V. Vattipalli, S. Prasad, A. Moini, W. F. Schneider, and R. Gounder, "Cooperative and Competitive Occlusion of Organic and Inorganic Structure-Directing Agents within Chabazite Zeolites Influences Their Aluminum Arrangement," Journal of the American Chemical Society **142**, 4807–4819 (2020).

[111] Available under the MIT License at `https://github.com/learningmatter-mit/Zeolite-Reproducibility-Binding`.