Development of Accurate and Reliable Methods for *In Silico* Modeling of Adsorption in Nanoporous Materials

Ontwikkeling van accurate and betrouwbare methodes voor het *in silico* modelleren van adsorptie in nanoporeuze materialen

Ruben Goeminne

Supervisors: prof. dr. ir. T. Verstraelen
Dissertation submitted in fulfillment of the requirements for the degree of Doctor (Ph.D.) of Science: Physics

Department of Physics and Astronomy
Department chair: prof. dr. Dimitri Van Neck
Ghent University
Academic year 2022–2023

GHENT UNIVERSITY

Members of the Examination Committee

Chair

prof. Jutho Haegeman (Ghent University)

Reading Committee

dr. Karen Leus (Ghent University)
dr. ir. Sven Rogge (Ghent University)
prof. dr. Rochus Schmid (Ruhr-Universität Bochum)
prof. dr. ir. Louis Vanduyfhuys (Ghent University)

Supervisor

prof. dr. ir. Toon Verstraelen (Ghent University)

This research has been conducted at the **Center for Molecular Modeling**.

Contents

Samenvatting

Menselijke technologische vooruitgang is gekenmerkt door ons gebruik van werktuigen om de omgeving om ons heen vorm te geven. In de afgelopen eeuwen hebben deze werktuigen ons toegang gegeven tot domeinen ver buiten de gebieden van menselijke ervaring, gaande van een begrip van de dynamica van hele clusters van sterrenstelsels op de grootste lengteschalen tot de structuur van quarks in protonen op de kleinste schalen.

Een werktuig dat in recente decennia wijdverspreid werd door de explosie van computationele kracht is het veld van moleculair modelleren. In dit veld worden materialen en moleculen computationeel gemodellereed op de atomaire schaal om inzicht te verwerven in hun eigenschappen. In deze doctoraatsthesis ontwikkelen we computationele methoden voor het modelleren van een klasse materialen die poriën op de nanoschaal bevatten, toepasselijk genaamd nanoporeuze materialen. De ontdekking van een subklasse hiervan, genaamd metaal-organische roosters (MOFs, afgeleid van de Engelstalige naamgeving metal-organic frameworks), heeft in de afgelopen decennia geleid tot wijdverspreide wetenschappelijke interesse. MOFs zijn kristallijne materialen bestaande uit metaaloxide bouwblokken die via organische linkers met elkaar zijn verbonden. Toen duidelijk werd dat deze MOFs een permanente porositeit konden behouden, werden ze snel geïdentificeerd als ideale materialen voor gasadsorptie en scheidingstoepassingen. Door het rationeel ontwerp van de poriestructuur van deze materialen kunnen ze op maat gemaakt worden om bij voorkeur met de ene molecule meer dan de andere te interageren, waardoor een schat aan potentiële toepassingen ontstaat.

Echter, MOFs zijn verre van statische, onveranderlijke kristallen. Hun dynamische reactie op externe stimuli heeft wetenschappers zelfs al decennia lang voor raadsels gesteld. Contra-intuïtieve verschijnselen zoals negatieve thermische uitzetting, negatieve gasadsorptie en faseovergangen die gepaard

gaan met grootschalige structurele veranderingen zijn zowel een opportuniteit voor nog fijnere controle van hun poriestructuur, maar ook een uitdaging voor de moleculaire modelleringsgemeenschap.

Een nauwkeurige computationele modellering van gasadsorptie in deze materialen steunt in wezen op twee componenten. De eerste is een betrouwbare methode om een MOF adsorbent te koppelen aan de extern opgelegde stimuli, zoals de mechanische druk, temperatuur en gasdruk van de adsorbaten, door middel van computationele *sampling* schema's. Op basis van de principes van de statistische fysica uiteengezet in **Hoofdstuk 2**, worden verschillende computationele schema's ontwikkeld en getest in **Hoofdstuk 3** op hun vermogen om de adsorbaat-geïnduceerde flexibiliteit van MIL-53(Al) te reproduceren, een prototypische MOF die faseovergangen ondergaat onder adsorptie, geassocieerd met volumeveranderingen tot 40%. Deze schema's berusten in wezen op een combinatie van moleculaire dynamica (MD) simulaties die de temperatuur en druk handhaven die op het rooster worden opgelegd, en groot-kanonisch Monte Carlo (GCMC) simulaties die het rooster koppelen aan het omringende reservoir van adsorbaat moleculen. Één daarvan is een nieuw ontwikkeld MC/MD schema dat gebruik maakt van behouden-volume MD simulaties, dat accuraat de faseovergangen van MIL-53(Al) onder CO_2, CH_4 en Xe adsorptie kan voorspellen (en het gebrek daaraan voor Ne en Ar) van de *large pore* (lp) fase naar de *narrow pore* (np) fase. Vervolgens werd dit computationele schema gebruikt om de recent ontdekte negatieve gasadsorptie (NGA) van CH_4 in DUT-49(Cu) te onderzoeken. Bij deze MOF leidt het verhogen van de methaangasdruk boven een druk van 25 kPa bij een temperatuur van 120 K tot een plotselinge ineenstorting van het rooster. Deze overgang gaat gepaard met een vrijlating van adsorbaten uit het rooster, een fenomeen dat voor het eerst in deze MOF werd waargenomen. Ons schema maakte een nauwkeurige karakterisering van de relatieve stabiliteit van de open en ingeklapte fasen mogelijk, waarbij werd aangetoond hoe de osmotische barrière tussen deze fasen het experimenteel waargenomen gedrag tijdens adsorptie en desorptie veroorzaakt. Bovendien onthulden simulaties uitgevoerd bij temperaturen van 90 K en 150 K waarom de NGA overgang alleen plaatsvindt in een klein temperatuursinterval. De verdienste van ons computationele schema is echter niet alleen beperkt tot de reproductie van experimenteel waargenomen verschijnselen. Belangrijker nog, het biedt inzicht in de drijvende krachten achter adsorptie-geïnduceerde flexibiliteit, waardoor het pad wordt geopend naar het rationeel ontwerp van MOFs, meer bepaald het precies afstemmen van materialen op onze gewenste toepassingen.

Het tweede essentiële onderdeel om adsorptie in MOFs computationeel te karakteriseren is een nauwkeurig model voor de interactie tussen gasmo-

leculen en het omringende rooster. Hoewel er zeer nauwkeurige *ab initio* methoden bestaan om deze rooster-adsorbaat interacties te berekenen, is hun computationele kost doorgaans te groot voor gebruik in MC of MD simulaties. Om deze reden hebben computationele karakteriseringen van adsorptie-eigenschappen in MOFs traditioneel gesteund op relatief eenvoudige, empirisch afgestemde interactiepotentialen, ook wel 'krachtvelden' genoemd. Hoewel deze benadering duidelijk zijn waarde heeft bewezen, bijvoorbeeld om de adsorptie-eigenschappen van grote aantallen MOFs te *screenen*, is hun soms beperkte nauwkeurigheid duidelijk aangetoond in de literatuur.

In **Hoofdstuk 4** worden twee totaal verschillende methodes onderzocht om de beschrijving van rooster-adsorbaat interacties te verbeteren. In een eerste aanpak ontwikkelen we meer nauwkeurige interactiepotentialen door deze af te leiden van *ab initio* vergelijkingen. Dit werk bouwt voort op het eerder ontwikkelde Monomer Electron Density Force Field (MEDFF), waarin transfereerbare interactiepotentialen werden ontwikkeld die functionalen zijn van de elektronendichtheid van de constituerende fragmenten (moleculen, rooster) van het systeem. In dit paarsgewijs additieve krachtveld dat slechts 3 empirisch bepaalde parameters bevat, werd bevonden dat de beschrijving van polarisatie de belangrijkste beperking was. Daarom werd in dit werk een nieuw polarisatiemodel ontwikkeld, gebaseerd op de basisprincipes van Density Functional Theory (DFT). In het resulterende model werd een vaak verwaarloosde *exchange-repulsion* interactie opgenomen. Vervolgens werd het model getest op zijn reproductie van moleculaire polariseerbaarheden en veeldeeltjes-energieën van clusters van moleculen, die aantonen hoe de inclusie van deze interactie de nauwkeurigheid aanzienlijk verbetert. Door dit nieuwe polarisatiemodel in MEDFF in te bedden, wordt een nieuw intermoleculair krachtveld verkregen, genaamd het Polarizable Monomer Electron Density Force Field (PMEDFF). Door PMEDFF te testen op zijn vermogen om dimeer interactie-energieën te reproduceren, tonen we aan hoe een nauwkeurigheid vergelijkbaar met of superieur aan gangbare krachtvelden in de literatuur kan worden verkregen, zonder de noodzaak van empirisch afgestemde parameters voor atomen in de relevante chemische omgevingen. Deze eerste resultaten zijn veelbelovend, maar er is verder werk nodig om de nauwkeurigheid van de andere interactiepotentialen in ons krachtveld te verbeteren, met aanwijzingen dat een opname van atomaire anisotropie van aanzienlijk belang kan zijn.

Deze eerste methode vertegenwoordigt een tak van ontwikkeling van krachtvelden waarbij fysisch inzicht wordt opgenomen in de interactiemodellen om zoveel mogelijk empirische parameters te elimineren (met als doel een algemeen toepasbaar tranfereerbare potentiaal te bereiken). Onze tweede

methode vertegenwoordigt de tegenovergestelde aanpak. Hierbij maken we gebruik van de recente ontwikkeling van zeer nauwkeurige en data-efficiënte Machine Learning Potentials (MLPs), waarbij geen *a priori* interactiepotentiaal wordt opgelegd. In plaats daarvan maken MLPs gebruik van een enorm groot aantal parameters (doorgaans honderdduizenden) die tijdens de training worden verfijnd aan de hand van een set referentie-energieën en krachten. In **Hoofdstuk 4** wordt een nieuwe methodologie voorgesteld en toegepast om een trainingsset te genereren die DFT intermoleculaire energieën en krachten van adsorbaten in MOFs bevat. Door training op deze set kunnen zeer nauwkeurige MLPs worden afgeleid voor gebruik in GCMC simulaties. Een *benchmark* op CO_2 adsorptie in ZIF-8 toont aan hoe adsorptie-eigenschappen met de nauwkeurigheid van DFT simulaties kunnen worden berekend met een computationele efficiëntie grootteordes groter dan DFT berekeningen. Hierdoor wordt een routine evaluatie van adsorptie-eigenschappen met een aanzienlijk verbeterde nauwkeurigheid mogelijk. Een onderzoek naar CO_2 adsorptie op de uitdagende open-metaal sites van Mg-MOF-74 toont verder aan hoe (met een geschikte dispersiecorrectie) zeer nauwkeurige adsorptie-eigenschappen kunnen worden verkregen met MLPs in chemische omgevingen waar krachtvelden doorgaans tekort schieten. Hoewel deze resultaten veelbelovend zijn, kan een getrainde MLP niet op dezelfde manier worden gebruikt als gewone krachtvelden. Voor krachtvelden wordt ervan uitgegaan dat parameters afgeleid voor atomen in bepaalde chemische omgevingen goed transfereerbaar zijn naar andere systemen die dezelfde omgevingen bevatten. Dergelijke aannames kunnen echter niet worden gemaakt voor MLPs. Dit wordt geïllustreerd in een onderzoek waarbij een MLP die is getraind op CO_2 adsorptie in ZIF-8 op andere MOFs in de ZIF-familie wordt toegepast. Voor deze andere MOFs worden grote systematische afwijkingen waargenomen in de adsorptie-isothermen en adsorptiewarmte, wat de noodzaak benadrukt van expliciete training op deze MOFs om betrouwbare resultaten te verkrijgen.

In alle voorgaande besprekingen werd met adsorptie in het algemeen fysisorptie bedoeld. Adsorptie kan echter ook verwijzen naar chemisorptie waarbij een chemische binding wordt gevormd tussen een adsorbaat en het adsorbent. Chemisorptie is geassocieerd met een activeringsbarrière en is vanuit een modelleringsperspectief moeilijk te beschrijven vanwege het ontbreken van algemeen toepasbare, nauwkeurige krachtvelden die chemische reacties kunnen modelleren. MLPs daarentegen zijn niet op dezelfde manier gelimiteerd, omdat ze in principe onafhankelijk zijn van het type intermoleculaire interacties. Als demonstratie van de haalbaarheid van het gebruik van MLPs voor chemische reacties, werd de prototypische *proton hopping* reactie tussen zuurstofatomen in de eerste coördinatiesfeer van een

Al^{3+} substitutie in de H-CHA zeoliet onderzocht. Vanwege de aanwezigheid van activeringsbarrières komen transities in reguliere MD simulaties echter zelden voor. Daarom werd een klein aantal DFT geavanceerde MD simulaties van de *proton hopping* reactie uitgevoerd bij hoge temperaturen, en gebruikt als trainingsset voor een MLP. Met de getrainde potentiaal konden zeer nauwkeurige geavanceerde MD simulaties worden uitgevoerd bij verschillende temperaturen. Bovendien maakte de MLP het gebruik van (anders computationeel onhaalbare) pad-integraal MD simulaties mogelijk om het kwantumgedrag van een proton te includeren. Een vergelijking tussen klassieke en kwantum-reactiesnelheden demonstreert de dramatische versnelling van de reactie als gevolg van het kwantumgedrag van de licht kern, een effect dat vaak genegeerd wordt in computationele studies van reacties.

Samengevat had deze doctoraatsthesis als doel om de nauwkeurigheid en betrouwbaarheid van de computationele modellering van adsorptie in flexibele nanoporeuze materialen te verbeteren. Enerzijds werden computationele *sampling* schema's onderzocht en ontwikkeld om de stabiliteit van flexibele MOFs onder adsorptie te onderzoeken. Één van deze schema's, een MC/MD schema dat de groot-kanonische Monte Carlo en behouden-volume MD technieken combineert, toonde duidelijk belofte in de modellering van de adsorptie-geïnduceerde flexibiliteit van MIL-53(Al) en DUT-49(Cu), waarbij het experimenteel waargenomen gedrag werd gereproduceerd en inzicht werd verkregen in de drijvende krachten erachter. Een tweede aspect dat noodzakelijk is voor een accurate modellering van adsorptie is de beschrijving van de rooster-adsorbaat interacties. In dit opzicht toonden twee zeer verschillende methodologieën belofte. De eerste maakt gebruik van *ab initio* afgeleide interactiepotentialen met zo min mogelijk empirische parameters, terwijl de tweede volledig op *ab initio* data berust, zonder een functionele vorm op te leggen. Hoewel de laatste aanpak de wind in de zeilen van recente ontwikkelingen van *machine learning* in zijn voordeel heeft, is het zeker niet vanzelfsprekend dat deze alle door fysica geïnspireerde potentialen zal vervangen. In plaats daarvan lijkt een samenvoeging van fysisch inzicht en *machine learning* de gouden weg voorwaarts. In ieder geval zal deze doctoraatsthesis hopelijk de lezer vertrouwen geven in de maturatie van het vakgebied van computationele modellering van adsorptie, dat steeds dichter bij de ware belofte van het vakgebied komt: een nauwkeurige en betrouwbare computationele karakterisering van alle eigenschappen van materialen die je ooit zou willen.

Summary

Human technological progress has been marked by our use of tools to shape the environment around us. In the last few centuries, these tools have given us access to domains far outside the realms of human experience, from an understanding of the dynamics of whole clusters of galaxies on the largest lengths scales, to the structures of quarks in protons on the smallest scales.

A tool that has become widely available in recent decades due to the explosion in computational power is the field of molecular modeling. Herein, materials and molecules are modeled *in silico* at the atomic scale to gain insight into their various properties. In this thesis, we develop computational tools to model a class of materials containing pores at the nanoscale, aptly named nanoporous materials. A subclass of these, coined metal-organic frameworks (MOFs), has in recent decades attracted widespread scientific interest. MOFs are crystalline materials composed of metal-oxide clusters connected through organic linkers. When it became clear that these MOFs could maintain a permanent porosity, they were quickly identified as ideal materials for gas adsorption and separation applications. Through the rational design of the pore structure of these materials, they can be tailored to interact preferentially with one molecule over another, opening up a wealth of potential applications.

However, MOFs are far from static, unchanging crystals. In fact, their dynamic response to external stimuli has continued to stump scientists for decades. Counterintuitive phenomena such as negative thermal expansion, negative gas adsorption and phase transitions associated with large-amplitude structural changes are both opportunities for even more fine-grained control of their pore structure, but also a challenge for the molecular modeling community.

An accurate computational modeling of gas adsorption in these materials relies essentially on two components. The first is a reliable sampling scheme

for coupling a MOF adsorbent to the externally applied conditions such as the mechanical pressure, temperature and gas pressure of the adsorbates. Based on the principles of statistical physics laid out in **Chapter 2**, different computational sampling schemes are developed and benchmarked in **Chapter 3** on their ability to reproduce the adsorbate-induced flexibility of MIL-53(Al), a prototypical MOF exhibiting phase transitions under adsorption associated with volume changes of up to 40%. These sampling schemes essentially rely on a combination of molecular dynamics (MD) simulations, which maintain the temperature and pressure imposed on the framework, and grand-canonical Monte Carlo (GCMC) simulations, which couple the framework to the surrounding reservoir of guest molecules. Among these, a newly developed MC/MD scheme making use of volume-constrained MD simulations was shown to accurately predict the phase transitions of MIL-53(Al) under CO_2, CH_4 and Xe adsorption (and lack thereof for Ne and Ar) from the large pore (lp) to the narrow pore (np) phase. Subsequently, this computational scheme was employed to rationalize the recent discovery of negative gas adsorption (NGA) of CH_4 in DUT-49(Cu). In this MOF, increasing the methane gas pressure above a pressure of 25 kPa at a temperature of 120 K results in a sudden collapse of the framework. This transition is associated with a release of adsorbates from the framework, a phenomenon which had not been encountered previously. Our sampling scheme allowed for a precise characterization of the relative stabilities of the 'open' and 'collapsed' phases, demonstrating how the osmotic barrier present between these phases gives rise to the experimentally observed behavior during adsorption as well as desorption. Additionally, simulations performed at temperatures of 90 K and 150 K uncovered why the NGA transition only occurs in a small temperature window. The merit of our computational scheme is, however, not only limited to the reproduction of experimentally observed phenomena. More importantly, it provides insight into the driving forces behind adsorption-induced flexibility, opening up the path towards the rational design of MOFs.

The second vital component to computationally characterize adsorption in MOFs is an accurate model for interaction between gas molecules and the surrounding framework. While highly accurate *ab initio* methods capturing these host–guest interactions exist, their computational cost is generally prohibitive for use in MC or MD simulations. For this reason, computational characterizations of adsorption properties in MOFs have traditionally relied on relatively simple empirically fitted interaction potentials, also referred to as 'force fields'. While this approach has clearly shown its merit, for example to screen the adsorption properties of vast sets of MOFs, their sometimes limited accuracy is well established in literature.

In **Chapter 4**, two very different pathways to an improved description of host–guest interactions are investigated. In a first approach, we aim to develop more accurate *ab initio* derived interaction potentials. This work builds on the previously developed Monomer Electron Density Force Field (MEDFF), in which transferable interaction potentials are developed which are functionals of the electron density of the constituting fragments (molecules, framework) of the system. In this pairwise additive force field containing only 3 empirically fitted parameters, the description of polarization was found to be its main limitation. Therefore, a new polarization model was developed, grounded in the fundamentals of Density Functional Theory (DFT). In the resulting model, a commonly neglected Pauli repulsion interaction was included. Subsequent benchmarks on the model's reproduction of molecular polarizabilities and cluster many-body energies demonstrate how the inclusion of this interaction significantly improves its accuracy. Embedding this new polarization model in MEDFF, a new intermolecular force field coined the Polarizable Monomer Electron Density Force Field (PMEDFF) is attained. Testing PMEDFF in its ability to reproduce dimer interaction energies, we demonstrate how an accuracy comparable or superior to common literature force fields can be attained, without the need for empirically fitted parameters for atoms in each encountered chemical environment. These initial results are promising, however further work is required to improve the accuracy of the other interaction potentials in our force field, with indications that an inclusion of atomic anisotropy can be of significant importance.

This first approach represents a branch of force field development in which physical insight is incorporated into the interaction models to eliminate as many empirical parameters as possible (with the aim to attain a generally applicable transferable potential). Our second approach represents the polar opposite. Leveraging the recent development of highly accurate and data-efficient Machine Learning Potentials (MLPs), no *a priori* interaction potential is imposed. Instead, MLPs employ a comparatively vast set of parameters (commonly hundreds of thousands) which are refined during training to a set of reference energies and forces. In **Chapter 4**, a new methodology is proposed and employed to generate a training set containing DFT intermolecular energies and forces of adsorbates in MOFs. Training on this set, highly accurate MLPs can be derived for use in GCMC simulations. A benchmark on CO_2 adsorption in ZIF-8 demonstrates how adsorption properties can be computed at DFT accuracy with a computational efficiency orders of magnitude larger than DFT calculations, unlocking a routine evaluation of adsorption properties with significantly improved accuracy. An investigation of CO_2 adsorption on the challenging open-metal sites of Mg-MOF-74 further demonstrates how (with an appropriately chosen dispersion

correction) highly accurate adsorption properties can be obtained with MLPs in chemical environments which are challenging for force fields to properly capture. While these results hold great promise, a trained MLP cannot be employed in the same way as common force fields. For the latter, parameters derived for atoms in certain chemical environments are assumed to transfer well to different systems containing the same environments. However, such assumptions cannot be made for MLPs. This is exemplified from an investigation employing an MLP trained on CO_2 adsorption in ZIF-8 to other MOFs in the ZIF family. For these, large systematic deviations on the adsorption isotherms and heats of adsorption are observed, highlighting the need for explicit training on these MOFs to obtain reliable results.

In all preceding discussions, adsorption was generally conflated with physisorption. However, adsorption can also refer to chemisorption in which a chemical bond is formed between the adsorbate and adsorbent. Chemisorption is associated with an activation barrier and is a challenging process from a modeling point of view due to the lack of generally applicable, accurate reactive force fields. Contrary to most force fields, MLPs are not limited in this way, being in principle agnostic to the type of interactions. As a demonstration for the feasibility of employing MLPs for reactive events, the prototypical proton hopping reaction between oxygens in the first coordination sphere of an Al^{3+} substitution in the H-CHA zeolite was investigated. Due to the presence of activation barriers, unbiased MD simulations only rarely sample transitions. Therefore, a small set of high-temperature DFT enhanced sampling simulations of the proton hopping reaction were performed, serving as training data for an MLP. With the trained potential, highly accurate enhanced sampling simulations could be performed at a range of temperatures. Moreover, the use of an MLP enabled the use of (otherwise prohibitively expensive) path-integral MD simulations to capture the quantum behavior of a proton. A comparison between classical and quantum reaction rate constants demonstrates the dramatic speed-up resulting from the quantum behavior of a light nucleus, an effect which has commonly been neglected in computation studies of reactive events.

In conclusion, this Ph.D. dissertation aimed at improving the accuracy and reliability of a computational modeling of adsorption in flexible nanoporous materials. On one hand, computational sampling schemes were investigated and developed to probe the stability of flexible MOFs under adsorption. Of these, an MC/MD scheme combining grand-canonical Monte Carlo moves with volume-constrained MD moves demonstrated clear promise in modeling the adsorption-induced flexibility of MIL-53(Al) and DUT-49(Cu), capturing the experimentally observed behavior and yielding insight into the driving forces behind it. On the other hand, an improved description of

the host–guest interactions giving rise to adsorption was investigated. In this regards, two very distinct approaches showed promise. One relies on *ab initio* derived interaction potentials with as little empirical parameters as possible, while the other relies purely on data, without imposing a functional form. While the latter has the recent wind in the sails of machine learning techniques in its favor, it is certainly not a given that it will replace all physically-inspired potentials. Instead, a merging between physical insight and machine learning appears like the golden path forwards. In any case, this dissertation will hopefully fill the reader with confidence about the ever maturing field of computational modeling of adsorption, coming closer to the field's true promise, an accurate and reliable *in silico* characterization of all material's properties your heart could ever desire.

List of Symbols

In this thesis, italicized variables indicate scalars and bold-faced variables indicate vectors or tensors.

Alphanumerical symbols	
C_n^{ij}	The n^{th} order dispersion coefficient between atoms i and j
E	Energy
E_0	Ground state energy
E_{kin}	Kinetic energy
E_n	Energy of the n^{th} eigenstate (general)
E_n^e	Electronic energy of the n^{th} eigenstate
E_{xc}	Exchange-correlation energy
$\hat{\mathcal{F}}$	Fock operator
$\boldsymbol{F}$	Fock matrix
$\boldsymbol{F}_i$	Force acting on the i^{th} atomic nucleus
F	Helmhotz free energy in the canonical ensemble
G	Gibbs free energy in the isobaric-isothermal ensemble
$\hat{h}$	A single-particle operator
$\boldsymbol{h}$	Unit cell matrix
$\boldsymbol{h}_0$	Unit cell shape matrix
$\hat{\mathcal{J}}$	Coulomb two-particle operator
$\hat{\mathcal{K}}$	Exchange two-particle operator
k_B	Boltzmann constant
M_a	Mass of the a^{th} nucleus
N_{g}	Number of guest molecules
N_{host}	Number of adsorbent atoms
$n(\boldsymbol{r})$	Electron density
$\hbar$	Reduced Planck constant
$\hat{\mathcal{H}}$	Hamiltonian operator
P	Pressure

$\boldsymbol{p}_i$	Momentum vector of the i^{th} electron
q_A^c	Atomic core charge on atom A
Q	Partition function in the canonical ensemble
Q_{st}	Isosteric heat of adsorption
$\boldsymbol{r}_i$	Position vector of the i^{th} electron
r_{iA}	Interatomic distance between the i^{th} electron (pair) and A^{th} nucleus
r_{ij}	Interatomic distance between electron pair i and j
R_{AB}	Interatomic distance between nucleus A and B
S	Entropy
t	Time
T	Kinetic interaction (in Section 2.1.2)
T_s	Kinetic interaction of a system of non-interaction electrons (in Section 2.1.2)
T	Temperature
T_{inst}	Instantaneous temperature
$\hat{\mathfrak{T}}_e$	Kinetic energy operator of the electrons
$\hat{\mathfrak{T}}_N$	Kinetic energy operator of the nuclei
V	Volume
V_{ext}	External potential
V_{ee}	Electron-electron interaction
$\mathcal{V}_{ee}$	Coulomb repulsion operator between electrons
$\mathcal{V}_{Ne}$	Coulomb attraction operator between electrons and nuclei
$\mathcal{V}_{NN}$	Coulomb repulsion operator between nuclei
V_H	Coulomb interaction or Hartree term
$\boldsymbol{x}_i$	Generalized coordinate including the positions and spin of the i^{th} electron
Z_A	Atomic number of the A^{th} nucleus
Z_{os}	Osmotic partition function

Greek symbols	
α_i	Isotropic atomic polarizability of atom i
$\boldsymbol{\alpha}_i$	Atomic polarizability tensor of atom i
Δ	Isobaric-isothermal partition function
ϵ_{ij}	van der Waals interaction parameter between atoms i and j
ζ	Grand canonical partition function
μ	Chemical potential
μ_i	Spin of the i^{th} electron
$\boldsymbol{\mu}_i$	Induced atomic dipole on atom i
ρ	Density
σ_A	Width of a Slater function on atom A
σ_{ij}	van der Waals distance parameter between atoms i and j
$\boldsymbol{\sigma}_i$	Instantaneous stress tensor
$\boldsymbol{\sigma}_{i,a}$	Instantaneous anisotropic stress tensor
$\lvert\chi_{nm}\rangle$	The nm^{th} nuclear wavefunction
$\lvert\phi_n\rangle$	The n^{th} general wavefunction
$\lvert\phi_n^e\rangle$	The n^{th} electronic wavefunction
$\lvert\psi\rangle$	Eigenfunction (general)
$\lvert\Psi(t)\rangle$	Time-dependent eigenfunction (general)
Ω	Grand-canonical potential
Ω_{os}	Osmotic potential

Mathematical symbols	
$\dot{a}$	Time derivative of a
∇a	Gradient of a
$a_1 \otimes a_2$	Outer product of vectors a_1 and a_2
A^{-1}	Matrix inverse of A
$\det(A)$	Determinant of the square matrix A
$\text{Tr}(A)$	Trace of the square matrix A

List of Abbreviations

AIM	Atoms in molecules
ACSF	Atom-centered symmetry functions
ASE	Atomic Simulation Environment
BBCDC	9,9'-([1,1'-biphenyl]-4,4'-diyl)bis(9H-carbazole-3,6-dicarboxylate
BDC	1,4-benzodicarboxylate
BHE-bpb	2,5-bis(2-hydroxyethoxy)-1,4-bis(4-pyridyl)benzene)
BO	Born-Oppenheimer
BTC	1,3,5-benzenetricarboxylate
CCSD(T)	Coupled-cluster theory with a full treatment of singles and doubles but perturbative triples
CFC	Chlorofluorocarbon
CN	Coordination number
cp	Closed pore
DFT	Density function theory
DMA	Distributed multipole analysis
DOBDC	2,5-dioxido-1,4-benzenedicarboxylate
DUT	Dresden University of Technology
FF	Force field
GAFF	General amber force field
GCMC	Grand-canonical Monte Carlo
GGA	Generalized-gradient approximation
HF	Hartree-Fock
HKUST	Hong Kong University of Science and Technology
HPC	High-performance computing
IRMOF	Isoreticular MOF
KIE	Kinetic isotope effect
LDA	Local-density approximation
lp	Large pore
MAE	Mean absolute error
MBIS	Minimal basis iterative stockholder
MC	Monte Carlo

MD	Molecular dynamics
ME	Mean error
MEDFF	Monomer Electron Density Force Field
MIL	Matériaux de L'Institut Lavoisier
MLP	Machine learning potential
MOF	Metal-organic framework
MP2	Møller–Plesset perturbation to the second order
MSE	Mean squared error
MTTK	Martyna–Tobias–Tuckerman–Klein
NequIP	Neural equivariant interatomic potential
NGA	Negative gas adsorption
NHC	Nosé-Hoover chain
np	Narrow pore
NQEs	Nuclear quantum effects
op	Open pore
PBE	Perdew-Ernzerhof-Burke
PCP	Porous coordination polymer
PES	Potential energy surface
PGA	Positive gas adsorption
PIMD	Path integral molecular dynamics
PZDC	2,3-pyrazinedicarboxylate
RDF	Radial distribution function
RMSPE	Root mean square percentage error
SBU	Secondary building unit
SPC	Soft porous crystal
SAPT	Symmetry-adapted perturbation theory
SCF	Self-consistent field
UiO	Universitetet i Oslo
UFF	Universal force field
XDM	Exchange-hole dipole moment
ZIF	Zeolitic imidazolate framework

Part I

Development of Accurate and Reliable Methods for *In Silico* Modeling of Adsorption in Nanoporous Materials

1

Introduction

What we know is a drop,
what we don't know is an ocean.
Isaac Newton (1643–1727)

Human progress throughout history can be defined by the use of tools. As far back as 2.5 million years, ancient humans used stone chopping tools to break open bones of deer to extract the nutritious bone marrow. [1] Primitive stone tools developed slowly into hand axes and blades, and by the Upper Paleolithic (50 000 to 12 000 years ago), fishing harpoons, nets and the bow and arrow had been invented. [2–4]

Especially in the last few centuries, spurred by the Enlightenment and the Industrial Revolution, progress has been remarkable. Our world has been transformed by inventions which were enabled by an ever-deepening understanding of materials and the fundamental forces through which they interact. Most of the seminal inventions of the last 150 years have arisen from a deeper understanding of different branches of physics. Electric refrigeration and the internal combustion engine arose from an understanding of electromagnetism and thermodynamics, the use of nuclear fission was enabled by new insights in subatomic physics and the transistor could only be invented through the development of solid-state physics.

Of course, this is not the end of our story. The search for materials tailored for specific applications has increased tremendously. On the semiconductor

front, the ever-increasing demand for faster, more power-efficient computation is driving the size of transistors to atomic length scales, and spurring the development of quantum computing. [5] On another front, nuclear fusion is an exciting technology under development for next-generation energy production. For this application, exotic new alloys are required for the reactor wall to withstand the high-energy neutron radiation, as well as the harsh mechanical, heat and electromagnetic environment. [6]

Although the pace of scientific innovation is unrelenting, not all developments have been for the betterment of humanity. Of the previously mentioned applications, the use of chlorofluorocarbons (CFCs) as a safer alternative to previously used refrigerants lead to a depletion of the ozone layer [7, 8] and the worldwide use of fossil fuels lead to climate change. [9] Especially the latter has been identified as one of the paramount challenges for the 21st century. To tackle this challenge, a fundamental paradigm shift away from carbon-intensive industries will be required. While this task appears daunting, several promising technologies to this end have already been identified. On the supply side, a transition to hydrogen as the carrier for a net-zero energy system holds promise to decarbonize parts of the transportation, industrial and heating sectors. [10] On the other hand, carbon capture and storage (CCS) is an already established technology in the oil refining and natural gas sectors. [10] However, commercially viable techniques such as amine-based chemical adsorption usually have the disadvantage of involving harmful toxic volatiles and are associated with a high energy penalty of the regeneration process.[11, 12] Clearly, further process and material innovations are required for carbon capture technologies to meet the proposed cost and performance goals which have been stated. [13–15] In this regard, physical adsorption in nanoporous materials appears as a prime candidate.

Before stating the goal of this dissertation in Section 1.4, this class of materials is introduced in Section 1.1, focusing specifically on metal-organic frameworks (MOFs). In Section 1.2, their extraordinary gas adsorption and separation properties are introduced. Finally, we arrive at the need for computational tools to model the adsorption and separations processes at operating conditions, described in Section 1.3, which will lead to the goals of this work.

1.1 Nanoporous materials

To understand the extent of scientific and industrial interest in nanoporous materials, it's essential to make their primary use case intuitively clear via analogy. For this, let us imagine a kitchen sponge. Whether natural or

synthetic, sponges contain pores ideally suited for absorbing liquids like water. Just as plants pull up water from their roots through capillary action, the pores of sponges are small enough for water to accumulate in. Analogously (but not exactly through the same process), what if we could employ materials with much smaller pores, at the length scales of individual molecules. These materials could store and separate gases, even act as catalysts for chemical reactions. Depending on the exact structure of the pores and building blocks from which these materials are constructed, a wealth of applications can be envisaged. This is precisely the role that nanoporous materials can fulfill.

Nanoporous materials are a class of materials characterized by pores with a width of 1 – 100 nanometer. Already in the 18th century, a naturally occurring nanoporous material was discovered by mineralogist Axel F. Cronstedt who observed a mineral releasing water upon heating, with the ability for reuptake at lower temperatures. [16] Combining the Greek words for "to boil" and "stone", he coined the term "zeolite" for this mineral. Zeolites are crystalline three-dimensional aluminosilicates formed by linked TO_4 (T = Si, Al, P, ...) tetrahedra combined with charge-compensating cations. [17] Despite the remarkable molecular sieve properties of zeolites being recognized in the 19th century, the lack systematic synthesis procedures and experimental identification methods such as X-ray diffraction prevented their industrial use. [17] Only after pioneering work by R. M. Barrer and the discovery of commercially significant zeolites A, X and Y by R. Milton, the path towards industrial use of zeolites was laid out. [18, 19] Nowadays, zeolites are routinely employed in organic synthesis and for catalytic cracking in the petrochemical industry. [20, 21] Even so, the tunability of zeolites for specific applications is limited, as the number of distinct zeolite structures is relatively small, and chemical functionalization has remained challenging. [22, 23]A new type of nanoporous material, coined metal-organic frameworks, therefore gained prominence at an opportune time.

1.2 Metal-organic frameworks for adsorption and separation

Metal-organic frameworks (MOFs), also referred to as porous coordination polymers (PCPs), are crystalline materials composed of metal-oxide secondary building units (SBUs) linked together by organic linkers. Although first reports on this class of materials date back to the 1960s,[30–32] interest in them accelerated greatly from the late 1990s when it became clear that highly crystalline and porous MOFs could be synthesized that did not collapse

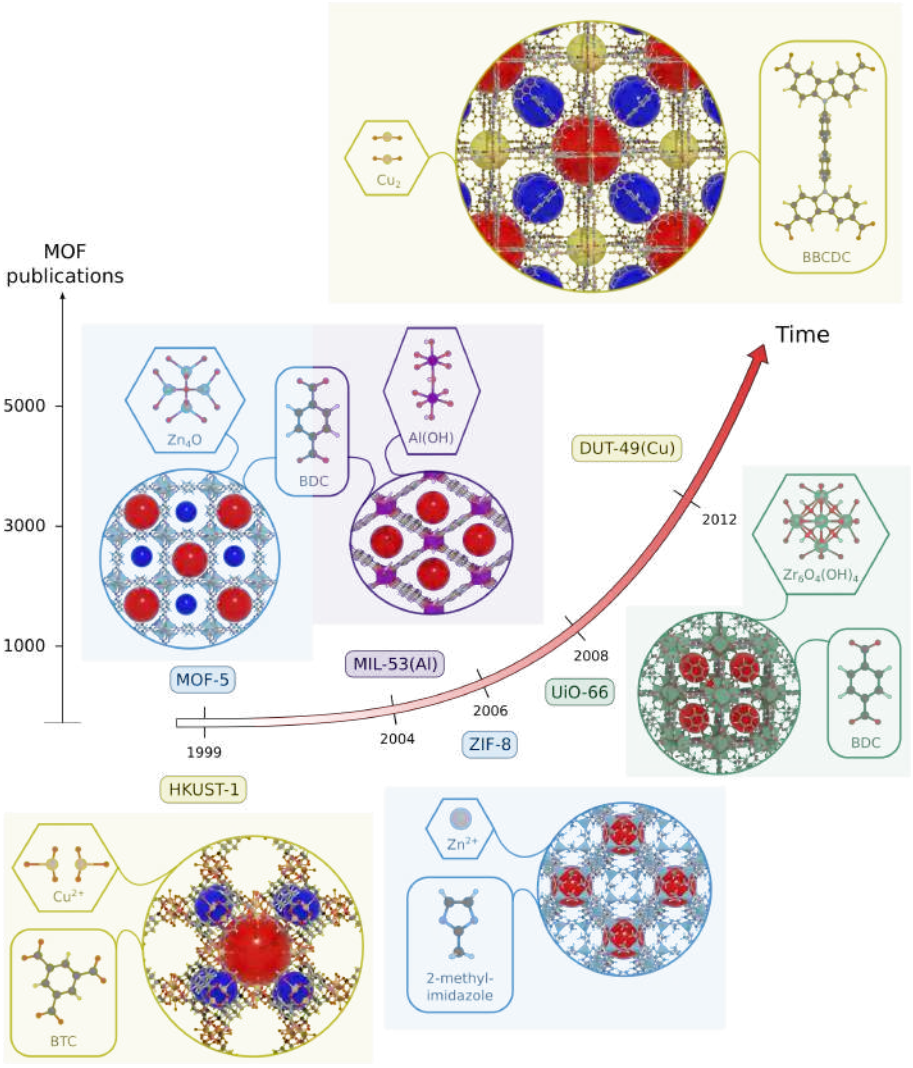


Figure 1.1: An overview of the first syntheses of six prototypical MOFs (MOF-5[24], HKUST-1[25], MIL-53(Al)[26], ZIF-8[27], UiO-66[28] and DUT-49(Cu)[29]) on a time line of the number MOF publications through time. For each MOF, the constituting metal brick and organic linkers are shown. Abbreviations of the linkers refer to 1,4-benzodicarboxylate for BDC, 1,3,5-benzenetricarboxylate for BTC and 9,9'-([1,1'-biphenyl]-4,4'-diyl)bis(9H-carbazole-3,6-dicarboxylate for BBCDC.

during the removal of solvent molecules. [24, 25, 33–35] Although a wealth of new techniques have been developed,[36–39] conventional synthesis of MOFs is typically achieved through solvothermal reactions, mixing organic ligands and metal salts in a polar solvent such as water, acetonitrile, dimethylformamide, dimethylsolfoxide or methanol. [40–42] However, the crystallinity and porosity of the synthesized MOFs depends critically on the temperature, solvent compositions, reagent ratios, concentrations and reaction times, necessitating a degree of trial-and-error. [40, 43–45] During synthesis, the solvent molecules occupy the pores and are subsequently removed by heating or application of a vacuum. The first MOFs maintaining porosity during this last step were MOF-5 and HKUST-1. [25, 35] MOF-5 is a highly porous MOF composed of Zn_4O clusters tetrahedrally connected by 1,4-benzenedicarboxylate (BDC) linkers, resulting in a cubic network with an internal surface area of 3800 m^2/g, significantly higher than that of common zeolites or activated carbon. [46] Since then, many MOFs of interest have been identified. In 2002, the first of a highly flexible class of permanently porous MOFs was synthesized, coined MIL-53 (MIL = Matériaux de l'Institut Lavoisier[47, 48]), consisting of $[M(OH)]_\infty$ chains connected with BDC ligands in a wine-rack topology. Of these, MIL-53(Al) was observed to reversibly transition between a large pore (lp) and narrow pore (np) phase upon hydration/dehydration.[26] This phenomenon was coined the *breathing* of the framework, and is characterized by a volume change of up to 40%. In 2006, zeolitic imidazolate frameworks (ZIFs) were first synthesized.[27] ZIFs are a class of MOFs consisting of zinc or cobalt metals connected tetrahedrally by imidazolate linkers. They share the same topologies as zeolites while having the advantage of being more flexible in surface modifications targeting applications like carbon capture. [27, 49–52] In Figure 1.1, six prototypical MOFs are shown on a time line of their first synthesis. Through the principles of isoreticular synthesis, many more MOFs have since been synthesized with specific applications in mind.

1.2.1 Isoreticular synthesis and rational design

As a given metal-organic framework can be characterized by its topology and constituting SBUs (the metals and organic linkers), one could imagine substituting one SBU with another one with the same coordination geometry, maintaining the same topology. This concept, referred to as isoreticular synthesis, paves the way for the rational design of MOFs, endowing them with the desired characteristics. One of the prototypical MOFs, MOF-5, is a prime example of this concept in action. In seminal work by the group of Yaghi, the pore size of MOF-5 could be incrementally varied from 3.8 to 28.8 Å by substitution of the linker with biphenyl, tetrahydropyrene, pyrene, and

terphenyl, giving rise to the IRMOF isoreticular series of MOFs.[53] Among this series, IRMOF-6 demonstrated state of the art methane storage capacity of 240 cm^3/g at standard temperature and pressure. On the other hand, MOF-5 as first synthesized in 1999, is unstable in the presence of water. However, later work showed that the addition of methyl or trifluoromethoxy groups to the BDC linker significantly enhances its stability. Due to these hydrophobic moieties, these modified MOF-5 frameworks retained their porosity for days when exposed to ambient air. [54–56]

Apart from tuning the porosity for increased gas storage capacity, the size of the pore window between neighboring pores can also be tuned to separate gases kinetically. For example, ZIF-8 and ZIF-67 have been identified for their large propane/propene selectivity due to the small differences in kinetic diameters of these gases. The pore window of both MOFs is ideally tuned for allowing propene to pass through while hindering propane. [57] The ability of MOFs to kinetically separate gases has also been identified for N_2/O_2, CO_2/CH_4 and CO_2/N_2 separation. [58–60] A second, more common way to separate gases is by means of adsorptive selectivity, also called thermodynamic separation. [43] In physisorptive adsorption, separation is achieved based on the distinct physical properties of gases. For example, the significant quadrupole moment of CO_2 ($13.4 \cdot 10^{-40}$ $C \cdot m^2$), compared to the lack of one for CH_4, results in a preferential interaction of CO_2 with the MIL-53 and MIL-47 frameworks. [61, 62] The addition of polar groups can further enhance selectivity. In UiO-66(Zr), the CO_2/CH_4 selectivity can be significantly increased by ligand functionalization with $-SO_3H$ and $-CO_2H$ groups.[63] The adsorption characteristics of a MOF can also be tuned by changing the metal site. In the widely studied M_2(dobdc) (dobdc = 2,5-dioxido-1,4-benzenedicarboxylate) framework, changing the open-metal site from Mg to Ni, Co, Fe, Mn, Zn or Cu systematically decreases the heat of adsorption of CO_2 from 43.5 $kJ \cdot mol^{-1}$ to between 38.6 $kJ \cdot mol^{-1}$ and 22.1 $kJ \cdot mol^{-1}$.[64] Many more MOFs have been modified with specific applications in mind, more than can be summarized here. For a thorough overview on selective gas adsorption in MOFs, the interested reader is referred to a review of Zhou *et al.* [65] However, a unique property of a subclass of MOFs endows them with a different pathway of tuning adsorption, that of flexibility.

1.2.2 Guest-induced flexibility

As outlined before, interest in metal-organic frameworks grew quickly after structures were identified that maintained their crystallinity and porosity after the removal of guest molecules. Kitagawa and Kondo coined this a transition from first generation to second generation compounds.[66, 67] Where

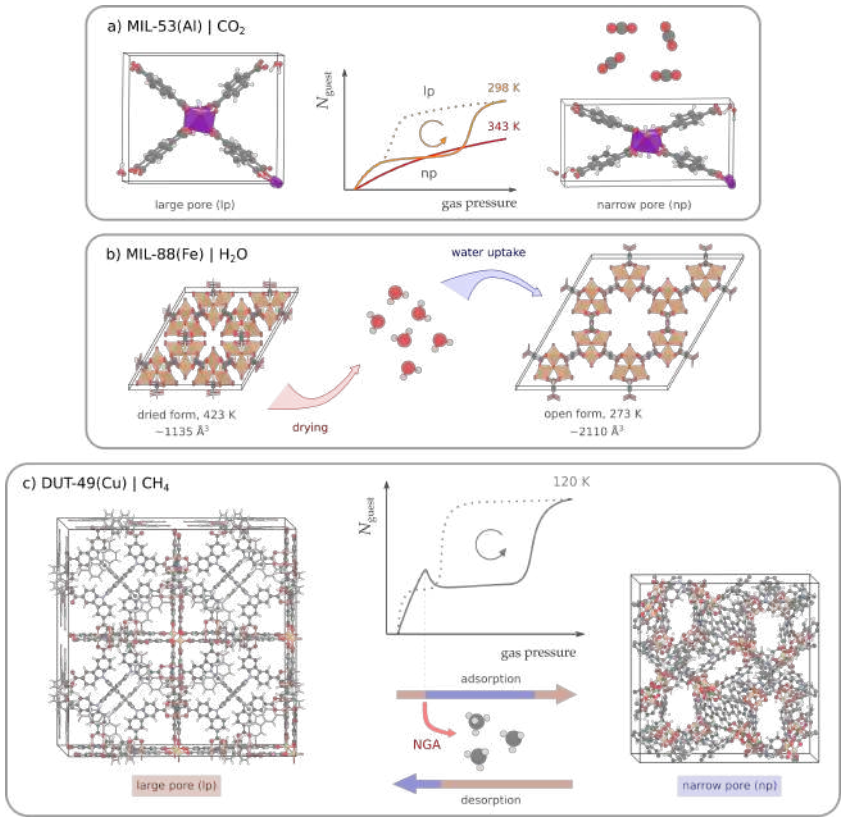


Figure 1.2: Examples of guest-induced flexibility in MOFs. (a) The breathing transition between the lp and np phases of MIL-53(Al) under carbondioxide adsorption, (b) the gradual swelling of MIL-88(Fe) under uptake of water and (c) the lp to np transition of DUT-49(Cu) associated with negative gas adsorption (NGA).

the first generation of materials collapse after the removal of guests, the second generation of MOFs maintain their porous structure before and after guest adsorption, paving the way to their use as functional adsorbents. Interestingly, they hypothesized a third generation of frameworks as well. These MOFs could dynamically respond to external stimuli such as temperature, pressure, light or guest adsorption. Such a class of materials was later identified, and termed soft porous crystals (SPCs). [68] They are characterized by their highly ordered network and a reversible transformability between two or more phases, of which at least one possesses space for guest molecules. [68] A prime example of this reversible transformation is the *breathing* transition in MIL-53(Al) shown in Figure 1.2(a). The empty framework can undergo a phase transformation between an orthorhombic

large-pore (lp) and monoclinic closed-pore (cp) phase under the influence of temperature or pressure.[69, 70] Futhermore, under methane, carbondioxide or xenon adsorption, an intermediate narrow-pore (np) phase can be reversibly reached, resulting in a distinct hysteresis between adsorption and desorption.[71, 72] In **Paper I**, computational schemes are developed to rationalize the breathing transition or lack thereof under guest adsorption of different molecules as a function of temperature and gas pressure. Instead of a phase transition between distinct phases, guest adsorption in MIL-88 is associated with a gradual increase in unit cell volume with increasing guest loading, up to 85% under water adsorption.[73] Such a transition is referred to as a *swelling* mode. Both the dried and hydrated forms of the framework are shown in Figure 1.2(b). Additionally, phase transitions which are not associated with a change in unit cell volume are possible. One such example is the phenomenon of gate-opening. The framework $[Cd_2(pzdc)_2(BHE\text{-}bpb)]_n$ (pzdc = 2,3-pyrazinedicarboxylate; BHE-bpb = 2,5-bis(2-hydroxyethoxy)-1,4-bis(4-pyridyl)benzene) consists of alkyloxy functionalized pyridyl linkers containing OH groups which can interact with neighboring linkers. These groups form a gate of the pore which can rotate and open the pore space when polar guest molecules interact with the Cd metal center, leading to a highly selective uptake.[74] Common to all these previous transitions is the monotonic increase in gas uptake with increasing gas pressures. However, in 2016, the counterintuitive phenomenon of negative gas adsorption (NGA) was discovered in DUT-49(Cu). [75] In this framework, a cubic-to-cubic phase transition is triggered by the presence of methane adsorbates (see Figure 1.2(c)), resulting in the sudden release of guests. In order to rationalize this phenomenon, research has focused on the linker dynamics of the framework,[76] the thermodynamic driving forces of the transition,[77] the dependence on temperature and adsorbate,[78] and the organic linker of the framework.[79] However, it was shown that an NGA transition must imply the involvement of a metastable state.[75] This necessitates a method to unequivocally determine the (meta)stability of phases as a function of temperature and gas pressure, which is the topic of **Paper II**.

Clearly, the inherent flexibility of this class of materials opens up another pathway through which rational design can be achieved, as the application of external stimuli can tune the pore structure, resulting in the highly selective capture of guests. Several studies have already leveraged the inherent flexibility of SPCs for CO_2/N_2, CO_2/CH_4 or C_2H_6/CH_4 separation. [80–83] Furthermore, possible applications for sensing by reversible color change upon a guest-induced phase transition or controlled drug release to a target organ have been identified. [84–87] Although many more applications for flexible MOFs can be imagined, significant roadblocks slowing down research

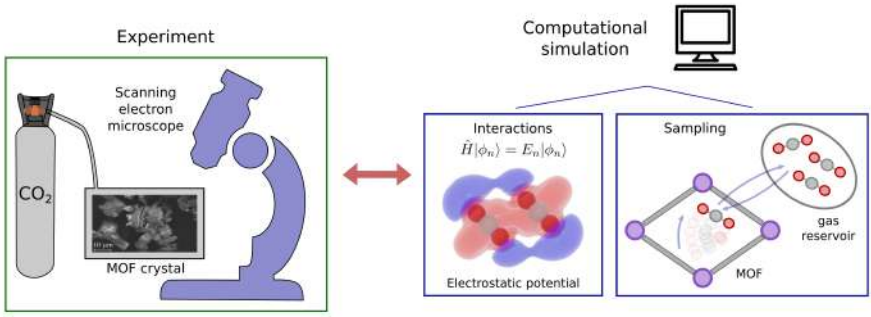


Figure 1.3: Experimental versus computational methods for studying MOFs and their adsorption properties. In experiments, a scanning electron microscope (SEM) can be used to visualize the single-crystal grain sizes. Connecting the MOF crystal with a gas reservoir, the adsorption isotherms of a MOF can be extracted. On the other hand, a computational simulation requires a representative system for the MOF, combined with both a method for calculating interactions as well as a sampling protocol which connects the MOF with the external gas reservoir. Figure for the experimental MIL-53(Al) crystal reproduced from Ref. [88] with permission from Elsevier.

exist. First of all, experimental synthesis and characterization of MOFs is costly both time-wise and money-wise, preventing a full exploration of the adsorption properties of the vast space of all MOFs. Secondly, to truly fulfill the promise held by rational design of materials, a thorough understanding of the kinetics, dynamics and thermodynamics of materials and adsorbed guests at the nanoscale is required. This is where the world of computational modeling is poised to take a leading role.

1.3 Computational modeling

The application of computational physics to materials can be seen as a branch of physics which neatly connects the long existing experimental and theoretical branches. However, it has firmly established itself as a full-fledged third branch of physics. Whereas experiments probe materials to understand their properties, theory has its roots in the fundamental quantum mechanical equations which can be employed to predict how materials interact. Both of these have their limitations, however. In experiments, perfect control of the material under study and the experimental conditions is difficult. On

the other hand, the quantum mechanical equations only reduce to a closed-form solution for the most simple cases. In contrast to this, computational physics as applied to molecules and materials aims at a numerical solution to (approximations of) the fundamental (or *ab initio*) quantum mechanical equations. With these approximate solutions, simulations are performed to replicate *in silico* the experimental conditions to which the system is exposed (by use of a sampling procedure). From these simulations, information is then gained about the properties of interest for the system under study. While this sounds promising, the success or failure of this approach relies critically both on the approximations made to the theory, as well as the way that the experimental conditions are replicated. Computational scientists therefore always have to keep both of these two ingredients (the approximated interactions and the sampling, see Figure 1.3) firmly in the back of their mind. While the theoretical foundations behind these two ingredients will be laid out in Chapter 2, it is informative to already give a short overview of how computational modeling has been applied to study the adsorption of guests in MOFs, which will lead to the goals of this dissertation.

To describe the interactions between atoms, non-polarizable *force fields* are most commonly used. Herein, the non-covalent interaction between adsorbates and a MOF are described by atom-centered empirical potentials, assigning a charge and van der Waals parameters to each atom. For the covalent interactions, separate potentials are used to describe bonds, bends, dihedrals and torsion (sometimes supplemented with more advanced terms). Many force fields following this recipe have been developed, combined with sets of parameters fitted to experimental data or *ab initio* calculations. Force fields can generally be distinguished based on their range of applicability. While some were developed with broad applicability to molecules, materials and liquids in mind (UFF, DREIDING, CHARMM), others were developed to describe individual or sets of MOFs (UFF4MOF, MOF-FF, BTW-FF, DWES, QuickFF).[89–97] Although these approaches have certainly been successful, many cases have been identified where the accuracy of the functional form used to describe the non-covalent interactions breaks down. In M-MOF-74, the interaction of CO_2 and small hydrocarbons with the open-metal sites of the framework cannot be accurately captured with these types of force fields, necessitating the need for more advanced polarizable force fields, although this has the disadvantage of introducing extra parameters which require careful fitting to agree with *ab initio* calculations. [98, 99] Also in other MOFs, the importance of electronic polarization has been clearly established. [100, 101] To improve the accuracy with which these effects are captured, a transferable polarization model is derived in **Paper III** and a method to train a machine leaning model on non-covalent interaction energies of adsorbates in MOFs

is developed in **Paper IV**.

Using an appropriate force field, adsorption simulations can subsequently be performed which replicate the experimental conditions in which a reservoir of gas molecules at a well-defined temperature T and pressure p is in contact with the system under study, allowing for an exchange in guests. This is usually achieved by means of grand-canonical Monte Carlo (GCMC) simulations. Herein, adsorbate molecules are inserted/deleted and translated/rotated in the framework as to sample the statistical ensemble associated with the experimental conditions (p, T). GCMC simulations have a long history in adsorption research, and have been employed in screening studies determining the adsorption properties of databases of MOFs. Early studies focused on the screening of MOFs for their uptake and heat of adsorption of hydrogen, carbon dioxide and methane. [102–104] Later studies identified MOFs suited for the selective capture of CO_2 over N_2, CO_2 over small hydrocarbons, ethylene from hydrocarbon mixtures and the performance of a set of 14 functional groups for the capture of toxic industrial chemicals. [105–108] In seminal works, Gómez–Gualdrón *et al.* identified NU-1103 out of a set of 13 512 MOFs with an outstanding hydrogen storage capacity at cryogenic temperatures, and Wilmer *et al.* identified 300 out of 137 953 MOFs for their state-of-the-art methane storage capacity.[109, 110] However, GCMC simulations do not take the flexibility of the framework into account, neglecting any flexibility-induced enhanced uptake and selectivity. In **Paper I** and **Paper II**, computational schemes are proposed and benchmarked to include this framework flexibility.

1.4 Goal and outline

As outlined in previous Sections, computationally modeling adsorption in nanoporous materials essentially relies on the quality of the used force field, as well as the sampling protocol with which adsorption simulations are performed. In this dissertation, contributions are made to improve the accuracy of these two necessary ingredients, as a step towards the goal of an accurate and reliable description of adsorption in MOFs and other nanoporous materials. To accomplish this goal, several intermediate goals are discussed in the following chapters.

Chapter 2 In Chapter 2, the theoretical foundations are laid out on which molecular modeling is based. Starting from the Schrödinger equation, the concept of the potential energy surface (PES) is introduced. Furthermore, methods to sample the PES will be

discussed. From these simulations, the key properties of interest of the system under study such as adsorption isotherms, heats of adsorption and the phase stability under adsorption can be extracted.

Chapter 3 In Chapter 3, several computational sampling schemes to include the framework flexibility in adsorption calculations are introduced. These schemes are applied to the adsorption of neon, argon, xenon, methane and carbon dioxide in the breathing MIL-53(Al). Each of these schemes yields complementary information, and capture the dynamic and thermodynamic features of the phase stability of MIL-53(Al) under gas adsorption. Subsequently, one of the schemes is applied to methane adsorption in DUT-49(Cu), capturing the negative gas adsorption (NGA) transition and the experimentally observed (meta)stability of both phases of the framework as a function of gas pressure and temperature, demonstrating its ability in describing guest-induced phase transitions.

Chapter 4 In Chapter 4, two different approaches for improving the description of framework–adsorbate interactions are developed. In the first, a transferable polarizable force field is developed that can accurately describe polarization effects with a single transferable parameter. In a second approach, relying on high-quality reference calculations, the framework–adsorbate interactions are learned with a machine learning potential (MLP) and employed in GCMC simulations to accurately calculate the carbon dioxide uptake and heat of adsorption in Mg-MOF-74, for which generic force fields break down. Lastly, the utility of employing MLPs to describe reactive events is investigated, providing great promise in the computational characterization of chemisorption and catalytic reactions.

Chapter 5 In Chapter 5, concluding remarks are made together with a personal perspective of possible further developments in methods for accurate adsorption calculations in nanoporous materials.

2

Computational methods

I insist upon the view that 'all is waves
Erwin Schrödinger (1887–1961)

To describe the interactions governing the behavior of atoms, molecules and materials at the nanoscale, we start from first principles (*ab initio*), provided by the Schrödinger equation.[111] This formulation is grounded in the quantum mechanical postulate that a system is fully characterized by its wavefunction. For a molecular system consisting of atomic nuclei and electrons, the wavefunction depends on the positions of the nuclei and the positions and spins of the electrons. Herein, it is assumed that nuclei (consisting of protons and neutrons) and electrons only interact through the electromagnetic interaction, neglecting gravitational and nuclear forces. Additionally, the nuclei are treated as point charges, and the relativistic motion of nuclei and electrons is neglected. These last two approximations are not generally correct, but can still approximately be taken into account *a posteriori*. From this starting point, several further approximations will be introduced, concluding in a force field description of the potential energy surface. Subsequently, insights from classical mechanics and statistical mechanics will lead to sampling techniques from which all properties of interest of the system can be extracted.

2.1 From first principles to molecular mechanics

2.1.1 The Schrödinger equation and Born-Oppenheimer approximation

In 1926, Schrödinger postulated that the dynamics of the wavefunction $|\Psi(t)\rangle$ is governed by the following equation

$$\mathrm{i}\hbar\frac{\partial}{\partial t}|\Psi(t)\rangle = \hat{\mathcal{H}}|\Psi(t)\rangle \tag{2.1}$$

with $\hat{\mathcal{H}}$ the Hamiltonian and $\hbar$ the reduced Planck constant. For a molecular system consisting of nuclei and electrons, the molecular Hamiltonian consists of the following kinetic and potential terms for the nuclei N and electrons e

$$\hat{\mathcal{H}} = \hat{\mathcal{T}}_N + \hat{\mathcal{T}}_e + \hat{\mathcal{V}}_{NN} + \hat{\mathcal{V}}_{Ne} + \hat{\mathcal{V}}_{ee} \tag{2.2}$$

where each operator is defined as follows

$$\hat{\mathcal{T}}_N = -\frac{1}{2}\sum_A \frac{\nabla_A^2}{M_A} \qquad \hat{\mathcal{V}}_{NN} = \sum_{A>B}\frac{Z_A Z_B}{R_{AB}}$$

$$\hat{\mathcal{T}}_e = -\frac{1}{2}\sum_i \nabla_i^2 \qquad \hat{\mathcal{V}}_{Ne} = \sum_A\sum_i \frac{Z_A}{r_{iA}}$$

$$\hat{\mathcal{V}}_{ee} = \sum_{i>j}\frac{1}{r_{ij}}$$

Herein, atomic units are employed, with M_A and Z_A representing the mass and charge of nucleus A and r_{AB}, r_{iA} and r_{ij} the nuclei–nuclei, electron–nuclei and electron–electron distances, respectively. As this Hamiltonian is time-independent, a separation of variables can be applied, reducing the time-dependent Schrödinger equation (Equation (2.1)) into a time-independent one

$$\hat{\mathcal{H}}|\phi_n\rangle = E_n|\phi_n\rangle \tag{2.3}$$

with E_n and $|\phi_n\rangle$ the eigenvalues and eigenstates, from which the solution to Equation (2.1) can be written as

$$|\Psi(t)\rangle = \sum_n c_n \exp(-\mathrm{i}E_n t)|\phi_n\rangle \tag{2.4}$$

In this Hamiltonian, a coupling between the nuclear and electronic motion exists through r_{iA}. To decouple the nuclear and electronic degrees of freedom, the Born-Oppenheimer (BO) approximation is made.[112] First,

note that the mass of electrons and nuclei differs by at least three orders of magnitude. The electrons move much faster, and can therefore equilibrate on shorter timescales than the nuclei. Neglecting nuclear motion, the following electronic eigenvalue equation is obtained

$$\left(\hat{\mathcal{T}}_e + \hat{\mathcal{V}}_{NN} + \hat{\mathcal{V}}_{Ne} + \hat{\mathcal{V}}_{ee}\right) |\phi_n^e\rangle = E_n^e \, |\phi_n^e\rangle \tag{2.5}$$

with $|\phi_n^e\rangle$ and E_n^e the electronic eigenvectors and eigenvalues, respectively. In this Equation, the nuclear coordinates only appear parametrically. In other words, the obtained eigenvalues $E_n^e = E_n^e(\{\boldsymbol{R}\})$ (with E_0^e the ground state) are high-dimensional surfaces as a function of the nuclear coordinates, often referred to as Born-Oppenheimer surfaces or potential energy surfaces (PES). Finally, by assuming that all excited electronic states are much higher in energy than the ground state, called the adiabatic approximation, the nuclear eigenfunctions $|\chi_{nm}\rangle$ are obtained from the following equation

$$\left(\hat{\mathcal{T}}_\mathrm{n} + E_n^e\right) |\chi_{nm}\rangle = E_{nm} \, |\chi_{nm}\rangle \tag{2.6}$$

In most common molecules and materials, the electronic energy levels are indeed separated by more than an electronvolt, larger then the separation between nuclear rotational and vibrational energy levels. The Born-Oppenheimer approximation is therefore used in the rest of this work, only occupying ourself with the calculation of the ground state energy $E_0^e(\{\boldsymbol{R}\})$. In the following sections, the ground state electronic wavefunction will be referred to as $|\psi\rangle$ for ease of notation. Finally, due to their large mass, it is possible to treat the nuclei as classical particles moving on the PES, avoiding the need to solve the nuclear eigenvalue Equation (2.6) explicitly. Only for the lightest nuclei like hydrogen does this approximation break down, although methods such as path integral molecular dynamics exist to take the quantum behavior of nuclei into account.[113]

2.1.2 Quantum mechanical methods

Solving Equation (2.5) for the molecules or materials of interest will now yield all the required insight into their constituting interactions and electronic structure. However, this is exactly where the crux of molecular modeling is situated. Equation (2.5) is a many-body problem, and even in one of the most simple cases of an isolated helium atom (consisting of one nuclei and two electrons) no analytical solution exists, analogous to the lack of a closed-form solution for the three-body problem in classical mechanics.[114] However, a wealth of methods for approximate solutions have been developed, many of which build on the Hartree-Fock formalism.[115–117]

Hartree-Fock

Given a system containing N electrons, it is tempting to write the electronic wavefunction as a product of one-electron wavefunctions

$$\psi\left(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N\right) = \phi_1\left(\boldsymbol{x}_1\right)\phi_2\left(\boldsymbol{x}_2\right)\cdots\phi_N\left(\boldsymbol{x}_N\right) \tag{2.7}$$

with $\boldsymbol{x}_i = (\boldsymbol{r}_i, \sigma_i)$ a generalized coordinate including both the position $\boldsymbol{r}_i$ and spin σ_i. However, this wavefunction does not satisfy the antisymmetric constraint of fermions under particle exchange. The Hartree-Fock (HF) method postulates a Slater determinant constructed from the one-electron orbitals instead which satisfies this constraint[118]

$$\psi\left(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N\right) = \frac{1}{\sqrt{N!}}\begin{vmatrix} \phi_1(\boldsymbol{x}_1) & \phi_1(\boldsymbol{x}_2) & \cdots & \phi_1(\boldsymbol{x}_N) \\ \phi_2(\boldsymbol{x}_1) & \phi_2(\boldsymbol{x}_2) & \cdots & \phi_2(\boldsymbol{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_N(\boldsymbol{x}_1) & \phi_N(\boldsymbol{x}_2) & \cdots & \phi_N(\boldsymbol{x}_N) \end{vmatrix} \tag{2.8}$$

Making use of the variational principle, together with the orthonormality constraint of the orbitals, the ground state energy follows from a minimization of the expectation value of $\hat{\mathcal{H}}$ with respect to the one-electron orbitals

$$E = \min_{\phi_i(\boldsymbol{x}_i)} \langle\psi|\hat{\mathcal{H}}|\psi\rangle \tag{2.9}$$

resulting in a set of equations for the one-electron orbitals

$$\hat{\mathcal{F}}\,\phi_i = \left(\hat{h} + \sum_j \left[\hat{\mathcal{J}}_j - \hat{\mathcal{K}}_j\right]\right)\phi_i = \epsilon_i\phi_i \tag{2.10}$$

with $\hat{\mathcal{F}}$ the Fock operator, consisting of a single-particle operator $\hat{h}$ and the Coulomb $\hat{\mathcal{J}}$ and exchange $\hat{\mathcal{K}}$ two-particle operators defined as follows

$$\begin{aligned} \hat{h} &= -\frac{1}{2}\nabla^2 - \sum_A \frac{Z_A}{|\boldsymbol{r} - \boldsymbol{R_A}|} = -\frac{1}{2}\nabla^2 + V_{ext} \\ \hat{\mathcal{J}}_j\,\phi_i(\boldsymbol{x}) &= \int \frac{\phi_j^*(\boldsymbol{x}')\phi_j(\boldsymbol{x}')}{|\boldsymbol{r} - \boldsymbol{r}'|}\phi_i(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}' \\ \hat{\mathcal{K}}_j\,\phi_i(\boldsymbol{x}) &= \int \frac{\phi_j^*(\boldsymbol{x}')\phi_i(\boldsymbol{x}')}{|\boldsymbol{r} - \boldsymbol{r}'|}\phi_j(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}' \end{aligned}$$

The Coulomb term represents the classical interaction between charge distributions. However, no such classical analogue exists for the exchange term, as it originates from the antisymmetry constraint of the wavefunction with

respect to particle exchange, also referred to as the Pauli exclusion principle. To solve Equation (2.10), the single-particle orbitals are expanded into a set of basisfunctions (most commonly Gaussian functions or plane waves)

$$\phi_i(\boldsymbol{x}) = \sum_j c_{ij} b_j(\boldsymbol{x}) \tag{2.11}$$

with c_{ij} the coefficients to be calculated, which transforms Equation (2.10) into the Hartree-Fock-Roothaan equation $\boldsymbol{FC} = \boldsymbol{SC}\boldsymbol{\epsilon}$, with $\boldsymbol{F}$ the Fock matrix, $\boldsymbol{C}$ the basis set coefficient matrix, $\boldsymbol{S}$ the basis set overlap matrix and $\boldsymbol{\epsilon}$ a diagonal matrix with the orbital energies.[117] While this looks like a standard eigenvalue equation, the Fock matrix depends on the orbitals (through $\boldsymbol{C}$), requiring a self-consistent iterative solution method until the obtained orbitals and energy converge. Due to the form of the Coulomb and exchange operators (electrons interact with the *average* distribution of other electrons), the HF approach is a *mean field* theory neglecting the correlation energy resulting from the instantaneous interaction between electrons, representing one of its most important limitations. However, post-HF methods which include these effects exist, with the most common ones being Møller–Plesset perturbation to the second order (MP2), configuration interaction (CI) and coupled cluster theory (CC). In MP2, the true Hamiltonian is used in a perturbation to the unperturbed Fock operator, recovering some of the correlation energy. [119, 120] On the other hand, coupled cluster theory postulates an exponential ansatz for the wavefunction,[121–123] in which a Slater determinant serves as the reference wavefunction $|\psi_0\rangle$:

$$|\psi\rangle = e^{\hat{T}} |\psi_0\rangle \tag{2.12}$$

with $\hat{T}$ the cluster operator. The variant of coupled cluster often considered the golden standard for quantum chemistry calculations consists of a truncation of the cluster operator at the second level (including single and double excitations) with a perturbative treatment of the triple excitations, referred to as CCSD(T). [124, 125] While such approaches to include electron correlation are highly accurate, their practical use is severely limited by their computational cost. A popular alternative to wavefunction based methods with more favorable computational scaling than CCSD(T) are the electron-density based methods.

Density-functional theory

Although the wavefunction $|\psi\rangle$ encodes all properties of a system, it is a complex high-dimensional function of the spatial and spin coordinates of

each electron. Consider instead the electron density

$$n(\boldsymbol{r}) = \langle\psi| \sum_i \delta(\boldsymbol{r} - \boldsymbol{r}_i)|\psi\rangle \tag{2.13}$$

$$= N \int |\psi(\boldsymbol{r}, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)|^2 \, \mathrm{d}\boldsymbol{r}_2 \ldots \mathrm{d}\boldsymbol{r}_N \tag{2.14}$$

which is a 3-dimensional scalar function. Herein, we restricting ourselves to the ground state of the system. At first sight, the electron density contains considerably less information than the wavefunction. However, two theorems proven by Hohenberg and Kohn in 1964 established this not to be the case.[126] First, the ground state electron density uniquely determines the external potential and therefore the Hamiltonian, making it possible to define the ground state energy E as a functional of the density n. Second, a variational principle for the density was established, proving that $E[n(\boldsymbol{r})]$ is minimal when $n(\boldsymbol{r})$ is the true ground state (subject to the constraint $N = \int n(\boldsymbol{r})\mathrm{d}\boldsymbol{r}$). Given that the Hamiltonian consists of a kinetic energy term $\hat{\mathcal{T}}_e$, an external potential $\hat{\mathcal{V}}_{Ne}$ and the electron-electron interaction $\hat{\mathcal{V}}_{ee}$ ($\hat{\mathcal{V}}_{NN}$ is a constant), the energy functional $E[n(\boldsymbol{r})]$ can be written as

$$E[n(\boldsymbol{r})] = T[n(\boldsymbol{r})] + V_{\mathrm{ext}}[n(\boldsymbol{r})] + V_{ee}[n(\boldsymbol{r})] \tag{2.15}$$

Herein, only the interaction with the external potential is known exactly, although significant contributions to the other functionals are known. First, an important part of the electron-electron interaction is known as the Coulomb interaction or Hartree term, which can be written as a functional of the density $V_H[n(\boldsymbol{r})]$. Second, Kohn and Sham made use of the following model for the kinetic term.[127] Consider a fictitious system of N non-interacting fermions described by a Slater determinant consisting of N fictitious orbitals ϕ_i which reproduce the true ground state density. For this system, a kinetic term $T_s(\boldsymbol{r})$ is known exactly:

$$V_{\mathrm{ext}}[n(\boldsymbol{r})] = \int V_{\mathrm{ext}}(\boldsymbol{r}) n(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r} \tag{2.16}$$

$$V_H[n(\boldsymbol{r})] = \frac{1}{2} \int \frac{n(\boldsymbol{r}_1) n(\boldsymbol{r}_2)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|} \, \mathrm{d}\boldsymbol{r}_1 \mathrm{d}\boldsymbol{r}_2 \tag{2.17}$$

$$T_s[n(\boldsymbol{r})] = -\frac{1}{2} \sum_i \langle\phi_i|\nabla^2|\phi_i\rangle \tag{2.18}$$

Using these terms, the energy functional can be rewritten as

$$E[n(\boldsymbol{r})] = T_s[n(\boldsymbol{r})] + V_{\mathrm{ext}}[n(\boldsymbol{r})] + V_H[n(\boldsymbol{r})] + E_{xc}[n(\boldsymbol{r})] \tag{2.19}$$

with $E_{xc}[n(\boldsymbol{r})]$ the exchange-correlation functional representing the difference between treating the electron-electron interaction classically and from using the non-interacting kinetic energy

$$E_{xc}[n(\boldsymbol{r})] = (T[n(\boldsymbol{r})] - T_s[n(\boldsymbol{r})]) + (V_{ee}[n(\boldsymbol{r})] - V_H[n(\boldsymbol{r})]) \tag{2.20}$$

The variational principle together with a Lagrange multiplier constraint on the density can now be used to minimize the functional

$$\frac{\delta}{\delta n(\boldsymbol{r})}\left[E[n(\boldsymbol{r})] - \mu \int n(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}\right] = \frac{\delta E[n(\boldsymbol{r})]}{\delta n(\boldsymbol{r})} - \mu = 0 \tag{2.21}$$

and therefore

$$\frac{\delta T_s[n(\boldsymbol{r})]}{\delta n(\boldsymbol{r})} + V_{\mathrm{ext}}(\boldsymbol{r}) + \int \frac{n(\boldsymbol{r})}{|\boldsymbol{r} - \boldsymbol{r}'|}\,\mathrm{d}\boldsymbol{r} + \frac{\delta E_{xc}[n(\boldsymbol{r})]}{\delta n(\boldsymbol{r})} = \mu \tag{2.22}$$

Again, making a comparison between the true system subject to an external potential V_{ext} and a system consisting of N non-interacting electrons with the same density, this hypothetical system should be subject to a different external potential V_s, for which the variational principle (from Equation (2.15)) leads to

$$\frac{\delta E[n(\boldsymbol{r})]}{\delta n(\boldsymbol{r})} = \frac{\delta T_s[n(\boldsymbol{r})]}{\delta n(\boldsymbol{r})} + V_s(\boldsymbol{r}) = \mu \tag{2.23}$$

Both systems described by Equations (2.22) and (2.23) are therefore identical in case

$$V_s(\boldsymbol{r}) = V_{\mathrm{ext}}(\boldsymbol{r}) + \int \frac{n(\boldsymbol{r})}{|\boldsymbol{r} - \boldsymbol{r}'|}\,\mathrm{d}\boldsymbol{r} + \frac{\delta E_{xc}[n(\boldsymbol{r})]}{\delta n(\boldsymbol{r})} \tag{2.24}$$

In other words, the ground state of the true system can be found by solving the fictitious non-interacting system with an external potential given by Equation (2.24), allowing us to use the HF formalism given by Equation (2.10) for the orbitals of the non-interacting system

$$\left(-\frac{1}{2}\nabla_i^2 + V_s(\boldsymbol{r})\right)\phi_i = \epsilon_i \phi_i \tag{2.25}$$

with the density $n(\boldsymbol{r})$ equal to

$$n(\boldsymbol{r}) = \sum_i f_i\,|\phi_i(\boldsymbol{r})|^2 \tag{2.26}$$

with f_i the occupancy of the orbitals ($0 \leq f_i \leq 2$). Although the complete derivation so far is exact, the form of the exchange-correlation energy E_{xc} is not known. However, many approximations exist with a varying degree of

empiricism, complexity and computational cost. The simplest approximation assumes E_{xc} to be a functional of $n(\boldsymbol{r})$, aptly named the local density approximation (LDA). Increasing in complexity, a dependence on gradient $\nabla n(\boldsymbol{r})$ and laplacian $\nabla^2 n(\boldsymbol{r})$ are included, termed the generalized-gradient approximation (GGA) and meta-GGA, respectively. The GGA family of functionals have been used most often, with the Perdew-Ernzerhof-Burke (PBE) functional being the workhorse of solid state physics.[128, 129] In even more advanced approaches, a part of the exact Hartree-Fock exchange computed from the occupied Kohn-Sham orbitals is added, termed hybrid functionals or hybrids. Despite the sound theoretical foundations of density-functional theory (DFT), significant errors remain due to the introduction of approximate exchange-correlation functionals. In near-degenerate systems which cannot be accurately be described with a single Slater determinant, static correlation errors crop up. Artificial delocalization of the electron density is also known affect DFT, with fractional charge occupation being artificially stabilized.[130] However, one of the most severe limitations affecting DFT when describing non-covalent interactions between molecules is the lack of long-range dispersion.[131–134] Due to the semi-local nature of the approximate functionals, the dynamic correlation of instantaneously interacting dipole fluctuations cannot be accurately captured. To alleviate this problem, empirical dispersion corrections exist based on the theoretical insight that the leading dispersion terms between atoms scale with the inverse power of six and eight in the distance.[135] Commonly, a D3 dispersion correction is added to the ground state energy

$$E_{\text{D3}} = -\frac{1}{2}\sum_{i,j}\sum_{n=6,8} f_n(r_{ij})\frac{C_n^{ij}}{r_{ij}^n} \tag{2.27}$$

with $f_n(r_{ij})$ a short-range damping function, C_n^{ij} the n^{th} order dispersion coefficient and r_{ij}^n the interatomic distance between atom i and j.[136] With proper parameterization, this empirical contribution recovers most of the missing long-range correlation energy. While DFT has witnessed wide-scale adoption in material science and computational chemistry, its computational cost still limits its use in applications where large system sizes are required, only recently being applied to protein-ligand binding or mesoporous materials.[137, 138] For truly large-scale simulations, force fields are required.

2.1.3 Force field methods

Before providing an overview of force field methods, it is instructive to remember the target applications for molecular modeling. In some applications

such as the determination of the band gap of a material or the excited states of a molecule, it is essential to accurately capture the electronic structure of the system at hand. However, in other types of applications such as adsorption or phase transformations in nanoporous materials, protein-ligand docking, or the modeling of viruses such as HIV-1, only the structure, dynamics or statistics of the nuclei is of interest.[139–141] Of course, nuclei interact with the surrounding electrons, but the exact electronic structure is not of interest. One could therefore imagine an approach which only retains the nuclei which interact through empirical potentials. This is exactly the force field approach. Grounded in the many-body expansion of the energy, a force field decomposes the total energy into n-body contributions

$$E = \sum_i E_i + \sum_{i<j} E_{ij} + \sum_{i<j<k} E_{ijk} + \sum_{i<j<k<l} E_{ijkl} + \cdots \tag{2.28}$$

with the indices $i, j, k, l, \cdots$ running over all atoms in the system. Typically, these methods employ different functional forms to describe the covalent (bonding) and non-covalent (non-bonding) interactions, each grounded in different physical considerations. Therefore, each will be discussed separately.

Covalent force fields

Covalent force fields describe the interaction of molecules or materials connected through chemical bonds. Generally, as the absolute force field energy has no physical meaning, the first term in the many-body expansion representing atomic energies is omitted. For the next terms, inspiration can be drawn from considering some simple molecular systems. First, consider a linear carbondioxide molecule. Scanning the potential energy surface by linearly displacing one of the oxygen atoms with respect to the rest length, the profile shown in blue in Figure 2.1(a) is obtained using the PBE-D3(BJ) functional and the def2-tzvp basis set. Fitting a harmonic term to this profile, a near perfect fit is obtained for this two-body term. Similarly, the three-body bending mode in water (shown in Figure 2.1(b)) can be represented by a harmonic term in the angle. Several different choices for representing the four-body term exist such as out-of-plane distance or torsion form. For example, the rotation of the methyl groups comprising ethane (see Figure 2.1(c)) can be represented by a torsion term with shifted cosine functional form with multiplicity m = 3.

While the preceding discussion demonstrates how covalent force fields can be derived from fitting to DFT calculations, early generic force fields followed

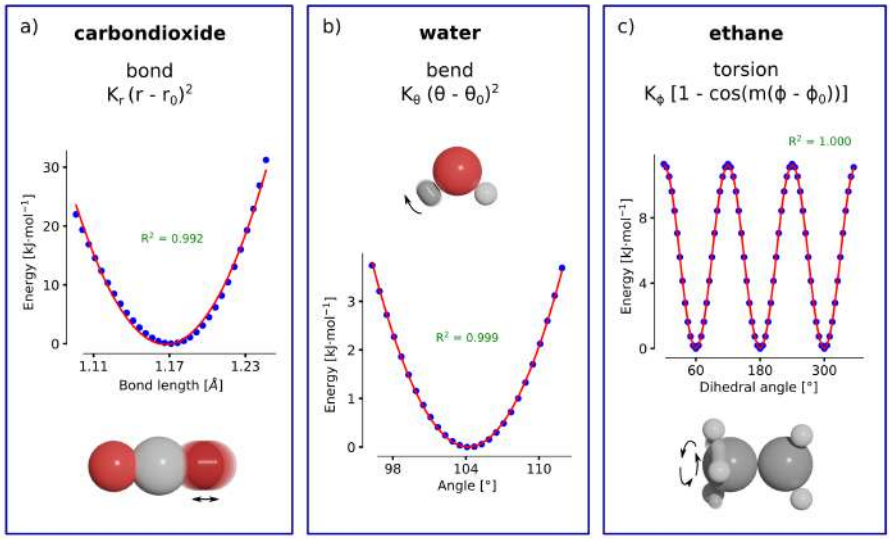


Figure 2.1: Scans of the PES obtained with PBE-D3(BJ)/def2-pvtz of (a) the carbondioxide bond, (b) the water bending mode and (c) torsion of a methyl group in ethane, together with analytical functions fitted to the DFT data, demonstrating good agreement.

different strategies. Force fields such as CHARMM, UFF and DREIDING make use of *atom types*, defined by their chemical symbol, hybridization and/or oxidation state.[89, 91, 92] For each atom type, bond distances and equilibrium angles were determined from experimental structures, while force constants were set to empirical values, or derived from the classical Coulomb interaction between effective charges.[142] Supplemented with torsion and inversion terms, these force field's reliance on atom types make them generally applicable to a wide range of systems. However, a trade off between wide applicability and accuracy usually exists, motivating the development of system-specific force fields such as UFF4MOF, which extends UFF with atom types occurring in typical MOFs, combined with a refitting of bond angles and radii. [90]

With the advent of high-performance computing (HPC) capabilities, force fields have started relying more on DFT calculations for fitting. In the *force matching* approach,[143] force field parameters are derived from a minimization of a cost function including computed energies, forces and/or stresses obtained from a set of representative configurations. Although this approach suffers from an ill-conditioning of the cost functions, several codes such as potfit, ForceFit and ForceBalance exist to perform this minimization.[144–146] In

another approach coined QuickFF, the Hessian matrix (consisting of second order derivatives of the energy) in equilibrium is used in conjunction with perturbation trajectories, avoiding this limitation.[96, 97] This methodology is of particular interest for this work, as it has been shown to accurately capture the framework flexibility of the empty MIL-53(Al), UiO-66 and DUT-49 MOFs. [147–150]

Non-covalent force fields

Essential for the description of liquids or adsorption of gases in MOFs are the non-covalent or non-bonded interactions between molecules or molecules with materials. Within these interactions, a division is usually made between long-range and short-range effects, as they have different physical origins. First, the long-range effects are considered. Consider two isolated molecules A and B, each described by a wavefunction (ψ_A, ψ_B) and a Hamiltonian ($\hat{\mathcal{H}}_A, \hat{\mathcal{H}}_B$). Suppose these molecules interact at a sufficiently large distance such that there is no overlap in the wavefunctions. In this case, the combined wavefunction can be written as a simple Hartree product of ψ_A and ψ_B, and the Hamiltonian can be written as $\hat{\mathcal{H}} = \hat{\mathcal{H}}_A + \hat{\mathcal{H}}_B + \hat{\mathcal{H}}'$ with $\hat{\mathcal{H}}'$ the electrostatic interaction between the particles of molecule A and molecule B. This intermolecular interaction $\hat{\mathcal{H}}'$ can be treated as a perturbation on the unperturbed Hamiltonian $\hat{\mathcal{H}}_0 = \hat{\mathcal{H}}_A + \hat{\mathcal{H}}_B$. Using Rayleigh-Schrödinger perturbation theory, the first-order energy term (the expectation value of $\hat{\mathcal{H}}'$ in the ground state) is simply equal to the intermolecular electrostatic interaction. The second-order perturbation term is a combination of the induction energy of molecule A (molecule A is in an excited state, molecule B is in the ground state), the induction energy of molecule B (molecule A is in the ground state, molecule A is in an excited state), and the dispersion energy (both molecules are excited).

When the wavefunctions of the separate molecules overlap, an antisymmetrization of the total wavefunction is required, leading to the need for symmetry-adapted perturbation theories such as SAPT.[151] A full discussion of the underlying SAPT theory would lead us too far. However, the most important short-range interaction which does not appear in the long-range effects can also be understood from HF, where the antisymmetry constraint leads to an exchange term, originating from the overlap between orbitals (see Equation (2.10)). As the orbitals decay exponentially with the distance, the so-called exchange-repulsion interaction also (to a good approximation) decays exponentially with the distance. For an essential overview of intermolecular interactions, the interested reader is referred to the book *The Theory of Intermolecular Forces* by Stone.[152]

From the preceding discussion, the development of intermolecular force fields can be grounded in a solid physical basis. To capture the electrostatic intermolecular interactions, most force fields make use of atomic charges, *allocating* the electronic charge density to the composing atoms of the system. This allocation procedure is commonly referred to as an atoms in molecules (AIM) or charge partitioning scheme. Many such schemes exist, which can be divided in methods which derive charges based on electronegativity or charge equilibration,[153–155] or from a preceding HF or DFT calculation. In the latter, different methods aim at reproducing the electrostatic potential around the molecule[156–159] or partitioning the density matrix or the spacial electron density.[160–166] However, as the atomic charge in a molecule is not an observable in quantum mechanics, a charge partitioning scheme cannot really be chosen for its 'correctness'. They are instead meant to be reliable and 'useful' for specific applications. Given a set of atomic charges, the electrostatic intermolecular interaction energy can be calculated from the classical Coulomb interaction. The dispersion interaction (in its leading term) is known to decay with the inverse sixth power. While it can be modeled separately from other terms as in Equation (2.27), most forcefields combine this interaction with a model for the exchange-repulsion (and other neglected) terms. The most common functional forms are the Lennard-Jones (LJ) or the Buckingham potential used in the MM3 force field

$$E_{\mathrm{LJ}} = \frac{1}{2}\sum_{i\neq j} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.29}$$

$$E_{\mathrm{MM3}} = \frac{1}{2}\sum_{i\neq j} \epsilon_{ij}\left[1.84\cdot 10^{5}\cdot e^{-\frac{12 r_{ij}}{\sigma_{ij}}} - 2.25\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.30}$$

with r_{ij} the distance between atom i and j. The van der Waals parameters σ_{ij} and ϵ_{ij} are usually determined with mixing rules from the atomic parameters ($\sigma_{ij} = 0.5(\sigma_i+\sigma_j)$ or $\sigma_{ij} = \sqrt{\sigma_i\sigma_j}$ and $\epsilon_{ij} = \sqrt{\epsilon_i\epsilon_j}$). In both potentials, the latter term represents the dispersion interaction with the physically justified $-r^{-6}$ form. The first term represents the exchange-repulsion. The MM3 potential has the appropriate exponential form while the Lennard-Jones has a r^{-12} form, which is less physically justified, chosen instead from the pragmatic standpoint that this term is easy to compute when the r^{-6} dispersion term has been calculated.[167]

Most force fields, such as the ones discussed in the previous Section on covalent force fields restrict themselves to the following two forms to capture intermolecular interactions: a pair-wise additive electrostatic and an LJ or

MM3 term. This *frozen electron* approach neglects any N–body ($N \geq 3$) terms. For example, it neglects the non-additivity of the previously introduced induction interaction, caused by the redistribution of the molecular electron densities due to their mutual interaction. Several methods to include these effects in force fields exist, usually grouped in three families. The first is the Drude oscillator model, in which atoms are modeled as a core charge $q + \Delta q$ and a negative shell $-\Delta q$ connected to the core charge with a harmonic spring.[168, 169] Due to the field generated by the surrounding charges, charge separation occurs between the core charge and shell, modeling the induction interaction. Another approach is that of fluctuating charges, using the previously mentioned electronegativity or charge equilibration on the fly during a simulation.[153, 154] The third, most used approach, is the inducible dipole model. This method introduces a dipole $\boldsymbol{\mu}_i$ at each atomic site,[170, 171] equal to the matrix product of the atomic polarizability tensor $\boldsymbol{\alpha}_i$ and the electric field at that site $\boldsymbol{E}_i$

$$\boldsymbol{\mu}_i = \boldsymbol{\alpha}_i \cdot \boldsymbol{E}_i = \boldsymbol{\alpha}_i \left(\boldsymbol{E}_{0,i} - \sum_{j \neq i} \boldsymbol{T}_{ij} \cdot \boldsymbol{\mu}_j \right) \tag{2.31}$$

with $\boldsymbol{T}_{ij}$ the dipole-dipole interaction tensor. Complicating the determination of the dipole moments is the fact that the total field $\boldsymbol{E}_i$ is composed of both the fields generated by surrounding charges ($\boldsymbol{E}_{0,i}$) and the induced dipoles (second term of Equation (2.31)), requiring a matrix inversion to solve Equation (2.31): $\boldsymbol{\mu} = \boldsymbol{A}^{-1}\boldsymbol{E}_0$ with the elements of $\boldsymbol{A}$ being

$$\boldsymbol{A}_{ij} = \begin{cases} \boldsymbol{\alpha}_i^{-1} & \text{if } i = j \\ \boldsymbol{T}_{ij} & \text{if } i \neq j \end{cases} \tag{2.32}$$

with these dipoles, the polarization energy is equal to

$$U_{\text{pol}} = -\frac{1}{2} \sum_i \boldsymbol{\mu}_i^T \boldsymbol{E}_{0,i} \tag{2.33}$$

what distinguishes different inducible dipole models is the form of the dipole-dipole interaction tensor $\boldsymbol{T}_{ij}$, and the static charge distribution. The most popular approaches use point dipoles or Gaussian dipoles[172, 173]

$$\boldsymbol{T}_{ij,\text{point}} = \frac{1}{r_{ij}^3}\boldsymbol{I} - \frac{3\boldsymbol{r}_{ij} \otimes \boldsymbol{r}_{ij}}{r_{ij}^5} \tag{2.34}$$

$$\boldsymbol{T}_{ij,\text{Gaussian}} = \beta_{ij}^3 \left(B_1(x)\boldsymbol{I} - \beta_{ij}^2 \boldsymbol{r}_{ij} \otimes \boldsymbol{r}_{ij} B_2(x) \right) \tag{2.35}$$

with $r_{ij} = |\boldsymbol{r}_{ij}|$, $\beta_{ij} = \frac{\beta_i \beta_j}{\sqrt{\beta_i^2 \beta_j^2}}$, $B_0(x) = \frac{\text{erf}(x)}{x}$, $B_1(x) = -\frac{1}{x}\frac{\mathrm{d}B_0(x)}{\mathrm{d}x}$, $B_2(x) = -\frac{1}{x}\frac{\mathrm{d}B_1(x)}{\mathrm{d}x}$, $x = \beta_{ij}\boldsymbol{r}_{ij}$ and β_i the exponent of the Gaussian function

on atom i. However, both models can suffer from the so-called polarization catastrophe.[170] For the point dipole model, the matrix $\boldsymbol{A}_{ij}$ of Equation (2.32) contains negative eigenvalues when a pair of dipoles interacts at a distance $r_{ij} < (4\alpha_i\alpha_j)^{1/6}$, leading to unphysical results. While divergence of the interaction for two interacting dipoles can be avoided by introducing damping functions in the dipole-dipole interaction tensor $\boldsymbol{T}_{ij}$,[170] positive definiteness is not assured for larger systems. The Gaussian dipole model avoids the polarization catastrophe by use of sufficiently small exponents. However, both of these models require the fitting of atomic polarizabilities for a set of atom types from experimental or computed molecular polarizabilities. Moreover, Gaussian functions are not the most appropriate description of the atomic response density. A Slater function based transferable polarization model derived from the monomer electron density is therefore developed in **Paper III**, together with a new interaction model inducing the dipoles. It should be noted that polarization models do not completely capture the induction energy, as both higher order polarization (quadrupole, octopole, ...) and charge transfer effects are generally neglected. This presents a major limitation, as charge transfer effects can even dominate over electrostatic and polarization effects, as shown for example for the solvation of biomolecules.[174]

Although the previously described functional forms for the description of non-covalent interactions are at least somewhat physically inspired, recent developments have focused on potentials more closely related to *ab initio* methods. SAPTFF was developed to decompose interactions in the same way as SAPT, giving rise to an exchange, electrostatic, induction and dispersion contribution.[175] Making use of the monomer electron density, atomic charges, static polarizabilities and dispersion coefficients can be derived from a distributed multipole analysis (DMA).[176] Similarly *ab initio* derived force fields include Slater-ISA, ISA-Pol, MEDFF, MASTIFF and the ACKS2 polarizable force field,[177–181] all with promise of being used as accurate next-generation non-covalent forcefields.

Machine learning potentials

Traditionally, force fields make use of physically inspired functional forms, together with a set of appropriately fitted parameters. Recently, a new approach has emerged. Instead of presupposing certain functional forms, the chemical environment of each atom is represented by atomic descriptors or *features*. The earliest methods made use of atomic descriptors. In 2007, Behler and Parrinello introduced atom-centered symmetry functions (ACSF).[182–184] In this method, the radial information of atoms surrounding

atoms a can be encoded using the following descriptor

$$G^{\text{rad}}_{a,\nu\mu} = \sum_b e^{-\nu(r_{ab}-\mu)^2} f_{\text{cutoff}}(r_{ab}) \tag{2.36}$$

with different values for μ and ν probing different regions surrounding the atom a and f_{cutoff} a cutoff function to ensure continuous first derivatives. As the sum runs over atoms b within a cutoff radius around atoms a, this descriptor encodes only local information. Angular information can be encoded in a similar way. Using a combination of these types of atomic descriptors for each atom $\boldsymbol{x}_a$, kernel methods can be used to learn the relation between the atomic descriptors and the energy. Using a kernel $K(\boldsymbol{x}_a, \boldsymbol{x}_b)$ the energy of each atom E_a is linearized as a sum of these descriptors in an abstract feature space with regression weights α_k[185]

$$E_a = \sum_k \alpha_k K(\boldsymbol{x}_a, \boldsymbol{x}_k) \tag{2.37}$$

A important limitation of kernel methods is their computational scaling. Due to the use of a kernel, the training complexity scales as $\mathcal{O}(N^2)$ with N the number of training samples. Alternative to this, deep neural networks can be used to learn the relation between the descriptors and the energy, scaling linearly with N. In its most basic form, neural networks represent a high-dimensional function (such as the energy of a system) by a series of *layers*, connected through linear combinations of the previous layer combined with non-linear activation functions. Several descriptor-based neural networks have been developed, such as ANI-1, DeepMD and Behler-Parrinello HDNNPs.[183, 186, 187] However, these descriptor-based neural networks are still limited by the choice of atomic descriptors. The recent development of message passing neural networks omit the use of these manually constructed descriptors.[188] Again, each atom is characterized by a feature vector $\boldsymbol{x}_a$, although these are not based on atomic descriptors. Instead, they are updated with a learned function U_t at step t

$$\boldsymbol{x}_a^{t+1} = U_t(\boldsymbol{x}_a^t, \boldsymbol{m}_a^{t+1}) \tag{2.38}$$

based on *messages* which are computed with a learned function M_t

$$\boldsymbol{m}_a^{t+1} = \sum_b M_t(\boldsymbol{x}_a^t, \boldsymbol{x}_b^t, \boldsymbol{e}_{ab}) \tag{2.39}$$

In this Equation, a sum is taken over the atoms b neighboring atom a to within a certain cutoff radius, making use of *edge features* $\boldsymbol{e}_{ab}$ between atom a and b (for example the interatomic distance r_{ab}). Lastly, the atomic energies

are computed from the features $\boldsymbol{x}_a^T$ at the final step $t = T$. Through these *messages*, the interaction between atoms can be propagated through iterations t from 0 to T. Popular architectures based on this message passing method include SchNet and NequiP, with the latter demonstrating state-of-the-art performance on a set of different benchmarks.[189, 190] While these machine learning potentials (MLPs) represent a major innovation in the field of quantum chemistry, high-quality reference data (energies, forces, stresses) is still required for training to the system of interest. However, once an MLP is trained, it can be used with a computational cost orders of magnitude smaller than HF, DFT, MP2 or CCSD(T) calculations. In **Paper IV**, a methodology for training an MLP on interaction energies will be developed and used to perform adsorption simulations of CO_2 in ZIF-8 and Mg-MOF-74 with DFT accuracy.

2.2 Sampling the potential energy surface

In Section 2.1, the concept of a potential energy surface (PES) was developed. However, reminding ourselves of the goal of molecular modeling, we aim to connect the detailed microscopic description of materials with macroscopic observations such as "What is the adsorption capacity of this material?" or "How does the (meta)stability of this material change with increasing gas pressure?". The link between the microscopic interactions which can be probed by simulations and macroscopic observations is provided by ensemble theory in statistical physics.[191, 192] Herein, ensembles are defined by their thermodynamic control variables such as temperature, pressure and chemical potential. Such macroscopic ensembles can be realized by a large number of microscopic states (defined by the atomic positions $\{\boldsymbol{r}_i\}$ and momenta $\{\boldsymbol{p}_i\}$). The space spanned by the positions and momenta of all particles is called the phase space.

The microcanonical ensemble

Central to statistical physics is the assumption that every microscopic state consistent with the macroscopic conditions of an isolated system is equally likely.[191, 192] Under this assumption, the probability p of each microstate in the microcanonical ensemble (N, V, E) can be obtained from the total number of microstates $\Omega(N, V, E)$

$$p = \frac{\delta(E - \mathcal{H}(\boldsymbol{r}, \boldsymbol{p}))}{\Omega(N, V, E)} \tag{2.40}$$

with $\mathcal{H}$ the Hamiltonian. By means of an integration over the phase space in the small energy range E to $E + E_0$, $\Omega(N, V, E)$ can be obtained as follows

$$\Omega(N, V, E) = \frac{E_0}{h^{3N} N!} \int \int \mathrm{d}\boldsymbol{r} \mathrm{d}\boldsymbol{p}\, \delta(E - \mathcal{H}(\boldsymbol{r}, \boldsymbol{p})) \tag{2.41}$$

The prefactor is merely a normalization constant and does not affect further properties. From Equation (2.41), the thermodynamic properties of entropy S and temperature T can be defined as

$$S = k_B \ln(\Omega) \tag{2.42}$$

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_{N,V} \tag{2.43}$$

with k_B the Boltzmann constant. While this is an important starting point, experimental conditions usually control the temperature, not the energy.

The canonical ensemble

The canonical or (N, V, T) ensemble can be realized by connecting the system with a thermal reservoir with which energy can be exchanged. The total number of microstates of the system $Q(N, V, T)$ (called the partition function) is equal to[191]

$$Q(N, V, T) = \frac{1}{h^{3N} N!} \int \int \mathrm{d}\boldsymbol{r} \mathrm{d}\boldsymbol{p}\, e^{-\beta \mathcal{H}(\boldsymbol{r}, \boldsymbol{p})} \tag{2.44}$$

$$= \frac{1}{\Lambda^{3N} N!} \int \mathrm{d}\boldsymbol{r}\, e^{-\beta \mathcal{U}(\boldsymbol{r})} \tag{2.45}$$

with $\beta = \frac{1}{k_B T}$. The integration over the momenta can be performed analytically, leaving only the integration over coordinates with a potential energy $\mathcal{U}$. The prefactor contains a term $\Lambda = \frac{h}{\sqrt{2\pi m k_B T}}$ which is called the thermal de Broglie length. With this partition function, the probability of encountering the system at coordinates $\boldsymbol{r}$ is

$$p(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} = \frac{1}{\Lambda^{3N} N!} \frac{e^{-\beta \mathcal{U}(\boldsymbol{r})}}{Q(N, V, T)} \mathrm{d}\boldsymbol{r} \tag{2.46}$$

From the partition function of Equation (2.45) the thermodynamic potential called the Helmholtz free energy of the canonical ensemble is

$$F(N, V, T) = -k_B T \ln(Q) \tag{2.47}$$

from which the pressure P, entropy S and chemical potential μ can be determined.

The grand canonical ensemble

The previous ensembles kept a fixed number of particles N. However, to study adsorption, an ensemble is needed in which particles can be exchanged with an external reservoir. This is achieved in the grand canonical ensemble, retaining the chemical potential μ (the free energy change from adding a particle adiabatically), the volume V and temperature T as control variables. The associated grand canonical partition function is

$$\zeta(\mu, V, T) = \sum_{N=0}^{+\infty} \frac{e^{\beta\mu N}}{\Lambda^{3N}} \int \mathrm{d}\boldsymbol{r} e^{-\beta \mathcal{U}(\boldsymbol{r})} \tag{2.48}$$

$$= \sum_{N=0}^{+\infty} Q(N, V, T) e^{\beta\mu N} \tag{2.49}$$

with $Q(N, V, T)$ the canonical partition function. The thermodynamic potential in the (μ, V, T) ensemble is equal to

$$\Omega(\mu, V, T) = -k_B T \ln(\zeta) = F(N(\mu), V, T) - N(\mu)\mu \tag{2.50}$$

The isobaric-isothermal ensemble

As discussed in Section 1.2.2, MOFs can exhibit significant flexibility of the volume and cell shape. Controlling the pressure P instead of the volume, together with the temperature T and number of particles N gives rise to the isobaric-isothermal ensemble. Consider first only isotropic cell fluctuations. The partition function (neglecting constant prefactors) in this ensemble is

$$\Delta(N, P, T) = \int \mathrm{d}V \int \mathrm{d}\boldsymbol{r}\, e^{-\beta P V} e^{-\beta \mathcal{U}(\boldsymbol{r})} \tag{2.51}$$

The associated thermodynamic potential is the Gibbs free energy

$$G(N, P, T) = -k_B T \ln(\Delta) = F(N, V(P), T) + PV(P) \tag{2.52}$$

To additionally include anisotropic cell fluctuation, as for example needed to describe the orthorhombic to monoclinic phase transformation of MIL-53(Al), the cell matrix $\mathbf{h}$ can be written as $\mathbf{h} = V^{1/3}\mathbf{h}_0$ with $\det(\mathbf{h}_0) = 1$. Restricting ourselves to an isotropic applied pressure P, the partition function can be written as[193]

$$\Delta(N, P, T) = \int \mathrm{d}V \int \mathrm{d}\mathbf{h}_0 \int \mathrm{d}\boldsymbol{r}\, e^{-\beta P V} e^{-\beta \mathcal{U}(\boldsymbol{r})}\, \delta(\det(\mathbf{h}_0) - 1) \tag{2.53}$$

Making use of the Equations $\mathbf{h} = V^{1/3}\mathbf{h}_0$, $\mathrm{d}\mathbf{h} = V^3\mathrm{d}\mathbf{h}_0$ and eliminating the volume with the δ-function, Equation (2.53) reduces to[191]

$$\Delta(N,P,T) = \int \mathrm{d}\mathbf{h} \int \mathrm{d}\boldsymbol{r}\; e^{-\beta P \det(\mathbf{h})}\, e^{-\beta \mathcal{U}(\boldsymbol{r})} \det(\mathbf{h})^{-2} \tag{2.54}$$

Lastly, an implicit volume-dependence is still present in the integration limits of the particle coordinates $\boldsymbol{r}$. Eliminating this dependence by introducing scaled coordinates $\boldsymbol{s} = \boldsymbol{r}/V^{1/3}$ results in

$$\Delta(N,P,T) = \int \mathrm{d}\mathbf{h} \int \mathrm{d}\boldsymbol{s}\; e^{-\beta P \det(\mathbf{h})}\, e^{-\beta \mathcal{U}(\boldsymbol{s},\mathbf{h})} \det(\mathbf{h})^{N-2} \tag{2.55}$$

Given this set of ensembles, our work might be considered finished. As each partition function can be obtained by an integration over particle coordinates (and/or the volume, the cell shape and the particle numbers), the thermodynamic variables of interest can also be calculated. Moreover, the ensemble average of an observable of interest $\langle A \rangle$ can be calculated, for example in the canonical ensemble, as

$$\langle A \rangle = \frac{\int \mathrm{d}\boldsymbol{r}\, e^{-\beta \mathcal{U}(\boldsymbol{r})} A(\boldsymbol{r})}{\int \mathrm{d}\boldsymbol{r}\, e^{-\beta \mathcal{U}(\boldsymbol{r})}} \tag{2.56}$$

in which A was assumed not to depend on the momenta $\boldsymbol{p}$. Unfortunately, this is only possible for the simplest systems. The integral in Equation (2.56) is $3N$-dimensional with N the number of particles, precluding it from being solved for realistic systems. However, most configurations occurring in the integral would yield negligible contributions due to the weighting with an exponential of the potential energy. In other words, methods are needed to *intelligently* sample the relevant parts of the phase space. Two prototypical simulation techniques to fulfill this role are molecular dynamics (MD) and Monte Carlo (MC).

2.2.1 Molecular dynamics

Given a method to compute the potential energy $\mathcal{U}$ of a system (see Section 2.1), the idea of molecular dynamics (MD) is to simply let the system evolve in time according to Newton's equations of motion. With the following Hamiltonian

$$\mathcal{H}(\boldsymbol{r},\boldsymbol{p}) = \sum_i \frac{\boldsymbol{p}_i^2}{2m_i} + \mathcal{U}(\boldsymbol{r}) \tag{2.57}$$

the equations of motion read

$$\dot{\boldsymbol{r}}_i = \nabla_{\boldsymbol{p}_i}\mathcal{H} = \frac{\boldsymbol{p}_i}{m_i} \tag{2.58a}$$

$$\dot{\boldsymbol{p}}_i = -\nabla_{\boldsymbol{r}_i}\mathcal{H} = \boldsymbol{F}_i \tag{2.58b}$$

with m_i the mass of particle i and $\boldsymbol{F}_i$ the force acting on this particle. Given a set of initial conditions, the equations are integrated numerically in time with an integration scheme. In each iteration, forces are evaluated and the positions and momenta of atoms are updated. Many such integration schemes exist. However, the most useful for applications are the symplectic, time reversible integration algorithms such as the velocity Verlet integrator.[191] The required time step for such integration depends on the highest frequency motions in the system, usually originating from hydrogen atoms. A typical time step is on the order of 0.5 fs.

From an MD simulation, a property of interest A can be computed by averaging over the simulation trajectory, a so-called *time average* $\overline{A}$. Such a time average clearly looks different from the *ensemble average* $\langle A \rangle$ of Equation (2.56). Indeed, it is not generally true that both coincide

$$\overline{A(\boldsymbol{r})} = \lim_{t\to\infty} \frac{1}{t} \int \mathrm{d}t' A(\boldsymbol{r}; t') \neq \langle A(\boldsymbol{r}) \rangle \tag{2.59}$$

The (often made) assumption that both are the same is called the *ergodic hypothesis*. While it often holds, the ergodic hypothesis can break down in systems with metastable phases. As materials exhibiting metastable phases will be investigated in this work, special care will have to be taken not to tacitly assume this hypothesis.

The straightforward application of Equation (2.58) leads to a sampling in the microcanonical or (N, V, E) ensemble. To perform simulations in the canonical or isobaric-isothermal ensembles, thermostats and/or barostats are used. Thermostats aim to control the temperature of a system. From the equipartition theorem, the instantaneous temperature T_{inst} can be related to the kinetic energy

$$\frac{1}{2} k_B N_f T_{\text{inst}} = E_{\text{kin}} = \frac{1}{2} \sum_i \frac{\boldsymbol{p}_i^2}{m_i} \tag{2.60}$$

with k_B the Boltzmann constant and N_f the degrees of freedom of the system. From this, a straight forward approach to thermostat a system is to intervene in the momenta $\boldsymbol{p}_i$ such that the average instantaneous temperature $\langle T_{\text{inst}} \rangle$ matches the desired temperature T. Common stochastic approaches to achieve this either make use of an artificial heat bath of particles interacting with the system through collisions and drag effects, or through a rescaling

of the velocities.[194–196] However, the introduced stochasticity induces time-irreversability. An alternative to this is the so-called extended phase space approach which supplements the phase space with additional variables which mimic the effect of a heat bath through a deterministic coupling. The prime example of this is the Nosé-Hoover chain (NHC) thermostat.[197]

Barostats aim to, analogously to thermostats, control the instantaneous internal stress $\boldsymbol{\sigma}_i$ of the system. As both isotropic and anisotropic cell fluctuations are allowed in the isobaric-isothermal ensemble, the internal stress is decomposed in an isotropic and anisotropic contribution

$$\boldsymbol{\sigma}_i = P_i \mathbf{1} + \boldsymbol{\sigma}_{i,a} \tag{2.61}$$

with $P_i = \frac{\mathrm{Tr}(\boldsymbol{\sigma}_i)}{3}$ and $\mathrm{Tr}(\boldsymbol{\sigma}_{i,a}) = 0$. Deterministic barostats were developed by Andersen, Berendsen and Hoover.[195, 198, 199] However, they only modify the volume of the cell $V = \det(\mathbf{h})$, leaving the cell shape $\mathbf{h}_0$ fixed. In 1994, the Martyna-Tuckerman-Tobias-Klein (MTTK) barostat was developed, coupling to both the volume and the cell shape, controlling the pressure P_i and anisotropic stress $\boldsymbol{\sigma}_a$.[193, 200] In 2015, Rogge *et al.* modified the original MTTK equations to separate the update of the cell volume V and cell shape $\mathbf{h}_0$, introducing an ensemble in which the cell shape can fluctuate without changing the volume, giving us all the necessary tools to study framework flexibility.[147]

2.2.2 Grand canonical Monte Carlo

In Section 2.2.1, the MD tools which can be employed to sample ensembles with fixed particle numbers N were discussed. However, the discrete nature of particle fluctuations does not neatly fit into the continuous MD framework.[191] Instead, the grand canonical Monte Carlo (GCMC) method is often used to exchange particles with an external reservoir in order to model adsorption. Within the grand canonical ensemble (see Equation (2.48)), the ensemble average of a property $\langle A \rangle$ can be written as

$$\langle A \rangle = \frac{\sum_{N=0}^{+\infty} \frac{e^{\beta\mu N}}{\Lambda^{3N}} \int \mathrm{d}\boldsymbol{r} e^{-\beta \mathcal{U}(\boldsymbol{r})} A(N, \boldsymbol{r})}{\sum_{N=0}^{+\infty} \frac{e^{\beta\mu N}}{\Lambda^{3N}} \int \mathrm{d}\boldsymbol{r} e^{-\beta \mathcal{U}(\boldsymbol{r})}} = \sum_{N=0}^{+\infty} \int \mathrm{d}\boldsymbol{r} f(N, \boldsymbol{r}) A(N, \boldsymbol{r}) \tag{2.62}$$

with $f(N, \boldsymbol{r})$ a distribution defined as follows

$$f(N, \boldsymbol{r}) = \frac{\frac{e^{\beta\mu N}}{\Lambda^{3N}} e^{-\beta \mathcal{U}(\boldsymbol{r})}}{\sum_{N'=0}^{+\infty} \frac{e^{\beta\mu N'}}{\Lambda^{3N'}} \int \mathrm{d}\boldsymbol{r} e^{-\beta \mathcal{U}(\boldsymbol{r})}} = \frac{\frac{e^{\beta\mu N}}{\Lambda^{3N}} e^{-\beta \mathcal{U}(\boldsymbol{r})}}{\zeta(\mu, V, T)} \tag{2.63}$$

with $\zeta(\mu, V, T)$ the grand canonical partition function. As discussed before, an explicit calculation of ζ is unfeasible. However, suppose an algorithm exists for generating M samples $\{N_i, \boldsymbol{r}_i\}$ from $f(N, \boldsymbol{r})$. The ensemble average could then be estimated from

$$\langle \hat{A} \rangle = \frac{1}{M} \sum_{i=1}^{M} A(N_i, \boldsymbol{r}_i) \tag{2.64}$$

From the central limit theorem, it is guaranteed that $\lim_{M\to\infty} \langle \hat{A} \rangle = \langle A \rangle$.[201] For this to be a viable method, an algorithm is required for generating samples from the distribution $f(N, \boldsymbol{r})$. Note that, while the absolute probability of a configuration $f(N_i, \boldsymbol{r}_i)$ cannot be determined, the *relative* probability of two configurations i and j *can* easily be calculated

$$\frac{f(x_i)}{f(x_j)} = \frac{\frac{e^{\beta\mu N_i}}{\Lambda^{3N_i}} e^{-\beta \mathcal{U}(\boldsymbol{r}_i)}}{\frac{e^{\beta\mu N_j}}{\Lambda^{3N_j}} e^{-\beta \mathcal{U}(\boldsymbol{r}_j)}} \tag{2.65}$$

With $x_i = \{N_i, \boldsymbol{r}_i\}$. Herein lies the basis of the Metropolis Monte Carlo method.[202] A Markov chain $x_1 \to x_2 \to \cdots \to x_M$ (here presumed symmetric) of configurations is generated sequentially from the previous configuration with an acceptance rate

$$\text{acc}(x_i \to x_j) = \begin{cases} f(x_j)/f(x_i) & \text{if } f(x_j) < f(x_j) \\ 1 & \text{if } f(x_j) \geq f(x_j) \end{cases} \tag{2.66}$$

This choice of acceptance criterion made by Metropolis satisfies the condition of detailed balance ($f(x_i)\text{acc}(x_i \to x_j) = f(x_j)\text{acc}(x_j \to x_i)$), guaranteeing a proper sampling in the grand canonical ensemble. In its basic form, the generation of a new configuration consists of attempting an insertion, deletion, translation or rotation move to an adsorbate. Subsequently, the ratio $f(x_j)/f(x_i)$ is calculated. If it is larger than 1, the trial move is accepted. If it is smaller than 1, a random number $0 \leq n \leq 1$ is generated. If n is smaller than the ratio, the move is also accepted. Otherwise, the move is rejected and the subsequent Monte Carlo iteration starts from the original configuration x_i. The required number of iterations M is generally system, property and condition-dependent. Therefore, convergence tests are required to determine an appropriate value for M.

From the resulting grand canonical Monte Carlo simulations, the adsorption capacity, selectivity and heat of adsorption can be calculated. In principle, the framework flexibility of a MOF could also be incorporated by means of trial moves acting on the atoms constituting the framework. However,

constructing a complete set of trial moves which efficiently sample flexible modes of the framework is cumbersome and system-dependent. A much more efficient method is to incorporate short MD runs which sample the framework flexibility in a GCMC simulation, coined hybrid Monte Carlo.[203]

In Section 3.1, different variants of such schemes are critically compared and applied to the adsorption of neon, argon, xenon, methane and carbon dioxide in the flexible MIL-53(Al). Based on the obtained insight, a novel computational scheme is proposed, supplementing GCMC simulations with MD moves in the $(N_{\text{host}}, N_g, \boldsymbol{\sigma}_a = \mathbf{0}, V, T)$ ensemble, described in **Paper I**. This new scheme is shown to accurately capture the breathing behavior of MIL-53(Al) under xenon, methane and carbon dioxide adsorption, demonstrating it to be a useful tool for providing insight into adsorption-induced framework flexibility.

Furthermore, the proposed computational scheme is applied in **Paper II** to gain insight into the NGA transition of DUT-49 under methane adsorption (see Section 3.2). From a determination of the osmotic landscape, both the observed transitions as well as the hysteresis between adsorption and desorption are reproduced. Moreover, the non-monotonic temperature dependence of NGA is rationalized, providing insight into the driving forces behind this novel phenomenon.

3

Computational Sampling Schemes to Model the Adsorption-Induced Flexibility of Metal-Organic Frameworks

Nature laughs
at the difficulties
of integration.
Pierre-Simon Laplace (1749–1827)

As outlined in the previous Chapter, two computational tools are ideally suited towards the key ingredients needed for a description of adsorption-induced framework flexibility. Molecular dynamics (MD) simulations controlling the temperature T, pressure P and anisotropic stress $\boldsymbol{\sigma}_a$ are ideally suited to capture the flexibility of a MOF in the presence of guest molecules. On the other hand, Monte Carlo simulations are ideally suited for controlling the chemical potential μ of guest molecules in the framework, allowing for an exchange of guests between the framework and an external reservoir. In this Chapter, both of these key ingredients (shown in Figure 3.1) will be combined in several computational schemes which through their differing approaches yield complementary information. The physical basis for these schemes lies in the osmotic or $(N_{\text{host}}, \boldsymbol{\sigma}_a = \mathbf{0}, \mu, P, T)$ ensemble. Guest-induced flexibility

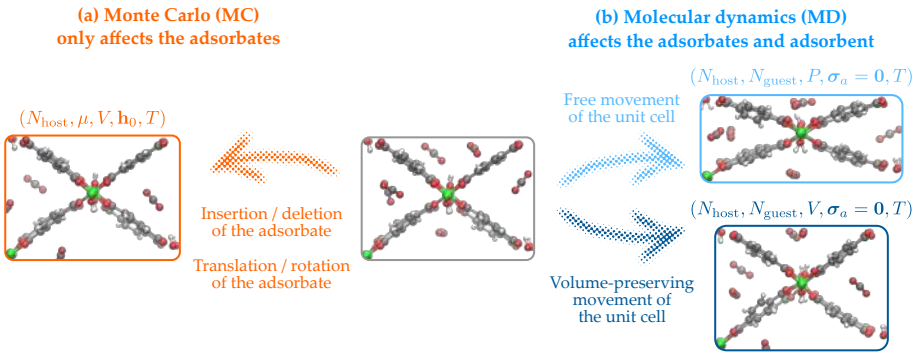


Figure 3.1: Key computational components to capture guest-induced flexibility. (a) Monte Carlo simulations control the chemical potential μ of the adsorbates, allowing for an exchange of guests with an external reservoir. (b) Molecular dynamics simulations are employed to sample the framework flexibility with guest molecules present in the framework, controlling the anisotropic stress $\boldsymbol{\sigma}_a$ and the pressure P or volume V. Figure reproduced from Ref. [208] with permission from John Wiley and Sons.

is most properly described in this ensemble.[204–207] It is characterized by the following partition function

$$Z_{\text{os}}(N_{\text{host}}, \boldsymbol{\sigma}_a = \mathbf{0}, \mu, P, T) = \tag{3.1}$$
$$\sum_{N_g=0}^{+\infty} \frac{\beta P e^{\beta \mu N_g}}{\Lambda^{3N_g}} \int\int \mathrm{d}V \mathrm{d}\mathbf{h}_0 \; e^{-\beta PV} \; \delta(\det(\mathbf{h}_0) - 1) \int\int \mathrm{d}\boldsymbol{r} \mathrm{d}\boldsymbol{R} e^{-\beta \mathcal{U}(\boldsymbol{r}, \boldsymbol{R})}$$

with N_g the number of guest molecules, $\mathbf{h}_0$ the cell shape, V the volume, $\boldsymbol{r}$ the coordinates of the host atoms and $\boldsymbol{R}$ the coordinates of the guest atoms. While this partition function appears complex at first sight, it can be simplified in terms of partition functions of the previously discussed isobaric-isothermal and grand canonical ensembles. Making use of Equation (2.52) and the following relation between the grand canonical potential Ω and the chemical potential μ[191, 209]

$$\left(\frac{\partial \Omega}{\partial \mu}\right)_{V,T} = -N_g \tag{3.2}$$

the osmotic potential associated with this ensemble can be written as[206]

$$\Omega_{\text{os}}(N_{\text{host}}, \boldsymbol{\sigma}_a = \mathbf{0}, P = p, T) = F(N_{\text{host}}, T) + PV - \int_{-\infty}^{\mu(P,T)} N_g(N_{\text{host}}, \mu', \boldsymbol{\sigma}_a = \mathbf{0}, T)\, \mathrm{d}\mu' \qquad (3.3)$$

with $F(N_{\text{host}}, T)$ the Helmholtz free energy of the empty framework. In this formulation, the mechanical pressure P exerted on the system is assumed to be equal to the gas pressure p, and the chemical potential is assumed to be a function of the pressure and temperature. While differing mechanical and gas pressures are possible, these situations are not considered further. From this osmotic potential, the relative stabilities of phases can be determined as a function of the applied gas pressure and the temperature. In Section 3.1, sampling schemes are developed and benchmarked to directly or indirectly (through Equation (3.3)) determine the stability of MIL-53(Al) under noble gas, methane and carbon dioxide adsorption. In Section 3.2, a newly proposed scheme is applied to methane adsorption in DUT-49, where it will be used to accurately predict the observed phase transitions under both adsorption and desorption as well as the observed hysteresis.

3.1 Benchmarking of sampling schemes on MIL-53(Al)

Our development of computational schemes modeling guest-induced flexibility is based on the observation that the volume V plays a vital role in distinguishing between the different phases in the majority of flexible MOFs, such as MIL-53(Al), MIL-88 and DUT-49.[71–73, 75] The volume can therefore be used as a *collective variable* or *reaction coordinate* along which phase transitions occur. Based on this, four computational schemes are developed to probe guest-induced flexibility. A schematic overview of each scheme is given in Figure 3.2. The used methodologies and obtained results are discussed in the following Sections. The MIL-53(Al) framework was modeled using a 1×2×1 cell containing 152 atoms, obtained by doubling the unit cell along the inorganic chain. The flexible force field was generated from periodic PBE-D3(BJ) data following the QuickFF procedure, accurately reproducing the underlying free energy of the empty framework. [97, 150] As discussed in Chapter 1, the empty MIL-53(Al) framework can undergo phase transitions between an orthorhombic large-pore (lp) and monoclinic closed-pore (cp) phase under the influence of temperature or pressure.[69, 70] These two phases are associated with a unit cell volume of approximately 1450 Å^3

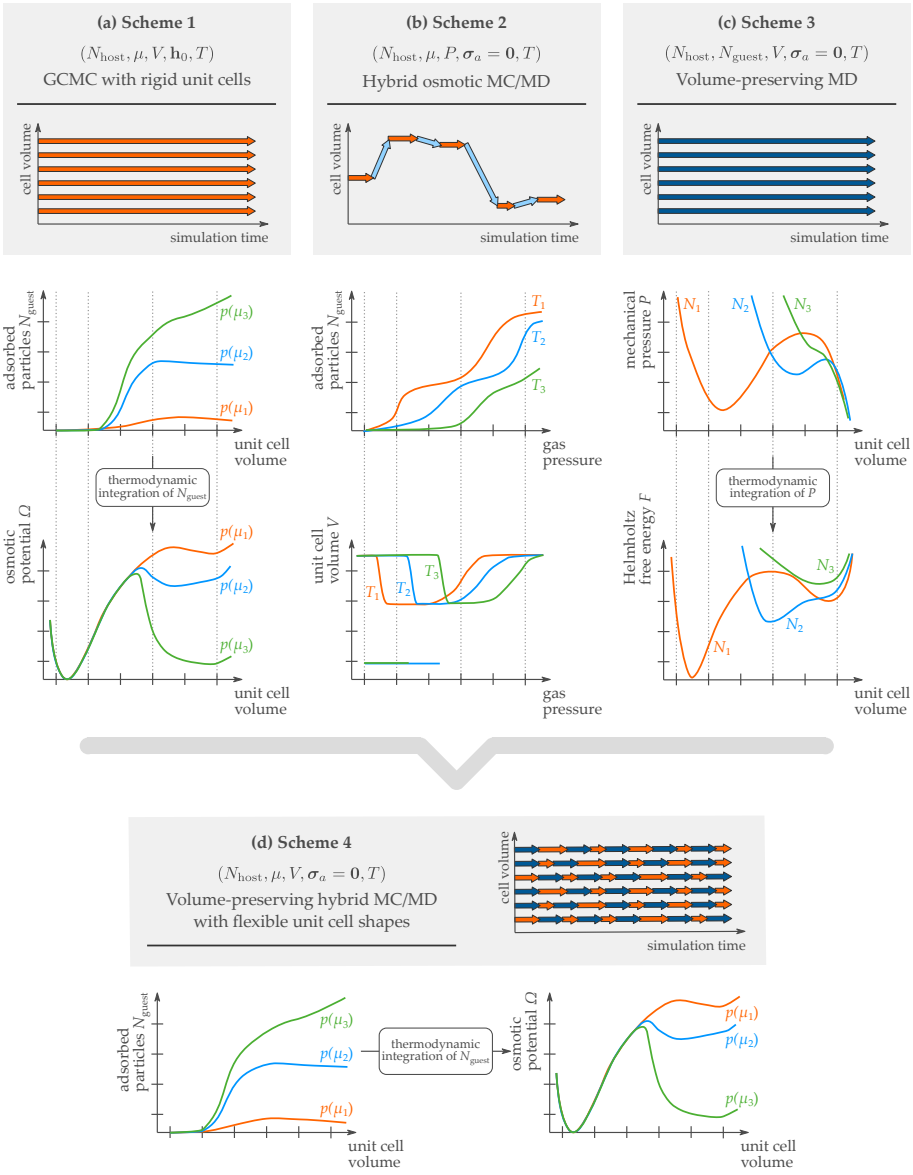


Figure 3.2: Various computational schemes to model the guest-induced flexibility of MOFs. The fourth scheme, proposed in this work, combines the desirable properties of the previous schemes. Figure reproduced from Ref. [208] with permission from John Wiley and Sons.

and 820 Å^3, respectively. Moreover, an intermediate narrow pore (np) phase can be reached under methane, carbon dioxide or xenon adsorption. Under neon or argon adsorption, no intermediate np phase is observed. Each of the investigated sampling schemes are therefore evaluated on their ability to reproduce these observations, as well the insight provided by each on the driving forces triggering guest-induced phase transitions.

3.1.1 Scheme I: Rigid-host GCMC in the restricted osmotic ensemble

To determine the phase stability of the framework along the collective variable, the osmotic potential is required as a function of the volume. The first scheme therefore consists of performing GCMC simulations for a range of framework volumes V between 850 Å^3 and 1500 Å^3 with a step size of 25 Å^3. As each individual GCMC simulation restricts the volume, the restricted osmotic ensemble is sampled.[206, 207, 210] To obtain these framework structures, an MD simulation in the $(N_{\text{host}}, \boldsymbol{\sigma}_a = \mathbf{0}, P, T)$ ensemble of the empty framework is performed at sufficiently high pressure to initiate an lp–to–cp transition. Subsequently, GCMC simulations are performed for each framework volume and a range of gas pressures. The chemical potential μ as a function of the gas pressure p and temperature T is obtained from the van der Waals equation of state of all of the considered guest species. As the cell shape $\mathbf{h}_0$ remains fixed during each GCMC calculation, the simulations give rise to what will be referred to as rigid-host adsorption isotherms $N_{\text{g}}(N_{\text{host}}, \mu, \mathbf{h}_0, T; V)$. By use of Equation (3.3), the rigid-host osmotic potential as a function of the gas pressure p, temperature T and framework volume V is obtained as follows

$$\begin{aligned}\Omega_{\text{os}}(N_{\text{host}}, \mathbf{h}_0, p, T; V) = F(N_{\text{host}}, T; V) + pV \\ - \int_{-\infty}^{\mu(p,T)} N_g(N_{\text{host}}, \mu', \mathbf{h}_0, T; V)\, \mathrm{d}\mu' \end{aligned} \tag{3.4}$$

The free energy of the empty host can be calculated from thermodynamic integration of the pressure–versus–volume equation of state $P(N_{\text{host}}, T; V)$. For this, the internal pressure was obtained from MD simulations in the $(N_{\text{host}}, \boldsymbol{\sigma}_a = \mathbf{0}, V, T)$ ensemble for a set of framework volumes, as described before. [147] Similarly, the integration of the isotherms N_g with respect to the chemical potential μ is performed numerically at a temperature of 300 K. The results, shown in Figure 3.3, are averaged over five independent GCMC runs (initialized from different framework snapshots). For both xenon, methane and carbon dioxide, the uptake in the framework at low to intermediate

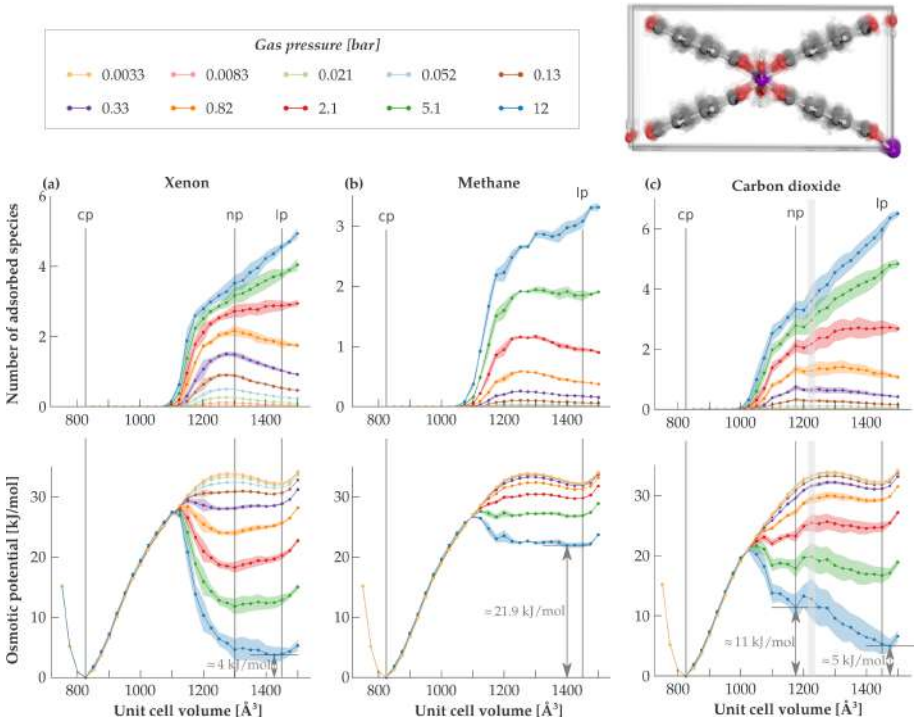


Figure 3.3: Scheme I – Uptake and osmotic potential with 1σ interval of (a) xenon, (b) methane and (c) carbon dioxide adsorption in MIL-53(Al) for a range of gas pressures at 300 K. Arrows denote the relative osmotic potential of the lp and/or np phase w.r.t. the cp phase at a gas pressure of 12 bar. For each of the adsorbates, the encountered cp, lp and/or np phase are annotated. The variation of framework geometries at a fixed volume V = 1225 Å^3 obtained from an MD simulation in the $(N_{\text{host}}, \boldsymbol{\sigma}_a = \mathbf{0}, V, T)$ ensemble is shown in the top right. Figure adapted from Ref. [208] with permission from John Wiley and Sons.

gas pressures is favored at volumes between 1200 Å^3 and 1350 Å^3 compared to the lp phase at approximately 1450 Å^3, due to the stronger interaction with the framework linkers. As seen from the osmotic potential, this leads to the stabilization of an np phase for xenon and carbon dioxide at these intermediate volumes. For methane, this never leads to a np phase which is more stable than the lp phase, although the osmotic potential is nearly flat at the highest gas pressure. As can be seen from the isotherms, further increasing the gas pressure favors the lp phase due to the larger pore size. The (empty) cp phase remains the most stable at gas pressures below 12 bar,

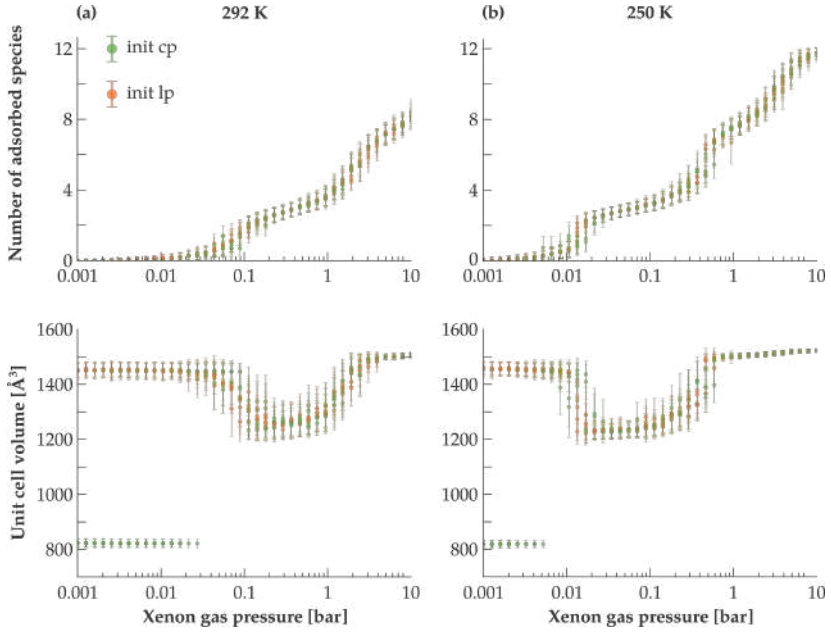


Figure 3.4: Scheme II – Xenon uptake as a function of the gas pressure at (a) 292 K and (b) 250 K, combined with the associated unit cell volumes. Ten independent simulations were performed at each pressure, with five each initialized in the cp (green bars) and lp phase (orange bars).

but an osmotic barrier remains between this phase and the lp (or np) phase exists, which could inhibit a transition to the cp phase. The stabilization of an np phase at intermediate xenon and carbon dioxide pressures at 300 K (and lack of one for methane) is in agreement with experiments. However, a quantitative prediction of the relative stabilities is hampered by the large 1σ. This uncertainty does not arise from a lack of convergence of the individual GCMC simulations. Instead, it arises from the use of five different framework snapshots at each volume. As shown on the top right of Figure 3.3 (obtained from an MD simulation in the $(N_{\text{host}}, \boldsymbol{\sigma}_a = \mathbf{0}, V, T)$ ensemble), even at a fixed volume a significant flexibility both of the BDC ligands and the cell shape remains, affecting the resulting uptake. To properly account for this, a coupling between the GCMC and MD runs is required.

3.1.2 Scheme II: Hybrid MC/MD in the osmotic ensemble

To include the flexibility of the framework in adsorption calculations, a direct sampling of the osmotic ensemble can be achieved by introducing an

additional trial move consisting of an MD run in the $(N_{\text{host}}, N_g, \boldsymbol{\sigma}_a = \mathbf{0}, P, T)$ ensemble to the GCMC simulations. The theoretical foundations for such a hybrid Monte Carlo method were first laid by Duane *et al.* in 1987 and later widely applied to Lennard-Jones fluids, silica, path integral methods, the breathing of MIL-53(Cr) under carbon dioxide adsorption and many more.[203, 211–214] From Equation (2.55), the acceptance probability of the MD trial move is as follows[191]

$$\text{acc}(\{\boldsymbol{r}, \mathbf{h}\} \rightarrow \{\boldsymbol{r}', \mathbf{h}'\}) = \min\left[1, \left(\frac{\det(\mathbf{h}')}{\det(\mathbf{h})}\right)^{N-2} e^{-\beta[\Delta\mathcal{U} + P(\det(\mathbf{h}') - \det(\mathbf{h}))]}\right] \tag{3.5}$$

with $\det(\mathbf{h}')$ and $\det(\mathbf{h})$ the final and initial volume and $\Delta\mathcal{U}$ the change in potential energy. Note that for this acceptance rule to satisfy the detailed balance condition, symmetry of the underlying Markov chain is assumed, necessitating the use of a reversible and measure-preserving integrator for the MD trajectories.[191] However, non-measure-preserving integrators can also be used by addition of the Jacobian in the acceptance rule of Equation (3.5).[200] The MD trial move was attempted with a frequency 400 times lower than the other GCMC moves with a time step of 0.5 fs and a total length of 200 fs. These settings were determined to be optimal for convergence of a simulation in the lp phase to transition to the intermediate np phase. With the possible hurdle of ergodicity in mind, five simulations were initialized in both the lp and cp phase for every combination of adsorbate, temperature and gas pressure. The results for xenon are shown in Figure 3.4. At 292 K, two steps in the isotherm are observed. For the lowest pressures, up to about 0.03 bar, the xenon uptake is negligible. For these lowest gas pressures, the phase observed during the simulation depends on the initialization: simulations initialized in the lp or cp state remain in the lp or cp state, respectively. While, in theory, phase transitions between the lp and cp phases could be observed, these transitions are rare events that are unlikely to occur as both phases are separated by a high osmotic energy barrier. When increasing the gas pressure above 0.03 bar at 292 K, a first step in the adsorption isotherm is apparent, as xenon starts to adsorb in the material. These adsorbates steer the framework from the cp or lp phase to the np phase with a volume of about 1250–1300 Å^3. This observation is in excellent agreement with the xenon-loaded np volume extracted from Scheme I (Figure 3.3). When further increasing the gas pressure, the uptake increases continuously to about four xenon atoms per unit cell, until reaching a pressure of 1–2 bar. At this pressure, a second step in the isotherm is observed, corresponding to a np–to–lp transition. Decreasing the temperature to 250 K, the predicted isotherms and unit cell volumes as a function of the gas pressure retain their 292 K shapes. Decreasing the temperature only results in a decrease in the critical

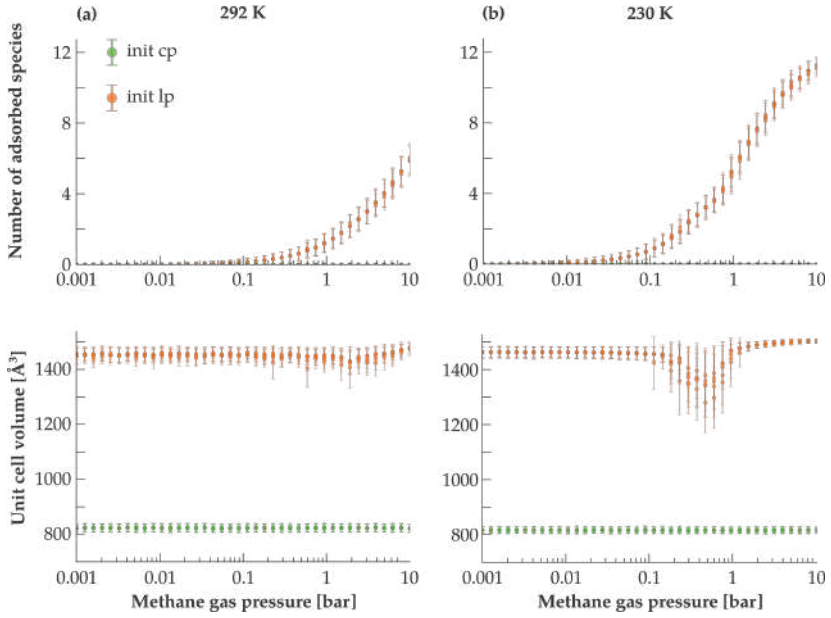


Figure 3.5: Scheme II – Methane uptake as a function of the gas pressure at (a) 292 K and (b) 230 K, combined with the associated unit cell volumes. Ten independent simulations were performed at each pressure, with five each initialized in the cp (green bars) and lp phase (orange bars).

pressures necessary for the cp/lp–to–np and the np–to–lp transitions, an observation which was also made by performing the simulations of Scheme I at this lower temperature. For the other noble gases (neon and argon), no phase transitions were observed at any temperature or pressure, consistent with experiments, and rationalized from their lower adsorption energy and thus uptake in the framework (see Supporting information of **Paper I**).

The adsorption isotherms and unit cell volumes under methane adsorption are shown in Figure 3.5. At 292 K a continuous adsorption isotherm is obtained without the presence of an intermediate np phase, in agreement with the literature.[61, 72] However, at pressures between 0.2 and 2 bar, simulations already start exploring smaller volumes during simulations. Lowering the temperature to 230 K, a similar (although less pronounced) stepwise behavior is encountered as for xenon. Inspecting the variation of the volume of the framework at intermediate gas pressures around 0.2–1 bar, spontaneous phase transitions back and forth between the np and lp phases are observed. This indicates that the np and lp phases are (meta)stable states with a very small osmotic barrier separating them, impeding an accurate determination

of the volumes and relative stability of these phases. For carbon dioxide (see Supporting information of **Paper I**), two-step isotherms akin to the ones observed for xenon are obtained, in agreement with Scheme I.

In conclusion, the hybrid MC/MD protocol of Scheme II yields direct access to the adsorption isotherms and (meta)stable MIL-53(Al) states at a given gas pressure, in qualitative agreement with experiments. However, a quantitative interpretation of these results is impeded due to two phenomena. First, premature phase transitions may occur if both phases are separated by very low barriers, such as for the lp–to–np transition under methane adsorption, making it difficult to exactly pinpoint the critical gas pressure necessary to induce this transition. Second, the phase space associated with the unit cell volume may not be sampled adequately if the (meta)stable states are separated by large osmotic barriers, such as for the cp–to–np transition under methane adsorption. In this case, the initial state of the simulation affects the simulation results, such that multiple independent simulations need to be performed, reducing the efficiency of the procedure. A similar limitation was encountered when studying the adsorption induced breathing of MIL-53(Cr).[214] Remedies for this limitation of a direct sampling in the osmotic ensemble exist. In replica exchange Monte Carlo, multiple replicas differing in one thermodynamic parameter (usually temperature, or alternatively chemical potential) are simulated. Coupling these systems through an exchange of coordinates (and cell parameters) with associated acceptance probability could overcome the osmotic free energy barriers present in the system.[191] Alternatively, a Wang–Landau based Monte Carlo sampling was applied to MIL-53(Al) by Bousquet *et al.* to efficiently sample the density of states in the (V, N_g) space, and with it calculate the osmotic potential.[191, 215] Although successful, such simulations require several orders of magnitude more iterations, limiting their use for more complex systems.

3.1.3 Scheme III: Volume-constrained guest-loaded MD

Instead of performing GCMC simulations to control the chemical potential μ, guest-loaded MD simulations in the $(N_{\text{host}}, N_g, \boldsymbol{\sigma}_a = \mathbf{0}, V, T)$ ensemble keeping the volume fixed can be performed with a discrete number of adsorbates present in the pore, constraining the volume. Subsequently, the Helmholtz free energy can be obtained from thermodynamic integration of the internal pressure

$$F(N_{\text{host}}, N_g, T; V) - F(N_{\text{host}}, N_g, T; V_{\text{ref}}) = -\int_{V_{\text{ref}}}^{V} P(N_{\text{host}}, N_g, T; V')\,\mathrm{d}V' \tag{3.6}$$

This procedure, originally employed for investigating pressure-induced phase transitions in empty MOFs, was previously extended to include guest loadings $N_g > 0$.[147, 150] From this, different (meta)stable states can be identified as a function of the guest loading. Furthermore, the osmotic potential can be obtained by means of a Legendre transform, as outlined before.[96] The resulting (meta)stability of the different phases obtained by Vanduyfhuys *et al.* are in agreement with the results from Scheme I and II. However, due to the control variable being a discrete number of adsorbates, more expensive supercell simulations would be required to probe fractional loadings. Additionally, the application of this scheme could be hindered by the presence of distinct pores in a framework, as adsorbates need to be distributed among them in an appropriate way.

3.1.4 Scheme IV: Flexible-host MC/MD in the restricted osmotic ensemble

In all of the previously investigated computational schemes, different hindrances arise, limiting their use. In summary, a computational scheme is desired which controls the chemical potential μ and samples the flexible modes of the framework under adsorption (also those orthogonal to the volume V), while not suffering from the lack of observed phase transitions due to the presence of large osmotic barriers. As seen from Scheme I and Scheme III, explicitly scanning along the collective variable V allows for a determination of both the stable, as well as metastable phases. With this in mind, starting from the procedure of Scheme I, MD trajectories are employed as an additional Monte Carlo move. However, contrary to Scheme II, the MD trajectories sample the $(N_{\text{host}}, N_g, \boldsymbol{\sigma}_a = \mathbf{0}, V, T)$ ensemble from Scheme III. Due to the proper sampling of both the positions of the framework and the cell shape, the artificial dependence of the results on the initial atomic positions and cell shape of the framework is eliminated, as was present in Scheme I. Additionally, the framework can adapt to the presence of guests. This hybrid MC/MD scheme therefore results in flexible-host isotherms $N_{\text{g}}(N_{\text{host}}, \mu, \mathbf{h}_0, T; V)$, contrary to the rigid-host isotherms of Scheme I. These isotherms are then transformed to the flexible-host osmotic potential

$$\begin{aligned}\Omega_{\text{os}}(N_{\text{host}}, \mathbf{h}_0, p, T; V) = {} & F(N_{\text{host}}, T; V) + pV \\ & - \int_{-\infty}^{\mu(p,T)} N_g(N_{\text{host}}, \mu', \mathbf{h}_0, T; V)\, \mathrm{d}\mu' \end{aligned} \tag{3.7}$$

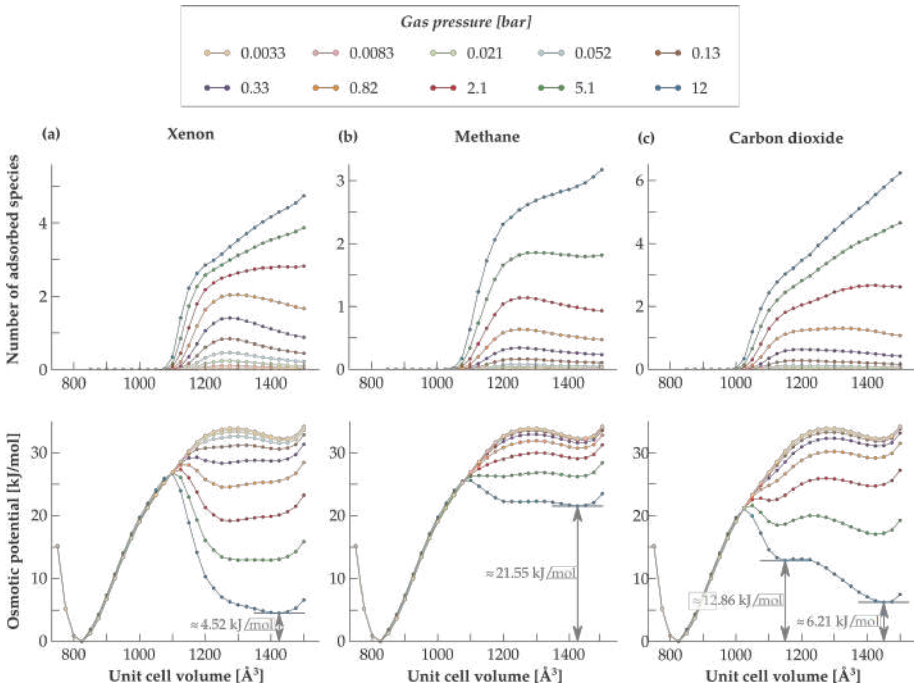


Figure 3.6: Scheme IV – Uptake and osmotic potential with 1σ interval of (a) xenon, (b) methane and (c) carbon dioxide adsorption in MIL-53(Al) calculated at a range of gas pressures at 300 K. The relative osmotic potential of the lp and/or np phase w.r.t. the cp phase are denoted at the largest gas pressure of 12 bar. Figure adapted from Ref. [208] with permission from John Wiley and Sons.

To investigate the performance of the newly proposed scheme, Scheme IV is adopted to predict the adsorption isotherms and corresponding osmotic potentials for all five guests in MIL-53(Al) at 300 K. In Figure 3.6, the results are shown for the adsorbates with the potential to induce phase transitions. Compared to Scheme I (see Figure 3.3), results are obtained with a much lower uncertainty. Five independent simulations were performed for every guest, gas pressure and volume, the results of which are virtually indistinguishable. This improved accuracy not only reduces the number of independent simulations that need to be carried out, but also allows for a quantitative prediction of the evolution of the adsorption isotherms and (meta)stable states with increasing gas pressure. As seen from Figure 3.6(a), the results unambiguously demonstrate the existence of an intermediate np phase under xenon adsorption at 300 K, in agreement with experiments.[71] For methane, although the difference in relative stability between the np and lp phase are minuscule at 5.1 and 12 bar, the lp phase remains the more stable state. This

is consistent with experimental observations that breathing under methane adsorption only occurs at temperatures lower than 300 K.[72] Undoubtedly, the increased accuracy of Scheme IV is most pronounced for carbon dioxide. The rigid-host isotherms and osmotic potentials displayed in Figure 3.3(c) revealed a large dependence on the choice of the fixed framework during the GCMC procedure. As a result, it was difficult to uniquely determine for which range of pressures an np phase is induced, as well as to determine the relative stability of the various metastable phases. As a substantially lower uncertainty is associated with the flexible-host isotherms and osmotic potentials obtained in Scheme IV, these questions can now be fully answered. Figure 3.6(c) reveals that an np phase starts to appear at a carbon dioxide pressure of 2.1 bar (somewhat larger than observed experimentally[72]), while it almost disappears when increasing the pressure to 12 bar. With increasing pressure, the relative stability of the np and lp phases with respect to the cp phase increases, yielding metastable lp and np phases at 12 bar which are about 6.2 and 12.9 $kJ \cdot mol^{-1}$ less stable than the stable cp phase.

To conclude, the proposed computational scheme shows clear promise in probing the guest-induced flexibility of MOFs. As both the relative stability of different phases as well as the osmotic energy barriers between phases can be tracked as a function of the applied gas pressure, it provides a tool for capturing hysteresis effects between adsorption and desorption. Moreover, it can provide insight into phase transitions which do not occur under global thermodynamic equilibrium in which the transition to a stable state is inhibited by the presence of barriers. It is therefore perfectly positioned to tackle one of the most puzzling phenomena in adsorption, as discussed in the following Section.

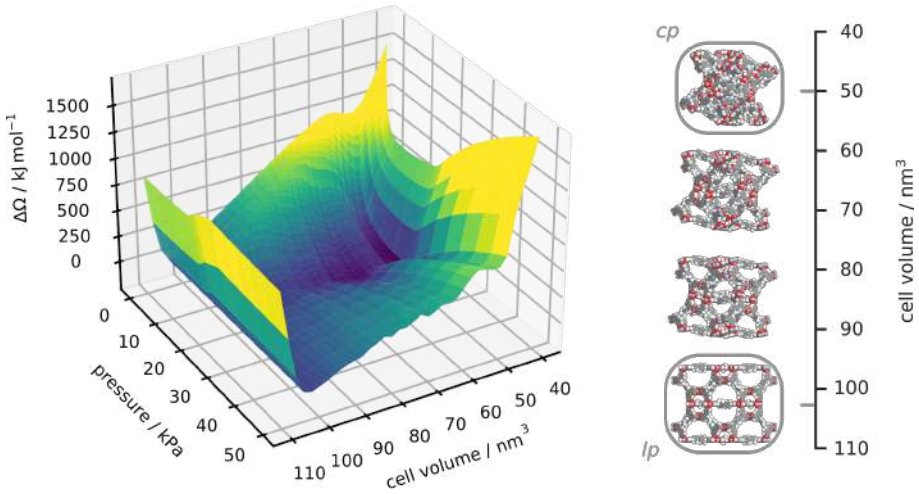


Figure 3.7: Computed osmotic potential profile of DUT-49(Cu) under methane adsorption at 120 K as a function of the unit cell volume and gas pressure. Figure adapted from Ref. [216] with permission from the American Chemical Society.

3.2 Modeling the negative gas adsorption phase transition of DUT-49(Cu)

In 2016, adsorption experiments in DUT-49(Cu) revealed a cubic–to–cubic phase transition of the framework triggered by an increase in methane gas pressure, associated with a massive contraction of the unit cell volume from 100 072 Å^3 to 47 282 Å^3. Not only is this transition remarkable for the large volume difference between phases, it is also accompanied by an expulsion of methane adsorbates out of the framework, coined negative gas adsorption (NGA).[75] For example, at a temperature of 120 K and the transition pressure of 25 kPa, 102 molecules per unit cell of the framework are released (ΔN_g = -102). The NGA transition under methane adsorption was found to only occur between 91 K and 130 K. In this temperature range, ΔN_g was observed to vary non-monotonically with temperature. Later work also demonstrated the existence of NGA transitions under ethane, ethane, propane, argon, krypton and xenon adsorption, each in a range of temperatures correlated with their critical temperature.[78] To demonstrate why the NGA transition cannot occur under global thermodynamic equilibrium,[75, 77] consider the difference

in osmotic potential between both phases

$$\Delta\Omega = \Delta F_{\text{host}} + p\Delta V - \int_{-\infty}^{\mu} \Delta N_g(\mu')\,\mathrm{d}\mu' \tag{3.8}$$

For a transition to occur from the lp to cp phase under thermodynamic equilibrium, two conditions have to be fulfilled: $\Delta\Omega = 0$ and $\frac{\partial\Delta\Omega}{\partial\mu} < 0$. Making use of Equation (3.8), the latter condition reduces to

$$\Delta V \frac{\partial p}{\partial \mu} - \Delta N_g < 0 \tag{3.9}$$

with $\rho = \left(\frac{\partial p}{\partial \mu}\right)_T > 0$ the density of the gas reservoir. However, as the ΔV term can generally be neglected (even for the large volume difference between the lp and cp phases of DUT-49(Cu), it is three orders of magnitude smaller than the adsorption term), this reduces to $\Delta N_g > 0$, precluding the possibility of an NGA transition. With this in mind, capturing the NGA transition necessitates not only the tracking of the relative stability of both phases, but also the barriers existing between these phases as a function of the applied gas pressure. Scheme IV of the previous Section is therefore used to obtain the osmotic potential as a function of the gas pressure and unit cell volume of the framework. Furthermore, simulations are performed at three temperatures as each probes a different experimentally observed behavior. At 90 K, a structural transformation occurs, although it is associated with a near zero value for ΔN_g. At 120 K, the previously described NGA transition occurs with ΔN_g = -102. Lastly, a temperature of 150 K was included, as no phase transition and thus no NGA is observed experimentally at this temperature. For a sufficiently fine-grained investigation, flexible-host isotherms were calculated for a series of unit cell volumes between 40 and 110 nm^3 with a spacing of 2 nm^3. At each temperature, between 46 and 75 gas pressures were considered, chosen manually to ensure convergence of the subsequently obtained osmotic potentials. As reported in **Paper II**, the resulting three-dimensional surface as a function of the unit cell volume and gas pressure at 120 K is shown in Figure 3.7. For clarity, the surface is normalized to the lp phase volume for each pressure. At the $p = \mu = 0$, the profile reduces to the empty host free energy profile with the lp phase as the global minimum and the cp phase as metastable state. However, at intermediate pressures, gas adsorption clearly favors the cp phase compared to the lp phase, which is the driving force behind the lp–to–cp phase transition. In order to gain insight into the NGA transition and hysteresis upon adsorption/desorption, let us now *walk* along this surface in two ways.

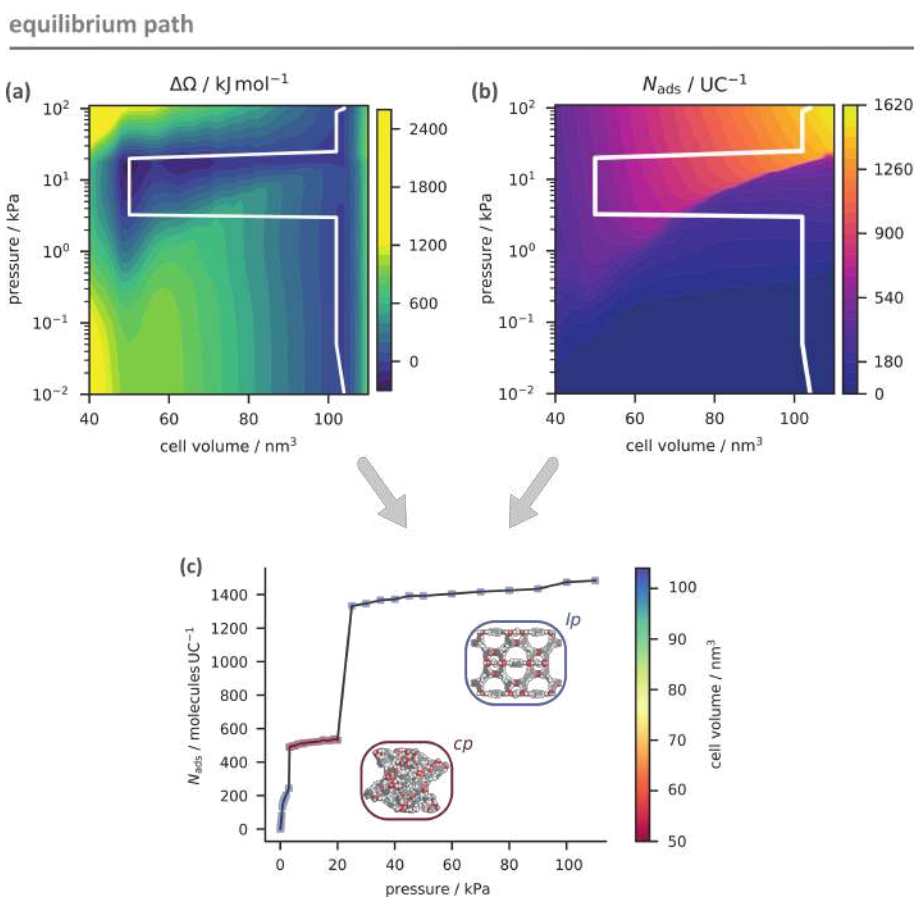


Figure 3.8: Contour plots of (a) the osmotic free energy profile and (b) the methane uptake in DUT-49(Cu) at 120 K as a function of the unit cell volume and gas pressure, together with a walk along the equilibrium phases (corresponding to the global thermodynamic equilibrium) identified with a white line. (c) The adsorption isotherm resulting from a determination of the equilibrium phases as a function of the gas pressure, together with insets of the structure of the lp and cp phases. Figure adapted from Ref. [216] with permission from the American Chemical Society.

3.2.1 The equilibrium path

First, the global thermodynamic equilibrium will be tracked as a function of the gas pressure. As discussed before, the equilibrium path cannot account for NGA, but it is nevertheless instructive to contrast this path with the *kinetic* path discussed further. To extract the equilibrium path, the osmotic

potential and gas uptake are shown in two-dimensional contour plots in Figure 3.8(a) and (b). Starting in the lp phase, the lp phase remains stable until the gas pressure is increased to 3.25 kPa. At this pressure, the thermodynamic equilibrium shifts to the cp phase. Further increasing the gas pressure to 25 kPa, the lp phase again becomes stabilized. Having determined the pressure range of stability of the lp and cp phases, the isotherms can be extracted from *walking* along the same path on the isotherm contours of Figure 3.8(b) as in Figure 3.8(a). The resulting one-dimensional isotherm is shown in Figure 3.8(c). Indeed, it is clear that both the lp–to–cp and cp–to–lp transitions are associated with a positive gas adsorption (PGA) step. Furthermore, as only the thermodynamic equilibrium is tracked as a function of the gas pressure, the same isotherm is obtained for desorption as for adsorption. Focusing back on the lp–to–cp phase transition in Figure 3.8(a), it is clear that a large osmotic free energy barrier is present between both phases which can act to prevent a transition. To take these barriers into account in our *walk* along the osmotic surface, a transition is now only allowed when the barrier separating phases has decreased sufficiently, yielding the kinetic path.

3.2.2 The kinetic path

The kinetic path is obtained from only allowing a phase transformation when the osmotic barriers separating phases has disappeared. However, as thermal fluctuations can allow the crossing of barriers, a transition was noted when the barrier was reduced to within 15 $k_{\mathrm{B}}T$. The same transitions are obtained for the 120 K surface when varying this value of the 'crossable' barrier between 0 and 25 $k_{\mathrm{B}}T$. Starting the 'adsorption walk' along the surface (full white line in Figure 3.9(a)), the system begins at equilibrium in the lp phase, where there is also a metastable cp phase (at a volume of 50 nm^3). At a pressure of 3.25 kPa, at which the cp phase becomes more stable than the lp phase, a large osmotic barrier of 591 $\mathrm{kJ \cdot mol^{-1}}$ remains, preventing a phase transition. Only when the pressure is further increased to 17.5 kPa does the barrier between phases disappear. Here, a phase transition between the metastable lp to the stable cp phase occurs, associated with a predicted NGA step of 86 molecules per unit cell (see Figure 3.9(b) and (c)). This prediction is in good agreement with the experimentally observed NGA step of 102 molecules per unit cell at a gas pressure of 25 kPa. Further increasing the gas pressure, the barrier between the cp and lp phases only disappears at 70 kPa, compared to the equilibrium transition at 25 kPa. Just as a walk across the osmotic surface can be performed towards increasing gas pressures to predict the adsorption branch, the same process can be done in reverse to track desorption. Starting at the largest gas pressure in

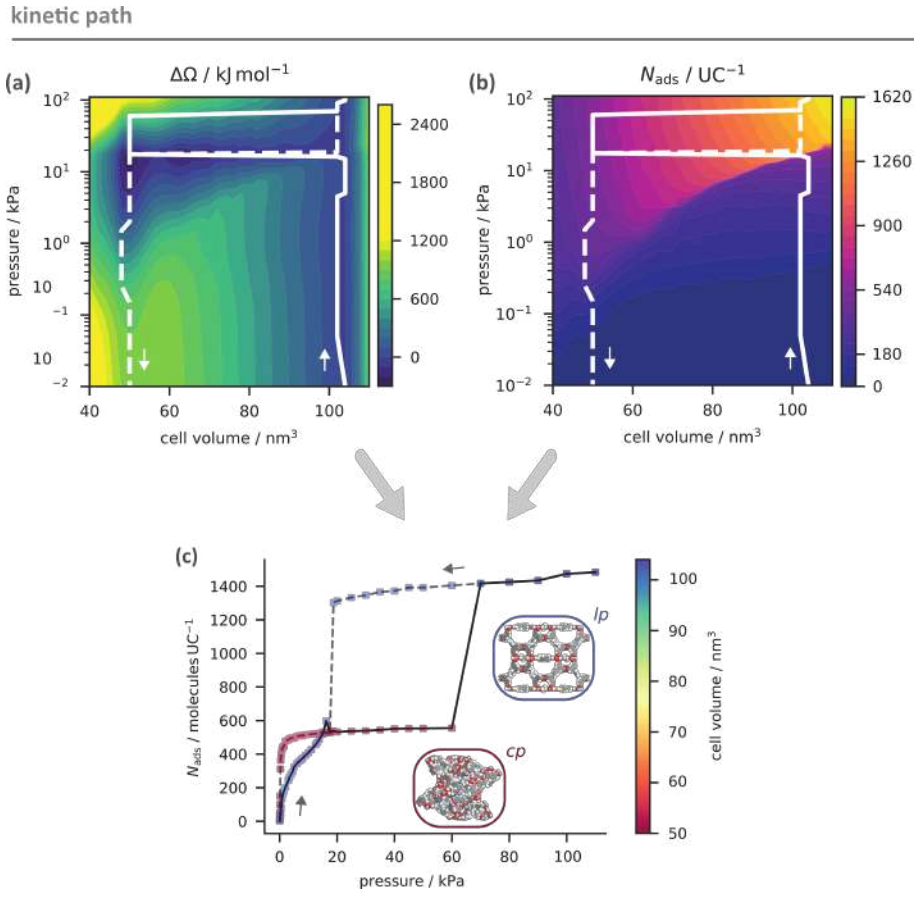


Figure 3.9: Contour plots of (a) the osmotic free energy profile and (b) the methane uptake in DUT-49(Cu) at 120 K as a function of the unit cell volume and gas pressure. The walk along the kinetically accessible phases (only allowing a transition when the osmotic barrier between phases disappears) is identified with a white line (full line for adsorption, dotted line for desorption). (c) The adsorption and desorption isotherm resulting from a determination of the kinetic path as a function of the gas pressure, together with insets of the structure of the lp and cp phases. Figure adapted from Ref. [216] with permission from the American Chemical Society.

the lp phase, the pressure is reduced until a barrier preventing the lp–to–cp transition disappears at 17.5 kPa. This prediction of an lp–to–cp transition during desorption at a gas pressure similar to that of the lp–to–cp transition during adsorption is in agreement with experiment, and can be attributed to the flatness of the osmotic landscape over the entire volume range at this pressure, seen in Figure 3.9(a). Interestingly, on decreasing the gas pressure further, it is predicted from this osmotic surface that the framework remains kinetically trapped in the metastable cp phase, exactly as observed experimentally. This is due to the barrier surrounding the cp local minimum of the osmotic potential at low gas pressures.

3.2.3 The temperature dependence of negative gas adsorption

Similar to this investigation of the osmotic landscape at 120 K, simulations were performed at 90 K and 150 K. The structural transitions derived from the kinetic pathway during adsorption and desorption at 90 K agree quantitatively with reported in situ powder X-ray diffraction experiments.[75] The presence of intermediate phases between the lp and cp phases at approximately 2 kPa are reproduced, evidencing the flatness of osmotic landscape at this gas pressure (see Supporting Information of **Paper II**). Furthermore, to investigate the non-monotonic trend of NGA with temperature,[78] the difference in osmotic potential between the lp and cp phases (solid lines) and the osmotic barrier from the lp to cp phase (dotted lines) are shown in Figure 3.10. As can be seen, the pressure window for stability of the cp phase (solid line above $\Delta\Omega = 0$) is significantly affected by the temperature. At the highest temperature of 150 K, the cp phase only becomes marginally more stable than the lp phase at a gas pressure of approximately 100 kPa. However, the barrier (dotted line) from the lp to cp phase has not disappeared at this pressure, preventing a phase transition. As no phase transition occurs, no NGA step is observed, consistent with experiments.[78] As discussed before, at 120 K, the barrier vanishes at a pressure of 17.5 kPa, allowing for a phase transition. Moreover, from Equation (3.8), the derivative of $\Delta\Omega$ with respect to the gas pressure (or equivalently, the chemical potential) is seen to be proportional with the difference in uptake between both phases. When the derivative is positive, more gas is adsorbed in the cp phase than the lp phase. Therefore, the slope of the difference in osmotic potential (solid lines in Figure 3.10) yields information of the difference in uptake between phases. At 120 K, the barrier preventing a transition (shown from the dotted line) disappears just as the slope of the solid line has turned negative, therefore resulting in negative gas adsorption. In contrast, at 90 K an lp–to–cp phase transition occurs at a gas pressure when the slope of the solid line is still positive, yielding a small

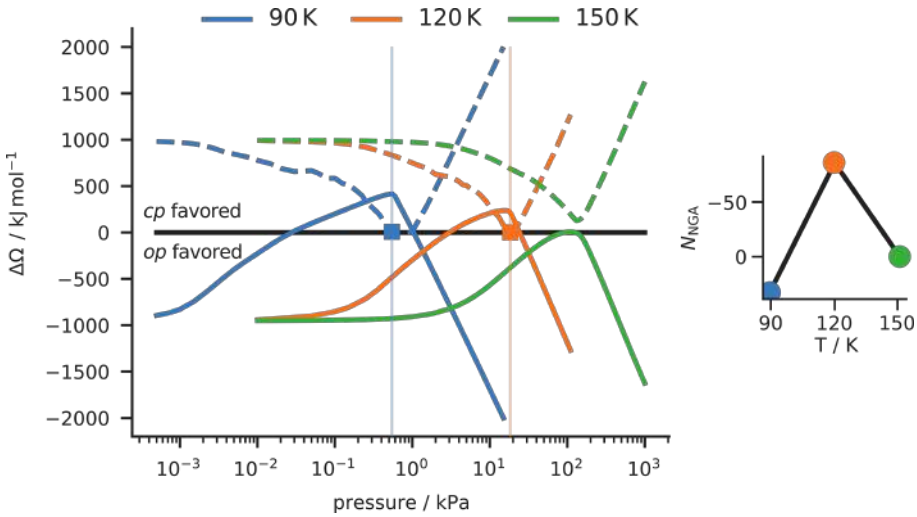


Figure 3.10: Difference in osmotic potential between the lp and cp phase (solid lines) and osmotic barrier present preventing a transition from the lp to cp phase (dotted ines) at 90 K, 120 K and 150 K. On the left, the resulting NGA step as a function of temperature is shown. Figure adapted from Ref. [216] with permission from the American Chemical Society.

PGA step, shown on the left of Figure 3.10. Although experiments still predict a small NGA step at this temperature, the obtained non-monotonic variation of the NGA step with temperature accurately reproduces the experimental observation. From this discussion it is clear that NGA is a very subtle phenomenon, as it can only occur when a metastable lp phase has been achieved through favorable gas adsorption of the other stable cp phase and the metastability of this phase remains maintained until the metastable lp phase starts to be re-favored in uptake (negative derivative of the solid line in Figure 3.10). This subtlety is likely a reason why MOFs exhibiting this phenomenon are so unique in the vast field of synthesized MOFs.

Through the benchmarking of several computation schemes to model adsorption in flexible metal-organic frameworks, a newly proposed scheme has shown to provide valuable insight into guest-induced phase transitions. By explicitly scanning the osmotic potential along the volume, insight cannot only be gained into the relative stability of different phases as a function of the applied gas pressure, but also into the osmotic barriers between phases which can prevent transitions. Guest-induced phase transitions and hysteresis between adsorption and desorption due to metastable phases can be quantitatively predicted. Moreover, the separation of the osmotic

potential into its terms (free energy of the host versus the adsorption term) yields insight into how materials could be altered to be optimized further for specific applications.

4

Non-Covalent Force Fields to Model Adsorbate–Adsorbent Interactions

Computers are fantastic.
In a few moments they can make a mistake
so great that it would take many men
many months to equal it.
Anonymous

In Chapter 3, computational sampling schemes were introduced and applied to model the adsorption-induced flexibility of two highly flexible MOFs. However, as outlined in Chapter 2, these investigations stand or fall entirely on the accuracy of used the force field description of the potential energy surface (PES). In this respect, much effort has gone into refining the covalent contribution of force fields from fitting to DFT reference data.[93–97, 143–146] For the non-covalent contributions, highly accurate force fields have been developed capturing the adsorbate–adsorbate interactions such as TraPPE, in which parameters are derived from fitting to critical temperatures and saturated liquid densities.[217] However, developing accurate non-covalent force fields for use in MOFs is more complicated. For the empty MIL-53(Al) framework, a rescaling of the MM3 ϵ parameter was needed to properly capture the relative stability of the cp and lp phases.[218] Moreover, describing the adsorbate–adsorbent or guest–host interactions has proven difficult. These contributions, although essential for a description of guest adsorption in MOFs, are usually not derived explicitly, but rather implicitly

from a combination of the individual MOF and adsorbate interaction potentials. Examples of use-cases where parameters needed tuning to improve the accuracy of the guest–host interactions are therefore plentiful in the literature. Boulanger *et al.* rescaled the general Amber force field (GAFF) van der Waals dispersion parameter by a factor of 1.115 to obtain accurate hydration free energies, Wu *et al.* rescaled the UFF LJ parameter ϵ by a factor of 0.635 to reproduce methane adsorption isotherms in ZIF-8 and similarly, Pérez-Pellitero *et al.* found that a rescaling of the LJ parameters was required for agreement with experimental adsorption isotherms of N_2 and CO_2 in ZIFs.[219–222] Moreover, non-polarizable force fields have been shown to significantly underestimate the interaction strength of adsorbates with open-metal sites present in many MOFs.[223] While more advanced polarizable force fields can be employed for these interactions, they require careful refitting to reference calculations.[98, 99, 224]

In this Chapter, we therefore introduce two (very different) pathways to an improved description of guest–host interactions. In the first, laid out in **Paper III** and Section 4.1, we aim to develop a physically inspired polarization model with a single transferable parameter. This approach builds on the Monomer Electron Density Force Field (MEDFF) which introduced four physically inspired functionals of the electron density of the constituting monomers in the system, together with 3 fitted parameters to model the non-covalent interactions. [179] In the resulting pairwise additive non-covalent force field, polarization was modeled by an overlap term, which was found to be the main limitation of the model. To improve upon this, we develop a new induced dipole polarization model based on the monomer electron density, employing one of the interaction parameters present in the MEDFF model. Performing benchmarks of the predicted molecular polarizabilities and many-body interaction energies of molecular clusters reveals an improved performance compared to commonly employed inducible dipole models. This first approach represents a branch of force field development in which physical insight is incorporated into the interaction model. Herein, the aim is to derive generally applicable models with only a limited set of fitted parameters. In the most optimistic of cases, this could lead to interaction models which transfer well to diverse systems to which parameters were not explicitly fitted.

In a second methodology, investigated in **Paper IV** and in Section 4.2, the polar opposite of this approach is taken. Instead of building as much (hopefully accurate) physical insight into a model as possible, one could imagine not imposing any *a priori* interaction potential. Instead, the potential is learned through training to reference energies (and/or forces, stresses) obtained with DFT or wavefunction-based methods. This approach, in stark contrast to

the first, is inspired by the recent development of highly accurate and data-efficient Machine Learning Potentials (MLPs). Whether MLPs can be trained on a small set of intermolecular interactions with sufficiently high accuracy for the envisaged adsorption applications will be explored in Section 4.2.

After both methodologies are explored in the following Sections, conclusions are drawn about the merit of both approaches. A short discussion is provided about further promising developments and applications to capture the adsorbate–adsorbent interactions on which an accurate investigation of the adsorption properties of nanoporous materials relies.

4.1 Development of a transferable polarizable force field

To model polarization with inducible dipoles, first a description of the *frozen* electron density of the constituting interacting fragments (e.g. the framework and adsorbates) of a system is required. For this, the electron density of each fragment is calculated and subsequently partitioned into atomic fragments with the Minimal Basis Iterative Stockholder (MBIS) scheme.[166] Herein, a core (point) charge q_A^c and valence Slater function is assigned to each atom A:

$$\rho_A^0(\boldsymbol{r}) = \frac{N_A}{8\pi {\sigma_A}^3} \exp\left(-\frac{|\boldsymbol{r} - \boldsymbol{R}_A|}{\sigma_A}\right) + q_A^{\mathrm{c}}\, \delta(\boldsymbol{r} - \boldsymbol{R}_A) \tag{4.1}$$

with N_A the population and σ_A the width of the Slater function. With this frozen electron density for each fragment of the system, the electrostatic, dispersion and exchange interaction can be calculated analytically through approximate interaction potentials.[179] However, when these isolated fragments approach one another, the electron density should be allowed to deform under their mutual interaction, giving rise to an additional stabilizing interaction. To demonstrate the physical basis for this, our polarization model is first grounded in density-functional theory.

4.1.1 Grounding in density-functional theory

To understand the approximations made in inducible dipole models common in the literature, it is essential to ground such models in a solid theoretical foundation like density-functional theory (DFT). The following derivations build upon the work of York *et al.* and Tabacchi *et al.*[225–227] Consider a system of N interacting subsystems, for example a set of molecules or a molecule in an adsorbent, each described by a nuclear charge distribution $\rho_i^n(\boldsymbol{r})$ and

ground state electron density $\rho_i^0(\boldsymbol{r})$ of the isolated subsystem. The total electron density can then be defined as a sum of the reference densities and the density response induced due to the interaction of the subsystems

$$\rho(\boldsymbol{r}) = \rho^0(\boldsymbol{r}) + \delta\rho(\boldsymbol{r}) \text{ with } \begin{cases} \rho^0(\boldsymbol{r}) = \sum_i \rho_i^0(\boldsymbol{r}) \\ \delta\rho(\boldsymbol{r}) = \sum_i \delta\rho_i(\boldsymbol{r}) \end{cases} \tag{4.2}$$

This system can be characterized by the energy functional of the total system (see Section 2.1.2)

$$E[\rho, v] = F[\rho] + \int \rho(\boldsymbol{r}) v(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r} \tag{4.3}$$

with $v(\boldsymbol{r})$ the external potential generated by the nuclei and $F[\rho]$ the universal functional. Substituting Equation (4.2) in the energy functional and expanding to second order in changes of the density and external potential yields

$$\begin{aligned} E[\rho^0 + \delta\rho, v^0 + \delta v] = E[\rho^0, v^0] &+ \int \frac{\delta E}{\delta\rho(\boldsymbol{r})} \delta\rho(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r} + \int \frac{\delta E}{\delta v(\boldsymbol{r})} \delta v(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r} \\ &+ \frac{1}{2} \int\int \delta\rho(\boldsymbol{r}) \left(\frac{\delta^2 E}{\delta\rho(\boldsymbol{r}) \delta\rho(\boldsymbol{r}')} \right) \delta\rho(\boldsymbol{r}') \, \mathrm{d}\boldsymbol{r} \mathrm{d}\boldsymbol{r}' \\ &+ \int\int \delta\rho(\boldsymbol{r}) \left(\frac{\delta^2 E}{\delta\rho(\boldsymbol{r}) \delta v(\boldsymbol{r}')} \right) \delta v(\boldsymbol{r}') \, \mathrm{d}\boldsymbol{r} \mathrm{d}\boldsymbol{r}' \\ &+ \frac{1}{2} \int\int \delta v(\boldsymbol{r}) \left(\frac{\delta^2 E}{\delta v(\boldsymbol{r}) \delta v(\boldsymbol{r}')} \right) \delta v(\boldsymbol{r}') \, \mathrm{d}\boldsymbol{r} \mathrm{d}\boldsymbol{r}' \end{aligned} \tag{4.4}$$

Performing the functional derivatives and making use of Equation (4.2) and (4.3), the following is obtained (see Ref. [226]):

$$\begin{aligned} E[\rho^0 + \delta\rho, v^0 + \delta v] = E[\rho^0, v^0] &+ \int \rho^0(\boldsymbol{r}) \delta v(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r} + \int \delta\rho(\boldsymbol{r}) v(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r} \\ &+ \sum_k \int \frac{\delta F}{\delta\rho(\boldsymbol{r})} \delta\rho_k(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r} \\ &+ \frac{1}{2} \sum_{k,l} \int\int \delta\rho_k(\boldsymbol{r}) \left(\frac{\delta^2 F}{\delta\rho(\boldsymbol{r}) \delta\rho(\boldsymbol{r}')} \right) \delta\rho_l(\boldsymbol{r}') \, \mathrm{d}\boldsymbol{r} \mathrm{d}\boldsymbol{r}' \end{aligned} \tag{4.5}$$

From this, all terms which depend on a change in density $\delta\rho$ can be gathered into a term called the polarization energy E_{PFF}. Furthermore, the external potential $v(\boldsymbol{r})$ is assumed to be generated by the nuclear charges $\rho^n(\boldsymbol{r})$ (no external fields are present)

$$v(\boldsymbol{r}) = \int \frac{\rho^n(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \, \mathrm{d}\boldsymbol{r}' \text{ with } \rho^n(\boldsymbol{r}) = \sum_k \rho_k^n(\boldsymbol{r}) \tag{4.6}$$

with this, E_{PFF} can be written as

$$E_{\text{PFF}} = \sum_k \int \left[\int \frac{\rho^n(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \, \mathrm{d}\boldsymbol{r}' + \frac{\delta F}{\delta \rho(\boldsymbol{r})} \right] \delta\rho_k(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r} + \frac{1}{2} \sum_{k,l} \int\int \delta\rho_k(\boldsymbol{r}) \left(\frac{\delta^2 F}{\delta\rho(\boldsymbol{r})\delta\rho(\boldsymbol{r}')} \right) \delta\rho_l(\boldsymbol{r}') \, \mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{r}' \tag{4.7}$$

This Equation can be seen as a starting point for the developments of inducible dipole models. Within a Kohn-Sham formulation of DFT, the universal functional $F[\rho]$ is decomposed as $F[\rho] = T_S[\rho] + V_H[\rho] + E_{xc}[\rho]$. Only the functional derivatives of the Hartree term are known exactly

$$\frac{\delta V_H[\rho]}{\delta\rho(\boldsymbol{r})} = \int \frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \, \mathrm{d}\boldsymbol{r}', \quad \frac{\delta^2 V_H[\rho]}{\delta\rho(\boldsymbol{r})\delta\rho(\boldsymbol{r}')} = \frac{1}{|\boldsymbol{r} - \boldsymbol{r}'|} \tag{4.8}$$

Therefore, classical inducible dipole models only retain this term ($F[\rho] \approx V_H[\rho]$). The density response is now expanded in a dipole basis set centered at each atom a, with $i \in \{x, y, z\}$:

$$\delta\rho(\boldsymbol{r}) = \sum_a \sum_{i=x,y,z} c_{a,i} d_a^i(\boldsymbol{r}) \tag{4.9}$$

Due to the choice of dipole basis set, the diagonal elements of the second functional derivative of $F[\rho]$ with respect to the density response can be seen to be equal to the inverse of the dipole polarizability.[228] Assuming isotropic atomic polarizabilities, Equation (4.7) reduces to

$$E_{\text{PFF,cl}} = \sum_a \sum_i c_{a,i} \int\int \frac{\rho^0_{\text{tot}}(\boldsymbol{r}') d_a^i(\boldsymbol{r})}{|\boldsymbol{r} - \boldsymbol{r}'|} \, \mathrm{d}\boldsymbol{r} \, \mathrm{d}\boldsymbol{r}' + \frac{1}{2} \sum_{a \neq b} \sum_{i,j} c_{a,i} c_{b,j} \int\int \frac{d_a^i(\boldsymbol{r}) d_b^j(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \, \mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{r}' + \frac{1}{2} \sum_{a,i} \frac{c_{a,i}^2}{\alpha_a} \tag{4.10}$$

with α_a the atomic polarizability of atom a and ρ^0_{tot} the sum of the nuclear charges ρ^n and the reference electron density ρ^0. Herein, the first term represents the classical interaction of the total charge density with a dipole on atom a, with the second term the classical interaction between two dipoles located at atoms a and b. Lastly, the third term represents the energy required to generate the dipoles. Minimizing $E_{\text{PFF,cl}}$ with respect the magnitude of the induced dipoles $c_{a,i}$, the inducible dipole model as discussed

in Section 2.1.3 is obtained (not considering the usually applied damping functions[170] and scaling factors for interactions between atoms separated by up to 4 bonds[229]). However, several approximations were introduced. First, the interactions are assumed to be purely Coulombic, neglecting the Pauli exclusion principle. Additionally, the density response is assumed to be captured purely by a dipole basis set. Furthermore, the functional form of the dipole basis set remains to be chosen. Most commonly, point dipoles or Gaussian dipoles are employed. In the following Section, improvements of both the interaction model and the basis functions are therefore investigated and benchmarked on their ability to reproduce molecular polarizabilities and many-body energies.

4.1.2 The inducible dipole model

As discussed in the previous Section, inducible dipole models are built up from essentially three components. The first is a determination of the atomic polarizabilities. The second is a choice of dipole basis set, and the last is the functional form of the interaction kernel through which dipoles interact mutually and with the ground state density. Each of these components are therefore discussed separately.

Atomic polarizabilities

Atomic polarizabilities are usually derived for a limited set of atom types from fitting to the resulting experimental[230] or *ab initio*[173] derived molecular polarizability tensor. Our approach differs from this, as the intent is to derive all model parameters except one interaction parameter from a single monomer calculation. Therefore, isolated free atom polarizabilities obtained from linear response time-dependent density functional theory are used.[231] To account for the chemical environment of the atom embedded in a molecule, the isotropic polarizability of each atom $\alpha_{A,\text{free}}$ is subsequently rescaled proportionally to the atom–in–molecule effective volume (V_{aim})

$$\alpha_A = \frac{V_{\text{aim}}}{V_{\text{free}}}\,\alpha_{A,\text{free}} = \frac{\int r^3 \rho_A(\boldsymbol{r})\; d\boldsymbol{r}}{\int r^3 \rho_{A,\text{free}}(\boldsymbol{r})\; d\boldsymbol{r}}\,\alpha_{A,\text{free}} \tag{4.11}$$

with V_{free} the effective volume of the free atom in vacuum. This rescaling of polarizabilities is inspired by the empirically established linear correlation between the polarizability and volume, as also employed in the exchange-hole dipole moment (XDM) and Tkatchenko-Scheffler dispersion models.[232–234] Herein, it is calculated from an MBIS partitioning of the monomer electron density into atomic fragments.[166]

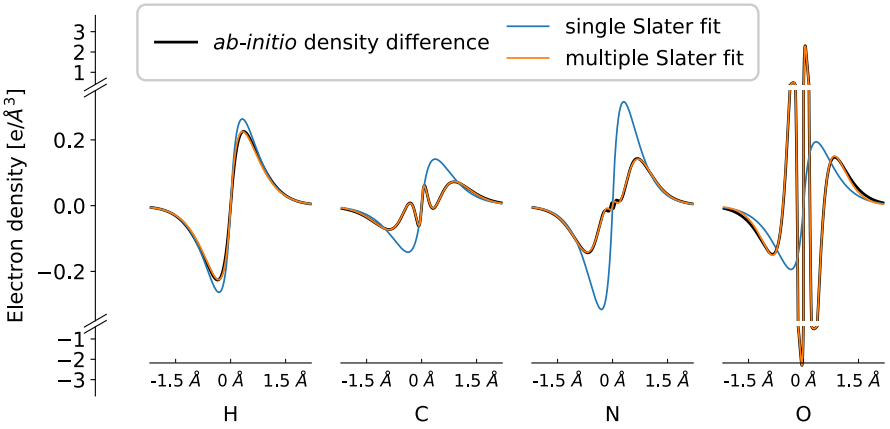


Figure 4.1: *Ab initio* density difference between the unperturbed hydrogen, carbon, nitrogen and oxygen atoms and the same atoms in a dipole field along the perturbation axis (in black), the single Slater function fit (in blue), and the least squares fit of multiple Slater functions to the *ab initio* density (in orange).

Slater dipole response functions

Commonly, point dipoles or Gaussian dipoles are used in inducible dipole models as response functions, consistent with the atomic point charges or Gaussian charge densities employed to represent the ground state charge distribution of the isolated fragments constituting the system. However, as the true electron density tails off exponentially, Slater functions are used in this work to represent the valence electron density, as in Equation (4.1). For consistency with this ground state density, Slater dipole functions are constructed as the gradient of a normalized Slater 1s function. However, as this function has no well-defined limit towards $\boldsymbol{R}_A$, instead a normalized 1s+2s function is employed:

$$\begin{aligned}\boldsymbol{\rho}_A^{\mathrm{p}}(\boldsymbol{r}) &= \boldsymbol{\nabla}_A \rho_A^{\mathrm{1s+2s}} \\ &= \boldsymbol{\nabla}_A \left[\frac{1}{32\pi {\sigma_{A,\mathrm{p}}}^3} \exp\left(-\frac{|\boldsymbol{r}-\boldsymbol{R}_A|}{\sigma_{A,\mathrm{p}}}\right)\left(1+\frac{|\boldsymbol{r}-\boldsymbol{R}_A|}{\sigma_{A,\mathrm{p}}}\right)\right] \\ &= \frac{1}{32\pi {\sigma_{A,\mathrm{p}}}^5} \exp\left(-\frac{|\boldsymbol{r}-\boldsymbol{R}_A|}{\sigma_{A,\mathrm{p}}}\right)(\boldsymbol{r}-\boldsymbol{R}_A) \end{aligned} \tag{4.12}$$

Physically, such a Slater dipole can be interpreted as the density difference between an unperturbed ground state electron density and a density perturbed by an electric dipole field. From this interpretation, the dipole width $\sigma_{A,\mathrm{p,free}}$ for each free atom can be determined by fitting a Slater dipole to

the normalized *ab initio* density difference of a free atom. Subsequently, to account for the chemical environment of each atom in a molecule or solid, the free atom dipole widths $\sigma_{A,\mathrm{p,free}}$ are scaled with the cube root of the ratio of its effective volume in a molecule compared to its effective volume in vacuum[234], similarly to the rescaling used for the polarizabilities:

$$\sigma_{A,\mathrm{p}} = \left(\frac{\int r^3 \rho_A(\boldsymbol{r})\, d\boldsymbol{r}}{\int r^3 \rho_{A,\mathrm{free}}(\boldsymbol{r})\, d\boldsymbol{r}} \right)^{1/3} \sigma_{A,\mathrm{p,free}} \tag{4.13}$$

A similar functional form to Eq. (4.12) has been proposed before, although dipole widths were fit for specific atom types instead of making use of *ab initio* atomic density differences.[235] From the interpretation of a Slater dipole as a density difference between a perturbed and unperturbed atoms, atom-specific Slater dipoles can be employed as the basis set of Equation (4.9), by fitting to *ab initio* calculations of the respective atoms. The density differences for hydrogen, carbon, nitrogen and oxygen between an atom perturbed with a dipole field and the unperturbed atom as computed at the CCSD(T)/aug-cc-pVQZ level of theory are shown in Figure 4.1. As seen from Figure 4.1, a single Slater dipole per atom cannot capture all fluctuations of the density, but is in reasonable agreement far from the nucleus. To investigate whether these fluctuations cause significant deviations in the energy of our polarization model, an additional fitting of a linear combination of Slater dipoles per atom was also determined, shown in orange in Figure 4.1. For these elements, a very accurate fit to the *ab initio* density difference is possible using a linear combination of at most 5 Slater dipoles.

Interaction model

As discussed in Section 4.1.1, the purely classical interaction between dipoles neglects the important Pauli exclusion principle, requiring antisymmetrization of the wavefunction with respect to an exchange in electrons. While this constraint clearly affects the electron-electron interaction $\hat{\mathcal{V}}_{ee}$ (see Equation (2.15)), it also affects the kinetic energy T. The antisymmetrization leads to more strongly oscillating wavefunctions, resulting in a higher kinetic energy.[236] While the exchange interaction has generally been neglected in inducible dipole models, empirical models for it have been included in chemical potential equalization methods before.[225] Additionally, the exchange interaction is known to be approximately proportional to the overlap between the electron densities, allowing for an inclusion in our model.[152, 179, 237, 238] This proportionality factor was determined previously from fitting to SAPT2+(3) exchange-repulsion energies of dimers in the S66x8 set, a set containing 66 dimers representing a wide range of different types of intermolecular

interactions, along 8 points on the dissociation curve. This resulted in a value of $U_{\text{ex}} = 8.13$ a.u.[179, 239, 240] Therefore, in addition to a purely classical Coulomb interaction, an overlap term is added to the model

$$K(\boldsymbol{r}, \boldsymbol{r'}) = \frac{\delta^2 F[\rho]}{\delta\rho(\boldsymbol{r})\delta\rho(\boldsymbol{r'})} \approx \frac{1}{|\boldsymbol{r} - \boldsymbol{r'}|} + U_{\text{ex}}\,\delta(\boldsymbol{r} - \boldsymbol{r'}) \tag{4.14}$$

In other words, instead of only making use of a Coulomb term to derive Equation (4.10) from (4.7), the new kernel of Equation (4.14) is employed. Clearly, long-range correlation effects are neglected in this model. However, including an exchange interaction should significantly increase the physical accuracy of the model by an improved description of this short-range effect. To test this, the model is now benchmarked on its ability to reproduce molecular polarizabilities and many-body interaction energies.

4.1.3 Model benchmarking

Molecular polarizabilities

A first validation of the inducible dipole model developed herein is the reproduction of the molecular polarizability tensor. Applying a uniform dipole field to an isolated molecule, atomic dipoles are induced in response. This change in polarization per unit on the applied field is equal to the molecular polarizability, which can be directly compared to a reference calculation (at the B3LYP/aug-cc-pVTZ level of theory). In this way, different approximations of the interaction kernel of Equation (4.14) can be compared. First, turning off the interaction kernel ($K = 0$), the obtained molecular polarizability is equal to the sum of the isotropic atomic polarizabilities as obtained from the rescaling of Equation (4.11). This can be compared with the situation in which only a Coulomb interaction ($U_{\text{ex}} = 0$ in Equation (4.14)) or the full interaction ($U_{\text{ex}} = 8.13$) is turned on. This validation was performed on the QM7 dataset which is composed of 7165 molecules containing up to 23 atoms (including up to 7 non-hydrogen atoms) which were extracted from the larger GDB-13 dataset.[242, 243] Both the monomer electron densities and dipole polarizabilities were obtained at the B3LYP/aug-cc-pVTZ level of theory using Gaussian16.[244] From the obtained molecular polarizability matrices, both the eigenvalues and fractional polarizability anisotropy (FPA) are calculated. The latter is a metric for the anisotropy ranging from 0 to 1, and is defined as follows[245]

$$\text{FPA} = \sqrt{\frac{1}{2}\frac{[(\alpha_{xx} - \alpha_{yy})^2 + (\alpha_{xx} - \alpha_{zz})^2 + (\alpha_{yy} - \alpha_{zz})^2]}{\alpha_{xx}^2 + \alpha_{yy}^2 + \alpha_{zz}^2}} \tag{4.15}$$

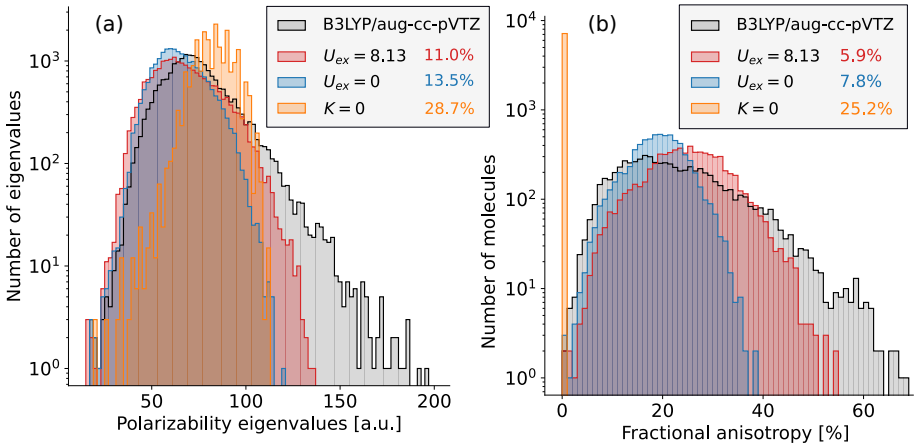


Figure 4.2: (a) Eigenvalues of the dipole polarizability matrices of molecules in the QM7 dataset obtained at the reference B3LYP/aug-cc-pVTZ level of theory and with the inducible dipole model with $U_{\text{ex}} = 8.13$ in red, $U_{\text{ex}} = 0$ in blue and $K = 0$ in orange. The root mean square percentage errors (RMSPEs) compared to the reference are shown in the legend. (b) Fractional polarizability anisotropy of the obtained dipole polarizability matrices. The RMSPEs compared to the reference level of theory are shown in the legend.

The resulting eigenvalues and FPEs of molecules in the QM7 dataset are shown in Figure 4.2(a) and 4.2(b), respectively. As can be seen on the left side, turning on the dipole–dipole interactions via a Coulomb term significantly improves the description of the molecular polarizability with respect to the case $K = 0$, reducing the root mean square percentage error (RMSPE) from 28.7% to 13.5%. Secondly, turning on the non-classical exchange term ($U_{\text{ex}} = 8.13$) further reduces the RMSPE to 11.0%. The difference between the various approximations to the interaction kernel is highlighted even more when considering the description of the anisotropy via Equation (4.15), as shown in Figure 4.2(b). The RMSPE on the FPA is decreased from 25.2% for $K = 0$ (no anisotropy) to 5.9% for $U_{\text{ex}} = 8.13$. Again, an improvement is seen by adding an exchange term with respect to the pure Coulomb term. The ability of the model to reproduce the response of a molecule to an external dipole field is already a first validation. A second validation consists in its ability to accurately predict intermolecular interaction energies, as will be benchmarked in the following Section.

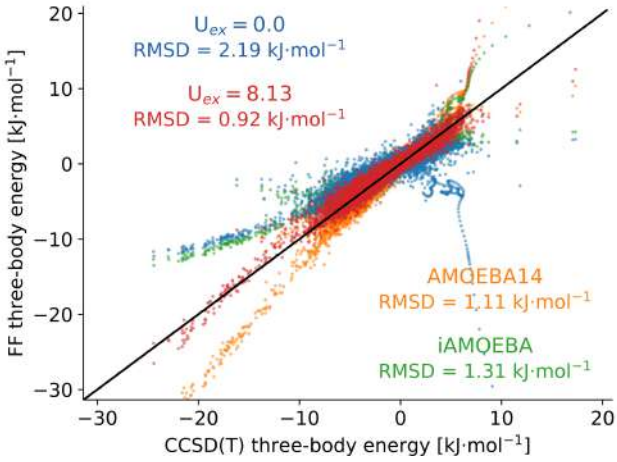


Figure 4.3: Parity plot of a set of 12 347 water trimer three-body energies calculated with the inducible dipole model with $U_{\text{ex}} = 0$ in blue (representing only the Coulomb interaction), $U_{\text{ex}} = 8.13$ in red (including the exchange interaction), AMOEBA14 in orange and iAMOEBA in green. Results are compared with reference CCSD(T) three-body energies.

Many-body interaction energies

To validate the developed polarizable force field, its prediction of intermolecular polarization energies is now investigated. In order to benchmark it separately from other intermolecular interactions, our attention is turned to three, four and five-body interaction energies. While two-body terms are composed of electrostatic, exchange, dispersion and induction terms, benchmark systems such as water clusters can be chosen in which the higher order terms are dominated by induction.[246, 247] In this way, the predicted energies can be explicitly compared with *ab initio* many-body energies. First, the intermolecular three-body energies of a set of 12 347 water trimers in a wide range of conformations are computed with the inducible dipole model using the multiple Slater fit of Figure 4.1 and compared to reference calculations performed at the CCSD(T)/aug-cc-pVTZ level of theory.[248] In Figure 4.3, a comparison is shown between the reference energies and those computed with the inducible dipole model without an exchange interaction ($U_{\text{ex}} = 0$) and with the exchange interaction ($U_{\text{ex}} = 8.13$). Significant deviations are observed for the pure Coulomb interaction, significantly underestimating the energy for low-lying trimers. This trend is reversed for trimers with repulsive

three-body energies, erroneously predicting stabilizing interactions. The inclusion of exchange significantly improves the predicted energies, reducing the root mean square error (RMSD) from 2.19 $kJ \cdot mol^{-1}$ to 0.92 $kJ \cdot mol^{-1}$. Only minor difference are observed when the single Slater dipole functions are employed instead of the multiple Slater functions, increasing the RMSD from 0.92 $kJ \cdot mol^{-1}$ to 1.06 $kJ \cdot mol^{-1}$ when including exchange. This suggests that the inclusion of the precise density fluctuations near the nucleus is not vital to obtain accurate three-body energies, and the single Slater fit can be used for improved computational efficiency. The trimer dataset for this comparison was previously constructed to fit a full-dimensional potential energy function for water (MB-pol).[248] As our model instead aims at being transferable without fitting to the interaction energies of the systems of interest, it cannot be expected to be competitive on the specific case of water with MB-pol. Instead, it is compared with the popular iAMOEBA and AMOEBA14 force fields from the literature.[146, 255] These force field result in an RMSD of 1.31 $kJ \cdot mol^{-1}$ and 1.11 $kJ \cdot mol^{-1}$, respectively. Especially trimers with negative three-body energies are described well by our model (RMSD of 0.79 $kJ \cdot mol^{-1}$), compared with iAMOEBA and AMOEBA14 (RMSD of 1.52 $kJ \cdot mol^{-1}$ and 1.29 $kJ \cdot mol^{-1}$, respectively). The main reason why AMOEBA14 can reproduce three-body energies adequately without the inclusion of an exchange interaction originates from its use of point dipoles. As these interact more strongly than delocalized dipoles, the missing exchange interaction is partly compensated. The performance on water three-body energies is an encouraging sign for the model developed herein, as both literature force fields contain parameters fitted specifically to reproduce water interaction energies and condensed phase properties, while no such fitting is performed for our model. In fact, our model contains no parameters fitted on interaction energies of water. Only the monomer electron density of a single optimized water molecule was used, together with a fit for the atomic dipole basis for each element and the U_{ex} parameter fitted previously on the dispersion-dominated dimers of the S66x8 set. As an even more direct validation of the physical validity of the developed model, four-body and five-body energies were predicted for a set of 8 water hexamer clusters.[249] These higher-order contributions are generally smaller than the two-body and three-body terms. As a consequence, these terms are usually not fitted to in the construction of specialized water force fields. Instead, they provide a look into the reliance of a force field on cancellation of errors. As shown in Figure 4.4, our model is competitive with the TTM4-F, WHBB5 and MB-pol force fields in the prediction of four-body energies, and even improves on those tailor-made force fields in the prediction of five-body energies. In comparison, AMOEBA14 predicts significantly over-bounded four-body and five-body energies.

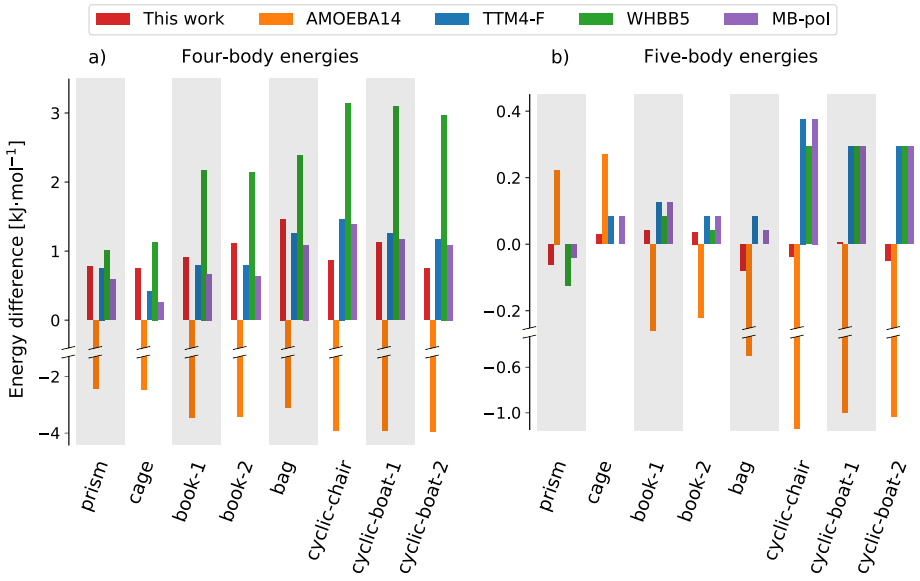


Figure 4.4: Error on the four-body (a) and five-body (b) interaction energies of 8 water hexamers as computed with our model and water potentials from the literature, with respect to reference calculations at the CCSD(T)-F12/cc-pVTZ level of theory.[248–254]

To extend the benchmark on three body energies to a wider range of intermolecular interactions, calculations were additionally performed on the 3b69 set of trimers.[256] Included in this set are 18 out of the 24 molecules which are not present in the S66x8 set to which the U_{ex} parameter was fitted. It therefore represent a test of how well the model transfers to interactions to which the parameter was not fitted. To test the many-body polarization component separately from many-body dispersion, interaction energies were recalculated at the MP2 level of theory extrapolated to the complete basis set. MP2 includes many-body induction effects, but the dispersion non-additivity only appears at the MP3 level, allowing for a comparison between the inducible dipole model and MP2 three-body energies.[256, 257] An RMSD of 0.88 kJ · mol^{-1} was obtained, demonstrating good performance across a range of different interactions. In comparison, AMOEBA three-body energies were obtained for the 22 out of the 69 trimers for which parameters are available.[258] For this subset, an RMSD of 0.67 kJ · mol^{-1} is obtained. However, the observation that the application of a regularly used polarizable force field from the literature is limited to certain interactions demonstrates the usefulness of a transferable model, as it can be applied to a wide range of interactions.

4.1.4 The polarizable monomer electron density force field

To predict total non-covalent interaction energies, the developed polarization model has to be combined with models for the other intermolecular interactions. For this, the monomer electron density force field (MEDFF) is ideally suited. In analogy to the *ab initio* energy decomposition made by symmetry-adapted perturbation theory (SAPT), MEDFF separates the total interaction energy in an electrostatic, dispersion, exchange-repulsion and induction interaction. Each one is a functional form of the electron density of the constituting monomers:

$$E_{\text{MEDFF}} = E_{\text{elst}} + E_{\text{disp}} + E_{\text{ex}} + E_{\text{ind}} \tag{4.16}$$

$$= \sum_{\text{A}} \sum_{\text{B}} \left(E_{\text{elst}}^{\text{AB}} + E_{\text{disp}}^{\text{AB}} + U_{\text{ex}} S^{\text{AB}} - U_{\text{ind}} S^{\text{AB}} \right) \tag{4.17}$$

with the sum over A and B running over all atoms in the first and second monomer, respectively, and S^{AB} the overlap integral between electron densities ρ of atoms A and B:

$$S^{\text{AB}} = \int \mathrm{d}\boldsymbol{r} \rho_A(\boldsymbol{r}) \rho_B(\boldsymbol{r}) \tag{4.18}$$

Both the exchange-repulsion and induction interactions are modeled as a proportionality of the overlap integral of electron densities of the monomers (with proportionality factors U_{ex} and U_{ind}). While the exchange-repulsion was shown to be accurately modeled with the overlap model, the picture is more complex for induction. This interaction consists of multiple effects like polarization and charge-transfer effects. Of these, charge-transfer will typically dominate in dispersion-dominated interactions where polarization plays a minor role. Moreover, charge-transfer is known to be proportional (to good approximation) to the overlap between electron densities. For this reason, MEDFF performs well on benchmarks of dispersion-dominated interactions. However, for hydrogen-bonded interactions in which polarization is dominant, significantly worse performance is observed. Clearly, polarization cannot be accurately captured by a pairwise additive overlap model. Herein lies the value of our previously developed polarization model. Incorporating our polarization model in MEDFF, the polarizable monomer electron density force field (PMEDFF) is obtained:

$$E_{\text{PMEDFF}} = E_{\text{elst}} + E_{\text{disp}} + E_{\text{ex\_ct}} + E_{\text{pol}} \tag{4.19}$$

$$= \sum_{\text{A}} \sum_{\text{B}} \left(E_{\text{elst}}^{\text{AB}} + E_{\text{disp}}^{\text{AB}} + U_{\text{ex\_ct}} S^{\text{AB}} \right) + E_{\text{pol}} \tag{4.20}$$

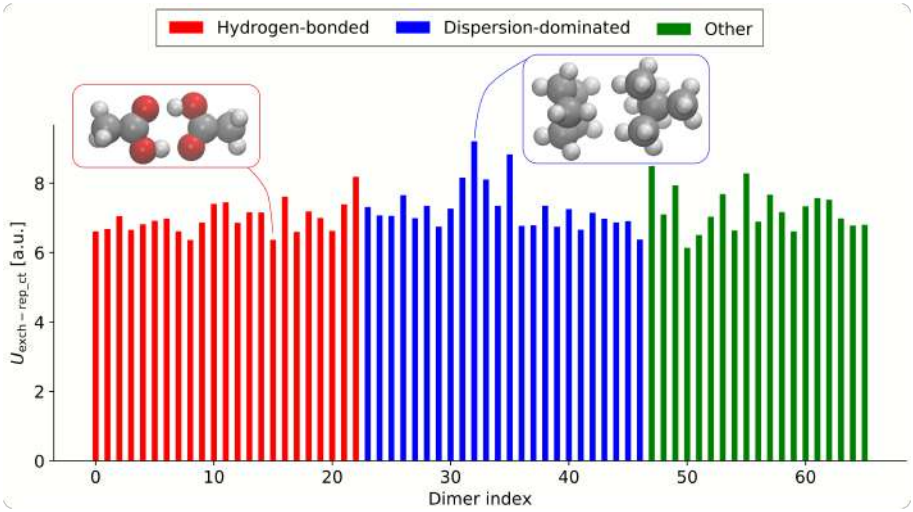


Figure 4.5: Fitted values of U_{ex_ct} for each of the 66 dimers present in the S66x10 set. For each dimer, U_{ex_ct} is fitted to the 10 dimer configurations along the dissociation curve. For the hydrogen-bonded dimers, acetic acid is shown as example, while the cyclopentane-neopentane dimer is shown among the dispersion-dominated structures.

with E_{pol} given by Equation (4.10). Herein, U_{ex_ct} represents the proportionality between the exchange-repulsion and charge transfer interactions and the electron density overlap. The common treatment of both these interactions is further justified by the observation that in the higher order terms of SAPT, exchange-repulsion and induction effects cannot be distinguished. This new proportionality factor U_{ex_ct} is therefore fitted to the sum of the SAPT2+(3) exchange-repulsion and induction energies minus the polarization energy E_{pol} for dimers of the S66x10 set. This set is an extension of the original S66x8 set, including smaller intermolecular distances (with rescaling factors of 0.7 and 0.8 with respect to the equilibrium distance).[259, 260] This fit results in a value of U_{ex_ct} = 7.16. However, some dependence of this parameter on the type of interaction remains, as can be seen from Figure 4.5 in which a fit was made for each separate dimer complex. On average, slightly lower values for U_{ex_ct} are observed for the hydrogen-bonded dimers, shown in red. For example, for the acetic acid dimer with a double hydrogen bond, a value of U_{ex_ct} = 6.37 is obtained. Among the dispersion-dominated structures shown in blue, the cyclopentane-neopentane dimer results in the largest value of U_{ex_ct} = 9.20. However, there is a clear indication that the inclusion of our polarizable model has captured an important part of the physical interactions

between these dimers better than the overlap model used in the original MEDFF (see Equation (4.17)). The standard deviation between fitted values of $U_{\text{ex\_ct}}$ in Figure 4.5 is equal to 0.59. This value can be interpreted as the degree variability across different intermolecular interactions, and therefore the degree to which the model is 'missing' some physics. Fitting the original U_{ex} and U_{ind} parameters in the same way, a standard deviation of 0.97 is obtained, indicating that our polarizable model more accurately captures the intermolecular interactions. Furthermore, a correlation is present between the fitted values for $U_{\text{ex\_ct}}$ and the original MEDFF values for U_{ex} which were fitted to the *ab initio* SAPT2+(3) exchange-repulsion energies. This indicates that an important remainder of the variability of our fitted $U_{\text{ex\_ct}}$ parameter originates from a breakdown of the relatively simple overlap model for exchange-repulsion. Clear indications already exist in the literature how such a model could be improved, relating mainly to a description of the atomic anisotropy of the interactions.[261] However, such extensions would introduce new fitting parameters, diminishing on our promise of a transferable force field with as little parameters as possible. While alternatively, deeper insight into *ab initio* derivations of the origin of exchange-repulsion could lead to a better force field description of this interaction, this will be left for further work.

Interaction energies of the hsg and hbc6 datasets

With our newly developed and fitted polarizable force field, interaction energies are now calculated for two datasets with dimers not included in the S66x10 set to which $U_{\text{ex\_ct}}$ was fitted. Both datasets and the associated interaction energies computed at the CCSD(T)/CBS level of theory were taken from the BioFragment Database.[262–264] The first is the hsg set, consisting of 21 dimer fragments (none of which are present in the S66x10 set) extracted from an HIV-II protease crystal structure with a bound ligand (indinavir), representing a wide range of interactions from dispersion-dominated to hydrogen-bonded. For this set, the general value of $U_{\text{ex\_ct}} = 7.16$ is used. In Figure 4.6, the resulting interaction energies obtained with PMEDFF are compared to the Merck molecular force field (MMFF)[265], AMOEBA[258], the generalized amber force field (GAFF)[219] and the original MEDFF[179]. Both AMOEBA and GAFF show a balanced performance across the range of interaction types, although the RMSD of 3.22 $\text{kJ} \cdot \text{mol}^{-1}$ obtained for AMOEBA only includes 13 out of the 21 complexes for which parameters are available. A significant overestimation of the magnitude of the interaction energy of the hydrogen-bonded complexes is observed for MMFF, resulting in an RMSD of 19.40 $\text{kJ} \cdot \text{mol}^{-1}$. MEDFF, on the other hand, slightly underesti-

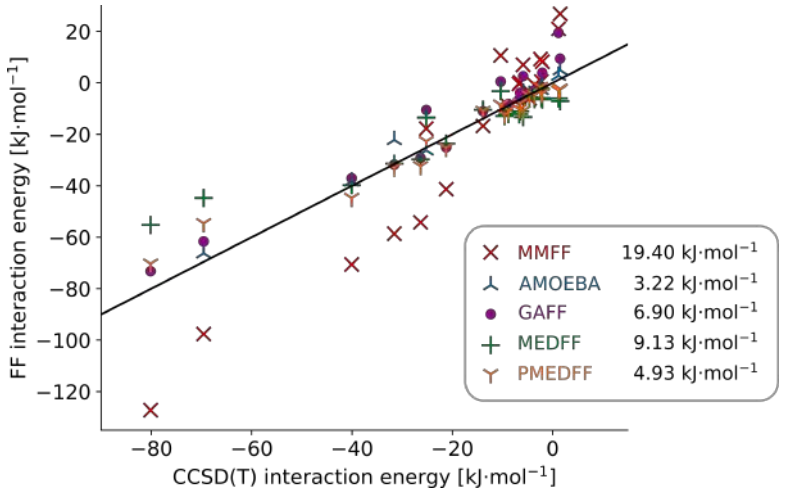


Figure 4.6: Parity plot of dimer interaction energies of the hsg set calculated with MMFF (red), AMOEBA (blue), GAFF (purple), MEDFF (green) and PMEDFF (orange). The RMSD for each force field is shown in the legend.

mates the interaction energy for some of these dimers, yielding an RMSD of 9.13 kJ · mol^{-1}. PMEDFF generally performs well on the whole set, with an RMSD of 4.93 kJ · mol^{-1}. This is encouraging, as no fitting was performed to any of the dimers present in this set.

The hbc6 dataset consists of doubly hydrogen bonded dimers extracted from 6 dissociation curves containing formic acid, formamide and formamidine.[266] Because of the presence of double hydrogen bonds, we opt to use the $U_{\text{ex\_ct}}$ parameter of 6.37 which was optimal for the double hydrogen bond containing acetic acid dimer (see Figure 4.5). The induction component of these doubly hydrogen bonded dimers is generally much larger than for dimers in the hsg set. This set therefore represents a more stringent test of our polarization model. In Figure 4.7, a comparison is made between PMEDFF, MMFF, AMOEBA and MEDFF. For AMOEBA, only 60 out of the 118 dimers containing formic acid and formamide were retained, as no parameters were available for formamidine. The comparison between force fields is further divided into intermolecular distances smaller or larger than that of the optimized dimer geometry. For intermolecular distances larger or equal to the optimized geometry, the RMSD of MMFF, AMOEBA, MEDFF and PMEDFF is equal to 8.95 kJ · mol^{-1}, 2.93 kJ · mol^{-1}, 8.24 kJ · mol^{-1} and 5.56 kJ · mol^{-1}, respectively. The dimers with rescaled intermolecular distances smaller than 1 are more challenging to predict, as hydrogen bonds are artificially compressed. The increase in error from the rescaled distances larger than

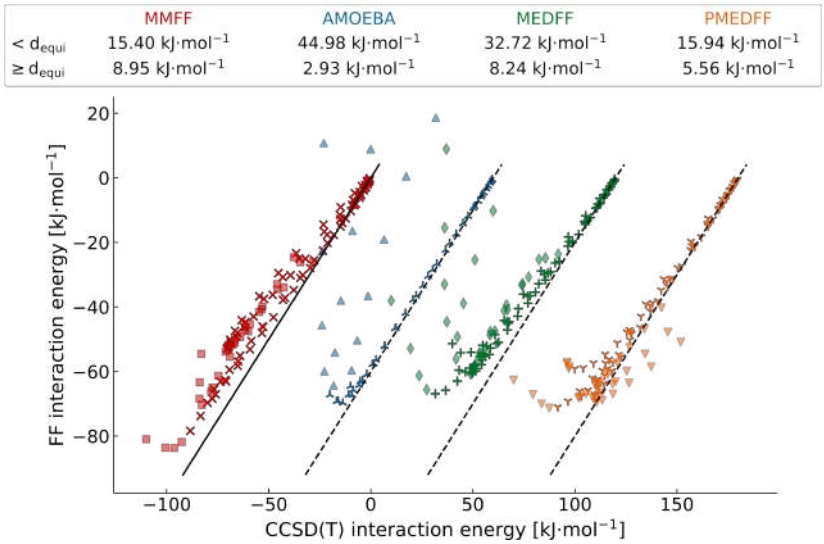


Figure 4.7: Parity plot of dimer interaction energies of the hbc set calculated with MMFF (red), AMOEBA (blue), MEDFF (green) and PMEDFF (orange). Dimers along the dissociation curve with rescaled intermolecular distances smaller than the equilibrium distance d_{equi} are displayed with filled symbols in a lighter shade. Results for all force fields except MMFF are shifted horizontally for clarity. The RMSD for each force field at distances < d_{equi} and $\geq d_{equi}$ are shown in the legend.

1 is most pronounced for AMOEBA. The error increases to 44.98 kJ · mol^{-1} due to the prediction of large repulsive interaction energies. Lower errors of 32.72 kJ · mol^{-1}, 15.40 kJ · mol^{-1} and 15.94 kJ · mol^{-1} are seen for MEDFF, MMFF and PMEDFF. Overall, the performance of PMEDFF against other force fields is encouraging, especially given that, as a test of transferability of our force field, no explicit fitting was performed to any of the dimers present in both the hsg and hbc6 set.

To conclude, it is clear from the preceding discussions how, starting from the ground state electron density of a molecule, the inclusion of usually neglected physical interactions in a polarization model can lead to an improved performance on many-body interactions. Moreover, a single fitted parameter in the model is shown to perform well across a range of intermolecular interactions, demonstrating its transferability. The inclusion of our polarization model in a complete polarizable non-covalent force field leads to a model that can reliably predict interaction energies of dimers. However, the U_{ex_ct} parameter is still slightly sensitive to the type of intermolecular interaction, indicating

that further developments are warranted to attain a truly transferable force field. In this respect, extending the model to include an anisotropic treatment of the atomic interactions has promise to improve our force field. Our strategy of including as much physical consideration into an interaction model to avoid as many fitting parameters as possible has promise in this regards, although the complete polar opposite of this approach has in recent year gained traction, spurred by the development of highly accurate machine learning potentials.

4.2 Deriving machine learning potentials for accurate adsorption simulations

While polarizable force fields can be employed to model guest adsorption in MOFs with a degree of accuracy higher than non-polarizable force fields, their accuracy can still be lacking in challenging cases such as open-metal site containing MOFs. In these cases, one would preferentially directly use density-functional theory or wavefunction-based methods. However, the computational cost associated with these calculations is still excessively high for the typical number of evaluations required in adsorption calculations, necessitating the use of force fields. However, recently, approaches to bridge this gap and directly incorporate DFT or wavefunction-based calculations in adsorption simulations in MOFs have been developed. Lee *et al.* evaluated the Henry regime (characterizing the low-pressure adsorption behavior) in M-MOF-74(M=Zn,Mg) for several adsorbents by performing biased single particle insertions, evaluated at the DFT level of theory.[267] In a similar work, Vandenbrande *et al.* proposed an importance sampling scheme based on preceding force field calculations to evaluate the Henry coefficients and heats of adsorption at infinite dilution for methane in UiO-66 and CO_2 in Mg-MOF-74 by means of single particle insertions.[268] Alternatively, Kundu *et al.* predicted the adsorption isotherms of CO_2 in Mg-MOF-74 from calculations of the interaction energy of an adsorbate on the open-metal site and linker with a correlated wavefunction-corrected DFT method.[269–271] With these adsorption sites, a lattice of sites was defined, on which grand canonical Monte Carlo simulations were performed. Although successful in predicting the adsorption isotherms at different temperatures, this method relies on an initial accurate determination of well-defined adsorption sites, as guest molecules are not considered to adsorb on other sites of the framework, limiting the general applicability of this method.

Therefore, in **Paper IV**, we make use of the recent developments of highly accurate and data-efficient machine learning potentials (MLPs) to describe the interaction of guest molecules in MOFs. In the message passing NequIP potential used herein, each atom is characterized by a feature vector which is refined during training through a series of interactions layers.[190] Tested on datasets of energies and forces of small molecules and periodic materials, it demonstrated state-of-the-art accuracy as well as exceptional data efficiency. While MLPs are usually trained on the totally energy and forces of the system, our interest is solely in intermolecular energies (and forces). First, a method for efficiently generating a dataset of structures and DFT intermolecular energies and forces is proposed. After training an MLP on this dataset, GCMC calculations can be performed to extract the adsorp-

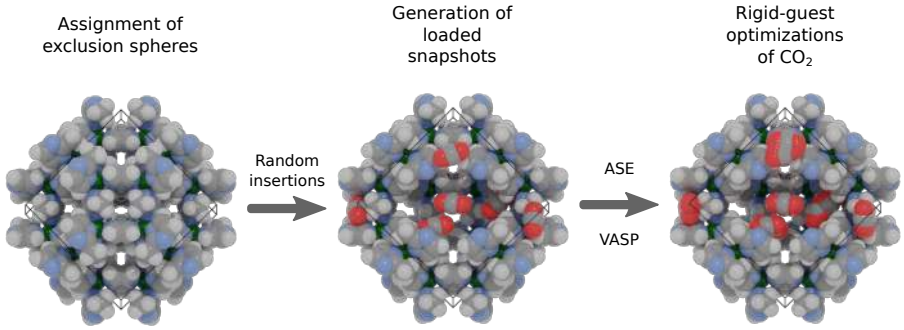


Figure 4.8: Proposed methodology for generating the dataset to train an MLP on intermolecular energies (and forces). First, snapshots of the framework loaded with adsorbates respecting minimum intermolecular distances defined by the exclusion spheres are generated. Subsequently, rigid-guest optimizations of the adsorbates are performed (with a fixed framework) to guide the adsorbates to favorable interaction sites in the framework.

tion isotherms and heats of adsorption, as well as structural information as captured for example by radial distribution functions (RDFs). The proposed methodology is benchmarked first on CO_2 adsorption in ZIF-8, after which it is employed to gain insight into CO_2 adsorption in the challenging open-metal site containing Mg-MOF-74. Finally the degree of transferability of a trained network is investigated by applying an MLP trained on ZIF-8 to other MOFs in the ZIF family.

4.2.1 CO_2 adsorption in ZIF-8

The requirements on an MLP to be used for GCMC simulations differs fundamentally from one with MD applications is mind. Through the integration of the equations of motion, an equilibrium canonical MD simulation samples regions of the potential energy surface (PES) proportional to a Boltzmann factor of the energy. While the exact same distribution is sampled from (canonical) Monte Carlo simulations, this is achieved by performing trial moves of the system to regions of the PES which might be very high in energy. Even though these trial moves will not be accepted and are therefore not relevant, they still need to be reliably rejected. A dataset for our envisaged applications therefore needs to be composed of both high energy snapshots, as well as the relevant low energy snapshots. However, repulsive interactions between atoms diverge as they approach each other, and the inclusion of these highly non-physical interactions would degrade an MLPs quality on

the relevant low energy interactions. Therefore, the procedure shown in Figure 4.8 is proposed to obtain a balanced dataset of loaded frameworks including both high energy and low energy structures. First, the van der Waals radii derived by Alvarez are assigned to each element present in the framework and adsorbates. [272] Their analysis of more than 5 million non-covalent distances shows that no non-bonded interactions occur at interatomic distances below the sum of two atomic radii minus 0.7 Å. Therefore, the van der Wall radius for each atom minus 0.35 Å can be seen as an 'exclusion sphere' within which no adsorbate-adsorbate or adsorbate-framework interactions occur. These exclusion spheres are also used further on (with proper validation) in GCMC simulations to reject any trial step in which an overlap between spheres occurs. Because of the use of these spheres, no explicit training of the MLP to these very short-ranged (repulsive) interaction is required.

Subsequently, a set of N guest-loaded snapshots constrained by the exclusion spheres are generated. As it is not known *a priori* how many guests can feasibly be adsorbed, snapshots are generated with a uniformly distributed number of guests N_{ads} between 0 and N_{max}, the maximum number of guests which can reliably be inserted in the framework with violating the constraints (equal to 32 CO_2 adsorbates for ZIF-8). Although the snapshots generated from this procedure respect the minimum intermolecular distances defined by the van der Waals radii, they still mostly represent high energy structures. Therefore, in a last step, each snapshot is optimized using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimizer from the Atomic Simulation Environment (ASE) at the relevant level of theory for N_{steps} iterations, keeping both the framework coordinates fixed and the adsorbates rigid.[273] After these optimizations, a dataset of $N \times N_{\text{steps}}$ snapshots is obtained. By subtracting the energies and forces of the constituting isolated framework and adsorbates from the total energies and forces, the intermolecular energies E_{inter} and forces $\boldsymbol{F}_{\text{inter}}$ are obtained. As the framework remains fixed, the energy and forces of the framework are the same for each snapshot, requiring only 1 single point calculation. Similarly, because of the use of rigid adsorbates, only a single calculation of the energy and forces of an isolated adsorbate are required. For CO_2 adsorption in ZIF-8, 30 optimization steps with a maximum step size of 0.5 Å each were performed. After 30 steps, the optimizations have steered the adsorbates from an adsorption energy of -1.1 $\pm$ 4.7 to -18.8 $\pm$ 2.1 kJ $\cdot$ mol^{-1} per adsorbate. Note that it is not vitally important for all optimization trajectories to fully converge. As long as sufficient favorable interactions are included in the dataset, the MLP is able to accurately represent the minima (see further).

To obtain both accurate and computationally efficient MLPs, the impact of

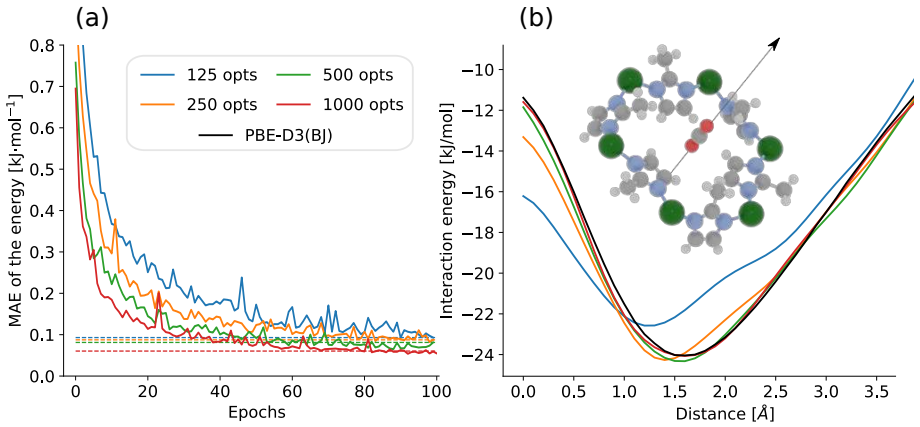


Figure 4.9: (a) The validation MAE on the energy during training of an MLP trained on 125, 250, 500 and 1000 optimization trajectories. (b) Dissociation curve of a single adsorbate translated along the normal vector on the plane of the pore window of ZIF-8, computed with all trained MLPs and recomputed at the reference PBE-D3(BJ) level of theory. For clarity, only the Zn atoms and imidazolate linkers of the six-membered pore ring are shown.

the important hyperparameters of the network were benchmarked in the Supporting Information of **Paper IV**. Vitally important for applications is the required number of optimizations N to obtain an accurate MLPs. As an investigation into this, MLPs were trained on 125, 250, 500 and 1000 optimizations (denoted further as MLP_{125}, MLP_{250}, MLP_{500} and MLP_{1000}). The resulting mean absolute errors (MAEs) on the interaction energy vary from 0.093 to 0.060 $kJ \cdot mol^{-1}$ per adsorbate, as shown in Figure 4.9(a). Even for just 125 optimization trajectories, the MAE is already remarkably low. However, it should be noted that, as the dataset contains clear correlations within each optimization trajectory, this validation error is artificially lowered compared to a test error. Indeed, evaluating the performance of each MLP on a test set containing 100 optimizations to which no training was performed, elevated errors of between 0.45 $kJ \cdot mol^{-1}$ and 0.15 $kJ \cdot mol^{-1}$ per adsorbate are obtained for MLP_{125} and MLP_{1000}, respectively. As a first validation of the derived MLPs, optimizations of a single adsorbate in the framework are performed with the MLP trained on 1000 optimization, retaining the geometry with the lowest energy. The strongest interaction site is a configuration in which the adsorbate interacts with a pore widow, perpendicular to the plane of the window (see Figure 4.9(b)). To explore the interaction around this minimum, the adsorbate was subsequently translated along the normal of the pore window plane. Relative to the configuration centered in the pore window,

the adsorbate is translated with distances between 0 and 4 Å with steps of 0.1 Å. For each structure, the interaction energy is recomputed with all MLPs and the reference PBE-D3(BJ) level of theory. As shown in Figure 4.9(b), significant deviations are observed for MLP_{125}. For the other MLPs, good agreement with the reference energies are obtained. Furthermore, as the optimizations were performed with the MLP, this demonstrates that the forces learned during training enable accurate optimizations towards the minimum. Additionally, to demonstrate that the exact choice of exclusion spheres does not affect further results, the heat of adsorption Q_{st} of CO_2 in ZIF-8 at 273 K was computed by means of Widom insertions in the framework. This was repeated with rescaled values for the exclusion radii with a factor between 0.80 and 1.20. As shown in **Paper IV**, Q_{st} is indeed insensitive to the factor with which the exclusion radii were rescaled, demonstrating that the region excluded by the exclusion spheres is indeed an irrelevant part of the phase space.

Subsequently, with each of the MLPs trained on varying dataset sizes, GCMC simulations were performed at a temperature of 273 K and a range of gas pressures between 0.1 bar and 50 bar. Each simulation consists of $5 \cdot 10^6$ MC steps, with equal probability for translations, rotations, insertions and deletions. The resulting uptake as a function of the gas pressure and heat of adsorption as a function of the guest loading are shown in Figure 4.10(a) and Figure 4.10(b). The adsorption isotherms obtained with the MLPs trained on between 125 and 1000 optimizations are nearly identical. Comparing these isotherms with experimental results obtained by Simmons *et al.*, Abraha *et al.* and Gracés *et al.*, good agreement between both is observed, as seen in Figure 4.10(a).[52, 274, 275] While a slight over-prediction of the uptake at low gas pressure and an under-prediction of the uptake at high gas pressures is observed, this could be due to the rigid-framework approximation employed in GCMC simulations, as the flexible behavior of ZIF-8 and its effects on adsorption have been well-established.[276] As shown in Figure 4.10(b), MLP_{1000} predicts an increase in the heat of adsorption of CO_2 in ZIF-8 from 18.0 $kJ \cdot mol^{-1}$ at the smallest loadings to approximately 25.8 $kJ \cdot mol^{-1}$ at the highest loadings, in good agreement with experimentally derived heats of adsorption.[274] This increase in heat of adsorption as a function of the CO_2 loading is indicative for the relatively weak host-guest interactions in ZIF-8 compared to the guest-guest interactions. Even though the most favorable interaction sites for an adsorbate in ZIF-8 are already occupied at higher guest loadings, the heat of adsorption still increases due to the guest-guest stabilization. To determine the accuracy of the performed GCMC simulations compared to the reference level of theory, 50 snapshots were taken from the GCMC simulations performed with each of the MLPs at each gas pressure. The

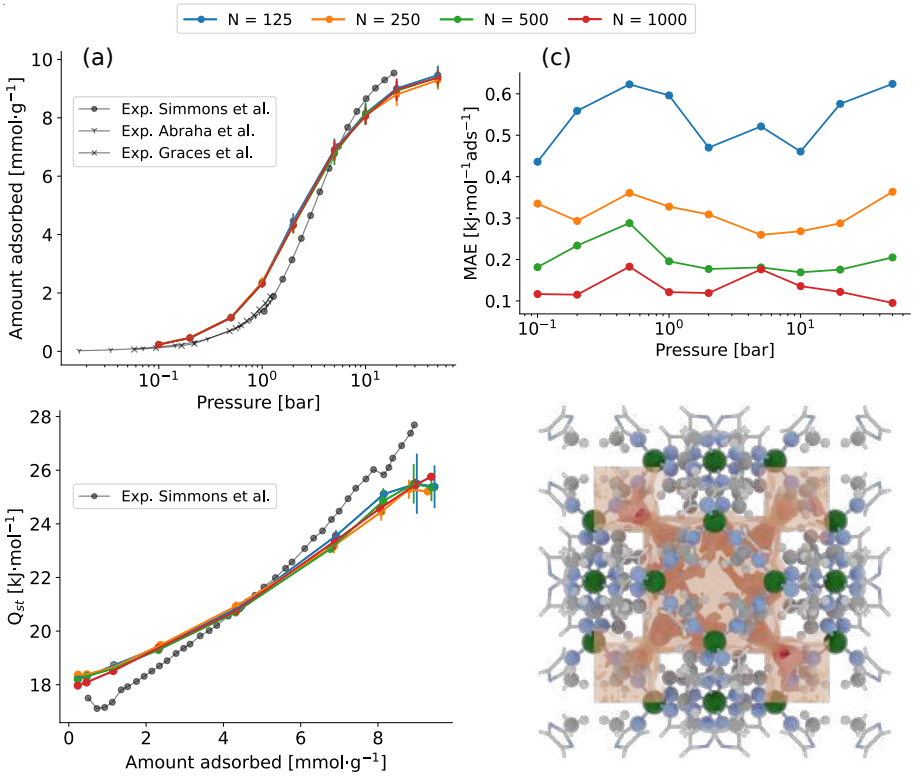


Figure 4.10: (a) The gas uptake and (b) isosteric heats of adsorption of CO_2 in ZIF-8 obtained with MLPs trained on between 125 and 1000 optimizations. Error bars are obtained from a block averaging procedure. (c) The test MAE on the energy of snapshots extracted from GCMC simulations as a function of the gas pressure and (d) a density plot of adsorbate oxygen atoms extracted from the MLP trained on 1000 optimizations as 10 bar.

interaction energy of these snapshots was recomputed with the reference PBE-D3(BJ) level of theory. As shown in Figure 4(c), the resulting average MAE on the energy decreases with increasing N from 0.48 to 0.17 kJ · mol^{-1} per adsorbate, similar to the test errors obtained previously. As argued before, these errors are somewhat higher than the validation MAE during training due to correlations in the dataset. However, for the MLP trained on 250 optimizations or more, the MAE is smaller than 0.25 kJ · mol^{-1}, which is certainly acceptable. As illustration, the density of adsorbate oxygen atoms extracted from the GCMC simulations using N=1000 at a pressure of 10 bar is shown in Figure 4(d).

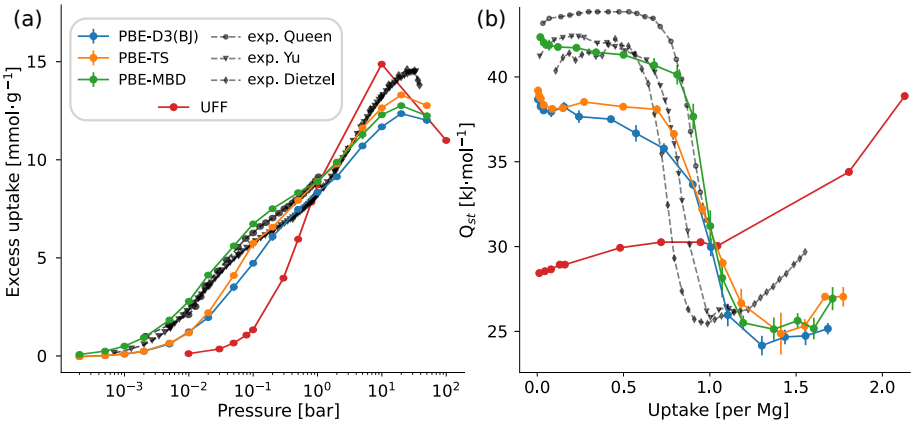


Figure 4.11: (a) The excess CO_2 uptake and (b) isosteric heats of adsorption in Mg-MOF-74 computed with MLPs trained on PBE-D3(BJ), PBE-TS and PBE-MBD reference optimizations, compared to experiments by Dietzel *et al.*[278], Yu *et al.*[279] and Queen *et al.*[64] and UFF simulations.[99] All MLP GCMC simulations were performed at 298 K and pressures between $2 \cdot 10^{-4}$ and 50 bar.

4.2.2 CO_2 adsorption in Mg-MOF-74

Having validated our proposed methodology, it is now applied to the challenging case of CO_2 adsorption in the open-metal sites containing Mg-MOF-74. Due to the undercoordination of the metal, CO_2 molecules can approach the Mg site in this MOF to within 2.2 Å, which is closer than would be allowed by the tabulated van der Waals radius for Mg.[99, 272] Therefore, the Mg van der Waals radius was set to 1.0 Å, allowing insertions of CO_2 in the framework to within 1.8 Å of the Mg site. This choice was validated *a posteriori* (see Supporting Information of **Paper IV**). From the literature, it is known that the interaction energy of a CO_2 molecule with the Mg open-metal site in Mg-MOF-74 depends significantly on the choice of dispersion correction.[277] Therefore, 1000 optimizations were performed in the framework at the PBE-D3(BJ), PBE-TS and PBE-MBD levels of theory. This selection of dispersion correction was inspired by previous work by Rehak *et al.* in which the adsorption energy on the Mg site obtained with different levels of theory was benchmarked.[277] Herein, the PBE-TS and PBE-MBD functionals most closely agreed with the experimentally derived adsorption energy. Therefore, these two functionals, as well as the commonly used PBE-D3(BJ) functional are considered.

The excess uptake of the framework and heats of adsorption obtained with GCMC simulations performed with an MLP trained to each of the levels of

theory at 298 K and gas pressures ranging from $2 \cdot 10^{-4}$ to 50 bar are shown in Figure 4.11(a) and 4.11(b), respectively. Results are compared with previous Universal Force Field (UFF) simulations, as well as experiments performed by Dietzel *et al.*, Yu *et al.* and Queen *et al.*[64, 99, 278, 279] UFF underestimates the uptake at low pressure and overestimates the uptake at the highest pressures. This is due to its lacking accuracy in describing the interaction strength with the Mg site (underestimating the heats of adsorption at low uptakes, see Figure 4.11(b)) as well as the mutual interaction between CO_2 guests (as seen from the heats of adsorption at the highest uptakes). GCMC simulations performed with the MLPs at each of the three levels of theory significantly improves the description of guest adsorption in the framework. Between these, PBE-MBD performs best, accurately capturing the shape of the adsorption isotherm at pressures below 1 bar. Moreover, as seen in Figure 4.11(b), the MLP trained on PBE-MBD reference data results in highly accurate heats of adsorption. Large differences between the heats of adsorption predicted by the MLPs trained on different levels of theory are observed, highlighting the importance in benchmarking the method at which reference calculations were performed. As shown by the differences in heats of adsorption in the low pressure regime, the choice of dispersion correction alone can affect the obtained heat of adsorption of CO_2 in Mg-MOF-74 by as much as 5 $kJ \cdot mol^{-1}$. Compared to the dependence of the heats of adsorption on the uptake seen for ZIF-8 (see Figure 4.10(b)), the opposite behavior is observed. At loadings below 1 adsorbate per Mg site, a heat of adsorption of approximately 42 $kJ \cdot mol^{-1}$ is obtained for the MLP trained on PBE-MBD reference data, in good agreement with experiments.[64, 278, 279] At higher loadings, the heat of adsorption suddenly drops to around 26 $kJ \cdot mol^{-1}$ due to the full occupation of Mg interaction sites, as also observed experimentally. This investigation demonstrates how, compared to UFF, a model which captures the interaction energy with the framework and other adsorbates with high accuracy can result in quantitative predictions of the adsorption properties in an open-metal site containing MOF.

Evaluating a test set consisting of 50 snapshots extracted from each GCMC simulation revealed a MAE on the energy of 0.41, 0.42 and 0.40 $kJ \cdot mol^{-1}$ for PBE-D3(BJ), PBE-TS and PBE-MBD, respectively. This is significantly higher than the test error for CO_2 adsorption in ZIF-8, demonstrating how 'learning' the diverse adsorption environment in Mg-MOF-74 resulting from high interaction energies on the Mg site and lower interaction energies elsewhere in the framework can be more difficult.

For additional insight in the adsorption behavior in the framework, as well as differences between levels of theory, radial distribution functions (RDFs) of the distances between the Mg site of the framework and adsorbate oxygen

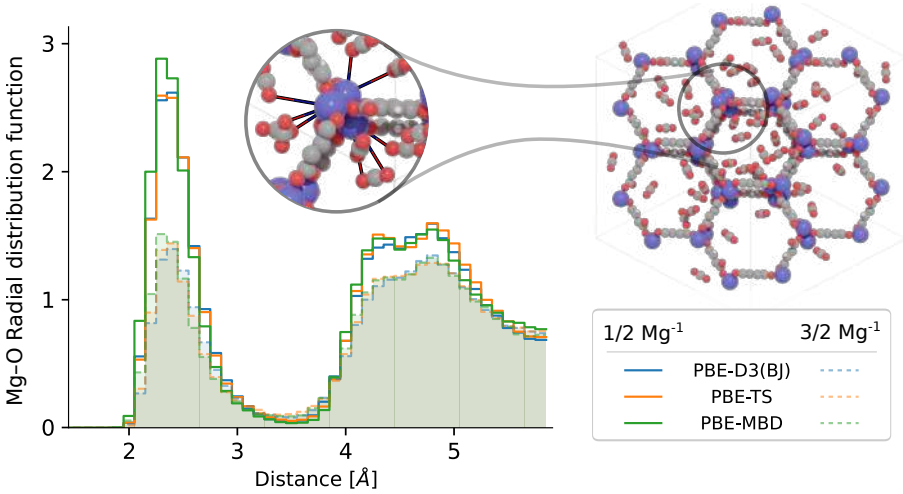


Figure 4.12: Radial distribution functions (RDFs) of magnesium–adsorbate oxygen atoms computed from GCMC simulations of CO_2 in Mg-MOF-74 performed with MLPs trained on PBE-D3(BJ), PBE-TS and PBE-MBD reference data. The full lines correspond to RDFs computed from a loading of half an adsorbate per Mg site, while the dotted lines correspond to a loading of one and a half adsorbates per Mg site.

atoms are computed, extracted from the performed GCMC calculations for each of the three levels of theory. In Figure 4.12, the RDFs are shown for framework loadings corresponding to both half (full lines) and one and a half adsorbate per Mg site (dotted lines). It is clear that adsorbates preferentially adsorbate at the Mg site. At a loading of half an adsorbates per Mg site, a significant peak occurs at a Mg–O distances of approximately 2.4 Å. A secondary more broad peak is seen between 4 and 5 Å, representing interaction with a neighboring Mg site. Compared to this, the first peak is significantly lowered at a framework loading of one and a half adsorbates per Mg site, as at this loading all sites are occupied, with adsorbates occupying the center of the pores, further away from the Mg sites. The difference between simulations performed with MLPs trained on each of the three levels of theory is minor. Only in the first peak, a slight shift of approximately 0.05 Å towards smaller distances is seen for PBE-MBD compared to the other levels of theory. This is consistent with the larger heat of adsorption on the Mg site obtained with PBE-MBD compared to PBE-D3(BJ) and PBE-TS, as seen from Figure 4.11(b).

While the MLPs trained and applied in the previous Sections perform very well, it would be a mistake to regard them in the same way as a classical

force field which, once fitted, are generally assumed to perform well when employed in materials with similar chemical environments; these possess a degree of transferability. The situation is more complicated with MLPs.

4.2.3 Transferability - a cautionary tale

Given the MLP used to model CO_2 adsorption in ZIF-8 from Section 4.2.1, a computational scientist who is used to working with force fields could be tempted to employ it to model CO_2 adsorption in a set of similar MOFs in the ZIF family. This approach is exemplified in most common force fields making use of atom types. Parameters derived for atoms in certain chemical environments are assumed to transfer well between different systems containing those environments. To investigate whether this concept of transferability can be applied to non-covalent interactions in MOFs modeled by MLPs, let us investigate how the MLP trained on CO_2 adsorption in ZIF-8 (denoted MLP_{ZIF8}) performs in other MOFs of the ZIF family: ZIF-3, ZIF-4 and ZIF-6. These different MOFs are shown in Figure 4.13(a). GCMC simulations of CO_2 in these ZIFs remained stable and showed no unphysical behavior. The resulting adsorption isotherms and heats of adsorption are shown with dotted lines in Figure 4.13(c). The predicted heats of adsorption of all three are comparable or lower than for ZIF-8. However, the uptakes in ZIF-3 and ZIF-6 can be higher at the largest considered gas pressures. This difference can be attributed to the differing pore structure between these frameworks. Using a nitrogen probe, the probe-occupiable pore volume computed with PoreBlazer[280] of ZIF-3 and ZIF-6 are equal to 0.61 $cm^3 \cdot g^{-1}$ and 0.80 $cm^3 \cdot g^{-1}$, compared to 0.22 $cm^3 \cdot g^{-1}$ and 0.52 $cm^3 \cdot g^{-1}$ for ZIF-4 and ZIF-8, correlating with the maximum uptake in these frameworks. However, the test error on the energies from snapshots extracted from GCMC simulation reveals a problem.

As shown in Table 4.1, the mean errors (ME) and mean absolute errors (MAE) are unacceptably high. Energies are consistently overestimated, leading to underestimated heats of adsorption and uptakes. As comparison to this, a single MLP was trained on 250 optimization trajectories in each of the four ZIFs (denoted $MLP_{ZIF\{3,4,6,8\}}$). The predicted uptakes and heats of adsorption using $MLP_{ZIF\{3,4,6,8\}}$ (shown in full lines in Figure 4.13(c)) differ significantly. Most notably, the heat of adsorption in the low pressure regime for ZIF-4 increases by almost 10 $kJ \cdot mol^{-1}$. For ZIF-3, the adsorption behavior fundamentally changes, displaying a minimum in the heats of adsorption at intermediate pressures, where this does not occur with MLP_{ZIF8}. Validating this newly trained MLP by recomputing snapshots at the reference PBE-D3(BJ) level of theory results in the MAEs and MEs summarized in Table 4.2. Interestingly,

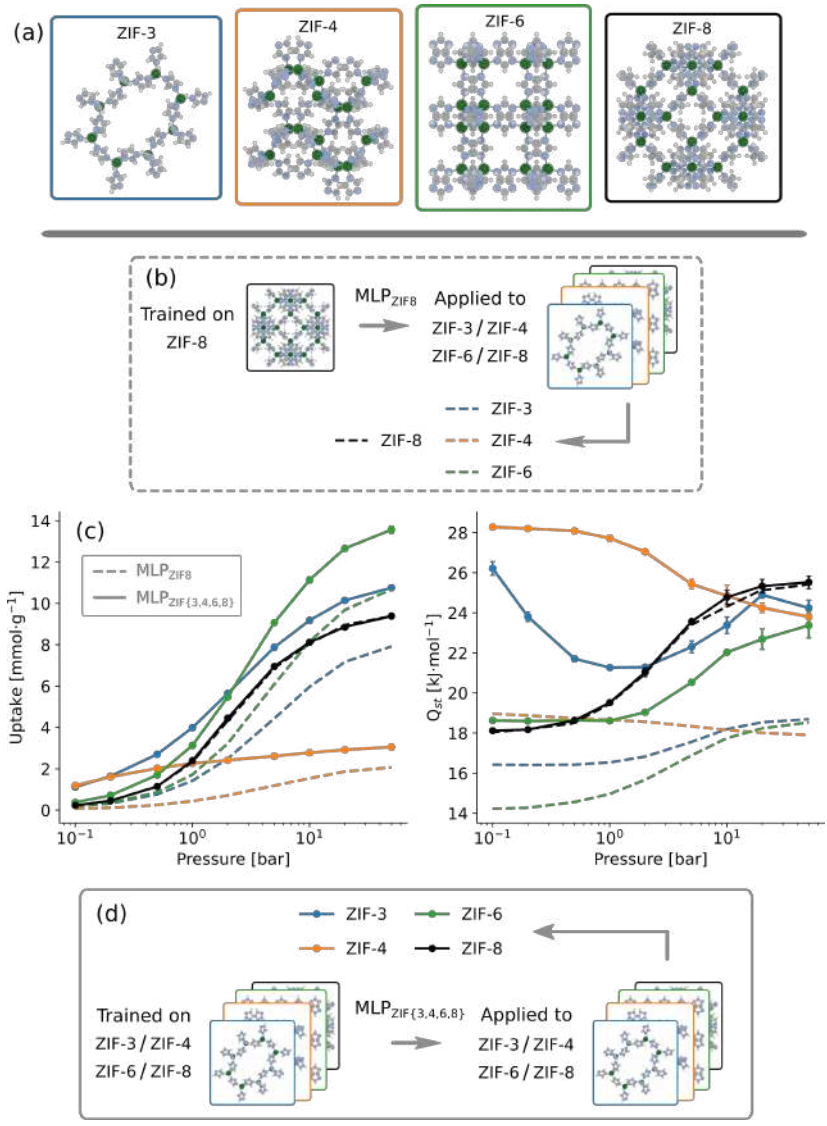


Figure 4.13: (a) Illustrations of ZIF-3, ZIF-4, ZIF-6 and ZIF-8. (b) An MLP trained on optimizations in ZIF-8 is applied to perform GCMC simulations in the other 3 ZIFs. (c) The uptake and heats of adsorption as a function of the gas pressure obtained from GCMC simulations obtained with both MLP_{ZIF8} in dotted lines and $MLP_{ZIF\{3,4,6,8\}}$ in full lines. (d) An MLP is trained on optimizations in all four ZIFs ($MLP_{ZIF\{3,4,6,8\}}$) and subsequently applied to each.

the MAE of $MLP_{ZIF\{3,4,6,8\}}$ for ZIF-8 is similar than that of a MLP_{ZIF8}, trained solely on 250 snapshots of guest-loaded ZIF-8 (equal to 0.31 $kJ \cdot mol^{-1}$). This demonstrates how learning different adsorption behaviors across different MOFs is non-trivial. It suggests that no significant information of the intermolecular interactions is shared by training to different ZIFs. Therefore, the required dataset to train an MLP to a large set of MOFs would likely scale linearly with the number of MOFs in the dataset. Only when very similar adsorption environments would be encountered between different MOFs in the dataset should this scaling decrease.

Table 4.1: Mean errors (ME) and mean absolute errors (MAE) of guest-loaded snapshots in ZIF-3, ZIF-4 and ZIF-6 extracted from GCMC simulations performed with an MLP trained on guest loadings in ZIF-8.

	ME [$kJ \cdot mol^{-1}$]	MAE [$kJ \cdot mol^{-1}$]
ZIF-3	2.29	2.38
ZIF-4	8.26	8.27
ZIF-6	0.52	0.97

Table 4.2: Mean errors (ME) and mean absolute errors (MAE) of guest-loaded snapshots in ZIF-3, ZIF-4, ZIF-6 and ZIF-8 extracted from GCMC simulations performed with an MLP trained on guest-loadings in all four ZIFs.

	ME [$kJ \cdot mol^{-1}$]	MAE [$kJ \cdot mol^{-1}$]
ZIF-3	-0.33	0.40
ZIF-4	0.21	0.39
ZIF-6	0.05	0.32
ZIF-8	-0.22	0.32

However, the issue of transferability is even more nuanced than certain errors on a test set, as a relatively acceptable MAE on the energy (as seen for ZIF-6 in Table 4.1) can deceptively hide further problems. Even with a MAE of 0.97 $kJ \cdot mol^{-1}$, the difference between heats of adsorption shown in Figure 4.13(c) for this ZIF using both MLPs is consistently higher than 4 $kJ \cdot mol^{-1}$. The reason for this discrepancy is the content of the test set. Due to the lacking ability of MLP_{ZIF8} to recognize the favorable adsorption locations in the ZIF-6 framework, the test set does not include these locations. Therefore, the test error is deceptively low, while the actual error at these sites is significantly higher. Only when explicitly training an MLP to reference data which includes these important sites is this failure alleviated. To validate whether

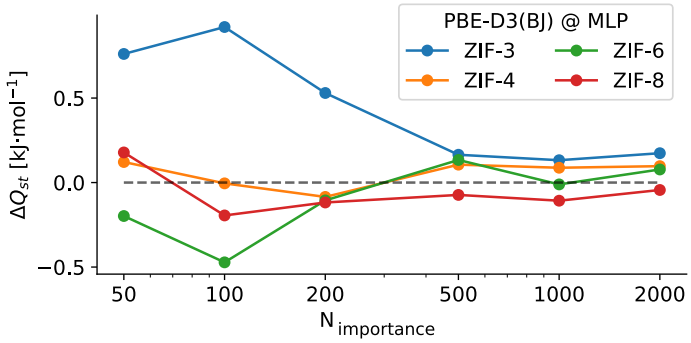


Figure 4.14: Difference between the $MLP_{ZIF\{3,4,6,8\}}$ and reference PBE-D3(BJ) heats of adsorption for all four ZIFs. MLP results are obtained from 10^6 Widom insertions, and the reference heats of adsorption are obtained from recomputing importance sampled configurations from the MLP insertions. Convergence with respect to the number of recomputed snapshots $N_{\text{importance}}$ is shown for $N_{\text{importance}}$ between 50 and 2000.

this statements holds true, it is possible to compute the heats of adsorption at infinite dilution at the reference PBE-D3(BJ) level of theory by means of an importance sampling approach. First, a set of 10^6 random insertions in each of the ZIFs are constructed for which the $MLP_{ZIF\{3,4,6,8\}}$ interaction energies are computed. Subsequently, in accordance with the procedure proposed by Vandenbrande *et al.*,[268] a set of $N_{\text{importance}}$ insertions are extracted from this wider set with a probability proportional to a Boltzmann factor. Recomputing those $N_{\text{importance}}$ snapshots, the heats of adsorption at the reference level of theory can be obtained from the following Equation

$$Q_{\text{st}} = \frac{\langle \Delta U e^{-\beta(\Delta U - \Delta \tilde{U})} \rangle}{\langle e^{-\beta(\Delta U - \Delta \tilde{U})} \rangle} \tag{4.21}$$

with $\Delta \tilde{U}$ and ΔU the interaction energies computed with the MLP and PBE-D3(BJ), respectively, $\beta = 1/k_B T$ and the ensemble average a canonical average over the MLP energies. In case the MLP accurately reproduces the reference PES, only a limited number of insertions $N_{\text{importance}}$ are required, keeping the computational cost for this procedure manageable. In Figure 4.14, the difference between the MLP and PBE-D3(BJ) importance sampled heats of adsorption obtained in this manner are shown. With $N_{\text{importance}} = 1000$, convergence within 0.10 kJ · mol^{-1} is obtained. As can be seen, the MLP heats of adsorption compare well to those computed from importance sampled PBE-D3(BJ) insertions. As can be seen, the heats of adsorption obtained with $MLP_{ZIF\{3,4,6,8\}}$ compare well to those computed from

importance sampled PBE-D3(BJ) insertions. For all four ZIFs, the error on the predicted heats of adsorption are lower than 0.20 $kJ \cdot mol^{-1}$, further validating the derived MLP.

This investigation into transferability highlights how even seemingly well-performing MLPs can fail in unforeseen ways when sampling parts of the PES on which it was not adequately trained. Such failures might also be subtly hidden when performing validations of snapshots from the performed GCMC simulations with the reference level of theory. While these failures can be detected by validation on optimizations performed at the reference DFT level of theory, this implies additional computationally demanding calculations.

However, the results obtained in Sections 4.2.1 and 4.2.2 clearly demonstrate the power of MLPs. By training on a limited set of DFT reference data, a 'best of both worlds' compromise is made between the accuracy of DFT and the computational efficiency of force fields, unlocking the potential for routine GCMC simulations at DFT accuracy. Moreover, it could be employed as a tool to benchmark computationally cheaper classical force fields where experimental reference data is not available. The proposed methodology is also generally applicable to nanoporous materials, not only to model adsorption in MOFs, but also other materials such as covalent organic frameworks (COFs), zeolites or graphene derivatives. It therefore holds clear promise to aid future accurate computational screening studies of adsorption in a wide range of nanoporous materials.

4.3 Modeling reactive events with machine learning potentials

In the preceding discussions, the term adsorption was purposefully conflated with physisorption, in which the electronic structure of the adsorbates is not strongly perturbed upon adsorption. However, a much more challenging case is that of chemisorption, in which chemical bonds can be formed and broken. A prototypical example involving chemisorption is heterogeneous catalysis in zeolites in which a reactant chemisorbs at an active site on the catalyst surface. Characterizing such processes through MD or MC simulations has traditionally been challenging due to the required accuracy of the description of the PES, as well as the length of the involved time scales. As chemisorption is associated with an activation barrier, unbiased simulation will only rarely sample the transition state between the reactant and product state. Secondly, as bond formation or breaking occurs, adequate generally applicable force field parameterizations of the PES are severely lacking. While reactive force

fields such as ReaxFF[281, 282] exist, they require manual fitting to the system at hand, usually involving expert knowledge. On the other hand, performing DFT calculations for the required number of iterations to converge MC or MD simulations is computationally extremely demanding. Again, the use of MLPs presents itself as an ideal compromise between the accuracy of DFT and the computational efficiency of force fields. In **Paper V**, the feasibility of modeling reactive events with MLPs was investigated by applying it to an archetypal proton-transfer reaction in the H-CHA zeolite. In this zeolite composed of SiO_4 tetrahedra, a fraction of the Si^{4+} ions is substituted by Al^{3+} together with a charge-compensating proton on one of the oxygens in the first coordination sphere of the Al substitution, as shown schematically in Figure 4.15(a) and 4.15(b). The charge compensating proton is however not confined to a specific oxygen in the Al tetrahedron, but can jump from one to the other, commonly known as the 'proton hopping' reaction. Although it is one of the most fundamental activated events in zeolite chemistry, its accurate computational characterization has been challenging due to the existence of nuclear quantum effects (NQEs). For the lightest nuclei such as hydrogen, a classical description of the nuclei moving on the underlying PES is known to break down, as the effect of NQEs becomes non-negligible. While techniques such as path integral molecular dynamics (PIMD) exist to model the quantum behavior of nuclei, it makes use of a number of replicas of the system. This results in an increase of the computational cost of typically more than an order of magnitude, precluding its common use in combination with DFT. However, the use of a computationally efficient MLP enables such simulations.

The computational workflow followed in **Paper V** is summarized in Figure 4.15(c). First, training data for all six hoppings between the four oxygens was obtained as training data for the MLP. For each hopping, a collective variable (CV) was defined to distinguish between the product and reactant states. Employing these CVs, a set of umbrella sampling (US) simulations were performed in which a quadratic bias potential is applied along certain values of the CVs. In this way, the whole transition path from one oxygen to another is sampled, overcoming the inefficient sampling of the transition state associated with unbiased MD simulations. Solely high temperature (873 K) classical DFT simulations with the ^1H isotope of hydrogen were performed in the gathering of training data, as these sample all regions of the PES encountered in classical MD and PIMD simulations at all considered temperatures between 273 K and 873 K for both isotopes of hydrogen (as verified *a posteriori* from validations of the MLP (PI)MD simulations).

With the extracted snapshots as training data, an MLP was trained to a validation error of 41.9 meV/Å on the forces using SchNetPack[189]. Subsequently,

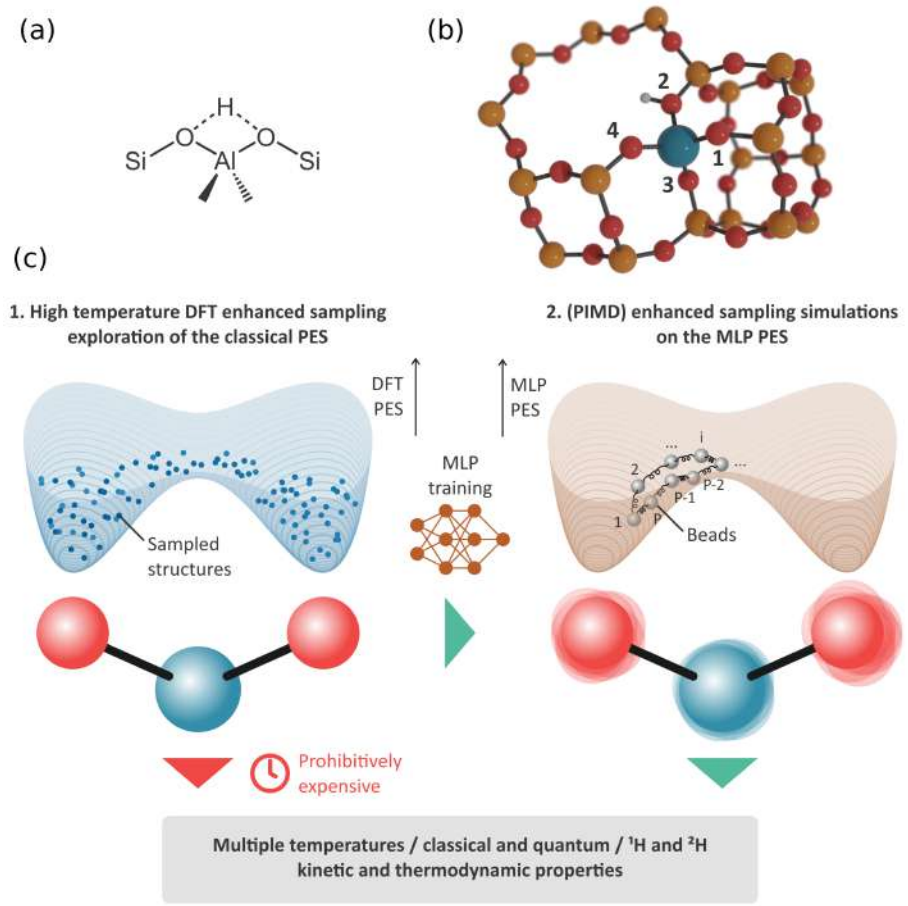


Figure 4.15: (a) Schematic depiction of the hopping reaction. (b) Part of the H-CHA unit cell with the conventional nomenclature of the oxygen atoms in the first coordination sphere of the Al defect. (c) Representation of the computational strategy. First the PES is sampled with high temperature classical DFT simulations, after which an MLP is trained on the extracted snapshots. With the computational speedup enabled by the MLP, both classical as well as path-integral MD simulations were performed at a range of temperatures, with both ^{1}H and ^{2}H isotopes of hydrogen.

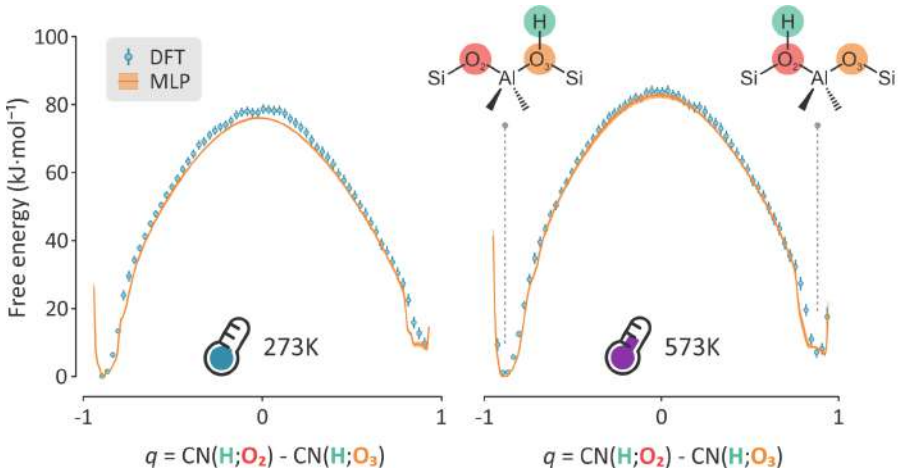


Figure 4.16: Comparison between the free energy barriers of the 2–3 hopping at 273 K and 573 K obtained with DFT US simulations (blue) and MLP US simulations (orange). The collective variable q is equal to the difference in coordination numbers (CN) of the proton with each of the oxygens.

owing to its computational efficiency, a large set of both classical MD and PIMD US simulations were performed at temperatures between 273 K and 873 K for both the ^{1}H and ^{2}H isotopes of hydrogen. As a demonstration that the MLP can indeed reproduce the classical DFT free energy profiles at temperatures other than 873 K at which the training data was gathered, an explicit comparison between both is given in Figure 4.16 for the 2–3 hopping of a proton at 273 K and 573 K, showing good agreement. Furthermore, the obtained free energy barriers were used to compute the kinetic rate constants for all six hoppings and were combined to a total overall hopping rate per second as a function of temperature. This was repeated for three separate cases; one in which an ^{1}H isotope is described classically, one in which NQEs are included and one in which an ^{2}H isotope is considered, including NQEs. All three are shown in an Arrhenius plot in Figure 4.17(a) in orange, green and purple, respectively. Activation energies were also computed from a best fit of the slope. An inclusion of NQEs is observed to significantly decrease the activation energy of the proton hopping reaction from 67.1 kJ $\cdot$ mol^{-1} to 55.8 kJ $\cdot$ mol^{-1}. The effect on the resulting speedup of the hopping reaction as a function of temperature is shown in Figure 4.17(b). As expected, the largest speedup is observed at the lowest temperatures, up to a factor of 65 at 273 K. While most zeolite-catalyzed reactions occur at elevated temperatures, our results indicate that an inclusion of NQEs is likely even important for reactions such as the aqueous cyclohexanol dehydration in H-Beta zeolite

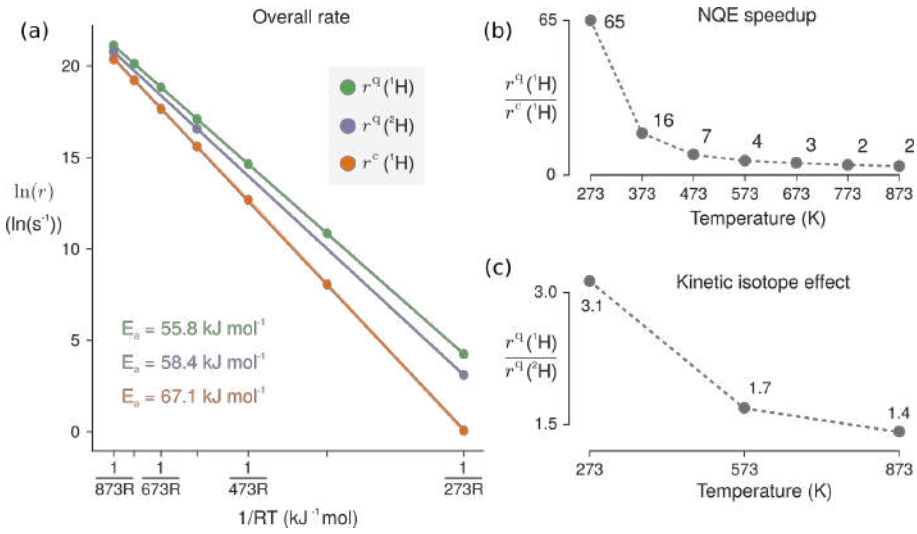


Figure 4.17: (a) Arrhenius plot of the total hopping rate of the ^{1}H isotope treated classically (orange) and with NQEs included (green) and the hopping rate of the ^{2}H isotope including NQEs (purple). Activation energies were extracted from the slope of a best fit through the data points. (b) The increase in hopping rates of the ^{1}H isotope from an inclusion of NQEs as a function of temperature. (c) The predicted kinetic isotope effect, defined as the ratio between the hopping rate of an ^{1}H isotope and an ^{2}H isotope.

at temperatures around 400 K, or the oxidation of methane to methanol in copper-loaded zeolites carried out at around 473 K.[283, 284]

While the difference between a classical and NQE-included treatment of proton-transfer reactions cannot be explicitly experimentally verified (a 'classical' proton does not exist), the difference in hopping rate between ^{1}H and the heavier ^{2}H isotope can in principle be probed by experiment. This effect, originating from both the altering of vibrational energy levels and tunneling, is commonly referred to as the kinetic isotope effect (KIE), and is shown in Figure 4.17(c). The computed ratio between both reaction rates varies from 3.1 at 273 K to 1.4 at 873 K. While no explicit experimental attempts have been made so far to measure the KIE for proton hopping in zeolites, the results are in line with theoretical considerations.[285]

These results demonstrate both the feasibility and exceeding added value of employing MLPs in describing reactive events, with applications both in catalysis as well as the characterization of chemisorption. By gathering high-

temperature classical DFT training data, a well-trained MLP unlocks the possibility of performing simulations at a wide range of different temperatures, including nuclear quantum effects, thereby strongly expanding the scope of the extracted information. MLPs can therefore be conceptualized as a 'force amplification' tool. An untrained MLP by itself does not provide any value. However, a limited set of DFT evaluations can be used to train it to 'mimic' the reference data at a vastly reduced computational cost. It therefore strongly amplifies the reach of that set of training data, enabling a whole set of new simulation tools. The future use of MLPs in molecular modeling can therefore be expected to grow quickly, likely becoming an important tool in most molecular modeler's toolboxes.

5

Conclusions and Perspectives

There is no real ending.
It's just the place where you stop the story
Frank Herbert (1920-1986)

5.1 Conclusions

Research interest in metal-organic frameworks has unequivocally flourished in the last few decades. MOFs are built up from interchangeable building units (metal-oxides and organic linkers), enabling them to be endlessly tuned for specific applications. They can be 'rationally designed'; with a target application in mind, the building units of a MOF can be selected for optimal performance. This promise of a diverse class of materials that can act as a platform for all desired gas adsorption and separation applications has, however, lead to a degree of 'cheerleading' from scientists. While the concept of rational design sounds very promising, fulfilling it is highly non-trivial. It requires detailed knowledge on the interplay between their mechanical, thermal, adsorption and separation properties, as well as how combinations of building units give rise to these properties; a daunting task. Moreover, designing MOFs that remain stable under exposure to water vapor and acid gases as would be encountered in post-combustion flue gases is far from trivial.[286–290] Therefore, to fulfill the promise of rational design, accurate and reliable computational tools are required for predictive simulations of

adsorption in MOFs. This work has aimed to contribute in sharpening these tools, approaching this task from two different angles.

A first important requirement for accurately modeling of adsorption in MOFs at operating conditions, explored in Chapter 3, is the computational sampling scheme that couples the MOF framework to the surrounding gas reservoir. While GCMC simulations have become the standard computational workhorse for this task, they neglect the framework flexibility that is inherent to many MOFs. To properly model phenomena like adsorption-induced phase transitions, which can be associated with large structural deformations, it is vital to include the flexibility of the host adsorbent in a simulation algorithm.

In Section 3.1, different computational sampling schemes combining Monte Carlo moves with molecular dynamics simulations are applied to model noble gas, methane and carbon dioxide adsorption in MIL-53(Al). Comparing the merits of these schemes, a new sampling scheme making use of volume-constrained molecular dynamics simulations is proposed in **Paper I**. With this new scheme, the relative stability of different phases of MIL-53(Al) under guest adsorption could be quantitatively predicted, demonstrating under what conditions xenon, methane and carbon dioxide can trigger phase transitions. In Section 3.2 and **Paper II**, this newly proposed scheme was applied to counter-intuitive phenomenon of negative gas adsorption (NGA) in DUT-49(Cu). In this MOF, a phase transition from the metastable large pore phase to the stable closed pore phase at a temperature of 120 K is associated with a sudden release of methane adsorbates from the framework. While previous research had highlighted how the relative stability of both phases could be altered by guest adsorption, the mechanism delaying the large pore to closed pore phase transition past the gas pressure at which the large pore remains stable was not fully understood yet. From the application of our computational scheme, we demonstrated how the system can remain kinetically trapped in the large pore phase, accurately reproducing the experimentally observed NGA step. Moreover, performing simulations at 90 K and 150 K revealed the underlying reasons why the NGA phenomenon only occurs in a small temperature window. These investigations demonstrate how modeling adsorption with sampling schemes that include framework flexibility can yield quantitative predictions of the relative stability of these materials at operating conditions, opening up their utility to the wider field of adsorption research in MOFs.

In Chapter 3, computational sampling schemes were applied to MIL-53(Al) and DUT-49(Cu) in which the guest-host interactions giving rise to adsorption could be satisfactorily captured by classical force fields. However, for many more strongly interacting adsorbates and MOFs containing open-

metal sites, the accuracy of such force fields can break down. To ensure wide applicability of these sampling schemes, more advanced methods to capture the host-guest interactions are clearly needed, as discussed in Chapter 4. To this end, two vastly different routes were explored, the 'physics' and the 'data' approach. In the first, as much physical insight as possible is incorporated in the interaction potentials. Herein, as little empirical fitting parameters as possible are used, generally leading to potentials which transfer well across a wide range of interactions. The second approach represents the polar opposite of the first. Herein, (almost) no physical insight is built into the interaction potentials. Instead, a large amount of reference (usually DFT or wavefunction-based) data is used to fit these potentials.

The 'physics' approach was discussed in Section 4.1 and **Paper III**. Herein, a transferable polarizable force field was developed which aims to model electronic polarization of adsorbates with a single transferable parameter. Benchmarking its reproduction of the dipole polarizabilities of a large set of molecules, we demonstrated how including the commonly neglected exchange interaction significantly improves the model's performance. Moreover, three, four and five-body interaction energies of water clusters could be reproduced with an accuracy equal or superior to common force fields which are explicitly fitted to water. Also including the polarization model in a complete non-covalent force field (coined PMEDFF), a performance competitive or superior to common literature force fields is observed on dimer interaction energies, even without fitting to any interaction energies of the dimers under study. While these initial results are promising, the embedding of a polarization model in a complete non-covalent force field is challenging due to the need to carefully balance the separate interaction potentials. While the polarization model developed in this work appears to improve on the accuracy of previous models, such a statement can only be convincingly proven from a comparison of total intermolecular interaction energies which include other interactions, obscuring our model's performance. In this respect, future research into the atomic anisotropy of the exchange-repulsion interaction holds promise to significantly improve our model.

A different approach to model host-guest interactions in MOFs has been enabled by the recent development of accurate and data-efficient machine learning potentials (MLPs). In this 'data' approach discussed in Section 4.2 and **Paper IV**, MLPs were trained on rigid-body DFT optimizations of CO_2 adsorbates in MOFs. Excellent reproduction of the experimental isotherms and heats of adsorption was demonstrated for ZIF-8 and the open-metal site containing Mg-MOF-74. However, for Mg-MOF-74, this performance was contingent on the proper choice of dispersion correction used in the construction of the training set. Lastly, poor transferability was observed

when applying the potential trained on ZIF-8 to other ZIFs. It was shown how significantly deviating adsorption isotherms and heats of adsorption can be obtained, even when performing satisfactorily on a test set. Only by training on all of the considered ZIFs, accurate results were obtained. This investigation showed how subtle errors can crop up when employing MLPs in adsorption simulations, highlighting the difference between empirical force fields and MLPs. While empirical force fields are limited in their accuracy, they often (are assumed to) transfer moderately well across different intermolecular interactions. On the other hand, the lack of physical insight in MLPs can lead to surprising failures when applied to adsorption environments only slightly different than those it was trained on. This apparent trade-off between accuracy and transferability will lead us into the perspectives for future research.

A important limitation of classical force fields is their difficulty in describing reactive events. Contrary to this, MLPs have no such inherent limitation. As a demonstration for this flexibility, an MLP was trained and employed to model a prototypical activated process in zeolites, namely the hopping of a proton between oxygens in the first coordination sphere of an Al substitution. As described in Section 4.3 and **Paper V**, training an MLP on high-temperature DFT umbrella sampling simulations allowed for an otherwise computationally excessively demanding path-integral MD characterization of the proton hopping reaction. Due to the reduced computational cost of the MLP compared to DFT simulations, the full temperature-dependence as well as the kinetic isotope effects (the different ^{1}H and ^{2}H isotope behavior) could be simulated. This use-case of characterizing a reactive event clearly demonstrates the 'force amplification' of MLPs in conjunction with standard DFT calculations. From a limited set of DFT evaluations, simulations can be performed at orders of magnitude lower cost, unlocking a vast suite of existing computational tools to physics and chemistry.

5.2 Perspectives

At the end of a doctoral thesis, a moment of self-reflection seems appropriate. What has yielded surprisingly good results, and what didn't work at all? How do we, as computational physicists, fulfill the promise of reliable, quantitative predictions of material properties? And which further developments are needed to bring us there? I will attempt to give a short personal view on this within the scope of adsorption in nanoporous materials, as illustrated in Figure 5.1.

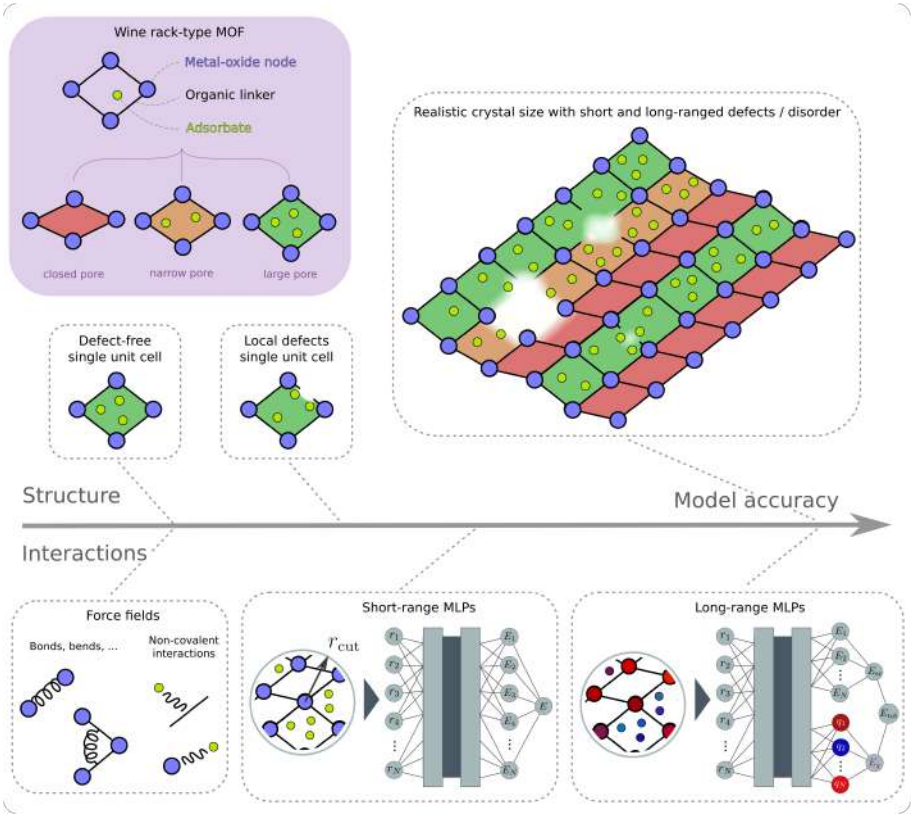


Figure 5.1: A perspective on the envisaged strides towards more accurate structural and interaction models for adsorption in MOFs, as exemplified for a wine rack-type MOF such as MIL-53(Al). Representative structural models should include both short and long-range defects and disorder encountered in MOFs. On the interaction side, the leap from a force field to short-range MLP description can be extended towards long-range interactions. From a prediction of physical interaction parameters (e.g. charges for electrostatics) these interactions can be fully captured.

The computational sampling schemes explored in Chapter 3 were consistently applied to perfectly crystalline, infinitely repeating single units cells which were assumed to be representative for the experimentally synthesized materials. However, there are plenty of indications that discrepancies exists between these idealized and the real materials. One of these is that infinitely repeating materials don't seem to exist. The finite crystal size of MOFs can have important implications. For example, reducing the DUT-49(Cu) crystal size below 1-2 μm can fully suppress the NGA transition observed for larger crystals.[291] To capture effects on these types of length scales, simulations containing on the order of 10^5 unit cells or 10^8 atoms would be required. While not inconceivable in the near term, MC/MD simulations on this scale would require carefully optimized implementations in high-performance codes such as LAMMPS or GROMACS, as well as generous allocations on petascale supercomputers.[292, 293] In the same vein, computational investigations of the effect of material defects on adsorption properties would benefit from such large-scale simulation tools, as defects can be introduced in materials at realistic concentrations.

On the potential energy side of the story, MLPs trained on DFT or wavefunction-based reference data clearly have a lot of promise. As shown in this work, highly accurate adsorption isotherms (of rigid frameworks) can be obtained with MLPs. The obvious next step forwards lies in the extending this work to incorporate the flexibility of the host framework in adsorption simulations. In this way, the developments discussed in Section 3 and Section 4 can be merged. However, this does not necessarily imply that classical force fields will be abandoned in favor of MLPs. As seen before, MLPs suffer from a lacking description of the long-range interactions (electrostatics, polarization, dispersion) between adsorbates and the host framework, limiting their degree of transferability. An integration of physically inspired interaction potentials with MLPs could alleviate this. As these interaction potentials (either with trained or empirical parameters) can generally capture long-range interactions quite well, their integration with MLPs would combine both their strong points; an accurate MLP description of short-range interactions combined with force field long-range interactions. In his review on the developments of MLPs, Behler coined this as a third generation of MLPs, improving upon the purely short-range MLPs of the second generation.[294] As shown on the right of Figure 5.1, additional to atomic energies, interaction parameters such as atomic charges can be predicted from the local environment. With these charges, a supplemental long-range classical electrostatic interaction is added to the total energy. While an improvement on the second generation of short-range MLPs, this approach is still limited by the local dependence of the atomic charges, failing

to capture nonlocal interaction that aren't (efficiently) captured by an MLP. In fourth-generation MLPs, global charge redistribution can be captured, for example by incorporation of a charge equilibration method. In this approach, local atomic electronegativities are predicted by the network.[295] Subsequently, the charge density of the system is determined by minimizing the electrostatic energy, allowing for long-range charge transfer. However, charge equilibration methods can suffer from erroneous predictions of the dipole polarizability scaling with system size, as well as assigning fractional charges to separated molecules.[180, 296, 297] Therefore, the development of our polarization model could be well-situated to fit in this fourth generation of MLPs. Local atomic polarizabilities could be predicted by the network, and employed in the interaction potential developed in Section 4.1. In any case, recent progress is undoubtedly bringing us ever closer to fulfilling the true promise of computational physics; an accurate and reliable *in silico* characterization of all material's properties your heart could ever desire.

Part II

Published Papers

A

Publications in International Peer–Reviewed Journals

Paper I **Modeling Gas Adsorption in Flexible Metal-Organic Frameworks via Hybrid Monte Carlo/Molecular Dynamics Schemes**

S.M.J. Rogge[†], R. Goeminne[†], R. Demuynck, J. J. Gutiérrez-Sevillano, S. Vandenbrande, L. Vanduyfhuys, M. Waroquier, T. Verstraelen, V. Van Speybroeck

Adv. Theory and Simul., **2019**, *2* (4), 1800177

[†] Authors contributed equally to this work

Paper II **Charting the Complete Thermodynamic Landscape of Gas Adsorption for a Responsive Metal–Organic Framework**

R. Goeminne, S. Krause, S. Kaskel, T. Verstraelen, J. D. Evans

J. Am. Chem. Soc, **2021**, *143* (11), 4143–4147

Paper III **Accurate Transferable Polarization Model Derived from the Monomer Electron Density**

R. Goeminne, T. Verstraelen

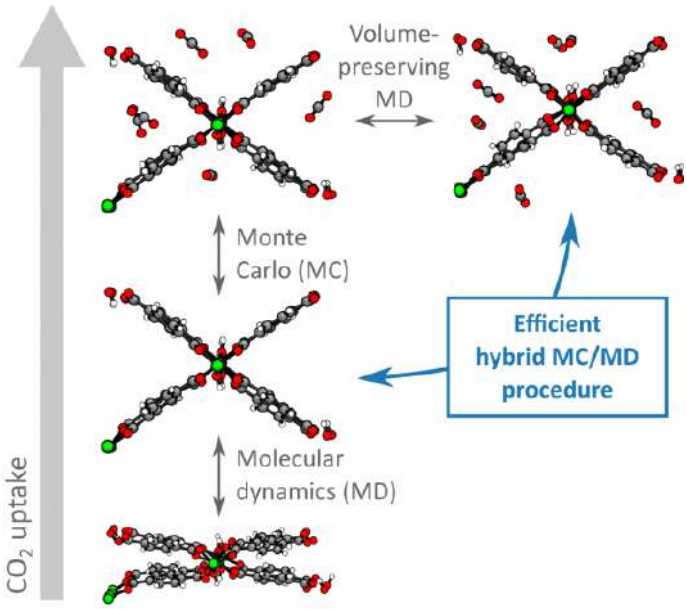
Manuscript in preparation

Paper IV **DFT-quality Adsorption Simulations in Metal-Organic Frameworks Enabled by Machine Learning Potentials**
R. Goeminne, L. Vanduyfhuys, V. Van Speybroeck, T. Verstraelen
Submitted to J. Chem. Theory Comput

Paper V **Nuclear Quantum Effects on Zeolite Proton Hopping Kinetics Explored with Machine Learning Potentials and Path Integral Molecular Dynamics**
M. Bocus†, R. Goeminne†, A. Lamaire, M. Cools-Ceuppens, T. Verstraelen, V. Van Speybroeck
Nat. Commun., **2023**, *14* (1008)

Paper I

Modeling Gas Adsorption in Flexible Metal-Organic Frameworks via Hybrid Monte Carlo/Molecular Dynamics Schemes



S.M.J. Rogge†, R. Goeminne†, R. Demuynck, J. J. Gutiérrez-Sevillano, S. Vandenbrande, L. Vanduyfhuys, M. Waroquier, T. Verstraelen, V. Van Speybroeck

Adv. Theory and Simul., **2019**, *2* (4), 1800177

† Authors contributed equally to this work.
R. Goeminne performed the simulations, interpreted the resulting data and contributed to the manuscript.

FULL PAPER
Gas Adsorption
ADVANCED THEORY AND SIMULATION
www.advtheorysimul.com

Modeling Gas Adsorption in Flexible Metal–Organic Frameworks via Hybrid Monte Carlo/Molecular Dynamics Schemes

*Sven M. J. Rogge, Ruben Goeminne, Ruben Demuynck, Juan José Gutiérrez-Sevillano, Steven Vandenbrande, Louis Vanduyfhuys, Michel Waroquier, Toon Verstraelen, and Veronique Van Speybroeck**

Herein, a hybrid Monte Carlo (MC)/molecular dynamics (MD) simulation protocol that properly accounts for the extraordinary structural flexibility of metal–organic frameworks (MOFs) is developed and validated. This is vital to accurately predict gas adsorption isotherms and guest-induced flexibility of these materials. First, the performance of three recent models to predict adsorption isotherms and flexibility in MOFs is critically investigated. While these methods succeed in providing qualitative insight in the gas adsorption process in MOFs, their accuracy remains limited as the intrinsic flexibility of these materials is very hard to account for. To overcome this challenge, a hybrid MC/MD simulation protocol that is specifically designed to handle the flexibility of the adsorbent, including the shape flexibility, is introduced, thereby unifying the strengths of the previous models. It is demonstrated that the application of this new protocol to the adsorption of neon, argon, xenon, methane, and carbon dioxide in MIL-53(Al), a prototypical flexible MOF, substantially decreases the inaccuracy of the obtained adsorption isotherms and predicted guest-induced flexibility. As a result, this method is ideally suited to rationalize the adsorption performance of flexible nanoporous materials at the molecular level, paving the way for the conscious design of MOFs as industrial adsorbents.

1. Introduction

Whether for the storage of natural gas[1–3] or hydrogen gas[3–5] in fuel cells, the capture of greenhouse gases,[3,6] or the capture and destruction of chemical warfare agents,[3,7] the rational design of new generation materials for adsorption applications requires adsorbents exhibiting a well-defined structural nanoporosity, a large internal surface area, and a high degree of tunability. In this respect, few materials can compete with the widespread attention metal–organic frameworks (MOFs) have drawn in the past three decades.[8–10] Composed of inorganic metal-oxide building blocks and organic ligands connected through relatively weak coordination bonds, MOFs form scaffold-like materials that usually exhibit a permanent nanoporosity.[11–13] Based on the concept of isoreticular synthesis, this porosity can moreover be modified by the incorporation of specific chemical functionalities into the material, either during synthesis or post-synthetically, without altering the underlying topology of the framework.[14,15] As a result of this tuneable and structural porosity, MOFs have been put forward as interesting platforms for gas and liquid adsorption with unprecedented storage and separation performances.[3,16–18] However, to enable the rational design and industrialization of MOFs as adsorbents, it is imperative to understand the fundamental interplay between the adsorbates and the host material that gives rise to this attractive behavior.[19] Given the challenges in the experimental characterization of adsorption in MOFs as well as the enormous amount of synthesized and hypothetical MOFs to be explored,[20] the development of computational techniques to reliably predict the adsorption performance of these materials is crucial as a complementary tool to experimental research.[21–24] Therefore, we introduce in this article a hybrid simulation protocol to predict adsorption isotherms for flexible MOFs with a substantial increase in accuracy compared to current state-of-the-art protocols. The key of this new model relies in the fact that both the gas adsorption process as well as the flexibility of the MOF, especially its shape, are properly taken into account.

From a computational point of view, modeling adsorption in MOFs is not a trivial task.[24–28] Owing to the open structures of these materials and the relatively weak coordination bonds between their organic and inorganic building blocks, most MOFs exhibit a degree of structural flexibility exceeding that of conventional materials. In particular, the subclass of the so-called flexible MOFs or soft porous crystals (SPCs) exhibits large-amplitude

Dr. S. M. J. Rogge, R. Goeminne, R. Demuynck,
Dr. J. J. Gutiérrez-Sevillano, S. Vandenbrande, Dr. L. Vanduyfhuys,
Prof. M. Waroquier, Prof. T. Verstraelen, Prof. V. Van Speybroeck
Center for Molecular Modeling
Ghent University
Technologiepark 46, 9052 Zwijnaarde, Belgium
E-mail: Veronique.VanSpeybroeck@UGent.be

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adts.201800177

DOI: 10.1002/adts.201800177

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

ADVANCED THEORY AND SIMULATIONS
www.advtheorysimul.com

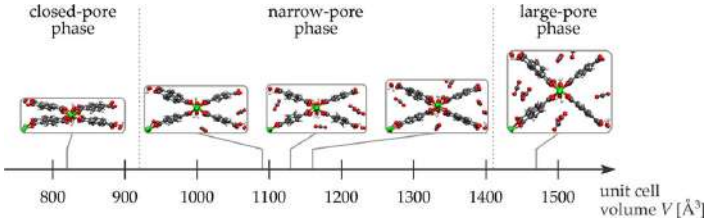


Figure 1. The closed-pore (cp), narrow-pore (np), and large-pore (lp) phases of MIL-53(Al) as encountered during CO_2 adsorption.

structural changes upon external stimuli such as pressure, guest adsorption, temperature, or light absorption.[22,29,30] While these structural changes may be local, for example, the gate opening observed in ZIF-8,[31] many flexible MOFs undergo large collective deformations upon external stimuli, such as the swelling in MIL-88[32,33] and the breathing in the MIL-53 family.[34] For instance, the typical wine-rack topology in MIL-53(Al), composed of parallel aluminum hydroxide $[Al(OH)]_\infty$ chains connected through 1,4–benzenedicarboxylate (BDC) ligands, endows this material with the topological freedom to undergo phase transitions between a large-pore (lp) and a closed-pore (cp) phase under the influence of temperature and pressure changes.[35,36] Furthermore, a narrow-pore (np) phase with an intermediate volume can be reached upon gas adsorption (see **Figure 1**).[34] To correctly model this flexibility and the corresponding activated phase transitions in MOFs, the development of sampling protocols specifically targeting the MOF's flexibility is a prerequisite.[37,38]

In this respect, some of us recently introduced a thermodynamic simulation protocol to model the mechanical stability of this challenging class of materials at operando conditions of temperature and pressure.[37,39] This simulation protocol relies on the observation that the unit cell volume V plays a vital role in distinguishing between the different phases in the majority of flexible MOFs and that their extraordinary flexibility limits the applicability of regular isothermal–isobaric molecular dynamics (MD) simulations, carried out in the (N_{host}, N_{guest}, P, $\boldsymbol{\sigma}_a = \mathbf{0}$, T) ensemble.[37,39] Therefore, we proposed an alternative sampling technique in which the so-called pressure-versus-volume equations of state of these materials are constructed in the dedicated (N_{host}, N_{guest}, V, $\boldsymbol{\sigma}_a = \mathbf{0}$, T) ensemble. In this last ensemble, the unit cell volume V rather than the pressure P is controlled next to the deviatoric stress $\boldsymbol{\sigma}_a = \mathbf{0}$, while still ensuring that the framework flexibility associated with the cell shape $\mathbf{h}_0$ is fully taken into account (see **Figure 2**b).[37,40] These pressure-versus-volume equations of state were observed to provide valuable microscopic insight into the mechanical stability of a variety of flexible MOFs, including MIL-53(Al),[19,37,38,40] DUT-49(Cu),[40,41] and DMOF-1(Zn).[19]

In a similar vein, several sampling techniques to model adsorption isotherms and adsorption-induced phase transitions in these flexible materials have been introduced,[19,42–47] which all rely on a sequence of Monte Carlo (MC) and/or MD simulation steps, as indicated in Figure 2. On the one hand, grand canonical MC (GCMC) trial moves are attractive as they allow not only to translate and rotate the adsorbates, but also to model adsorption by exchanging guest species between the adsorbent and an external reservoir, as indicated in Figure 2a.[24,47] However, as the adsorbent is assumed to remain rigid during this GCMC step, GCMC simulations are limited by the flexibility exhibited by the studied material.[48] On the other hand, MD moves such as those depicted in Figure 2b integrate the equations of motion for the guest-loaded framework as a whole, directly sampling the flexibility of the material. While these MD moves have attracted recent interest to extensively sample the flexibility in MOFs,[19,37,38] the number of adsorbed particles remains unchanged during this MD step such that adsorption isotherms cannot be directly extracted from these simulations.

Therefore, one usually adopts a well-chosen simulation protocol that combines these different MC and MD simulation steps to investigate gas adsorption in flexible materials. In the literature, three distinct simulation protocols prevail, which are schematically depicted as Schemes 1 to 3 in **Figure 3**a–c. A first scheme is based on conventional GCMC simulations, which were traditionally employed to model gas adsorption in rigid porous materials but neglect the potential flexibility of the material.[24,47] This shortcoming is partially addressed in Scheme 2, a hybrid MC/MD procedure in which the GCMC simulations are sequentially followed by short MD runs.[42–44] This allows the framework to deform under guest adsorption, and was successfully adopted to predict the structural transformations of MIL-53(Cr) under CO_2 adsorption.[49] However, the method may be expected to fail when encountering high free energy barriers between the different metastable states of the host material, such as for the flexible MIL-53(Al). Finally, a third scheme was recently introduced by some of the present authors, and consists of MD simulations in the (N_{host}, N_{guest}, V, $\boldsymbol{\sigma}_a = \mathbf{0}$, T) ensemble, fixing the unit cell volume but allowing the cell shape to fluctuate.[19,37] The pressure-versus-volume equations of state obtained within this scheme can be integrated to obtain free energy profiles, which enables one to examine the evolution of the various (meta)stable states as a function of the varying guest loading. However, this method has until now not been adopted in a hybrid MC/MD simulation protocol and it is therefore unclear whether it may be used to further increase the accuracy of the obtained adsorption isotherms.

In this article, we therefore introduce a new hybrid MC/MD simulation protocol combining MC steps with MD trajectories in the (N_{host}, N_{guest}, V, $\boldsymbol{\sigma}_a = \mathbf{0}$, T) ensemble (Scheme 4 in Figure 3d), which is specifically designed to handle flexible adsorbents. The performance of this newly developed hybrid MC/MD scheme to predict gas adsorption isotherms in flexible adsorbents such as MOFs and associated adsorption-induced phase transitions is critically assessed and compared to current

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

ADVANCED THEORY AND SIMULATIONS
www.advtheorysimul.com

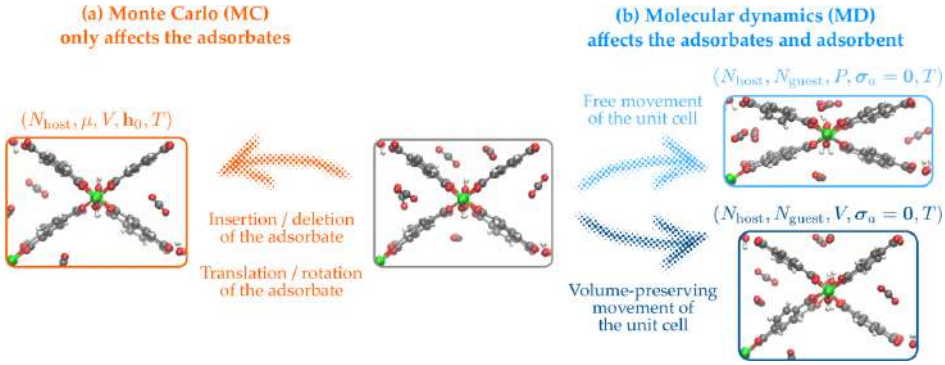


Figure 2. a) Monte Carlo (orange) and b) molecular dynamics (blue) simulation steps adopted to sample guest adsorption in flexible materials. For the MD steps, a distinction is made between steps that preserve the unit cell volume (dark blue) and those that affect the unit cell volume (light blue).

state-of-the-art simulation protocols embedded in Schemes 1–3 of Figure 3a–c. Its applicability is illustrated by modeling the adsorption of the noble gases neon, argon, and xenon, as well as methane and carbon dioxide, in the flexible MIL-53(Al). This choice of adsorbates is motivated by the experimentally observed breathing of MIL-53(Al) under the influence of xenon,[50] methane,[51] and carbon dioxide.[51,52] A critical comparison of the four schemes highlights the superior performance of the newly introduced scheme. As a result, this newly derived scheme is ideally suited for the further computational exploration of flexible MOFs as interesting adsorbents for a variety of gas adsorption applications.

2. Hybrid Monte Carlo/Molecular Dynamics Schemes

The four simulation protocols discussed in this article are schematically outlined in Figure 3. Before comparing their performance in the prediction of adsorption isotherms and adsorption-induced flexibility of the host material, the theory underlying these methods is briefly summarized here. In this discussion, we adopt the notation introduced in ref. [37] to uniquely define the different thermodynamic ensembles for flexible framework materials such as MOFs. In this notation, the external stress $\boldsymbol{\sigma}$ is split into an isotropic component, the pressure $P = \text{Tr}(\boldsymbol{\sigma})/3$, and an anisotropic component, the deviatoric stress $\boldsymbol{\sigma}_a = \boldsymbol{\sigma} - P\mathbf{1}$. Likewise, the full cell matrix $\mathbf{h}$, containing the three cell vectors defining the periodic boundary conditions, is factorized into the cell volume $V = \det(\mathbf{h})$ and the cell shape $\mathbf{h}_0$ such that $\mathbf{h} = V^{1/3}\mathbf{h}_0$.

2.1. Scheme 1: GCMC Simulations with a Rigid Unit Cell in the Restricted Osmotic Ensemble

The first scheme, depicted in Figure 3a, finds its roots in conventional GCMC simulations. These simulations, carried out in the $(N_{\text{host}}, \mu, V, \mathbf{h}_0, T)$ ensemble, were traditionally employed to model gas adsorption in rigid porous media such as rigid MOFs at a given chemical potential μ and temperature T.[24,47] To also simulate flexible materials, exhibiting multiple (meta)stable states, this original protocol was embedded into Scheme 1, in which a series of these GCMC simulations are performed in parallel for a limited set of different adsorbed configurations (see Figure 3a).[45] For many flexible materials, such as MIL-53(Al), these configurations can be defined by the unit cell volume of the material. As indicated in Figure 3a, Scheme 1 gives rise to adsorption isotherms in the so-called restricted osmotic ensemble.[21,45,46] As the cell shape $\mathbf{h}_0$ is kept fixed during these simulations, these isotherms will be referred to as rigid-host isotherms $N_{\text{guest}}(N_{\text{host}}, \mu, \mathbf{h}_0, T; V)$. These rigid-host isotherms can afterward be transformed toward the osmotic potential by thermodynamic integration of the isotherms with respect to the chemical potential:

$$\Omega(N_{\text{host}}, p, \mathbf{h}_0, T; V) = F_{\text{host}}(N_{\text{host}}, T; V) + pV - \int_{-\infty}^{\mu(p,T)} N_{\text{guest}}(N_{\text{host}}, \mu', \mathbf{h}_0, T; V)\mathrm{d}\mu' \quad (1)$$

A similar procedure was followed in the derivation of the OFAST model by Coudert and co-workers.[21,45,46] From this osmotic potential, the (meta)stable phases of the system at a given temperature and gas pressure as well as its adsorption capacity can be extracted (see Figure 3a).[21]

2.2. Scheme 2: Hybrid MC/MD Simulations in the Osmotic Ensemble

In the second scheme, depicted in Figure 3b, the osmotic or $(N_{\text{host}}, \mu, P, \boldsymbol{\sigma}_a = \mathbf{0}, T)$ ensemble is sampled.[42–44] In this thermodynamic ensemble,[53] not only the number of adsorbed guest molecules may vary during the simulation, as in Scheme 1, but also the framework is allowed to deform under guest adsorption and external stress $\boldsymbol{\sigma} = P\mathbf{1} + \boldsymbol{\sigma}_a$. To sample the osmotic

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

ADVANCED THEORY AND SIMULATIONS
www.advtheorysimul.com

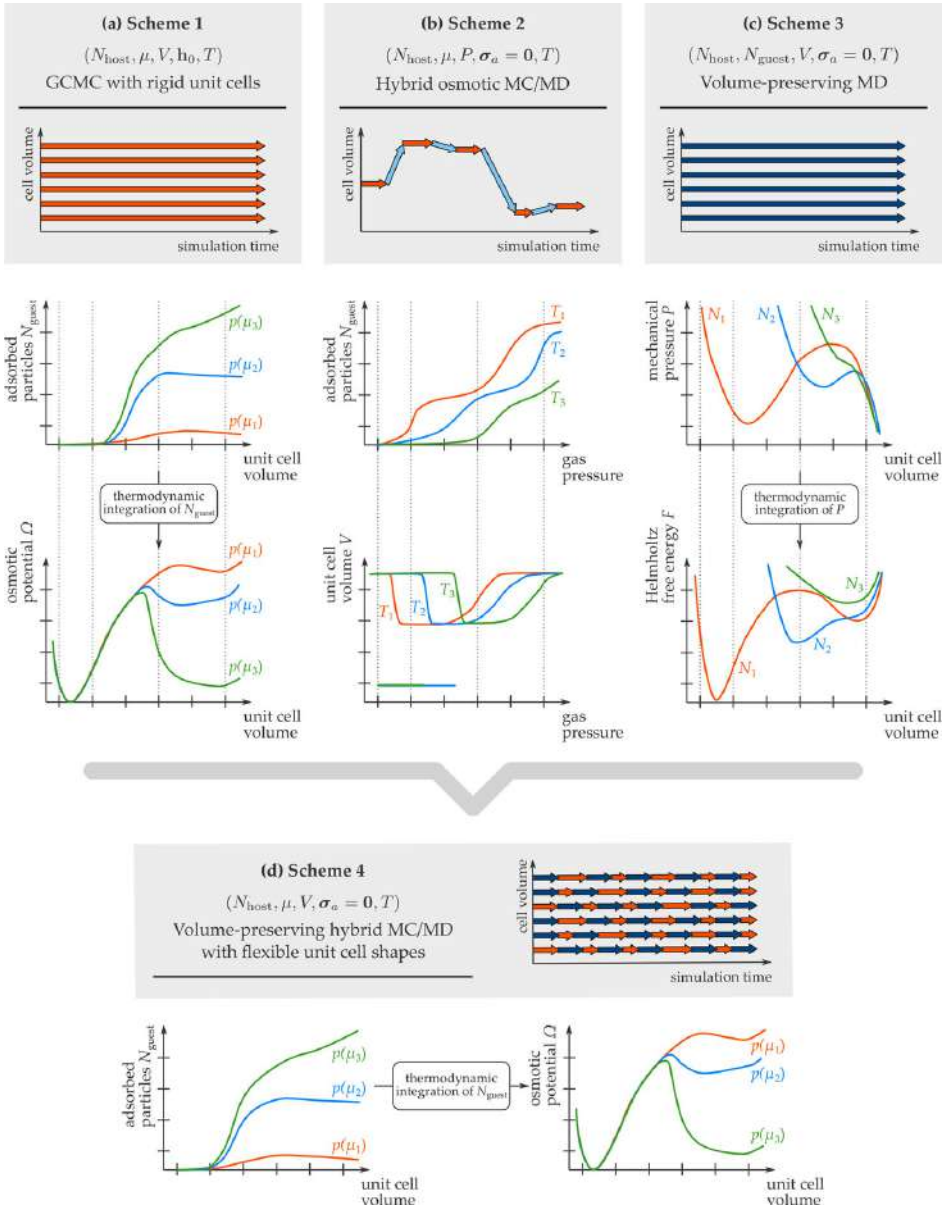


Figure 3. Schematic overview of the different combinations of MD and MC building blocks of Figure 2 that lead to the four simulation schemes discussed in this work and the thermodynamic information that can be extracted from them.

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

ADVANCED THEORY AND SIMULATIONS
www.advtheorysimul.com

ensemble, a hybrid MC/MD procedure is considered, depicted schematically in Figure 3b. Besides the translation, rotation, insertion, and deletion of the guest molecules considered in the GCMC steps (orange arrows in Figure 3b), which are again carried out in the (N_{host}, $\boldsymbol{\mu}$, V, $\mathbf{h}_0$, T) ensemble, also short MD simulations in which the guest–host system is propagated in time according to the (N_{host}, N_{guest}, $P = p$, $\boldsymbol{\sigma}_a = \mathbf{0}$, T) ensemble (light blue arrows in Figure 3b) are considered as MC trial moves in this hybrid scheme.[42–44] Here, the external pressure P is controlled to equal the gas pressure p, which is related to the chemical potential $\boldsymbol{\mu}$ through the van der Waals equation of state. This hybrid (osmotic) Monte Carlo (HOMC) approach was successfully adopted before to qualitatively predict the structural transitions and adsorption isotherms of MIL-53(Cr) upon CO_2 adsorption.[49] However, the presence of unsurmountable free energy barriers between the different adsorption-induced metastable states of the host material as well as large volume fluctuations of the MOF unit cell during these simulations may limit the accuracy of this method.[37,38]

2.3. Scheme 3: Advanced MD Simulations of the Guest-Loaded Framework at Constant Volume

A third method to predict guest-induced flexibility in MOFs was recently introduced by some of the present authors.[19,37] As shown in Figure 3c, this method relies on a series of MD simulations in the (N_{host}, N_{guest}, V, $\boldsymbol{\sigma}_a = \mathbf{0}$, T) ensemble that only fix the cell volume while allowing the cell shape to fluctuate.[37,40] In these simulations, the cell volume V is controlled, whereas the cell shape $\mathbf{h}_0$ is allowed to fluctuate so as to counteract, on average, the zero deviatoric stress $\boldsymbol{\sigma}_a = \mathbf{0}$. At equilibrium, the average internal pressure during such a simulation indicates the pressure the material can withstand at the given volume. By repeating this procedure for a range of volumes, the finite-temperature pressure-versus-volume equation of state at a given guest loading, $P(N_{\text{guest}}, T; V)$, can be extracted (see Figure 3c). From these pressure profiles, the corresponding Helmholtz free energy profiles $F(N_{\text{guest}}, T; V)$—the finite-temperature analogue of the energy-versus-volume equations of state—can be derived via thermodynamic integration:[37,54]

$$F(N_{\text{guest}}, T; V) - F(N_{\text{guest}}, T; V_{\text{ref}}) = -\int_{V_{\text{ref}}}^{V} P(N_{\text{guest}}, T; V')\,\mathrm{d}V' \quad (2)$$

By determining the (local) minima in the free energy profile, this method identifies the different (meta)stable states at a given guest loading even if they are separated by substantial free energy barriers. While this method was originally introduced to investigate pressure-induced phase transitions in empty MOFs,[37] the procedure was extended in ref. [19] to also account for adsorption-induced transitions. In this modified procedure, the (N_{host}, N_{guest}, V, $\boldsymbol{\sigma}_a = \mathbf{0}$, T) simulations are carried out not only for the empty framework ($N_{\text{guest}} = 0$), but also for the adsorbate-loaded framework ($N_{\text{guest}} > 0$). By tracking the different metastable states as a function of the guest loading, this method allows for the identification of adsorption-induced flexibility in MOFs. Furthermore, the osmotic potential can be obtained by considering the Legendre transform of the Helmholtz free energy profile, as outlined in more detail in ref. [55].

2.4. Scheme 4: Hybrid MC/MD Simulations with a Flexible Unit Cell in the Restricted Osmotic Ensemble

To circumvent the disadvantages associated with the previous three schemes, a hybrid MC/MD procedure is introduced in this article. This new hybrid procedure, depicted as Scheme 4 in Figure 3d, samples the (N_{host}, $\boldsymbol{\mu}$, V, $\boldsymbol{\sigma}_a = \mathbf{0}$, T) ensemble. To this end, one starts from the procedure outlined in Scheme 1, which samples the rigid-host or (N_{host}, $\boldsymbol{\mu}$, V, $\mathbf{h}_0$, T) ensemble. To go beyond this approach and ensure a proper sampling of the atomic positions and the cell shape, an MD trajectory is considered as an additional MC trial move in Scheme 4, akin to Scheme 2. However, while the MD trajectory in Scheme 2 is performed in the (N_{host}, N_{guest}, $P = p$, $\boldsymbol{\sigma}_a = \mathbf{0}$, T) ensemble, here, the (N_{host}, N_{guest}, V, $\boldsymbol{\sigma}_a = \mathbf{0}$, T) ensemble of Scheme 3 is sampled, so that the volume remains conserved while fully sampling the cell shape and atomic positions. This hybrid MC/MD scheme therefore results in flexible-host isotherms, $N_{\text{guest}}(N_{\text{host}}, \boldsymbol{\mu}, \boldsymbol{\sigma}_a = \mathbf{0}, T; V)$, in contrast to the rigid-host isotherms of Scheme 1. These isotherms can then be transformed to the flexible-host osmotic potential:

$$\Omega(N_{\text{host}}, p, \boldsymbol{\sigma}_a = \mathbf{0}, T; V) = F_{\text{host}}(N_{\text{host}}, T; V) + pV - \int_{-\infty}^{\mu(p,T)} N_{\text{guest}}(N_{\text{host}}, \boldsymbol{\mu}', \boldsymbol{\sigma}_a = \mathbf{0}, T; V)\,\mathrm{d}\boldsymbol{\mu}' \quad (3)$$

3. Results and Discussion

3.1. Scheme 1: GCMC Simulations with a Rigid Unit Cell in the Restricted Osmotic Ensemble

3.1.1. Noble Gas Adsorption

Given the experimentally observed breathing of MIL-53(Al) under the influence of xenon adsorption at room temperature,[50] the 300 K rigid-host isotherms of Scheme 1 were first constructed for neon, argon, and xenon. The resulting adsorption isotherms are visualized in **Figure 4**a,b and Figure S1, Supporting Information. At the lowest MIL-53(Al) cell volumes, no uptake is observed for any of the three guest species. Adsorption only starts to take place once the cell volume is increased above 950, 1025, and 1100 Å^3 for neon, argon, and xenon, respectively. At these volumes, the diameter of the MIL-53 channels starts to exceed the kinetic diameters of these guest species, as indicated in Figure S3, Supporting Information, allowing for the adsorbates to be accommodated in the channels.

However, even for sufficiently large unit cell volumes, the uptake of neon and argon in MIL-53(Al) remains small, amounting to only 0.4 atoms (neon) or 1.5 atoms (argon) per unit cell. This low uptake can be explained by the low adsorption energies for

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

ADVANCED THEORY AND SIMULATIONS
www.advtheorysimul.com

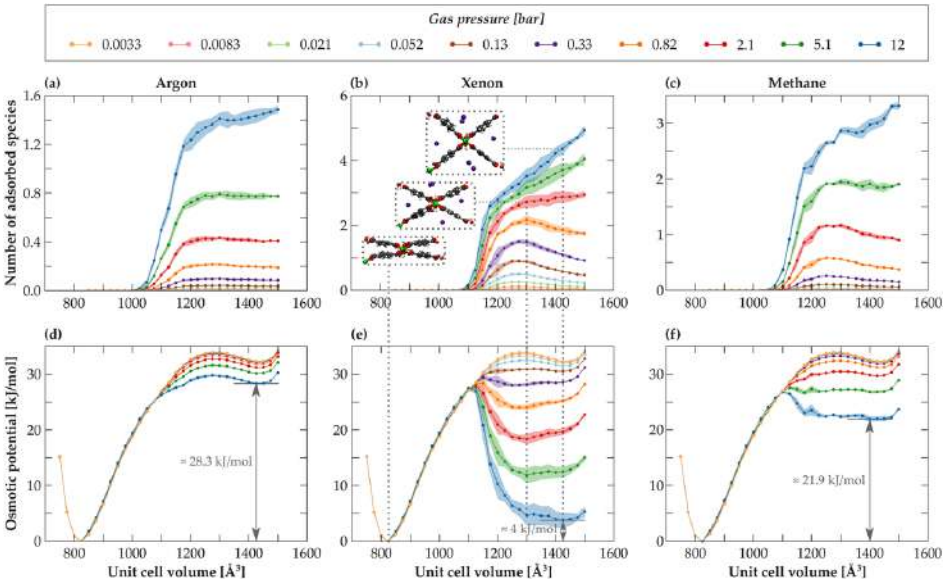


Figure 4. Rigid-host adsorption isotherms for a) argon, b) xenon, and c) methane as a function of the MIL-53(Al) unit cell volume V at 300 K for a range of pressures between 3.3×10^{-3} and 12 bar. d–f) Corresponding osmotic potentials Ω as a function of the unit cell volume V for these three guest species. Shaded areas denote the 1σ uncertainty interval as determined over five independent simulations.

both gases in MIL-53(Al), as illustrated in Figure S4, Supporting Information. As a result of this low uptake, the corresponding osmotic potentials, shown in Figure 4d, exhibit two minima. These minima correspond with a stable, guest-free closed-pore (cp) phase at small volumes, which does not change with increasing gas pressure as no guests are adsorbed, and a metastable lp phase that is significantly less stable. Since the adsorption of neon nor argon is able to significantly affect the relative stability of the empty-host (meta)stable states of MIL-53(Al), Scheme 1 predicts that these guest species do not have the potential to induce structural transitions in MIL-53(Al) at room temperature.

For xenon, in contrast, Scheme 1 predicts an appreciable gas uptake. Depending on the gas pressure or, equivalently, the chemical potential, two types of xenon isotherms may be distinguished in Figure 4b. For low to intermediate pressures, the uptake reaches a maximum for intermediate states with a cell volume around 1250–1300 Å^3. At this volume, the diameter of the MIL-53(Al) 1D channel amounts to about 5 Å (see Figure S3, Supporting Information), which is slightly larger than the kinetic diameter of xenon. This ensures that each xenon atom can interact favorably with the organic linkers delimiting the 1D channel, thereby minimizing the adsorption energy (see insets in Figure 4b and Figure S4, Supporting Information). Consequently, xenon adsorption leads to the stabilization of a xenon-filled narrow-pore (np) phase, inducing a local minimum in the osmotic potentials of Figure 4e at this volume. At a gas pressure of 5.1 bar, for instance, xenon induces a np metastable minimum at a volume of about 1300 Å^3.

On further increasing the gas pressure, additional xenon atoms are forced inside the framework. As the xenon uptake at these highest pressures can only be increased further by increasing the unit cell volume, the adsorption isotherms at these high gas pressures no longer show a local maximum, but rather continue to increase with increasing cell volume. This is also revealed in the osmotic potentials, in which the np metastable minimum disappears in favor of the lp metastable minimum at ≈1425 Å^3, which is only about 4 kJ mol^{-1} less stable than the cp phase at a gas pressure of 12 bar. The xenon-filled np and lp phases can both be stabilized further by decreasing the temperature, as discussed in Figure S2, Supporting Information, and the lp phase becomes the most stable phase at a gas pressure of 12 bar for temperatures below 250 K. Scheme 1 therefore correctly predicts that xenon may induce a phase transition in MIL-53(Al), in qualitative agreement with experiment.[50] However, the inaccuracy associated with Scheme 1 prohibits a quantitative prediction of the adsorption-induced lp volume. For instance, at a xenon gas pressure of 12 bar, all volumes in between 1300 and 1475 Å^3 lie within the 1σ uncertainty interval of the lp state.

3.1.2. Methane and Carbon Dioxide Adsorption

Experimentally, methane adsorption only induces phase transitions in MIL-53(Al) for temperatures lower than about 250 K, whereas carbon dioxide has been shown to induce phase transitions for temperatures up to 350 K.[51,52] When constructing the

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

ADVANCED THEORY AND SIMULATIONS
www.advtheorysimul.com

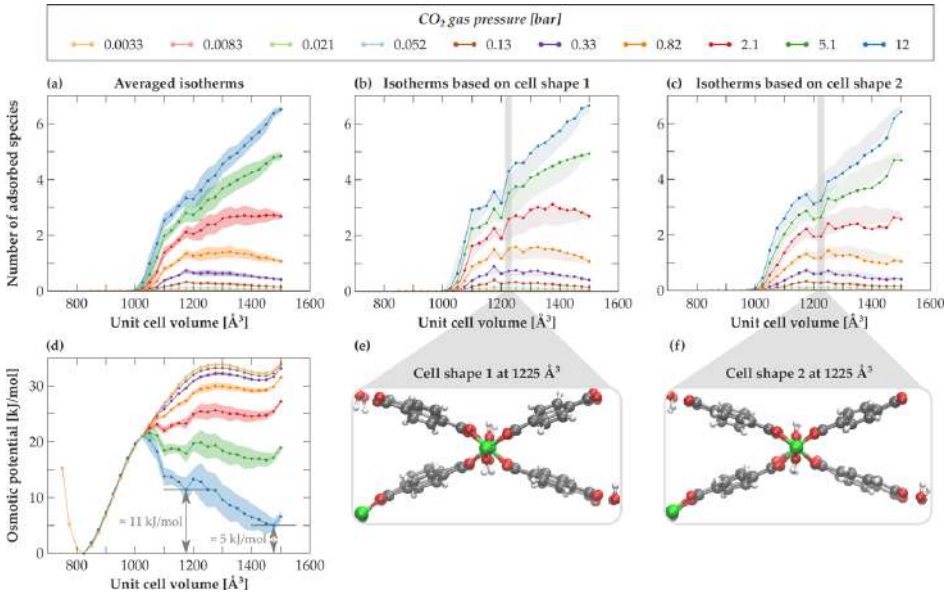


Figure 5. Rigid-host adsorption isotherms for carbon dioxide as a function of the MIL-53(Al) unit cell volume V at 300 K for a range of pressures between 3.3×10^{-3} and 12 bar, obtained by b,c) considering two different fixed cell shapes $\mathbf{h}_0$ and a) averaged over GCMC simulations starting from five different cell shapes. d) Corresponding osmotic potential Ω as a function of the unit cell volume V for the averaged isotherms. e–f) The two MIL-53(Al) cell structures at 1225 Å^3 that give rise to the adsorption capacity shown in grey in panes (b) and (c). Shaded areas in panes (a) and (d) denote the 1σ uncertainty interval as determined over five independent simulations, and are repeated in panes (b) and (c) as a guide for the eye.

methane rigid-host isotherms, shown in Figure 4c, one indeed observes a behavior in between that of argon and xenon. While an intermediate maximum in the methane rigid-host isotherms is found for low to intermediate gas pressures, similar to xenon adsorption, the total uptake is only half of that of xenon at the same gas pressure. As a result, the corresponding osmotic potentials of Figure 4f show an np or lp metastable minimum that is about 22 kJ mol^{-1} less stable than the cp global minimum, so that one expects the cp phase to be retained during methane adsorption at 300 K. However, also here the 1σ uncertainty intervals limit the accuracy with which this np or lp volume can be determined.

For carbon dioxide, the rigid-host isotherms and corresponding osmotic potentials, depicted in panes a) and d) of **Figure 5**, respectively, are similar to those observed for xenon. The intermediate np metastable minimum for carbon dioxide adsorption occurs at a volume of about 1100–1200 Å^3, which is lower than the volume of the np phase induced by xenon adsorption (Figure 4e). This is in line with the smaller kinetic diameter of carbon dioxide with respect to xenon as well as the minimum in adsorption energy appearing at a lower volume than for xenon in Figure S4, Supporting Information. For the highest gas pressures, the lp phase is again only 5 kJ mol^{-1} less stable than the cp phase, so that one expects that both phases may be encountered during carbon dioxide adsorption in MIL-53(Al) at room temperature, in qualitative correspondence with experimental literature.[51,52]

However, when contrasting the carbon dioxide rigid-host isotherms and osmotic potentials to those obtained for the other adsorbates, an even larger variability and therefore inaccuracy is apparent for the former as indicated by the larger 1σ uncertainty intervals. To investigate the origin of this larger variability, panes b) and c) of Figure 5 depict the rigid-host isotherms obtained by considering only one cell shape per unit cell volume instead of averaging over five cell shapes as for pane a). Comparing panes b) and c), one observes that the two different cell shapes give rise to adsorption capacities that differ significantly for some unit cell volumes. At a volume of 1225 Å^3 (shaded in grey) and at a pressure of 12 bar, for instance, the two simulations predict an average uptake of, respectively, 4.3 and 3.2 carbon dioxide molecules per unit cell, a difference of about 25%. Note that this difference does not arise from imprecision, as Section S1.4, Supporting Information indicates that the results are well converged. Rather, the observed discrepancy is only associated with insufficient sampling of the adsorption-induced flexibility of the framework in the (N_{host}, $\boldsymbol{\mu}$, V, $\mathbf{h}_0$, T) ensemble of Scheme 1. Indeed, visualizing the two MIL-53(Al) structures at 1225 Å^3 in panes e) and f) of Figure 5 shows a seemingly slight rotation of the phenyl moieties in the top right corner, which allows for a substantially more favorable interaction with the adsorbed carbon dioxide molecules for cell shape 1 and therefore an increased gas uptake. This observation confirms that the rigid-host simulations of Scheme 1 are too crude to accurately sample adsorption in flexible MOFs, as it

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

ADVANCED THEORY AND SIMULATIONS
www.advtheorysimul.com

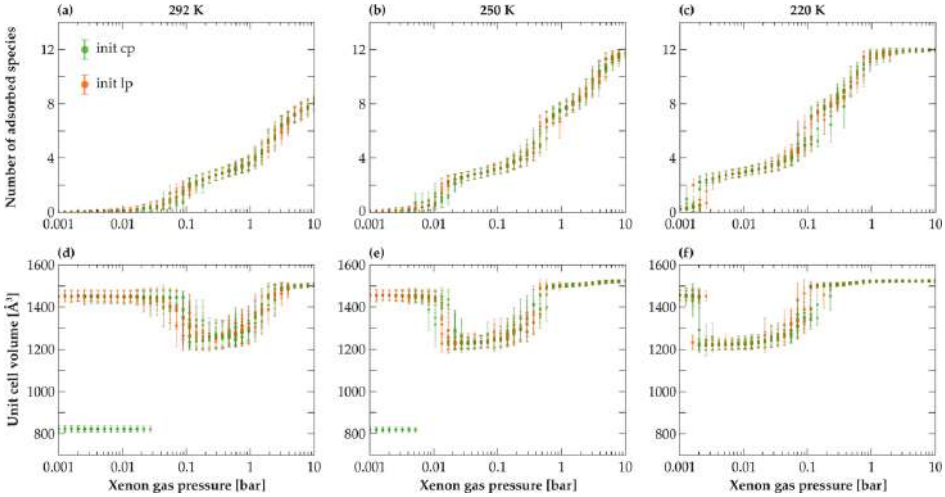


Figure 6. a–c) Predicted number of adsorbed xenon atoms N_{Xe} and d–f) MIL-53(Al) unit cell volumes V as a function of the xenon gas pressure p for three different temperatures: (a,d) $T = 292$ K, (b,e) $T = 250$ K, and (c,f) $T = 220$ K. At each temperature and gas pressure, ten simulations were performed: five starting in the cp state (green) and five starting in the lp state (orange). Error bars indicate the 1σ standard deviations of each separate simulation.

is in general impossible to quantitatively determine the volume of the adsorption-induced (meta)stable phases.

3.2. Scheme 2: Hybrid MC/MD Simulations in the Osmotic Ensemble

3.2.1. Noble Gas Adsorption

Figure 6 displays the results of Scheme 2 for xenon adsorption at three different temperatures. At 292 K, one observes two steps in the isotherm of Figure 6a. For the lowest pressures, up to about 0.03 bar, the xenon uptake is negligible. Correspondingly, the predicted (meta)stable states in Figure 6d are the (empty) cp and lp states. For these lowest gas pressures, the phase observed during the simulation depends on the initialization: simulations initialized in the lp or cp state remain in the lp or cp state, respectively, as illustrated by the ten independent simulations shown in Figure 6. While, in theory, phase transitions between the lp and cp phases can be observed in this scheme, these phase transitions are rare events that are unlikely to occur during the simulation as both phases are separated by a high free energy barrier.

When increasing the gas pressure above 0.03 bar at 292 K, a first step in the adsorption isotherm is apparent, as xenon starts to adsorb in the material at this pressure. These adsorbates steer the framework from the cp or lp phase to the np phase with a volume of about 1250–1300 Å^3, as indicated in Figure 6d. This observation is in excellent agreement with the xenon-loaded np volume extracted from Scheme 1 (Figure 4e). When further increasing the gas pressure, the uptake increases continuously to about four xenon atoms per unit cell, until reaching a pressure of 1–2 bar. At this pressure, a second step in the isotherm is observed, corresponding to a phase transition from the xenon-loaded np phase to the xenon-loaded lp phase. At the maximum pressure of 10 bar considered here, a maximum loading of eight xenon atoms per unit cell is reached at 292 K.

When decreasing the temperature in Figure 6 to 250 K (panes b) and e)) and 220 K (panes c) and f)), the predicted isotherms and unit cell volumes as a function of the gas pressure retain their 292 K shapes. Decreasing the temperature only results in a decrease in the critical pressures necessary for the cp-to-np and the np-to-lp transitions, confirming the observations in Figure S2, Supporting Information that decreasing the temperature facilitates xenon-induced transitions in MIL-53(Al).

For neon and argon, no phase transitions are observed, as shown in Figures S7 and S8, Supporting Information, confirming the results of Scheme 1. For both adsorbates and at both temperatures, the MIL-53(Al) unit cell remains in the initial cp or lp state, irrespective of the gas pressure, confirming that transitions between both phases are activated processes.

3.2.2. Methane and Carbon Dioxide Adsorption

Figure 7a reveals a continuous methane adsorption isotherm at 292 K. The corresponding cell volumes in Figure 7d indeed show that MIL-53(Al) retains its original phase upon adsorption, with no intermediate np phase, in qualitative agreement with the literature.[51,52] However, at methane pressures between 0.2 and 2 bar, one observes that the simulations initialized in the lp phase already start to explore smaller volumes during the simulation,

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

ADVANCED THEORY AND SIMULATIONS
www.advtheorysimul.com

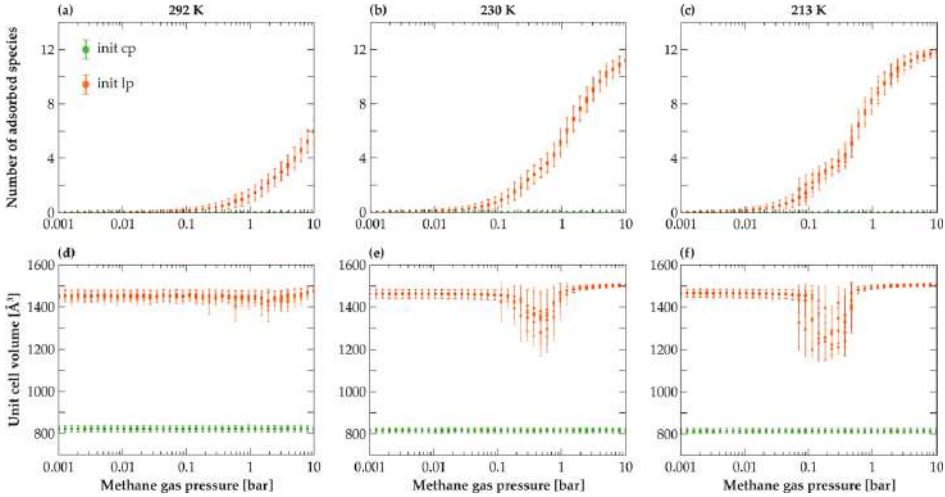


Figure 7. a–c) Predicted number of adsorbed methane molecules N_{CH_4} and d–f) MIL-53(Al) unit cell volumes V as a function of the methane gas pressure p for three different temperatures: (a,d) $T = 292$ K, (b,e) $T = 230$ K, and (c,f) $T = 213$ K. At each temperature and gas pressure, ten simulations were performed: five starting in the cp state (green) and five starting in the lp state (orange). Error bars indicate the 1σ standard deviations of each separate simulation.

with the lower limit of the 1σ uncertainty interval extending to about 1325 Å^3. By lowering the temperature to 230 K, the same stepwise behavior as encountered with xenon adsorption is observed. At this temperature, methane induces an lp-to-np transition at a pressure of ≈0.2 bar, while an np-to-lp transition is induced at a pressure of ≈1 bar. However, while Scheme 1 indicated that the methane-induced np phase would occur near a volume of about 1200 Å^3, the unit cell volumes at intermediate methane gas pressures vary here between 1200 Å^3 and 1500 Å^3. Visualizing the unit cell volume throughout the hybrid MC/MD procedure at such an intermediate gas pressure reveals spontaneous phase transitions between the np phase, at a volume of about 1200 Å^3, and the lp phase, at a volume of about 1475 Å^3 (see Figure S10, Supporting Information). These spontaneous phase transitions indicate that both the np and lp phases are (meta)stable states at this gas pressure, with a very small free energy barrier separating them, resulting in a large variation in the sampled unit cell volume during such a simulation. This impedes an accurate determination of the volumes associated with these (meta)stable states.

At 213 K, the stepwise behavior of the methane isotherm is even more pronounced. Both critical pressures necessary to induce a phase transition decrease with respect to their 230 K values. At this lower temperature, the frequency of the np-to-lp and lp-to-np phase transitions during a single simulation as observed at intermediate gas pressures also decreases due to an increased free energy barrier between both states and a lower thermal energy of the atoms. Furthermore, neither at 230 K nor at 213 K a cp-to-np transition is observed at these intermediate gas pressures, despite the metastability of the np phase. This is a direct consequence of the large energy barrier between both states, preventing these transitions from occurring spontaneously during the simulation.

For carbon dioxide, visualized in Figure S9, Supporting Information, two-step isotherms akin to the ones observed for xenon are obtained, in agreement with Scheme 1. Also here, decreasing the temperature results in a decrease in the critical gas pressures necessary to induce these phase transitions, in qualitative correspondence with experimental observations.[51]

In conclusion, the hybrid MC/MD protocol of Scheme 2 yields direct access to the adsorption isotherms and (meta)stable MIL-53(Al) states at a given gas pressure, in qualitative agreement with experiment.[51] However, a quantitative interpretation of these results is impeded due to two phenomena. First, premature phase transitions may occur if both phases are separated by very low barriers, such as for the lp-to-np transition under methane adsorption, making it impossible to exactly pinpoint the critical gas pressure necessary to induce this transition. Second, the phase space associated with the unit cell volume may not be sampled adequately if the (meta)stable states are separated by large free energy barriers, such as for the cp-to-np transition under methane adsorption. In that case, the initial state of the simulation affects the simulation results, such that multiple independent simulations need to be performed, reducing the efficiency of the procedure.

3.3. Scheme 3: Advanced MD Simulations of the Guest-Loaded Framework at Constant Volume

Figure 8a,b displays the pressure-versus-volume equations of state and corresponding free energy profiles for various xenon

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

ADVANCED THEORY AND SIMULATIONS
www.advtheorysimul.com

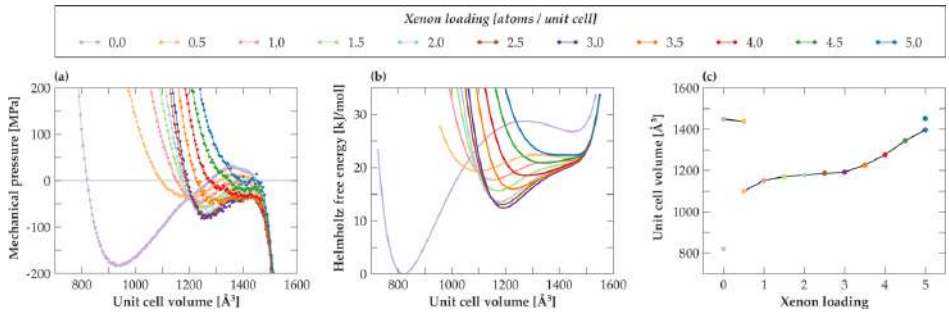


Figure 8. a) Predicted pressure-versus-volume equation of state of MIL–53(Al) with various fixed xenon loadings at 300 K. b) Corresponding Helmholtz free energy profiles at 300 K, and c) unit cell volumes of the (meta)stable state(s) with increasing xenon loading. Reproduced from ref. [19] with permission from Springer Nature.

loadings in MIL-53(Al) at 300 K, obtained by application of Scheme 3. These profiles reveal the unit cell volumes of the (meta)stable states, which are plotted as a function of the xenon loading in Figure 8c. The empty framework exhibits the well-known free energy profile with a stable cp phase at ≈820 Å^3 and a metastable lp phase at ≈1450 Å^3.[37,38] Upon forcing xenon into the framework, the global cp minimum vanishes immediately, while an intermediate np phase at a volume of about 1100 Å^3 is formed next to a very shallow lp phase. This lp phase disappears completely upon further increasing the xenon loading, while the stable np phase shifts to larger unit cell volumes reaching up to 1400 Å^3. Finally, when a loading of five xenon atoms per unit cell is reached, the lp state re-emerges in co-existence with the np state.

At first instance, these observations may seem to contradict the results of Schemes 1 and 2. First, both previous schemes predicted a stable cp phase also for low but nonzero xenon gas pressures, while here the cp minimum is only encountered for a zero xenon loading. However, this is a direct consequence of the guest loading being the control variable in this scheme rather than the gas pressure in Schemes 1 and 2, with a nontrivial relationship between both variables. Indeed, the xenon isotherms of Scheme 1 (Figure 4b,e) and Scheme 2 (Figure 6) predict that no xenon is adsorbed in the stable cp phase at low xenon gas pressures. Moreover, Figure 6 clearly reveals that a cp-to-np phase transition is triggered as soon as the cp initialized simulations start to adsorb xenon, confirming the observation of Scheme 3 that a stable cp phase cannot exist for a nonzero xenon loading. Second, the procedure outlined here predicts an np volume which starts to increase appreciably for xenon loadings between 3.5 and 5 atoms per unit cell, whereas the xenon-loaded np volume was found to be largely independent of the gas pressure in Figures 4 and 6. Once again, this results from fixing the xenon loading rather than the gas pressure or chemical potential. Figure 6a reveals that a xenon loading between 3.5 and 5 atoms per unit cell near 300 K corresponds to a very narrow range of gas pressures, between 1.0 and 1.5 bar, coinciding with the pressure necessary to induce an np-to-lp phase transition at this temperature. As a result, Figure 8c disproportionally highlights the volume dependence of the np phase.

In conclusion, while Scheme 3 succeeds in identifying all (meta)stable states independent of the free energy barriers between them, it only provides direct insight in the evolution of these (meta)stable states as a function of the guest loading rather than the experimentally controlled gas pressure or chemical potential. To access the osmotic potential, one would need to further correlate the chemical potential with the number of adsorbed species at each unit cell volume, as outlined in ref. [55]. Furthermore, given the discrete and constant number of guest species present in the unit cell, one has to rely on more expensive supercell simulations to predict the different (meta)stable states at fractional loadings, such as the $1 \times 2 \times 1$ supercell employed here to access half-integer loadings per unit cell.

3.4. Scheme 4: Hybrid MC/MD Simulations with a Flexible Unit Cell in the Restricted Osmotic Ensemble

Based on the results obtained so far, it is clear that none of the previous three schemes provides an ideal procedure to study gas adsorption in flexible MOFs. The rigid-host GCMC simulations of Scheme 1 are too crude to accurately predict adsorption isotherms and osmotic potentials, while the occurrence of premature phase transitions and high free energy barriers, which may not be surmounted in regular MD simulations, hinder the efficient application of Scheme 2. Finally, in Scheme 3, only the gas loading instead of the experimentally accessible chemical potential or gas pressure could be controlled, preventing a direct comparison with experiment. For this reason, we developed and validated a fourth scheme that alleviates these issues by explicitly accounting for the shape flexibility of the material. Compared to Scheme 1, the procedure of Scheme 4 can be expected to be more accurate and efficient. First off, by allowing fluctuations in the framework atoms in Scheme 4, the inaccuracy due to the artificial dependence on the initial atomic positions and cell shape that was present in Scheme 1 should be eliminated. Second, given that Scheme 4 is a completely hybrid MC/MD procedure, the MIL-53(Al) framework can adapt to accommodate for the different guest species during the simulation. In Scheme

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

ADVANCED THEORY AND SIMULATIONS
www.advtheorysimul.com

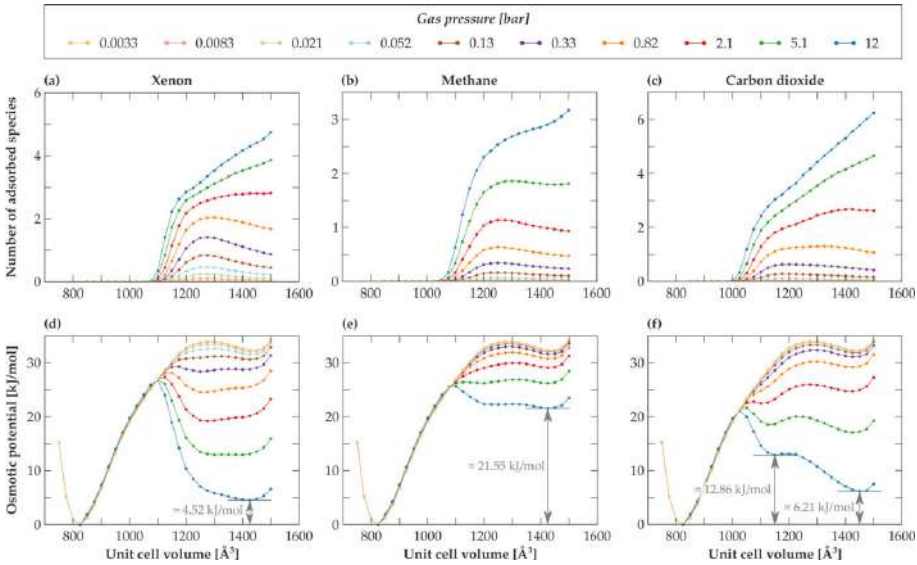


Figure 9. Flexible-host adsorption isotherms for a) xenon, b) methane, and c) carbon dioxide as a function of the MIL-53(Al) unit cell volume V at 300 K for a range of pressures between 3.3×10^{-3} and 12 bar as predicted from the newly introduced hybrid MC/MD scheme. d–f) Corresponding osmotic potentials Ω as a function of the unit cell volume V for these three guest molecules. Shaded areas denote the 1σ uncertainty interval as determined over five independent simulations, which are negligible here.

1, this is inherently impossible, as one has to rely on one fixed framework per volume.

3.4.1. Noble Gas Adsorption

To investigate the performance of the newly proposed scheme, Scheme 4 is adopted to predict the adsorption isotherms and corresponding osmotic potentials for each of the five guest species in MIL-53(Al) at 300 K. In **Figure 9**, only the gases with the potential to induce phase transitions in MIL-53(Al)—xenon, methane, and carbon dioxide—are considered, whereas the results for neon and argon are displayed in Figure S14, Supporting Information.

Comparing the xenon flexible-host isotherm of Figure 9a with the corresponding rigid-host isotherm of Figure 4b immediately reveals that, while both types of isotherms share the same shape, an appreciably lower uncertainty is associated with the flexible-host isotherms of Scheme 4. Similar to the rigid-host isotherms, five independent flexible-host isotherms were constructed starting from five different cell shapes for each unit cell volume. In contrast to Scheme 1, however, our newly introduced scheme yields five sets of results which are virtually indistinguishable, as indicated by the very small 1σ uncertainty interval in Figure 9d. For the metastable lp phase, this uncertainty is reduced almost 100-fold. This confirms that the new procedure outlined here succeeds in fully accounting for the flexibility in the atomic positions and cell shape, so that the choice of initial snapshot does no longer artificially influence the obtained results. As a result, Scheme 4 requires only one hybrid MC/MD simulation per unit cell volume to obtain accurate results, improving the efficiency of the method.

This substantially improved accuracy does not only reduce the number of independent simulations that need to be carried out to get significant results, but also allows to quantitatively predict the evolution of the adsorption isotherm and (meta)stable states with increasing gas pressure. As an example, while it was impossible to extract from Figure 4e the exact metastable lp volume at a xenon gas pressure of 12 bar—only a broad volume range between 1300 and 1475 Å^3 could be obtained—the osmotic potential of Figure 9d clearly indicates that the lp volume is found around 1425 Å^3. Moreover, Figure 4e hinted toward a possible metastable np phase at a pressure of 12 bar, with a volume around 1300 Å^3. However, when properly accounting for the flexibility in the host material, it becomes clear from Figure 9d that this np minimum only results from the large inaccuracy of Scheme 1.

3.4.2. Methane and Carbon Dioxide Adsorption

For methane adsorption at 300 K, the staggered osmotic potential of Figure 4f revealed multiple minima between 1100 and 1400 Å^3 at the highest gas pressure, with an energy that fell within the 1σ uncertainty interval of the metastable lp minimum. In contrast, thanks to the increased accuracy of Scheme 4, Figure 9e demonstrates that only the lp state at a volume of 1425 Å^3 is a metastable state at 12 bar.

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

ADVANCED THEORY AND SIMULATIONS
www.advtheorysimul.com

Undoubtedly, the increased accuracy of Scheme 4 is most pronounced for carbon dioxide. The rigid-host isotherms and osmotic potentials displayed in Figure 5 revealed a large dependence on the choice of the fixed framework during the GCMC procedure. As a result, it is difficult to uniquely determine from Figure 5d for which range of pressures an np phase is induced, as well as to determine the relative stability of the various metastable phases. As a substantially lower uncertainty is associated with the flexible-host isotherms and osmotic potentials obtained in Scheme 4, which are displayed in Figure 9c,f, these questions can now be fully answered. Figure 9f reveals that an np phase starts to appear at a carbon dioxide pressure of ≈2.1 bar, while it almost disappears when increasing the pressure to 12 bar. With increasing pressure, the relative stability of the np and lp phases with respect to the cp phase increases, yielding metastable lp and np states at 12 bar which are about 6.21 and 12.86 kJ mol^{-1} less stable than the stable cp state at 12 bar.

In conclusion, these observations confirm the improved accuracy of the flexible-host isotherms, obtained by explicitly accounting for the flexibility of the host material in Scheme 4. This allows to quantitatively predict the number of adsorbed guest species as well as the different metastable states of the adsorbent at a given gas pressure. Furthermore, Scheme 4 prevents the need for multiple independent runs to improve the accuracy of the prediction, as one hybrid MC/MD simulation per unit cell volume is sufficient to obtain accurate results in good agreement with experiment. As a result, adoption of this scheme, which is freely available in the in-house `Yaff` software code,[56] to model gas adsorption in flexible adsorbents has the potential to further speed up the computer-aided discovery of interesting nanoporous materials for adsorption applications.

4. Conclusions

To adopt flexible MOFs for applications in gas storage, it is crucial to understand at a microscopic level how the interactions between the adsorbate and the adsorbent can be tuned in order to design the optimal MOF for a given application. In this respect, the insights provided by computational research could be unparalleled, given the existence of a computational procedure that can efficiently and accurately predict the adsorption isotherms and adsorption-induced flexibility in MOFs. While multiple schemes have been proposed for this purpose, the extreme flexibility present in some MOFs may limit the applicability and accuracy of the results obtained with these schemes.

Therefore, the performances of four hybrid MC/MD simulation protocols to predict the adsorption of neon, argon, xenon, methane, and carbon dioxide in MIL-53(Al) were critically compared, focussing on their accuracy and efficiency. In a first scheme, rigid-host isotherms and associated osmotic potentials as a function of the unit cell volume were constructed. While this procedure is fast and solely relies on a set of GCMC simulations, it does not properly account for the flexibility in the atomic positions and cell shape of the host structure. As a consequence, the results obtained with this scheme were found to depend on the choice of the fixed framework used as input for the GCMC procedure, yielding differences of up to 25% in predicted carbon dioxide uptake caused by a small rotation in the organic linkers of MIL-53(Al). In a second scheme, hybrid MC/MD simulations were performed to obtain the equilibrium loadings and unit cell volume at a given gas pressure. This scheme, which directly samples the osmotic ensemble, was hindered by both premature phase transitions—a consequence of very small free energy barriers and internal pressure fluctuations during the simulations—and suppressed phase transitions—a consequence of the metastable states being separated by too high an energy barrier. Finally, in the third scheme, the Helmholtz free energy and equilibrium volumes as a function of the guest loading were determined. The main disadvantage of this last scheme was the lack of direct information on the osmotic potential, as the procedure samples at constant guest loading rather than at the experimentally controlled constant chemical potential or gas pressure.

Given the absence of a satisfactory procedure to study gas adsorption in flexible MOFs, a new hybrid procedure that is specifically designed to account for the flexibility of the host material was derived. In this novel scheme, flexible-host isotherms were constructed via hybrid MC/MD simulations during which the atomic positions and cell shape were allowed to fluctuate, while still controlling the cell volume. Application of this procedure to each of the five guest species indicated that the accuracy obtained with this scheme substantially outperforms that of the other three schemes, as the artificial dependence on the fixed framework of Scheme 1 is no longer present. As a result, only one set of hybrid MC/MD simulations was sufficient to accurately predict the potential of these guest molecules to induce phase transitions in MIL-53(Al), furthermore obtaining a strong agreement with experiment. Moreover, this increased accuracy also results in a quantitative prediction of the different adsorption-induced metastable states at a given gas pressure. Therefore, this new and efficient scheme has the potential to further accelerate the computer-aided discovery of flexible MOFs for adsorption applications.

5. Computational Details

The four simulation schemes discussed here were implemented and tested in Yaff, our freely available in-house developed software package,[56] and validated extensively using RASPA[57] (see Section S4, Supporting Information). The temperature was controlled using a single Nosé–Hoover chain consisting of three beads and with a relaxation time $\tau_T = 0.1$ ps, which was coupled to both the particles and the barostat if applicable.[58–61] If necessary, the pressure P and/or deviatoric stress $\boldsymbol{\sigma}_a$ were controlled using a Martyna–Tobias–Tuckerman–Klein (MTTK) barostat with a relaxation time of 1 ps.[62,63] This combination of relaxation times promotes a complete yet efficient sampling of the accessible phase space.[37] To efficiently evaluate the long-range van der Waals and electrostatic interactions, a smooth cutoff at 15 Å was introduced and supplemented with analytical tail corrections. Specifically, the electrostatic interactions were calculated using an Ewald summation with a splitting parameter α of 0.213 Å^{-1} and a reciprocal space cutoff of 0.32 Å^{-1}.[64] For the MD simulations, a time step of 0.5 fs was employed to ensure energy conservation.

MIL-53(Al) was modeled using a $1 \times 2 \times 1$ cell containing 152 atoms, obtained by doubling the unit cell along the inorganic

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

ADVANCED THEORY AND SIMULATIONS
www.advtheorysimul.com

chain, and employing a force field generated from periodic ab initio data following the QuickFF procedure.[65] For the non-covalent MIL-53(Al) force field parameters, atomic charges were obtained according to the Minimal Basis Iterative Stockholder (MBIS) partitioning scheme[66] as implemented in HORTON.[67] The all-electron density necessary for the MBIS scheme was obtained via a GPAW calculation.[68–70] For the GCMC simulations in Schemes 1 and 4, the van der Waals interactions were described using the Lennard-Jones (LJ) 12–6 interatomic potentials that are strongly positive for small internuclear distances, which is important for the insertion trial moves.[71] These LJ parameters were obtained from the MM3 parameters found in the literature by requiring that the location and depth of the potential well are the same in both descriptions. In accordance with the earlier work on Scheme 3,[19,37] the MM3 Buckingham potentials were adopted to describe the van der Waals interactions in Schemes 2 and 3.[72] To prevent the insertion of particles in the unphysical regions at very small internuclear distances in Scheme 2, the MM3 potential was modified to reach a plateau equal to the maximum in the non-modified MM3 potential. The MM3 parameters, σ and ε, for both the MIL-53(Al) framework atoms and the guest species were obtained from ref. [72] and were modified to the LJ form if necessary. For methane, a neutral united-atom model was employed, following the description of ref. [73]. Furthermore, an atomic point charge model was adopted for carbon dioxide, with partial charges of $0.6512|e|$ and $-0.3256|e|$ for the carbon and oxygen atoms, respectively, which complement the atomic MM3 parameters.[74] In the MC simulations, the carbon dioxide molecules were assumed to be rigid, whereas a flexible model was employed in the MD simulations. For the latter, the covalent force field terms were obtained using the QuickFF procedure.[65] As shown in Section S5, Supporting Information, the predicted gas adsorption isotherms are robust with respect to small changes in the host–guest interactions.

In Scheme 1, the GCMC simulations in the (N_{host}, μ, V, $\mathbf{h}_0$, T) ensemble were performed for a range of framework volumes V between 850 and 1500 Å^3 with a step size of 25 Å^3, and for a series of chemical potentials μ. The snapshots of the empty framework structures were extracted from prior MD simulations in the (N_{host}, $N_{\text{guest}} = 0$, P, $\boldsymbol{\sigma}_a = \mathbf{0}$, T) ensemble, starting in the lp phase and performed at a sufficiently high pressure in order to initiate an lp-to-cp transition. During these MC simulations, four types of MC moves were tried: insertion of the guest molecule in the framework, deletion of the guest molecule from the framework, translation of the guest molecule, and rotation of the guest molecule in the case of CO_2. As indicated in Section S1.3, Supporting Information, an equal probability of proposing these moves leads to an optimal convergence. A total number of 750 000 MC steps were performed for xenon, methane, and carbon dioxide, whereas this was decreased to 500 000 and 250 000 MC steps for argon and neon, respectively. As discussed in Section S1.4, Supporting Information, this simulation length ensures a full convergence of the obtained results. To shed light on the possible influence of the flexibility of the framework on the obtained results, five different structures—with differing cell shapes $\mathbf{h}_0$ and atomic positions but with the same volume—were considered at each volume point between 850 and 1500 Å^3, and the results reported here were averaged over these five simulations with associated 1σ uncertainty intervals unless indicated otherwise.

In Scheme 2, the rate of convergence of the hybrid procedure was maximized by considering MD runs of 400 steps (0.2 ps) each and proposing this move with a probability of 1:400 compared to each of the other MC moves (insertion, deletion, translation, and, if necessary, rotation), as outlined in more detail in Sections S2.2 and S2.3, Supporting Information. To improve the probability of acceptance and hence accelerate the convergence of the method, acceptance and rejection windows rather than single acceptance and rejection snapshots are considered for the MD trajectories, as introduced in ref. [75] and outlined in more detail in Section S2.4, Supporting Information. Here, the size of these windows was chosen to equal halve the total simulation length, so that each step in the MD simulation can be selected as the new state in the MC procedure. To determine the variability in the obtained results, ten independent simulations were performed for each chemical potential μ, or equivalently, gas pressure p. Five of these simulations started from the empty lp phase, while the other five started from the empty cp phase. A total number of 3 000 000 MC steps were performed for argon, xenon, and methane, while 2 000 000 and 5 000 000 MC steps were performed for neon and carbon dioxide, respectively.

In Scheme 3, the (N_{host}, N_{guest}, V, $\boldsymbol{\sigma}_a = \mathbf{0}$, T) simulations were performed for various guest loadings N_{guests} varying between 0 and 5 xenon atoms. As a $1 \times 2 \times 1$ supercell is used, it is also possible to probe at half-integer guest loadings. For all simulations of Scheme 3, a production run of 700 ps was considered, which was preceded by a 100 ps equilibration run.

In Scheme 4, five structures with differing initial cell shapes $\mathbf{h}_0$ and atomic positions but with the same volume were considered at each volume point between 850 and 1500 Å^3, and the results reported here were averaged over these five simulations, similar to Scheme 1. As in Scheme 2, a total number of 3 000 000 MC steps were performed for argon, xenon, and methane, while 2 000 000 and 5 000 000 MC steps were performed for neon and carbon dioxide, respectively.

Finally, to relate the chemical potential μ and the gas pressure p at a given temperature T, the free energy of a van der Waals gas is adopted:

$$\mu(p, T) = k_B T \ln \frac{b\rho}{1 - b\rho} + k_B T \frac{b\rho}{1 - b\rho} - 2a\rho + \mu_0 - k_B T \ln \frac{p_0 b}{k_B T} \tag{4a}$$

$$p(\rho, T) = k_B T \frac{\rho}{1 - b\rho} - a\rho^2 \tag{4b}$$

In this formula, k_B is the Boltzmann's constant and $\rho(p, T)$ is the number density of the gas, obtained from the van der Waals equation of state at the given pressure and temperature. The parameters μ_0 and p_0 are the reference chemical potential and pressure, and a and b describe the interactions between the guest molecules and the volume taken up by these guest molecules, respectively. The a and b parameters for the adsorbates discussed in this work are summarized in **Table 1**.

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

ADVANCED THEORY AND SIMULATIONS
www.advtheorysimul.com

Table 1. The van der Waals parameters a and b entering Equation (4) for the various adsorbates discussed in this work.[76]

Guest species	$a[\ell^2 \cdot \text{bar mol}^{-1}]$	$b[\ell \text{ mol}^{-1}]$
Neon	0.214	0.01710
Argon	1.355	0.03200
Xenon	4.250	0.05105
Methane	2.283	0.04280
Carbon dioxide	3.640	0.04270

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

S.M.J.R. and R.G. contributed equally to this work. This research was funded by the Fund for Scientific Research—Flanders (FWO), the Research Board of Ghent University (BOF), and the European Union's Horizon 2020 research and innovation programme (consolidator ERC grant agreement number 647755–DYNPOR [2015–2020]). The computational resources and services used in this work were provided by the Flemish Supercomputer Center (VSC), funded by the Hercules foundation and the Flemish Government—department EWI. J.J.G.-S. acknowledges the Research Council of Ghent University for his BOF-postdoctoral fellowship.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

breathing, flexibility, gas adsorption, hybrid Monte Carlo/molecular dynamics models, metal–organic frameworks, osmotic ensemble, phase transitions

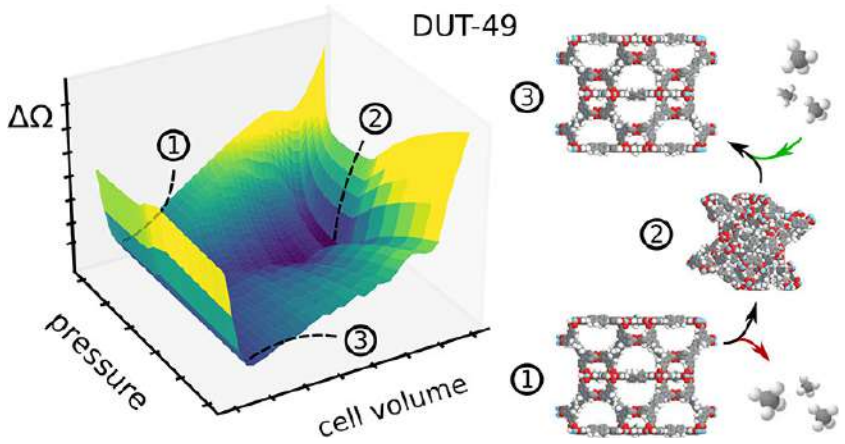
Received: November 12, 2018
Revised: December 20, 2018
Published online:

[1] J. A. Mason, M. Veenstra, J. R. Long, *Chem. Sci.* **2014**, *5*, 32.
[2] Y. He, W. Zhou, G. Qian, B. Chen, *Chem. Soc. Rev.* **2014**, *43*, 5657.
[3] H. Li, K. Wang, Y. Sun, C. T. Lollar, J. Li, H.-C. Zhou, *Mater. Today* **2018**, *21*, 108.
[4] L. J. Murray, M. Dincă, J. R. Long, *Chem. Soc. Rev.* **2009**, *38*, 1294.
[5] M. P. Suh, H. J. Park, T. K. Prasad, D.-W. Lim, *Chem. Rev.* **2012**, *112*, 782.
[6] J. Liu, P. K. Thallapally, B. P. McGrail, D. R. Brown, J. Liu, *Chem. Soc. Rev.* **2012**, *5*, 2308.
[7] J. E. Mondloch, M. J. Katz, W. C. Isley, III, P. Ghosh, P. Liao, W. Bury, G. W. Wagner, M. G. Hall, J. B. DeCoste, G. W. Peterson, R. Q. Snurr, C. J. Cramer, J. T. Hupp, O. K. Farha, *Nat. Mater.* **2015**, *14*, 512.
[8] H. Li, M. Eddaoudi, M. O'Keeffe, O. M. Yaghi, *Nature* **1999**, *402*, 276.
[9] G. Férey, *Chem. Mater.* **2001**, *13*, 3084.
[10] S. Kitagawa, R. Kitaura, S.-I. Noro, *Angew. Chem., Int. Ed.* **2004**, *43*, 2334.
[11] H.-C. Zhou, J. R. Long, O. M. Yaghi, *Chem. Rev.* **2012**, *112*, 673.
[12] H. Furukawa, K. E. Cordova, M. O'Keeffe, O. M. Yaghi, *Science* **2013**, *341*, 1230444.
[13] G. Maurin, C. Serre, A. Cooper, G. Férey, *Chem. Soc. Rev.* **2017**, *46*, 3104.
[14] M. Eddaoudi, J. Kim, N. Rosi, D. Vodak, J. Wachter, M. O'Keeffe, O. M. Yaghi, *Science* **2002**, *295*, 469.
[15] O. M. Yaghi, M. O'Keeffe, N. W. Ockwig, H. K. Chae, M. Eddaoudi, J. Kim, *Nature* **2003**, *423*, 705.
[16] J.-R. Li, J. Sculley, H.-C. Zhou, *Chem. Rev.* **2012**, *112*, 869.
[17] X. Li, Y. Liu, J. Wang, J. Gascon, J. Li, B. Van der Bruggen, *Chem. Soc. Rev.* **2017**, *46*, 7124.
[18] Z. Kang, L. Fan, D. Sun, *J. Mater. Chem. A* **2017**, *5*, 10073.
[19] L. Vanduyfhuys, S. M. J. Rogge, J. Wieme, S. Vandenbrande, G. Maurin, M. Waroquier, V. Van Speybroeck, *Nat. Commun.* **2018**, *9*, 204.
[20] P. Chowdhury, C. Bikkina, D. Meister, F. Dreisbach, S. Gumma, *Microporous Mesoporous Mater.* **2009**, *117*, 406.
[21] F.-X. Coudert, A. Boutin, M. Jeffroy, C. Mellot-Draznieks, A. H. Fuchs, *ChemPhysChem* **2011**, *12*, 247.
[22] A. Schneemann, V. Bon, I. Schwedler, I. Senkovska, S. Kaskel, R. A. Fischer, *Chem. Soc. Rev.* **2014**, *43*, 6062.
[23] F.-X. Coudert, A. H. Fuchs, *Coord. Chem. Rev.* **2016**, *307*, 211.
[24] S. Vandenbrande, T. Verstraelen, J. J. Gutiérrez-Sevillano, M. Waroquier, V. Van Speybroeck, *J. Phys. Chem. C* **2017**, *121*, 25309.
[25] L.-C. Lin, A. H. Berger, R. L. Martin, J. Kim, J. A. Swisher, K. Jariwala, C. H. Rycroft, A. S. Bhown, M. W. Deem, M. Haranczyk, B. Smit, *Nat. Mater.* **2012**, *11*, 633.
[26] M. Fischer, J. R. B. Gomes, M. Fröba, M. Jorge, *Langmuir* **2012**, *28*, 8537.
[27] H. J. Fang, H. Demir, P. Kamakoti, D. S. Sholl, *J. Mater. Chem. A* **2014**, *2*, 274.
[28] S. Vandenbrande, M. Waroquier, V. Van Speybroeck, T. Verstraelen, *J. Chem. Theory Comput.* **2018**, *14*, 6359
[29] G. Férey, C. Serre, *Chem. Soc. Rev.* **2009**, *38*, 1380.
[30] F.-X. Coudert, *Chem. Mater.* **2015**, *27*, 1905.
[31] D. Fairen-Jimenez, S. A. Moggach, M. T. Wharmby, P. A. Wright, S. Parsons, T. Düren, *J. Am. Chem. Soc.* **2011**, *133*, 8900.
[32] C. Mellot-Draznieks, C. Serre, S. Surblé, N. Audebrand, G. Férey, *J. Am. Chem. Soc.* **2005**, *127*, 16273.
[33] C. Serre, C. Mellot-Draznieks, S. Surblé, N. Audebrand, Y. Filinchuk, G. Férey, *Science* **2007**, *315*, 1828.
[34] T. Loiseau, C. Serre, C. Huguenard, G. Fink, F. Taulelle, M. Henry, T. Bataille, G. Férey, *Chem. Eur. J.* **2004**, *10*, 1373.
[35] Y. Liu, J.-H. Her, A. Dailly, A. J. Ramirez-Cuesta, D. A. Neumann, C. M. Brown, *J. Am. Chem. Soc.* **2008**, *130*, 11813.
[36] I. Beurroies, M. Boulhout, P. L. Llewellyn, B. Kuchta, G. Férey, C. Serre, R. Denoyel, *Angew. Chem., Int. Ed.* **2010**, *49*, 7526.
[37] S. M. J. Rogge, L. Vanduyfhuys, A. Ghysels, M. Waroquier, T. Verstraelen, G. Maurin, V. Van Speybroeck, *J. Chem. Theory Comput.* **2015**, *11*, 5583.
[38] R. Demuynck, S. M. J. Rogge, L. Vanduyfhuys, J. Wieme, M. Waroquier, V. Van Speybroeck, *J. Chem. Theory Comput.* **2017**, *13*, 5861.
[39] S. M. J. Rogge, M. Waroquier, V. Van Speybroeck, *Acc. Chem. Res.* **2018**, *51*, 138.
[40] S. M. J. Rogge, S. Caroes, R. Demuynck, M. Waroquier, V. Van Speybroeck, A. Ghysels, *J. Chem. Theory Comput.* **2018**, *14*, 1186.
[41] J. D. Evans, F.-X. Coudert, *J. Phys. Chem. Lett.* **2017**, *8*, 1578.
[42] S. Duane, A. D. Kennedy, B. J. Pendleton, D. Roweth, *Phys. Lett. B* **1987**, *195*, 216.
[43] B. Mehlig, D. W. Heermann, B. M. Forrest, *Phys. Rev. B* **1992**, *45*, 679.
[44] R. Faller, J. J. d. Pablo, *J. Chem. Phys.* **2002**, *116*, 55.
[45] M. Jeffroy, A. H. Fuchs, A. Boutin, *Chem. Commun.* **2008**, 3275.
[46] F.-X. Coudert, M. Jeffroy, A. H. Fuchs, A. Boutin, C. Mellot-Draznieks, *J. Am. Chem. Soc.* **2008**, *130*, 14294.

[47] J. A. Greathouse, T. L. Kiniburgh, M. D. Allendorf, *Ind. Eng. Chem. Res.* **2009**, *48*, 3425.
[48] M. Witman, S. Ling, S. Jawahery, P. G. Boyd, M. Haranzcyk, B. Slater, B. Smit, *J. Am. Chem. Soc.* **2017**, *139*, 5547.
[49] A. Ghoufi, G. Maurin, *J. Phys. Chem. C* **2010**, *114*, 6496.
[50] A. Boutin, M.-A. Springuel-Huet, A. Nossov, A. Gédéon, T. Loiseau, C. Volkringer, G. Férey, F.-X. Coudert, A. H. Fuchs, *Angew. Chem., Int. Ed.* **2009**, *48*, 8314.
[51] A. Boutin, F.-X. Coudert, M.-A. Springuel-Huet, A. V. Neimark, G. Férey, A. H. Fuchs, *J. Phys. Chem. C* **2010**, *114*, 22237.
[52] S. Bourrelly, P. L. Llewellyn, C. Serre, F. Millange, T. Loiseau, G. Férey, *J. Am. Chem. Soc.* **2005**, *127*, 13519.
[53] M. Mehta, D. A. Kofke, *Chem. Eng. Sci.* **1994**, *49*, 2633.
[54] J. G. Kirkwood, *J. Chem. Phys.* **1935**, *3*, 300.
[55] L. Vanduyfhuys, A. Ghysels, S. M. J. Rogge, R. Demuynck, V. Van Speybroeck, *Mol. Simulat.* **2015**, *41*, 1311.
[56] T. Verstraelen, L. Vanduyfhuys, S. Vandenbrande, S. M. J. Rogge, *Yaff, Yet Another Force Field*, http://molmod.ugent.be/software/
[57] D. Dubbeldam, S. Calero, D. E. Ellis, R. Q. Snurr, *Mol. Simul.* **2016**, *42*, 81.
[58] S. Nosé, *Mol. Phys.* **1984**, *52*, 255.
[59] S. Nosé, *J. Chem. Phys.* **1984**, *81*, 511.
[60] W. G. Hoover, *Phys. Rev. A* **1985**, *31*, 1695.
[61] G. J. Martyna, M. L. Klein, M. E. Tuckerman, *J. Chem. Phys.* **1992**, *97*, 2635.
[62] G. J. Martyna, D. J. Tobias, M. L. Klein, *J. Chem. Phys.* **1994**, *101*, 4177.
[63] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, M. L. Klein, *Mol. Phys.* **1996**, *87*, 1117.
[64] P. P. Ewald, *Ann. Phys.* **1921**, *369*, 253.
[65] L. Vanduyfhuys, S. Vandenbrande, J. Wieme, M. Waroquier, T. Verstraelen, V. Van Speybroeck, *J. Comput. Chem.* **2018**, *39*, 999.
[66] T. Verstraelen, S. Vandenbrande, F. Heidar-Zadeh, L. Vanduyfhuys, V. Van Speybroeck, M. Waroquier, P. W. Ayers, *J. Chem. Theory Comput.* **2016**, *12*, 3894.
[67] T. Verstraelen, K. Boguslawski, P. Tecmer, F. Heidar-Zadeh, M. Chan, T. D. Kim, T. Zhao, S. Vandenbrande, D. Yang, C. E. González-Espinoza, P. A. Limacher, D. Berrocal, A. Malek, P. W. Ayers, HORTON 2.0.0, **2015**, http://theochem.github.com/horton/
[68] J. J. Mortensen, L. B. Hansen, K. W. Jacobsen, *Phys. Rev. B* **2005**, *71*, 035109.
[69] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, L.-O. Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Häkkinen, G. K. H. Madsen, R. M. Nieminen, et al., *J. Phys. Condens. Matter* **2010**, *22*, 253202.
[70] S. R. Bahn, K. W. Jacobsen, *Comput. Sci. Eng.* **2002**, *4*, 56.
[71] J. E. Jones, *Proc. R. Soc. A* **1924**, *106*, 463.
[72] J. H. Lii, N. L. Allinger, *J. Am. Chem. Soc.* **1989**, *111*, 8576.
[73] M. G. Martin, J. I. Siepmann, *J. Phys. Chem. B* **1998**, *102*, 2569.
[74] F. Salles, A. Ghoufi, G. Maurin, R. G. Bell, C. Mellot-Draznieks, G. Férey, *Angew. Chem., Int. Ed.* **2008**, *47*, 8487.
[75] R. M. Neal, *J. Comput. Phys.* **1994**, *111*, 194.
[76] J. R. Rumble, *CRC Handbook of Chemistry and Physics*, 98th ed., CRC Press, Boca Raton, Florida, USA **2017**.

Paper II

Charting the Complete Thermodynamic Landscape of Gas Adsorption for a Responsive Metal–Organic Framework



R. Goeminne, S. Krause, S. Kaskel, T. Verstraelen, J.D. Evans

J. Am. Chem. Soc, **2021**, *143* (11), 4143–4147

R. Goeminne performed the simulations, interpreted the resulting data and contributed to the manuscript.

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Communication

Charting the Complete Thermodynamic Landscape of Gas Adsorption for a Responsive Metal–Organic Framework

Ruben Goeminne, Simon Krause, Stefan Kaskel, Toon Verstraelen,* and Jack D. Evans*

Cite This: *J. Am. Chem. Soc.* 2021, 143, 4143–4147 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: New nanoporous materials have the ability to revolutionize adsorption and separation processes. In particular, materials with adaptive cavities have high selectivity and may display previously undiscovered phenomena, such as negative gas adsorption (NGA), in which gas is released from the framework upon an increase in pressure. Although the thermodynamic driving force behind this and many other counterintuitive adsorption phenomena have been thoroughly investigated in recent years, several experimental observations remain difficult to explain. This necessitates a comprehensive analysis of gas adsorption akin to the conformational free energy landscapes used to understand the function of proteins. We have constructed the complete thermodynamic landscape of methane adsorption on DUT-49. Traversing this complex landscape reproduces the experimentally observed structural transitions, temperature dependence, and the hysteresis between adsorption and desorption. The complete thermodynamic description presented here provides unparalleled insight into adsorption and provides a framework to understand other adsorbents that challenge our preconceptions.

Advanced porous materials can deliver exciting developments to industrial processes[1] and the flexibility of porous adsorbents could address complex separations similar to biological systems.[2] For example, extreme changes in adsorption behavior produce improved selectivity, storage capacity, and thermal management.[3−5] Responsive adsorbents, such as metal–organic frameworks (MOFs), can significantly change pore structure and volume in response to external stimuli such as pressure, temperature, gas adsorption or light absorption.[6]

One particularly exciting MOF that exhibits negative gas adsorption (NGA) is DUT-49. This MOF undergoes a structural transition associated with a 50% contraction of the unit cell upon increasing methane gas pressure, resulting in spontaneous desorption of methane.[7] This leads to pressure amplification in a closed system, opening the door to new applications.[8] The lack of thermodynamic understanding of the NGA phenomenon has motivated a number of computational studies.[9,10] However, efforts have mostly relied on two-state models that are unable to capture the complete phase-space during adsorption and desorption.[11]

The most appropriate description of the thermodynamics of responsive frameworks under gas adsorption is provided by the osmotic ensemble. Simulations in this ensemble impose a chemical potential of the adsorbed fluid (μ), a mechanical pressure (p), and a temperature (T). Under these conditions, the equilibrium framework minimizes the osmotic potential (Ω) through structural relaxation, changes in cell shape and volume, and adsorption of guest molecules.[12] However, atomistic simulations within the osmotic ensemble remain challenging because dynamics must be combined with particle insertion/deletion, limiting its application to a handful of systems.[13,14] Descriptions of the osmotic potential are often achieved through analytical treatment of gas adsorption.[15] This can ascertain the relative stability of two distinct and porous phases.[16] For example, the process of "breathing" between an open (*op*) and contracted porous phase (*cp*), such as that observed for MIL-53,[17] can be understood by their difference in osmotic potential, $\Omega_{op} - \Omega_{cp}$ (Figure 1).[12] The difference in osmotic potential reveals which phase represents the global thermodynamic equilibrium as a function of gas pressure.[18] However, this methodology is not sufficient to describe NGA as it must involve metastable states.[7] Moreover, the

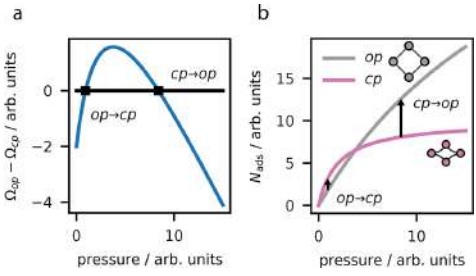


Figure 1. Osmotic potential (a) applied to transitions between two phases (*op* and *cp*) of an adsorbent. This can describe the "breathing" adsorption phenomenon (b).

Received: January 15, 2021
Published: March 15, 2021

https://doi.org/10.1021/jacs.1c00522
J. Am. Chem. Soc. 2021, 143, 4143–4147

Journal of the American Chemical Society pubs.acs.org/JACS Communication

experimentally observed hysteresis and temperature dependence of NGA cannot be reproduced by simply considering the difference in osmotic potential.[19] Experiments demonstrate that the *op* to *cp* transition at 90 K produces a much smaller NGA step than 120 K, whereas at 150 K, no transitions are observed.[19] To explain this temperature dependence and other complex adsorption processes, knowledge of not only the osmotic potential but also the osmotic barriers that exist between phases that act to prevent transformations, are required.

Free-energy landscapes of chemical reactions and conformational dynamics are expressed in terms of collective variables; similarly, barriers of "breathing" MOFs in the osmotic ensemble appear at specific cell volumes. We employed molecular simulations to treat the cell volume (V) as a "collective variable" in the osmotic ensemble. The osmotic potential, as a function of volume, reveals the (meta)stable states of interest and the barriers between them for any set of thermodynamic parameters μ, p, and T. With this method, detailed landscapes of the osmotic potential, as a function of cell volume of DUT-49 and methane gas pressure, have been computed at 90, 120, and 150 K. This complete thermodynamic analysis provides comprehensive insight into phase stability under adsorption.

We employed a force field derived from ab initio data of the inorganic and organic units of the DUT-49 framework[10] with the TraPPE[20] parametrization of methane. In our hybrid MC[14,21−23] simulations, molecular dynamics (MD) trajectories in the (N, V, $\boldsymbol{\sigma}_a = 0$, T) ensemble are new moves in a grand canonical Monte Carlo scheme. The MD simulations impose a fixed volume V and anisotropic contribution $\boldsymbol{\sigma}_a$ to the total stress tensor $\boldsymbol{\sigma} = P\mathbf{1} + \boldsymbol{\sigma}_a$,[24] allowing the cell shape to adapt to the fluctuating number of adsorbates resulting in flexible-host adsorption isotherms, $N_{guest}(N_{host}, \mu, \boldsymbol{\sigma}_a = \mathbf{0}, T; V)$. Isotherms were calculated for a series of unit cell volumes between 40 and 110 nm^3 with a spacing of 2 nm^3 (Figure 2). Subsequently, the flexible-host isotherms were used to construct the osmotic potential as eq 1.

$$\Omega(N_{host}, p, \boldsymbol{\sigma}_a = 0, T; V) = F_{host}(N_{host}, T; V) + pV - \int_{-\infty}^{\mu(p,T)} N_{guest}(N_{host}, \mu', \boldsymbol{\sigma}_a = 0, T; V)\, d\mu' \quad (1)$$

The empty-host free energy F_{host} was obtained by thermodynamic integration of the internal pressure, extracted from MD simulations of the empty framework in the (N, V, $\boldsymbol{\sigma}_a = 0$, T) ensemble at different unit cell volumes at 90, 120, and 150 K. The free energy was chosen to be 0 kJ mol^{-1} in the *op* phase at a unit cell volume of 104 nm^3, comparable to the experimentally observed unit cell volume of 100 nm^3.[7] This produces a meaningful comparison of the empty-host free energy (and thus osmotic potential). Simulations were performed between 46 and 75 values of gas pressure and the chemical potential at each temperature and gas pressure was calculated with the Peng−Robinson equation of state.[25] The above approach produces the osmotic potential at different methane gas pressures and cell volumes. The osmotic landscape is shown relative to the *op* phase at each gas pressure in Figure 2 to highlight differences between the phases. In contrast to previous approaches (Figure 1a), this surface offers a tremendous amount of additional information.

To clearly outline the effect of osmotic barriers on gas adsorption, we can chart paths[26] on the osmotic landscape that

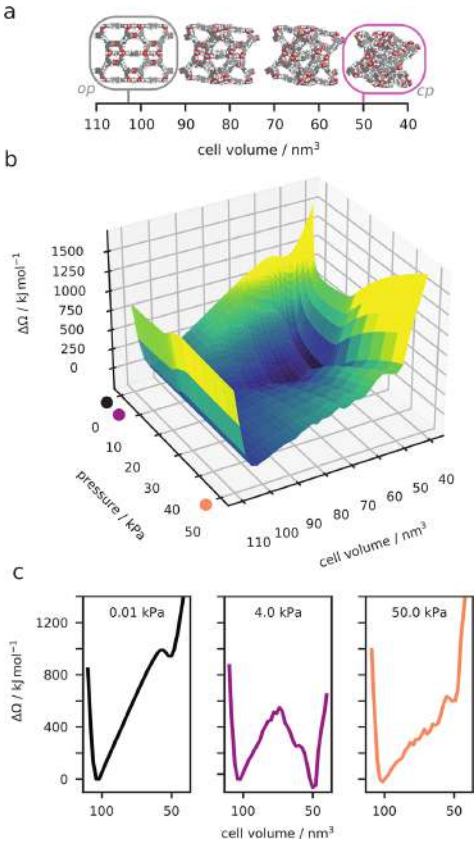


Figure 2. Computed osmotic surface of methane adsorption on DUT-49 at 120 K, as a function of unit cell volume (a) and methane gas pressure (b). Examples of the 1D osmotic surface at specific gas pressures (c).

are taken with adsorption and desorption along with the associated amount adsorbed. Figure 3a−c shows the path along which the system is in global thermodynamic equilibrium at each gas pressure, i.e., the path of lowest osmotic potential. Alternatively, Figure 3d−f shows the kinetic path along which transitions can occur only when the osmotic barrier between phases disappears. However, as thermal fluctuations can allow the crossing of barriers, a transition was noted when the barrier was reduced to 15 k_BT. The same transitions are obtained for the computed 120 K surface when varying this value of "crossable" barriers between 0 and 25 k_BT. The appropriate value of barrier heights that can be crossed in typical time spans of experiments at a given temperature is difficult to obtain. For example, a fluctuation energy of 6 k_BT per unit cell was proposed for a gate-opening MOF.[27]

At zero gas pressure, the system begins at equilibrium in the *op* phase, where there is also a metastable *cp* phase (at 50 nm^3). By increasing gas pressure, the *cp* phase is stabilized with respect to the *op* phase, until the difference in osmotic

https://doi.org/10.1021/jacs.1c00522
J. Am. Chem. Soc. 2021, 143, 4143−4147

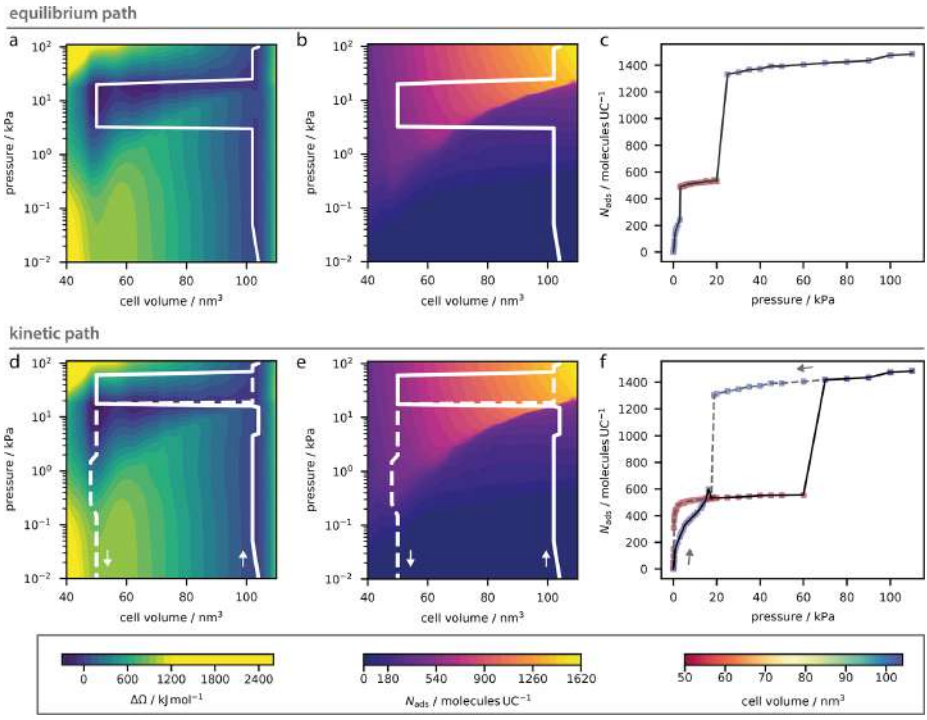


Figure 3. Osmotic potential (a) and the quantity adsorbed (b) surfaces of methane adsorption on DUT-49 at 120 K, with the thermodynamic equilibrium path, resulting in the equilibrium adsorption isotherm (c). The adsorption (solid line) and desorption (dashed line) path for kinetically accessible states on the same surfaces (d and e) that produce the adsorption isotherm (f).

potential between both phases vanishes at a gas pressure of 3.25 kPa. At this point, a transition occurs along the equilibrium path to the *cp* phase (Figure 3a). This contraction produces an increase in amount adsorbed, a step of 236 molecules per unit cell (molecules UC^{-1}). Conversely, no transitions occur along the kinetic path at this pressure, as a barrier between the *op* and *cp* phase of 591 kJ mol^{-1} remains. This barrier only vanishes at a gas pressure of 17.5 kPa, at which the *op* → *cp* transition occurs, associated with a NGA step of 86 molecules UC^{-1}. This predicted transition pressure is remarkably close to the observed value of 25 kPa. Although the predicted NGA step is underestimated compared to experiments (102 molecules UC^{-1}), the observation that the NGA phenomenon can be captured by atomistic simulation is remarkable in itself. On further increasing the gas pressure to 25 kPa, the *op* phase becomes the equilibrium phase, shown by the second transition in Figure 3b. However, the barrier between these phases only disappears at a higher gas pressure of 70 kPa (Figure 3d).

Just as a trajectory can be walked toward increasing gas pressures, the same process can be done in reverse to track desorption. The same path is obtained for the equilibrium path, the experimentally observed hysteresis is absent with this approach. Kinetic paths are different in adsorption and desorption. Osmotic potential energy barriers during desorption, which prevent an *op* → *cp* transition, only disappear when the gas pressure decreases to 17.5 kPa. This prediction of an *op* → *cp* transition during desorption at a gas pressure similar to that of the *op* → *cp* transition during adsorption is in agreement with experiments, attributed to the flat osmotic potential landscape over the entire volume range at this pressure. Interestingly, on decreasing the gas pressure further, it is predicted from this surface that the framework remains locked into the *cp* phase, as observed experimentally. This is due to the barrier surrounding the *cp* local minimum of the osmotic potential landscape at low gas pressures.

The structural transitions derived from the kinetic pathway during adsorption and desorption at 90 K agree quantitatively with reported in situ powder X-ray diffraction experiments.[7] The presence of intermediate phases between the *op* and *cp* phases at approximately 2 kPa are reproduced, evidencing the flatness of osmotic landscape at this gas pressure.

Recently, we reported methane adsorption on DUT-49 for the range 91−190 K, observing a non-monotonic trend of NGA amount with temperature and a distinct temperature window.[19] To provide insight into this trend, simulations were completed for three temperatures (90, 120, and 150 K). The difference in osmotic potential energy between the *op* and *cp* phases (corresponding to 104 and 50 nm^3, respectively) is directly computed from these simulations, as well as the

https://doi.org/10.1021/jacs.1c00522
J. Am. Chem. Soc. 2021, 143, 4143−4147

osmotic potential energy barrier from *op* to *cp* (Figure 4a). The pressure window for stability of the *cp* phase (where $\Omega_{op} - \Omega_{cp} > 0$) is significantly affected by temperature. This results in the *op* → *cp* phase transitions becoming unfavorable at 150 K, in line with the absence of experimentally observed phase transitions at this temperature and beyond.[19] We can alternatively consider these trends with respect to the difference in adsorbed amount in the *op* and *cp* phases (Figure 4b), representing the amount of NGA (N_{NGA}). The barrier present at 90 K disappears before the isotherm crossing point, implying that NGA is suppressed, whereas at 120 K, the barrier only disappears after the *op* state has more methane adsorbed than the *cp* state, necessitating the release of gas and the NGA step. By considering the amount adsorbed in each of the phases when the osmotic potential barrier for the *op* → *cp* transition disappears, we can predict the amount of NGA (Figure 4b, inset). This qualitative description outlines how NGA must be a function of not only the difference in osmotic potential but also the barrier to produce a temperature window.

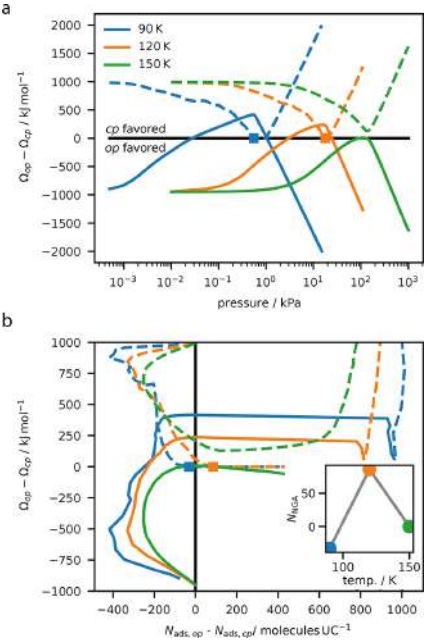


Figure 4. Difference in osmotic potential between *op* and *cp* phases (solid lines) and the barrier height between these phases (dashed lines) (a). The same difference in osmotic potential and barriers displayed with respect to the difference in amount adsorbed (b). The point where the osmotic potential barrier disappears is marked by a square. Predicted amount of NGA in units of molecules UC^{-1} (b, inset).

We have outlined an approach to investigate a responsive gas adsorption process using atomistic simulations to create an osmotic potential surface. Roaming this surface using equilibrium or kinetic trajectories produces very different isotherms. It is only by considering a kinetic pathway that accounts for osmotic potential barriers that the salient features of complex adsorption processes are produced. We used this model to understand the effects of temperature, demonstrating the previously poorly understood non-monotonic trend of NGA magnitude. The presented methodology enables ab initio prediction of NGA and permits investigation of similar counterintuitive adsorption processes. This is an important step toward predicting the evolution of conformational energy landscapes of flexible porous materials upon guest adsorption.[28] In particular, this new insight into the barriers present in responsive adsorption processes is of utmost importance for the intended application of these materials in kinetic selectivity and recognition. Many separation process are emphasized for flexible porous materials, but hysteresis is rarely addressed and poorly understood.[29] This approach highlights how desorption can access an entirely different pathway than adsorption. Flexible systems can wander beyond equilibrium, and it is only by mapping the complete energy surface that we can understand these unique adsorption phenomena and pave the way for new adsorbent technologies.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.1c00522.

Details of the classical potentials, the simulation procedures and additional figures of the computed surfaces and adsorption isotherms (PDF)

Movie showing the osmotic landscape at 120 K (MP4)

■ AUTHOR INFORMATION

Corresponding Authors

Toon Verstraelen − *Center for Molecular Modeling, Ghent University, 9052 Zwijnaarde, Belgium;* orcid.org/0000-0001-9288-5608; Email: toon.verstraelen@ugent.be

Jack D. Evans − *Department of Inorganic Chemistry, Technische Universität Dresden, 01062 Dresden, Germany;* orcid.org/0000-0001-9521-2601; Email: jack.evans@tu-dresden.de

Authors

Ruben Goeminne − *Center for Molecular Modeling, Ghent University, 9052 Zwijnaarde, Belgium*

Simon Krause − *Stratingh Institute for Chemistry, Faculty of Mathematics and Natural Sciences, University of Groningen, 9747 AG Groningen, The Netherlands;* orcid.org/0000-0001-9504-8514

Stefan Kaskel − *Department of Inorganic Chemistry, Technische Universität Dresden, 01062 Dresden, Germany*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.1c00522

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

J.D.E. and S. Krause acknowledge the support of the Alexander von Humboldt foundation. High-performance computing platforms were provided by the Center for Information Services and High Performance Computing (ZIH) at TU Dresden. R.G. and T.V. acknowledge the financial support

https://doi.org/10.1021/jacs.1c00522
J. Am. Chem. Soc. 2021, 143, 4143−4147

from the Research Board of Ghent University (BOF), the computational resources (Stevin Supercomputer Infrastructure), and services provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, FWO, and the Flemish Government, department EWI. S. Kaskel received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 742743).

REFERENCES

(1) Davis, M. E. Ordered porous materials for emerging applications. *Nature* **2002**, *417*, 813−821.
(2) Rabone, J.; Yue, Y.-F.; Chong, S. Y.; Stylianou, K. C.; Bacsa, J.; Bradshaw, D.; Darling, G. R.; Berry, N. G.; Khimyak, Y. Z.; Ganin, A. Y.; Wiper, P.; Claridge, J. B.; Rosseinsky, M. J. An Adaptable Peptide-Based Porous Material. *Science* **2010**, *329*, 1053−1057.
(3) Mason, J. A.; Oktawiec, J.; Taylor, M. K.; Hudson, M. R.; Rodriguez, J.; Bachman, J. E.; Gonzalez, M. I.; Cervellino, A.; Guagliardi, A.; Brown, C. M.; Llewellyn, P. L.; Masciocchi, N.; Long, J. R. Methane storage in flexible metal−organic frameworks with intrinsic thermal management. *Nature* **2015**, *527*, 357−361.
(4) Nijem, N.; Wu, H.; Canepa, P.; Marti, A.; Balkus, K. J.; Thonhauser, T.; Li, J.; Chabal, Y. J. Tuning the gate opening pressure of Metal-Organic Frameworks (MOFs) for the selective separation of hydrocarbons. *J. Am. Chem. Soc.* **2012**, *134*, 15201−15204.
(5) Zhang, J.-P.; Zhou, H.-L.; Zhou, D.-D.; Liao, P.-Q.; Chen, X.-M. Controlling flexibility of metal−organic frameworks. *National Science Review* **2018**, *5*, 907−919.
(6) Coudert, F.-X. Responsive Metal-Organic Frameworks and Framework Materials: Under Pressure, Taking the Heat, in the Spotlight, with Friends. *Chem. Mater.* **2015**, *27*, 1905−1916.
(7) Krause, S.; Bon, V.; Senkovska, I.; Stoeck, U.; Wallacher, D.; Többens, D. M.; Zander, S.; Pillai, R. S.; Maurin, G.; Coudert, F.-X.; Kaskel, S. A pressure-amplifying framework material with negative gas adsorption transitions. *Nature* **2016**, *532*, 348−352.
(8) Bon, V.; Krause, S.; Senkovska, I.; Grimm, N.; Wallacher, D.; Többens, D. D.; Kaskel, S. Massive Pressure Amplification by Stimulated Contraction of Mesoporous Frameworks. *ChemRxiv*, December 1, 2021, ver. 3. DOI: 10.26434/chemrxiv.12758015.v3 (accessed 2021−03−01).
(9) Evans, J. D.; Bocquet, L.; Coudert, F.-X. Origins of Negative Gas Adsorption. *Chem.* **2016**, *1*, 873−886.
(10) Vanduyfhuys, L.; Van Speybroeck, V. Unraveling the thermodynamic conditions for negative gas adsorption in soft porous crystals. *Commun. Phys.* **2019**, *2*, 102.
(11) Krause, S.; Hosono, N.; Kitagawa, S. Chemistry of Soft Porous Crystals: Structural Dynamics and Gas Adsorption Properties. *Angew. Chem., Int. Ed.* **2020**, *59*, 15325−15341.
(12) Coudert, F. X.; Jeffroy, M.; Fuchs, A. H.; Boutin, A.; Mellot-Draznieks, C. Thermodynamics of guest-induced structural transitions in hybrid organic-inorganic frameworks. *J. Am. Chem. Soc.* **2008**, *130*, 14294−14302.
(13) Zang, J.; Nair, S.; Sholl, D. S. Osmotic ensemble methods for predicting adsorption-induced structural transitions in nanoporous materials using molecular simulations. *J. Chem. Phys.* **2011**, *134*, 184103.
(14) Rogge, S. M. J.; Goeminne, R.; Demuynck, R.; Gutiérrez-Sevillano, J. J.; Vandenbrande, S.; Vanduyfhuys, L.; Waroquier, M.; Verstraelen, T.; Van Speybroeck, V. Modeling Gas Adsorption in Flexible Metal−Organic Frameworks via Hybrid Monte Carlo/Molecular Dynamics Schemes. *Adv. Theory Simul.* **2019**, *2*, 1800177.
(15) Dunne, L. J.; Manos, G. Exact matrix treatment of an osmotic ensemble model of adsorption and pressure induced structural transitions in metal organic frameworks. *Dalton Transactions* **2016**, *45*, 4213−4217.
(16) Boutin, A.; Coudert, F.-X.; Springuel-Huet, M.-A.; Neimark, A. V.; Férey, G.; Fuchs, A. H. The Behavior of Flexible MIL-53(Al) upon CH4 and CO2 Adsorption. *J. Phys. Chem. C* **2010**, *114*, 22237−22244.
(17) Llewellyn, P. L.; Maurin, G.; Devic, T.; Loera-Serna, S.; Rosenbach, N.; Serre, C.; Bourrelly, S.; Horcajada, P.; Filinchuk, Y.; Férey, G. Prediction of the Conditions for Breathing of Metal Organic Framework Materials Using a Combination of X-ray Powder Diffraction, Microcalorimetry, and Molecular Simulation. *J. Am. Chem. Soc.* **2008**, *130*, 12808−12814.
(18) The gas pressure equals the mechanical pressure and also affects the chemical potential of the fluid.
(19) Krause, S.; Evans, J. D.; Bon, V.; Senkovska, I.; Coudert, F.-X.; Többens, D. M.; Wallacher, D.; Grimm, N.; Kaskel, S. The role of temperature and adsorbate on negative gas adsorption transitions of the mesoporous metal−organic framework DUT-49. *Faraday Discuss.* **2021**, *225*, 168−183.
(20) Martin, G. M.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of *n*-Alkanes. *J. Phys. Chem. B* **1998**, *102*, 2569.
(21) Faller, R.; de Pablo, J. J. Constant pressure hybrid molecular dynamics-Monte Carlo simulations. *J. Chem. Phys.* **2002**, *116*, 55−59.
(22) Banaszak, B. J.; Faller, R.; de Pablo, J. J. Simulation of the effects of chain architecture on the sorption of ethylene in polyethylene. *J. Chem. Phys.* **2004**, *120*, 11304−11315.
(23) Ghoufi, A.; Maurin, G. Hybrid monte carlo simulations combined with a phase mixture model to predict the structural transitions of a porous Metal-organic framework material upon adsorption of guest molecules. *J. Phys. Chem. C* **2010**, *114*, 6496−6502.
(24) Rogge, S. M. J.; Caroes, S.; Demuynck, R.; Waroquier, M.; Van Speybroeck, V.; Ghysels, A. The Importance of Cell Shape Sampling To Accurately Predict Flexibility in Metal−Organic Frameworks. *J. Chem. Theory Comput.* **2018**, *14*, 1186−1197.
(25) Peng, D.-Y.; Robinson, D. B. A New Two-Constant Equation of State. *Ind. Eng. Chem. Fundam.* **1976**, *15*, 59−64.
(26) Evans, J. D.; Krause, S.; Kaskel, S.; Sweatman, M. B.; Sarkisov, L. Exploring the thermodynamic criteria for responsive adsorption processes. *Chemical Science* **2019**, *10*, 5011−5017.
(27) Numaguchi, R.; Tanaka, H.; Watanabe, S.; Miyahara, M. T. Simulation study for adsorption-induced structural transition in stacked-layer porous coordination polymers: Equilibrium and hysteretic adsorption behaviors. *J. Chem. Phys.* **2013**, *138*, 054708.
(28) Katsoulidis, A. P.; Antypov, D.; Whitehead, G. F. S.; Carrington, E. J.; Adams, D. J.; Berry, N. G.; Darling, G. R.; Dyer, M. S.; Rosseinsky, M. J. Chemical control of structure and guest uptake by a conformationally mobile porous material. *Nature* **2019**, *565*, 213−217.
(29) Rahman, S.; Arami-Niya, A.; Yang, X.; Xiao, G.; Li, G. K.; May, E. F. Temperature dependence of adsorption hysteresis in flexible metal organic frameworks. *Commun. Chem.* **2020**, *3*, 186.

NOTE ADDED AFTER ISSUE PUBLICATION

This article was initially published with an incorrect copyright statement and was corrected on or around May 5, 2021.

https://doi.org/10.1021/jacs.1c00522
J. Am. Chem. Soc. 2021, 143, 4143−4147

Paper III

Accurate Transferable Polarization Model Derived from the Monomer Electron Density

R. Goeminne, T. Verstraelen

R. Goeminne performed the simulations, interpreted the resulting data and wrote the manuscript.

Manuscript in preparation

PCCP

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Accurate transferable polarization model derived from the monomer electron density†

Ruben Goeminne,[a] and Toon Verstraelen*[a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Force field have for decades proven to be an indispensable tool for molecular simulations which are out of reach for *ab initio* methods. Recent efforts to improve the accuracy of these simulations have focused on the inclusion of many-body interactions in force fields. In this regard, we propose a transferable inducible dipole model which requires only the monomer electron density as input, without the need for atom type specific parameters. Slater dipoles are introduced, the widths of which are derived from the *ab initio* monomer density. An additional exchange-repulsion interaction is introduced in our model, modeled as an overlap of densities. This interaction, previously neglected in point dipole models, can be grounded in a second order expansion of the universal DFT functional in changes of the density and external potential. The inclusion of this interaction is shown to significantly improve the prediction of molecular polarizabilities and three-body energies. Our model is incorporated in a previously proposed non-covalent force field and is benchmarked on interaction energies of dimers contained in the hsg and hbc6 datasets, demonstrating competitive accuracy with common fitted force fields. The inherent transferability of our model makes it widely applicable, for example to systems for which no specifically fitted parameters for polarization models are available in the literature.

1 Introduction

The importance of non-additive effects in molecular simulations has long been established. Most efforts to include these many-body effects have focused on modeling the polarization component. This interaction is known to be important for the solvation free energy of salt ions and amines[1–3], cation-π interactions[4], the modeling of polarizable organic compounds[5–10] and the various anomalous properties of water.[11–15] Models to include electronic polarization can generally be divided into the Drude oscillator,[16–20] fluctuating charge[21] and inducible dipole models.[22] The Drude oscillator or charge-on-spring model introduces an auxiliary charged particle which is attached to the polarizable center by a parabolic restraint. In the fluctuating charge model, the magnitude of the charge on the polarizable center itself is allowed to fluctuate. The inducible dipole model, on the other hand, does not alter the monopoles but introduces a basis of atomic dipole response functions at each atomic site. The dipole response has previously been modeled by point dipoles, in which case damping functions are required to avoid the polarization catastrophe[15,23], or Gaussian dipoles, for which widths of Gaussian charge distributions are required. Furthermore, the parametrization of polarizabilities usually requires the introduction of a set of atom types, dependent on both the element and its chemical environment.[24,25] The atomic polarizabilities are then determined by fitting to the experimental or *ab initio* polarizability tensors for a set of molecules or functional groups which contain the atom type of interest, or by probing molecules with point charges or external electric fields.[23,24,26] However, these methods limit the model's applicability to atoms in molecules with similar chemical environments as in the training set.

This work is based on the inducible dipole model. Compared to models in the literature, we aim foremost at a fully transferable polarization model. It is transferable in a different sense than previous inducible dipole models. Usually, the term refers to the ability of a model to generalize to molecules outside the training set, which are however composed of atom types included in the training set. In contrast, our model makes use of the ground state electron density, obtained from a single *ab initio* calculation of the monomer(s) of interest. Making used of a single interaction parameter, good transferability is observed to molecules outside the training set, making the model easily applicable to new systems without the need to define atom types.

Our model introduces Slater dipoles as response functions, and includes an additional exchange-repulsion interaction, together with the previously used classical electrostatic interaction. The exchange-repulsion interaction is modeled as a proportionality

[a] *Center for Molecular Modeling (CMM), Ghent University, Technologiepark 46, 9052 Zwijnaarde, Belgium; E-mail: Toon.Verstraelen@UGent.be*

† Electronic Supplementary Information (ESI) available: [The fitting procedure and results for the Slater dipole widths and amplitudes.]. See DOI: 10.1039/cXCP00000x/

with the overlap between electron densities, based on previous observations of this proportionality.[27–30] After the inducible dipole model is introduced in Section 2, and Slater widths for each chemical element are determined in Section 3.1, we validate our model by its performance in the prediction of a large set of molecular polarizabilities in Section 3.2 and cluster many-body interaction energies in Section 3.3. In this way, we can directly compare the non-additive component of our model to *ab initio* many-body energies, ensuring our model exhibits the correct many-body behavior. In Section 3.4, the inducible dipole model is included in the non-covalent monomer electron density force field (MEDFF), which decomposes the interaction energy in four interaction terms in accordance with SAPT.[31] Similar to the newly proposed polarization model, MEDFF only requires the monomer electron density as input. It comprises 3 interaction parameters which were fit to a set of dimer interaction energies, one of which is replaced by including the new inducible dipole model with its single interaction parameter. The resulting polarizable monomer electron density force field (PMEDFF) is then benchmarked on dimer interactions to which no fitting was performed in Section 3.5. For several dimers, no parameters were available for the commonly used AMOEBA polarizable force field,[11] demonstrating the utility of our model's transferability to novel intermolecular interactions.

2 Inducible dipole model

A physical inducible dipole model is built up from essentially three components. The first is a determination of the atomic polarizabilities. The second is a representation of the ground state electron density and the functional form of the atomic dipole response functions. The last component is a model for the interactions of the induced dipoles with the ground state and for the interactions between the induced dipoles. These interactions have previously been approximated as purely electrostatic contributions. As the proposed polarization model differs in its three components from frequently-used inducible dipole models in the literature, these separate components are introduced first.

2.1 Atomic polarizabilities

Atomic polarizabilities are usually derived by defining a limited set of atom types, after which the respective atomic polarizabilities are fit to reproduce the experimental[32] or *ab initio*[24] derived molecular polarizability tensor. Our approach differs from this methodology, as our intent is to derive all model parameters except a single interaction parameter from an *ab initio* calculation of the molecule of interest. Therefore, we start from the free atom polarizabilities obtained from linear response time-dependent density functional theory.[33] Subsequently, to account for the chemical environment, the isotropic polarizability of each atom $\alpha_{i,\text{free}}$ is rescaled proportionally to the effective volume (V_{aim}) of the atom-in-molecule

$$\alpha_i = \frac{V_{\text{aim}}}{V_{\text{free}}}\,\alpha_{i,\text{free}} = \frac{\int r^3 \rho_i(\vec{r})\,d\vec{r}}{\int r^3 \rho_{i,\text{free}}(\vec{r})\,d\vec{r}}\,\alpha_{i,\text{free}} \tag{1}$$

with V_{free} the effective volume of the free atom in vacuum. This rescaling of polarizabilities was proposed previously in the exchange-hole dipole moment (XDM) and Tkatchenko-Scheffler dispersion models.[34,35] For consistency with MEDFF, all ground state densities were calculated at the spin-polarized B3LYP/aug-cc-pVTZ level of theory. Similarly, for consistency with the Slater dipoles derived in the following Section, the minimal basis iterative stockholder (MBIS) scheme was used to partition the molecular density into atomic fragments.[36]

2.2 Slater dipole response functions

Previously, both point charges and Gaussian charge distributions have been used to represent the ground state electron density, together with point dipoles and Gaussian dipoles as response functions. However, as the true electron density tails off exponentially, atom-centered Slater functions are used in this work to represent the ground state.[37–39] To this end, the MBIS scheme[36] is used to partition the *ab initio* molecular density into atom-centered core charges q_A^{c} and valence 1s Slater functions of the form:

$$\rho_A^{\text{s}}(\vec{r}) = \frac{N_A}{8\pi{\sigma_{A,\text{s}}}^3}\exp\left(-\frac{\left|\vec{r}-\vec{R}_A\right|}{\sigma_{A,\text{s}}}\right) + q_A^{\text{c}}\,\delta(\vec{r}-\vec{R}_A) \tag{2}$$

with N_A the population and $\sigma_{A,\text{s}}$ the width of the distribution, both of which are fitted to the *ab initio* density by minimizing the Kullback-Leibler divergence. This partitioning has been show to accurately reproduce dimer electrostatic interactions, while being robust with respect to small changes in the electronic structure calculation from which it is derived[36]. A Slater dipole can now be constructed as the gradient of a normalized 1s Slater function. However, as this function has no well defined limit towards $\vec{R}_A$, we alternatively use a normalized 1s+2s function:

$$\vec{\rho}_A^{\text{p}}(\vec{r}) = \vec{\nabla}_A\,\rho_A^{\text{1s+2s}} \tag{3}$$

$$= \vec{\nabla}_A\left[\frac{1}{32\pi{\sigma_{A,\text{p}}}^3}\exp\left(-\frac{\left|\vec{r}-\vec{R}_A\right|}{\sigma_{A,\text{p}}}\right)\left(1+\frac{\left|\vec{r}-\vec{R}_A\right|}{\sigma_{A,\text{p}}}\right)\right] \tag{4}$$

$$= \frac{1}{32\pi{\sigma_{A,\text{p}}}^5}\exp\left(-\frac{\left|\vec{r}-\vec{R}_A\right|}{\sigma_{A,\text{p}}}\right)\left(\vec{r}-\vec{R}_A\right) \tag{5}$$

This Slater dipole can be interpreted as the density difference between an unperturbed ground state electron density and a density perturbed by an electric dipole field. From this interpretation, the dipole width $\sigma_{A,\text{p,free}}$ for each free atom can be determined by fitting a Slater dipole to the normalized *ab initio* density difference of a free atom. More details are provided in Section 3.1. To account for the chemical environment of each atom in a molecule or solid, every free atom dipole width $\sigma_{A,\text{p,free}}$ is subsequently scaled with the cube root of the ratio of its effective volume in a molecule compared to its effective volume in vacuum[35], similarly to the rescaling used for the polarizabilities in Section 2.1:

$$\sigma_{A,\text{p}} = \left(\frac{\int r^3\rho_A(\vec{r})\,d\vec{r}}{\int r^3\rho_{A,\text{free}}(\vec{r})\,d\vec{r}}\right)^{1/3}\sigma_{A,\text{p,free}} \tag{6}$$

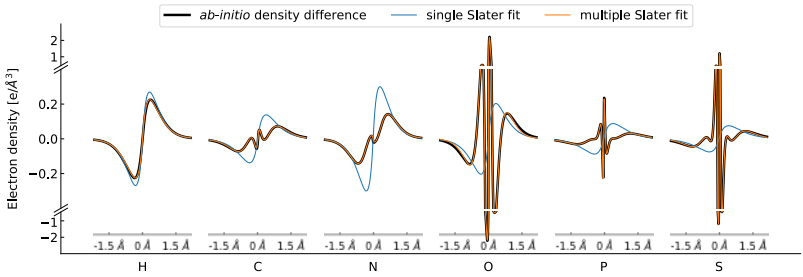


Fig. 1 *Ab initio* density difference between the unperturbed free atoms and atoms in a dipole field along the p-axis (in black) with $p \in \{\mathbf{x},\mathbf{y},\mathbf{z}\}$ that minimizes the energy, the Slater function fit to the *ab initio* expectation value of p^3 (in blue), and the least squares fit of multiple Slater functions to the *ab initio* density (in orange).

A similar functional form to Eq. 5 has been proposed before, although dipole widths were fit for specific atom types instead of making use of *ab initio* atomic density differences.[20]

2.3 Interaction model

To understand the approximations made in inducible dipole models,[9,24,26,40–42], let us ground the inducible dipole model in density-functional theory (DFT). These derivations build upon the work of York *et al.* and Tabacchi *et al.*[43,44] Consider a system of N interacting subsystems, for example a set of molecules or a molecule with an adsorbent, each described by a nuclear charge distribution $\rho_i^n(\vec{r})$ and ground state electron density $\rho_i^0(\vec{r})$ of the isolated subsystem. The total electron density can be defined as a sum of the reference densities and the density response induced due to the interaction of the subsystems

$$\rho(\vec{r}) = \rho^0(\vec{r}) + \delta\rho(\vec{r}) \text{ with } \begin{cases} \rho^0(\vec{r}) = \sum_i \rho_i^0(\vec{r}) \\ \delta\rho(\vec{r}) = \sum_i \delta\rho_i(\vec{r}) \end{cases} \tag{7}$$

This system can be characterized by the energy functional:

$$E[\rho, v] = F[\rho] + \int \rho(\vec{r}) v(\vec{r})\, \mathrm{d}\vec{r} \tag{8}$$

with $v(\vec{r})$ the external potential generated by the nuclei and $F[\rho]$ the universal functional. Substituting Equation 7 in the energy functional and expanding to second order in changes of the density and external potential (see Tabacchi *et al.* [44]) yields

$$\begin{aligned} E[\rho^0 + \delta\rho, v^0 + \delta v] = E[\rho^0, v^0] + \int \rho^0(\vec{r}) \delta v(\vec{r})\, \mathrm{d}\vec{r} + \int \delta\rho(\vec{r}) v(\vec{r})\, \mathrm{d}\vec{r} \\ + \sum_k \int \frac{\delta F}{\delta\rho(\vec{r})} \delta\rho_k(\vec{r})\, \mathrm{d}\vec{r} \\ + \frac{1}{2} \sum_{k,l} \int\int \delta\rho_k(\vec{r}) \left(\frac{\delta^2 F}{\delta\rho(\vec{r}) \delta\rho(\vec{r}')} \right) \delta\rho_l(\vec{r}')\, \mathrm{d}\vec{r}\mathrm{d}\vec{r}' \end{aligned} \tag{9}$$

From this, all terms which depend on a change in density $\delta\rho$ can be gathered into a term called the polarization energy E_{PFF}. Furthermore, the external potential $v(\vec{r})$ is assumed to be generated by the nuclear charges $\rho^n(\vec{r})$ (no external fields are present)

$$v(\vec{r}) = \int \frac{\rho^n(\vec{r}')}{|\vec{r} - \vec{r}'|}\, \mathrm{d}\vec{r}' \text{ with } \rho^n(\vec{r}) = \sum_k \rho_k^n(\vec{r}) \tag{10}$$

with this, E_{PFF} can be written as

$$\begin{aligned} E_{\mathrm{PFF}} = \sum_k \int \left[\int \frac{\rho^n(\vec{r}')}{|\vec{r} - \vec{r}'|}\, \mathrm{d}\vec{r}' + \frac{\delta F}{\delta\rho(\vec{r})} \right] \delta\rho_k(\vec{r})\, \mathrm{d}\vec{r} \\ + \frac{1}{2} \sum_{k,l} \int\int \delta\rho_k(\vec{r}) \left(\frac{\delta^2 F}{\delta\rho(\vec{r}) \delta\rho(\vec{r}')} \right) \delta\rho_l(\vec{r}')\, \mathrm{d}\vec{r}\mathrm{d}\vec{r}' \end{aligned} \tag{11}$$

This Equation can be seen as a starting point for the developments of inducible dipole models. Within a Kohn-Sham formulation of DFT, the universal functional $F[\rho]$ is decomposed as $F[\rho] = T[\rho] + V_{ee}[\rho] = T_S[\rho] + V_H[\rho] + E_{xc}[\rho]$. Only the functional derivatives of the Hartree term are known exactly

$$\frac{\delta V_H[\rho]}{\delta\rho(\vec{r})} = \int \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|}\, \mathrm{d}\vec{r}', \quad \frac{\delta^2 V_H[\rho]}{\delta\rho(\vec{r}) \delta\rho(\vec{r}')} = \frac{1}{|\vec{r} - \vec{r}'|} \tag{12}$$

Therefore, classical inducible dipole models only retain this term ($F[\rho] \approx V_H[\rho]$). The density response is now expanded in a dipole basis set centered at each atom a, with $i \in \{x, y, z\}$:

$$\delta\rho(\vec{r}) = \sum_a \sum_{i=x,y,z} c_{a,i} d_a^i(\vec{r}) \tag{13}$$

Due to the choice of dipole basis set, the diagonal elements of the second functional derivative of $F[\rho]$ with respect to the density response are equal to the inverse of the dipole polarizability.[45] Assuming isotropic atomic polarizabilities, Equation 11 reduces

to

$$E_{\mathrm{PFF,cl}} = \sum_a \sum_i c_{a,i} \int\int \frac{\rho^0_{\mathrm{tot}}(\vec{r}')d^i_a(\vec{r})}{|\vec{r}-\vec{r}'|}\,\mathrm{d}\vec{r}\,\mathrm{d}\vec{r}' + \frac{1}{2}\sum_{a\neq b}\sum_{i,j} c_{a,i}c_{b,j}\int\int \frac{d^i_a(\vec{r})d^j_b(\vec{r}')}{|\vec{r}-\vec{r}'|}\,\mathrm{d}\vec{r}\mathrm{d}\vec{r}' + \frac{1}{2}\sum_{a,i}\frac{c^2_{a,i}}{\alpha_a} \quad (14)$$

with α_a the atomic polarizability of atom a and ρ^0_{tot} the sum of the nuclear charges ρ^n and the reference electron density ρ^0. Herein, the first term represents the classical interaction of the total charge density with a dipole on atom a, with the second term the classical interaction between two dipoles located on atoms a and b. Lastly, the third term represents the energy required to generate the dipoles. Minimizing $E_{\mathrm{PFF,cl}}$ with respect to the the magnitude of the induced dipoles $c_{a,i}$, the classical inducible dipole model is obtained (not considering the usually applied damping functions[23] and scaling factors for interactions between atoms separated by up to 4 bonds[46]). However, in this derivation, the interactions were assumed to be purely Coulombic, neglecting the Pauli exclusion principle.[47] While this constraint requiring antisymmetrization of the wavefunction with respect to the exchange of two electrons clearly affects the electron-electron interaction $V_{ee}[\rho]$, it also affects the kinetic energy T. This is due to the antisymmetrization leading to more strongly oscillating wavefunctions, leading to an increase in kinetic energy.[48] The exchange interaction has previously been neglected due to the fact that it cannot be represented as a functional of the interacting electron densities. However, an approximate proportionality between this interaction and the overlap of the electron densities has been observed previously.[27–30] Recently, the proportionality factor was determined by fitting to SAPT2+3 exchange-repulsion interaction energies of the dispersion dominated dimers in the S66x8 set, and is on average equal to 8.13 a.u.[30,31,49]. This overlap model reproduces SAPT2+3 exchange-repulsion energies with a root mean square error of 1.22 kcal/mol. This additional interaction is now included in our interaction kernel:

$$K(\vec{r},\vec{r}') = \frac{\delta^2 F[\rho]}{\delta\rho(\vec{r})\delta\rho(\vec{r}')} \approx \frac{1}{|\vec{r}-\vec{r}'|} + U_{\mathrm{ex}}\,\delta(\vec{r}-\vec{r}') \quad (15)$$

where U_{ex} represents the proportionality between the overlap in electron density and the exchange-repulsion energy. Note that the need for an additional interaction which induces the dipoles has been noted before, with the magnitude of induced dipoles being consistently underestimated at intermediate distances.[26,30] With this new interaction kernel, the following polarizable model is obtained, in contrast to the classical model from Equation 14:

$$E_{\mathrm{PFF}} = \sum_a \sum_i c_{a,i}\left[\int\int \frac{\rho^{a^*}_{\mathrm{tot}}(\vec{r}')d^i_a(\vec{r})}{|\vec{r}-\vec{r}'|}\,\mathrm{d}\vec{r}\,\mathrm{d}\vec{r}' + U_{\mathrm{ex}}\int \rho^{a^*}_{\mathrm{tot}}(\vec{r}')d^i_a(\vec{r})\,\mathrm{d}\vec{r}\right] + \frac{1}{2}\sum_{a\neq b}\sum_{i,j} c_{a,i}c_{b,j}\left[\int\int \frac{d^i_a(\vec{r})d^j_b(\vec{r}')}{|\vec{r}-\vec{r}'|}\,\mathrm{d}\vec{r}\mathrm{d}\vec{r}' + U_{\mathrm{ex}}\int d^i_a(\vec{r})d^j_b(\vec{r})\,\mathrm{d}\vec{r}\right] + \frac{1}{2}\sum_{a,i}\frac{c^2_{a,i}}{\alpha_a} \quad (16)$$

With $\rho^{a^*}_{\mathrm{tot}}$ the sum of the nuclear charges ρ^n and monomer electron density ρ^0 of all monomers except the monomer to which atom a belongs. Herein, the nuclear charges and electron density of the monomer containing atom a are omitted, as the intramolecular polarization effects are already mostly captured in the *ab initio* monomer density used as input. In this polarization model, Slater dipoles can thus be induced to minimize both the electrostatic and exchange-repulsion energy between the ground state electron densities of molecules. Clearly, long-range correlation effects are neglected in this model. However, including the often neglected exchange interaction should significantly increase the physical accuracy of the model by an improved description of this short-range effect.

3 Results and Discussion

3.1 Free atom dipole widths

Before the new inducible dipole model proposed in this work is benchmarked, the free atom dipole widths are determined for the following set of elements: H, C, N, O, F, P, S, Cl and Zn. The response in electron density was determined by applying a dipole field of 0.0001 a.u. to the free atom along the axis $p \in \{\mathbf{x},\mathbf{y},\mathbf{z}\}$ that minimizes the energy. This field strength is small enough to prevent higher order effects, but large enough to prevent numerical instabilities. Electron densities of both the perturbed and unperturbed free atoms were calculated with the CCSD method[50] and the aug-cc-pVQZ basis set[51] using Gaussian09[52]. The basis sets was chosen large enough to obtain sufficiently accurate density tails.

The resulting density differences are shown in Figure 1 for H, C, N, O, P and S. As large fluctuations of the density close to the nucleus are present, a least squares fit of a single Slater dipole yields unreliable results. Therefore, the Slater widths were determined by fitting the third moment $\langle p^3\rangle$ of the Slater dipole along the perturbation axis p to that of the *ab initio* density difference, with $p \in \{\mathbf{x},\mathbf{y},\mathbf{z}\}$ which minimizes the energy. More details are provided in the Supporting information. As seen from Figure 1, this single Slater dipole per atom doesn't capture all fluctuations of the density, but is in reasonable agreement far from the nucleus. To investigate whether these fluctuations cause significant deviations in the energy of our polarization model, an additional fitting of a linear combination of Slater dipoles per atom is also determined. In this case, a least squares fit was used, as it is desired to fit to the exact density fluctuations. Both the amplitudes and Slater widths were fit, and the lowest number of Slater dipoles able to capture all density fluctuations is used. The result

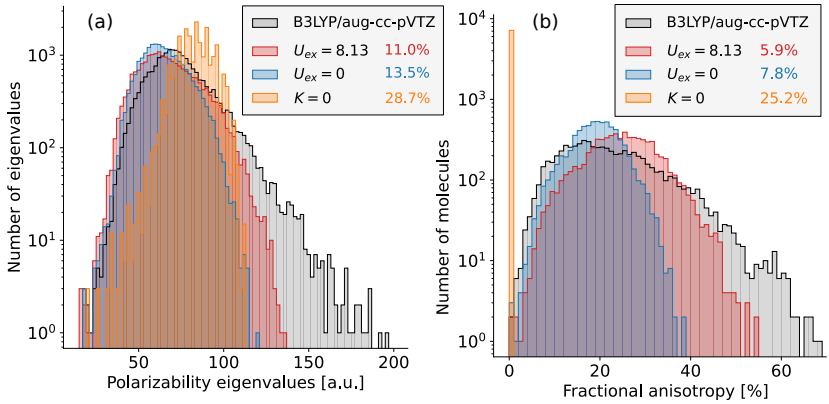


Fig. 2 (a) Eigenvalues of the dipole polarizability matrices of molecules in the QM7 dataset obtained at the reference B3LYP/aug-cc-pVTZ level of theory and with the inducible dipole model with U_{ex}=8.13 in red, U_{ex}=0 in blue and K=0 in orange. The root mean square percentage errors (RMSPEs) compared to the reference are shown in the legend. (b) Fractional polarizability anisotropy of the obtained dipole polarizability matrices. The RMSPEs compared to the reference level of theory are shown in the legend.

are shown in orange in Figure 1. For these elements, a very accurate fit to the *ab initio* density difference is possible using a linear combination of at most 5 Slater dipoles. Note that the widths of the derived atomic Slater dipoles in a molecule are subsequently rescaled proportional to the cube root of the effective volume according to Equation 6 in order to account for their chemical environment.

3.2 Reproduction of the molecular polarizability

A first validation of the polarizable force field model developed in this work is the reproduction of the molecular polarizability tensor. By adding a uniform electric field $V_E = \vec{E}\cdot\vec{r}$ to the external potential in Equation 10, the induced atomic dipoles can be calculated and summed together to equal the induced molecular dipole. If the interaction kernel is turned off ($K(\vec{r},\vec{r'}) = 0$ in Equation 15), the molecular polarizability is simply equal to the sum of the atomic polarizabilities, obtained with Equation 1. Therefore, as we make use of isotropic atomic polarizabilities, the molecular polarizability is isotropic in this case. In contrast, turning on the interaction kernel $K(\vec{r},\vec{r'})$, dipoles on neighboring atoms can interact, giving rise to anisotropic contributions in the molecular polarizability matrix. Equation 15 can therefore be validated by the degree to which it gives rise to the correct molecular anisotropy. To perform this validation, the QM7 dataset is used, composed of 7165 molecules containing up to 23 atoms (including up to 7 non-hydrogen atoms) which were extracted from the larger GDB-13 dataset.[53,54] Both the monomer electron densities and reference dipole polarizabilities were obtained at the B3LYP/aug-cc-pVTZ level of theory using Gaussian16.[55] From the obtained molecular polarizability matrices, both the eigenvalues and fractional polarizability anisotropy (FPA) are calculated. The latter is a metric for the anisotropy ranging from 0 to 1, and is defined as follows[56]

$$\mathrm{FPA} = \sqrt{\frac{1}{2}\frac{\left[(\alpha_{xx}-\alpha_{yy})^2+(\alpha_{xx}-\alpha_{zz})^2+(\alpha_{yy}-\alpha_{zz})^2\right]}{\alpha_{xx}^2+\alpha_{yy}^2+\alpha_{zz}^2}} \qquad (17)$$

The resulting eigenvalues and FPEs of molecules of the QM7 dataset are shown in Figure 2(a) and 2(b), respectively. Three different approximations to the interaction kernel are investigated. In the first, with $K(\vec{r},\vec{r'}) = 0$, the root mean square percentage error (RMSPE) on the polarizability eigenvalues is equal to 28.7%. As shown on in Figure 2(b), this approximation results in isotropic molecular polarizabilities, resulting in a RMSPE on the FPEs of 25.2%. In contrast, only turning on the Coulomb interaction ($K(\vec{r},\vec{r'}) = 1/|\vec{r}-\vec{r'}|$), results in RMSPEs of 13.5% and 7.8% on the eigenvalues and FPEs, respectively. Finally, adding the exchange-repulsion interaction of Equation 15, RMSPEs of 11.0% and 5.9% are obtained on the eigenvalues and FPEs. Clearly, an improvement is seen by adding the exchange-repulsion term with respect to the pure Coulomb term. The ability of the model to reproduce the response of a molecule to an external dipole field is already a first validation. A second validation consists in its ability to accurately predict intermolecular interaction energies, as will be benchmarked in the following Section.

3.3 Benchmarking of many-body energies

3.3.1 Water three-body energies

To validate the developed polarizable force field, its prediction of intermolecular polarization energies is now investigated. In order to benchmark it separately from other intermolecular interactions, our attention is turned to three, four and five-body interaction energies. While two-body terms are composed of electrostatic, exchange, dispersion and induction terms, benchmark systems such as water clusters can be chosen in which the

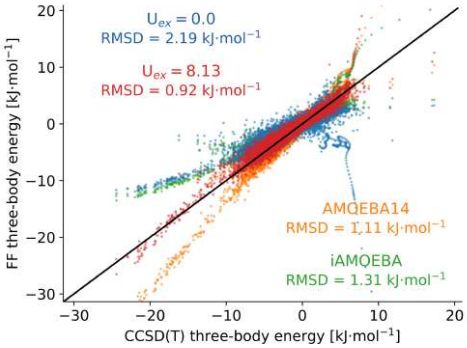


Fig. 3 Parity plot of a set of 12 347 water trimer three-body energies calculated with the inducible dipole model with $U_{ex}=0$ in blue (representing only the Coulomb interaction), $U_{ex}=8.13$ in red (including the exchange interaction), AMOEBA14 in orange and iAMOEBA in green. Results are compared with reference CCSD(T) three-body energies.

higher order terms are dominated by induction.[57,58] In this way, the predicted energies can be explicitly compared with *ab initio* many-body energies. First, the intermolecular three-body energies of a set of 12 347 water trimers in a wide range of conformations are computed with the inducible dipole model using the multiple Slater fit of Figure 1 and compared to reference calculations performed at the CCSD(T)/aug-cc-pVTZ level of theory.[59] In Figure 3, a comparison is shown between the reference energies and those computed with the inducible dipole model without the exchange-repulsion interaction ($U_{ex}=0$) and with the exchange-repulsion interaction ($U_{ex}=8.13$). Significant deviations are observed for the pure Coulomb interaction, significantly underestimating the energy for low-lying trimers. This trend is reversed for trimers with repulsive three-body energies, erroneously predicting stabilizing interactions. The inclusion of exchange-repulsion significantly improves the predicted energies, reducing the root mean square deviation (RMSD) from 2.19 kJ·mol^{-1} to 0.92 kJ·mol^{-1}. Only minor difference are observed when the single Slater dipole functions are employed instead of the multiple Slater functions, increasing the RMSD from 0.92 kJ·mol^{-1} to 1.06 kJ·mol^{-1} when including exchange. This suggests that the inclusion of the precise density fluctuations near the nucleus is not vital to obtain accurate three-body energies, and the single Slater fit can be used for improved computational efficiency. The trimer dataset for this comparison was previously constructed to fit a full-dimensional potential energy function for water (MB-pol).[59] As our model instead aims at being transferable without fitting to the interaction energies of the systems of interest, it cannot be expected to be competitive on the specific case of water with MB-pol. Instead, it is compared with the popular iAMOEBA and AMOEBA14 force fields from the literature.[11,66] These force field result in an RMSD of 1.31 kJ·mol^{-1} and 1.11 kJ·mol^{-1}, respectively. Especially trimers with negative three-body energies are described well by our model (RMSD of 0.79 kJ·mol^{-1}), compared with iAMOEBA and AMOEBA14 (RMSD of 1.52 kJ·mol^{-1} and 1.29 kJ·mol^{-1}, respectively). The main reason why AMOEBA14 can reproduce three-body energies adequately without the inclusion of an exchange interaction originates from its use of point dipoles. As these interact more strongly than delocalized dipoles, the missing exchange interaction is partly compensated. The performance on water three-body energies is an encouraging sign for the model developed herein, as both literature force fields contain parameters fitted specifically to reproduce water interaction energies and condensed phase properties, while no such fitting is performed for our model. In fact, our model contains no parameters fitted on interaction energies of water. Only the monomer electron density of a single optimized water molecule was used, together with a fit for the atomic dipole basis for each element and the U_{ex} parameter fitted previously on the dispersion-dominated dimers of the S66x8 set.

3.3.2 Four and five-body energies of water hexamers

An even more direct validation of the many-body behaviour of our polarization model is the prediction of four and five-body energies. It is generally difficult to fit potentials to higher order contributions in the many-body expansion either implicitly or explicitly, as these contributions are usually significantly smaller than the lower order contributions, and an explicit fitting would require computationally demanding *ab initio* reference data for clusters of 4 and 5 molecules. Therefore, these higher order contributions are a good benchmark of the underlying physics of the polarization model as well as the degree to which a potential relies on error cancellation between terms. For this comparison we use a set of 8 MP2/CBS optimized water hexamer clusters[60] for which the many-body decomposition was calculated by Medders *et al*. with CCSD(T)-F12, TTM3-F, TTM4-F, WHBB5 and MB-pol.[59,61–65,67] Additionally, calculations were performed with AMOEBA14 in Tinker.[7,68] The four and five-body energies are shown in Figure 4(a) and 4(b), respectively. The magnitude of these interactions is significantly overestimated by AMOEBA14, which is likely due to a degree of error cancellation with the two-body contribution. The four and five-body energies predicted by our model are similar to those predicted by TTM4-F and MB-pol. However, the five-body energies are predicted especially well by our model, improving on some small deviations of all models observed for the cyclic chair and cyclic boat hexamers. This is especially noteworthy as our model has not been developed with the specific case of water in mind, and no parameters are present which were fitted to clusters of water molecules, whereas all other models except AMOEBA14 were fitted to interaction energies of clusters containing at least 3 monomers.

3.3.3 3b69 trimer three-body energies

To extend the benchmark on three body energies to a wider range of intermolecular interactions, calculations were additionally performed on the 3b69 set of trimers.[69] Included in this set are 18 out of 24 molecules which are not present in the S66x8 set to which the U_{ex} parameter was fitted. It therefore represent a test of how well the model transfers to interactions to which the parame-

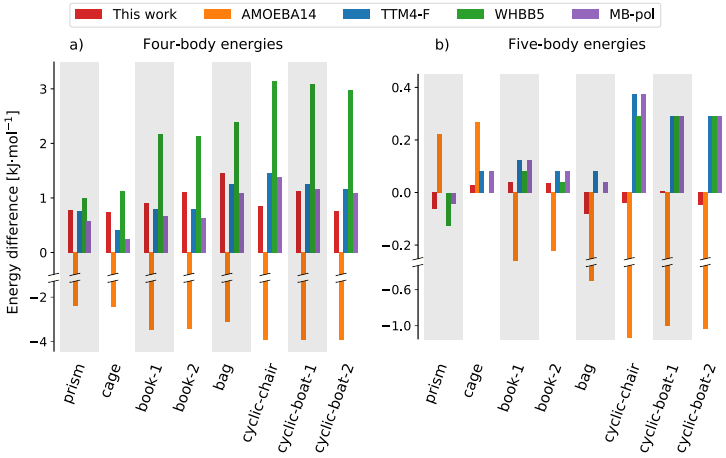


Fig. 4 Error on the four-body (a) and five-body (b) interaction energies of 8 water hexamers as computed with our model and water potentials from the literature, with respect to reference calculations at the CCSD(T)-F12/cc-pVTZ level of theory.[59–65]

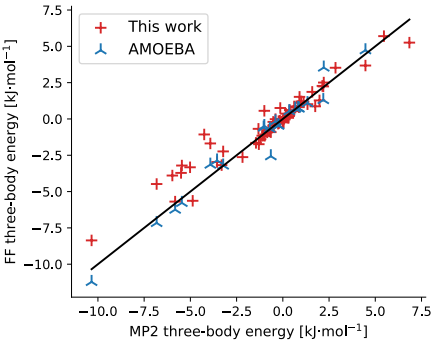


Fig. 5 Parity plot of three-body energies of trimers in the 3b69 set calculated at the MP2/CBS level of theory, compared with the polarization model proposed in this work (red) and the AMOEBA force field (blue).[7]

ter was not fitted. To test the many-body polarization component separately from many-body dispersion, interaction energies were recalculated at the MP2 level of theory extrapolated to the complete basis set. MP2 includes many-body induction effects, but the dispersion non-additivity only appears at the MP3 level, allowing for a comparison between the inducible dipole model and MP2 three-body energies.[69,70] An RMSD of 0.88 kJ·mol^{-1} was obtained, demonstrating good performance across a range of different interactions. In comparison, AMOEBA three-body energies were obtained for the 22 out of the 69 trimers for which parameters are available.[7] For this subset, an RMSD of 0.67 kJ·mol^{-1} is obtained. However, the observation that the application of a regularly used polarizable force field from the literature is limited to certain interactions demonstrates the usefulness of a transferable model, as it can be applied to a wide range of interactions.

3.4 Development of a full non-covalent force field

To predict total non-covalent interaction energies, the developed polarization model has to be combined with models for the other intermolecular interactions. For this, the monomer electron density force field (MEDFF) is ideally suited. In analogy to the *ab initio* energy decomposition made by symmetry-adapted perturbation theory (SAPT), MEDFF separates the total interaction energy in an electrostatic, dispersion, exchange-repulsion and induction interaction. Each one is a functional form of the electron density of the constituting monomers:

$$E_{\mathrm{MEDFF}} = E_{\mathrm{elst}} + E_{\mathrm{disp}} + E_{\mathrm{ex}} + E_{\mathrm{ind}} \tag{18}$$

$$= \sum_{\mathrm{A}} \sum_{\mathrm{B}} \left(E^{\mathrm{AB}}_{\mathrm{elst}} + E^{\mathrm{AB}}_{\mathrm{disp}} + U_{\mathrm{ex}} S^{\mathrm{AB}} - U_{\mathrm{ind}} S^{\mathrm{AB}} \right) \tag{19}$$

with the sum over A and B running over all atoms in the first and second monomer, respectively, and S^{AB} the overlap integral between electron densities ρ of atoms A and B:

$$S^{\mathrm{AB}} = \int \mathrm{d}\vec{r} \rho_A(\vec{r}) \rho_B(\vec{r}) \tag{20}$$

Both the exchange-repulsion and induction interactions are modeled as a proportionality of the overlap integral of electron densities of the monomers (with proportionality factors U_{ex} and U_{ind}). While the exchange-repulsion was shown to be accurately modeled with the overlap model, the picture is more complex for induction. This interaction consists of multiple effects like polarization and charge-transfer effects. Of these, charge-transfer will typically dominate in dispersion-dominated interactions where

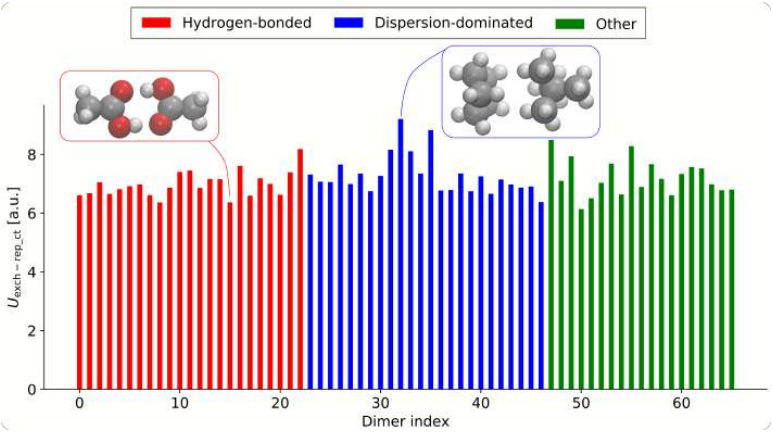


Fig. 6 Fitted values of U_{ex_ct} for each of the 66 dimers present in the S66x10 set. For each dimer, U_{ex_ct} is fitted to the 10 dimer configurations along the dissociation curve. For the hydrogen-bonded dimers, acetic acid is shown as example, while the cyclopentane-neopentane dimer is shown among the dispersion-dominated structures.

polarization plays a minor role. Moreover, charge-transfer is known to be proportional (to good approximation) to the overlap between electron densities. For this reason, MEDFF performs well on benchmarks of dispersion-dominated interactions. However, for hydrogen-bonded interactions in which polarization is dominant, significantly worse performance is observed. Clearly, polarization cannot be accurately captured by a pairwise additive overlap model. Herein lies the value of our previously developed polarization model. Incorporating our polarization model in MEDFF, the polarizable monomer electron density force field (PMEDFF) is obtained:

$$E_{\mathrm{PMEDFF}} = E_{\mathrm{elst}} + E_{\mathrm{disp}} + E_{\mathrm{ex\_ct}} + E_{\mathrm{pol}} \tag{21}$$

$$= \sum_{\mathrm{A}} \sum_{\mathrm{B}} \left(E_{\mathrm{elst}}^{\mathrm{AB}} + E_{\mathrm{disp}}^{\mathrm{AB}} + U_{\mathrm{ex\_ct}} S^{\mathrm{AB}} \right) + E_{\mathrm{pol}} \tag{22}$$

with E_{pol} given by Equation 16. Herein, $U_{\mathrm{ex\_ct}}$ represents the proportionality between the exchange-repulsion and charge transfer interactions and the electron density overlap. The common treatment of both these interactions is further justified by the observation that in the higher order terms of SAPT, exchange-repulsion and induction effects cannot be distinguished. This new proportionality factor $U_{\mathrm{ex\_ct}}$ is therefore fitted to the sum of the SAPT2+(3) exchange-repulsion and induction energies minus the polarization energy E_{pol} for dimers of the S66x10 set. This set is an extension of the original S66x8 set, including smaller intermolecular distances (with rescaling factors of 0.7 and 0.8 with respect to the equilibrium distance).[71,72] This fit results in a value of $U_{\mathrm{ex\_ct}} = 7.16$. However, some dependence of this parameter on the type of interaction remains, as can be seen from Figure 6 in which a fit was made for each separate dimer complex. On average, slightly lower values for $U_{\mathrm{ex\_ct}}$ are observed for the hydrogen-bonded dimers, shown in red. For example, for the acetic acid dimer with a double hydrogen bond, a value of $U_{\mathrm{ex\_ct}} = 6.37$ is obtained. Among the dispersion-dominated structures shown in blue, the cyclopentane-neopentane dimer results in the largest value of $U_{\mathrm{ex\_ct}} = 9.20$. However, there is a clear indication that the inclusion of our polarizable model has captured an important part of the physical interactions between these dimers better than the overlap model used in the original MEDFF (see Equation 19). The standard deviation between fitted values of $U_{\mathrm{ex\_ct}}$ in Figure 6 is equal to 0.59. This value can be interpreted as the degree variability across different intermolecular interactions, and therefore the degree to which the model is 'missing' some physics. Fitting the original U_{ex} and U_{ind} parameters in the same way, a standard deviation of 0.97 is obtained, indicating that our polarizable model more accurately captures the intermolecular interactions. Furthermore, a correlation is present between the fitted values for $U_{\mathrm{ex\_ct}}$ and the original MEDFF values for U_{ex} which were fitted to the *ab initio* SAPT2+(3) exchange-repulsion energies. This indicates that an important remainder of the variability of our fitted $U_{\mathrm{ex\_ct}}$ parameter originates from a breakdown of the relatively simple overlap model for exchange-repulsion. Clear indications already exist in the literature how such a model could be improved, relating mainly to a description of the atomic anisotropy of the interactions.[73] However, such extensions would introduce new fitting parameters, diminishing on our promise of a transferable force field with as little parameters as possible. While alternatively, deeper insight into *ab initio* derivations of the origin of exchange-repulsion could lead to a better force field description of this interaction, this will be left for further work.

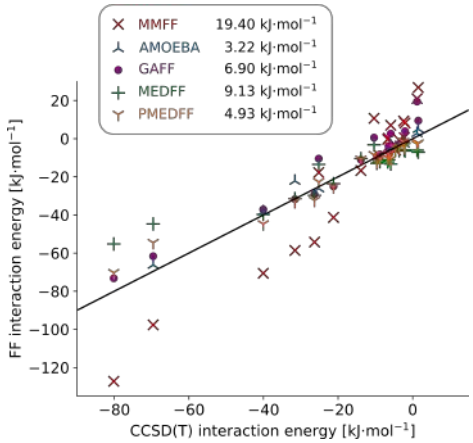


Fig. 7 Parity plot of dimer interaction energies of the hsg set calculated with MMFF (red), AMOEBA (blue), GAFF (purple), MEDFF (green) and PMEDFF (orange). The RMSD for each force field is shown in the legend.

3.5 Benchmarking of dimer interaction energies

To validate our force field, we test its performance on two datasets of dimers which are not present in the S66x8 set to which the interaction parameters were fit. Both datasets and the associated interaction energies computed at the CCSD(T)/CBS level of theory were taken from the BioFragment Database.[74–76] The hsg set contains 21 dimer fragments, extracted from an HIV-II protease crystal structure with a bound ligand (indinavir), representing a wide range of interactions (from dispersion to electrostatically dominated).[76] Therefore, the universal value for U_{ex_ct} of 7.16 a.u. was used for PMEDFF. A comparison was made with the Merck molecular force field[77] (MMFF), AMOEBA[7], the generalized amber force field[78] (GAFF) and MEDFF[30]. As can be seen in Figure 7, PMEDFF performs well on the whole set, with an RMSD of 4.93 kJ·mol^{-1}. Both GAFF and AMOEBA also show a balanced performance across the range of interaction types, although the RMSD of 3.22 kJ·mol^{-1} obtained for AMOEBA only includes 13 out of the 21 complexes for which parameters are available. A significant overestimation of the magnitude of the interaction energy of the electrostatically dominated complexes is observed for MMFF, resulting in an RMSD of 19.40 kJ·mol^{-1}. MEDFF, on the other hand, slightly underestimates the interaction energy of some of the electrostatically dominated complexes, yielding an RMSD of 9.13 kJ·mol^{-1}.

The hbc6 set consists of doubly hydrogen bonded dimers extracted from 6 dissociation curves containing formic acid, formamide and formamidine.[79] Because of the interaction type, the U_{ex_ct} parameter was set to 6.37 a.u., obtained from the fit of the double hydrogen bonded acetic acid dimer (see Figure 6) Due to the double hydrogen bonds, the induction component of the interaction for complexes around the equilibrium intermolecular distances is much larger than for dimers in the hsg set.[74] This set is therefore a more stringent test of the induction component of our force field. We compare with AMOEBA, MMFF and MEDFF. Only 60 out of the 118 dimers containing formic acid and formamide were retained for AMOEBA, as no parameters were available for formamidine. We divide our comparisons with other force fields between dimers with an intermolecular distance larger than and smaller than that of the optimized dimer geometry, shown in Figure 8. For large intermolecular distances, the RMSD of MMFF, AMOEBA, MEDFF and PMEDFF is equal to 8.95 kJ·mol^{-1}, 2.93 kJ·mol^{-1}, 8.24 kJ·mol^{-1} and 5.56 kJ·mol^{-1}, respectively. The dimers with rescaled intermolecular distances smaller than 1 are more challenging to predict, as hydrogen bonds are artificially compressed. The increase in error from the rescaled distances larger than 1 is most pronounced for AMOEBA. The error increases to 44.98 kJ·mol^{-1}, due to the prediction of large repulsive interaction energies. Lower errors of 32.72 kJ·mol^{-1}, 15.40 kJ·mol^{-1} and 15.94 kJ·mol^{-1} are seen for MEDFF, MMFF and PMEDFF. Overall, the performance of PMEDFF against other force fields is encouraging, especially given that, as a test of transferability of our force field, no fitting was performed to any of the dimers present in both the hsg and hbc6 set.

4 Conclusions and outlook

A new transferable polarization model was developed, based on the often-used inducible dipole model. It is transferable in the sense that only the ground state electron density of the molecule or periodic structure in question is required, together with a single proportionality factor which was fitted to the S66x8 set. Slater dipoles were introduced as the dipole response functions, and exchange-repulsion was included as an additional interaction inducing the Slater dipoles. The resulting model performs well on the prediction of molecular polarizabilities of molecules in the QM7 dataset and three-body energies of trimers containing a wide range of intermolecular interactions. Importantly, this is also the case for molecules not present in the S66x8 set and molecules for which no specifically fitted parameters are available for induction models in the literature. The performance is especially noteworthy for the prediction of higher order many-body contributions in a set of water hexamers. A complete polarizable non-covalent force field, coined the polarizable monomer electron density force field (PMEDFF), was developed by including our polarization model in the previously proposed monomer electron density force field (MEDFF). A benchmark on two dimer datasets revealed a performance comparable or better than force fields in the literature, without the need for molecule specific parameters.

To further improve upon the model developed in this work, a more high quality two-body potential for the exchange term would be required which includes anisotropic atomic parameters. The current exchange term retains isotopic parameters for reasons of transferability, although certainly anisotropic effects are important, for example in the description of water.[73] The inclusion of these anisotropic effects in a transferable way is out of the scope of this work, but will be investigated in future work. Moreover, our model retains a pair-wise additive dispersion model, neglecting any many-body dispersion interactions. These effects

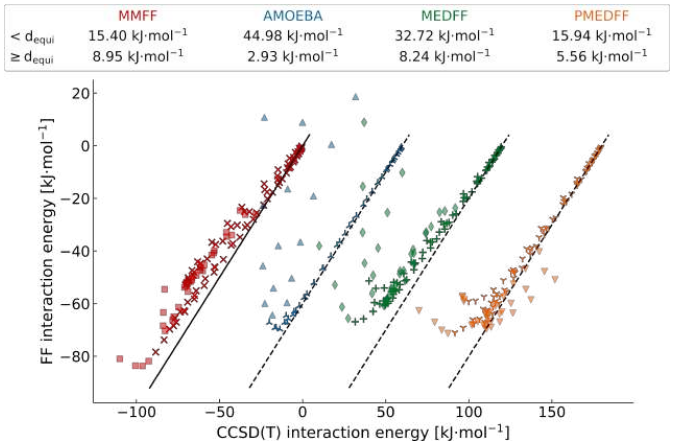


Fig. 8 Parity plot of dimer interaction energies of the hbc set calculated with MMFF (red), AMOEBA (blue), MEDFF (green) and PMEDFF (orange). Dimers along the dissociation curve with rescaled intermolecular distances smaller than the equilibrium distance d_{equi} are displayed with filled symbols in a lighter shade. Results for all force fields except MMFF are shifted horizontally for clarity. The RMSD for each force field at distances $< d_{equi}$ and $\geq d_{equi}$ are shown in the legend.

certainly become important in the condensed phase, and future extensions of our force field will focus on including an appropriate model to include these interactions.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This research was funded by the Research Board of Ghent University (BOF). The computational resources and services used were provided by Ghent University (Stevin Supercomputer Infrastructure).

Notes and references

1 J. C. Wu, J. P. Piquemal, R. Chaudret, P. Reinhardt and P. Ren, *Journal of Chemical Theory and Computation*, 2010, **6**, 2059–2070.

2 D. Jiao, C. King, A. Grossfield, T. A. Darden and P. Ren, *Journal of Physical Chemistry B*, 2006, **110**, 18553–18559.

3 E. C. Meng, J. W. Caldwell and P. A. Kollman, *Journal of Physical Chemistry*, 1996, **100**, 2367–2371.

4 J. W. Caldwell and P. A. Kollman, *Journal of the American Chemical Society*, 1995, **117**, 4177–4178.

5 J. Soetens, G. Jansen and C. Millot, *Molecular Physics*, 1999, **96**, 1003–1012.

6 J. Wang, P. Cieplak, J. Li, J. Wang, Q. Cai, M. Hsieh, H. Lei, R. Luo and Y. Duan, *Journal of Physical Chemistry B*, 2011, **115**, 3100–3111.

7 P. Ren, C. Wu and J. W. Ponder, *Journal of Chemical Theory and Computation*, 2011, **7**, 3143–3161.

8 M. Masella and P. Cuniasse, *Journal of Chemical Physics*, 2003, **119**, 1866–1873.

9 G. A. Kaminski, H. A. Stern, B. J. Berne, R. A. Friesner, Y. X. Cao, R. B. Murphy, R. Zhou and T. A. Halgren, *Journal of Computational Chemistry*, 2002, **23**, 1515–1531.

10 P. Cieplak, F. Y. Dupradeau, Y. Duan and J. Wang, *Polarization effects in molecular mechanical force fields*, 2009.

11 M. L. Laury, L. P. Wang, V. S. Pande, T. Head-Gordon and J. W. Ponder, *Journal of Physical Chemistry B*, 2015, **119**, 9423–9437.

12 P. Ren and J. W. Ponder, 2003.

13 B. Chen, J. Xing and J. I. Siepmann, *Journal of Physical Chemistry B*, 2000, **104**, 2391–2401.

14 L. X. Dang, *Journal of Physical Chemistry B*, 1998, **102**, 620–624.

15 J. Wang, P. Cieplak, Q. Cai, M.-J. Hsieh, J. Wang, Y. Duan and R. Luo, *The Journal of Physical Chemistry B*, 2012, **116**, 7999–8008.

16 G. Lamoureux, A. D. MacKerell and B. Roux, *Journal of Chemical Physics*, 2003, **119**, 5185–5197.

17 H. Yu, T. Hansson and W. F. Van Gunsteren, *Journal of Chemical Physics*, 2003, **118**, 221–234.

18 G. V. Lewis and C. R. A Catlow, *Journal of Physics C: Solid State Physics*, 1985, **18**, 1149–1161.

19 A. Albaugh and T. Head-Gordon, *Journal of Chemical Theory and Computation*, 2017, **13**, 5207–5216.

20 M. M. Ghahremanpour, P. J. Van Maaren, C. Caleman, G. R. Hutchison and D. Van Der Spoel, *Journal of Chemical Theory and Computation*, 2018, **14**, 5553–5566.

21 S. W. Rick, S. J. Stuart and B. J. Berne, *The Journal of Chemical Physics*, 1994, **101**, 6141–6156.

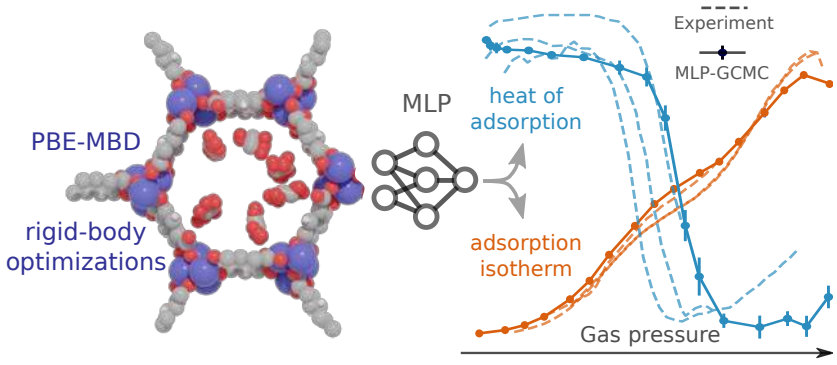
22 J. Caldwell, P. A. Kollman and L. X. Dang, *Journal of the Amer-*

ican Chemical Society, 1990, **112**, 9144–9147.
23 B. Thole, *Chemical Physics*, 1981, **59**, 341–350.
24 D. Elking, T. Darden and R. J. Woods, *Journal of computational chemistry*, 2007, **28**, 1261–74.
25 A. G. Donchev, M. G. Galkin, A. A. Illarionov, O. V. Khoruzhii, M. A. Olevanov, V. D. Ozrin, M. V. Subbotin and V. I. Tarasov, *Proceedings of the National Academy of Sciences of the United States of America*, 2006, **103**, 8613–8617.
26 M. Masia, M. Probst and R. Rey, *The Journal of Chemical Physics*, 2005, **123**, 164505.
27 Y. S. Kim, S. K. Kim and W. D. Lee, *Chemical Physics Letters*, 1981, **80**, 574–575.
28 A. J. Misquitta and A. J. Stone, *Journal of Chemical Theory and Computation*, 2016, **12**, 4184–4208.
29 A. Stone, *The Theory of Intermolecular Forces*, Oxford University Press, 2013.
30 S. Vandenbrande, M. Waroquier, V. V. Speybroeck and T. Verstraelen, *Journal of Chemical Theory and Computation*, 2017, **13**, 161–179.
31 B. Jeziorski, R. Moszynski and K. Szalewicz, *Chemical Reviews*, 1994, **94**, 1887–1930.
32 J. Wang, P. Cieplak, J. Li, T. Hou, R. Luo and Y. Duan, *The Journal of Physical Chemistry B*, 2011, **115**, 3091–3099.
33 X. Chu and A. Dalgarno, *Journal of Chemical Physics*, 2004, **121**, 4083–4088.
34 A. D. Becke and E. R. Johnson, *Journal of Chemical Physics*, 2006, **124**, year.
35 A. Tkatchenko and M. Scheffler, *Physical Review Letters*, 2009, **102**, 073005.
36 T. Verstraelen, S. Vandenbrande, F. Heidar-Zadeh, L. Vanduyfhuys, V. Van Speybroeck, M. Waroquier and P. W. Ayers, *Journal of Chemical Theory and Computation*, 2016, **12**, 3894–3912.
37 R. Ahlrichs, *Chemical Physics Letters*, 1972, **15**, 609–612.
38 A. J. O'Connor, *Comm. Math. Phys.*, 1973, **32**, 319–340.
39 M. Hoffmann-Ostenhof and T. Hoffmann-Ostenhof, *Physical Review A*, 1977, **16**, 1782–1785.
40 P. Cieplak, J. Caldwell and P. Kollman, *Journal of Computational Chemistry*, 2001, **22**, 1048–1057.
41 T. D. Rasmussen, P. Ren, J. W. Ponder and F. Jensen, *International Journal of Quantum Chemistry*, 2007, **107**, 1390–1395.
42 W. Xie, J. Pu, A. D. MacKerell and J. Gao, *Journal of Chemical Theory and Computation*, 2007, **3**, 1878–1889.
43 D. M. York and W. Yang, *Journal of Chemical Physics*, 1996, **104**, 159–172.
44 G. Tabacchi, J. Hutter and C. J. Mundy, *Journal of Chemical Physics*, 2005, **123**, 074108.
45 A. Wadehra and S. K. Ghosh, *Journal of Chemical Sciences*, 2005, **117**, 401–409.
46 Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder and P. Ren, *Journal of Chemical Theory and Computation*, 2013, **9**, 4046–4063.
47 P. Echenique and J. L. Alonso, *Molecular Physics*, 2007, **105**, 3057–3098.
48 S. P. Jarvis, A. M. Sweetman, L. Kantorovich, E. Mcglynn and P. Moriarty, *Pauli's Principle in Probe Microscopy*.
49 J. Rezáč, K. E. Riley and P. Hobza, *Journal of Chemical Theory and Computation*, 2011, **7**, 2427–2438.
50 G. D. Purvis and R. J. Bartlett, *The Journal of Chemical Physics*, 1982, **76**, 1910–1918.
51 K. A. Peterson, D. E. Woon and T. H. Dunning, *The Journal of Chemical Physics*, 1994, **100**, 7410–7415.
52 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian09 Revision D.01*, 2013.
53 L. C. Blum and J. L. Reymond, *Journal of the American Chemical Society*, 2009, **131**, 8732–8733.
54 M. Rupp, A. Tkatchenko, K. R. Müller and O. A. Von Lilienfeld, *Physical Review Letters*, 2012, **108**, 058301.
55 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, M. B., G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, A. P. Raghavachari, K. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. Daniels, O. D.; Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian16 (Revision A.03)*, 2016.
56 A. Tkatchenko, R. A. DiStasio, R. Car and M. Scheffler, *Physical Review Letters*, 2012, **108**, 236402.
57 A. Milet, R. Moszynski, P. E. Wormer and A. Van der Avoird, *Journal of Physical Chemistry A*, 1999, **103**, 6811–6819.
58 R. Podeszwa and K. Szalewicz, *Journal of Chemical Physics*, 2007, **126**, year.
59 V. Babin, G. R. Medders and F. Paesani, *Journal of Chemical Theory and Computation*, 2014, **10**, 1599–1607.

60 D. M. Bates and G. S. Tschumper, *Journal of Physical Chemistry A*, 2009, **113**, 3555–3559.
61 G. R. Medders, A. W. Götz, M. A. Morales, P. Bajaj and F. Paesani, *Journal of Chemical Physics*, 2015, **143**, 104102.
62 C. J. Burnham, D. J. Anick, P. K. Mankoo and G. F. Reiter, *Journal of Chemical Physics*, 2008, **128**, 154519.
63 Y. Wang and J. M. Bowman, *Journal of Chemical Physics*, 2011, **134**, 154510.
64 V. Babin, C. Leforestier and F. Paesani, *Journal of Chemical Theory and Computation*, 2013, **9**, 5395–5403.
65 G. R. Medders, V. Babin and F. Paesani, *Journal of Chemical Theory and Computation*, 2014, **10**, 2906–2910.
66 L. P. Wang, T. Head-Gordon, J. W. Ponder, P. Ren, J. D. Chodera, P. K. Eastman, T. J. Martinez and V. S. Pande, *Journal of Physical Chemistry B*, 2013, **117**, 9956–9972.
67 G. S. Fanourgakis and S. S. Xantheas, *Journal of Chemical Physics*, 2008, **128**, 074506.
68 J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardère, M. J. Schnieders, J. P. Piquemal, P. Ren and J. W. Ponder, *Journal of Chemical Theory and Computation*, 2018, **14**, 5273–5289.
69 J. Rezáč, Y. Huang, P. Hobza and G. J. O. Beran, *Journal of Chemical Theory and Computation*, 2015, **11**, 3065–3079.
70 G. Chalasiński, M. M. SzczÈĽśniak and R. A. Kendall, *The Journal of Chemical Physics*, 1994, **101**, 8860–8869.
71 B. Brauer, M. K. Kesharwani, S. Kozuch and J. M. Martin, *Physical Chemistry Chemical Physics*, 2016, **18**, 20905–20925.
72 D. G. Smith, L. A. Burns, K. Patkowski and C. D. Sherrill, *Journal of Physical Chemistry Letters*, 2016, **7**, 2197–2203.
73 M. J. Van Vleet, A. J. Misquitta and J. R. Schmidt, *Journal of Chemical Theory and Computation*, 2018, **14**, 739–758.
74 L. A. Burns, J. C. Faver, Z. Zheng, M. S. Marshall, D. G. A. Smith, K. Vanommeslaeghe, A. D. MacKerell, K. M. Merz and C. D. Sherrill, *The Journal of Chemical Physics*, 2017, **147**, 161727.
75 L. A. Burns, M. S. Marshall and C. D. Sherrill, *Journal of Chemical Physics*, 2014, **141**, 234111.
76 J. C. Faver, M. L. Benson, X. He, B. P. Roberts, B. Wang, M. S. Marshall, M. R. Kennedy, C. D. Sherrill and K. M. Merz, *Journal of Chemical Theory and Computation*, 2011, **7**, 790–797.
77 T. A. Halgren, *Journal of Computational Chemistry*, 1999, **20**, 720–729.
78 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *Journal of Computational Chemistry*, 2004, **25**, 1157–1174.
79 K. S. Thanthiriwatte, E. G. Hohenstein, L. A. Burns and C. D. Sherrill, *Journal of Chemical Theory and Computation*, 2011, **7**, 88–96.

Paper IV

DFT-quality Adsorption Simulations in Metal-Organic Frameworks Enabled by Machine Learning Potentials



R. Goeminne, L. Vanduyfhuys, V. Van Speybroeck, T. Verstraelen

Submitted to J. Chem. Theory Comput

R. Goeminne performed the simulations, interpreted the resulting data and wrote the manuscript.

DFT-quality adsorption simulations in metal-organic frameworks enabled by machine learning potentials

Ruben Goeminne, Louis Vanduyfhuys, Veronique Van Speybroeck, and Toon Verstraelen*

Center for Molecular Modeling (CMM), Ghent Univeristy, Technologiepark 46, 9052 Zwijnaarde, Belgium

E-mail: Toon.Verstraelen@UGent.be

Abstract

Nanoporous materials such as metal-organic frameworks (MOFs) have been extensively studied for their potential for adsorption and separation applications. In this respect, grand canonical Monte Carlo (GCMC) simulations have become a well-established tool for computational screenings of the adsorption properties of large sets of MOFs. However, their reliance on empirical force field potentials have limited the accuracy with which this tool can be applied to MOFs with challenging chemical environments such as open-metal sites. On the other hand, density-functional theory (DFT) is too computationally demanding to be routinely employed in GCMC simulations due to the excessive number of required function evaluations. Therefore, we propose in this paper a protocol for training machine leaning potentials (MLPs) on a limited set of DFT intermolecular interaction energies (and forces) of CO_2 in ZIF-8 and the open-metal site containing Mg-MOF-74, and use the MLPs to derive adsorption isotherms from first principles. We make use of equivariant NequIP model which

has demonstrated excellent data efficiency, and as such an error on the interaction energies below 0.2 kJ·mol^{-1} per adsorbate in ZIF-8 was attained. Its use in GCMC simulations results in highly accurate adsorption isotherms and heats of adsorption. For Mg-MOF-74, a large dependence of the obtained results on the used dispersion correction was observed, where PBE-MBD performs the best. Lastly, to test the transferability of the MLP trained on ZIF-8, it was applied to ZIF-3, ZIF-4 and ZIF-6, which resulted in large deviations on the predicted adsorption isotherms and heats of adsorption. Only when explicitly training on data for all ZIFs, accurate adsorption properties were obtained. As the proposed methodology is widely applicable to guest adsorption in nanoporous materials, it opens up the possibility for training general-purpose MLPs to perform highly accurate investigations of guest adsorption.

Introduction

Porous materials such as metal-organic frameworks (MOFs) have drawn widespread interest in the last few decades for their potential for gas adsorption and separation applications.[1–3] Consisting of metal-oxide building blocks and organic ligands, their permanent porosity under guest removal was first established in the late 1990s.[4–8] Through their high degree of versatility, rational design of these materials for targeted adsorption properties became possible through the concept of isoreticular design. In seminal work by the group of Yaghi, the pore size of MOF-5 could be incrementally varied from 3.8 to 28.8 Å by substitution of the constituting organic linker, giving rise to the IRMOF isoreticular series of MOFs. Of these, IRMOF-6 demonstrated state-of-the-art methane storage capacity at the time.[9] Furthermore, linkers can be functionalized to tune the host-guest interactions.[10–12] Clearly, the chemical space of MOFs arising from combining linkers, metal-oxide bricks, topologies and linker functionalizations is vast, precluding an exhaustive search of this space through experimental characterization. Herein lies the prime value of computational screening studies.[13–23] Through grand-canonical Monte Carlo simulations, the adsorption properties of large sets of MOF structures can be quickly extracted. Early studies following this approach focused on the screening of MOFs for their uptake and heat of adsorption of hydrogen, carbon dioxide and methane,[13–15] with later studies investigating separation properties.[16–19] Prime examples of the success of such computational screenings include the identification of NU-1103 out of a set of 13 512 MOFs with outstanding hydrogen storage capacity at cryogenic temperatures and the identification of 300 out of a set of 137 953 MOFs with exceptional methane storage capacity.[20,21] Lin *et al.* leveraged GPU hardware to significantly speed up adsorption simulations, enabling them to screen hundreds of thousands of zeolites and ZIFs (zeolitic imidazolate framework; a class of MOFs with isomorphic topology to zeolites) for carbon capture applications.[22,23] Furthermore, recent developments in machine learning (ML) models have been leveraged to predict the adsorption characteristics in MOFs.[24–30] By training such models on reference calculations or experimental uptakes, the adsorption properties of

novel MOFs can be predicted with high accuracy (compared to the reference calculations or experiments).

While all of the aforementioned screening studies have clearly demonstrated their merit in identifying promising nanoporous for applications, they critically rely on an accurate description of the underlying potential energy surface (PES). Preferentially, DFT or wavefunction-based methods would be used to provide a highly accurate description of the PES. However, such methods are excessively computationally demanding for routine usage in GCMC simulations, due to the required number of evaluations needed to converge a typical simulation. Therefore, computationally cheap force fields are usually employed. To describe the host-guest and guest-guest interactions, non-covalent force field potentials are employed. Usually, a Lennard-Jones (LJ) or Buckingham potential is used, supplemented with an electrostatic interaction arising from atomic charges assigned to the framework and adsorbates. Examples of force fields used for this task are the Universal Force Field (UFF), DREIDING, MM3 and the general Amber force field (GAFF).[31–34] In later years, these force fields were adapted for specific use with MOFs, including parameterizations of transition metals commonly encountered in these materials.[35,36] However, many examples exist in the literature where force fields required re-parameterization due to poor agreement with experiments. Boulanger *et al.* rescaled the GAFF van der Waals dispersion parameters by a factor of 1.115 to obtain accurate hydration free energies, Wu *et al.* rescaled the UFF LJ parameter ϵ by a factor of 0.635 to reproduce methane adsorption isotherms in ZIF-8 and similarly, Pérez-Pellitero *et al.* found that a rescaling of the LJ parameters was required for agreement with experimental adsorption isotherms of N_2 and CO_2 in ZIFs.[37–39] Moreover, force fields such as UFF and DREIDING have been shown to significantly underestimate the interaction strength of adsorbates with open-metal sites present in many MOFs.[40] This deficiency has spurred the development of more advanced polarizable force fields in which polarization of the adsorbate on open-metal sites is modeled with the induced dipole method.[41–46] However, even when employing these more advanced force fields, significant deviations between the experimental

and predicted CO_2 adsorption isotherms and heats of adsorption remain in the open-metal site containing M-MOF-74 (M=Co,Fe,Mg,Mn,Ni,Zn) series of MOFs.[46] Such limits in the attainable accuracy could be related to the functional form of the employed force fields, which might not fully capture the interaction of adsorbates with the open-metal site.

The preceding discussion demonstrates the need for methods that can bridge the gap between a force field and DFT description of the intermolecular interactions between adsorbates and a MOF adsorbent. On one hand, DFT calculations are prohibitively expensive for use in GCMC simulations due to the required number of iterations needed for convergence.[47] On the other hand, force field based GCMC simulations are computationally (relatively) cheap, but are limited in accuracy when applied to challenging intermolecular interactions such as those present in open-metal site containing MOFs. To bridge this gap, different approaches have been investigated. Lee *et al.* evaluated the Henry regime (characterizing the low-pressure adsorption behavior) in M-MOF-74(M=Zn,Mg) for several adsorbents by performing biased single particle insertions, evaluated at the DFT level of theory.[48] In a similar work, Vandenbrande *et al.* proposed an importance sampling scheme based on preceding force field calculations to evaluate the Henry coefficients and heats of adsorption at the DFT level of theory at infinite dilution for methane in UiO-66 and CO_2 in Mg-MOF-74 by means of single particle insertions.[49] However, the efficiency of this method (the number of required DFT evaluations) still depends critically on the quality of the used force field. Moreover, only the low-pressure regime can be characterized from this method. Alternatively, Kundu *et al.* predicted the adsorption isotherms of CO_2 in Mg-MOF-74 from calculations of the interaction energy of an adsorbate on the open-metal site and linker with a correlated wavefunction-corrected DFT method.[50–52] With these adsorption sites, a lattice of sites was defined, on which grand canonical Monte Carlo simulations were performed. Although successful in predicting the adsorption isotherms at different temperatures, this method relies on an initial accurate determination of well-defined adsorption sites, as guest molecules are not considered to adsorb on other sites of the framework, limiting the general

applicability of this method.

Hence, this work aims at a generally applicable methodology to perform GCMC calculations in MOFs at DFT accuracy. For this, we leverage the recent developments of highly accurate and data-efficient equivariant machine learning potentials (MLPs) to describe the interaction of guest molecules in MOFs. In the message passing NequIP potential used in this work, each atom is characterized by a feature vector which is refined during training through a series of interactions layers.[53] Tested on datasets of energies and forces of small molecules and periodic materials, it demonstrated state-of-the-art accuracy as well as exceptional data efficiency. In recent years, MLPs have been employed to study the mechanical, diffusion and phase transition properties of MOFs by means of MD simulations.[54–57] In these applications, the training set consists of DFT energies and atomic forces. However, for our application of employing an MLP in GCMC simulations, the requirements on the training set are fundamentally different. Through the integration of the equations of motion, an equilibrium canonical MD simulation samples regions of the PES proportional to a Boltzmann factor of the potential energy. While the exact same distribution is sampled from (canonical) Monte Carlo simulations, this is achieved by performing trial moves of the system to regions of the PES which might be very high in energy. Even though these trial moves will not be accepted and are therefore not relevant, they still need to be reliably rejected. This implies that, contrary to the case of MD simulations, the MLP needs to be explicitly trained on structures high on the PES. Secondly, only the intermolecular or non-covalent energy is of importance in GCMC simulations, as both the framework and adsorbates are generally considered rigid.

To address these considerations, a methodology is proposed to efficiently generate datasets of frameworks loaded with guests which will serve as the input data to train an MLP. Subsequently, tests are performed to determine both the hyperparameters of the network, as well as the required data set size to derive an accurate model for CO_2 in ZIF-8. For adsorption in this MOF, high quality experimental data is available, allowing for a unambiguous

comparison with the results obtained with our proposed methodology.[58] With the trained MLP, GCMC calculations are performed to extract the isotherms and heats of adsorption, demonstrating excellent agreement with experiments. The proposed methodology is also applied to model CO_2 adsorption in the open-metal site containing Mg-MOF-74 with reference data generated at the PBE-D3(BJ), PBE-TS and PBE-MBD level of theory, highlighting the important differences between the used dispersion corrections.[59–63] Lastly, the ability of an MLP trained on ZIF-8 to model adsorption in ZIF-3, ZIF-4 and ZIF-6 was investigated, as this type of transferability of interactions to similar chemical environments is usually implicitly assumed in classical force fields. The MLP is shown to be unable to capture the stronger interactions present in ZIF-3 and ZIF-4. Even in ZIF-6, where the error on the test set is acceptable, significant errors in the predicted heats of adsorption are obtained, highlighting the risk of extrapolation for MLPs. Only when explicitly training on all ZIFs, accurate results are obtained.

Methodology

Non-covalent dataset generation

As mentioned before, the requirements on a dataset to train an MLP employed in GCMC simulations are different than the ones needed for performing MD simulations, due to the need to reliably reject trial moves that are high in potential energy. On the other hand, MLPs can generally be trained to higher accuracy when the training set is not excessively unbalanced, as would be the case when combining repulsive interactions order of magnitudes stronger than the attractive interactions. Therefore, an inclusion in the dataset of extremely repulsive interactions as encountered when atoms approach each other closely, is not desirable. As a compromise between these considerations, the method illustrated in Figure 1 for generating a dataset of non-covalent interactions is proposed.

First, the van der Waals radii derived by Alvarez are assigned to each element present in

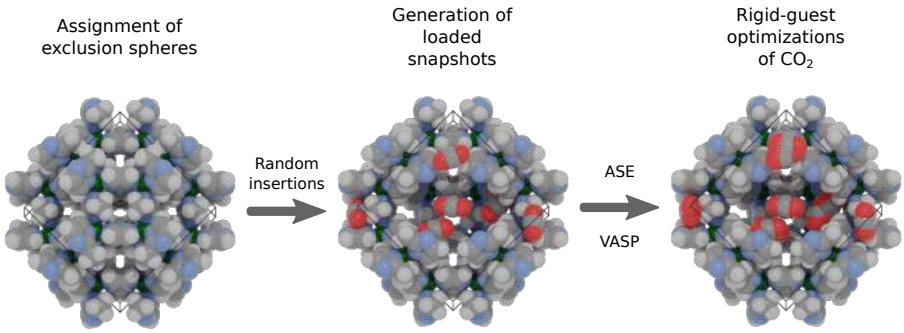


Figure 1: Proposed methodology for generating the dataset to train an MLP on intermolecular energies (and forces). First, snapshots of the framework loaded with adsorbates respecting minimum intermolecular distances defined by the exclusion spheres are generated. Subsequently, rigid-guest optimizations of the adsorbates are performed (with a fixed framework) to guide the adsorbates to favorable interaction sites in the framework.

the framework and adsorbates.[64] Their analysis of more than 5 million non-covalent distances shows that no non-bonded interactions occur at inter-atomic distances below the sum of two atomic radii minus 0.7 Å. Therefore, the van der Waals radius for each atom minus 0.35 Å can be seen as an 'exclusion sphere' within which no adsorbate-adsorbate or adsorbate-framework interactions occur. The exclusion spheres will be used further on (with proper validation) in GCMC simulations to reject any trial step in which an overlap occurs. Because of this, there is no need to explicitly train the MLP to these very short-ranged (repulsive) interactions.

Subsequently, a set of N guest-loaded snapshots constrained by the exclusion spheres are generated by means of random insertions, implemented in a custom python script. The number of adsorbates in each snapshot N_{ads} is distributed uniformly between 0 and N_{max}, with N_{max} the maximum number of guests which can reliably be inserted without violating the constraints. Although the snapshots generated from this procedure respect the minimum intermolecular distances defined by the exclusion spheres, they still mostly represent high energy structures. Therefore, in a last step, each snapshot is optimized using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimizer from the Atomic Simulation Environ-

ment (ASE) at the relevant level of theory for N_{steps} iterations, keeping both the framework coordinates fixed and the adsorbates rigid ($N_{\text{steps}} = 30$ in this work, as will be shown further).[65] While extensions to flexible frameworks and adsorbates are possible, these will not be considered in this work. Every snapshot from the optimizations is retained in the dataset, for a total of $N \times N_{\text{steps}}$ snapshots. By retaining all snapshots, both the high energy (at the beginning of the optimization), as well as the low energy structures (at the end of the optimization) are sampled. By subtracting the energies and forces of the constituting isolated framework and adsorbates from the total energies and forces, the intermolecular energies E_{inter} and forces $\mathbf{F}_{\text{inter}}$ are obtained. As the framework remains fixed, the energy and forces of the framework are the same for each snapshot, requiring only 1 single point calculation. In this work, the experimentally resolved structures for all MOFs are used. Similarly, because of the use of rigid adsorbates, only a single calculations of the energy and forces of an isolated adsorbate are required. Note that this calculation of intermolecular energies and forces is not subject to a basis set superposition error, as plane waves were used as basis set in this work. Further details on the DFT settings are given in the Computational Details.

MLP training

The Neural Equivariant Interatomic Potential (NequIP) is used in this work as MLP to represent the intermolecular interaction energy and forces of guest-loaded frameworks, chosen for its excellent accuracy and data efficiency.[53] However, future more advanced models could also be employed as simple drop-in replacements. To train the network, an appropriate cost function is required. In this work, a slight adaptation of the existing cost functions in NequIP is made. As N_{ads} varies across snapshots in the training set and the intermolecular energy is an extensive property, the energy cost is normalized with N_{ads} (see Equation 1). Neglecting to do so would artificially bias training towards snapshots with higher N_{ads}. Additionally, the intermolecular forces are also included in the cost function. While not strictly needed, including forces during training is known to improve data efficiency. An L1 cost function is

used for increased robustness with respect to outliers:

$$\mathcal{C}(\boldsymbol{\theta}) = \frac{\lambda_E}{N}\sum_{i=1}^{N}\left\|\frac{E^{(i)}(\boldsymbol{\theta}) - \hat{E}^{(i)}_{\text{inter}}}{N_{\text{ads,i}}}\right\| + \frac{\lambda_F}{3NN_a}\sum_{i=1}^{N}\sum_{a=1}^{N_a^{(i)}}\left\|\boldsymbol{F}_a^{(i)}(\boldsymbol{\theta}) - \hat{\boldsymbol{F}}^{(i)}_{\text{inter,a}}\right\| \tag{1}$$

with the sum over i running over the snapshots, the sum over a running over atoms, $N_a^{(i)}$ the number of atoms in each snapshot i and $N_a = \sum_i N_a^{(i)}$. In this cost function, the parameters of the network $\boldsymbol{\theta}$ are to be optimized to minimize the difference between the predicted energies and forces ($E^{(i)}(\boldsymbol{\theta})$ and $\boldsymbol{F}_a^{(i)}(\boldsymbol{\theta})$) and the reference energies and forces ($\hat{E}^{(i)}_{\text{inter}}$ and $\hat{\boldsymbol{F}}^{(i)}_{\text{inter,a}}$). The hyperparameters λ_E and λ_F determine the relative weight of the energies and forces. The effect of different hyperparameters of the network such as the cutoff radius, the number of interaction layers, the number of features and the maximum rotation order are investigated in the Supporting Information.

GCMC calculations

With the trained MLPs, GCMC calculations are performed using the same exclusion spheres as for the dataset generation to reject unphysical trial moves. Care is taken that the exclusion spheres were not too large such that energetically favorable moves would be rejected (see further). In other words, the exclusion spheres should strictly encompass regions of the pores associated with unfavorable insertion energies. The Peng-Robinson equation of state is used to convert the gas pressure to a fugacity used in the simulations. Apart from the gas uptake, the isosteric heat of adsorption Q_{st} as a function of the gas pressure can be derived from GCMC calculations from the following Equation:[66]

$$Q_{\text{st}} = -\frac{\langle E \cdot N\rangle - \langle E\rangle\langle N\rangle}{\langle N^2\rangle - \langle N\rangle^2} + RT \tag{2}$$

with N and E the instantaneous number of adsorbates and intermolecular energy, the brackets $\langle\cdots\rangle$ denoting an ensemble average and R and T are the molar gas constant and tem-

perature. In this Equation, the first term represents the difference in interaction energy between a MOF loaded with $N + 1$ molecules, compared to N molecules, and the second term represents the work required to compress an ideal gas at a fixed pressure and temperature. Alternatively, the isosteric heat of adsorption at infinite dilution $Q_{st,\text{dilution}}$ can also be computed using the Widom insertion method, performing a set of random insertions of the adsorbate in the framework and computing the interaction energy ΔE:[67]

$$Q_{st,\text{dilution}} = -\frac{\langle \Delta E\, e^{-\beta \Delta E} \rangle}{\langle e^{-\beta \Delta E} \rangle} + RT \tag{3}$$

Results

Carbon dioxide adsorption in ZIF-8

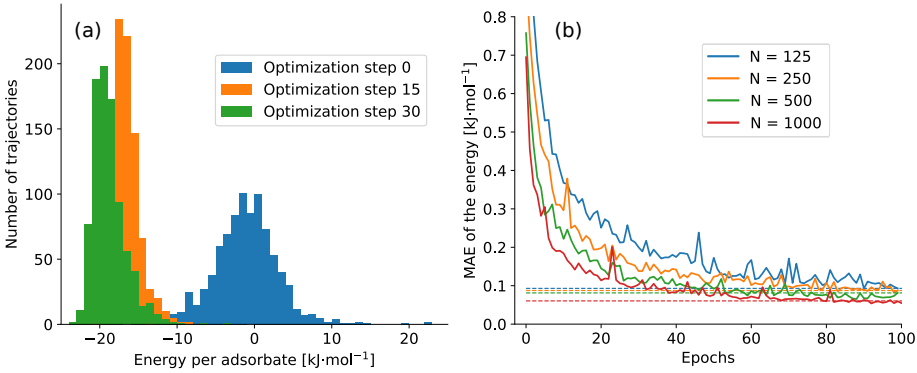


Figure 2: (a) Histogram of the interaction energy per adsorbate along three steps in optimization trajectories (step 0, 15 and 30) out of a total of 1000 optimizations, performed at the PBE-D3(BJ) level of theory. (b) The mean absolute error (MAE) on the energy of the validation set during training for different sizes of the dataset (between 125 and 1000 optimizations).

With the procedure outlined before, 1000 snapshots containing between 0 and 32 CO_2 adsorbates were generated after which each was optimized at the PBE-D3(BJ) level of theory for a total of 30 optimization steps. To determine whether the geometric optimizations have

steered the adsorbates to the favorable interaction sites, the histograms of the interaction energy per adsorbate are shown at the start of the optimization trajectories (step 0), in the middle (step 15) and at the end (step 30) in Figure 2(a). As can be seen, the initial snapshots represent mostly configurations between -10 and 10 kJ·mol^{-1} per adsorbate. Optimizing the snapshots, most have reached interaction sites between -20 and -15 kJ·mol^{-1} per adsorbate at the end. Note that it is not vitally important for all optimization trajectories to fully converge. As long as sufficient favorable interactions are included in the dataset, the MLP is able to accurately represent the minima (see further). To determine the required number of trajectories to train an accurate MLP, the total dataset of 1000 optimization trajectories was subdivided into sets containing N = 125, 250, 500 and 1000 trajectories. For each set, a separate MLP was trained (denoted further as MLP$_{125}$, MLP$_{250}$, MLP$_{500}$ and MLP$_{1000}$), in which 80% of the data is used for training and 20% for validation. The validation mean absolute error (MAE) is shown in Figure 2(b). As can be seen, the MAE on the interaction energies is low even for the dataset containing 125 trajectories. Increasing N from 125 to 1000 decreases the MAE from 0.093 to 0.060 kJ·mol^{-1} per adsorbate. However, evaluating each model on a test set containing 100 optimizations to which no training was performed, elevated errors of 0.45 to 0.15 kJ·mol^{-1} per adsorbate are obtained for MLP$_{125}$ and MLP$_{1000}$.

As a first validation of the derived MLPs, their accuracy in reproducing interaction energies of a single adsorbate in the framework is investigated. A set of 50 rigid-body optimizations of an adsorbate in ZIF-8 performed with MLP$_{1000}$ revealed that the strongest interaction site of CO_2 is a configuration in which the adsorbate interacts with a pore window, perpendicular to the plane of the window (see Figure 3(b)). To explore the interaction around this minimum, the adsorbate was subsequently translated along the normal of the pore window plane. Relative to the configuration centered in the pore window, the adsorbate is translated with distances between 0 and 4 Å with steps of 0.1 Å. For each structure, the interaction energy is recomputed with all MLPs and the reference PBE-D3(BJ) level of theory. As shown in Figure 3(a), significant deviations are observed for MLP$_{125}$. For the

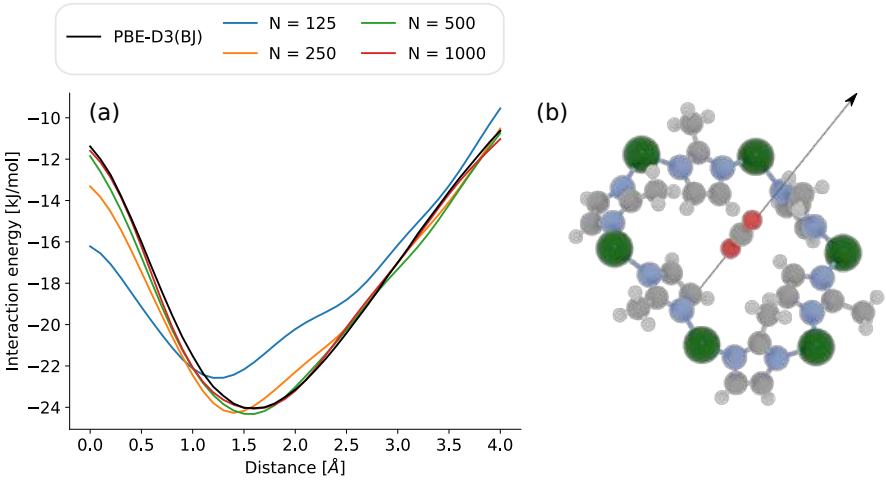


Figure 3: (a) Interaction energies of an adsorbate with the pore window in ZIF-8, perpendicular to the plane of the window, computed with all four MLPs and the reference PBE-D3(BJ) level of theory. The adsorbate is translated along the normal (see (b)) with distances between 0 and 4 Å. (b) Visualization of the adsorbate interacting with the pore window of ZIF-8, and the translation vector along the normal on the pore window. For clarity, only the Zn atoms and imidazolate linkers of the six-membered pore ring are shown.

other MLPs, good agreement with the reference energies are obtained. Furthermore, as the optimizations were performed with an MLP, this demonstrates that the forces learned during training enable accurate optimizations towards the minimum.

While this first validation demonstrates that an MLP trained on a sufficient number of optimizations can accurately reproduce interaction energies around minima on the potential energy surface, an additional validation of the used exclusion spheres is vital to ensure that their exact values do not influence the adsorption properties obtained further. To this end, the heats of adsorption Q_{st} of CO_2 in ZIF-8 were computed at 273 K by means of Widom insertions with rescaled values of the radii of the exclusion spheres between a factor of 0.80 and 1.20. In case the exclusion spheres would erroneously exclude relevant regions of the potential energy surface, Q_{st} would be sensitive to rescaling factors smaller than 1. As shown in Figure 4, Q_{st} is indeed insensitive to the exact choice of the radii for the exclusion spheres

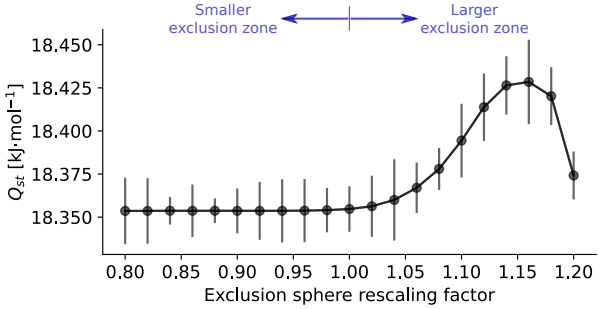


Figure 4: Isosteric heat of adsorption Q_{st} of CO_2 in ZIF-8 obtained from $5 \cdot 10^6$ Widom insertions using MLP_{1000} with rescaled exclusion sphere radii. Errors are obtained from dividing the set of $5{\cdot}10^6$ Widom insertions in 5 subsets, and computing the standard deviation between Q_{st} obtained for each subset.

(with only small deviations occurring from rescaling factors larger than 1), demonstrating that the region excluded by the exclusion spheres is indeed an irrelevant part of the phase space.

With each of the MLPs trained on varying dataset sizes, GCMC calculations were subsequently performed at a temperature of 273 K and a range of gas pressures between 0.1 bar and 50 bar. Each simulation consists of $5 \cdot 10^6$ MC steps, with equal probability for translations, rotations, insertions and deletions. The resulting uptake as a function of the gas pressure and heat of adsorption as a function of the guest loading are shown in Figure 5(a) and 5(b). The adsorption isotherms obtained with the MLPs trained on between 125 and 1000 optimizations are nearly identical. Comparing these isotherms with experimental results obtained by Simmons *et al.*, Abraha *et al.* and Gracés *et al.*, good agreement between both is observed, as seen in Figure 5(a).[58,68,69] While a slight over-prediction of the uptake at low gas pressure and an under-prediction of the uptake at high gas pressures is observed, this could be due to the rigid-framework approximation employed in GCMC simulations, as the flexible behavior of ZIF-8 and its effects on adsorption have been well-established.[70] As shown in Figure 5(b), MLP_{1000} predicts an increase in the heat of adsorption of CO_2 in ZIF-8 from 18.0 kJ·mol^{-1} at the smallest loadings to approximately 25.8 kJ·mol^{-1} at the highest

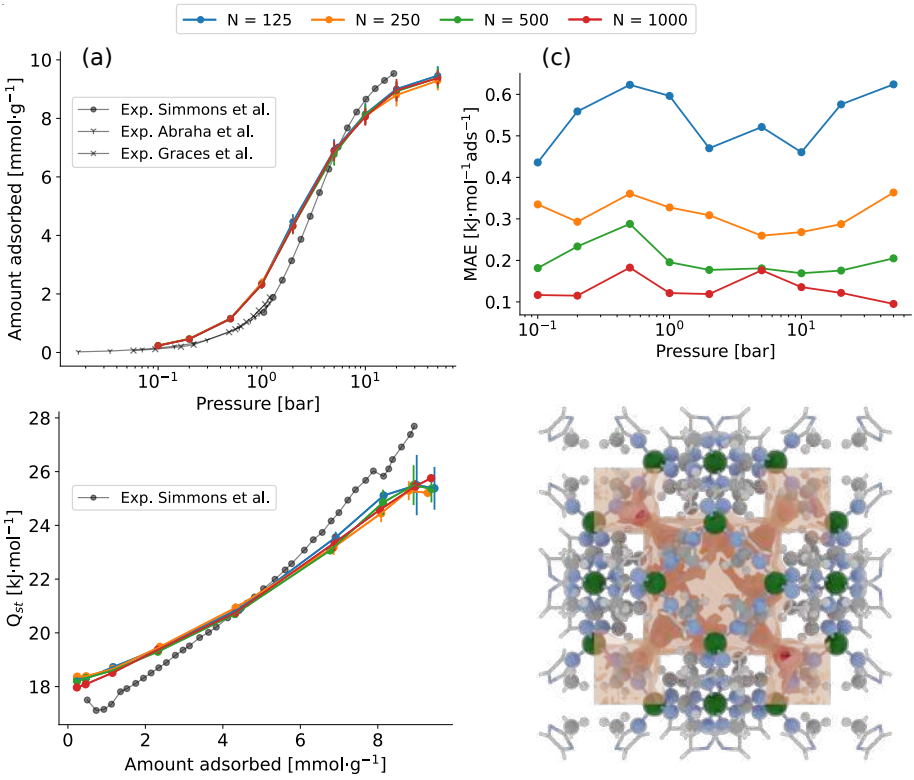


Figure 5: (a) The gas uptake and (b) isosteric heats of adsorption of CO_2 in ZIF-8 obtained with MLPs trained on between 125 and 1000 optimizations. Error bars are obtained from block averaging. (c) The test MAE on the energy of snapshots extracted from GCMC simulations as a function of the gas pressure and (d) a density plot of adsorbate atoms extracted from the MLP trained on 1000 optimizations at 10 bar.

loadings, in good agreement with experimentally derived heats of adsorption.[58] This increase in heat of adsorption as a function of the CO_2 loading is indicative for the relatively weak host-guest interactions in ZIF-8 compared to the guest-guest interactions. Even though the most favorable interaction sites for an adsorbate in ZIF-8 are already occupied at higher guest loadings, the heat of adsorption still increases due to the guest-guest stabilization.

To determine the accuracy of the performed GCMC simulations compared to the reference level of theory, 50 snapshots were taken from the GCMC simulations performed with each of the MLPs at each gas pressure. The interaction energy of these snapshots was recomputed with the reference PBE-D3(BJ) level of theory. As shown in Figure 5(c), the resulting average MAE on the energy decreases with increasing N from 0.54 to 0.13 kJ·mol^{-1} per adsorbate. Already for $N = 250$ (representing 7500 total DFT evaluations), a MAE smaller than 0.40 kJ·mol^{-1} per adsorbate is attained. As illustration, the density of adsorbate oxygen atoms extracted from the GCMC simulations using $N = 1000$ at a pressure of 10 bar is shown in Figure 5(d).

Carbon dioxide adsorption in Mg-MOF-74

Having validated our proposed methodology, it is now applied to the challenging case of CO_2 adsorption on the open-metal sites present in Mg-MOF-74. Due to the undercoordination of the metal, CO_2 molecules can approach the Mg site in this MOF to within 2.2 Å, which is closer than would be allowed by the tabulated van der Waals radius for Mg.[46,64] Therefore, the Mg van der Waals radius was set to 1.0 Å, allowing insertions of CO_2 in the framework to within 1.8 Å of the Mg metal site. This choice was validated *a posteriori*, as was done for ZIF-8 in Figure 4, and is shown in Section 3 of the Supporting Information.

From the literature, it is known that the interaction energy of a CO_2 molecule with the Mg open-metal site in Mg-MOF-74 depends significantly on the choice of dispersion correction.[71] Therefore, 1000 optimizations were performed in the framework at the PBE-D3(BJ), PBE-TS and PBE-MBD levels of theory. This selection of dispersion corrections

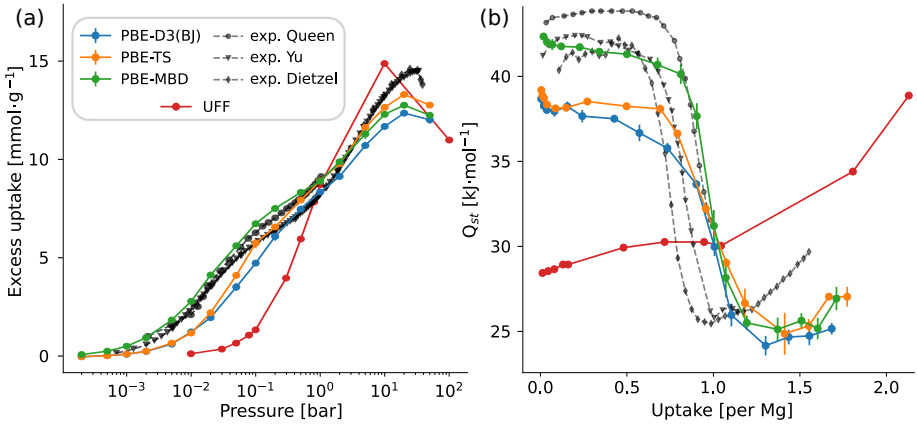


Figure 6: (a) The excess CO_2 uptake and (b) isosteric heats of adsorption in Mg-MOF-74 computed with MLPs trained on PBE-D3(BJ), PBE-TS and PBE-MBD reference optimizations, compared to experiments by Dietzel *et al.*,[72] Yu *et al.*[73] and Queen *et al.*[74] and UFF simulations.[46] All MLP GCMC simulations were performed at 298 K and pressures between $2{\cdot}10^{-4}$ and 50 bar.

was inspired by previous work by Rehak *et al.* in which the adsorption energy on the Mg site obtained with different levels of theory was benchmarked.[71] Therein, the PBE-TS and PBE-MBD functionals most closely agreed with the experimentally derived adsorption energy. Therefore, these two functionals, as well as the commonly used PBE-D3(BJ) functional are considered further.

Training an MLP on the energies and forces results in a validation MAE on the energies of 0.16, 0.20 and 0.16 kJ·mol^{-1} per adsorbate for the PBE-D3(BJ), PBE-TS and PBE-MBD functionals, respectively. Subsequently, GCMC simulations were performed for $2.5{\cdot}10^{6}$ iterations at 298 K and pressures between $2{\cdot}10^{-4}$ and 50 bar. The results are shown in Figure 6 and compared to experiments performed by Dietzel, Yu and Queen *et al.*, as well as simulations performed previously with the Universal Force Field (UFF).[46,72–75] As seen in Figure 6(a), UFF underestimates the uptake at low pressures and overestimates the uptake at the highest pressures. This is due to its limited accuracy in describing the interaction strength with the Mg site (underestimating the heats of adsorption at low uptakes, see Figure 6(b)), as well as

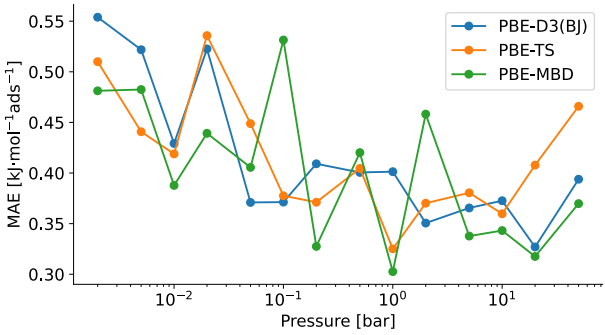


Figure 7: The test MAE on the energy of snapshots extracted from GCMC simulations as a function of the gas pressure for the MLPs trained on PBE-D3(BJ), PBE-TS and PBE-MBD reference optimizations, respectively.

the mutual interaction between CO_2 adsorbates (as seen from the heats of adsorption at the highest uptakes). GCMC simulations performed with the MLPs at each of the three levels of theory significantly improves the description of guest adsorption in the framework. Between these, PBE-MBD performs best, accurately capturing the shape of the adsorption isotherm at pressures below 1 bar. Moreover, as seen in Figure 6(b), the MLP trained on PBE-MBD reference data results in highly accurate heats of adsorption. Compared to the dependence of the heats of adsorption on the uptake seen for ZIF-8 (see Figure 5(b)), the opposite behavior is observed. At loadings below 1 adsorbate per Mg site, a heat of adsorption of approximately 42 kJ·mol^{-1} is obtained, in good agreement with experiments.[72–74] At higher loadings, the heat of adsorption suddenly drops to around 26 kJ·mol^{-1} due to the full occupation of Mg interaction sites, as also observed experimentally. This investigation demonstrates how, compared to UFF, a model which captures the interaction energy with the framework and other adsorbates with high accuracy can result in quantitative predictions of the adsorption properties in an open-metal site containing MOF.

As before, 50 snapshots were extracted from simulations at each gas pressure for the three levels of theory and recomputed. As shown in Figure 7, the MAEs are equal to 0.41, 0.42 and

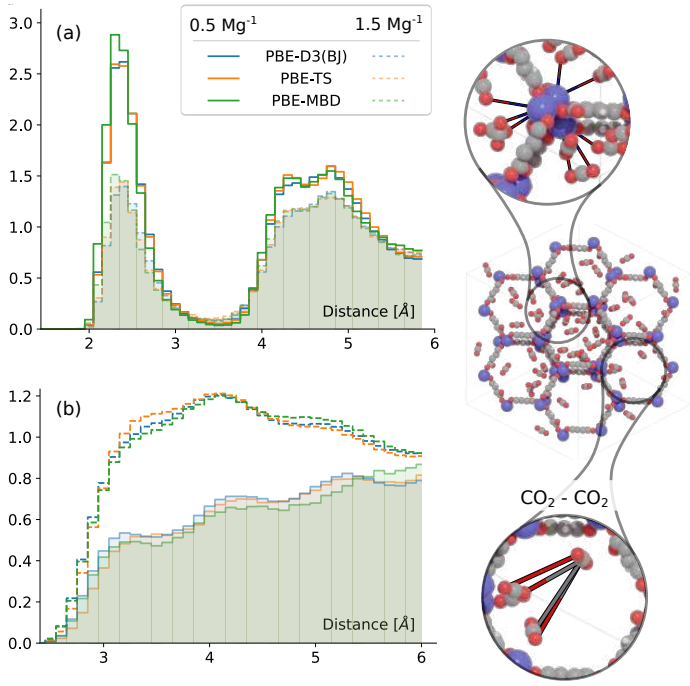


Figure 8: (a) RDF of magnesium–adsorbate oxygen atoms and (b) RDF of the adsorbate–adsorbate atoms computed from GCMC calculations in Mg-MOF-74 at each of the three levels of theory. The full lines correspond to a loading of half an adsorbate per magnesium site, and the dotted lines correspond to a loading of one and a half adsorbates per magnesium site.

0.40 kJ·mol^{-1} per adsorbate for PBE-D3(BJ), PBE-TS and PBE-MBD, respectively. A slight increase in error is seen at the lowest pressures, as for these snapshots all CO_2 molecules are adsorbed on Mg sites, on which the interaction is most difficult to capture. The errors are significantly higher than for CO_2 adsorption in ZIF-8, and can be rationalized by the more diverse adsorption environment in Mg-MOF-74, with significantly stronger interactions on the Mg site compared to the linkers.[52] However, these errors are still deemed acceptable, as differences between the levels of theory can be significantly larger (as seen from the heats of adsorption shown in Figure 6(b)).

To investigate the difference in structure of adsorbed CO_2 in the region of high heats of adsorption at low loadings and lower heats of adsorption at higher loadings (see Figure 6(b)) as well as differences between functionals, radial distribution functions (RDFs) of the magnesium–adsorbate oxyen atoms and adsorbate–adsorbate atoms are computed and shown in Figures 8(a) and 8(b), respectively. Comparing the Mg–O RDFs at low and high adsorbate loadings (0.5 adsorbates versus 1.5 adsorbates per Mg site), shown in Figure 8(a), it is clear that at low loadings guests preferentially adsorb at the Mg site, with a peak in the RDF around 2.4 Å and a secondary broad peak between 4 and 5 Å, representing the interaction with a neighboring Mg site. Compared to this, the high framework loading exhibits a significantly lower first peak as more adsorbates are located in the middle of the pores, further away from the Mg sites. The difference between simulations performed with MLPs trained on each of the three levels of theory is minor. Only in the first peak, a slight shift of approximately 0.05 Å is seen for PBE-MBD compared to the other functionals. In Figure 8(b), the RDF of adsorbate atoms is shown at low and high loadings. As seen from the low occupation of distances between 3 and 4 Å at low loadings, adsorbates are mostly separated from each other, located at each Mg site. Contrary to this, due to the filling of the center of the pore at higher loadings, interatomic distances between 3 and 4 Å are much more represented, with a peak around 4 Å. Again, only minor differences are observed between different functionals.

An investigation into transferability

Given an MLP trained on interaction energies of an adsorbate in a MOF consisting of certain building blocks, one could be interested in modeling adsorption in MOFs comprised of the same building blocks but with different topologies. This approach is exemplified in most common force fields making use of atom types: parameters derived for atoms in certain chemical environments are assumed to transfer well between different systems containing those environments. To investigate whether this concept of transferability can be applied to non-covalent interactions in MOFs modeled by MLPs, GCMC simulations were performed on ZIF-3, ZIF-4 and ZIF-6 (see Figure 9(a)) by use of the MLP trained on 1000 optimizations of CO_2–loaded ZIF-8 (denoted MLP_{ZIF8}). This approach is shown schematically in Figure 9(b), with the resulting isotherms and heats of adsorption shown as dotted curves in Figure 9(c). While the results seem acceptable at first sight, a problem arises when recomputing extracted snapshots at the PBE-D3(BJ) level of theory. MAEs of 2.38, 8.27 and 0.97 $kJ \cdot mol^{-1}$ are obtained for ZIF-3, ZIF-4 and ZIF-6.

As comparison to this, a single MLP was trained instead on 250 optimization in each of the four ZIFs (denoted $MLP_{ZIF\{3,4,6,8\}}$). Note that the same number of total optimizations (1000) are performed for this MLP as for MLP_{ZIF8}. This approach is shown in Figure 9(d), and the resulting adsorption isotherms and heats of adsorption are shown in Figure 9(c) with solid lines. A significantly different adsorption behavior is observed, with consistently higher heats of adsorption and uptakes. For $MLP_{ZIF\{3,4,6,8\}}$, recomputing the extracted snapshots results in MAEs of 0.40, 0.39, 0.32 and 0.32 $kJ \cdot mol^{-1}$ for ZIF-3, ZIF-4, ZIF-6 and ZIF-8. From the difference in heats of adsorption obtained from both methods, it is clear that MLP_{ZIF8} is not capable of reproducing the adsorption behavior in the other ZIFs. This can be rationalized by their substantially different density, probe-occupiable pore volume and surface area (computed using a nitrogen probe with PoreBlazer,[76] see Table 1). For example, ZIF-4 is the densest of the four MOFs, resulting in a significantly stronger interaction of an adsorbate with the adsorbent (as seen from the heats of adsorption

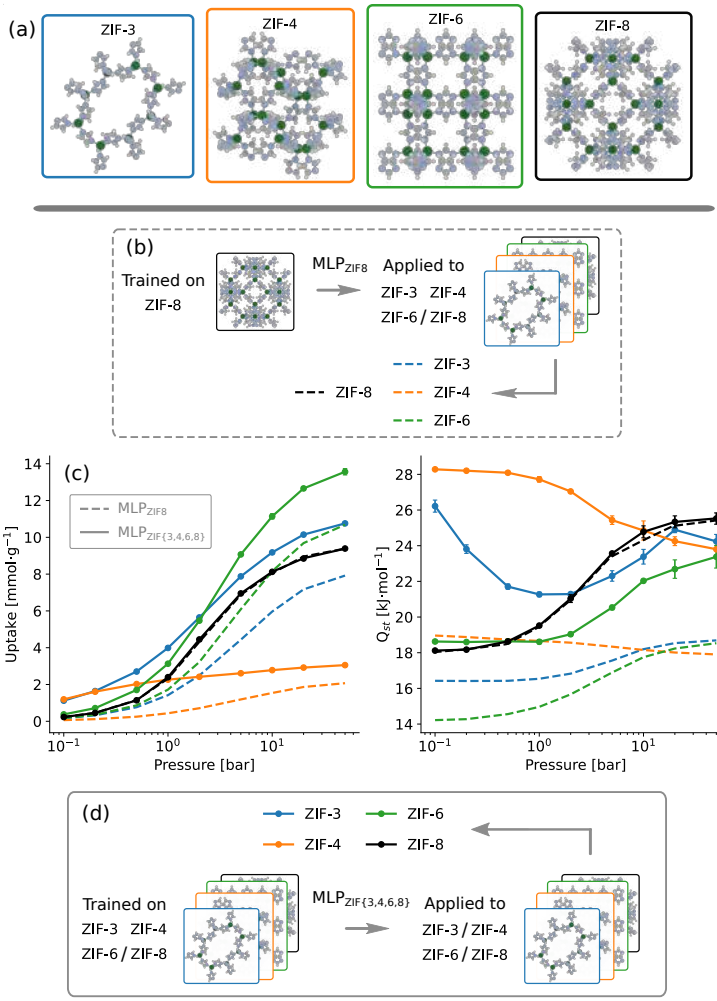


Figure 9: (a) Illustrations of ZIF-3, ZIF-4, ZIF-6 and ZIF-8. (b) An MLP trained on optimizations in ZIF-8 (MLP_{ZIF8}) is applied to perform GCMC simulations in all of the four ZIFs. (c) The uptake and heats of adsorption as a function of the gas pressure obtained from GCMC simulations obtained with both MLP_{ZIF8} in dotted lines and $MLP_{ZIF\{3,4,6,8\}}$ in full lines. (d) An MLP is trained on optimizations in all four ZIFs ($MLP_{ZIF\{3,4,6,8\}}$) and subsequently applied to each.

Table 1: Topology, density, probe-occupiable pore volume (V_{PO}) and surface area of all four ZIFs, computed with PoreBlazer.[76]

	Topology	Density [g/cm^3]	V_{PO} [cm^3/g]	Surface area [m^2/g]
ZIF-3	dft	0.88	0.61	1476
ZIF-4	cag	1.22	0.22	166
ZIF-6	gls	0.76	0.80	2612
ZIF-8	sod	0.92	0.52	1173

in the low pressure regime from Figure 9(c)). Even for ZIF-6, on which MLP_{ZIF8} exhibited a MAE of 0.97 kJ·mol^{-1}, heats of adsorption consistently off by approximately 4 kJ·mol^{-1} are obtained compared to $MLP_{ZIF\{3,4,6,8\}}$. The reason for this discrepancy is the content of the test set. Due to the lacking ability of MLP_{ZIF8} to recognize the favorable adsorption locations in the ZIF-6 framework, the test set does not include these locations. Therefore, the test error is deceptively low, while the actual error at these sites is significantly higher. Only when explicitly training an MLP to reference data which includes these important sites, this failure is alleviated.

To validate whether this statement holds true, it is possible to compute the heats of adsorption at infinite dilution at the reference PBE-D3(BJ) level of theory by means of an importance sampling approach, proposed by Vandenbrande *et al.*[49] First, a set of 10^6 random insertions in each of the ZIFs are constructed for which the $MLP_{ZIF\{3,4,6,8\}}$ interaction energies are computed. Subsequently, a set of $N_{\text{importance}}$ insertions are extracted from this wider set with a probability proportional to a Boltzmann factor. Recomputing those $N_{\text{importance}}$ snapshots, the heats of adsorption at the reference level of theory can be obtained from the following Equation:

$$Q_{st} = \frac{\langle \Delta U \exp\left(-\beta(\Delta U - \Delta\tilde{U})\right)\rangle}{\langle \exp\left(-\beta(\Delta U - \Delta\tilde{U})\right)\rangle} \tag{4}$$

with $\Delta\tilde{U}$ and ΔU the interaction energies computed with the MLP and PBE-D3(BJ), respectively, $\beta = 1/k_B T$ and the ensemble average a canonical average over the MLP energies. In case the MLP accurately reproduces the reference PES, only a limited number of insertions $N_{\text{importance}}$ are required, keeping the computational cost for this procedure manageable. In

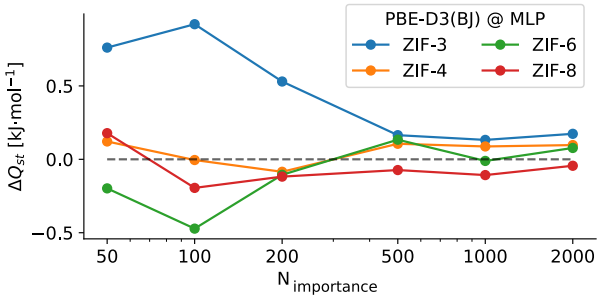


Figure 10: Difference between the MLP and reference PBE-D3(BJ) heats of adsorption for all four ZIFs. MLP results are obtained from 10^6 Widom insertions, and the reference heats of adsorption are obtained from recomputing importance sampled configurations from the MLP insertions. Convergence with respect to the number of recomputed snapshots $N_{\text{importance}}$ is shown for $N_{\text{importance}}$ between 50 and 2000.

Figure 10, the difference between the MLP and PBE-D3(BJ) importance sampled heats of adsorption obtained in this manner are shown. With $N_{\text{importance}} = 1000$, convergence within 0.10 kJ·mol^{-1} is obtained. As can be seen, the heats of adsorption obtained with the MLP compare well to those computed from importance sampled PBE-D3(BJ) insertions. For all four ZIFs, the error on the predicted heats of adsorption are lower than 0.20 kJ·mol^{-1}, further validating the derived MLP.

This investigation highlights how even seemingly well-performing MLPs can fail in unforeseen ways when sampling parts of the PES on which it was not adequately trained. While such failures can be detected by validation on optimizations performed at the reference DFT level of theory, this implies additional computationally demanding calculations. However, from the results obtained in this work, it is clear that accurate MLPs can be derived from a dataset containing between 250 and 1000 optimizations performed at the DFT level of theory (between 7500 and 30000 single point calculations, respectively). Performing GCMC calculations at 10 gas pressures to construct an isotherm at the reference DFT level of theory would require between 3 and 4 orders of magnitude more single point calculations. On the other hand, with an MLP trained for 27 hours on a single NVIDIA V100 GPU, a typical GCMC calculation consisting of $5 \cdot 10^6$ iterations can be performed in approximately 50 hours

on a single NVIDIA V100 GPU. It is therefore clear that properly trained MLPs can vastly extend the applicability of DFT calculations, and can be expected to be of significant help in modeling adsorption in challenging systems such as open-metal site containing MOFs.

Conclusions

In this work, we proposed a generally applicable methodology for performing GCMC simulations by use of MLPs trained on DFT intermolecular reference energies and forces. First, an appropriate dataset was constructed to train the MLPs. For this, snapshots of guest-loaded frameworks are generated based on van der Waals 'exclusion spheres' and subsequently (rigidly) optimized using the relevant level of theory. With 7500 single point calculations (250 trajectories, with 30 optimizations steps each) a test MAE on the energy of less than 0.4 $kJ \cdot mol^{-1}$ per adsorbate can be achieved, as calculated from snapshots extracted from GCMC simulations performed with the MLP. Having initially benchmarked the methodology on CO_2 adsorption in ZIF-8, it was subsequently applied to model CO_2 adsorption in Mg-MOF-74, an open-metal site containing MOF which has attracted attention for its interesting adsorption properties, as well as the challenge in modeling these properties computationally using force fields. Comparing three different dispersion corrections combined with the PBE functional for modeling adsorption in this MOF, PBE-MBD showed excellent agreement between the computed and experimentally obtained adsorption isotherms and heats of adsorption at 298 K. The characteristic drop in the heat of adsorption of CO_2 in this MOF between low and high uptakes from approximately 42 $kJ \cdot mol^{-1}$ to 26 $kJ \cdot mol^{-1}$ is quantitatively predicted from the performed GCMC simulations. Lastly, also the transferability of an MLP trained on ZIF-8 towards other ZIFs was tested. In this case, significant under-predictions of the uptakes and heats of adsorption were obtained for a set of other ZIFs composed of the same building blocks. This investigation highlighted the difficulty in training transferable MLPs, warranting caution when MLPs are employed to model ad-

sorption in MOFs with structural differences to those/that in the training set. Only when explicitly training to data for all ZIFs, accurate results are obtained.

The general methodology proposed here has potential to be widely applicable not only to model adsorption in MOFs, but also in other nanoporous materials such as covalent organic frameworks (COFs), zeolites or graphene derivatives. By training an MLP on limited set of DFT reference data, a 'best of both worlds' compromise is made between the accuracy of DFT and the computational efficiency of force fields, unlocking the potential for routine GCMC simulations at DFT accuracy. Moreover, it could be employed as a tool to benchmark computationally cheaper classical force fields where experimental reference data is not available. Therefore, this work holds potential to aid accurate computational screening of adsorption properties in nanoporous materials, supporting application-focused future research.

Care should still be taken in choosing an accurate reference level of theory. As shown in this work, the choice of dispersion correction alone can affect the obtained heat of adsorption of CO_2 in Mg-MOF-74 by as much as 5 kJ·mol^{-1}. For MOFs containing transition metals, the presence of multiple magnetic configurations could further complicate the generation of accurate reference data.[49,73] Secondly, this work only considered rigid frameworks. However, it could easily be extended to flexible adsorbents by including different framework geometries and training on the total energy and atomic forces (not just the intermolecular energy and forces). With this, hybrid GCMC/MD schemes could be employed to properly sample the adsorbate-induced flexibility of frameworks.[77]

Computational details

DFT optimizations

All single point calculations were performed with VASP, using the projected augmented wavefunction (PAW) PBE potentials.[78–81] The electronic convergence threshold for the en-

ergy was set to 10^{-5} eV. For ZIF-8, a single k-point was used, while for Mg-MOF-74 a $2\times2\times2$ k-point mesh was used for all PBE-D3(BJ) and PBE-TS calculations and a $3\times3\times3$ k-point mesh for all PBE-MBD calculations (see Supporting Information). An appropriate value for the cutoff energy for CO_2 adsorption in ZIF-8 was determined by performing 20 optimizations of guest-loaded frameworks for 30 steps using a cutoff of 700 eV. The final 20 structures were recomputed with cutoff energies between 350 eV and 650 eV. As shown in the Supporting Information, the interaction energy is converged to within 0.05 kJ·mol^{-1} using a cutoff energy of 500 eV. For Mg-MOF-74, a cutoff energy of 600 eV was used in accordance with previous work.[71] Furthermore, as the difference between calculations performed with the precision set to `normal` compared to `accurate` is also within 0.05 kJ·mol^{-1}, the precision was set to `normal`. For the optimizations, the BFGS optimizer from ASE was used with a maximum step size of 0.5 Å.[65]

MLP hyperparameters

The effect of different hyperparameters of the network on both the accuracy and computational efficiency are tested and are included in the Supporting Information. The used hyperparameters are as follows: a cutoff radius of 5 Å is used, combined with 4 interaction layers, 32 features and a maximum rotation order of 1. The weights $\lambda_E = 50$ and $\lambda_F = 1$ are used to train the network. Although forces are not strictly necessary (only the energy is required in GCMC simulations), including forces is known to improve data efficiency during training.

Acknowledgement

The authors acknowledge the financial support from the Research Board of Ghent University (BOF), the Flemish Government and Flanders Innovation & Entrepreneurship (VLAIO) through the Moonshot project MOONRISE, the computational resources (Stevin Supercom-

puter Infrastructure) and services provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, FWO, and the Flemish Government, department EWI.

Supporting Information Available

Benchmarking of the used DFT settings, determination of the MLP hyperparameters and validation of the used van der Waals radius for the Mg site in Mg-MOF-74. A zip file containing the simulation data and software implementations is archived on Zenodo (10.5281/zenodo.7904959).

References

(1) Li, J. R.; Kuppler, R. J.; Zhou, H. C. Selective gas adsorption and separation in metal-organic frameworks. *Chemical Society Reviews* **2009**, *38*, 1477–1504.

(2) Li, J. R.; Ma, Y.; McCarthy, M. C.; Sculley, J.; Yu, J.; Jeong, H. K.; Balbuena, P. B.; Zhou, H. C. Carbon dioxide capture-related gas adsorption and separation in metal-organic frameworks. 2011.

(3) Sumida, K.; Rogow, D. L.; Mason, J. A.; McDonald, T. M.; Bloch, E. D.; Herm, Z. R.; Bae, T. H.; Long, J. R. Carbon dioxide capture in metal-organic frameworks. *Chemical Reviews* **2012**, *112*, 724–781.

(4) Li, H.; Eddaoudi, M.; Groy, T. L.; Yaghi, O. M. Establishing microporosity in open metal-organic frameworks: Gas sorption isotherms for Zn(BDC) (BDC = 1,4-benzenedicarboxylate) [28]. *Journal of the American Chemical Society* **1998**, *120*, 8571–8572.

(5) Yaghi, O. M.; Li, H.; Davis, C.; Richardson, D.; Groy, T. L. *Synthetic Strategies,*

Structure Patterns, and Emerging Properties in the Chemistry of Modular Porous Solids ; 1998.

(6) Kepert, C. J.; Rosseinsky, M. J. Zeolite-like crystal structure of an empty microporous molecular framework. *Chemical Communications* **1999**, 375–376.

(7) Li, H.; Eddaoudi, M.; O'Keeffe, M.; Yaghi, O. M. Design and synthesis of an exceptionally stable and highly porous metal- organic framework. *Nature* **1999**, *402*, 276–279.

(8) Chui, S. S.; Lo, S. M.; Charmant, J. P.; Orpen, A. G.; Williams, I. D. A chemically functionalizable nanoporous material [Cu3(TMA)2 (H2O)3](n). *Science* **1999**, *283*, 1148–1150.

(9) Eddaoudi, M.; Kim, J.; Rosi, N.; Vodak, D.; Wachter, J.; O'Keeffe, M.; Yaghi, O. M. Systematic design of pore size and functionality in isoreticular MOFs and their application in methane storage. *Science* **2002**, *295*, 469–472.

(10) Klontzas, E.; Mavrandonakis, A.; Tylianakis, E.; Froudakis, G. E. Improving hydrogen storage capacity of MOF by functionalization of the organic linker with lithium atoms. *Nano Letters* **2008**, *8*, 1572–1576.

(11) Lyu, J.; Zhang, X.; Otake, K. I.; Wang, X.; Li, P.; Li, Z.; Chen, Z.; Zhang, Y.; Wasson, M. C.; Yang, Y.; Bai, P.; Guo, X.; Islamoglu, T.; Farha, O. K. Topology and porosity control of metal-organic frameworks through linker functionalization. *Chemical Science* **2019**, *10*, 1186–1192.

(12) Ali Akbar Razavi, S.; Morsali, A. Linker functionalized metal-organic frameworks. *Coordination Chemistry Reviews* **2019**, *399*, 213023.

(13) Düren, T.; Bae, Y. S.; Snurr, R. Q. Using molecular simulation to characterise metalorganic frameworks for adsorption applications. *Chemical Society Reviews* **2009**, *38*, 1237–1247.

(14) Han, S. S.; Mendoza-Cortés, J. L.; Goddard, W. A. Recent advances on simulation and theory of hydrogen storage in metalorganic frameworks and covalent organic frameworks. *Chemical Society Reviews* **2009**, *38*, 1460–1476.

(15) Getman, R. B.; Bae, Y. S.; Wilmer, C. E.; Snurr, R. Q. Review and analysis of molecular simulations of methane, hydrogen, and acetylene storage in metal-organic frameworks. *Chemical Reviews* **2012**, *112*, 703–723.

(16) Yang, Q.; Liu, D.; Zhong, C.; Li, J. R. Development of computational methodologies for metal-organic frameworks and their application in gas separations. *Chemical Reviews* **2013**, *113*, 8261–8323.

(17) Wu, D.; Wang, C.; Liu, B.; Liu, D.; Yang, Q.; Zhong, C. Large-scale computational screening of metal-organic frameworks for CH4/H2 separation. *AIChE Journal* **2012**, *58*, 2078–2084.

(18) Colón, Y. J.; Snurr, R. Q. High-throughput computational screening of metal-organic frameworks. *Chemical Society Reviews* **2014**, *43*, 5735–5749.

(19) Yu, J.; Xie, L. H.; Li, J. R.; Ma, Y.; Seminario, J. M.; Balbuena, P. B. CO2 Capture and Separations Using MOFs: Computational and Experimental Studies. *Chemical Reviews* **2017**, *117*, 9674–9754.

(20) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Large-scale screening of hypothetical metal-organic frameworks. *Nature Chemistry* **2012**, *4*, 83–89.

(21) Gómez-Gualdrón, D. A.; Colón, Y. J.; Zhang, X.; Wang, T. C.; Chen, Y. S.; Hupp, J. T.; Yildirim, T.; Farha, O. K.; Zhang, J.; Snurr, R. Q. Evaluating topologically diverse metal-organic frameworks for cryo-adsorbed hydrogen storage. *Energy and Environmental Science* **2016**, *9*, 3279–3289.

(22) Lin, L. C.; Berger, A. H.; Martin, R. L.; Kim, J.; Swisher, J. A.; Jariwala, K.; Rycroft, C. H.; Bhown, A. S.; Deem, M. W.; Haranczyk, M.; Smit, B. In silico screening of carbon-capture materials. *Nature Materials* **2012**, *11*, 633–641.

(23) Kim, J.; Martin, R. L.; Rübel, O.; Haranczyk, M.; Smit, B. High-Throughput Characterization of Porous Materials Using Graphics Processing Units. *Journal of Chemical Theory and Computation* **2012**, *8*, 1684–1693.

(24) Ahmed, A.; Siegel, D. J. Predicting hydrogen storage in MOFs via machine learning. *Patterns* **2021**, *2*, 100291.

(25) Altintas, C.; Altundal, O. F.; Keskin, S.; Yildirim, R. Machine Learning Meets with Metal Organic Frameworks for Gas Storage and Separation. 2021; `https://doi.org/10.1021/acs.jcim.1c00191`.

(26) Guo, W.; Liu, J.; Dong, F.; Chen, R.; Das, J.; Ge, W.; Xu, X.; Hong, H. Deep Learning Models for Predicting Gas Adsorption Capacity of Nanomaterials. *Nanomaterials* **2022**, *12*, 3376.

(27) Borboudakis, G.; Stergiannakos, T.; Frysali, M.; Klontzas, E.; Tsamardinos, I.; Froudakis, G. E. Chemically intuited, large-scale screening of MOFs by machine learning techniques. *npj Computational Materials* **2017**, *3*, 40.

(28) Shi, K.; Li, Z.; Anstine, D. M.; Tang, D.; Colina, C. M.; Sholl, D. S.; Siepmann, J. I.; Snurr, R. Q. Two-Dimensional Energy Histograms as Features for Machine Learning to Predict Adsorption in Diverse Nanoporous Materials. *Journal of Chemical Theory and Computation* **2023**,

(29) Gheytanzadeh, M.; Baghban, A.; Habibzadeh, S.; Esmaeili, A.; Abida, O.; Mohaddespour, A.; Muhammad, .; Munir, T. Towards estimation of CO2 adsorption on highly porous MOF-based adsorbents using gaussian process regression approach. *Scientific Reports* — **2021**, *11*, 15710.

(30) Hung, T. H.; Xu, Z. X.; Kang, D. Y.; Lin, L. C. Chemistry-Encoded Convolutional Neural Networks for Predicting Gaseous Adsorption in Porous Materials. *Journal of Physical Chemistry C* **2022**, *126*, 2813–2822.

(31) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *Journal of the American Chemical Society* **1992**, *114*, 10024–10035.

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

(33) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 1. *Journal of the American Chemical Society* **1989**, *111*, 8551–8566.

(34) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.

(35) Addicoat, M. A.; Vankova, N.; Akter, I. F.; Heine, T. Extension of the universal force field to metal-organic frameworks. *Journal of Chemical Theory and Computation* **2014**, *10*, 880–891.

(36) Bureekaew, S.; Amirjalayer, S.; Tafipolsky, M.; Spickermann, C.; Roy, T. K.; Schmid, R. MOF-FF - A flexible first-principles derived force field for metal-organic frameworks. *Physica Status Solidi (B) Basic Research* **2013**, *250*, 1128–1141.

(37) Boulanger, E.; Huang, L.; Rupakheti, C.; Mackerell, A. D.; Roux, B. Optimized Lennard-Jones Parameters for Druglike Small Molecules. *Journal of Chemical Theory and Computation* **2018**, *14*, 3121–3131.

(38) Wu, X.; Huang, J.; Cai, W.; Jaroniec, M. Force field for ZIF-8 flexible frameworks:

Atomistic simulation of adsorption, diffusion of pure gases as CH4, H2, CO 2 and N2. *RSC Advances* **2014**, *4*, 16503–16511.

(39) Pérez-Pellitero, J.; Amrouche, H.; Siperstein, F. R.; Pirngruber, G.; Nieto-Draghi, C.; Chaplais, G.; Simon-Masseron, A.; Bazer-Bachi, D.; Peralta, D.; Bats, N. Adsorption of CO2, CH4, and N2 on zeolitic imidazolate frameworks: Experiments and simulations. *Chemistry - A European Journal* **2010**, *16*, 1560–1571.

(40) Fischer, M.; Gomes, J. R.; Jorge, M. Computational approaches to study adsorption in MOFs with unsaturated metal sites. *Molecular Simulation* **2014**, *40*, 537–556.

(41) Belof, J. L.; Stern, A. C.; Space, B. An Accurate and Transferable Intermolecular Diatomic Hydrogen Potential for Condensed Phase Simulation.

(42) Belof, J. L.; Stern, A. C.; Eddaoudi, M.; Space, B. On the mechanism of hydrogen storage in a metal-organic framework material. *Journal of the American Chemical Society* **2007**, *129*, 15202–15210.

(43) Franz, D. M.; Dyott, Z. E.; Forrest, K. A.; Hogan, A.; Pham, T.; Space, B. Simulations of hydrogen, carbon dioxide, and small hydrocarbon sorption in a nitrogen-rich rht-metal-organic framework. *Physical Chemistry Chemical Physics* **2018**, *20*, 1761–1777.

(44) Becker, T. M.; Heinen, J.; Dubbeldam, D.; Lin, L. C.; Vlugt, T. J. Polarizable Force Fields for CO2 and CH4 Adsorption in M-MOF-74. *Journal of Physical Chemistry C* **2017**, *121*, 4659–4673.

(45) Becker, T. M.; Luna-Triguero, A.; Vicent-Luna, J. M.; Lin, L. C.; Dubbeldam, D.; Calero, S.; Vlugt, T. J. Potential of polarizable force fields for predicting the separation performance of small hydrocarbons in M-MOF-74. *Physical Chemistry Chemical Physics* **2018**, *20*, 28848–28859.

(46) Becker, T. M.; Lin, L. C.; Dubbeldam, D.; Vlugt, T. J. Polarizable Force Field for CO2 in M-MOF-74 Derived from Quantum Mechanics. *Journal of Physical Chemistry C* **2018**, *122*, 24488–24498.

(47) Chang, C.; Deringer, V. L.; Katti, K. S.; Van Speybroeck, V.; Wolverton, C. M. Simulations in the era of exascale computing. *Nature Reviews Materials* **2023**, 1–5.

(48) Lee, Y.; Poloni, R.; Kim, J. Probing gas adsorption in MOFs using an efficient ¡i¿ab initio¡/i¿ widom insertion Monte Carlo method. *Journal of Computational Chemistry* **2016**, *37*, 2808–2815.

(49) Vandenbrande, S.; Waroquier, M.; Van Speybroeck, V.; Verstraelen, T. Ab Initio Evaluation of Henry Coefficients Using Importance Sampling. *Journal of Chemical Theory and Computation* **2018**, *14*, 6359–6369.

(50) Kundu, A.; Piccini, G.; Sillar, K.; Sauer, J. Ab Initio Prediction of Adsorption Isotherms for Small Molecules in Metal-Organic Frameworks. *Journal of the American Chemical Society* **2016**, *138*, 14047–14056.

(51) Sillar, K.; Kundu, A.; Sauer, J. Ab Initio Adsorption Isotherms for Molecules with Lateral Interactions: CO2 in Metal-Organic Frameworks. *Journal of Physical Chemistry C* **2017**, *121*, 12789–12799.

(52) Kundu, A.; Sillar, K.; Sauer, J. Ab Initio Prediction of Adsorption Isotherms for Gas Mixtures by Grand Canonical Monte Carlo Simulations on a Lattice of Sites. *Journal of Physical Chemistry Letters* **2017**, *8*, 2713–2718.

(53) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications* **2022**, *13*, 1–11.

(54) Eckhoff, M.; Behler, J. From Molecular Fragments to the Bulk: Development of a Neural Network Potential for MOF-5. *Journal of Chemical Theory and Computation* **2019**, *15*, 3793–3809.

(55) Achar, S. K.; Wardzala, J. J.; Bernasconi, L.; Zhang, L.; Johnson, J. K. Combined Deep Learning and Classical Potential Approach for Modeling Diffusion in UiO-66. *Journal of Chemical Theory and Computation* **2022**, *18*, 3593–3606.

(56) Yu, Y.; Zhang, W.; Mei, D. Artificial Neural Network Potential for Encapsulated Platinum Clusters in MOF-808. *Journal of Physical Chemistry C* **2022**, *126*, 1204–1214.

(57) Vandenhaute, S.; Cools-Ceuppens, M.; DeKeyser, S.; Verstraelen, T.; Van Speybroeck, V. Machine learning potentials for metal-organic frameworks using an incremental learning approach. *npj Computational Materials* **2023**, *9*, 1–8.

(58) Simmons, J. M.; Wu, H.; Zhou, W.; Yildirim, T. Carbon capture in metal-organic frameworks - A comparative study. *Energy and Environmental Science* **2011**, *4*, 2177–2185.

(59) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Physical Review Letters* **1996**, *77*, 3865–3868.

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

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

(62) Tkatchenko, A.; Scheffler, M. Accurate molecular van der Waals interactions from

ground-state electron density and free-atom reference data. *Physical Review Letters* **2009**, *102*, 073005.

(63) Ambrosetti, A.; Reilly, A. M.; Distasio, R. A.; Tkatchenko, A. Long-range correlation energy calculated from coupled atomic response functions. *Journal of Chemical Physics* **2014**, *140*, 18A508.

(64) Alvarez, S. A cartography of the van der Waals territories. *Dalton Transactions* **2013**, *42*, 8617–8636.

(65) Hjorth Larsen, A. et al. The atomic simulation environment - A Python library for working with atoms. *Journal of Physics Condensed Matter* **2017**, *29*, 30.

(66) Nicholson, D.; Parsonage, N. G. *Computer simulation and the statistical mechanics of adsorption*; Academic Press: London, 1982.

(67) Widom, B. Some topics in the theory of fluids. *The Journal of Chemical Physics* **1963**, *39*, 2808–2812.

(68) Abraha, Y. W.; Tsai, C.-W.; Niemantsverdriet, J. W. H.; Langner, E. H. G. Optimized CO 2 Capture of the Zeolitic Imidazolate Framework ZIF-8 Modified by Solvent-Assisted Ligand Exchange. **2021**,

(69) Garces, S. I.; Villarroel-Rocha, J.; Sapag, K.; Korili, S. A.; Gil, A. Comparative Study of the Adsorption Equilibrium of CO 2 on Microporous Commercial Materials at Low Pressures. **2013**,

(70) Fairen-Jimenez, D.; Moggach, S. A.; Wharmby, M. T.; Wright, P. A.; Parsons, S.; D Uren, T. Opening the Gate: Framework Flexibility in ZIF-8 Explored by Experiments and Simulations. *J. Am. Chem. Soc* **2011**, *133*, 8900–8902.

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

(72) Dietzel, P. D.; Besikiotis, V.; Blom, R. Application of metal-organic frameworks with coordinatively unsaturated metal sites in storage and separation of methane and carbon dioxide. *Journal of Materials Chemistry* **2009**, *19*, 7362–7370.

(73) Yu, D.; Yazaydin, A. O.; Lane, J. R.; Dietzel, P. D.; Snurr, R. Q. A combined experimental and quantum chemical study of CO2 adsorption in the metal-organic framework CPO-27 with different metals. *Chemical Science* **2013**, *4*, 3544–3556.

(74) Queen, W. L. et al. Comprehensive study of carbon dioxide adsorption in the metal-organic frameworks M2(dobdc) (M = Mg, Mn, Fe, Co, Ni, Cu, Zn). *Chemical Science* **2014**, *5*, 4569–4581.

(75) Rappé, A. K.; Goddard, W. A. Charge equilibration for molecular dynamics simulations. *Journal of Physical Chemistry* **1991**, *95*, 3358–3363.

(76) Sarkisov, L.; Bueno-Perez, R.; Sutharson, M.; Fairen-Jimenez, D. Materials Informatics with PoreBlazer v4.0 and the CSD MOF Database. *Chemistry of Materials* **2020**, *32*, 9849–9867.

(77) Rogge, S. M. J.; Goeminne, R.; Demuynck, R.; GutiérrezSevillano, J. J.; Vandenbrande, S.; Vanduyfhuys, L.; Waroquier, M.; Verstraelen, T.; Van Speybroeck, V. Modeling Gas Adsorption in Flexible MetalOrganic Frameworks via Hybrid Monte Carlo/Molecular Dynamics Schemes. *Advanced Theory and Simulations* **2019**, *2*, 1800177.

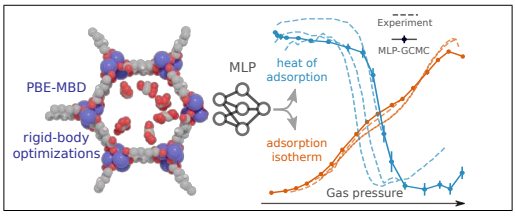
(78) Kresse, G.; Hafner, J. Ab initio molecular dynamics for liquid metals. *Physical Review B* **1993**, *47*, 558–561.

(79) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B - Condensed Matter and Materials Physics* **1996**, *54*, 11169–11186.

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

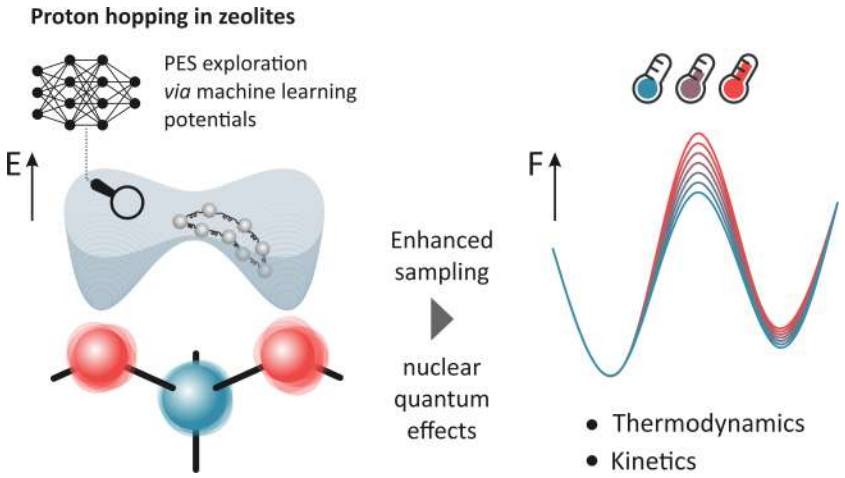
(81) Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B - Condensed Matter and Materials Physics* **1999**, *59*, 1758–1775.

Graphical TOC Entry



Paper V

Nuclear Quantum Effects on Zeolite Proton Hopping Kinetics Explored with Machine Learning Potentials and Path Integral Molecular Dynamics



M. Bocus†, R. Goeminne†, A. Lamaire, M. Cools-Ceuppens, T. Verstraelen, V. Van Speybroeck

Nat. Commun., **2023**, *14* (1008)

† Authors contributed equally to this work.
R. Goeminne trained the MLP and performed all related simulations, analyzed the results and contributed to the manuscript.

nature communications

Article

https://doi.org/10.1038/s41467-023-36666-y

Nuclear quantum effects on zeolite proton hopping kinetics explored with machine learning potentials and path integral molecular dynamics

Received: 26 July 2022

Accepted: 10 February 2023

Published online: 23 February 2023

Check for updates

Massimo Bocus[1,2], Ruben Goeminne[1,2], Aran Lamaire[1], Maarten Cools-Ceuppens[1], Toon Verstraelen[1] & Veronique Van Speybroeck[1]

Proton hopping is a key reactive process within zeolite catalysis. However, the accurate determination of its kinetics poses major challenges both for theoreticians and experimentalists. Nuclear quantum effects (NQEs) are known to influence the structure and dynamics of protons, but their rigorous inclusion through the path integral molecular dynamics (PIMD) formalism was so far beyond reach for zeolite catalyzed processes due to the excessive computational cost of evaluating all forces and energies at the Density Functional Theory (DFT) level. Herein, we overcome this limitation by training first a reactive machine learning potential (MLP) that can reproduce with high fidelity the DFT potential energy surface of proton hopping around the first Al coordination sphere in the H-CHA zeolite. The MLP offers an immense computational speedup, enabling us to derive accurate reaction kinetics beyond standard transition state theory for the proton hopping reaction. Overall, more than 0.6 μs of simulation time was needed, which is far beyond reach of any standard DFT approach. NQEs are found to significantly impact the proton hopping kinetics up to ~473 K. Moreover, PIMD simulations with deuterium can be performed without any additional training to compute kinetic isotope effects over a broad range of temperatures.

Brønsted-acidic zeolites are versatile, resistant catalysts that for decades have been recognized as the workhorse of the petrochemical industry[1]. Furthermore, they are also expected to play a vital role in next-generation biorefineries for the conversion of non-fossil feedstocks[2]. From a theoretical point of view, zeolites belong to the most studied materials in the field of heterogeneous computational catalysis[3]. In their ideal, defect-free crystalline form, Brønsted-acidic zeolites are composed of interconnected SiO_4 tetrahedra, where a fraction of the Si^{4+} ions is substituted by Al^{3+}. The excess of negative charge is compensated by the addition of a proton—the Brønsted Acid Site (BAS)—on one of the oxygens in the first coordination sphere of the Al substitution. Interestingly, the BAS is not confined to a specific oxygen of the Al tetrahedron, but it can jump from one oxygen atom to another in what is commonly known as the 'proton hopping' reaction. This process is one of the most fundamental activated events within zeolite chemistry (Fig. 1a) and represents the archetypal proton-transfer reaction which is at the base of any Brønsted acid-catalyzed reaction.

Because of its apparent simplicity, proton hopping is an ideal case study for both experiment and theory, hence various methods have been used to investigate the process kinetics. Experimentally, Nuclear Magnetic Resonance (NMR)[4–6], Impedance Spectroscopy (IS)[7] and

[1]Center for Molecular Modeling, Ghent University, Technologiepark 46, 9052 Zwijnaarde, Belgium. [2]These authors contributed equally: Massimo Bocus, Ruben Goeminne. e-mail: Veronique.Vanspeybroeck@Ugent.be

Article https://doi.org/10.1038/s41467-023-36666-y

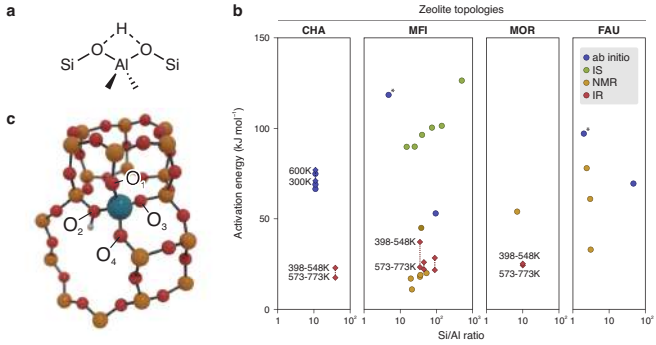


Fig. 1 | Poor agreement is found in the available literature for the activation energy of the proton hopping reaction. a Schematic depiction of the proton hopping reaction. **b** Activation energy for the proton hopping process as function of the Si/Al ratio for multiple zeolites, as derived from the available literature[75]. The data was obtained from ab initio calculations (blue, the points with an asterisk correspond to cluster calculations and, therefore, the Si/Al ratio is meaningless) or IR (red), NMR (yellow) and IS (green) spectroscopies. If more values are available for different temperatures, they are reported with diamonds interconnected by a dotted line. For more details about the reported values and a full list of references, the interested reader is referred to Supplementary Note 1. **c** Part of the H-CHA unit cell, showing the conventional nomenclature of the oxygen atoms in the first coordination sphere of the Al defect adopted herein (Si is in yellow, O in red, Al in blue and H in white).

InfraRed spectroscopy (IR)[8] have been employed to retrieve the activation energies for the proton hopping process. From the theoretical side, the reaction has been tackled with various methodologies ranging from static simulations[9–15] to enhanced-sampling techniques based on molecular dynamics (MD)[16]. Given this plethora of scientific reports, it would be tempting to assume that every detail of the proton hopping reaction is now revealed. However, when surveying the available literature, a huge spread in both the theoretically and experimentally obtained activation energies for proton hopping barriers can be found (Fig. 1b).

In general, activation energies derived from NMR spectroscopy are lower than the theoretical ones. From IR experiments, two different activation energies were obtained for two different temperature ranges (398-548 and 573–773 K, see red diamonds in Fig. 1), a fact that was attributed to the switch from intra-site hopping to inter-site hopping[8]. However, a more recent investigation has disproven such interpretation and indicated active site proximity effects as the main cause for the observed change in activation energy[17]. Inter-site hopping was also suggested to be responsible for the high activation energies retrieved with IS[7].

To understand this lack of consistency, it is important to consider the main possible sources of discrepancy between the proton hopping barriers from literature. First, the residual presence of water in the catalyst is often indicated as the main source for the—in general—low experimental barriers[10], as it is almost impossible to achieve a completely dry material with routine drying procedures[18]. Moreover, the presence of defective sites like extra-framework aluminum species is known to alter the BAS' acidity compared to the pristine material[19]. On the theoretical side, most of the calculations performed so far did not explicitly account for the quantum nature of the hydrogen nucleus. Instead, the nuclei in the system are treated as classical particles moving on an underlying Potential Energy Surface (PES), which is obtained by solving the electronic many-body problem using quantum many-body techniques. This is normally done using Density Functional Theory (DFT) for the sake of computational efficiency. In what follows, the terminology 'classical DFT PES' will be used to refer to nuclei that are treated as classical particles on a DFT-determined PES, thus the electronic degrees of freedom are treated quantum mechanically whereas the nuclei are treated as classical particles. To include Nuclear Quantum Effects (NQEs), approximative methods have been used. For example, tunneling corrections have sometimes been applied to account for NQEs[12]. To more rigorously account for NQEs, one should resort to computationally more expensive methods such as the Path Integral Molecular Dynamics (PIMD) approach, which relies on Feynman's path integral formulation of quantum mechanics. Within PIMD simulations, the statistics of quantum particles are retrieved using a classical ring polymer consisting of P replicas of the system[20]. Each replica runs on the classical DFT PES, making PIMD P times more expensive than a standard MD simulation. This is because an independent DFT-level energy and force evaluation must be performed every MD step for each replica. Within the field of heterogeneous-catalyzed reactions such simulations have so far been mostly unfeasible due to the high computational cost of each PIMD step, as at least 10 replicas are usually required to achieve converged results—making the simulations prohibitively expensive[21]. Nonetheless, there is clear evidence that NQEs may have a significant impact on the physicochemical properties of systems containing light atoms[20,22–24]. For example, it is well-known that they can significantly affect the strength of hydrogen bonds in a variety of systems[25,26]. NQEs have never been explicitly included in zeolite-related reactions and thus it remains unclear to what extent they would affect the rate of proton hopping and by extension any proton-transfer reactions within the field of zeolite catalysis.

To fill this gap of knowledge, we present in the current contribution a methodology that may allow to systematically include NQEs when investigating activated hydrogen transfer events. To this end, proton hopping in H-CHA with isolated active sites is used as a case study (Fig. 1c), for which we first trained an accurate Machine Learning Potential (MLP) based on an underlying set of high-temperature DFT Umbrella Sampling (US) simulations. The use of enhanced-sampling simulations is essential to explore in an efficient way the less-probable highly energetic regions of the PES, which are typically associated with reactive events. The underlying DFT simulations at finite temperatures serve as input to train a deep neural network MLP (Fig. 2). Once an accurate MLP is constructed, an enormous computational speedup can be achieved, which allowed to: (i) compute the Free Energy Surfaces (FESs) of all possible hoppings around an isolated Al defect in the temperature range 273–873 K with a large number of umbrellas and long simulation times to obtain well-converged results, (ii) explicitly include NQEs through the PIMD approach, (iii) derive accurate kinetic

Article https://doi.org/10.1038/s41467-023-36666-y

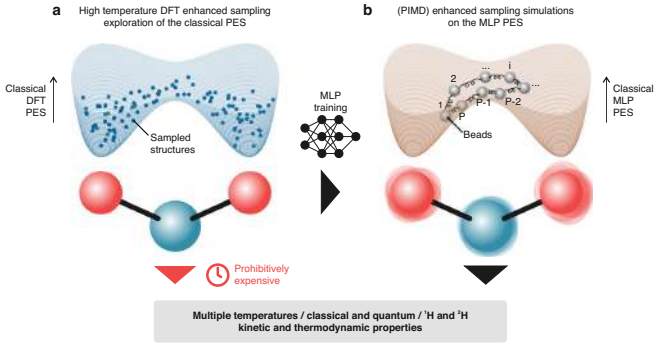


Fig. 2 | Kinetic and thermodynamic properties with and without NQEs over a wide range of temperatures can be computed using an MLP trained on high-temperature enhanced-sampling DFT simulations. The pictures show a simplified schematic representation of the PES experienced by the proton when hopping between two oxygens around the Al site (red and blue spheres, respectively), sampled with DFT simulations (**a**) and subsequently learned with the MLP (**b**). In **b** a schematic view of the hydrogen ring polymer with P beads running on the classical MLP PES is shown.

constants beyond the Transition State Theory (TST) approximation, taking barrier recrossing into account via the reactive flux formalism[27] and (iv) perform an additional set of simulations with deuterium instead of protium to explicitly compute the Kinetic Isotope Effect (KIE) on the reaction.

We show that even at catalytically relevant temperatures (>400 K) NQEs may still be important to consider when computing reaction kinetics and their relevance is not restricted to the absolute low temperature regime. While the work performed here is illustrative for the most basic proton hopping reaction in zeolites, it provides the means to routinely include NQEs and explicitly calculate KIEs when studying any proton-transfer event in heterogeneous catalysis.

Results

Construction of a reactive MLP with DFT accuracy

To train an accurate MLP, a sufficiently large set of DFT datapoints is required, which should cover the relevant regions of the reaction phase space. To this end, high-temperature (873 K) DFT US simulations were performed on a CHA conventional cell containing 36 T Si atoms, where 1 silicon is replaced by Al to give a final Si/Al ratio of 35. The temperature choice of 873 K is arbitrary but, in general, on the higher end of typical zeolite-catalyzed processes[1]. In the CHA topology, all T atoms are equivalent. However, the four O atoms in the first coordination sphere of the Al defect are not (Fig. 1c). This leads to 6 distinct hopping paths, which are all considered in this work. To assess whether any path could be significantly disfavored, activation free energies were initially screened with static calculations. The results suggest that all 6 possible hopping paths have relatively similar activation free energies (within ~20 kJ mol^{-1}) and no single one is strongly (dis)favored (Supplementary Note 2). Therefore, 6 separate DFT US simulations at 873 K were performed to sample all the possible hoppings. A difference in coordination numbers between the proton and the two oxygen atoms involved in the hopping was used as main collective variable to bias the system (see "Methods" section and Supplementary Note 3). One-dimensional umbrellas were used to sample the reaction path and, if needed, additional two-dimensional umbrellas were added to improve the sampling of scarcely visited regions of the phase space (more details are reported in Supplementary Note 3.2). A full overview of the DFT US results is reported in Supplementary Note 4.

Energies and forces were extracted every 5 fs from the DFT US trajectories, yielding a total of ~1,200,000 structures which were used to train an MLP with the SchNetPack package (see "Methods" section and Fig. 2)[28,29]. Performing MD simulations with the MLP provides a dramatic speedup in computational time, going from ~8.3 s/step on 56 Xeon E5-2680v4 CPUs@2.4 GHz cores to ~0.01 s/step on a single NVIDIA Volta V100 GPU. As part of the MLP validation, well-converged 873 K DFT FESs were generated to compare them with the MLP-derived ones within a reasonably small uncertainty. To this end, about 50 ps of simulation time was required for each DFT umbrella. Considering that 19 umbrellas are needed to sample each of the 6 hoppings, this was a computationally demanding task. On the other hand, it also provided us with a very large number of DFT datapoints, hence the large number of structures used to train the MLP. With the acquired knowledge that a mean absolute error on the force of about 40 meV Å^{-1} is sufficient to obtain very accurate FESs (vide infra), we also tested the performance of newer and more data efficient equivariant neural networks[30], where preliminary results indicate that a few hundred fs per umbrella are sufficient to achieve converged results (Supplementary Note 10), providing an enormous computational saving in the DFT data generation.

To further validate the trained MLP, we also tested whether it could reproduce FESs at lower temperatures than the training one. To this end, three additional sets of DFT US simulations were performed. The 2–3 hopping was tested at 573 and 273 K, while the 1–4 hopping was tested at 273 K. With this choice, the hoppings with the smallest activation energies are tested and all four oxygens are considered at the lowest temperature. For the sake of clarity, a detailed comparison between the MLP and DFT results is presented in Supplementary Note 6.1, while here only the 2–3 hopping is discussed in detail. As shown in Fig. 3, the DFT and MLP FESs exhibit an almost perfect overlap, with most variations contained well within the error bars. The free energy barrier exhibits a clear increase with temperature, which is in line with a rigid transition state associated with a negative entropy variation. It must be pointed out that, thanks to the large computational speedup enabled by the MLP, longer simulation times (100 ps vs. 50 ps per umbrella) and a larger number of umbrellas (39 vs. 19 per hopping) were easily achievable. This led to a vastly improved sampling of the reaction PES, thereby obtaining much better converged FESs. Moreover, all MLP simulations were repeated three separate times starting from different initial velocities and the associated results and uncertainties were obtained by averaging over these three independent runs. Initially, the 573 K DFT profile presented a moderate spike in the transition state region, which was not present in the MLP profile. Therefore, 2 additional umbrellas were added in the proximity of the transition state and an additional 40 ps of simulation was performed in every umbrella, for a total of 90 ps. The final DFT profile

Article https://doi.org/10.1038/s41467-023-36666-y

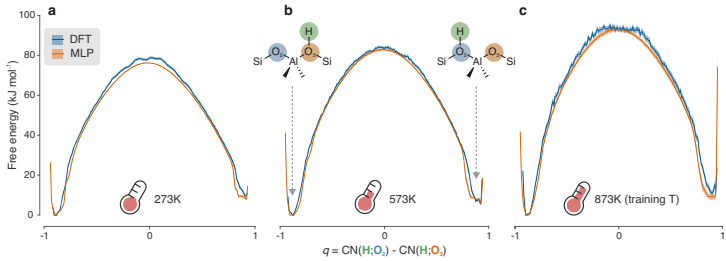


Fig. 3 | The trained MLP can reproduce DFT FESs with high accuracy. The MLP and DFT FESs (in orange and blue, respectively) of the 2–3 hopping at 273 (**a**), 573 (**b**) and 873 K (**c**) are almost perfectly superimposable, even though the MLP model is only trained on snapshots at the highest temperature. A comparison with the other available FESs is shown in Supplementary Fig. 10. For a more detailed definition of the collective variable (q) used to build the FESs, the interested reader is referred to the "Methods" section. Uncertainties (shaded areas around the FESs) on the MLP FESs are obtained by averaging over three independent runs (Supplementary Note 6.1) while for DFT they are estimated according to the procedure summarized in Supplementary Note 3.1.3. Source data are provided as Source Data File.

reaches almost perfect agreement with the MLP one, highlighting how the (small) differences between MLP and DFT FESs are almost certainly caused by sampling issues rather than by significant deviations in the underlying PES. The results show that (i) the trained MLP is effectively capable of encoding chemical reactivity and (ii) high accuracy on the computed FESs is retained also for temperatures lower than the training one, offering thermodynamic transferability in terms of operating conditions.

While directly superimposing FESs provides an intuitive visual means of comparison, the FES itself is not experimentally measurable. The final macroscopic quantity of interest is the kinetic constant of the reaction, which does not depend on the choice of the collective variable used to represent the FES[31,32]. By means of classical (TST), the forward and backward kinetic constants for the 6 high-temperature hoppings, the 2–3 hopping at 573 K and the 1–4 and 2–3 hoppings at 273 K were retrieved (see "Methods" section). Fig 4 reports a graphical comparison between the DFT and MLP rates, where the computed kinetic constants are converted to a corresponding phenomenological barrier using Eyring's equation[32]. The sole purpose of the latter is to provide a more tangible equivalent to the kinetic constant, without comparing values that can span multiple orders of magnitude (more details are provided in the "Methods" section). None of the computed barriers differ more than ~5 kJ mol^{-1} and, for most of the hoppings, the MLP values lie within the error bars of the DFT ones. These results indicate that the MLP accurately reproduces the DFT PES underlying the proton hopping reaction in H-CHA and can therefore be used to compute reaction rates at any temperature of interest and to explicitly introduce NQEs through the PIMD approach (Fig. 2).

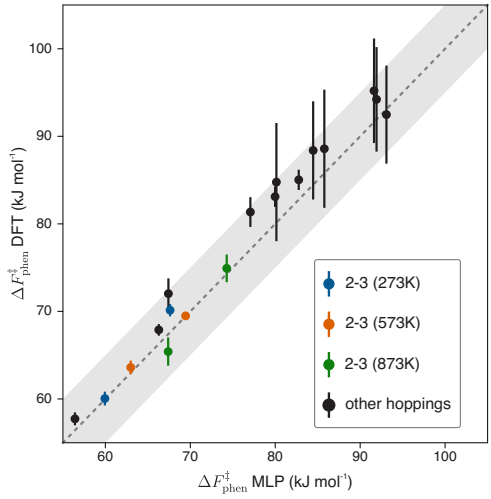


Fig. 4 | The DFT and MLP phenomenological barriers are in excellent agreement. The available DFT phenomenological barriers (12 forward and backward barriers for the 873 K hoppings, the forward and backward barriers of the 2–3 hopping at 573 K and the forward and backward barrier of the 1–4 and 2–3 hoppings at 273 K) are all within about ±5 kJ mol^{-1} from the MLP ones, as highlighted by the gray shaded area. The phenomenological barriers of the 2–3 hopping at various temperatures, corresponding with the free energy profiles of Fig. 3, are highlighted in different colors (273 K in blue, 573 K in orange and 873 K in green). Uncertainties on the MLP barriers are obtained by averaging over three independent runs (although their magnitude is so low that they are barely visible in the figure) while for DFT they are estimated according to the procedure summarized in Supplementary Note 3.3. Source data are provided as Source Data File.

Full characterization of the hopping kinetics

Having validated the MLP to faithfully reproduce the proton hopping FESs over a broad temperature range (273-873 K), additional US simulations were performed to retrieve the full reaction kinetics considering all hopping paths. Moreover, NQEs can be systematically included in the reaction investigation as the PIMD formalism becomes accessible, thanks to the large computational efficiency of the MLP. To obtain well-converged FESs, at least 16 system replicas (also referred to as beads) are required in the ring polymer (Supplementary Note 7.1). A graphical visualization of the spread of the beads around the transition state region compared with the classical case is shown in Fig. 5, where it becomes clear that quite some uncertainty is present on the proton's position compared to the classical deterministic trajectories. A full overview of the classical and quantum FESs is reported in Supplementary Notes 6.1 and 7.3, respectively. Introducing NQEs leads to a general decrease of the free energy barriers compared to the case where nuclei are treated classically. This effect tends to lessen with increasing temperatures, in accordance with the expected convergence between the quantum and classical behavior for high temperatures.

Not only does the MLP allow to include in a reliable—yet computationally feasible—way NQEs, but it also allows to determine reaction kinetics beyond classical TST and explicitly include barrier recrossing through the reactive flux formalism (obtaining the true

Article https://doi.org/10.1038/s41467-023-36666-y

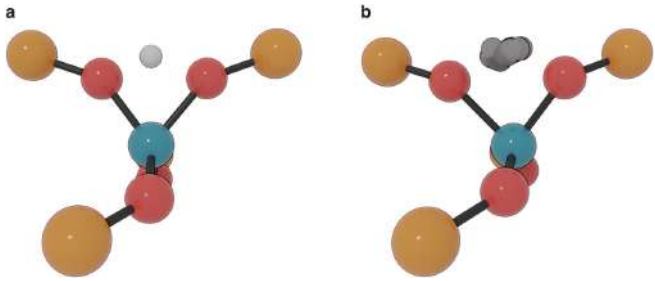


Fig. 5 | The use of PIMD leads to a significant spread in the proton location. These two snapshots, arbitrarily extracted from the transition state umbrella of the 2–3 hopping, highlight how the system beads in PIMD (**b**) can be spread quite significantly in space with respect to the classical case (**a**). For the sake of clarity, only the H-SSZ-13 atoms up to the second coordination sphere around the Al site are shown and in **b** a superposition of all beads is only present for the proton. Si is depicted in yellow, O in red, Al in blue and H in white.

kinetic constant of the reaction)[27,33], which is in its turn part of the Bennett-Chandler reaction rate theory[34,35]. In this approach, multiple unbiased simulations (5000 in this case) are initialized atop the transition state and monitored through time, to retrieve whether they end up in the product basin or whether they recross the barrier towards the reactant basin (see "Methods" section). This approach is most appropriate when NQEs are included, as quantum TST approximations such as ring polymer molecular dynamics (RPMD) TST do not yield a strict upper bound for the quantum rate (more details can be found in Supplementary Note 7.4)[36]. Overall, the MLP US simulations allowed to compute three different kinetic constants for all hopping paths and all temperatures: a classical TST-based one (k^{c}_{TST}), derived from classical MD and the TST approximation, a classical Bennett-Chandler one (k^{c}_{BC}), where barrier recrossing is now explicitly taken into account and, finally, a quantum Bennett-Chandler one (k^{q}_{BC}), analogous to k^{c}_{BC} but derived from the RPMD simulations and thus including NQEs. Remark that the amount of data used to obtain them is well beyond the reach of any pure ab initio methodology where all energy and force evaluations are performed at the DFT level. Even when excluding the thousands of short trajectories required to obtain well-converged k_{BC} values, computing the quantum FESs requires 42 sets of US simulations (6 hopping paths at 7 different temperatures), each consisting of 39 umbrellas simulated with 16 parallel beads—for a total of more than 0.6 μs of simulation time. Such simulation lengths are clearly beyond the limit of any DFT-based MD simulation.

While demonstrating the impact of NQEs on the reaction rate is important to highlight the cases in which NQEs cannot be neglected and should thus be accounted for computationally, the resulting 'quantum speedup' is not experimentally measurable as NQEs are an intrinsic part of nature. What is often measured experimentally, on the other hand, is the KIE—namely the change in rate when the hydrogen atoms in the system are substituted with deuterium (other isotopic substitutions are of course also possible[37]). Interestingly, the MLP trained on ^{1}H simulations can be directly used for KIE calculations, as the reaction PES does not depend on the atomic masses but only on their charge. An additional set of PIMD simulations was therefore performed at 273, 573 and 873 K with the BAS mass set to 2 a.m.u. Given the linear behavior of $\ln(k^{q}_{BC})$ over the whole temperature range (vide infra, Fig. 7), the intermediate temperatures were no longer considered. A full overview of the simulations' results is reported in Supplementary Note 7.5. The error on the MLP forces with respect to DFT remains basically unaffected by the change in the hydrogen mass, confirming that both simulations sample analogous PES regions (Supplementary Fig. 20). The reactive flux kinetic constant for the PIMD simulations with deuterium will be indicated with $k^{q}_{BC}(^{2}H)$, while for protium simulations the isotope label will be omitted.

Using the computed kinetic constants, the equilibrium coverage of the 4 oxygen sites (θ_i, $i = 1-4$) was determined as a function of temperature (more details can be found in Supplementary Note 8). The results for ^{1}H are shown in Fig. 6a–c. When considering the k^{c}_{TST} and k^{c}_{BC} kinetic constants, similar equilibrium populations are obtained within the limits of uncertainty (Fig. 6a, b), which is a consequence of the similar recrossing rate between the forward and backward barriers. In general, O_1 and O_3 are the most populated sites at any temperature, followed by O_2 and O_4. In the classical case, O_3 has the largest population up to 373 K, while at higher temperatures its population becomes nearly identical to O_1. Significantly different results are obtained when the quantum kinetic constants (k^{q}_{BC}, Fig. 6c) are considered, where θ_3 remains significantly larger than θ_1 at all temperatures. A similar trend is obtained for the PIMD simulations with deuterium (Supplementary Fig. 19). When the proton is on O_3, it finds itself oriented towards the center of the 6 T atoms ring (Fig. 6d) and can, therefore, interact with the oxygens on the opposite side. To understand more profoundly the impact of these intra-framework interactions, we performed a 273 K classical MD and PIMD simulations of the zeolite with the proton located on O_3. We then analyzed the radial distribution functions (RDFs) of the proton with the 6 oxygens sharing the Si and Al with O_3 ('adjacent' in Fig. 6e) and all the other oxygens in the unit cell ('others'). We found that—as expected— the BAS lies within 2–3 Å from the oxygens on the other side of the 6-Si ring and will therefore interact with them. By comparing the classical and quantum RDFs (Fig. 6e), it can be seen how the maximum in the RDF H-O (others) occurs at slightly shorter distances in the quantum case and, moreover, shorter distances—in the order of ~2 Å—are explored more often. Based on these findings, it appears that when the quantum nature of the hydrogen nucleus is considered the weak interaction between the non-adjacent framework oxygens and the proton becomes stronger and, as a result, O_3 becomes further stabilized with an increase in θ_3. No other site interacts with other framework oxygens within 3 Å (Supplementary Fig. 18). Previous reports in the literature, based on geometrical considerations concerning the crystallographic zeolite unit cell, suggested that none of the four BAS locations are suited to form H-bonds with other oxygens in the framework[38]. This no longer seems to be the case when temperature effects and NQEs are explicitly taken into account. The number of zeolite frameworks presenting this type of intra-framework interaction could thus be higher than previously thought[38] at realistic operando conditions.

Once all the equilibrium coverages as a function of temperature are known, the overall hopping rate can be computed using the

Article https://doi.org/10.1038/s41467-023-36666-y

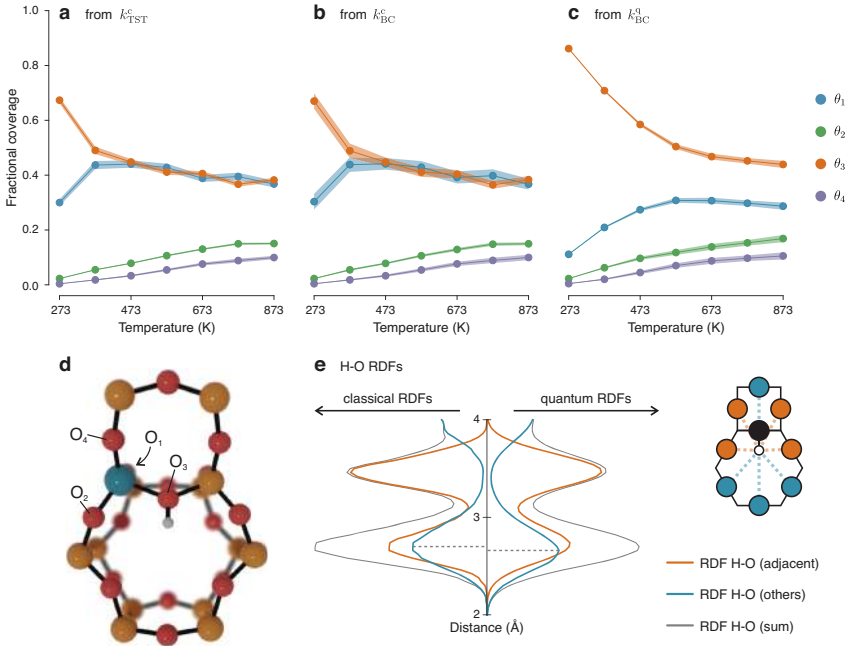


Fig. 6 | Including NQEs can change the computed equilibrium proton coverages. a–c Equilibrium coverages (θ_i) of the 4 oxygen sites around the Al defect as function of temperature, computed from the k^c_{TST}, k^c_{BC} and k^q_{BC} kinetic constants. **d** Portion of the H-CHA framework as seen along the c cell vector, showing the optimized structure of the BAS on O_3. Si is depicted in yellow, O in red, Al in blue and H in white. **e** RDFs between the BAS lying on O_3 and the 6 oxygens sharing the same Si and Al as O_3 (orange line) and the BAS and all other oxygens in the unit cell (blue line), from classical and NQE simulations. The schematic drawing is seen from the same perspective as d and highlights some of the oxygens belonging to 'adjacent' and 'others'. Source data are provided as Source Data File.

formula:

$$r = \sum_{i=1}^{4} \sum_{j \neq i}^{4} k_{ij} \theta_i, \quad (1)$$

where k_{ij} is the kinetic constant of the hopping from O_i to O_j and θ_i the coverage of O_i. From this, an Arrhenius plot for the hopping rate as a function of the temperature is computed (Fig. 7a), whose activation energy should be comparable with experiment. First, we analyzed in how far each of the hoppings is contributing to the overall rate. In all cases, only two hopping paths dominate the rate kinetics (Fig. 7b), namely the $1 \leftrightarrow 4$ and $2 \leftrightarrow 3$ paths, as one could expect based on their low free energy barriers. Note that the forward and backward rates have similar contributions, as a higher free energy of the minimum corresponds to both a lower coverage and a lower free energy barrier to hop towards a stabler minimum. These two factors tend to cancel each other when computing $k_{ij}\theta_i$. Minor contributions are given by the $1 \leftrightarrow 2$ and $2 \leftrightarrow 4$ paths, while the remaining two paths ($1 \leftrightarrow 3$ and $3 \leftrightarrow 4$) only have noticeable contributions at the highest temperatures. In the deuterium case, the $2 \leftrightarrow 3$ path becomes even more dominant at the expenses of $1 \leftrightarrow 4$ as it appears that the transition state energy is not shifted consistently by the isotope substitution (Supplementary Fig. 18).

By considering the slope of the best fit lines in the Arrhenius plots (Fig. 7a) it is possible to retrieve an effective activation energy for the proton hopping reaction. The k^c_{TST} results yield an activation energy of 67.1 kJ mol^{-1}. Going beyond the TST approximation and explicitly including recrossing (k^c_{BC}) does not significantly change the results, with a consistent—but almost negligible—decrease in the rate across the whole temperature range. When NQEs are included (k^q_{BC}), in contrast, the activation energy decreases with about 11 kJ mol^{-1} due to the possibility of the proton to tunnel through the potential energy barriers. When analyzing the rates related to a specific hopping (Supplementary Fig. 17), it was noticed that this effect is not constant, and becomes more prominent in the hopping paths with a more sharply peaked FES around the transition state region (Supplementary Figs. 15 and 16). This is because a narrower barrier increases the probability of tunneling, which in practice means that the beads of the ring polymer are easily located on both sides of the potential energy barrier experiencing on average a lower free energy. These results show that the impact of NQEs is not systematic in nature and can therefore be challenging to capture with ad hoc corrections. Indeed, previous investigations which included NQEs through an a posteriori tunneling correction suggested that above room temperature no significant effect should be observed[12]. Our results, on the other hand, show that the reaction proceeds 65 times faster at 273 K if NQEs are included and, even at 373 K, a 16-fold increase in the computed rate is still present (Fig. 7c). At 473 K, the reaction remains 7 times faster while the speedup, as expected, tends to become negligible at higher temperatures. It appears therefore that for zeolite-catalyzed processes conducted at milder conditions, among which the ones related to biomass conversion are a predominant example[2], NQEs might be non-negligible when computing the kinetics of proton-transfer steps. A few examples where these effects might be important are the aqueous cyclohexanol

Article https://doi.org/10.1038/s41467-023-36666-y

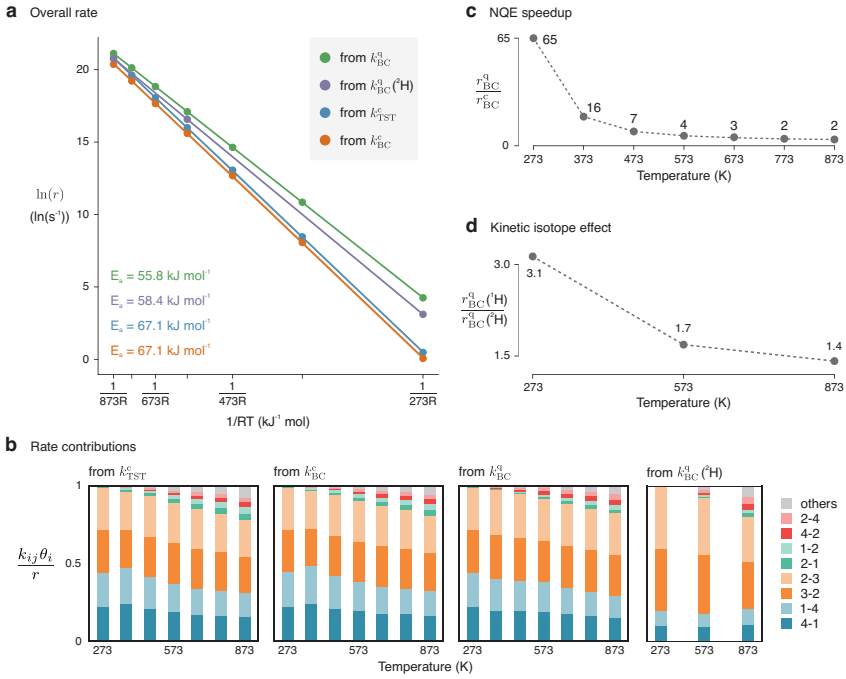


Fig. 7 | NQEs have a significant effect on the computed hopping rate. **a** Arrhenius plot showing the overall hopping rate as function of 1/RT. **b** Contributions to the overall rate from the different hopping paths, as function of temperature and methodology used. **c** Speedup in the rate due to the inclusion of the NQEs. **d** KIE as function of reaction temperature. Source data are provided as Source Data File.

dehydration in H-Beta zeolite at temperatures of about 400 K[39] and the oxidation of methane to methanol in copper-loaded zeolites carried out at ~473 K[40]. These two cases serve as examples of the relevant application area and conditions influenced by NQEs in zeolites. However, it is clear that many more cases could be affected by the inclusion of NQEs.

When protium is substituted by deuterium ($k^q_{BC}(^2H)$) the activation energy becomes 58.4 kJ mol^{-1}, yielding a small increase of 2.6 kJ mol^{-1}. This is in line with the modest magnitude of the predicted KIE (given by the ratio between the total rate with protium and with deuterium, Fig. 7d), which equals 3.1 at 273 K and decreases to 1.4 at 873 K, in line with standard Bigeleisen-Mayer KIE theory[41,42]. It can be noted how the deuterium rates remain significantly larger than the classical k^c_{BC} ones, indicating that the increase in mass is not sufficient to fully suppress the quantum behavior. Unfortunately, there has not been any experimental attempt so far to measure the KIE for proton hopping in zeolites. Theoretically, an early investigation based on static cluster calculations was performed by Fermann and Auerbach[43]. According to their semiclassical TST model, the differences between ^{1}H and ^{2}H are limited above room temperature, in line with our results.

So far, the study focused on H-CHA, which has a single indistinguishable T-site and a small unit cell. To broaden our scope, it is important to assess the MLP capability of describing other zeolite frameworks on which the MLP was not trained. To this end, the transferability of the MLP to other zeolite topologies was investigated. More specifically, we selected five all-silica frameworks from the international zeolite (IZA) database[44] that are of interest for catalytic applications[3] (AFX, CHA, FER, MFI and MOR) and performed a 100 ps NVT DFT MD simulation using the crystallographic unit cell parameters reported in the IZA database (see Supplementary Note 9). The CHA topology was included as a control system, to ensure that the MLP is robust with respect to changes in the unit cell volume. None of the MLP simulations presented obvious instabilities and the error on the forces is not excessive, even for frameworks that do not share any secondary building unit with CHA, varying between 196 meV Å^{-1} for MOR and 258 meV Å^{-1} for MFI. The quality of the zeolite trajectories, monitored through the Si–O and Si–Si RDFs, remains reasonably good with only small long-range differences for MFI (Supplementary Fig. 19). Testing the proton hopping reactivity in a systematic way for more frameworks would require a further set of expensive ab initio US simulations and, therefore, is outside the scope of this work. The results obtained on the all-silica frameworks, nonetheless, still indicate that the MLP can capture to a large extent the chemistry of Si–O–Si bonds and, therefore, we expect that not many additional DFT simulations would be needed to retrain it and extend its accuracy to new zeolite frameworks, for instance building on the transferable MLP for siliceous frameworks by Erlebach et al.[45] towards aluminum-containing zeolites of catalytic interest.

Discussion

Proton-transfer reactions are of primordial importance within zeolite catalysis. Thus far, it was unclear in how far NQEs affect the barriers and rates of proton hopping processes at realistic operating conditions, as their explicit inclusion through PIMD was prohibitively expensive if the underlying classical PES is evaluated at a DFT level of theory. Herein, we showed that a reactive MLP can be trained based on

Article https://doi.org/10.1038/s41467-023-36666-y

underlying high-temperature (873 K) US simulations at the DFT level, that provides kinetic results with a similar accuracy as the underlying DFT data. However, thanks to the enormous computational speedup gained by describing the PES based on the MLP compared to the original DFT energy and force calculations, the MLP can be used to perform virtually any type of simulation that relies on the classical PES of the considered reaction(s) over a broad range of temperatures. The proposed methodology thus not only succeeds in reproducing the underlying DFT simulations but comes with a series of advantages that were so far unreachable due to the prohibitively excessive computational cost.

First, the convergence of the free energy surfaces obtained from enhanced-sampling techniques can be improved by using many more umbrellas and by simulating for a longer time. Secondly, PIMD can be employed to explicitly account for the quantum nature of the nuclei in the system. While the inclusion of NQEs through MLPs has already been proposed in the literature[22,46,47], the application of PIMD/MLP to an activated event in heterogenous catalysis was still unexplored. We remark that for simulations at cryogenic temperatures the number of beads required to achieve converged results could become very large even for the MLP. This problem can be mitigated by coupling the MLP simulations with path integral acceleration techniques[20]. Thirdly, it also becomes possible to go beyond the TST approximation and explicitly include barrier recrossing via the reactive flux formalism, thereby obtaining the true kinetic constant of the reaction. Because of the thousands of short MD trajectories that have to be initialized atop the transition state, this type of calculation was so far too expensive to be performed at a DFT level of theory. The more efficient methodology for describing the forces and energies may also open the window to use methods like transition path sampling within the field of zeolite catalysis, which were thus far not truly accessible due to the large number of paths that needs to be sampled at the DFT level[48]. Finally, KIEs can be explicitly computed if the PIMD simulations are performed with different nuclear masses, as this does not affect the underlying PES learned by the MLP.

Our results show that the expected Arrhenius activation energy for the hopping process, considering all six hoppings and the coverages of the four oxygen sites, is 67.1 kJ mol^{-1} in the absence of NQEs, whereas including the quantum nature of the proton brings the activation energy down to 55.8 kJ mol^{-1}. When quantitively comparing this activation energy to experimental results, it is important to note that this study makes use of the revPBE-D3 level of theory, which is known to underestimate the activation energies of chemical reactions[16,49,50]. In this sense, our barriers will present a lower boundary for the chemically accurate activation energy. Because of the large improvement in data efficiency of newer MLP architectures (Section S10 of the Supplementary Information), we believe that training an accurate model based on a more expensive albeit more reliable exchange-correlation functionals should become feasible. The computed activation energy remains relatively higher than the experimentally available ones. The most likely source of discrepancy lies in the perfect crystalline nature of the adopted zeolite model. The presence of residual water molecules, defects (EFAL species, for instance) and an heterogeneous aluminum distribution are basically unavoidable at the macroscale and all these factors are known to potentially affect the behavior of protons in zeolites[17,19,51]. According to the simulations, a primary KIE of about 3 is expected at 273 K but no experimental evidence is available thus far to corroborate this result.

This proof-of-concept study presents a general scheme to obtain MLP models that can simulate proton hoppings and activated processes in zeolite catalysis with improved realism. The proposed methodology is, in principle, extendible to additional reactions and reactive environments, making it a valuable tool for studying a wide range of catalytic phenomena[52].

Methods

Umbrella sampling simulations

The hopping of the H-CHA BAS between the oxygens in the first coordination sphere of the Al defect was studied by means of umbrella sampling simulations[53,54]. In this approach, quadratic bias potentials (the 'umbrellas') are placed along a certain collective variable (q) which should smoothly vary between reactants and products. The bias has the form $V_i(q) = 1/2K_i(q - q_{0,i})^2$, where K_i is the force constant of the ith umbrella and $q_{0,i}$ its center. An MD simulation is then performed within each umbrella. To study the proton hopping, the chosen collective variable is a difference of coordination numbers (CNs) between the BAS and the two oxygens involved in the hopping:

$$q = \mathrm{CN}(\mathrm{O}_i;\mathrm{H}) - \mathrm{CN}\left(\mathrm{O}_j;\mathrm{H}\right) = \frac{1-\left(\frac{r_{\mathrm{O}_i\mathrm{H}}}{r_0}\right)^N}{1-\left(\frac{r_{\mathrm{O}_i\mathrm{H}}}{r_0}\right)^{2N}} - \frac{1-\left(\frac{r_{\mathrm{O}_j\mathrm{H}}}{r_0}\right)^N}{1-\left(\frac{r_{\mathrm{O}_j\mathrm{H}}}{r_0}\right)^{2N}} \quad (2)$$

The specific values of the N and r_0 parameters were adapted based on the reaction conditions, more information can be found in Supplementary Note 5.2. The bias potential was applied using PLUMED[55,56] and the final statistical analysis of the data was performed with our in-house developed ThermoLIB library[57]. For some of the hoppings, additional wall potentials were required to prevent undesired side reactions; further details are reported in Supplementary Note 3.1.4

DFT molecular dynamics

To perform the DFT MD simulations, the CP2K software package (version 7.1)[58,59] was employed to compute energies and forces at a revPBE-D3/TZVP[60–62] level of theory. Because of the mixed plane waves −atom-centered orbitals approach[63] used by CP2K, the plane waves energy cutoff was set to 350 Ry and GTH pseudopotentials[64] were used to smooth the electron density in the proximity of the nuclei. A significant dependency of the forces on the plane waves cutoff was found, but this was shown to have a negligible impact on the final FESs when much higher settings are used (Supplementary Note 3.1.5). The time step for the integration of the equations of motion was set to 0.5 fs. After equilibration of the unit cell (Supplementary Note 3.3.1), production runs were performed in the NVT ensemble using a Nosé-Hoover thermostat with a chain consisting of five beads[65,66] to control the temperature and a time constant of 334 fs (100 cm^{-1}).

MLP training and usage

A SchNet MLP was trained with the SchNetPack package on the DFT energies and forces which were extracted every 5 fs from the DFT US simulations at 873 K[28,29]. First, the energies and forces were unbiased by subtracting the bias potential applied in the US simulations with PLUMED[55,56]. The unbiased DFT datapoints were randomly divided in a training and validation set with a 80:20 ratio. Subsequently, the MLP was trained with a cutoff of 6 Å, 128 features, 50 gaussians and 6 interaction blocks. The resulting MAE on the validation set is 41.9 meV/Å. More details on the training are provided in Supplementary Note 5.1. Classical unbiased and US simulations with the trained MLP were performed with our in-house code YAFF[67] using a time step of 0.5 fs and a Nosé-Hoover thermostat with three beads for temperature control[65,66]. PIMD simulations were performed with the i-PI driver[68] using a time step of 0.25 fs and a PILE thermostat[69] with a time constant of 100 fs for temperature control. Because of the harmonic repulsion between the beads, some of them might explore regions of the phase space that are not necessarily well-sampled in classical DFT US. Therefore, we also performed an extra DFT PIMD US simulation for the 2–3 hopping (Supplementary Note 7.2) and the resulting FES shows an excellent agreement with the MLP one. It is important to remark that this agreement is very likely not generalizable to other systems or reactions and should always be tested appropriately[70]. In both the classical and PIMD US simulations, PLUMED was used to apply the bias.

Article https://doi.org/10.1038/s41467-023-36666-y

Kinetic rate constant calculation

The plain activation free energy obtained from a FES is largely dependent on the choice of collective variable[31,32]. To remove such dependency, it is necessary to move towards a more general macroscopic property of the process under study, namely the kinetic rate constant. In the Bennett-Chandler approach to transition state theory[34,35], the rate constant of a reaction can be written as[71]:

$$k_{\mathrm{BC}}(t) = \langle \dot{q}(0)\theta(q(t) - q^*)\rangle_{q(0)=q^*} \frac{e^{-\beta F(q^*)}}{\int_{-\infty}^{q^*} e^{-\beta F(q)} \mathrm{d}q}, \quad (3)$$

where the first term is the ensemble average of the time derivatives of q for trajectories that, starting atop the transition state ($q(0) = q^*$), end up in the product basin (as imposed by the Heavyside function $\theta(q(t) - q^*)$). With the MLP, it is possible to explicitly compute the first term by performing a large number of unbiased MD simulations (5000 in this case) starting on the transition state and monitor how many of them effectively end up in the product basin[27,33]. The true rate constant is, in principle, given by $k_{\mathrm{BC}} = \lim_{t \to +\infty} k_{\mathrm{BC}}(t)$. Luckily, its value quickly reaches a plateau and 50 fs of simulation were sufficient to obtain well-converged results (Supplementary Note 6.2). The rate constant calculated in this manner is referred to as the Bennett-Chandler one (k_{BC}).

In general, this approach is too expensive, especially for the DFT case, so that only the approximate transition state theory constant ($k_{\mathrm{TST}} = \lim_{t \to 0^+} k_{\mathrm{BC}}(t)$) can be computed from the US trajectories, thereby avoiding the need for additional simulations. While k_{TST} represents an upper limit of the true kinetic constant, assuming a recrossing probability equal to zero, it can be used to compare the DFT and MLP results. Further details are reported in Supplementary Note 3.3. To calculate the quantum rate constants, taking NQEs into account, the approximate technique of RPMD was used (see Supplementary Note 7.4)[72]. Although this approximation can only capture short-time quantum effects, it has been shown to yield good quantum rates in comparison with other approximations[73] or quantum mechanical calculations[74].

As the kinetic constant values can span several orders of magnitude, we often make use of Eyring's equation to convert them into phenomenological barriers, which encode the same information while being – in our opinion – more tangible than a reaction rate.

$$\Delta F^{\ddagger}_{\mathrm{phen}} = -\frac{1}{\beta} \ln(\beta h k) \quad (4)$$

Data availability

The complete training set, examples of input files, processing scripts and the trained MLP have been deposited in the Zenodo database (https://zenodo.org/record/7267913#.Y2U8tHbMK3A). Any additional data is available from the authors upon request. An extended discussion of the results can be found in the Supplementary Information. Source data are provided with this paper.

Code availability

CP2K (https://github.com/cp2k/cp2k), PLUMED (https://github.com/plumed/plumed2), SchNetPack (https://github.com/atomistic-machine-learning/schnetpack) and YAFF (https://github.com/molmod/yaff) are all open source and freely available at the provided links. ThermoLIB is available upon request at https://molmod.ugent.be/software/thermolib.

References

1. Weissermel, K. & Arpe, H.-J. *Industrial Organic Chemistry* (Wiley-VCH Verlag GmbH, Weinheim, Germany, 2003).
2. Ennaert, T. et al. Potential and challenges of zeolite chemistry in the catalytic conversion of biomass. *Chem. Soc. Rev.* **45**, 584–611 (2016).
3. Van Speybroeck, V. et al. Advances in theory and their application within the field of zeolite chemistry. *Chem. Soc. Rev.* **44**, 7044–7111 (2015).
4. Baba, T., Inoue, Y., Shoji, H., Uematsu, T. & Ono, Y. Temperature-dependent lineshape of 1H magic-angle spinning nuclear magnetic resonance spectra of acidic hydroxyl groups in zeolites. *Microporous Mater.* **3**, 647–655 (1995).
5. Sarv, P., Tuherm, T., Lippmaa, E., Keskinen, K. & Root, A. Mobility of the acidic proton in Brønsted sites of H-Y, H-mordenite, and H-ZSM-5 zeolites, studied by high-temperature 1H MAS NMR. *J. Phys. Chem.* **99**, 13763–13768 (1995).
6. Baba, T., Komatsu, N. & Ono, Y. Mobility of the acidic protons in H-ZSM-5 as studied by variable temperature ^{1}H MAS NMR. *J. Phys. Chem. B* **102**, 804–808 (1998).
7. Franke, M. E. & Simon, U. Proton mobility in H-ZSM5 studied by impedance spectroscopy. *Solid State Ion.* **118**, 311–316 (1999).
8. Osuga, R., Yokoi, T., Doitomi, K., Hirao, H. & Kondo, J. N. Infrared investigation of dynamic behavior of Brønsted acid sites on zeolites at high temperatures. *J. Phys. Chem. C.* **121**, 25411–25420 (2017).
9. Fermann, J. T., Blanco, C. & Auerbach, S. Modeling proton mobility in acidic zeolite clusters. I. Convergence of transition state parameters from quantum chemistry. *J. Chem. Phys.* **112**, 6779–6786 (2000).
10. Ryder, J. A., Chakraborty, A. K. & Bell, A. T. Density functional theory study of proton mobility in zeolites: Proton migration and hydrogen exchange in ZSM-5. *J. Phys. Chem. B* **104**, 6998–7011 (2000).
11. Sierka, M. & Sauer, J. Finding transition structures in extended systems: a strategy based on a combined quantum mechanics-empirical valence bond approach. *J. Chem. Phys.* **112**, 6983–6996 (2000).
12. Sierka, M. & Sauer, J. Proton mobility in chabazite, faujasite, and ZSM-5 zeolite catalysts. Comparison based on ab initio calculations. *J. Phys. Chem. B* **105**, 1603–1613 (2001).
13. Tuma, C. & Sauer, J. A hybrid MP2/planewave-DFT scheme for large chemical systems: proton jumps in zeolites. *Chem. Phys. Lett.* **387**, 388–394 (2004).
14. Wang, Y. et al. Density functional theory study of proton hopping in MCM-22 zeolite. *Chem. Phys. Lett.* **388**, 363–366 (2004).
15. Franke, M. E., Sierka, M., Simon, U. & Sauer, J. Translational proton motion in zeolite H-ZSM-5. Energy barriers and jump rates from DFT calculations. *Phys. Chem. Chem. Phys.* **4**, 5207–5216 (2002).
16. Bučko, T., Gešvandtnerová, M. & Rocca, D. Ab initio calculations of free energy of activation at multiple electronic structure levels made affordable: an effective combination of perturbation theory and machine learning. *J. Chem. Theory Comput.* **16**, 6049–6060 (2020).
17. Kester, P. M., Crum, J. T., Li, S., Schneider, W. F. & Gounder, R. Effects of Brønsted acid site proximity in chabazite zeolites on OH infrared spectra and protolytic propane cracking kinetics. *J. Catal.* **395**, 210–226 (2021).
18. Huo, H., Peng, L. & Grey, C. P. Low temperature 1H MAS NMR spectroscopy studies of proton motion in zeolite HZSM-5. *J. Phys. Chem. C.* **113**, 8211–8219 (2009).
19. Li, G. & Pidko, E. A. The nature and catalytic function of cation sites in zeolites: a computational perspective. *ChemCatChem* **11**, 134–156 (2019).
20. Markland, T. E. & Ceriotti, M. Nuclear quantum effects enter the mainstream. *Nat. Rev. Chem.* **2**, 0109 (2018).
21. Li, C. & Voth, G. A. Using machine learning to greatly accelerate path integral ab initio molecular dynamics. *J. Chem. Theory Comput.* **18**, 599–604 (2022).

22. Rossi, M. Progress and challenges in ab initio simulations of quantum nuclei in weakly bonded systems. *J. Chem. Phys.* **154**, 170902 (2021).
23. Sauceda, H. E., Vassilev-Galindo, V., Chmiela, S., Müller, K. R. & Tkatchenko, A. Dynamical strengthening of covalent and non-covalent molecular interactions by nuclear quantum effects at finite temperature. *Nat. Commun.* **12**, 442 (2021).
24. Ceriotti, M. et al. Nuclear quantum effects in water and aqueous systems: experiment, theory, and current challenges. *Chem. Rev.* **116**, 7529–7550 (2016).
25. Li, X. Z., Walker, B. & Michaelides, A. Quantum nature of the hydrogen bond. *Proc. Natl Acad. Sci. USA* **108**, 6369–6373 (2011).
26. Ceriotti, M., Cuny, J., Parrinello, M. & Manolopoulos, D. E. Nuclear quantum effects and hydrogen bond fluctuations in water. *Proc. Natl Acad. Sci. USA* **110**, 15591–15596 (2013).
27. Peters, B. Reactive flux. In *Reaction Rate Theory and Rare Events Simulations,* 335–362 (Elsevier Inc., Amsterdam, 2017).
28. Schütt, K. T., Sauceda, H. E., Kindermans, P. J., Tkatchenko, A. & Müller, K. R. SchNet - a deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
29. Schütt, K. T. et al. SchNetPack: a deep learning toolbox for atomistic systems. *J. Chem. Theory Comput.* **15**, 448–455 (2019).
30. Batzner, S. et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
31. Bailleul, S. et al. Ab initio enhanced sampling kinetic study on MTO ethene methylation reaction. *J. Catal.* **388**, 38–51 (2020).
32. Bučko, T., Chibani, S., Paul, J. F., Cantrel, L. & Badawi, M. Dissociative iodomethane adsorption on Ag-MOR and the formation of AgI clusters: an ab initio molecular dynamics study. *Phys. Chem. Chem. Phys.* **19**, 27530–27543 (2017).
33. Anderson, J. B. Statistical theories of chemical reactions. Distributions in the transition region. *J. Chem. Phys.* **58**, 4684–4692 (1973).
34. Bennett, C. H. Molecular dynamics and transition state theory: the simulation of infrequent events. In *Algorithms for Chemical Computations,* 63–97 (American Chemical Society, 1977).
35. Chandler, D. Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *J. Chem. Phys.* **68**, 2959–2970 (1978).
36. Hele, T. J. H. & Althorpe, S. C. Derivation of a true (t → 0+) quantum transition-state theory. I. Uniqueness and equivalence to ring-polymer molecular dynamics transition-state-theory. *J. Chem. Phys.* **138**, 084108 (2013).
37. Dale, H. J. A., Leach, A. G. & Lloyd-Jones, G. C. Heavy-atom kinetic isotope effects: primary interest or zero point? *J. Am. Chem. Soc.* **143**, 21079–21099 (2021).
38. Schroeder, C. et al. Hydrogen bond formation of Brønsted acid sites in zeolites. *Chem. Mater.* **32**, 1564–1574 (2020).
39. Vjunov, A. et al. Following solid-acid-catalyzed reactions by MAS NMR spectroscopy in liquid phase - Zeolite-catalyzed conversion of cyclohexanol in water. *Angew. Chem. Int. Ed.* **53**, 479–482 (2014).
40. Tomkins, P., Ranocchiari, M. & Van Bokhoven, J. A. direct conversion of methane to methanol under mild conditions over Cu-zeolites and beyond. *Acc. Chem. Res.* **50**, 418–425 (2017).
41. Bigeleisen, J. & Mayer, M. G. Calculation of equilibrium constants for isotopic exchange reactions. *J. Chem. Phys.* **15**, 261–267 (1947).
42. Bigeleisen, J. & Wolfsberg, M. Theoretical and experimental aspects of isotope effects in chemical kinetics. In *Advances in Chemical Physics* (ed. Prigogine, I.) 15–76 (John Wiley & Sons, Ltd, 1957).
43. Fermann, J. T. & Auerbach, S. Modeling proton mobility in acidic zeolite clusters. II. Room temperature tunneling effects from semiclassical rate theory. *J. Chem. Phys.* **112**, 6787–6794 (2000).
44. IZA Database of Zeolite Structures (accessed 28 March 2022). http://www.iza-structure.org/databases/.
45. Erlebach, A., Nachtigall, P. & Grajciar, L. Accurate large-scale simulations of siliceous zeolites by neural network potentials. *npj Comput. Mater.* **8**, 174 (2022).
46. Cendagorta, J. R., Shen, H., Bačić, Z. & Tuckerman, M. E. Enhanced sampling path integral methods using neural network potential energy surfaces with application to diffusion in hydrogen hydrates. *Adv. Theory Simul.* **4**, 2000258 (2021).
47. Hellström, M., Ceriotti, M. & Behler, J. Nuclear quantum effects in sodium hydroxide solutions from neural network molecular dynamics simulations. *J. Phys. Chem. B* **122**, 10158–10171 (2018).
48. Bolhuis, P. G., Chandler, D., Dellago, C. & Geissler, P. L. Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **53**, 291–318 (2002).
49. Sauer, J. Ab initio calculations for molecule-surface interactions with chemical accuracy. *Acc. Chem. Res.* **52**, 3502–3510 (2019).
50. Goncalves, T. J., Plessow, P. N. & Studt, F. On the accuracy of density functional theory in zeolite catalysis. *ChemCatChem* **11**, 4368–4376 (2019).
51. Liu, P. & Mei, D. Identifying free energy landscapes of proton-transfer processes between Brønsted acid sites and water clusters inside the zeolite pores. *J. Phys. Chem. C.* **124**, 22568–22576 (2020).
52. Ma, S. & Liu, Z. P. Machine learning potential era of zeolite simulation. *Chem. Sci.* **13**, 5055–5068 (2022).
53. Torrie, G. M. & Valleau, J. P. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.* **28**, 578–581 (1974).
54. Torrie, G. M. & Valleau, J. P. Monte Carlo study of a phase-separating liquid mixture by umbrella sampling. *J. Chem. Phys.* **66**, 1402–1408 (1977).
55. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: new feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).
56. The PLUMED consortium. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **16**, 667–673 (2019).
57. ThermoLIB | Center for Molecular Modeling (accessed 19 November 2021) https://molmod.ugent.be/software/thermolib.
58. Vandevondele, J. et al. Quickstep: fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **167**, 103–128 (2005).
59. Kühne, T. D. et al. CP2K: An electronic structure and molecular dynamics software package -Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **152**, 194103 (2020).
60. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
61. Yang, K., Zheng, J., Zhao, Y. & Truhlar, D. G. Tests of the RPBE, revPBE, τ -HCTHhyb, ωb97X-D, and MOHLYP density functional approximations and 29 others against representative databases for diverse bond energies and barrier heights in catalysis. *J. Chem. Phys.* **132**, 1–10 (2010).
62. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010).
63. Lippert, G., Hutter, J. & Parrinello, M. The Gaussian and augmented-plane-wave density functional method for ab initio molecular dynamics simulations. *Theor. Chem. Acc.* **103**, 124–140 (1999).
64. Goedecker, S., Teter, M. & Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B - Condens. Matter Mater. Phys.* **54**, 1703–1710 (1996).
65. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268 (1984).
66. Martyna, G. J., Klein, M. L. & Tuckerman, M. Nosé-Hoover chains: the canonical ensemble via continuous dynamics. *J. Chem. Phys.* **97**, 2635–2643 (1992).

Article https://doi.org/10.1038/s41467-023-36666-y

67. Verstraelen, T., Vanduyfhuys, L., Vandenbrande, S. & Rogge, S. M. J. Yaff, Yet Another Force Field. http://molmod.github.io/yaff/index.html.
68. Kapil, V. et al. i-PI 2.0: A universal force engine for advanced molecular simulations. *Comput. Phys. Commun.* **236**, 214–223 (2019).
69. Ceriotti, M., Parrinello, M., Markland, T. E. & Manolopoulos, D. E. Efficient stochastic thermostatting of path integral molecular dynamics. *J. Chem. Phys.* **133**, 124104 (2010).
70. Li, C., Paesani, F. & Voth, G. A. Static and dynamic correlations in water: comparison of classical ab initio molecular dynamics at elevated temperature with path integral simulations at ambient temperature. *J. Chem. Theory Comput.* **18**, 2124–2131 (2022).
71. Frenkel, D. & Smit, B. Rare events. In *Understanding Molecular Simulation* 431–464 (Academic Press, 2002).
72. Craig, I. R. & Manolopoulos, D. E. A refined ring polymer molecular dynamics theory of chemical reaction rates. *J. Chem. Phys.* **123**, 034102 (2005).
73. Voth, G. A., Chandler, D. & Miller, W. H. Rigorous formulation of quantum transition state theory and its dynamical corrections. *J. Chem. Phys.* **91**, 7749–7760 (1989).
74. Suleimanov, Y. V., Javier Aoiz, F. & Guo, H. Chemical reaction rate coefficients from ring polymer molecular dynamics: Theory and practical applications. *J. Phys. Chem. A* **120**, 8488–8502 (2016).
75. Bocus, M., Neale, S. E., Cnudde, P. & Van Speybroeck, V. Dynamic evolution of catalytic active sites within zeolite catalysis. In *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering* (Elsevier Inc., Amsterdam, 2021).

Acknowledgements

This work was supported by the Fund for Scientific Research Flanders (FWO, BioFact Excellence of Science project G0H0918N, ID EOS: 30902231, and project G024019N), the Flanders Industry Innovation Moonshot (ARCLATH II, No. HBC.2021.0254) and the Research Board of Ghent University (BOF). The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, FWO, and the Flemish Government – department EWI.

Author contributions

M.B. and V.V.S. initiated the discussion and designed the scope of the project. M.B. performed the classical DFT simulations. A.L. provided technical support for the PIMD simulations and performed the PIMD DFT simulations. Under the supervision of T.V., R.G. trained the MLP and performed the related simulations, with support from M.C.C. M.B., R.G. and A.L. analyzed the results, which were discussed among all authors. M.B. and V.V.S. wrote the manuscript, with contributions from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-023-36666-y.

Correspondence and requests for materials should be addressed to Veronique Van Speybroeck.

Peer review information *Nature Communications* thanks Donghai Mei and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

B

List of Publications

Publications in international peer–reviewed journals

1. **Modeling Gas Adsorption in Flexible Metal-Organic Frameworks via Hybrid Monte Carlo/Molecular Dynamics Schemes**
   Sven M.J. Rogge†, Ruben Goeminne†, Ruben Demuynck, Juan José Gutiérrez-Sevillano, Steven Vandenbrande, Louis Vanduyfhuys, Michel Waroquier, Toon. Verstraelen, Veronique Van Speybroeck
   Adv. Theory and Simul., **2019**, *2* (4), 1800177
   † Authors contributed equally to this work

2. **Charting the Complete Thermodynamic Landscape of Gas Adsorption for a Responsive Metal–Organic Framework**
   Ruben Goeminne, Simon Krause, Stefan Kaskel, Toon Verstraelen, Jack D. Evans
   J. Am. Chem. Soc, **2021**, *143* (11), 4143–4147

3. **Accurate Transferable Polarization Model Derived from the Monomer Electron Density**
   Ruben Goeminne, Toon Verstraelen
   Manuscript in preparation

4. **DFT-quality Adsorption Simulations in Metal-Organic Frameworks Enabled by Machine Learning Potentials**
   Ruben Goeminne, Louis Vanduyfhuys, Veronique Van Speybroeck, Toon Verstraelen
   Submitted to J. Chem. Theory Comput

5. **Accurately Determining the Phase Transition Temperature of CsPbI3 via Random-Phase Approximation Calculations and Phase-Transferable Machine Learning Potentials**
   Toon Braeckevelt, Ruben Goeminne, Sander Vandenhaute, Sander Borgmans, Toon Verstraelen, Julian A. Steele, Maarten B. J. Roeffaers, Johan Hofkens, Sven M.J. Rogge, Veronique Van Speybroeck
   Chem. Mater., **2022**, *34* (19), 8561-8576

6. **Nuclear Quantum Effects on Zeolite Proton Hopping Kinetics Explored with Machine Learning Potentials and Path Integral Molecular Dynamics**
   Massimo Bocus†, Ruben Goeminne†, Aran Lamaire, Maarten Cools-Ceuppens, Toon Verstraelen, Veronique Van Speybroeck
   Nat. Commun., **2023**, *14* (1008)
   † Authors contributed equally to this work

Conference contributions

Oral presentations

1. **The influence of nuclear quantum effects on proton hopping kinetics in the H-SSZ-13 zeolite through ab initio derived machine learning potentials**
   Massimo Bocus, Ruben Goeminne, Aran Lamaire, Maarten Cools-Ceuppens, Toon Verstraelen, and Veronique Van Speybroeck
   NCCC XXIII, Noordwijkerhout, The Netherlands, May 9 – May 11, 2022

2. **Accurately determining the transition temperature of metal halide perovskites via RPA calculations and phase-transferable machine learning potentials**

Tom Braeckevelt, Ruben Goeminne, Sander Vandenhaute, Sander Borgmans, Toon Verstraelen, Julian A. Steele, Maarten B.J. Roeffaers, Johan Hofkens, Sven M.J. Rogge and Veronique Van Speybroeck

DFT2022, Brussels, Belgium, August 29 – September 2, 2022

Invited talks

1. **Methane adsorption in DUT-49: an exploration of the osmotic landscape**

   Ruben Goeminne, Simon Krause, Stefan Kaskel, Toon Verstraelen, and Jack D. Evans

   Inorganic chemistry group, Dresden, Germany, April 1, 2021

Poster presentations

1. **Modeling gas adsorption in flexible metal-organic frameworks via hybrid Monte Carlo / molecular dynamics**

   Ruben Goeminne, Sven M.J. Rogge, Ruben Demuynck, Juan José Gutiérrez-Sevillano, Steven Vandenbrande, Louis Vanduyfhuys, Michel Waroquier, Toon Verstraelen, Veronique Van Speybroeck

   MOFSIM 2019, Ghent, Belgium, April 10–12, 2019

Master's thesis

The development of hybrid MC/MD schemes to model the adsorption of guest molecules in flexible metal-organic frameworks

Ruben Goeminne

Master's thesis performed at the Center for Molecular Modeling (CMM), Ghent University, 2017–2018

Supervisor: prof. dr. ir. Veronique Van Speybroeck and prof. dr. ir. Toon Verstraelen

Counsellors: dr. ir. Sven Rogge, prof. dr. ir. Louis Vanduyfhuys, dr. ir. Ruben Demuynck, dr. ir. Juan José Gutiérrez-Sevillano and dr. ir. Steven Vandenbrande

C

List of Software Packages

In the context of this Ph.D. dissertation, various software packages have been used extensively. Below, an overview of these software packages is provided.

CP2K

CP2K is a quantum chemistry and solid state program. It allows for DFT calculations using a mixed plane wave and atomic basis set. CP2K was adopted in this PhD to perform first-principles calculations on zeolites. More information about this software is available in Ref. [298].

Gaussian 09 and Gaussian 16

Gaussian 09 and Gaussian 16 are the latest in the Gaussian series of programs. They provide state-of-the-art capabilities for electronic structure modeling. Gaussian 09 and Gaussian 16 were adopted in this Ph.D. dissertation to calculate the *ab initio* Hessians of MOF cluster models, which were subsequently used to fit the force fields. More information about this software is available in references [244, 299].

HORTON

HORTON is a Helpful Open–source Research TOol for N–fermion systems. HORTON is not intended to be a replacement for existing quantum chemistry software, but a helpful supplement to it. HORTON was adopted in

this Ph.D. dissertation to partition the molecular charge density into atomic contributions with the MBIS scheme.[166] This partitioning of the monomer electron density lies at the basis of our polarizable force field. More information about this software is available at http://theochem.github.io/horton/.

i-PI

i-PI is a universal force engine interface written in Python, designed to be used together with an *ab initio*, machine-learned, or force field based evaluation of the interactions between the atoms. i-PI was adopted in this PhD in conjunction with SchNetPack to perform MLP-based enhanced sampling path-integral MD simulations to describe the the proton hopping reaction in H-CHA. More information about this software is available in Ref. [300].

MolMod

MolMod is a Python library with many auxiliary modules that are useful to write molecular modeling programs. It is the underlying library for many software packages developed at the Center for Molecular Modeling. MolMod was adopted in this Ph.D. dissertation as the underlying software package for Yaff. More information about this software is available at http://molmod.github.io/molmod/.

NequIP

NequIP is a Python library for building E(3)-equivariant deep neural networks for the prediction of potential energy surfaces. NequIP was adopted in this Ph.D. dissertation to train MLPs to describe the non-covalent interaction energy and forces of adsorbates in MOFs, and adapted to perform GCMC calculations. More information about this software is available in Ref. [190] and at https://github.com/mir-group/nequip.

PLUMED

PLUMED is an open-source, community-developed library that provides a wide range of different methods, including enhanced sampling algorithms. PLUMED was adopted in this Ph.D. dissertation to bias MD simulations with an umbrella bias to characterize the proton hopping reaction in H-CHA. More information about this software is available in Ref. [301] and at https://github.com/plumed/plumed2.

QuickFF

QuickFF is a software package developed at the Center for Molecular Modeling, designed to derive accurate force fields for isolated and complex molecular systems in a quick and easy manner. The force field parameters for the covalent interactions are derived from *ab initio* data, i.e., an equilibrium structure and a Hessian matrix, which need to be produced by the user. QuickFF was adopted in this Ph.D. dissertation to derive the force fields for the different MOFs. More information about this software is available under this link and in references [96, 97].

SchNetPack

SchNetPack is a Python library to function as a toolbox for the development and application of deep neural networks for the prediction of potential energy surfaces. SchNetPack was adopted in this Ph.D. dissertation to train and employ MLPs to describe the proton hopping reaction in H-CHA. It was coupled together with Yaff in order to perform enhanced sampling (classical) MLP-MD simulations, and used in conjunction with i-PI to perform enhanced sampling path-integral MLP-MD simulations. More information about this software is available in Ref. [189] and at https://github.com/atomistic-machine-learning/schnetpack.

Thermolib

ThermoLIB is a Python/Cython library to construct and manipulate the free energy surfaces (FES) as a function of a (set of) priori chosen collective variable(s) from output of molecular simulations. ThermoLIB was adopted in this Ph.D. dissertation to compute reaction rate constants for the proton hopping reaction in H-CHA. More information about this software is available at https://molmod.ugent.be/software/thermolib.

VASP

The Vienna Ab initio Simulation Package (VASP) is a program for first principles electronic structure calculations. VASP computes an approximate solution to the many-body Schrödinger equation, either within density functional theory (DFT) or within the Hartree-Fock (HF) approximation. VASP was adopted in this Ph.D. dissertation to calculate the interaction energies of adsorbates in MOFs using a plane wave basis set, used for the construction of machine learning potential (MLP) training sets. More information about this software is available in Ref. [302].

WHAM

A memory efficient implementation of the Weighted Histogram Analysis Method (WHAM) was used to compute free energy profiles from umbrella sampling simulations. WHAM was adopted in this Ph.D. dissertation to calculate these profiles for the proton hopping reaction in H-CHA. More information about this software is available at http://membrane.urmc.rochester.edu/?page_id=126.

Yaff

Yaff is a Python library developed at the Center for Molecular Modeling containing a variety of tools to impose the temperature and pressure during an MD simulation. Yaff was adopted in this Ph.D. dissertation as the simulation engine to perform all MD simulations, for both force field simulations on MOFs as well as MLP simulations on zeolites. More information about this software is available at https://github.com/molmod/yaff.

Zeo++

Zeo++ is a software package for the analysis of crystalline porous materials. Zeo++ was adopted in this Ph.D. dissertation to compute the probe-occupiable volume fraction of nanoporous materials. More information about this software is available in Ref. [303] at http://www.zeoplusplus.org/.

Bibliography

[1] F. Venditti, A. Agam, J. Tirillò, S. Nunziante-Cesaro, and R. Barkai, "An integrated study discloses chopping tools use from Late Acheulean Revadim (Israel)," *PLoS ONE*, vol. 16, p. e0245595, jan 2021.

[2] R. Corbey, A. Jagich, K. Vaesen, and M. Collard, "The acheulean handaxe: More like a bird's song than a beatles' tune?," *Evolutionary Anthropology*, vol. 25, pp. 6–19, jan 2016.

[3] É. Tartar, "Origin and Development of Aurignacian Osseous Technology in Western Europe: a Review of Current Knowledge," *Palethnologie*, dec 2015.

[4] F. W. Marlowe, "Hunter-gatherers and human evolution," *Evolutionary Anthropology*, vol. 14, pp. 54–67, apr 2005.

[5] A. Chatterjee, P. Stevenson, S. De Franceschi, A. Morello, N. P. de Leon, and F. Kuemmeth, "Semiconductor qubits in practice," mar 2021.

[6] M. Victoria, N. Baluc, and P. Spätig, "Structural materials for fusion reactors," *Nuclear Fusion*, vol. 41, pp. 1047–1053, aug 2001.

[7] R. S. Stolarski and R. J. Cicerone, "Stratospheric Chlorine: a Possible Sink for Ozone," *Canadian Journal of Chemistry*, vol. 52, no. 8, pp. 1610–1615, 1974.

[8] T. Ogawa, "Chemistry of Stratospheric Chlorine," *Journal of the Meteorological Society of Japan. Ser. II*, vol. 54, pp. 294–307, oct 1976.

[9] D. J. Griggs and M. Noguer, "Climate change 2001: The scientific basis. Contribution of working group I to the third assessment report of the intergovernmental panel on climate change," Tech. Rep. 8, 2002.

[10] J. Rissman, C. Bataille, E. Masanet, N. Aden, W. R. Morrow, N. Zhou, N. Elliott, R. Dell, N. Heeren, B. Huckestein, J. Cresko, S. A. Miller,

J. Roy, P. Fennell, B. Cremmins, T. Koch Blank, D. Hone, E. D. Williams, S. de la Rue du Can, B. Sisson, M. Williams, J. Katzenberger, D. Burtraw, G. Sethi, H. Ping, D. Danielson, H. Lu, T. Lorber, J. Dinkel, and J. Helseth, "Technologies and policies to decarbonize global industry: Review and assessment of mitigation drivers through 2070," may 2020.

[11] S. Wang, S. Yan, X. Ma, and J. Gong, "Recent advances in capture of carbon dioxide using alkali-metal-based oxides," *Energy and Environmental Science*, vol. 4, pp. 3805–3819, oct 2011.

[12] G. Cui, J. Wang, and S. Zhang, "Active chemisorption sites in functionalized ionic liquids for carbon capture," *Chemical Society Reviews*, vol. 45, pp. 4307–4339, aug 2016.

[13] B. Hu and H. Zhai, "The cost of carbon capture and storage for coal-fired power plants in China," *International Journal of Greenhouse Gas Control*, vol. 65, pp. 23–31, oct 2017.

[14] H. A. Patel, J. Byun, and C. T. Yavuz, "Carbon Dioxide Capture Adsorbents: Chemistry and Methods," *ChemSusChem*, vol. 10, pp. 1303–1317, apr 2017.

[15] E. Rivard, M. Trudeau, and K. Zaghib, "Hydrogen storage for mobility: A review," *Materials*, vol. 12, p. 1973, jun 2019.

[16] A. F. Cronstedt, "Natural zeolite and minerals," *Svenska Vetenskaps Akademiens Handlingar Stockholm*, vol. 17, p. 120, 1756.

[17] J. Weitkamp and L. Puppe, eds., *Catalysis and Zeolites*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999.

[18] R. M. Barrer, "33. Synthesis of a zeolitic mineral with chabazite-like sorptive properties," *Journal of the Chemical Society (Resumed)*, pp. 127–132, jan 1948.

[19] E. M. Flanigen, "Chapter 2 Zeolites and Molecular Sieves an Historical Perspective," *Studies in Surface Science and Catalysis*, vol. 58, pp. 13–34, jan 1991.

[20] F. N. Guerzoni and J. Abbot, "Catalytic cracking of a gippsland reduced crude on zeolite catalysts," *Journal of Catalysis*, vol. 147, no. 2, pp. 393–403, 1994.

[21] Y. N. Nayak, S. Nayak, Y. Nadaf, N. S. Shetty, and S. L. Gaonkar, "Zeolite Catalyzed Friedel-Crafts Reactions: A Review," *Letters in Organic Chemistry*, vol. 17, pp. 491–506, jul 2019.

[22] M. A. Carreon and S. R. Venna, *Metal-organic framework membranes for molecular gas separations*, vol. 06 of *Series on Chemical Engineering*. WORLD SCIENTIFIC (EUROPE), aug 2020.

[23] Z. Wang, J. Yu, and R. Xu, "Needs and trends in rational synthesis of zeolitic materials," *Chemical Society Reviews*, vol. 41, pp. 1729–1741, feb 2012.

[24] H. Li, M. Eddaoudi, T. L. Groy, and O. M. Yaghi, "Establishing microporosity in open metal-organic frameworks: Gas sorption isotherms for Zn(BDC) (BDC = 1,4-benzenedicarboxylate) [28]," *Journal of the American Chemical Society*, vol. 120, pp. 8571–8572, aug 1998.

[25] S. S. Chui, S. M. Lo, J. P. Charmant, A. G. Orpen, and I. D. Williams, "A chemically functionalizable nanoporous material [Cu3(TMA)2 (H2O)3](n)," *Science*, vol. 283, pp. 1148–1150, feb 1999.

[26] T. Loiseau, C. Serre, C. Huguenard, G. Fink, F. Taulelle, M. Henry, T. Bataille, and G. Férey, "A Rationale for the Large Breathing of the Porous Aluminum Terephthalate (MIL-53) Upon Hydration," *Chemistry - A European Journal*, vol. 10, pp. 1373–1382, mar 2004.

[27] K. S. Park, Z. Ni, A. P. Côté, J. Y. Choi, R. Huang, F. J. Uribe-Romo, H. K. Chae, M. O'Keeffe, and O. M. Yaghi, "Exceptional chemical and thermal stability of zeolitic imidazolate frameworks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, pp. 10186–10191, jul 2006.

[28] J. H. Cavka, S. Jakobsen, U. Olsbye, N. Guillou, C. Lamberti, S. Bordiga, and K. P. Lillerud, "A new zirconium inorganic building brick forming metal organic frameworks with exceptional stability," *Journal of the American Chemical Society*, vol. 130, pp. 13850–13851, oct 2008.

[29] U. Stoeck, S. Krause, V. Bon, I. Senkovska, and S. Kaskel, "A highly porous metal–organic framework, constructed from a cuboctahedral super-molecular building block, with exceptionally high methane uptake," *Chemical Communications*, vol. 48, pp. 10841–10843, oct 2012.

[30] E. A. Tomic, "Thermal stability of coordination polymers," *Journal of Applied Polymer Science*, vol. 9, pp. 3745–3752, nov 1965.

[31] B. P. Block, S. H. Rose, C. W. Schaumann, E. S. Roth, and J. Simkin, "Coordination Polymers with Inorganic Backbones Formed by Double-Bridging of Tetrahedral Elements," 1962.

[32] O. M. Yaghi, G. Li, and H. Li, "Selective binding and removal of guests in a microporous metal–organic framework," *Nature*, vol. 378, pp. 703–706, dec 1995.

[33] O. M. Yaghi, H. Li, C. Davis, D. Richardson, and T. L. Groy, "Synthetic Strategies, Structure Patterns, and Emerging Properties in the Chemistry of Modular Porous Solids," *Accounts of Chemical Research*, vol. 31, no. 8, pp. 474–484, 1998.

[34] C. J. Kepert and M. J. Rosseinsky, "Zeolite-like crystal structure of an empty microporous molecular framework," *Chemical Communications*, no. 4, pp. 375–376, 1999.

[35] H. Li, M. Eddaoudi, M. O'Keeffe, and O. M. Yaghi, "Design and synthesis of an exceptionally stable and highly porous metal- organic framework," *Nature*, vol. 402, pp. 276–279, nov 1999.

[36] Z. Ni and R. I. Masel, "Rapid production of metal-organic frameworks via microwave-assisted solvothermal synthesis," *Journal of the American Chemical Society*, vol. 128, pp. 12394–12395, sep 2006.

[37] M. Klimakow, P. Klobes, A. F. Thünemann, K. Rademann, and F. Emmerling, "Mechanochemical synthesis of metal-organic frameworks: A fast and facile approach toward quantitative yields and high specific surface areas," *Chemistry of Materials*, vol. 22, pp. 5216–5221, sep 2010.

[38] L. Chen, Y. Shen, J. Bai, and C. Wang, "Novel symmetrical coralloid Cu 3D superstructures: Solid-state synthesis from a Cu-carboxylate MOF and their in-situ thermal conversion," *Journal of Solid State Chemistry*, vol. 182, pp. 2298–2306, aug 2009.

[39] H. Al-Kutubi, J. Gascon, E. J. Sudhölter, and L. Rassaei, "Electrosynthesis of Metal-Organic Frameworks: Challenges and Opportunities," *ChemElectroChem*, vol. 2, pp. 462–474, apr 2015.

[40] N. Stock and S. Biswas, "Synthesis of metal-organic frameworks (MOFs): Routes to various MOF topologies, morphologies, and composites," feb 2012.

[41] X. Feng, T. Wu, and M. A. Carreon, "Synthesis of ZIF-67 and ZIF-8 crystals using DMSO (Dimethyl Sulfoxide) as solvent and kinetic transformation studies," *Journal of Crystal Growth*, vol. 455, pp. 152–156, dec 2016.

[42] S. Leubner, R. Stäglich, J. Franke, J. Jacobsen, J. Gosch, R. Siegel, H. Reinsch, G. Maurin, J. Senker, P. G. Yot, and N. Stock, "Solvent Impact on the Properties of Benchmark Metal–Organic Frameworks: Acetonitrile-Based Synthesis of CAU-10, Ce-UiO-66, and Al-MIL-53," *Chemistry - A European Journal*, vol. 26, pp. 3877–3883, mar 2020.

[43] K. Sumida, D. L. Rogow, J. A. Mason, T. M. McDonald, E. D. Bloch, Z. R. Herm, T. H. Bae, and J. R. Long, "Carbon dioxide capture in metal-organic frameworks," *Chemical Reviews*, vol. 112, no. 2, pp. 724–781, 2012.

[44] Y. Sun and H. C. Zhou, "Recent progress in the synthesis of metal-organic frameworks," sep 2015.

[45] Y. R. Lee, J. Kim, and W. S. Ahn, "Synthesis of metal-organic frameworks: A mini review," *Korean Journal of Chemical Engineering*, vol. 30, no. 9, pp. 1667–1680, 2013.

[46] H. Furukawa and O. M. Yaghi, "Storage of hydrogen, methane, and carbon dioxide in highly porous covalent organic frameworks for clean energy applications," *Journal of the American Chemical Society*, vol. 131, pp. 8875–8883, jul 2009.

[47] F. Millange, C. Serre, and G. Férey, "Synthesis, structure determination and properties of MIL-53as and MIL-53ht: The first CrIII hybrid inorganic-organic microporous solids: CrIII(OH)·(O2C-C6H4-CO2)·(HO2C-C6H4-CO2H)x," *Chemical Communications*, vol. 2, pp. 822–823, apr 2002.

[48] C. Serre, F. Millange, C. Thouvenot, M. Noguès, G. Marsolier, D. Louër, and G. Férey, "Very large breathing effect in the first nanoporous chromium(III)-based solids: MIL-53 or CrIII(OH)·O2C-C6H4- CO2·HO2C-C6H4 -CO2Hx·H2Oy," *Journal of the American Chemical Society*, vol. 124, pp. 13519–13526, nov 2002.

[49] R. Banerjee, A. Phan, B. Wang, C. Knobler, H. Furukawa, M. O'Keeffe, and O. M. Yaghi, "High-throughput synthesis of zeolitic imidazolate frameworks and application to CO2 capture," *Science*, vol. 319, pp. 939–943, feb 2008.

[50] A. Phan, C. J. Doonan, F. J. Uribe-Romo, C. B. Knobler, M. Okeeffe, and O. M. Yaghi, "Synthesis, structure, and carbon dioxide capture properties of zeolitic imidazolate frameworks," *Accounts of Chemical Research*, vol. 43, pp. 58–67, jan 2010.

[51] Y. Pan, Y. Liu, G. Zeng, L. Zhao, and Z. Lai, "Rapid synthesis of zeolitic imidazolate framework-8 (ZIF-8) nanocrystals in an aqueous system," *Chemical Communications*, vol. 47, pp. 2071–2073, feb 2011.

[52] Y. W. Abraha, C. W. Tsai, J. W. Niemantsverdriet, and E. H. Langner, "Optimized CO2Capture of the Zeolitic Imidazolate Framework ZIF-8 Modified by Solvent-Assisted Ligand Exchange," *ACS Omega*, vol. 6, pp. 21850–21860, aug 2021.

[53] M. Eddaoudi, J. Kim, N. Rosi, D. Vodak, J. Wachter, M. O'Keeffe, and O. M. Yaghi, "Systematic design of pore size and functionality in isoreticular MOFs and their application in methane storage," *Science*, vol. 295, pp. 469–472, jan 2002.

[54] X. Zhang, Z. Chen, X. Liu, S. L. Hanna, X. Wang, R. Taheri-Ledari, A. Maleki, P. Li, and O. K. Farha, "A historical overview of the activation and porosity of metal-organic frameworks," oct 2020.

[55] T. Wu, L. Shen, M. Luebbers, C. Hu, Q. Chen, Z. Ni, and R. I. Masel, "Enhancing the stability of metal-organic frameworks in humid air by incorporating water repellent functional groups," *Chemical Communications*, vol. 46, pp. 6120–6122, sep 2010.

[56] J. Yang, A. Grzech, F. M. Mulder, and T. J. Dingemans, "Methyl modified MOF-5: A water stable hydrogen storage material," *Chemical Communications*, vol. 47, pp. 5244–5246, apr 2011.

[57] Y. Pramudya, S. Bonakala, D. Antypov, P. M. Bhatt, A. Shkurenko, M. Eddaoudi, M. J. Rosseinsky, and M. S. Dyer, "High-throughput screening of metal-organic frameworks for kinetic separation of propane and propene," *Physical Chemistry Chemical Physics*, vol. 22, pp. 23073–23082, oct 2020.

[58] M. I. H. Mohideen, B. Xiao, P. S. Wheatley, A. C. McKinlay, Y. Li, A. M. Slawin, D. W. Aldous, N. F. Cessford, T. Düren, X. Zhao, R. Gill, K. M. Thomas, J. M. Griffin, S. E. Ashbrook, and R. E. Morris, "Protecting group and switchable pore-discriminating adsorption properties of a hydrophilic-hydrophobic metal-organic framework," *Nature Chemistry*, vol. 3, pp. 304–310, apr 2011.

[59] Z. Bao, S. Alnemrat, L. Yu, I. Vasiliev, Q. Ren, X. Lu, and S. Deng, "Kinetic separation of carbon dioxide and methane on a copper metal-organic framework," *Journal of Colloid and Interface Science*, vol. 357, pp. 504–509, may 2011.

[60] W. M. Bloch, R. Babarao, M. R. Hill, C. J. Doonan, and C. J. Sumby, "Post-synthetic structural processing in a metal-organic framework material as a mechanism for exceptional CO2/N2 selectivity," *Journal of the American Chemical Society*, vol. 135, pp. 10441–10448, jul 2013.

[61] S. Bourrelly, P. L. Llewellyn, C. Serre, F. Millange, T. Loiseau, and G. Férey, "Different adsorption behaviors of methane and carbon dioxide in the isotypic nanoporous metal terephthalates MIL-53 and MIL-47," *Journal of the American Chemical Society*, vol. 127, pp. 13519–13521, oct 2005.

[62] C. Graham, J. Pierrus, and R. E. Raab, "Measurement of the electric quadrupole moments of CO2, CO and N2," *Molecular Physics*, vol. 67, no. 4, pp. 939–955, 1989.

[63] Q. Yang, A. D. Wiersum, P. L. Llewellyn, V. Guillerm, C. Serre, and G. Maurin, "Functionalizing porous zirconium terephthalate UiO-66(Zr) for natural gas upgrading: A computational exploration," *Chemical Communications*, vol. 47, pp. 9603–9605, sep 2011.

[64] W. L. Queen, M. R. Hudson, E. D. Bloch, J. A. Mason, M. I. Gonzalez, J. S. Lee, D. Gygi, J. D. Howe, K. Lee, T. A. Darwish, M. James, V. K. Peterson, S. J. Teat, B. Smit, J. B. Neaton, J. R. Long, and C. M. Brown, "Comprehensive study of carbon dioxide adsorption in the metal-organic frameworks M2(dobdc) (M = Mg, Mn, Fe, Co, Ni, Cu, Zn)," *Chemical Science*, vol. 5, pp. 4569–4581, dec 2014.

[65] J. R. Li, R. J. Kuppler, and H. C. Zhou, "Selective gas adsorption and separation in metal-organic frameworks," *Chemical Society Reviews*, vol. 38, pp. 1477–1504, apr 2009.

[66] S. Kitagawa and M. Kondo, "Functional Micropore Chemistry of Crystalline Metal Complex-Assembled Compounds," *Bulletin of the Chemical Society of Japan*, vol. 71, pp. 1739–1753, jun 1998.

[67] A. Schneemann, V. Bon, I. Schwedler, I. Senkovska, S. Kaskel, and R. A. Fischer, "Flexible metal-organic frameworks," *Chemical Society Reviews*, vol. 43, pp. 6062–6096, aug 2014.

[68] S. Horike, S. Shimomura, and S. Kitagawa, "Soft porous crystals," *Nature Chemistry*, vol. 1, pp. 695–704, dec 2009.

[69] Y. Liu, J. H. Her, A. Dailly, A. J. Ramirez-Cuesta, D. A. Neumann, and C. M. Brown, "Reversible structural transition in MIL-53 with large temperature hysteresis," *Journal of the American Chemical Society*, vol. 130, pp. 11813–11818, sep 2008.

[70] I. Beurroies, M. Boulhout, P. L. Llewellyn, B. Kuchta, G. Férey, C. Serre, and R. Denoyel, "Using pressure to provoke the structural transition of metal-organic frameworks," *Angewandte Chemie - International Edition*, vol. 49, pp. 7526–7529, aug 2010.

[71] A. Boutin, M. A. Springuel-Huet, A. Nossov, A. Gédéon, T. Loiseau, C. Volkringer, G. Férey, F. X. Coudert, and A. H. Fuchs, "Breathing transitions in MIL-53(A1) metal-organic framework upon Xenon adsorption," *Angewandte Chemie - International Edition*, vol. 48, pp. 8314–8317, oct 2009.

[72] A. Boutin, F. X. Coudert, M. A. Springuel-Huet, A. V. Neimark, G. Férey, and A. H. Fuchs, "The behavior of flexible MIL-53(Al) upon CH4 and CO2 adsorption," *Journal of Physical Chemistry C*, vol. 114, pp. 22237–22244, dec 2010.

[73] C. Mellot-Draznieks, C. Serre, S. Surblé, N. Audebrand, and G. Férey, "Very large swelling in hybrid frameworks: A combined computational and powder diffraction study," *Journal of the American Chemical Society*, vol. 127, pp. 16273–16278, nov 2005.

[74] J. Seo, R. Matsuda, H. Sakamoto, C. Bonneau, and S. Kitagawa, "A pillared-layer coordination polymer with a rotatable pillar acting as a molecular gate for guest molecules," *Journal of the American Chemical Society*, vol. 131, pp. 12792–12800, sep 2009.

[75] S. Krause, V. Bon, I. Senkovska, U. Stoeck, D. Wallacher, D. M. Többens, S. Zander, R. S. Pillai, G. Maurin, F. X. Coudert, and S. Kaskel, "A pressure-amplifying framework material with negative gas adsorption transitions," *Nature*, vol. 532, pp. 348–352, apr 2016.

[76] J. D. Evans, L. Bocquet, and F. X. Coudert, "Origins of Negative Gas Adsorption," *Chem*, vol. 1, pp. 873–886, dec 2016.

[77] L. Vanduyfhuys and V. Van Speybroeck, "Unraveling the thermodynamic conditions for negative gas adsorption in soft porous crystals," *Communications Physics*, vol. 2, pp. 1–10, dec 2019.

[78] S. Krause, J. D. Evans, V. Bon, I. Senkovska, F. X. Coudert, D. M. Többens, D. Wallacher, N. Grimm, and S. Kaskel, "The role of temperature and adsorbate on negative gas adsorption transitions of the mesoporous metal-organic framework DUT-49," *Faraday Discussions*, vol. 225, pp. 168–183, feb 2021.

[79] S. Krause, J. D. Evans, V. Bon, I. Senkovska, P. Iacomi, F. Kolbe, S. Ehrling, E. Troschke, J. Getzschmann, D. M. Többens, A. Franz, D. Wallacher, P. G. Yot, G. Maurin, E. Brunner, P. L. Llewellyn, F. X. Coudert, and S. Kaskel, "Towards general network architecture design criteria for negative gas adsorption transitions in ultraporous frameworks," *Nature Communications*, vol. 10, pp. 1–12, dec 2019.

[80] Y. Inubushi, S. Horike, T. Fukushima, G. Akiyama, R. Matsuda, and S. Kitagawa, "Modification of flexible part in Cu2+ interdigitated framework for CH4/CO2 separation," *Chemical Communications*, vol. 46, pp. 9229–9231, dec 2010.

[81] T. Fukushima, S. Horike, Y. Inubushi, K. Nakagawa, Y. Kubota, M. Takata, and S. Kitagawa, "Solid solutions of soft porous coordination polymers: Fine-tuning of gas adsorption properties," *Angewandte Chemie - International Edition*, vol. 49, pp. 4820–4824, jun 2010.

[82] S. Horike, Y. Inubushi, T. Hori, T. Fukushima, and S. Kitagawa, "A solid solution approach to 2D coordination polymers for CH 4/CO2 and CH4/C2H6 gas separation: Equilibrium and kinetic studies," *Chemical Science*, vol. 3, pp. 116–120, jan 2012.

[83] P. Mishra, H. P. Uppara, B. Mandal, and S. Gumma, "Adsorption and separation of carbon dioxide using MIL-53(Al) metal-organic framework," *Industrial and Engineering Chemistry Research*, vol. 53, pp. 19747–19753, dec 2014.

[84] P. Horcajada, R. Gref, T. Baati, P. K. Allan, G. Maurin, P. Couvreur, G. Férey, R. E. Morris, and C. Serre, "Metal-organic frameworks in biomedicine," *Chemical Reviews*, vol. 112, pp. 1232–1268, feb 2012.

[85] A. C. McKinlay, J. F. Eubank, S. Wuttke, B. Xiao, P. S. Wheatley, P. Bazin, J. C. Lavalley, M. Daturi, A. Vimont, G. De Weireld, P. Horcajada, C. Serre, and R. E. Morris, "Nitric oxide adsorption and delivery in flexible MIL-88(Fe) metal-organic frameworks," *Chemistry of Materials*, vol. 25, pp. 1592–1599, may 2013.

[86] N. Klein, C. Herzog, M. Sabo, I. Senkovska, J. Getzschmann, S. Paasch, M. R. Lohe, E. Brunner, and S. Kaskel, "Monitoring adsorption-induced switching by 129Xe NMR spectroscopy in a new metal-organic framework Ni2(2,6-ndc)2(dabco)," *Physical Chemistry Chemical Physics*, vol. 12, pp. 11778–11784, sep 2010.

[87] Q. Chen, Z. Chang, W. C. Song, H. Song, H. B. Song, T. L. Hu, and X. H. Bu, "A controllable gate effect in cobalt(II) organic frameworks by re-

versible structure transformations," *Angewandte Chemie - International Edition*, vol. 52, pp. 11550–11553, oct 2013.

[88] V. I. Isaeva, A. L. Tarasov, V. V. Chernyshev, and L. M. Kustov, "Control of morphology and size of microporous framework MIL-53(Al) crystals by synthesis procedure," *Mendeleev Communications*, vol. 25, pp. 466–467, nov 2015.

[89] A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, "UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations," *Journal of the American Chemical Society*, vol. 114, pp. 10024–10035, dec 1992.

[90] M. A. Addicoat, N. Vankova, I. F. Akter, and T. Heine, "Extension of the universal force field to metal-organic frameworks," *Journal of Chemical Theory and Computation*, vol. 10, pp. 880–891, feb 2014.

[91] S. L. Mayo, B. D. Olafson, and W. A. Goddard, "DREIDING: A generic force field for molecular simulations," *Journal of Physical Chemistry*, vol. 94, no. 26, pp. 8897–8909, 1990.

[92] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations," *Journal of Computational Chemistry*, vol. 4, no. 2, pp. 187–217, 1983.

[93] S. Bureekaew, S. Amirjalayer, M. Tafipolsky, C. Spickermann, T. K. Roy, and R. Schmid, "MOF-FF - A flexible first-principles derived force field for metal-organic frameworks," *Physica Status Solidi (B) Basic Research*, vol. 250, pp. 1128–1141, jun 2013.

[94] J. K. Bristow, D. Tiana, and A. Walsh, "Transferable force field for metal-organic frameworks from first-principles: BTW-FF," *Journal of Chemical Theory and Computation*, vol. 10, pp. 4644–4652, oct 2014.

[95] D. Dubbeldam, K. S. Walton, D. E. Ellis, and R. Q. Snurr, "Exceptional negative thermal expansion in isoreticular metal-organic frameworks," *Angewandte Chemie - International Edition*, vol. 46, pp. 4496–4499, jun 2007.

[96] L. Vanduyfhuys, S. Vandenbrande, T. Verstraelen, R. Schmid, M. Waroquier, and V. Van Speybroeck, "QuickFF: A program for a quick and easy derivation of force fields for metal-organic frameworks from ab initio input," *Journal of Computational Chemistry*, vol. 36, pp. 1015–1027, may 2015.

[97] L. Vanduyfhuys, S. Vandenbrande, J. Wieme, M. Waroquier, T. Verstraelen, and V. Van Speybroeck, "Extension of the QuickFF force field protocol for an improved accuracy of structural, vibrational, mechanical and thermal properties of metal–organic frameworks," *Journal of Computational Chemistry*, vol. 39, pp. 999–1011, jun 2018.

[98] T. M. Becker, A. Luna-Triguero, J. M. Vicent-Luna, L. C. Lin, D. Dubbeldam, S. Calero, and T. J. Vlugt, "Potential of polarizable force fields for predicting the separation performance of small hydrocarbons in M-MOF-74," *Physical Chemistry Chemical Physics*, vol. 20, pp. 28848–28859, nov 2018.

[99] T. M. Becker, A. Luna-Triguero, J. M. Vicent-Luna, L. C. Lin, D. Dubbeldam, S. Calero, and T. J. Vlugt, "Potential of polarizable force fields for predicting the separation performance of small hydrocarbons in M-MOF-74," *Physical Chemistry Chemical Physics*, vol. 20, pp. 28848–28859, nov 2018.

[100] J. Cirera, J. C. Sung, P. B. Howland, and F. Paesani, "The effects of electronic polarization on water adsorption in metal-organic frameworks: H 2O in MIL-53(Cr)," *Journal of Chemical Physics*, vol. 137, p. 054704, aug 2012.

[101] J. G. McDaniel, S. Li, E. Tylianakis, R. Q. Snurr, and J. R. Schmidt, "Evaluation of force field performance for high-throughput screening of gas uptake in metal-organic frameworks," *Journal of Physical Chemistry C*, vol. 119, pp. 3143–3152, feb 2015.

[102] T. Düren, Y. S. Bae, and R. Q. Snurr, "Using molecular simulation to characterise metal–organic frameworks for adsorption applications," *Chemical Society Reviews*, vol. 38, pp. 1237–1247, apr 2009.

[103] S. S. Han, J. L. Mendoza-Cortés, and W. A. Goddard, "Recent advances on simulation and theory of hydrogen storage in metal–organic frameworks and covalent organic frameworks," *Chemical Society Reviews*, vol. 38, pp. 1460–1476, apr 2009.

[104] R. B. Getman, Y. S. Bae, C. E. Wilmer, and R. Q. Snurr, "Review and analysis of molecular simulations of methane, hydrogen, and acetylene storage in metal-organic frameworks," *Chemical Reviews*, vol. 112, pp. 703–723, feb 2012.

[105] T. Watanabe and D. S. Sholl, "Accelerating applications of metal-organic frameworks for gas adsorption and separation by computa-

tional screening of materials," *Langmuir*, vol. 28, pp. 14114–14128, oct 2012.

[106] O. T. Qazvini, R. Babarao, and S. G. Telfer, "Selective capture of carbon dioxide from hydrocarbons using a metal-organic framework," *Nature Communications*, vol. 12, pp. 1–8, dec 2021.

[107] Y. Zhou, X. Zhang, T. Zhou, and K. Sundmacher, "Computational Screening of Metal-Organic Frameworks for Ethylene Purification from Ethane/Ethylene/Acetylene Mixture," *Nanomaterials*, vol. 12, mar 2022.

[108] K. C. Kim, D. Fairen-Jimenez, and R. Q. Snurr, "Computational screening of functional groups for capture of toxic industrial chemicals in porous materials," *Physical Chemistry Chemical Physics*, vol. 19, pp. 31766–31772, dec 2017.

[109] D. A. Gómez-Gualdrón, Y. J. Colón, X. Zhang, T. C. Wang, Y. S. Chen, J. T. Hupp, T. Yildirim, O. K. Farha, J. Zhang, and R. Q. Snurr, "Evaluating topologically diverse metal-organic frameworks for cryo-adsorbed hydrogen storage," *Energy and Environmental Science*, vol. 9, pp. 3279–3289, oct 2016.

[110] C. E. Wilmer, K. C. Kim, and R. Q. Snurr, "An extended charge equilibration method," *Journal of Physical Chemistry Letters*, vol. 3, pp. 2506–2511, sep 2012.

[111] E. Schrödinger, "An undulatory theory of the mechanics of atoms and molecules," *Physical Review*, vol. 28, pp. 1049–1070, dec 1926.

[112] M. Born and R. Oppenheimer, "Zur Quantentheorie der Molekeln," *Annalen der Physik*, vol. 389, pp. 457–484, jan 1927.

[113] T. E. Markland and M. Ceriotti, "Nuclear quantum effects enter the mainstream," *Nature Reviews Chemistry*, vol. 2, pp. 1–14, mar 2018.

[114] Z. E. Musielak and B. Quarles, "The three-body problem," *Reports on Progress in Physics*, vol. 77, p. 30, jun 2014.

[115] D. R. Hartree and W. Hartree, "Self-consistent field, with exchange, for beryllium," *Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences*, vol. 150, pp. 9–33, may 1935.

[116] J. C. Slater, "A simplification of the Hartree-Fock method," *Physical Review*, vol. 81, pp. 385–390, feb 1951.

[117] J. Thijssen, *Computational physics, second edition*, vol. 9780521833. Cambridge University Press, jan 2007.

[118] J. C. Slater, "The theory of complex spectra," *Physical Review*, vol. 34, pp. 1293–1322, nov 1929.

[119] C. Møller and M. S. Plesset, "Note on an approximation treatment for many-electron systems," *Physical Review*, vol. 46, pp. 618–622, oct 1934.

[120] M. Head-Gordon, J. A. Pople, and M. J. Frisch, "MP2 energy evaluation by direct methods," *Chemical Physics Letters*, vol. 153, pp. 503–506, dec 1988.

[121] F. Coester and H. Kümmel, "Short-range correlations in nuclear wave functions," *Nuclear Physics*, vol. 17, pp. 477–485, jun 1960.

[122] I. Jir and E. K. Ĉiz, "On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods," *The Journal of Chemical Physics*, vol. 45, no. 11, pp. 4256–4266, 1966.

[123] G. D. Purvis and R. J. Bartlett, "A full coupled-cluster singles and doubles model: The inclusion of disconnected triples," *The Journal of Chemical Physics*, vol. 76, pp. 1910–1918, feb 1982.

[124] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, "A fifth-order perturbation comparison of electron correlation theories," *Chemical Physics Letters*, vol. 157, pp. 479–483, may 1989.

[125] R. J. Bartlett, J. D. Watts, S. A. Kucharski, and J. Noga, "Non-iterative fifth-order triple and quadruple excitation energy corrections in correlated methods," *Chemical Physics Letters*, vol. 165, pp. 513–522, feb 1990.

[126] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Physical Review*, vol. 136, p. B864, nov 1964.

[127] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Physical Review*, vol. 140, p. A1133, nov 1965.

[128] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Physical Review Letters*, vol. 77, pp. 3865–3868, oct 1996.

[129] J. P. Perdew, M. Ernzerhof, and K. Burke, "Rationale for mixing exact exchange with density functional approximations," *Journal of Chemical Physics*, vol. 105, pp. 9982–9985, dec 1996.

[130] A. J. Cohen, P. Mori-Sánchez, and W. Yang, "Insights into current limitations of density functional theory," *Science*, vol. 321, pp. 792–794, aug 2008.

[131] P. Hobza, J. šponer, and T. Reschel, "Density functional theory and molecular clusters," *Journal of Computational Chemistry*, vol. 16, pp. 1315–1325, nov 1995.

[132] S. Kristyán and P. Pulay, "Can (semi)local density functional theory account for the London dispersion forces?," *Chemical Physics Letters*, vol. 229, pp. 175–180, oct 1994.

[133] J. M. Pérez-Jordá and A. D. Becke, "A density-functional study of van der Waals forces: rare gas diatomics," *Chemical Physics Letters*, vol. 233, pp. 134–137, feb 1995.

[134] S. Grimme, "Density functional theory with London dispersion corrections," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 1, pp. 211–228, mar 2011.

[135] F. London, "Zur Theorie und Systematik der Molekularkräfte," *Zeitschrift für Physik*, vol. 63, pp. 245–279, mar 1930.

[136] 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," *Journal of Chemical Physics*, vol. 132, p. 154104, apr 2010.

[137] N. D. Yilmazer and M. Korth, "Recent progress in treating protein–ligand interactions with quantum-mechanical methods," *International Journal of Molecular Sciences*, vol. 17, may 2016.

[138] M. Gierada, I. Petit, J. Handzlik, and F. Tielens, "Hydration in silica based mesoporous materials: A DFT model," *Physical Chemistry Chemical Physics*, vol. 18, pp. 32962–32972, dec 2016.

[139] J. R. Perilla and K. Schulten, "Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations," *Nature Communications*, vol. 8, pp. 1–10, jul 2017.

[140] X. Du, Y. Li, Y. L. Xia, S. M. Ai, J. Liang, P. Sang, X. L. Ji, and S. Q. Liu, "Insights into protein–ligand interactions: Mechanisms, models, and methods," *International Journal of Molecular Sciences*, vol. 17, jan 2016.

[141] J. Lyu, X. Gong, S. J. Lee, K. Gnanasekaran, X. Zhang, M. C. Wasson, X. Wang, P. Bai, X. Guo, N. C. Gianneschi, and O. K. Farha, "Phase

Transitions in Metal-Organic Frameworks Directly Monitored through in Situ Variable Temperature Liquid-Cell Transmission Electron Microscopy and in Situ X-ray Diffraction," *Journal of the American Chemical Society*, vol. 142, pp. 4609–4615, mar 2020.

[142] R. F. Borkman, G. Simons, and R. G. Parr, "Simple Bond-Charge Model for Potential-Energy Curves of Heteronuclear Diatomic Molecules," *The Journal of Chemical Physics*, vol. 50, pp. 58–65, jan 1969.

[143] F. Ercolesi and J. B. Adams, "Interatomic potentials from first-principles calculations: The force-matching method," *Epl*, vol. 26, pp. 583–588, jun 1994.

[144] P. Brommer and F. Gähler, "Effective potentials for quasicrystals from ab-initio data," *Philosophical Magazine*, vol. 86, pp. 753–758, feb 2006.

[145] B. Waldher, J. Kuta, S. Chen, N. Henson, and A. E. Clark, "ForceFit: A code to fit classical force fields to quantum mechanical potential energy surfaces," *Journal of Computational Chemistry*, vol. 31, pp. 2307–2316, sep 2010.

[146] L. P. Wang, T. Head-Gordon, J. W. Ponder, P. Ren, J. D. Chodera, P. K. Eastman, T. J. Martinez, and V. S. Pande, "Systematic improvement of a classical molecular model of water," *Journal of Physical Chemistry B*, vol. 117, pp. 9956–9972, aug 2013.

[147] S. M. Rogge, L. Vanduyfhuys, A. Ghysels, M. Waroquier, T. Verstraelen, G. Maurin, and V. Van Speybroeck, "A Comparison of Barostats for the Mechanical Characterization of Metal-Organic Frameworks," *Journal of Chemical Theory and Computation*, vol. 11, pp. 5583–5597, oct 2015.

[148] S. M. Rogge, J. Wieme, L. Vanduyfhuys, S. Vandenbrande, G. Maurin, T. Verstraelen, M. Waroquier, and V. Van Speybroeck, "Thermodynamic Insight in the High-Pressure Behavior of UiO-66: Effect of Linker Defects and Linker Expansion," *Chemistry of Materials*, vol. 28, pp. 5721–5732, aug 2016.

[149] R. Demuynck, S. M. Rogge, L. Vanduyfhuys, J. Wieme, M. Waroquier, and V. Van Speybroeck, "Efficient Construction of Free Energy Profiles of Breathing Metal-Organic Frameworks Using Advanced Molecular Dynamics Simulations," *Journal of Chemical Theory and Computation*, vol. 13, pp. 5861–5873, dec 2017.

[150] L. Vanduyfhuys, S. M. Rogge, J. Wieme, S. Vandenbrande, G. Maurin, M. Waroquier, and V. Van Speybroeck, "Thermodynamic insight into

stimuli-responsive behaviour of soft porous crystals," *Nature Communications*, vol. 9, pp. 1–9, dec 2018.

[151] K. Szalewicz, "Symmetry-adapted perturbation theory of intermolecular forces," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 2, pp. 254–272, mar 2012.

[152] A. Stone, *The Theory of Intermolecular Forces*. Oxford University Press, may 2013.

[153] W. J. Mortier, S. K. Ghosh, and S. Shankar, "Electronegativity Equalization Method for the Calculation of Atomic Charges in Molecules," *Journal of the American Chemical Society*, vol. 108, no. 15, pp. 4315–4320, 1986.

[154] A. K. Rappé and W. A. Goddard, "Charge equilibration for molecular dynamics simulations," *Journal of Physical Chemistry*, vol. 95, no. 8, pp. 3358–3363, 1991.

[155] C. E. Wilmer, K. C. Kim, and R. Q. Snurr, "An extended charge equilibration method," *Journal of Physical Chemistry Letters*, vol. 3, pp. 2506–2511, sep 2012.

[156] U. C. Singh and P. A. Kollman, "An approach to computing electrostatic charges for molecules," *Journal of Computational Chemistry*, vol. 5, pp. 129–145, apr 1984.

[157] C. M. Breneman and K. B. Wiberg, "Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis," *Journal of Computational Chemistry*, vol. 11, pp. 361–373, apr 1990.

[158] H. Hu, Z. Lu, and W. Yang, "Fitting molecular electrostatic potentials from quantum mechanical calculations," *Journal of Chemical Theory and Computation*, vol. 3, pp. 1004–1013, may 2007.

[159] C. Campañá, B. Mussard, and T. K. Woo, "Electrostatic potential derived atomic charges for periodic systems using a modified error functional," *Journal of Chemical Theory and Computation*, vol. 5, pp. 2866–2878, oct 2009.

[160] A. E. Reed, R. B. Weinstock, and F. Weinhold, "Natural population analysis," *The Journal of Chemical Physics*, vol. 83, pp. 735–746, jul 1985.

[161] F. L. Hirshfeld, "Bonded-atom fragments for describing molecular charge densities," *Theoretica Chimica Acta*, vol. 44, pp. 129–138, jun 1977.

[162] P. Bultinck, C. Van Alsenoy, P. W. Ayers, and R. Carbó-Dorca, "Critical analysis and extension of the Hirshfeld atoms in molecules," *Journal of Chemical Physics*, vol. 126, p. 144111, apr 2007.

[163] T. Verstraelen, P. W. Ayers, V. Van Speybroeck, and M. Waroquier, "Hirshfeld-E partitioning: AIM charges with an improved trade-off between robustness and accurate electrostatics," *Journal of Chemical Theory and Computation*, vol. 9, pp. 2221–2225, may 2013.

[164] T. C. Lillestolen and R. J. Wheatley, "Redefining the atom: Atomic charge densities produced by an iterative stockholder approach," *Chemical Communications*, vol. 45, pp. 5909–5911, nov 2008.

[165] T. A. Manz and D. S. Sholl, "Chemically meaningful atomic charges that reproduce the electrostatic potential in periodic and nonperiodic materials," *Journal of Chemical Theory and Computation*, vol. 6, pp. 2455–2468, aug 2010.

[166] T. Verstraelen, S. Vandenbrande, F. Heidar-Zadeh, L. Vanduyfhuys, V. Van Speybroeck, M. Waroquier, and P. W. Ayers, "Minimal Basis Iterative Stockholder: Atoms in Molecules for Force-Field Development," *Journal of Chemical Theory and Computation*, vol. 12, pp. 3894–3912, aug 2016.

[167] R. A. Buckingham, "The classical equation of state of gaseous helium, neon and argon," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 168, pp. 264–283, oct 1938.

[168] J. A. Lemkul, J. Huang, B. Roux, and A. D. Mackerell, "An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications," *Chemical Reviews*, vol. 116, pp. 4983–5013, may 2016.

[169] S. J. Rick and S. W. Stuart, "Potentials and algorithms for incorporating polarizability in computer simulations," in *Reviews in Computational Chemistry*, vol. 18, pp. 89–146, Hoboken, New Jersey, USA: John Wiley and Sons, Inc., 2002.

[170] B. T. Thole, "Molecular polarizabilities calculated with a modified dipole interaction," *Chemical Physics*, vol. 59, pp. 341–350, aug 1981.

[171] J. Caldwell, P. A. Kollman, and L. X. Dang, "Implementation of Nonadditive Intermolecular Potentials by Use of Molecular Dynamics: Development of a Water-Water Potential and Water-Ion Cluster Interactions," *Journal of the American Chemical Society*, vol. 112, no. 25, pp. 9144–9147, 1990.

[172] J. Applequist, J. R. Carl, and K. K. Fung, "An Atom Dipole Interaction Model for Molecular Polarizability. Application to Polyatomic Molecules and Determination of Atom Polarizabilities," *Journal of the American Chemical Society*, vol. 94, pp. 2952–2960, may 1972.

[173] D. Elking, T. O. Darden, and R. J. Woods, "Gaussian induced dipole polarization model," *Journal of Computational Chemistry*, vol. 28, pp. 1261–1274, may 2007.

[174] A. Van Der Vaart and K. M. Merz, "The role of polarization and charge transfer in the solvation of biomolecules," *Journal of the American Chemical Society*, vol. 121, pp. 9182–9190, oct 1999.

[175] J. G. McDaniel and J. R. Schmidt, "Next-Generation Force Fields from Symmetry-Adapted Perturbation Theory," *Annual Review of Physical Chemistry*, vol. 67, pp. 467–488, may 2016.

[176] A. J. Stone and M. Alderton, "Distributed multipole analysis methods and applications," *Molecular Physics*, vol. 56, no. 5, pp. 1047–1064, 1985.

[177] M. J. Van Vleet, A. J. Misquitta, A. J. Stone, and J. R. Schmidt, "Beyond Born-Mayer: Improved Models for Short-Range Repulsion in ab Initio Force Fields," *Journal of Chemical Theory and Computation*, vol. 12, pp. 3851–3870, jun 2016.

[178] A. J. Misquitta and A. J. Stone, "ISA-Pol: distributed polarizabilities and dispersion models from a basis-space implementation of the iterated stockholder atoms procedure," *Theoretical Chemistry Accounts*, vol. 137, jun 2018.

[179] S. Vandenbrande, M. Waroquier, V. Van Speybroeck, and T. Verstraelen, "The monomer electron density force field (MEDFF): A Physically Inspired Model for Noncovalent Interactions," *Journal of Chemical Theory and Computation*, vol. 13, pp. 161–179, jan 2017.

[180] T. Verstraelen, P. W. Ayers, V. Van Speybroeck, and M. Waroquier, "ACKS2: Atom-condensed Kohn-Sham DFT approximated to second order," *Journal of Chemical Physics*, vol. 138, p. 074108, feb 2013.

[181] T. Verstraelen, S. Vandenbrande, and P. W. Ayers, "Direct computation of parameters for accurate polarizable force fields," *Journal of Chemical Physics*, vol. 141, p. 194114, nov 2014.

[182] J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," *Physical Review Letters*, vol. 98, p. 146401, apr 2007.

[183] J. Behler, "Atom-centered symmetry functions for constructing high-dimensional neural network potentials," *Journal of Chemical Physics*, vol. 134, p. 074106, feb 2011.

[184] J. Behler, "Perspective: Machine learning potentials for atomistic simulations," *Journal of Chemical Physics*, vol. 145, p. 170901, nov 2016.

[185] B. Schölkopf, A. Smola, and K. R. Müller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," *Neural Computation*, vol. 10, pp. 1299–1319, jul 1998.

[186] L. Zhang, J. Han, H. Wang, R. Car, and E. Weinan, "Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics," *Physical Review Letters*, vol. 120, jul 2018.

[187] J. S. Smith, O. Isayev, and A. E. Roitberg, "ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost," *Chemical Science*, vol. 8, pp. 3192–3203, mar 2017.

[188] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," *34th International Conference on Machine Learning, ICML 2017*, vol. 3, pp. 2053–2070, apr 2017.

[189] K. T. Schütt, H. E. Sauceda, P. J. Kindermans, A. Tkatchenko, and K. R. Müller, "SchNet - A deep learning architecture for molecules and materials," *Journal of Chemical Physics*, vol. 148, p. 241722, jun 2018.

[190] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, "E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials," *Nature Communications*, vol. 13, dec 2022.

[191] M. E. Tuckerman, *Statisical Mechanics: Theory and Molecular Simulation*. Oxford University Press, 2010.

[192] D. Frenkel, B. Smit, J. Tobochnik, S. R. McKay, and W. Christian, *Understanding Molecular Simulation*, vol. 11. AIP Publishing, 1997.

[193] G. J. Martyna, D. J. Tobias, and M. L. Klein, "Constant pressure molecular dynamics algorithms," *The Journal of Chemical Physics*, vol. 101, pp. 4177–4189, sep 1994.

[194] D. S. Lemons and A. Gythiel, " Paul Langevin's 1908 paper "On the Theory of Brownian Motion" ["Sur la théorie du mouvement brownien," C. R. Acad. Sci. (Paris) 146 , 530–533 (1908)] ," *American Journal of Physics*, vol. 65, pp. 1079–1081, nov 1997.

[195] H. C. Andersen, "Molecular dynamics simulations at constant pressure and/or temperature," *The Journal of Chemical Physics*, vol. 72, pp. 2384–2393, feb 1980.

[196] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *Journal of Chemical Physics*, vol. 126, p. 014101, jan 2007.

[197] G. J. Martyna, M. L. Klein, and M. Tuckerman, "Nosé-Hoover chains: The canonical ensemble via continuous dynamics," *The Journal of Chemical Physics*, vol. 97, pp. 2635–2643, aug 1992.

[198] H. J. Berendsen, J. P. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of Chemical Physics*, vol. 81, pp. 3684–3690, oct 1984.

[199] W. G. Hoover, "Constant-pressure equations of motion," *Physical Review A*, vol. 34, pp. 2499–2500, sep 1986.

[200] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein, "Explicit reversible integrators for extended systems dynamics," *Molecular Physics*, vol. 87, pp. 1117–1157, apr 1996.

[201] H. Fischer, *A History of the Central Limit Theorem.* Springer New York, 2011.

[202] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, pp. 1087–1092, jun 1953.

[203] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid Monte Carlo," *Physics Letters B*, vol. 195, pp. 216–222, sep 1987.

[204] J. K. Brennan and W. G. Madden, "Phase coexistence curves for off-lattice polymer-solvent mixtures: Gibbs-ensemble simulations," *Macromolecules*, vol. 35, pp. 2827–2834, mar 2002.

[205] B. J. Banaszak, R. Faller, and J. J. De Pablo, "Simulation of the effects of chain architecture on the sorption of ethylene in polyethylene," *Journal of Chemical Physics*, vol. 120, pp. 11304–11315, jun 2004.

[206] F. X. Coudert, M. Jeffroy, A. H. Fuchs, A. Boutin, and C. Mellot-Draznieks, "Thermodynamics of guest-induced structural transitions in hybrid organic-inorganic frameworks," *Journal of the American Chemical Society*, vol. 130, pp. 14294–14302, oct 2008.

[207] F. X. Coudert, A. Boutin, M. Jeffroy, C. Mellot-Draznieks, and A. H. Fuchs, "Thermodynamic methods and models to study flexible metal-organic frameworks," *ChemPhysChem*, vol. 12, pp. 247–258, feb 2011.

[208] S. M. Rogge, R. Goeminne, R. Demuynck, J. J. Gutiérrez-Sevillano, S. Vandenbrande, L. Vanduyfhuys, M. Waroquier, T. Verstraelen, and V. Van Speybroeck, "Modeling Gas Adsorption in Flexible Metal–Organic Frameworks via Hybrid Monte Carlo/Molecular Dynamics Schemes," *Advanced Theory and Simulations*, vol. 2, p. 1800177, apr 2019.

[209] B. K. Peterson and K. E. Gubbins, "Phase transitions in a cylindrical pore grand canonical monte carlo, mean field theory and the kelvin equation," *Molecular Physics*, vol. 62, no. 1, pp. 215–226, 1987.

[210] M. Jeffroy, A. H. Fuchs, and A. Boutin, "Structural changes in nanoporous solids due to fluid adsorption: Thermodynamic analysis and Monte Carlo simulations," *Chemical Communications*, pp. 3275–3277, jul 2008.

[211] B. Mehlig, D. W. Heermann, and B. M. Forrest, "Hybrid Monte Carlo method for condensed-matter systems," *Physical Review B*, vol. 45, pp. 679–685, jan 1992.

[212] F. A. Brotz and J. J. De Pablo, "Hybrid Monte Carlo simulation of silica," *Chemical Engineering Science*, vol. 49, pp. 3015–3031, sep 1994.

[213] M. E. Tuckerman, B. J. Berne, G. J. Martyna, and M. L. Klein, "Efficient molecular dynamics and hybrid Monte Carlo algorithms for path integrals," *The Journal of Chemical Physics*, vol. 99, pp. 2796–2808, aug 1993.

[214] A. Ghoufi and G. Maurin, "Hybrid monte carlo simulations combined with a phase mixture model to predict the structural transitions of a porous Metal-organic framework material upon adsorption of guest molecules," *Journal of Physical Chemistry C*, vol. 114, pp. 6496–6502, apr 2010.

[215] D. Bousquet, F. X. Coudert, and A. Boutin, "Free energy landscapes for the thermodynamic understanding of adsorption-induced deformations and structural transitions in porous materials.," *The Journal of chemical physics*, vol. 137, p. 044118, jul 2012.

[216] R. Goeminne, S. Krause, S. Kaskel, T. Verstraelen, and J. D. Evans, "Charting the Complete Thermodynamic Landscape of Gas Adsorp-

tion for a Responsive Metal-Organic Framework," *Journal of the American Chemical Society*, vol. 143, pp. 4143–4147, mar 2021.

[217] M. G. Martin and J. I. Siepmann, "Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes," *Journal of Physical Chemistry B*, vol. 102, no. 14, pp. 2569–2577, 1998.

[218] L. Vanduyfhuys, T. Verstraelen, M. Vandichel, M. Waroquier, and V. Van Speybroeck, "Ab initio parametrized force field for the flexible metal-organic framework MIL-53(Al)," *Journal of Chemical Theory and Computation*, vol. 8, pp. 3217–3231, sep 2012.

[219] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general Amber force field," *Journal of Computational Chemistry*, vol. 25, pp. 1157–1174, jul 2004.

[220] E. Boulanger, L. Huang, C. Rupakheti, A. D. Mackerell, and B. Roux, "Optimized Lennard-Jones Parameters for Druglike Small Molecules," *Journal of Chemical Theory and Computation*, vol. 14, pp. 3121–3131, jun 2018.

[221] X. Wu, J. Huang, W. Cai, and M. Jaroniec, "Force field for ZIF-8 flexible frameworks: Atomistic simulation of adsorption, diffusion of pure gases as CH4, H2, CO 2 and N2," *RSC Advances*, vol. 4, pp. 16503–16511, apr 2014.

[222] J. Pérez-Pellitero, H. Amrouche, F. R. Siperstein, G. Pirngruber, C. Nieto-Draghi, G. Chaplais, A. Simon-Masseron, D. Bazer-Bachi, D. Peralta, and N. Bats, "Adsorption of CO2, CH4, and N2 on zeolitic imidazolate frameworks: Experiments and simulations," *Chemistry - A European Journal*, vol. 16, pp. 1560–1571, feb 2010.

[223] M. Fischer, J. R. Gomes, and M. Jorge, "Computational approaches to study adsorption in MOFs with unsaturated metal sites," *Molecular Simulation*, vol. 40, pp. 537–556, aug 2014.

[224] T. M. Becker, J. Heinen, D. Dubbeldam, L. C. Lin, and T. J. Vlugt, "Polarizable Force Fields for CO2 and CH4 Adsorption in M-MOF-74," *Journal of Physical Chemistry C*, vol. 121, pp. 4659–4673, mar 2017.

[225] D. M. York and W. Yang, "A chemical potential equalization method for molecular simulations," *Journal of Chemical Physics*, vol. 104, pp. 159–172, jan 1996.

[226] G. Tabacchi, C. J. Mundy, J. Hutter, and M. Parrinello, "Classical polarizable force fields parametrized from ab initio calculations," *Journal of Chemical Physics*, vol. 117, pp. 1416–1433, jul 2002.

[227] G. Tabacchi, J. Hutter, and C. J. Mundy, "A density-functional approach to polarizable models: A Kim-Gordon response density interaction potential for molecular simulations," *Journal of Chemical Physics*, vol. 123, p. 074108, aug 2005.

[228] A. Wadehra and S. K. Ghosh, "A density functional theory-based chemical potential equalisation approach to molecular polarizability," *Journal of Chemical Sciences*, vol. 117, no. 5, pp. 401–409, 2005.

[229] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, and P. Ren, "Polarizable atomic multipole-based AMOEBA force field for proteins," *Journal of Chemical Theory and Computation*, vol. 9, pp. 4046–4063, sep 2013.

[230] J. Wang, P. Cieplak, J. Li, T. Hou, R. Luo, and Y. Duan, "Development of polarizable models for molecular mechanical calculations I: Parameterization of atomic polarizability," *Journal of Physical Chemistry B*, vol. 115, pp. 3091–3099, mar 2011.

[231] X. Chu and A. Dalgarno, "Linear response time-dependent density functional theory for van der Waals coefficients," *Journal of Chemical Physics*, vol. 121, pp. 4083–4088, sep 2004.

[232] T. Brinck, J. S. Murray, and P. Politzer, "Polarizability and volume," mar 1993.

[233] A. D. Becke and E. R. Johnson, "Exchange-hole dipole moment and the dispersion interaction: High-order dispersion coefficients," *Journal of Chemical Physics*, vol. 124, no. 1, 2006.

[234] A. Tkatchenko and M. Scheffler, "Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data," *Physical Review Letters*, vol. 102, p. 073005, feb 2009.

[235] M. M. Ghahremanpour, P. J. Van Maaren, C. Caleman, G. R. Hutchison, and D. Van Der Spoel, "Polarizable Drude Model with s-Type Gaussian or Slater Charge Density for General Molecular Mechanics Force Fields," *Journal of Chemical Theory and Computation*, vol. 14, pp. 5553–5566, nov 2018.

[236] S. P. Jarvis, A. M. Sweetman, L. Kantorovich, E. Mcglynn, and P. Moriarty, "Pauli's Principle in Probe Microscopy," tech. rep.

[237] Y. S. Kim, S. K. Kim, and W. D. Lee, "Dependence of the closed-shell repulsive interaction on the overlap of the electron densities," *Chemical Physics Letters*, vol. 80, pp. 574–575, jun 1981.

[238] A. J. Misquitta and A. J. Stone, "Ab Initio Atom-Atom Potentials Using C am CASP: Theory and Application to Many-Body Models for the Pyridine Dimer," *Journal of Chemical Theory and Computation*, vol. 12, pp. 4184–4208, sep 2016.

[239] J. Řezáč, K. E. Riley, and P. Hobza, "S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures," *Journal of Chemical Theory and Computation*, vol. 7, pp. 2427–2438, aug 2011.

[240] B. Jeziorski, R. Moszynski, and K. Szalewicz, "Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes," *Chemical Reviews*, vol. 94, pp. 1887–1930, nov 1994.

[241] M. Masia, M. Probst, and R. Rey, "On the performance of molecular polarization methods. II. Water and carbon tetrachloride close to a cation," *Journal of Chemical Physics*, vol. 123, p. 164505, oct 2005.

[242] L. C. Blum and J. L. Reymond, "970 Million druglike small molecules for virtual screening in the chemical universe database GDB-13," *Journal of the American Chemical Society*, vol. 131, pp. 8732–8733, jul 2009.

[243] M. Rupp, A. Tkatchenko, K. R. Müller, and O. A. Von Lilienfeld, "Fast and accurate modeling of molecular atomization energies with machine learning," *Physical Review Letters*, vol. 108, p. 058301, jan 2012.

[244] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, M. B., G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, A. P. Raghavachari, K. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. Daniels, O. D.; Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, "Gaussian16, Wallingford, CT," *Gaussian16 (Revision A.03)*, 2016.

[245] A. Tkatchenko, R. A. Distasio, R. Car, and M. Scheffler, "Accurate and efficient method for many-body van der Waals interactions," *Physical Review Letters*, vol. 108, p. 236402, jun 2012.

[246] A. Milet, R. Moszynski, P. E. Wormer, and A. Van der Avoird, "Hydrogen Bonding in Water Clusters: Pair and Many-Body Interactions from Symmetry-Adapted Perturbation Theory," *Journal of Physical Chemistry A*, vol. 103, pp. 6811–6819, aug 1999.

[247] R. Podeszwa and K. Szalewicz, "Three-body symmetry-adapted perturbation theory based on Kohn-Sham description of the monomers," *Journal of Chemical Physics*, vol. 126, no. 19, 2007.

[248] V. Babin, G. R. Medders, and F. Paesani, "Development of a "first principles" water potential with flexible monomers. II: Trimer potential energy surface, third virial coefficient, and small clusters," *Journal of Chemical Theory and Computation*, vol. 10, pp. 1599–1607, apr 2014.

[249] D. M. Bates and G. S. Tschumper, "CCSD(T) complete basis set limit relative energies for low-lying water hexamer structures," *Journal of Physical Chemistry A*, vol. 113, pp. 3555–3559, apr 2009.

[250] G. R. Medders, A. W. Götz, M. A. Morales, P. Bajaj, and F. Paesani, "On the representation of many-body interactions in water," *Journal of Chemical Physics*, vol. 143, p. 104102, sep 2015.

[251] C. J. Burnham, D. J. Anick, P. K. Mankoo, and G. F. Reiter, "The vibrational proton potential in bulk liquid water and ice," *Journal of Chemical Physics*, vol. 128, p. 154519, apr 2008.

[252] Y. Wang and J. M. Bowman, "Ab initio potential and dipole moment surfaces for water. II. Local-monomer calculations of the infrared spectra of water clusters," *Journal of Chemical Physics*, vol. 134, p. 154510, apr 2011.

[253] V. Babin, C. Leforestier, and F. Paesani, "Development of a "first principles" water potential with flexible monomers: Dimer potential energy surface, VRT spectrum, and second virial coefficient," *Journal of Chemical Theory and Computation*, vol. 9, pp. 5395–5403, dec 2013.

[254] G. R. Medders, V. Babin, and F. Paesani, "Development of a "first-principles" water potential with flexible monomers. III. Liquid phase properties," *Journal of Chemical Theory and Computation*, vol. 10, pp. 2906–2910, aug 2014.

[255] M. L. Laury, L. P. Wang, V. S. Pande, T. Head-Gordon, and J. W. Ponder, "Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model," *Journal of Physical Chemistry B*, vol. 119, pp. 9423–9437, jul 2015.

[256] J. Řezáč, Y. Huang, P. Hobza, and G. J. Beran, "Benchmark Calculations of Three-Body Intermolecular Interactions and the Performance of Low-Cost Electronic Structure Methods," *Journal of Chemical Theory and Computation*, vol. 11, pp. 3065–3079, jul 2015.

[257] G. Chałasiński, M. M. Szcześniak, and R. A. Kendall, "Supermolecular approach to many-body dispersion interactions in weak van der Waals complexes: He, Ne, and Ar trimers," *The Journal of Chemical Physics*, vol. 101, pp. 8860–8869, nov 1994.

[258] P. Ren, C. Wu, and J. W. Ponder, "Polarizable atomic multipole-based molecular mechanics for organic molecules," *Journal of Chemical Theory and Computation*, vol. 7, pp. 3143–3161, oct 2011.

[259] B. Brauer, M. K. Kesharwani, S. Kozuch, and J. M. Martin, "The S66x8 benchmark for noncovalent interactions revisited: Explicitly correlated: Ab initio methods and density functional theory," *Physical Chemistry Chemical Physics*, vol. 18, pp. 20905–20925, aug 2016.

[260] D. G. Smith, L. A. Burns, K. Patkowski, and C. D. Sherrill, "Revised Damping Parameters for the D3 Dispersion Correction to Density Functional Theory," *Journal of Physical Chemistry Letters*, vol. 7, pp. 2197–2203, jun 2016.

[261] M. J. Van Vleet, A. J. Misquitta, and J. R. Schmidt, "New Angles on Standard Force Fields: Toward a General Approach for Treating Atomic-Level Anisotropy," *Journal of Chemical Theory and Computation*, vol. 14, pp. 739–758, feb 2018.

[262] L. A. Burns, J. C. Faver, Z. Zheng, M. S. Marshall, D. G. Smith, K. Vanommeslaeghe, A. D. MacKerell, K. M. Merz, and C. D. Sherrill, "The BioFragment Database (BFDb): An open-data platform for computational chemistry analysis of noncovalent interactions," *Journal of Chemical Physics*, vol. 147, p. 161727, oct 2017.

[263] L. A. Burns, M. S. Marshall, and C. D. Sherrill, "Appointing silver and bronze standards for noncovalent interactions: A comparison of spin-component-scaled (SCS), explicitly correlated (F12), and specialized wavefunction approaches," *Journal of Chemical Physics*, vol. 141, p. 234111, dec 2014.

[264] J. C. Faver, M. L. Benson, X. He, B. P. Roberts, B. Wang, M. S. Marshall, M. R. Kennedy, C. D. Sherrill, and K. M. Merz, "Formal estimation of errors in computed absolute interaction energies of protein-ligand complexes," *Journal of Chemical Theory and Computation*, vol. 7, pp. 790–797, mar 2011.

[265] T. A. Halgren, "MMFF VI. MMFF94s option for energy minimization studies," *Journal of Computational Chemistry*, vol. 20, pp. 720–729, may 1999.

[266] K. S. Thanthiriwatte, E. G. Hohenstein, L. A. Burns, and C. D. Sherrill, "Assessment of the performance of DFT and DFT-D methods for describing distance dependence of hydrogen-bonded interactions," *Journal of Chemical Theory and Computation*, vol. 7, pp. 88–96, jan 2011.

[267] Y. Lee, R. Poloni, and J. Kim, "Probing gas adsorption in MOFs using an efficient ab initio widom insertion Monte Carlo method," *Journal of Computational Chemistry*, vol. 37, pp. 2808–2815, dec 2016.

[268] S. Vandenbrande, M. Waroquier, V. Van Speybroeck, and T. Verstraelen, "Ab Initio Evaluation of Henry Coefficients Using Importance Sampling," *Journal of Chemical Theory and Computation*, vol. 14, pp. 6359–6369, dec 2018.

[269] A. Kundu, G. Piccini, K. Sillar, and J. Sauer, "Ab Initio Prediction of Adsorption Isotherms for Small Molecules in Metal-Organic Frameworks," *Journal of the American Chemical Society*, vol. 138, pp. 14047–14056, oct 2016.

[270] K. Sillar, A. Kundu, and J. Sauer, "Ab Initio Adsorption Isotherms for Molecules with Lateral Interactions: CO2 in Metal-Organic Frameworks," *Journal of Physical Chemistry C*, vol. 121, pp. 12789–12799, jun 2017.

[271] A. Kundu, K. Sillar, and J. Sauer, "Ab Initio Prediction of Adsorption Isotherms for Gas Mixtures by Grand Canonical Monte Carlo Simulations on a Lattice of Sites," *Journal of Physical Chemistry Letters*, vol. 8, pp. 2713–2718, jun 2017.

[272] S. Alvarez, "A cartography of the van der Waals territories," *Dalton Transactions*, vol. 42, pp. 8617–8636, jun 2013.

[273] A. Hjorth Larsen, J. JØrgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin,

E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. SchiØtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, "The atomic simulation environment - A Python library for working with atoms," *Journal of Physics Condensed Matter*, vol. 29, p. 30, jun 2017.

[274] J. M. Simmons, H. Wu, W. Zhou, and T. Yildirim, "Carbon capture in metal-organic frameworks - A comparative study," *Energy and Environmental Science*, vol. 4, pp. 2177–2185, jun 2011.

[275] S. I. Garces, J. Villarroel-Rocha, K. Sapag, S. A. Korili, and A. Gil, "Comparative Study of the Adsorption Equilibrium of CO 2 on Microporous Commercial Materials at Low Pressures," 2013.

[276] D. Fairen-Jimenez, S. A. Moggach, M. T. Wharmby, P. A. Wright, S. Parsons, and T. D€ Uren, "Opening the Gate: Framework Flexibility in ZIF-8 Explored by Experiments and Simulations," *J. Am. Chem. Soc*, vol. 133, pp. 8900–8902, 2011.

[277] 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*, vol. 22, pp. 7577–7585, apr 2020.

[278] P. D. Dietzel, V. Besikiotis, and R. Blom, "Application of metal-organic frameworks with coordinatively unsaturated metal sites in storage and separation of methane and carbon dioxide," *Journal of Materials Chemistry*, vol. 19, pp. 7362–7370, oct 2009.

[279] D. Yu, A. O. Yazaydin, J. R. Lane, P. D. Dietzel, and R. Q. Snurr, "A combined experimental and quantum chemical study of CO2 adsorption in the metal-organic framework CPO-27 with different metals," *Chemical Science*, vol. 4, pp. 3544–3556, jul 2013.

[280] L. Sarkisov, R. Bueno-Perez, M. Sutharson, and D. Fairen-Jimenez, "Materials Informatics with PoreBlazer v4.0 and the CSD MOF Database," *Chemistry of Materials*, vol. 32, pp. 9849–9867, dec 2020.

[281] A. C. Van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, "ReaxFF: A reactive force field for hydrocarbons," *Journal of Physical Chemistry A*, vol. 105, pp. 9396–9409, oct 2001.

[282] T. P. Senftle, S. Hong, M. M. Islam, S. B. Kylasa, Y. Zheng, Y. K. Shin, C. Junkermeier, R. Engel-Herbert, M. J. Janik, H. M. Aktulga, T. Ver-

straelen, A. Grama, and A. C. Van Duin, "The ReaxFF reactive force-field: Development, applications and future directions," mar 2016.

[283] A. Vjunov, M. Y. Hu, J. Feng, D. M. Camaioni, D. Mei, J. Z. Hu, C. Zhao, and J. A. Lercher, "Following solid-acid-catalyzed reactions by MAS NMR spectroscopy in liquid phase - Zeolite-catalyzed conversion of cyclohexanol in water," *Angewandte Chemie - International Edition*, vol. 53, pp. 479–482, jan 2014.

[284] P. Tomkins, M. Ranocchiari, and J. A. Van Bokhoven, "Direct Conversion of Methane to Methanol under Mild Conditions over Cu-Zeolites and beyond," *Accounts of Chemical Research*, vol. 50, pp. 418–425, feb 2017.

[285] J. Blgeleisen and M. G. Mayer, "Calculation of equilibrium constants for isotopic exchange reactions," *The Journal of Chemical Physics*, vol. 15, pp. 261–267, may 1947.

[286] J. Liu, A. I. Benin, A. M. Furtado, P. Jakubczak, R. R. Willis, and M. D. Levan, "Stability effects on CO2 adsorption for the DOBDC series of metal-organic frameworks," *Langmuir*, vol. 27, pp. 11451–11456, sep 2011.

[287] K. Yu, K. Kiesling, and J. R. Schmidt, "Trace flue gas contaminants poison coordinatively unsaturated metal-organic frameworks: Implications for CO2 adsorption and separation," *Journal of Physical Chemistry C*, vol. 116, pp. 20480–20488, sep 2012.

[288] S. Han, Y. Huang, T. Watanabe, S. Nair, K. S. Walton, D. S. Sholl, and J. Carson Meredith, "MOF stability and gas adsorption as a function of exposure to water, humid air, SO2, and NO2," *Microporous and Mesoporous Materials*, vol. 173, pp. 86–91, jun 2013.

[289] B. S. Gelfand, J.-B. Lin, and G. K. H. Shimizu, "Design of a Humidity-Stable Metal-Organic Framework Using a Phosphonate Monoester Ligand," *Inorg. Chem*, vol. 17, p. 53, 2015.

[290] S. M. Rogge, M. Waroquier, and V. Van Speybroeck, "Reliably Modeling the Mechanical Stability of Rigid and Flexible Metal-Organic Frameworks," *Accounts of Chemical Research*, vol. 51, pp. 138–148, jan 2018.

[291] S. Krause, V. Bon, I. Senkovska, D. M. Többens, D. Wallacher, R. S. Pillai, G. Maurin, and S. Kaskel, "The effect of crystallite size on pressure amplification in switchable porous solids," *Nature Communications*, vol. 9, dec 2018.

[292] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, "LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales," *Computer Physics Communications*, vol. 271, p. 108171, feb 2022.

[293] S. Páll, M. J. Abraham, C. Kutzner, B. Hess, and E. Lindahl, "Tackling exascale software challenges in molecular dynamics simulations with GROMACS," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8759, pp. 3–27, Springer Verlag, 2015.

[294] J. Behler, "Four Generations of High-Dimensional Neural Network Potentials," aug 2021.

[295] S. Faraji, S. A. Ghasemi, S. Rostami, R. Rasoulkhani, B. Schaefer, S. Goedecker, and M. Amsler, "High accuracy and transferability of a neural network potential through charge equilibration for calcium fluoride," *Physical Review B*, vol. 95, p. 104105, mar 2017.

[296] J. Cioslowski and B. B. Stefanov, "Electron flow and electronegativity equalization in the process of bond formation," *The Journal of Chemical Physics*, vol. 99, pp. 5151–5162, oct 1993.

[297] J. Chen and T. J. Martínez, "QTPIE: Charge transfer with polarization current equalization. A fluctuating charge model with correct asymptotics," *Chemical Physics Letters*, vol. 438, pp. 315–320, apr 2007.

[298] T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffmann, D. Golze, J. Wilhelm, S. Chulkov, M. H. Bani-Hashemian, V. Weber, U. Borštnik, M. Taillefumier, A. S. Jakobovits, A. Lazzaro, H. Pabst, T. Müller, R. Schade, M. Guidon, S. Andermatt, N. Holmberg, G. K. Schenter, A. Hehn, A. Bussy, F. Belleflamme, G. Tabacchi, A. Glöß, M. Lass, I. Bethune, C. J. Mundy, C. Plessl, M. Watkins, J. VandeVondele, M. Krack, and J. Hutter, "CP2K: An electronic structure and molecular dynamics software package -Quickstep: Efficient and accurate electronic structure calculations," *Journal of Chemical Physics*, vol. 152, no. 19, p. 194103, 2020.

[299] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov,

J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, "Gaussian09 Revision E.01."

[300] V. Kapil, M. Rossi, O. Marsalek, R. Petraglia, Y. Litman, T. Spura, B. Cheng, A. Cuzzocrea, R. H. Meißner, D. M. Wilkins, B. A. Helfrecht, P. Juda, S. P. Bienvenue, W. Fang, J. Kessler, I. Poltavsky, S. Vandenbrande, J. Wieme, C. Corminboeuf, T. D. Kühne, D. E. Manolopoulos, T. E. Markland, J. O. Richardson, A. Tkatchenko, G. A. Tribello, V. Van Speybroeck, and M. Ceriotti, "i-PI 2.0: A universal force engine for advanced molecular simulations," *Computer Physics Communications*, vol. 236, pp. 214–223, mar 2019.

[301] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, "PLUMED 2: New feathers for an old bird," *Computer Physics Communications*, vol. 185, pp. 604–613, feb 2014.

[302] G. Kresse and J. Furthmüller, "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set," *Physical Review B - Condensed Matter and Materials Physics*, vol. 54, pp. 11169–11186, oct 1996.

[303] T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza, and M. Haranczyk, "Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials," *Microporous and Mesoporous Materials*, vol. 149, pp. 134–141, feb 2012.

This research was supported by the extraordinary research fund (BOF) through a starting grant (Grant No. 01N06617).

The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, the Research Foundation - Flanders (FWO) and the Flemish Government.