Technische Universität Dresden
Faculty of Mechanical Science and Engineering

Investigation of Surface Phenomena in Metal-Organic Frameworks using Molecular Simulation Methods

Dissertation

in Partial Fulfillment of the Requirements for the degree of

Doctor of Engineering (Dr.-Ing.)

submitted to the Faculty of Mechanical Science and Engineering at Technische Universität Dresden

by

M.Sc. Alexander von Wedelstedt
born on 1 December 1992 in Freital, Germany

Examining Board:

Prof. Dr. rer. nat. habil. Cornelia Breitkopf (TU Dresden)
Prof. Dr. rer. nat. habil. Grit Kalies (HTW Dresden)

Submission date: 24 June 2022
Defense date: 12 December 2022

Acknowledgments

I would like to express my gratitude to all those who supported and motivated me during the preparation of this thesis.

I would like to express my deepest thanks to Prof. Dr. Grit Kalies for her continued support throughout my dissertation, for the opportunity to conduct my research in her group, as well as for her professional and personal support. I also thank Prof. Dr. Cornelia Breitkopf for her help as cooperative supervisor and for the organizational support.

Many thanks go to the members of the research group of Applied and Theoretical Thermodynamics at HTW University of Applied Sciences Dresden, Dr.-Ing. Mandy Klauck, Dr. Matthias Hauser, and Thomas Hähnel, for our helpful discussions and their support in experimental issues.

My special regards go to Prof. Dr. Guillaume Maurin, Prof. Dr. Randall Q. Snurr, and Prof. Dr. Berend Smit for the opportunity to be part of their research group and expand my knowledge in the field of computational chemistry.

I would also like to express my gratitude to Benjamin Majid, Miriam Pougin, and Natalie Thomas for proofreading and editing the first version of this text. My special regards go to Lou D. Chukman. As a native English speaker, he helped me to significantly improve the linguistic as well as stylistic quality of this thesis.

Finally, I'd also like to acknowledge the financial support from the European Union through the European Social Fund, the State of Saxony, and the German Academic Exchange Service.

Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Meaning
ADMM	Auxiliary density matrix method
B3LYP	Becke, 3-parameter, Lee–Yang–Parr
BET	Brunauer-Emmett-Teller
BLYP	Becke, Lee-Yang-Parr
ButOH	n-Butanol
CAU	Christian-Albrecht-Universität zu Kiel
CAVE	Cave automatic virtual environments
CBMC	Configurational-bias Monte Carlo
CFCMC	Continuous fractional component Monte Carlo
CIF	Crystallographic information file
CSVR	Canonical sampling through velocity rescaling
DDEC	Density derived electrostatic and chemical
DFT	Density functional theory
DNP	Double numerical basis set with polarization functions
DZVP	Double-zeta valence polarized
EQeq	Extended charge equilibration
ESP	Electrostatic potential
EtOH	Ethanol
Fig	Figure
GCMC	Grand canonical Monte Carlo
GDS	Gibbs dividing surface
GGA	Generalized gradient approximation
GTH	Goedecker-Teter-Hutter
HTW	Hochschule für Technik und Wirtschaft
HMD	Head-mounted displays
IAST	Ideal adsorbed solution theory
INC	Institut für Nichtklassische Chemie
IUPAC	International Union of Pure and Applied Chemistry
LCAO	Linear combination of atomic orbitals
LDA	Local-density approximation
MBBB	Molecular building block-based
MIL	Materials Institute Lavoisier
MOF	Metal-organic framework
MSD	Mean-squared displacement
NVT	Canonical ensemble
PBE	Perdew-Burke-Ernzerhof
PDB	Protein databank
PPPM	Particle-Particle Particle-Mesh
Qeq	Charge equilibration

RAST	Real adsorbed solution theory
RDF	Radial distribution function
SE	Statistical error
ToBaCCo	Topologically based crystal constructor
TraPPE	Transferable potentials for phase equilibria
TZVP	Triple-zeta valence polarized
UA	United atom
UFF	Universal force field
UiO	Universitetet i Oslo
UNIFAC (Do)	Universal quasichemical functional group activity coefficients (Dortmund)
VR	Virtual reality

List of Symbols

Symbol	Meaning	Unit
$\tilde{A}$	Helmholtz free energy	J
A	Area	m^2
a	Number of systems in a microscopic state in an ensemble	-
$\tilde{a}$	van-der-Waals parameter	$J{\cdot}m^3/mol^2$
$\check{a}$	Activity coefficient	-
$\vec{a}$	Acceleration	m/s^2
b	Parameter for the dual-site Langmuir-Sips model	1/Pa
$\tilde{b}$	van-der-Waals parameter	m^3/mol
C	Constant	-
c	Concentration	mol/l
d	Distance	m
E	Energy	J
e	Elementary charge ($1.602 \cdot 10^{-19}$ C)	C
$\vec{F}$	Force	N
f	Fraction	-
$\check{f}_i$	Fugacity	Pa
$\tilde{f}$	Function	-
G	Gibbs free energy	J
$\vec{g}$	Constraining intramolecular force	N
$\tilde{h}$	Planck's constant ($6.626 \cdot 10^{-34}$ Js)	Js
$\check{h}$	Fitting parameter of the Bi-Langmuir function	-
h	Single-electron Hamiltonian	J
I	Integral	-
i	Loop variable (e.g. numerical value, undefined chemical species, etc.)	-
j	Loop variable (e.g. numerical value, undefined chemical species, etc.)	-
J	Idempotential	-
K	Fitting parameter of the Bi-Langmuir function	-
K_H	Henry's constant	1/Pa
K_L	Langmuir's constant	1/Pa
k	Number of adsorption sites a k-mer occupies	-
k_B	Boltzmann's constant ($1.381 \cdot 10^{-23}$ J/K)	J/K
$\tilde{k}$	Number of trial positions in CBMC	-
$\vec{k}$	Reciprocal space vector	1/Å
L	Length of contact line	m
$L\{\vec{r}_N\}$	Total number of configurations	-
$\vec{L}$	Angular velocity	rad/s
l	Number of adsorption sites a l-mer occupies	-
$\check{l}$	Number of simulation box replica	-
$\tilde{l}$	Number of trial positions in CBMC	-
$l\{\vec{r}_N\}$	Number of occurrences of any configuration	-

M	Mechanical property of a system	-
$\tilde{M}$	Number of adsorption sites	-
$\check{M}$	Number of chain segments in CBMC	-
m	Mass	kg
N	Number of objects	-
n	Amount of substance	-
$\tilde{n}$	Quantum number	-
$\vec{n}$	Box vector	Å
P	Probability	-
$\mathrm{P_{acc}}$	Acceptance probability	-
P	Momentum	kg·m/s
$\vec{P}$	Momentum vector	kg·m/s
p	Pressure	Pa
p_0	Vapor pressure (above flat interface)	Pa
Q	Fictitious mass	kg
q	Quaternion	-
$\tilde{q}$	Partial charge	C
R	Universal gas constant	J/(K·mol)
r	Radius	m
$\tilde{r}$	Roughness ratio	-
$\bar{r}$	Mean of compensation charge cloud	Å
$\vec{r}$	Position	Å
S	Entropy	J/K
$\bar{s}$	Additional degree of freedom in the Nosé-Hoover thermostat	-
$\tilde{s}$	Integer in the Ewald summation	-
T	Temperature	K
$\vec{T}$	Torque	N·m
t	Time	s
U	Internal energy	J
u	One-electron potential	J
V	Volume	$\mathrm{m^3}$
v	Speed	m/s
$\vec{v}$	Velocity	m/s
W	Total number of available microscopic states	-
$\check{W}$	Wiener noise	-
w	Weight	-
$\tilde{w}$	Rosenbluth weight for chain segment	
x	Mole fraction	-
Z	System partition function	-
z	Particle partition function	-
α	Proposal probability	-
$\alpha_{\mathcal{L}}$	Lagrangian multiplier	-
$\beta_{\mathcal{L}}$	Lagrangian multiplier	1/J
Γ	Mass specific adsorbed amount	mol/kg
$\gamma_{\mathcal{L}}$	Lagrangian multiplier	1/particle

δ	Wildcard (numerical value)	-
ϵ_0	Vacuum permittivity ($8.854 \cdot 10^{-12}$ $A^2s^4/(kg \cdot m^3)$)	$A^2s^4/(kg \cdot m^3)$
ϵ	Lennard-Jones well-depth	J/mol
ε	Particle energy	J
ζ	Random number	-
Θ	Contact angle	deg
$\tilde{\Theta}$	Occupancy of adsorption sites	-
θ	Bond angle	deg
κ	Scaling factor in velocity rescaling thermostats	-
Λ	de-Broglie wavelength	m
λ	Scaling factor	-
$\lambda_{\mathcal{L}}$	Lagrangian multipliers	-
μ	Chemical potential	J/particle
ν	Number of adsorbate molecules that can simultaneously occupy an adsorption site	-
ξ	Friction coefficient in the Nosé-Hoover thermostat	1/s
$\tilde{\pi}$	Spreading pressure	J/m^2
$\bar{\pi}$	Transition probability	-
ϖ	Ewald splitting parameter	1/Å
ρ	Density	kg/m^3
σ	Surface tension	J/m^2
$\bar{\sigma}$	Holonomic constraint	-
$\tilde{\sigma}$	Lennard-Jones distance	Å
τ	Coupling parameter in the CSVR algorithm	s
Υ	Position of segment for o in CBMC	Å
Φ	Torsion angle	deg
φ	Atomic orbital	-
ϕ	Electrostatic potential	J
χ	Electronegativity	-
Ψ	Eigenfunction describing the quantum state of a system	-
ψ	Molecular orbital/Eigenfunction	-
Ω	Entropy factor	-
ω	Rotational angle	deg
$\omega\{N_s\}$	Number of realizations	-
$\vec{\omega}$	Angular velocity	1/s
$\mathcal{A}$	Number of systems in Gibbs' thought experiment	-
$\mathcal{E}$	Total energy of the collection of systems	J
$\mathcal{H}$	Hamiltonian	J
$\hat{\mathcal{H}}$	Hamilton operator	-
$\mathcal{K}$	Kinetic energy	J
$\hat{\mathcal{K}}$	Kinetic energy operator	-
$\mathcal{L}$	Lagrangian auxiliary function	-
$\mathcal{N}$	Total number of systems in the collection	-
$\mathcal{U}$	Potential energy	J
$\hat{\mathcal{U}}$	Potential energy operator	-
$\mathcal{W}$	Rosenbluth weight for fully grown chain	-

Subscripts

Subscript	Meaning
0	Ground state
A	Chemical species A
a	Adsorbate
acc	Accuracy
app	Apparent
B	Chemical species B
b	Beginning
bend	Bond angle bending
bond	Bond stretching
C	Chemical species C
cc	Circle center
class	Classical
com	Center of mass
cut	Cutoff
D	Chemical species D
Dc	Convex droplet curvature
Drop	Droplet
e^-	Electron
e	Ending
el	Electronic
elec	Electrostatic
eq	Equilibrium
ex	Excess
f	Degrees of freedom
g	Gas
H	Hartree
i	Loop variable
i	interface
ic	Ideal chain
id	Ideal gas
intra	Intramolecular interactions
j	Loop variable
k	k-mer
known	Known contributions to energy
LJ	Lennard-Jones
l	Liquid
l	l-mer
m	Molar
m	Segment number in CBMC
max	Maximum
mix	Mixture

N	Number of particles
Nosé	Nosé thermostat
nb	Non-bonded
$\tilde{n}$	Quantum number
n^+	Nuclei
P	Momentum
p_i	Grid point
r	Position
rot	Rotation
S	System state
s	Solid
s	Particle state
$\bar{s}$	Additional degree of freedom in the Nosé-Hoover thermostat
sf	Surface
sp	Specific
sys	System
tor	Torsion
t	Target
trans	Translation
tot	Total
v	vapor
vdW	van-der-Waals
XC	Exchange-correlation energy contribution
x	x-component
y	y-component
z	z-component
$\tilde{z}$	Position of GDS
λ	CFCMC scaling factor
θ	Bond angle bending

Superscripts

Superscript	Meaning
calc	Predicted
e-gas	Uniform electron gas
ext	Intermolecular interactions
g	Gas phase
int	Intramolecular interactions
m	Saturation
α	Bulk phase of a system
β	Bulk phase of a system
γ	Interfacial Layer
σ	Adsorption excess
$\sigma(n)$	Reduced surface excess
$\sigma(s)$	Relative surface excess
0	Pure
I	Phase I in GDS formalism
II	Phase II in GDS formalism
$\prime$	Reference system
$*$	Most Probable distribution
†	Indistinguishable

1 Motivation

> *I would like to start by emphasizing the importance of surfaces. It is at a surface where many of our most interesting and useful phenomena occur. We live for example on the surface of a planet. It is at a surface where the catalysis of chemical reactions occur. It is essentially at a surface of a plant that sunlight is converted to a sugar. [...]*
>
> Walter H. Brattain, in his Noble Lecture, 1956

This extract from the Nobel Lecture by Walter H. Brattain, who together with William B. Shockley and John Bardeen received the Nobel Prize in Physics for their work on semiconductors, gives an impression of the importance of surfaces and the phenomena that occur on them. Surface phenomena are an integral part of everyday life – whether in the appearance of bubbles in the sink after washing one's hands or in the design of water-repellent clothing. Surface phenomena also find application in industrial processes, such as catalysis, fluid purification, or separation. For industrial application, materials with huge surface-to-volume ratios are preferred. Solids with pores in the nanometer range (i.e. *nanoporous solids*) are such materials, and of these, metal-organic frameworks are the most versatile class. Metal-organic frameworks have already received a high level of attention. The modular structure – MOFs consist of inorganic nodal building blocks that are connected by organic linking building blocks – allows almost continuous adjustment of pore size, shape, and environment. However, many aspects of surface phenomena in or on metal-organic frameworks are not yet fully understood. For example, it is known that entropy favors the accumulation of smaller guest molecules in nanoporous solids at high loading. But does entropy also favor the accumulation of water in metal-organic frameworks with internal hydrophobicity? Speaking of which, how is the hydrophobicity of the internal and external surface of metal-organic frameworks related? And how can modern visualization techniques, such as virtual reality, help in studying metal-organic frameworks and the guest molecules within them?

This thesis aims to shed light on these questions using classical molecular simulations. Molecular simulations are a helpful tool for studying surface phenomena, because they can complement experiments by providing insights at the microscopic level, and offer the possibility of exploring surface phenomena that can only rarely be investigated in experiments, plus help to improve the efficiency of experiments by predicting metal-organic frameworks with desired properties.

After describing the theoretical background in Section 2, the following research studies are presented:

1. The investigation of the impact of entropic effects on the separation of aqueous alcohol mixtures in CAU-10 (Section 3) using Monte Carlo simulations.

2. The study of the relationship between internal and external hydrophobicity of metal-organic frameworks (Section 4) using Monte Carlo and molecular dynamics simulations.

3. The development of MOF-VR: a virtual reality tool to study guest molecules in metal-organic frameworks (Section 5).

Each research study is introduced by a detailed motivation and an overview of the state of the research. Finally, the results are summarized in relation to each other in Section 6.

2 Theory Section

2.1 Interfacial Phenomena

Interfaces – whether fluid-fluid, fluid-solid, or solid-solid interfaces – give rise to a number of phenomena, which arise from the peculiar energy state of the surface particles shown in Fig. 1. In condensed phase, each particle is surrounded by a large number of other particles with which it shares attractive interactions (see Fig. 1a). The net force acting on any particle in the solid is zero. If we were now to suddenly remove the top layers of the solid, the particles in the resulting surface region would lack part of their attractive interactions (see Fig. 1b), and the net force acting on these particles would be unbalanced. As a result, the particles in the surface region would now be pulled inward until the net force acting on each particle is balanced by the force due to the compression resistance of the bulk solid (see Fig. 1c).

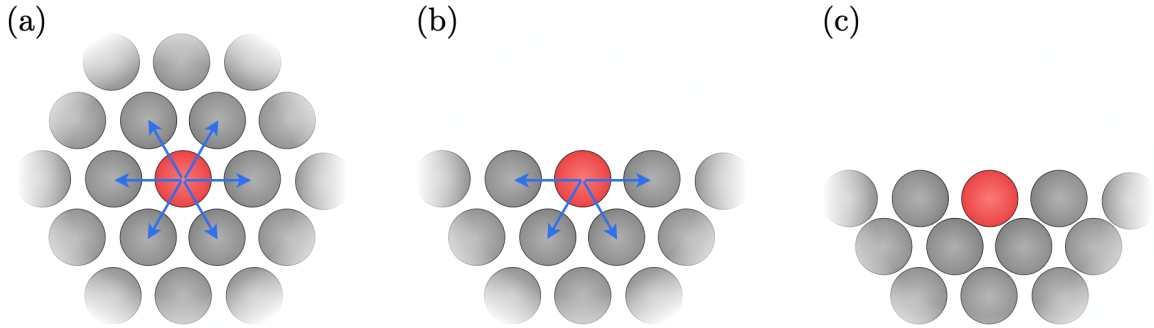

Fig. 1: Illustration of a solid surface. (a) The net force acting on any particle (colored red) in the bulk solid is zero (blue force vectors; only shown for the nearest neighbors). (b) If the top layers of the solid were suddenly removed, the surface particles would lack part of their attractive interactions, and (c) be pulled inward. Source: author's illustration inspired by Coussy, 2010, p. 108 [1].

The driving force for minimizing the interfacial area is the reduction of the internal energy of the system,

$$\mathrm{d}U = T\mathrm{d}S - p\mathrm{d}V + \sigma\mathrm{d}A_\mathrm{i} + \sum_i \mu_i \mathrm{d}n_i, \tag{1}$$

where T is the temperature, S is the entropy, p is the pressure, V is the volume, σ is the surface tension, A_i is the interfacial area, μ_i is the chemical potential of chemical species i, and n_i is the amount of i. The term $\sigma\mathrm{d}A_\mathrm{i}$ represents the work performed by changing the interfacial area. Many interfacial phenomena can be traced back to surface tension. In the following sections, we will discuss:

1) the adsorption of gases on solid surfaces (Sec. 2.1.1),

2) the adsorption of liquid mixtures on solid surfaces (Sec. 2.1.2), and

3) the wetting of solid surfaces (Sec. 2.1.3).

2.1.1 Adsorption of Gases on Solid Surfaces

Adsorption is a ubiquitous phenomenon that leads to an increase in fluid density near an interface and is driven by the reduction of surface tension. Industrial applications of adsorption include drying of gases, separation of fluids, or use in catalysis. As the extent of adsorption depends on the size of the interface, industrial adsorption materials possess large surface-to-volume ratios, such as nanoporous solids. Below, we will establish a relationship between adsorption and surface tension before turning to the special case of adsorption of gases on solid surfaces. In the process, we will draw on terms that are commonly used in the field, which are listed in Tab 1.

Tab. 1: Common terms used in the field of adsorption on solid surfaces [2, 3].

Term	**Explanation**
adsorption/desorption	accumulation/exclusion of matter near interfaces
adsorbent	solid on which adsorption occurs
adsorptive	substance that can be adsorbed (e.g. gas)
adsorbate	adsorbed matter
micropore	pore with width $\leq$ 2 nm
ultramicropore	micropore with width $<$ 0.7 nm
mesopore	pore with width $>$ 2 nm and $<$ 50 nm
macropore	pore with width $\geq$ 50 nm
nanopore	pore with width $\leq$ 100 nm
external surface	surface outside pores of porous adsorbent
internal surface	surface of all pores of porous adsorbent

Note that we will focus on *physisorption*, which deals only with adsorption effects that do not involve the formation of chemical bonds – in contrast to *chemisorption*. We will establish a relationship between adsorption and surface tension by using the example of a heterogeneous system that consists of a solid phase, α, a gas phase, β, and an interfacial layer, γ. The integral version of Eq. (1) vanishes for the bulk phases (i.e. α or β), but not for the interfacial layer, γ ([4], p. 14),

$$U^\alpha - TS^\alpha + pV^\alpha - \sum_i \mu_i n_i^\alpha = 0, \tag{2}$$

$$U^\beta - TS^\beta + pV^\beta - \sum_i \mu_i n_i^\beta = 0, \tag{3}$$

$$U^\gamma - TS^\gamma + pV^\gamma - \sum_i \mu_i n_i^\gamma = A_\mathrm{i}\sigma. \tag{4}$$

By differentiation of Eq. (4),

$$A_\mathrm{i}\mathrm{d}\sigma + \sigma\mathrm{d}A_\mathrm{i} = \mathrm{d}U^\gamma - S^\gamma\mathrm{d}T - T\mathrm{d}S^\gamma + V^\gamma\mathrm{d}p + p\mathrm{d}V^\gamma - \sum_i \mu_i \mathrm{d}n_i^\gamma - \sum_i n_i^\gamma \mathrm{d}\mu_i, \tag{5}$$

and comparison with the interfacial analog of Eq. (1),

$$\sigma\mathrm{d}A_\mathrm{i} = \mathrm{d}U^\gamma - T\mathrm{d}S^\gamma + p\mathrm{d}V^\gamma - \sum_i \mu_i \mathrm{d}n_i^\gamma, \tag{6}$$

we obtain:

$$0 = A_\mathrm{i}\mathrm{d}\sigma = -S^\gamma\mathrm{d}T + V^\gamma\mathrm{d}p - \sum_i n_i^\gamma \mathrm{d}\mu_i, \tag{7}$$

which applies to any kind of interfacial layer. For the adsorption of a pure gas on a solid surface at constant p and T, Eq. (7) reduces to:

$$A_\mathrm{i}\mathrm{d}\sigma = -n_\mathrm{g}^\gamma \mathrm{d}\mu_\mathrm{g}, \tag{8}$$

where n_g^γ is the amount of gas located in the interfacial layer, μ_g is the chemical potential of the gas, and the solid is assumed to be inert ($\mathrm{d}\mu_\mathrm{s} = 0$). By dividing Eq. (8) by the surface area, we obtain the relationship between adsorption and surface tension – often referred to as *Gibbs adsorption isotherm* ([4], p. 30),

$$\mathrm{d}\sigma = -\left[n_\mathrm{g}^\gamma\right] \mathrm{d}\mu_\mathrm{g}, \tag{9}$$

where $\left[n_\mathrm{g}^\gamma\right]$ is the areal amount of gas in the interfacial layer.

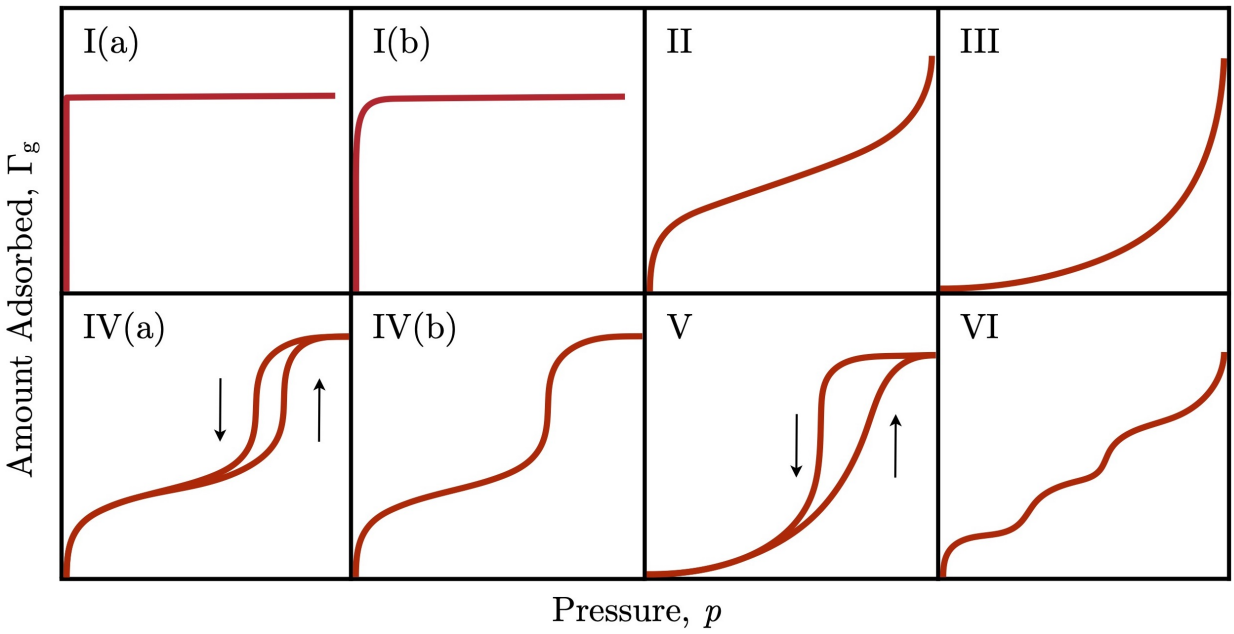

Fig. 2: Classification of physisorption isotherms according to IUPAC. Source: author's illustration inspired by IUPAC, 2015, p. 8 [3].

Gas adsorption on solid surfaces is commonly expressed quantitatively by plotting the pressure-dependent increase in the (mass) specific adsorbed amount of gas,

$$\Gamma_{\mathrm{g}} = \left[n_{\mathrm{g}}^{\gamma}\right] A_{\mathrm{sp}}, \tag{10}$$

where A_{sp} is the specific surface. These plots are recorded at constant temperature and thus called *adsorption isotherms*. The *International Union of Pure and Applied Chemistry* (IUPAC) [3] distinguishes six (plus two) types of adsorption isotherms (see Fig. 2):

- **Type I** isotherms are often found for the adsorption of gases on solids with narrow pores. This type of isotherm is characterized by a sharp uptake of adsorbate at low pressure (due to the high adsorption energy in narrow pores) and the formation of a plateau when approaching the vapor pressure, p_0, of the adsorptive (due to the low external surface area). Type I isotherms can be divided into

  - Type I(a) isotherms for narrow micropores (< 1 nm), and
  - Type I(b) isotherms for wider micropores and narrow mesopores (< 2.5 nm),

  which differ primarily in the steepness of the slope for low pressures. A type I isotherm can be found for the adsorption of ethanol in CAU-10 (see Sec. 3.4.1).

- **Type II** isotherms are often found for the adsorption of gases on non-porous or macroporous solids. This type of isotherm is characterized by a sharp uptake at low pressures (due to the formation of an adsorbate monolayer) and a subsequent linear range up to pressures close to p_0. In this linear range, the accumulation of matter increases progressively (due to the formation of a multilayer of adsorbate). A type II isotherm can be found for the adsorption of nitrogen in CAU-10 [5].

- **Type III** isotherms are found for the adsorption of gases on solid surfaces with weak adsorbate-adsorbent interactions. This type of isotherm is characterized by a poor increase at low pressures (indicating that the formation of a monolayer is not preferred) and a steep increase for pressures close to p_0 (multilayer formation). Type III isotherms are rare.

- **Type IV** isotherms are often found for the adsorption of gases in mesoporous solids. They are similar to type II isotherms, but differ in the occurrence of a plateau at high pressures. Type IV isotherms can be divided into:

  - Type IV(a) isotherms for mesopores exceeding a critical width (e.g. about 4 nm for nitrogen), and
  - Type IV(b) isotherms for small mesopores,

  which differ in the occurrence of *adsorption hysteresis* – i.e. adsorption and desorption branch are not congruent – caused by capillary condensation in wider mesopores. A type IV isotherm can be found for the adsorption of nitrogen in SBA-15 [6].

- **Type V** isotherms are often found for the adsorption of gases on porous materials with weak adsorbate-adsorbent interactions. This type of isotherm is similar to a type III isotherm, except for the plateau at high pressures and the occurrence of adsorption hysteresis. A type V isotherm can be found for the adsorption of water in CAU-10 [7].
- **Type VI** isotherms are found for the adsorption of gases on highly uniform solid surfaces. This type of isotherm is characterized by a stepwise formation of adsorbate layers. Type VI isotherms are rare.

Determining Γ_g requires knowledge on the composition of the interfacial layer. For adsorption in nanoporous solids, we can safely neglect the influence of the external surface. In adsorption simulations, we can thus determine Γ_g^γ by counting the particles in the pores, providing n_g^γ, and dividing the result by the adsorbent's mass. But this is hardly feasible in the experiment. When we wish to compare simulated gas adsorption isotherms with experimental ones, we need to define what is measured in the experiment.

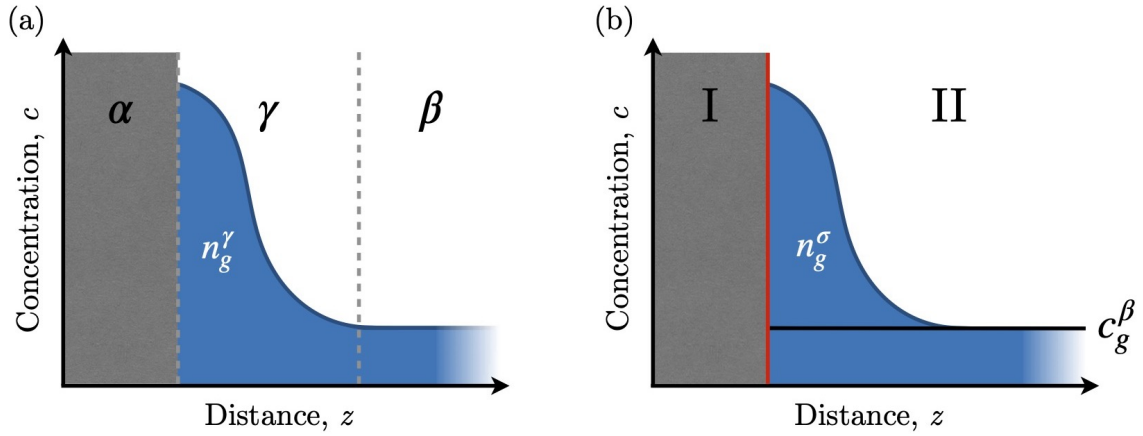

Fig. 3: Accumulation of gas on a solid surface. (a) Real system arbitrarily divided in solid phase, α, interfacial layer, γ, and gas phase, β. (b) The Gibbs dividing surface (red) separates the system in phase I and II. Source: author's illustration inspired by Rouquerol et al., 1999, p. 30 [2].

Fig. 3 illustrates the concentration profile for a single component gas near a solid surface. In consistency with the conventions used in Eqs. (2) - (4), we divide the system into three phases: a solid phase, α, a gas phase, β, and an interfacial layer, γ. In the following, we will assume that there is no gas accumulation in the solid phase – i.e. there is no *absorption*. The concentration profile in Fig. 3a shows an increased gas concentration in the interfacial layer, γ, which decreases progressively with increasing distance from the solid surface until the concentration of the bulk gas phase is reached. We can define n_g^γ as being the difference of the total amount of gas in the system, n_g, and the amount of gas in the bulk gas phase, $c_g^\beta V^\beta$,

$$n_g^\gamma = n_g - c_g^\beta V^\beta, \tag{11}$$

where c_g^β is the gas concentration profile and V^β is the volume of the bulk gas phase, β. Eq. (11) is difficult to evaluate because we are often unable to determine the concentration profile and/or unambiguously separate the interfacial layer, γ, and the gas phase, β, due to

the asymptotic approximation of the concentration profile to the bulk gas concentration. As a consequence, we often apply a technique called *Gibbs dividing surface* (GDS) illustrated in Fig. 3b. The fundamental idea is to compare the real system with a reference system of same volume in which c_{g}^{β} is constant up to the GDS, i.e. there is no gas accumulation near the solid surface in the reference system. The difference between the total amount of gas in the real system, n_{g}, and the total amount of gas in the reference system, n'_{g}, is dubbed as the *surface excess amount of gas* ([4], p. 8),

$$n_{\mathrm{g}}^{\sigma} = n_{\mathrm{g}} - n'_{\mathrm{g}} = n_{\mathrm{g}} - (c_{\mathrm{g}}^{\mathrm{I}} V^{\mathrm{I}} + c_{\mathrm{g}}^{\mathrm{II}} V^{\mathrm{II}}), \tag{12}$$

where $c_{\mathrm{g}}^{\mathrm{I}}$ and $c_{\mathrm{g}}^{\mathrm{II}}$ are the gas concentrations, and V^{I} and V^{II} are the volumes of phase I or II of the reference system. Eq. (12) indicates that the surface excess amount of gas depends on the position of the GDS and is thus no unambiguously defined physical quantity. However, certain linear combinations of the surface excess amount for two of more components are independent of the location of the GDS. One of them is referred to as the *relative adsorption excess* of one component with respect to another component, and is often used for the quantitative expression of gas adsorption on solid surfaces. We define the relative adsorption excess by determining the surface excess amount of solid,

$$n_{\mathrm{s}}^{\sigma} = n_{\mathrm{s}} - (c_{\mathrm{s}}^{\mathrm{I}} V^{\mathrm{I}} + c_{\mathrm{s}}^{\mathrm{II}} V^{\mathrm{II}}), \tag{13}$$

where n_{s} is the total amount of solid in the system, and $c_{\mathrm{s}}^{\mathrm{I}}$ and $c_{\mathrm{s}}^{\mathrm{II}}$ are the concentration of solid in phase I or II. Using $V^{\mathrm{I}} = V - V^{\mathrm{II}}$, Eq. (12) and Eq. (13) reduce to:

$$n_{\mathrm{g}}^{\sigma} = n_{\mathrm{g}} - c_{\mathrm{g}}^{\mathrm{I}} V + (c_{\mathrm{g}}^{\mathrm{I}} - c_{\mathrm{g}}^{\mathrm{II}}) V^{\mathrm{II}}, \tag{14}$$

$$n_{\mathrm{s}}^{\sigma} = n_{\mathrm{s}} - c_{\mathrm{s}}^{\mathrm{I}} V + (c_{\mathrm{s}}^{\mathrm{I}} - c_{\mathrm{s}}^{\mathrm{II}}) V^{\mathrm{II}}, \tag{15}$$

in which only V^{II} depends on the position of the GDS. To eliminate this dependency, we multiply Eq. (15) with $(c_{\mathrm{g}}^{\mathrm{I}} - c_{\mathrm{g}}^{\mathrm{II}})/(c_{\mathrm{s}}^{\mathrm{I}} - c_{\mathrm{s}}^{\mathrm{II}})$,

$$n_{\mathrm{s}}^{\sigma} \frac{c_{\mathrm{g}}^{\mathrm{I}} - c_{\mathrm{g}}^{\mathrm{II}}}{c_{\mathrm{s}}^{\mathrm{I}} - c_{\mathrm{s}}^{\mathrm{II}}} = (n_{\mathrm{s}} - c_{\mathrm{s}}^{\mathrm{I}} V) \frac{c_{\mathrm{g}}^{\mathrm{I}} - c_{\mathrm{g}}^{\mathrm{II}}}{c_{\mathrm{s}}^{\mathrm{I}} - c_{\mathrm{s}}^{\mathrm{II}}} + (c_{\mathrm{g}}^{\mathrm{I}} - c_{\mathrm{g}}^{\mathrm{II}}) V^{\mathrm{II}}, \tag{16}$$

and subtract the result from Eq. (14),

$$n_{\mathrm{g}}^{\sigma} - n_{\mathrm{s}}^{\sigma} \frac{c_{\mathrm{g}}^{\mathrm{I}} - c_{\mathrm{g}}^{\mathrm{II}}}{c_{\mathrm{s}}^{\mathrm{I}} - c_{\mathrm{s}}^{\mathrm{II}}} = n_{\mathrm{g}} - c_{\mathrm{g}}^{\mathrm{I}} V - (n_{\mathrm{s}} - c_{\mathrm{s}}^{\mathrm{I}} V) \frac{c_{\mathrm{g}}^{\mathrm{I}} - c_{\mathrm{g}}^{\mathrm{II}}}{c_{\mathrm{s}}^{\mathrm{I}} - c_{\mathrm{s}}^{\mathrm{II}}}. \tag{17}$$

The right-hand side of Eq. (17) consists of unambiguously defined quantities, which are independent of the position of the GDS. As a result, the relative surface excess of gas with respect to the solid,

$$n_{\mathrm{g}}^{\sigma(s)} = n_{\mathrm{g}}^{\sigma} - n_{\mathrm{s}}^{\sigma} \frac{c_{\mathrm{g}}^{\mathrm{I}} - c_{\mathrm{g}}^{\mathrm{II}}}{c_{\mathrm{s}}^{\mathrm{I}} - c_{\mathrm{s}}^{\mathrm{II}}}, \tag{18}$$

is also independent of the position of the GDS. Division by the mass of the adsorbent leads to the *specific relative surface excess* of gas with respect to the solid ([4], p. 9),

$$\Gamma_{\mathrm{g}}^{\sigma(s)} = \Gamma_{\mathrm{g}}^{\sigma} - \Gamma_{\mathrm{s}}^{\sigma} \frac{c_{\mathrm{g}}^{\mathrm{I}} - c_{\mathrm{g}}^{\mathrm{II}}}{c_{\mathrm{s}}^{\mathrm{I}} - c_{\mathrm{s}}^{\mathrm{II}}}. \tag{19}$$

Commonly, the position of the GDS, $\tilde{z}$, is chosen so that $\Gamma_{\mathrm{s}}^{\sigma} = 0$ and, as a consequence,

$$\Gamma_{\mathrm{g}}^{\sigma(s)} = \left(\Gamma_{\mathrm{g}}^{\sigma}\right)_{\tilde{z}}, \tag{20}$$

which corresponds to placing the GDS directly on the solid surface, so that the difference between n_{g}^{γ} (determined in adsorption simulations) and $n_{\mathrm{g}}^{\sigma(s)}$ (measured in adsorption experiments) is provided by:

$$n_{\mathrm{g}}^{\gamma} - n_{\mathrm{g}}^{\sigma} = c_{\mathrm{g}}^{\beta} V^{\gamma}. \tag{21}$$

Unless otherwise stated, gas adsorption isotherms shown in the following sections plot the absolute adsorbed amount of a gas as a function of pressure.

2.1.2 Adsorption from Liquid Mixtures at Solid Surfaces

Another common interfacial phenomenon is the adsorption from liquid mixtures, where we often observe the preferential adsorption of one or more components of the mixture. The preferential adsorption is driven by the reduction of surface tension,

$$\mathrm{d}\sigma = -\sum_i [n_i^{\gamma}]\, \mathrm{d}\mu_i, \tag{22}$$

which is tantamount to Eq. (7) for constant p and T. In Eq. (22) we are to determine the composition of the interfacial layer (given by $n_1^{\gamma}, n_2^{\gamma}, \dots$) that minimizes the surface tension. Similar to gas adsorption, it is often impossible to determine the composition of the interfacial layer in an experiment. We use thus the Gibbs dividing surface to define the *reduced surface excess*. Again, we compare the real system to a hypothetical one in which the concentrations of the components in the bulk liquid phase are assumed to be constant up to the GDS. The reduced surface excess is obtained by considering the surface excess of component i,

$$n_i^{\sigma} = n_i - c_i^{\mathrm{I}} V + (c_i^{\mathrm{I}} - c_i^{\mathrm{II}}) V^{\mathrm{II}}, \tag{23}$$

where n_i is the total amount of i in the real system and c_i^{I} and c_i^{II} are the concentrations of i in phase I or II of the reference system, and the total surface excess,

$$n^{\sigma} = n - c^{\mathrm{I}} V + (c^{\mathrm{I}} - c^{\mathrm{II}}) V^{\mathrm{II}}, \tag{24}$$

where n is the total amount of matter in the real system, and c^{I} and c^{II} are the total concentrations of the components in phase I or II of the reference system. In both equations, only V^{II}, depends on the position of the GDS. To eliminate this dependency, we multiply

Eq. (24) by $(c_i^{\mathrm{I}} - c_i^{\mathrm{II}})/(c^{\mathrm{I}} - c^{\mathrm{II}})$,

$$n^\sigma \frac{c_i^{\mathrm{I}} - c_i^{\mathrm{II}}}{c^{\mathrm{I}} - c^{\mathrm{II}}} = (n - c^{\mathrm{I}} V) \frac{c_i^{\mathrm{I}} - c_i^{\mathrm{II}}}{c^{\mathrm{I}} - c^{\mathrm{II}}} + (c_i^{\mathrm{I}} - c_i^{\mathrm{II}}) V^{\mathrm{II}}, \tag{25}$$

and subtract the result from Eq. (23),

$$n_i^\sigma - n^\sigma \frac{c_i^{\mathrm{I}} - c_i^{\mathrm{II}}}{c^{\mathrm{I}} - c^{\mathrm{II}}} = n_i - c_i^{\mathrm{I}} V - (n - c^{\mathrm{I}} V) \frac{c_i^{\mathrm{I}} - c_i^{\mathrm{II}}}{c^{\mathrm{I}} - c^{\mathrm{II}}}. \tag{26}$$

The right-hand side of Eq. (26) contains only unambiguously defined quantities which are independent of the position of the GDS. Thus, the reduced surface excess of i with respect to the total amount of matter,

$$n_i^{\sigma(n)} = n_i^\sigma - n^\sigma \frac{c_i^{\mathrm{I}} - c_i^{\mathrm{II}}}{c^{\mathrm{I}} - c^{\mathrm{II}}}, \tag{27}$$

is also independent of the position of the GDS. Division by the mass of the adsorbent leads to the *specific reduced surface excess* of i ([4], p. 9),

$$\Gamma_i^{\sigma(n)} = \Gamma_i^\sigma - \Gamma^\sigma \frac{c_i^{\mathrm{I}} - c_i^{\mathrm{II}}}{c^{\mathrm{I}} - c^{\mathrm{II}}}. \tag{28}$$

Commonly, the position of the GDS, $\tilde{z}$, is chosen so that

$$\Gamma^\sigma = \sum_i \Gamma_i^\sigma = 0, \tag{29}$$

and, thus,

$$\Gamma_i^{\sigma(n)} = (\Gamma_i^\sigma)_{\tilde{z}}\,. \tag{30}$$

This choice of the GDS has three major consequences:

1. Reference and real system have the same total amount of fluid.

2. The specific reduced surface excess of i is tantamount to the difference of the specific total amount of i in the real system, Γ_i, and the specific total amount of i in the reference system, Γ_i',

   $$\Gamma_i^{\sigma(n)} = \Gamma_i - \Gamma_i' = \frac{n}{m_{\mathrm{a}}}(x_i - x_i'), \tag{31}$$

   where n is the total amount of fluid in the mixture, x_i is the (total) mole fraction of i in the real system – i.e. corresponding to the mole fraction of i in the pores (simulation) or the initial liquid mixture (experiment) –, and x_i' is the mole fraction of i in the reference system – i.e. the equilibrium mole fraction of i in the liquid mixture.

3. For binary liquid mixtures, the surface excess of component 1 provides the surface excess of component 2,

   $$\Gamma_1^\sigma = -\Gamma_2^\sigma. \tag{32}$$

The reduced surface excess for binary liquid mixtures is commonly plotted against the equilibrium mole fraction of one of the components in the liquid phase. IUPAC [8] classifies these *excess isotherms* into two main groups: U-shaped and S-shaped isotherms. U-shaped isotherms indicate the preferential adsorption of one component over the whole concentration range, while S-shaped isotherms indicate a selectivity reversal.

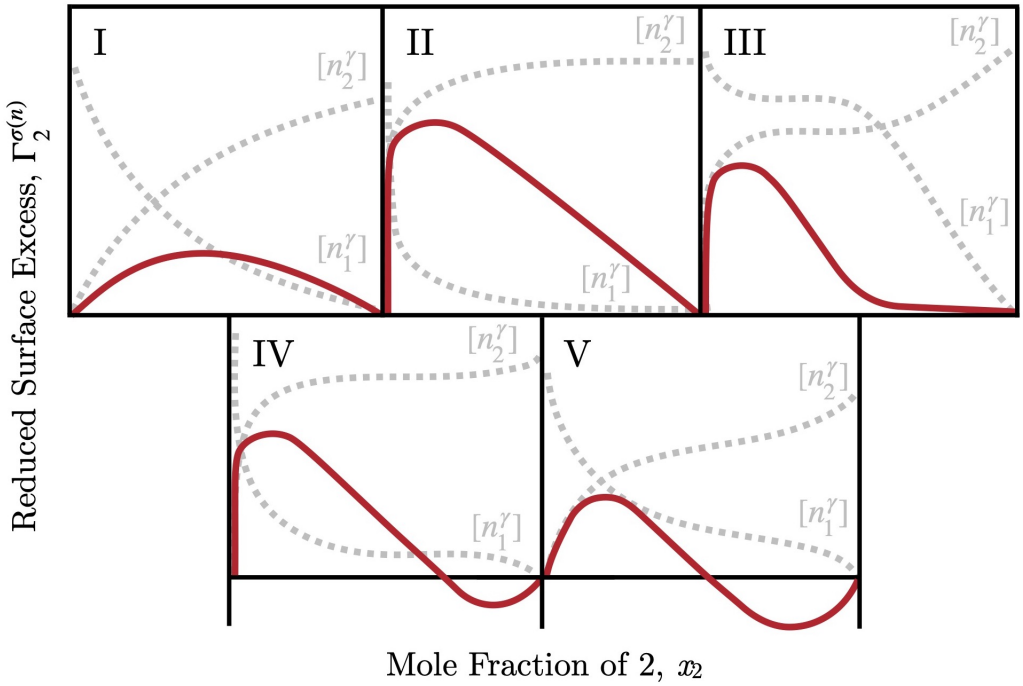

Fig. 4: Reduced surface excess isotherms (red) of binary liquid mixtures and exemplary amount n_1^γ and n_2^γ in the interfacial layer (gray). Source: author's illustration inspired by Nagy and Schay, 1963, p. 370 [9].

Nagy and Schay [9] further divided U-shaped and S-shaped excess isotherms into a total of five types (see Fig. 4):

- **Type I** excess isotherms are U-shaped (Fig. 4a). They indicate the preferential adsorption of one component in the mixture over the whole concentration range. A type I excess isotherm can be found for the adsorption of a mixture of benzene (1) and 1,2-dichloroethane (2) on alumina [9].

- **Type II** excess isotherms are U-shaped (Fig. 4b). They indicate a strong preference for the adsorption of one component in the mixture. Within a certain concentration range, the preferential adsorption of one component leads to the exclusion of the other one from the adsorbed layer. A type II excess isotherm can be found for the adsorption of a mixture of n-octane (1) and ethanol (2) on SBA-16 [10].

- **Type III** excess isotherms are U-shaped (Fig. 4c). They indicate the preferential adsorption of one component in the mixture. Within a certain concentration range, the composition of the mixture in the interfacial layer approaches the composition of the mixture in the bulk liquid phase. A type III excess isotherm can be found for the adsorption of a mixture of water (1) and ethanol (2) on charcoal [9].

- **Type IV** excess isotherms are S-shaped (Fig. 4d). Within a certain concentration range, the composition of the mixture in the interfacial layer is constant until a reversal of selectivity occurs. The zero of the excess isotherm is called the *azeotropic point*. At this point, the composition of the interfacial layer equals the one in the bulk liquid layer. A type IV excess isotherm can be found for the adsorption of a mixture of ethanol (1) and benzene (2) on charcoal [9].

- **Type V** excess isotherms are S-shaped (Fig. 4e). The composition of the mixture in the interfacial layer is not constant in any part of the concentration range. A type V excess isotherm can be found for the adsorption of a mixture of methyl acetate (1) and benzene (2) on charcoal [9].

2.1.3 Contact Angles of Liquids at Solid Surfaces

Understanding and controlling the wettability of solid surfaces is important in many industries, such as in the production of paints, stain-resistant textiles, or water-repellent eyeglass lenses ([11], pp. 1–2). Wettability can be evaluated by bringing a liquid droplet into contact with a solid surface. The liquid droplet will either spread completely, partially, or not at all. In the case of partial spreading, a sessile droplet with a certain angle between the tangent of the droplet and the solid surface forms at equilibrium. This angle is often referred to as *contact angle*, Θ, and is the result of three competing forces (illustrated in Fig. 5a), namely the desire to decrease:

1. the solid-liquid interface to reduce the solid-liquid surface tension, $\sigma_{s/l}$,
2. the solid-vapor interface to reduce the solid-vapor surface tension, $\sigma_{s/v}$, and
3. the liquid-vapor interface to reduce the liquid-vapor surface tension, $\sigma_{l/v}$.

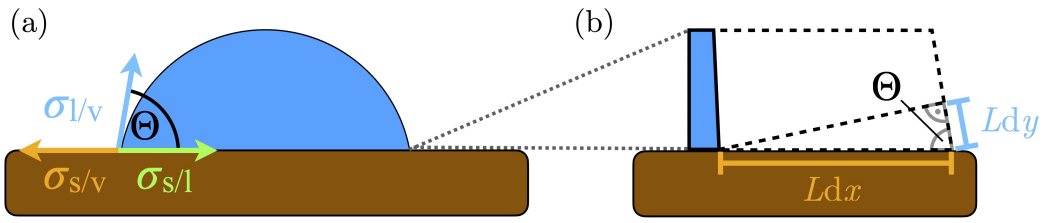

Fig. 5: (a) The contact angle of a sessile droplet is the result of three competing surface tensions: the solid-liquid surface tension, $\sigma_{s/l}$, the solid-vapor surface tension, $\sigma_{s/v}$, and the liquid-vapor surface tension, $\sigma_{l/v}$. (b) The relationship between surface tension and contact angle can be derived by displacing the contact line of the droplet by dx. An area $L\mathrm{d}x$ of the solid surface would be uncovered to the vapor phase, while the liquid-vapor phase would be reduced by an area $L\mathrm{d}y = L\cos\Theta\mathrm{d}x$. Source: author's illustration inspired by (a) Law and Zhao, 2016, p. 36 [11], and (b) the explanations by Safran, 2003, pp. 98–99. [12].

In 1805, Thomas Young [13] derived the relationship between the contact angle and the surface tensions,

$$\cos\Theta = \frac{\sigma_{s/v} - \sigma_{s/l}}{\sigma_{l/v}}, \tag{33}$$

based on the assumption of *mechanical equilibrium*.[1] But Young's equation can also be derived based on thermodynamic reasoning, as shown by Safran ([12], pp. 98–99). For this, we assume the system to be at constant p, T, and n. How would the Gibbs free energy change if we displaced the contact line by dx in the x-direction? This situation is illustrated in Fig. 5b. There would be cost due to uncovering a fraction of the solid surface,

$$\mathrm{d}G_{\mathrm{cost}} = (\sigma_{\mathrm{s/v}} - \sigma_{\mathrm{s/l}})L\mathrm{d}x, \tag{34}$$

where L is the length of the contact line in y-direction, but also gain,

$$\mathrm{d}G_{\mathrm{gain}} = L\sigma_{\mathrm{l/v}} \cos\Theta \mathrm{d}x, \tag{35}$$

due to reducing the liquid-vapor interface. At thermodynamic equilibrium, cost and gain are balanced ([12], p. 99),

$$(\sigma_{\mathrm{s/v}} - \sigma_{\mathrm{s/l}})L\mathrm{d}x = L\sigma_{\mathrm{l/v}} \cos\Theta \mathrm{d}x, \tag{36}$$

paving the way toward Eq. (33).

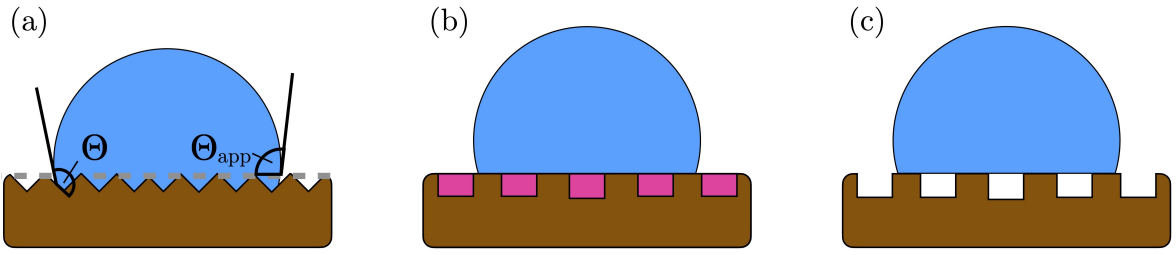

Fig. 6: Three concepts of wetting real solid surfaces: (a) Wenzel's concept of wetting rough, chemically homogeneous surfaces with: projected surface (dashed gray), ideal contact angle, Θ, and apparent contact angle, Θ_{app}, (b) Cassie's concept of wetting smooth, chemically heterogeneous surfaces (two chemistries: brown and purple), and (c) Cassie-Baxter's concept of wetting porous surfaces. Source: author's illustration inspired by Law and Zhao, 2016, p. 55 [11].

Young's equation is based on the assumption of an *ideal solid surface*, which is smooth, rigid, insoluble, non-reactive, and chemically homogeneous. Few experimentally available solid surfaces meet these conditions. Often, they are rough and/or chemically heterogeneous. Thus, contact angle experiments measure an *apparent* contact angle, Θ_{app}, rather than the (ideal) contact angle, Θ [14]. Fig. 6 illustrates three concepts to determine Θ_{app} from Θ:

1. *Wenzel's concept* [15] of wetting rough, chemically homogeneous solid surfaces (see Fig. 6a),

2. *Cassie's concept* [16] of wetting smooth, chemically heterogeneous solid surfaces (see Fig. 6b), and

[1] His famous paper «An essay on the cohesion of fluids»– where Young describes Eq. (33) in words – was published about 120 years *after* the invention of classical mechanics by Isaac Newton and 45 years *before* Rudolf Clausius explicitly mentioned the first law of thermodynamics.

3. *Cassie-Baxter's concept* [17] of wetting porous surfaces (see Fig. 6c).

1. The concept of Wenzel is based on the observation that roughening smooth solid surfaces increases wetting if the liquid preferentially wets the smooth surface, and decreases wetting if the liquid does not preferentially wet the surface [15]. The relationship between the apparent contact angle on rough, chemically homogeneous solid surfaces and the ideal contact angle is provided by the *Wenzel equation* [14],

$$\cos\Theta_{\text{app}} = \tilde{r}\cos\Theta, \tag{37}$$

where $\tilde{r}$ is the roughness ratio (i.e. the ratio of the actual and projected surface area).

2. The concept of Cassie was developed for smooth, chemically heterogeneous solid surfaces. In this case, the surface tension varies for different locations and Θ has a different value for each of them [16]. The *Cassie equation* [14],

$$\cos\Theta_{\text{app}} = f_1\cos\Theta_1 + f_2\cos\Theta_2, \tag{38}$$

where f_1 and f_2 are the area fractions of different chemistry, and Θ_1 and Θ_2 are the corresponding ideal contact angles, provides the relationship between the apparent and the ideal contact angles. Eq. (38) can be generalized for more than two different surface chemistries.

3. The concept of Cassie and Baxter is a special case of Cassie's concept for wetting of chemically homogeneous porous materials. It is based on the assumption that upon contact of the liquid droplet with the porous solid surface, air bubbles are trapped inside the pores. Cassie and Baxter assumed two chemistries: the material surrounding the pores (1) and the air inside the pores (2) [17]. Under the assumption that the liquid does not wet air,

$$\Theta_2 = 180^\circ, \tag{39}$$

the Cassie equation reduces to the Cassie-Baxter equation [14],

$$\cos\Theta_{\text{app}} = f_1\cos\Theta_1 - f_2. \tag{40}$$

2.2 Fundamentals of Statistical Thermodynamics

Statistical thermodynamics aims to express the properties of macroscopic systems in terms of the properties of their microscopic constituents, such as atoms or molecules (*particles*). While the properties of macroscopic systems can be described by classical thermodynamics – an elegant and self-consistent theory that makes no reference to the microscopic constituents in the system –, the properties of particles can be described by classical mechanics or quantum mechanics. In classical mechanics, the state of a particle is determined by its position and momentum, and the energy of a particle is a continuous function of these properties.

In quantum mechanics, a particle can only occupy discrete (quantum) states characterized by eigenfunctions, ψ_s, with eigenvalues of energy, ε_s. Quantum mechanics arose from the realization that it is impossible to precisely determine both the position and momentum of a particle simultaneously (Heisenberg's uncertainty principle). A naive approach for providing a microscopic foundation for classical thermodynamics would be to apply the laws of classical mechanics or quantum mechanics to each particle in a macroscopic system. But this attempt proves futile because of the large number of particles in a macroscopic system for which the laws of classical mechanics or quantum mechanics would have to be applied. The field of statistical thermodynamics arose from the realization that statistical methods are required to obtain a microscopic foundation for classical thermodynamics.

We are particularly interested in the concepts of statistical thermodynamics that constitute the foundation for Monte Carlo simulations (see Section 2.3.1). These concepts are based on a continuous description of particle states (classical mechanics), but are easier to derive for a discrete one (quantum mechanics). For this reason, the basics of statistical thermodynamics are first elaborated in Sections 2.2.1 and 2.2.2 assuming a quantum mechanical description of particle states, before the results are applied to a classical mechanical description of particle states in Section 2.2.3. In using quantum states, we will make the following assumptions:

1. The quantum mechanical nature of quantum states is irrelevant to the concepts of statistical thermodynamics described hereafter.

2. Quantum states with the same eigenvalue of energy possess the same probability.

3. Each particle in a system can occupy only one quantum state at a time. Quantum superposition (stating that any linear combination of eigenfunctions with the same eigenvalues of energy is also a valid eigenfunction) is neglected.

In more down-to-earth terms: we are only interested in the discrete description of states.

2.2.1 Probability and Combinatorics

The laws of probability and combinatorics form the foundation for statistical thermodynamics. We will illustrate these laws using the example of a model system that consists of four distinguishable and independent particles – A, B, C, and D. Each particle can occupy one of two available quantum states – described by ψ_1 or ψ_2 – with the same eigenvalue of energy. In this example, we will neglect any physical or quantum mechanical effect and only look at the statistical description of the model system. In principle, we could just as well be flipping four coins (named A, B, C, or D) showing either heads (ψ_1) or tails (ψ_2). Each sequence of particles in ψ_1 and ψ_2 represents a microscopic state of the model system. These microscopic states can also be described by eigenfunctions, Ψ_S, with eigenvalues of energy, E_S. For independent particles, Ψ_S can be represented by the product of single-particle eigenfunctions,

ψ_s. For example,

$$\Psi_S = \psi_1(\mathrm{A})\psi_1(\mathrm{B})\psi_1(\mathrm{C})\psi_1(\mathrm{D}), \tag{41}$$

characterizes the microscopic state of the model system in which the quantum state of each particle is described by ψ_1. The total number W of available microscopic states is given by:

$$W = N_\psi(A) \cdot N_\psi(B) \cdot N_\psi(C) \cdot N_\psi(D), \tag{42}$$

where $N_\psi(A), N_\psi(B), N_\psi(C)$, or $N_\psi(D)$ are the number of quantum states accessible to each particle. Thus, the model system has a total of $W = 16$ microscopic states to choose from:

$$\begin{array}{ll}
\psi_1(\mathrm{A})\ \psi_1(\mathrm{B})\ \psi_1(\mathrm{C})\ \psi_1(\mathrm{D}) & \psi_2(\mathrm{A})\ \psi_1(\mathrm{B})\ \psi_1(\mathrm{C})\ \psi_1(\mathrm{D}) \\
\psi_1(\mathrm{A})\ \psi_1(\mathrm{B})\ \psi_1(\mathrm{C})\ \psi_2(\mathrm{D}) & \psi_2(\mathrm{A})\ \psi_1(\mathrm{B})\ \psi_1(\mathrm{C})\ \psi_2(\mathrm{D}) \\
\psi_1(\mathrm{A})\ \psi_1(\mathrm{B})\ \psi_2(\mathrm{C})\ \psi_1(\mathrm{D}) & \psi_2(\mathrm{A})\ \psi_1(\mathrm{B})\ \psi_2(\mathrm{C})\ \psi_1(\mathrm{D}) \\
\psi_1(\mathrm{A})\ \psi_1(\mathrm{B})\ \psi_2(\mathrm{C})\ \psi_2(\mathrm{D}) & \psi_2(\mathrm{A})\ \psi_1(\mathrm{B})\ \psi_2(\mathrm{C})\ \psi_2(\mathrm{D}) \\
\\
\psi_1(\mathrm{A})\ \psi_2(\mathrm{B})\ \psi_1(\mathrm{C})\ \psi_1(\mathrm{D}) & \psi_2(\mathrm{A})\ \psi_2(\mathrm{B})\ \psi_1(\mathrm{C})\ \psi_1(\mathrm{D}) \\
\psi_1(\mathrm{A})\ \psi_2(\mathrm{B})\ \psi_1(\mathrm{C})\ \psi_2(\mathrm{D}) & \psi_2(\mathrm{A})\ \psi_2(\mathrm{B})\ \psi_1(\mathrm{C})\ \psi_2(\mathrm{D}) \\
\psi_1(\mathrm{A})\ \psi_2(\mathrm{B})\ \psi_2(\mathrm{C})\ \psi_1(\mathrm{D}) & \psi_2(\mathrm{A})\ \psi_2(\mathrm{B})\ \psi_2(\mathrm{C})\ \psi_1(\mathrm{D}) \\
\psi_1(\mathrm{A})\ \psi_2(\mathrm{B})\ \psi_2(\mathrm{C})\ \psi_2(\mathrm{D}) & \psi_2(\mathrm{A})\ \psi_2(\mathrm{B})\ \psi_2(\mathrm{C})\ \psi_2(\mathrm{D})
\end{array}$$

Each of these microscopic states has the same total energy, E. Their probability is given by the first postulate of statistical thermodynamics ([18], p. 5):

> 1. **Postulate:** For a system with constant total energy, E, all available microscopic states have the same probability,
>
> $$\mathrm{P}(\Psi_S) = \frac{1}{W}. \tag{43}$$

For a statistical description of our model system, we assign each microscopic state to a statistical distribution of particles over quantum states, $\{N_s\}$, where N_s is the number of particles found in each quantum state for any microscopic state. In our example, there exist five distinct distributions of particles over quantum states:

1) $\boldsymbol{\{4\ \psi_1 \mid 0\ \psi_2\}}$: All particles occupy state ψ_1.

2) $\boldsymbol{\{3\ \psi_1 \mid 1\ \psi_2\}}$: Three particles occupy state ψ_1 and one particle occupies state ψ_2.

3) $\boldsymbol{\{2\ \psi_1 \mid 2\ \psi_2\}}$: Two particles occupy state ψ_1 and two particles occupy state ψ_2.

4) $\boldsymbol{\{1\ \psi_1 \mid 3\ \psi_2\}}$: One particle occupies state ψ_1 and three particles occupy state ψ_2.

5) $\boldsymbol{\{0\ \psi_1 \mid 4\ \psi_2\}}$: All particles occupy state ψ_2.

Each microscopic state of the model system can be assigned to one (and only one) of these distributions. The number of microscopic states that belong to any of these distributions can be calculated using a multinomial distribution,

$$w\{N_s\} = \frac{N!}{\prod_s N_s!}, \tag{44}$$

where $w\{N_s\}$ is often referred to as the *number of realizations* of any distribution and N is the total number of particles. The probability of any distribution, $\mathrm{P}\{N_s\}$, is obtained by dividing the number of realizations by the total number of possible microscopic states,

$$\mathrm{P}\{N_s\} = \frac{w\{N_s\}}{W}, \tag{45}$$

where

$$W = \sum_s w\{N_s\}. \tag{46}$$

Using Eq. (44), we obtain the following numbers of realizations:

1) $\{4\ \psi_1 \mid 0\ \psi_2\}$: 1

2) $\{3\ \psi_1 \mid 1\ \psi_2\}$: 4

3) $\{2\ \psi_1 \mid 2\ \psi_2\}$: 6

4) $\{1\ \psi_1 \mid 3\ \psi_2\}$: 4

5) $\{0\ \psi_1 \mid 4\ \psi_2\}$: 1

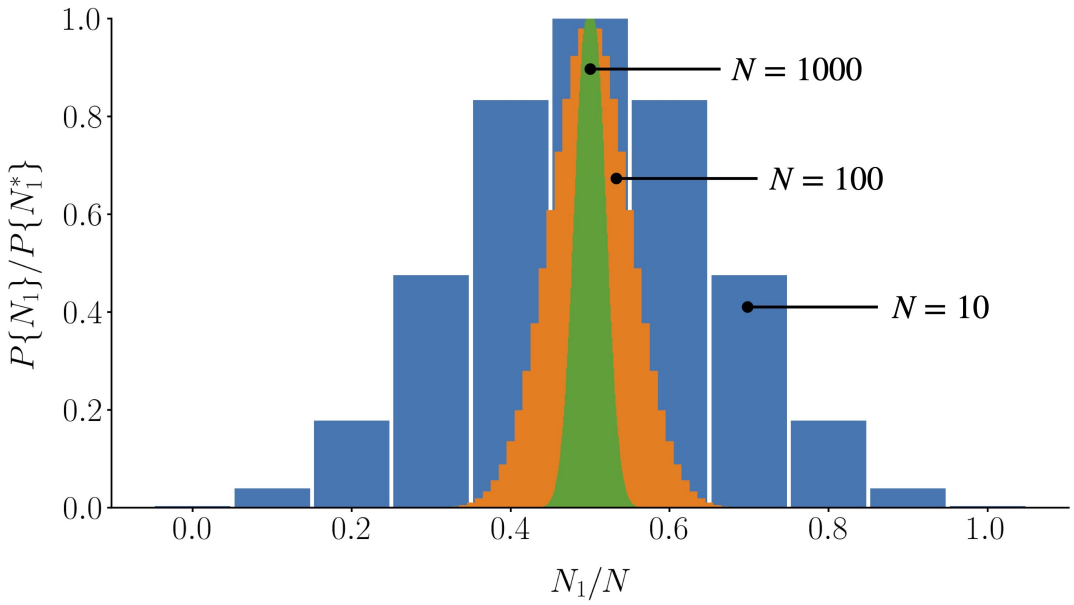

Fig. 7: Bar charts of the relative probability, $\mathrm{P}\{N_s\}/\mathrm{P}\{N_s^*\}$, against the relative number of particles in state ψ_1, N_1/N, for increasing particle numbers, N. Each particle can occupy one of two accessible quantum states (ψ_1 or ψ_2). Source: author's illustration inspired by Dill and Bromberg, 2011, p. 32 [19].

We notice that the equipartition of particles over quantum states, $\{2\ \psi_1 \mid 2\ \psi_2\}$, has the highest number of realizations and is according to Eq. (45) the most probable distribution of particles over quantum states, denoted by $\{N_s^*\}$. The importance of the most probable distribution increases significantly with increasing particle number. This is illustrated in Fig. 7, which plots the relative probability, $\mathrm{P}\{N_1\}/\mathrm{P}\{N_1^*\}$, against the relative number of particles in state ψ_1, N_1/N. For increasing N, the density of the relative probability increases sharply around the equipartition. More strictly, the *maximum term method* ([18], pp. 11–13),

$$\lim_{N\to\infty} (\ln W) \approx \ln w\{N_s^*\}, \tag{47}$$

suggests that any deviation from the most probable distribution, $\{N_s^*\}$, can be ignored for a system with a very high number of particles – as is common for a macroscopic system.

2.2.2 Statistical Ensembles

The attempt of reconciling thermodynamics with the microscopic laws of matter was decisively advanced by the concept of *statistical ensembles* originally introduced by the American physicist Josiah W. Gibbs [20]. For illustration, imagine a closed thermodynamic system of constant volume that is in thermal equilibrium with a heat bath, has a constant size, and contains only one type of particles. While the macroscopic state of the system is unambiguously determined by N, V, and T, the microscopic state of the system is not. In fact, the macroscopic state of the thermodynamic system at hand can be replicated by a large number of microscopic states. The collection of these microscopic states is referred to as the statistical ensemble of the system. Statistical ensembles are often classified based on the state variables that determine the macroscopic state of the system. Important statistical ensembles are:

- the **microcanonical ensemble**, which contains the microscopic states for an isolated system (i.e. N, V, and E are constant),
- the **canonical ensemble**, which contains the microscopic states for a closed system of constant volume (i.e. N, V, and T are constant), or
- the **grand canonical ensemble**, which contains the microscopic states for an open system of constant volume (i.e. μ, V, and T are constant).

The grand canonical ensemble is of special interest to this work, because of its application in adsorption simulations. It contains the microscopic states of a system with immovable, diathermic, and permeable boundaries ([21], p. 8). As a consequence, the system can not exchange energy through work with its surroundings, but energy through heat and matter through material exchange. In classical thermodynamics, this is an open system with constant volume. In equilibrium with a heat bath and a particle reservoir, an open system of constant volume has a constant chemical potential, μ, volume, V, and temperature, T. At constant T and constant μ, a system can occupy microscopic states, $\Psi_{N,S}$, of arbitrary

energy, $E_{N,S}$, and particle number, N, but only certain energies and particle numbers will have a significant probability. The probability distribution of the microscopic states for the grand canonical ensemble can be illustrated by a thought experiment that can be traced back to Gibbs ([20], pp. 32–56). In this thought experiment, we imagine the system of being surrounded by a nearly infinite number of identical copies. The fictive collection of systems is brought to the desired temperature and chemical potential and subsequently isolated from its surroundings. The systems in the collection have heat and particle permeable walls and will now distribute over the accessible microscopic states in a manner that is characteristic for the given chemical potential, volume, and temperature. We will denote this distribution of systems over microscopic states by $\{a_{N,S}\}$, where $a_{N,S}$ is the number of systems in the collection that occupy a certain microscopic state, $\Psi_{N,S}$. At equilibrium, $\{a_{N,S}\}$ corresponds to the distribution of systems over microscopic states that can be found in the grand canonical ensemble (see Fig. 8). The isolated collection of systems in our thought experiment has a constant number of systems, $\mathcal{A}$, a constant total energy, $\mathcal{E}$, and a constant number of particles, $\mathcal{N}$. Any distribution, $\{a_{N,S}\}$, must thus meet the following constraints:

$$\sum_N \sum_S a_{N,S} = \mathcal{A}, \tag{48}$$

$$\sum_N \sum_S a_{N,S} E_{N,S} = \mathcal{E}, \text{ and} \tag{49}$$

$$\sum_N \sum_S a_{N,S} N = \mathcal{N}. \tag{50}$$

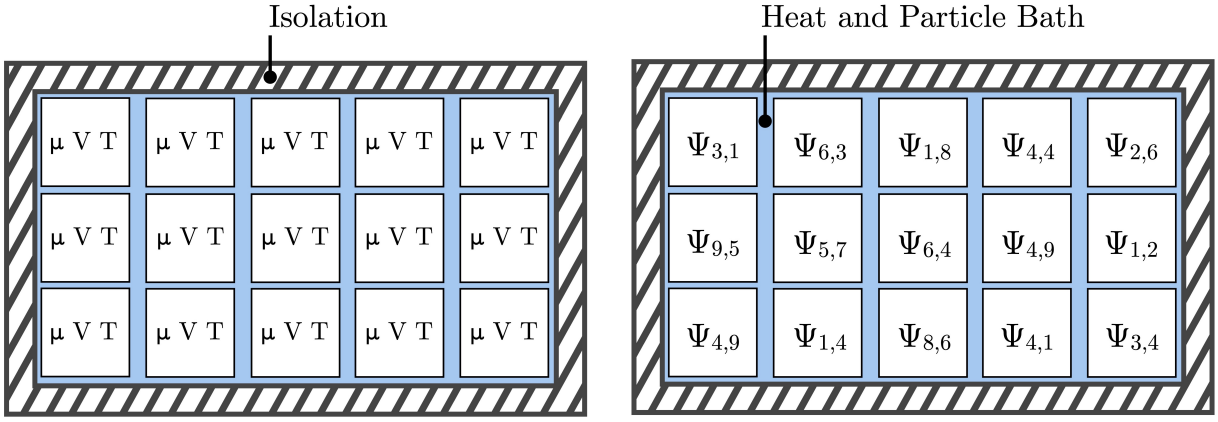

Fig. 8: Illustration of an isolated collection of $\mathcal{A} = 15$ systems. Each system in the collection has the same macroscopic state determined by μ, V, and T (left), but differs in the occupied microscopic state, $\Psi_{N,S}$ (right). At equilibrium, the distribution of systems over microscopic states in the collection is the same as that of the grand canonical ensemble. $\mathcal{A}$ can be chosen arbitrarily large. Source: author's illustration inspired by Raabe, 2017, p. 15 [21].

The number of realizations for any distribution of systems over microscopic states is given by:

$$\omega\{a_{N,S}\} = \frac{\mathcal{A}!}{\prod_{N,S} a_{N,S}!}. \tag{51}$$

The number of systems in the imaginary collection can be chosen arbitrarily large. Any deviation from the most probable distribution, $\{a^*_{N,S}\}$, can thus be neglected, as indicated by the maximum term method in Eq. (47). To use the maximum term method, we must take the logarithm of Eq. (51),

$$\ln w\{a_{N,S}\} = \ln \mathcal{A}! - \sum_{N,S} \ln a_{N,S}!, \tag{52}$$

which can be simplified into:

$$\ln w\{a_{N,S}\} = \mathcal{A} \ln \mathcal{A} - \mathcal{A} - \sum_{N,S} (a_{N,S} \ln a_{N,S} - a_{N,S}) \tag{53}$$

by using Stirling's approximation, which holds for large values of $a_{N,S}$. We find $\{a^*_{N,S}\}$ by determining the maximum of Eq. (53). For this, we will use the method of *Lagrange multipliers* (see [18], pp. 52–53). First, we equate Eq. (48) to (50) with zero,

$$\mathcal{A} - \sum_N \sum_S a_{N,S} = 0, \tag{54}$$

$$\mathcal{E} - \sum_N \sum_S a_{N,S} E_{N,S} = 0, \text{ and} \tag{55}$$

$$\mathcal{N} - \sum_N \sum_S a_{N,S} N = 0. \tag{56}$$

Then, we multiply the resulting expressions in Eq. (54) to (56) with the (yet undefined) Lagrangian multipliers $\alpha_\mathcal{L}$, $\beta_\mathcal{L}$, or $\gamma_\mathcal{L}$, and finally add the resulting terms to Eq. (53). The result is an auxiliary function,

$$\begin{aligned} \mathcal{L}\{a_{N,S}\} = \mathcal{A} \ln \mathcal{A} - \mathcal{A} - \sum_N \sum_S (a_{N,S} \ln a_{N,S} - a_{N,S}) + \alpha_\mathcal{L} \left(\mathcal{A} - \sum_N \sum_S a_{N,S} \right) \\ + \beta_\mathcal{L} \left(\mathcal{E} - \sum_N \sum_S a_{N,S} E_{N,S} \right) + \gamma_\mathcal{L} \left(\mathcal{N} - \sum_N \sum_S a_{N,S} N \right), \end{aligned} \tag{57}$$

whose total differential vanishes for $\{a^*_{N,S}\}$,

$$\mathrm{d}\mathcal{L}\{a^*_{N,S}\} = \sum_N \sum_S \left(\frac{\partial \mathcal{L}\{a^*_{N,S}\}}{\partial a_{N,S}} \right) \mathrm{d}a_{N,S} = 0. \tag{58}$$

For the non-trivial solution,

$$\frac{\partial \mathcal{L}\{a^*_{N,S}\}}{\partial a_{N,S}} = 0, \tag{59}$$

we obtain:

$$\frac{\partial \mathcal{L}\{a^*_{N,S}\}}{\partial a_{N,S}} = -\ln a_{N,S} - \alpha_\mathcal{L} - \beta_\mathcal{L} E_{N,S} - \gamma_\mathcal{L} N = 0. \tag{60}$$

Solving Eq. (60) for $a_{N,S}$ yields the number of systems in any microscopic state for $\{a^*_{N,S}\}$,

$$a^*_{N,S} = e^{-\alpha_{\mathcal{L}}} e^{-\beta_{\mathcal{L}} E_{N,S}} e^{-\gamma_{\mathcal{L}} N}. \tag{61}$$

The probability, $\mathrm{P}(\Psi_{N,S})$, of finding an open system in any microscopic state, $\Psi_{N,S}$, is thus ([22], p. 53):

$$\mathrm{P}(\Psi_{N,S}) = \frac{a^*_{N,S}}{\mathcal{A}} = \frac{e^{-\beta_{\mathcal{L}} E_{N,S}} e^{-\gamma_{\mathcal{L}} N}}{\sum_N \sum_S e^{-\beta_{\mathcal{L}} E_{N,S}} e^{-\gamma_{\mathcal{L}} N}}. \tag{62}$$

The Lagrangian multipliers $\beta_{\mathcal{L}}$ or $\gamma_{\mathcal{L}}$ are evaluated in Appendix A and found to be ([22], pp. 53–54):

$$\beta_{\mathcal{L}} = \frac{1}{k_{\mathrm{B}} T}, \tag{63}$$

$$\gamma_{\mathcal{L}} = -\frac{\mu}{k_{\mathrm{B}} T}, \tag{64}$$

where k_{B} is the Boltzmann factor. During a sufficiently long observation period, a system will visit all possible microscopic states many times. The time spent in each microscopic state is proportional to the probability, $\mathrm{P}(\Psi_S)$, of the respective microscopic state. This is often referred to as the *ergodic hypothesis* and forms the basis for the second postulate of statistical thermodynamics ([18], p. 25):

> **2. Postulate:** The time average of a mechanical variable, M, of a system is equal to the ensemble average, $\langle M \rangle$, of the corresponding ensemble, e.g.:
>
> $$\langle M \rangle := \sum_N \sum_S \mathrm{P}(\Psi_{N,S}) M_{N,S}, \tag{65}$$
>
> where $M_{N,S}$ is a mechanical variable that depends on $\Psi_{N,S}$.

The second postulate provides us with a method for determining the macroscopic observables of a system, $\langle M \rangle$, by performing averages over the statistical ensemble.

2.2.3 Classical Ensemble Averages

In Sections 2.2.1 and 2.2.2 we used the discrete nature of particle states in quantum mechanics to derive the concepts of statistical mechanics. Now, let us apply the results to a classical mechanical description of particle states to provide the foundation for Monte Carlo simulations; a technique that allows the approximation of ensemble averages. In classical mechanics, the energy of a system is a continuous function of the positions, $\{\vec{r}_N\}$, and momenta, $\{\vec{P}_N\}$, of all of its N particles,

$$\mathcal{H}\left(\{\vec{r}_N\}, \{\vec{P}_N\}\right) = \mathcal{U}(\{\vec{r}_N\}) + \mathcal{K}(\{\vec{P}_N\}), \tag{66}$$

where $\mathcal{U}$ and $\mathcal{K}$ are the potential or kinetic energy, and $\{\vec{r}_N\} = \vec{r}_1, \cdots, \vec{r}_N$ or $\{\vec{P}_N\} = \vec{P}_1, \cdots, \vec{P}_N$. To obtain a continuous description of the system's energy, we replace the sum over quantum states in Eq. (65) by an integral over the positions and momenta of the particles in the system,

$$\langle M \rangle := \frac{\sum_N \frac{1}{\tilde{h}^{3N}} \int_0^{d_{\text{sys}}} \cdots \int_0^{d_{\text{sys}}} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} M \exp\left(-\frac{\mathcal{H}(\{\vec{r}_N\},\{\vec{P}_N\}) - \mu N}{k_{\text{B}} T}\right) \, \mathrm{d}\{\vec{r}_N\}\mathrm{d}\{\vec{P}_N\}}{Z_{\text{class}}(\mu, V, T)}, \tag{67}$$

where d_{sys} is the dimension of the system in any of the three spatial directions, $Z_{\text{class}}(\mu, V, T)$ is the classical partition function for the grand canonical ensemble, M is a property of the system that depends on $\{\vec{r}_N\}$ and $\{\vec{P}_N\}$, and $\mathrm{d}\{\vec{r}_N\}$ as well as $\mathrm{d}\{\vec{P}_N\}$ are short-hand notations for:

$$\mathrm{d}\{\vec{r}_N\} = \mathrm{d}\vec{r}_{1,x}, \mathrm{d}\vec{r}_{1,y}, \mathrm{d}\vec{r}_{1,z}, \cdots, \mathrm{d}\vec{r}_{N,x}, \mathrm{d}\vec{r}_{N,y}, \mathrm{d}\vec{r}_{N,z}, \text{ or} \tag{68}$$

$$\mathrm{d}\{\vec{P}_N\} = \mathrm{d}\vec{P}_{1,x}, \mathrm{d}\vec{P}_{1,y}, \mathrm{d}\vec{P}_{1,z}, \cdots, \mathrm{d}\vec{P}_{N,x}, \mathrm{d}\vec{P}_{N,y}, \mathrm{d}\vec{P}_{N,z}. \tag{69}$$

In addition, we used the definitions for $\beta_{\mathcal{L}}$ and $\gamma_{\mathcal{L}}$ provided in Eq. (63) or Eq. (64) and introduced *Planck's constant*, $\tilde{h}$. In Appendix B, we show that Planck's constant is required to ensure that the classical partition function in Eq. (67) is dimensionless. The integral over momenta in Eq. (67) can be solved analytically. First, we use Eq. (66) to separate the integral in Eq. (67) into an integral over positions,

$$I_r = \int_0^{d_{\text{sys}}} \cdots \int_0^{d_{\text{sys}}} \exp\left(-\frac{\mathcal{U}(\{\vec{r}_N\})}{k_{\text{B}} T}\right) \, \mathrm{d}\{\vec{r}_N\}, \tag{70}$$

and an integral over momenta,

$$I_P = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp\left(-\frac{\mathcal{K}(\{\vec{P}_N\})}{k_{\text{B}} T}\right) \, \mathrm{d}\{\vec{P}_N\}. \tag{71}$$

Note that we omit M for the time being, as well as all terms that are independent of $\{\vec{r}_N\}$ or $\{\vec{P}_N\}$. Using the mechanical definition of kinetic energy,

$$\mathcal{K}(\{\vec{P}_N\}) = \sum_{i=1}^{N} \frac{\|\vec{P}_i\|^2}{2m_i}, \tag{72}$$

where $\|\vec{P}_i\|^2 = \|\vec{P}_{i,x}\|^2 + \|\vec{P}_{i,y}\|^2 + \|\vec{P}_{i,z}\|^2$, we obtain:

$$I_P = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \prod_{i=1}^{N} \exp\left(-\frac{\|\vec{P}_{i,x}\|^2 + \|\vec{P}_{i,y}\|^2 + \|\vec{P}_{i,z}\|^2}{2m_i k_{\text{B}} T}\right) \, \mathrm{d}\{\vec{P}_N\}. \tag{73}$$

For identical particles, Eq. (73) can be reduced to:

$$I_P = \left[\int_{-\infty}^{+\infty} \exp\left(-\frac{P^2}{2mk_\mathrm{B}T} \right) \, \mathrm{d}P \right]^{3N}, \tag{74}$$

because the integrands as well as the components, P, of the momentum vectors can independently assume any numerical value. Eq. (74) possesses an analytical solution, i.e. ([18], p. 65):

$$I_P = \sqrt{(2\pi mk_\mathrm{B}T)^{3N}}. \tag{75}$$

With these results, we can simplify the classical ensemble average into:

$$\langle M \rangle := \frac{\sum_N \frac{\exp\left(\frac{\mu N}{k_\mathrm{B}T}\right)}{\Lambda^{3N}} \int_0^{d_\mathrm{sys}} \cdots \int_0^{d_\mathrm{sys}} M \exp\left(-\frac{\mathcal{U}(\{\vec{r}_N\})}{k_\mathrm{B}T} \right) \, \mathrm{d}\{\vec{r}_N\}}{Z_\mathrm{class}(\mu, V, T)}, \tag{76}$$

where Λ is the *de-Broglie wavelength*,

$$\Lambda = \sqrt{\frac{h^2}{2\pi mk_\mathrm{B}T}}. \tag{77}$$

Note that Eq. (76) only provides the ensemble average of properties that depend on the *configuration* of the system, $\{\vec{r}_N\}$. As a consequence, Monte Carlo simulations are not suited for obtaining dynamic properties of a system. From Eq. (76) we obtain the probability of finding a system in any configuration,

$$\mathrm{P}(\{\vec{r}_N\}) = \frac{\frac{1}{\Lambda^{3N}} \exp\left(-\frac{\mathcal{U}(\{\vec{r}_N\}) - \mu N}{k_\mathrm{B}T} \right) \, \mathrm{d}\{\vec{r}_N\}}{Z_\mathrm{class}(\mu, V, T)}, \tag{78}$$

where the denominator is referred to as the *classical grand canonical partition function.*

2.3 Principles of Molecular Simulations

In theory, the ensemble average provides us with an expression for determining the properties of a macroscopic system from the properties of its microscopic constituents. However, a multidimensional integral such as occurring in Eq. (76) can only in a few instances be solved analytically. Moreover, numerical quadrature often proves futile, because of the large number of possible configurations for which we would need to determine the integrand. Molecular simulations are often the only way to approximate the ensemble average. Classical molecular simulation methods aim to evaluate the ensemble average by generating a representative sample of the microscopic states that replicate the macroscopic properties of a thermodynamic system ([23], pp. 163–164). In Sections 2.3.1 and 2.3.2, we will discuss two methods

for generating a representative sample of microscopic states: Monte Carlo simulations and molecular dynamics simulations, and in Section 2.3.3, we will explain the force field approach for calculating potential energies or forces in molecular simulations.

2.3.1 Monte Carlo Simulations

Monte Carlo simulations owe their name to the world-famous *Casino de Monte-Carlo* in the Principality of Monaco. Stanisław Ulam and John von Neumann – the pioneers of the Monte Carlo method – chose this name in allusion to the gambling-like nature of the algorithm, which is based on sequences of random numbers [24]. In the field of molecular modeling, Monte Carlo simulations aim to approximate the ensemble average by generating a representative sample of configurations for a system. This sample is representative if the probability of finding a configuration, $\{\vec{r}_N\}$, in the sample is equivalent to the probability of finding the configuration in the corresponding ensemble,

$$\mathrm{P}(\{\vec{r}_N\}) \approx \frac{l\{\vec{r}_N\}}{L\{\vec{r}_N\}}, \tag{79}$$

where $l\{\vec{r}_N\}$ is the number of any configuration within the sample and $L\{\vec{r}_N\}$ is the size of the sample. Generating a representative sample of configurations is easier said than done. We are often unable to draw configurations directly from $\mathrm{P}(\{\vec{r}_N\})$, because we are unable to solve the multidimensional integral in the partition function. The *Metropolis-Hastings algorithm* [25] does not require the determination of the partition function for drawing configurations from $\mathrm{P}(\{\vec{r}_N\})$. The idea is to set up a Markov chain of successive configurations in which any new configuration of the system, n, is obtained by a transition from the current configuration, o. Transitions are designed to obey *detailed balance*,

$$\mathrm{P}(o)\bar{\pi}(o \to n) = \mathrm{P}(n)\bar{\pi}(n \to o), \tag{80}$$

which states that any transition from o to (any) n is immediately balanced by a transition from (any) n to o. We are able to separate the transition probability, $\bar{\pi}$, into the probability of proposing the transition and the probability of accepting the proposal,

$$\bar{\pi}(o \to n) = \alpha(o \to n)\mathrm{P}_{\mathrm{acc}}(o \to n), \tag{81}$$

where $\alpha(o \to n)$ is the proposal probability and $\mathrm{P}_{\mathrm{acc}}(o \to n)$ is the acceptance probability. For the reverse transition holds:

$$\bar{\pi}(n \to o) = \alpha(n \to o)\mathrm{P}_{\mathrm{acc}}(n \to o). \tag{82}$$

Proposal probabilities are determined by the manner of how the transition is performed. Substituting Eq. (81) and Eq. (82) in Eq. (80) and solving for the acceptance probabilities

provides the key equation of the Metropolis-Hastings algorithm ([26], p. 29):

$$\frac{\mathrm{P}_{\mathrm{acc}}(o \to n)}{\mathrm{P}_{\mathrm{acc}}(n \to o)} = \frac{\alpha(n \to o)}{\alpha(o \to n)} \frac{\mathrm{P}(n)}{\mathrm{P}(o)}, \tag{83}$$

where the partition functions in $\mathrm{P}(n)$ and $\mathrm{P}(o)$ cancel out. Transitions that satisfy Eq. (83) ensure that equilibrium is maintained once it is reached. It is thus the starting point for developing transition algorithms, which are often dubbed as *trial moves*. Trial moves that are used in the following studies are:

1. the *translation* of atoms or molecules,
2. the *rotation* of rigid molecules,
3. the *identity change* of molecules,
4. the *configurational bias Monte Carlo* (CBMC) move,
5. the *insertion* or *deletion* of atoms or molecules, including the *continuous fractional component Monte Carlo* (CFCMC) move [27].

In what follows, we will illuminate the above mentioned trial moves for the grand canonical ensemble and derive the proposal and acceptance probability for each of them.

Trial Move 1: Translation

The *translational trial move* is illustrated in Fig. 9. In essence, it consists of randomly shifting the center of mass of a molecule. A random displacement can be achieved by adding small numerical values to the spatial coordinates of a molecule,

$$\vec{r}_x(n) = \vec{r}_x(o) + \Delta_{\max}\vec{r}(2\zeta_x - 1), \tag{84}$$

$$\vec{r}_y(n) = \vec{r}_y(o) + \Delta_{\max}\vec{r}(2\zeta_y - 1), \tag{85}$$

$$\vec{r}_z(n) = \vec{r}_z(o) + \Delta_{\max}\vec{r}(2\zeta_z - 1), \tag{86}$$

where $\vec{r}_x(o), \vec{r}_y(o), \vec{r}_z(o)$ and $\vec{r}_x(n), \vec{r}_y(n), \vec{r}_z(n)$ are the vector components of the current or new positions of the molecule, $\Delta_{\max}\vec{r}$ is the maximum possible spatial displacement, and ζ_x, ζ_y, and ζ_z are three independent random numbers between 0 and 1 drawn from a uniform distribution. Due to the limited floating-point precision of computers, there is a finite number of uniformly distributed new positions to choose from, so that

$$\alpha(o \to n) = \alpha(n \to o). \tag{87}$$

We obtain thus:

$$\frac{\mathrm{P}_{\mathrm{acc}}(o \to n)}{\mathrm{P}_{\mathrm{acc}}(n \to o)} = \exp\left(-\frac{\mathcal{U}(n) - \mathcal{U}(o)}{k_{\mathrm{B}}T}\right). \tag{88}$$

Any criterion used to determine whether a proposed translation is accepted or rejected must satisfy this equation. Commonly, the Metropolis-Hastings criterion is used ([26], p. 30),

$$\mathrm{P}_{\mathrm{acc}}(o \to n) = \min\left[1, \exp\left(-\frac{\mathcal{U}(n) - \mathcal{U}(o)}{k_{\mathrm{B}}T}\right)\right]. \tag{89}$$

For $\mathcal{U}(n) \leq \mathcal{U}(o)$, the proposed translation is always accepted. For $\mathcal{U}(n) > \mathcal{U}(o)$, the term in the second argument of Eq. (89) is determined and compared to a random number between 0 and 1. If the term is larger than the random number, the proposed translation is accepted. Otherwise it is rejected and the molecule remains in its current position.

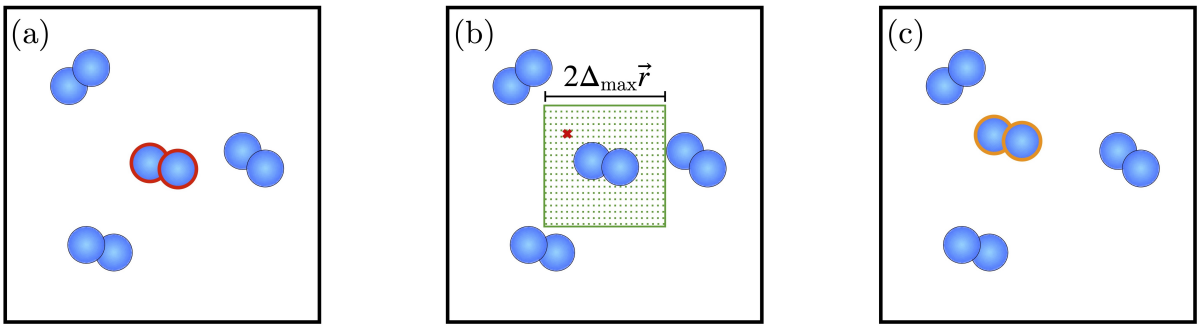

Fig. 9: Illustration of the translational trial move. (a) One of the N available molecules is chosen at random (bordered in red). (b) A new position is randomly proposed (red cross). We can imagine that we span a 2 $\Delta_{\max}\vec{r} \times 2$ $\Delta_{\max}\vec{r}$ mesh of equidistant points from which the new position is chosen. (c) The molecule is translated to the proposed position. The proposed position (bordered in orange) is accepted or rejected according to the acceptance criterion in Eq. (89). Source: author's illustration inspired by Allen and Tildesley, 1987, p. 119 [28].

Trial Move 2: Rotation

There are many ways to perform a *rotational trial move*. The simplest one is illustrated in Fig. 10 and is based on a coordinate transformation. Here, one of the Cartesian axes is selected and the coordinate system is rotated by a random angle $\delta\omega$ around the chosen axis ([29], p. 420). We obtain the following set of equations to compute the new positions of the atoms:

$$\begin{aligned}
\vec{r}_x(n) &= d_{\mathrm{bond}}\cos(\omega + \delta\omega), \\
&= d_{\mathrm{bond}}\cos(\omega)\cos(\delta\omega) - d_{\mathrm{bond}}\sin(\omega)\sin(\delta\omega), \\
&= \vec{r}_x(o)\cos(\delta\omega) - \vec{r}_y(o)\sin(\delta\omega),
\end{aligned} \tag{90}$$

$$\begin{aligned}
\vec{r}_y(n) &= d_{\mathrm{bond}}\sin(\omega + \delta\omega), \\
&= d_{\mathrm{bond}}\cos(\omega)\sin(\delta\omega) + d_{\mathrm{bond}}\sin(\omega)\cos(\delta\omega), \\
&= \vec{r}_x(o)\sin(\delta\omega) - \vec{r}_y(o)\cos(\delta\omega),
\end{aligned} \tag{91}$$

where d_{bond} is the bond distance and $\delta\omega$ is drawn randomly from the interval $(\delta\omega_{\min}, \delta\omega_{\max})$, leading to $\alpha(o \to n) = \alpha(n \to o)$. We obtain thus the acceptance criterion in Eq. (89).

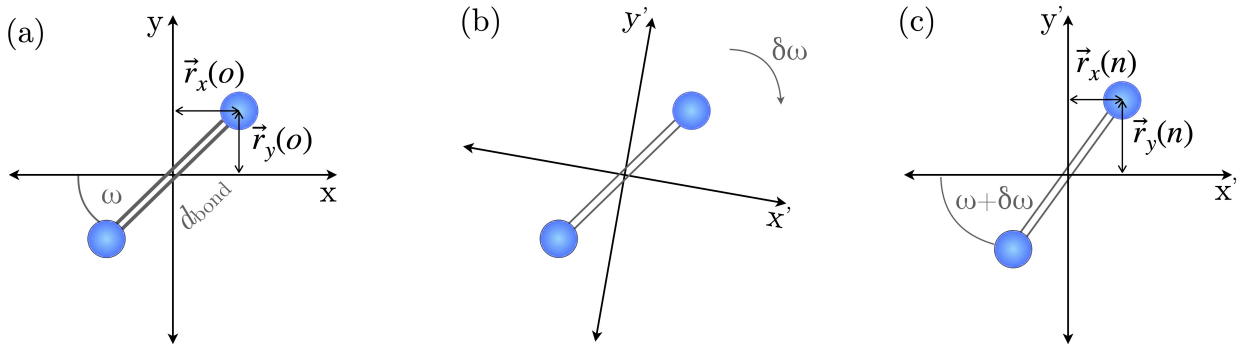

Fig. 10: Illustration of the rotational trial move for a diatomic molecule. (a) The positions of the atoms in the molecule are determined. (b) A coordinate transformation is performed by rotating the coordinate system by a random angle $\delta\omega$ around the z-axis. (c) The new positions of the atoms in the diatomic molecule are calculated using trigonometry. The proposed orientation is accepted or rejected according to the acceptance criterion in Eq. (89).

Trial Move 3: Identity Change

Translation and rotation are two physically meaningful trial moves. But the real strength of the Metropolis-Hastings algorithm lies in its ability to allow unphysical trial moves – as long as they satisfy detailed balance. The *identity change trial move* is an example for such an unphysical trial move, where a molecule of type A is replaced by a molecule of type B. This is illustrated in Fig. 11. Choosing any of the N_A molecules of type A in o possesses the probability:

$$\alpha(o \to n) = \frac{1}{N_\mathrm{A}}, \tag{92}$$

while choosing any of the $N_\mathrm{B} + 1$ molecules in n possesses the probability:

$$\alpha(n \to o) = \frac{1}{N_\mathrm{B} + 1}. \tag{93}$$

The probability of finding o in the grand canonical ensemble is provided by:

$$\mathrm{P}(o) = \frac{\Lambda_\mathrm{A}^{-3N_\mathrm{A}} \Lambda_\mathrm{B}^{-3N_\mathrm{B}} \exp\left(-\frac{\mathcal{U}(o) - (\mu_\mathrm{A} N_\mathrm{A} + \mu_\mathrm{B} N_\mathrm{B})}{k_\mathrm{B} T}\right) \mathrm{d}\vec{r}_1 \cdots \mathrm{d}\vec{r}_N}{Z(\mu_\mathrm{A}, \mu_\mathrm{B}, V, T)}, \tag{94}$$

where μ_A and μ_B are the chemical potentials of A or B and $Z(\mu_\mathrm{A}, \mu_\mathrm{B}, V, T)$ is the partition function of the system. The probability of finding n is provided by:

$$\mathrm{P}(n) = \frac{\Lambda_\mathrm{A}^{-3(N_\mathrm{A}-1)} \Lambda_\mathrm{B}^{-3(N_\mathrm{B}+1)} \exp\left(-\frac{\mathcal{U}(n) - (\mu_\mathrm{A}(N_\mathrm{A}-1) + \mu_\mathrm{B}(N_\mathrm{B}+1))}{k_\mathrm{B} T}\right) \mathrm{d}\vec{r}_1 \cdots \mathrm{d}\vec{r}_N}{Z(\mu_\mathrm{A}, \mu_\mathrm{B}, V, T)}. \tag{95}$$

Substituting Eqs. (92) to (95) in Eq. (83) yields:

$$\frac{\mathrm{P}_\mathrm{acc}(o \to n)}{\mathrm{P}_\mathrm{acc}(n \to o)} = \frac{N_\mathrm{A} \Lambda_\mathrm{A}^3}{(N_\mathrm{B} + 1) \Lambda_\mathrm{B}^3} \exp\left(\frac{\mu_\mathrm{B} - \mu_\mathrm{A}}{k_\mathrm{B} T}\right) \exp\left(-\frac{\mathcal{U}(n) - \mathcal{U}(o)}{k_\mathrm{B} T}\right), \tag{96}$$

which any acceptance criterion must satisfy. The Metropolis-Hastings criterion is [30]:

$$\mathrm{P}_{\mathrm{acc}}(o \to n) = \min\left[1, \frac{N_{\mathrm{A}}\Lambda_{\mathrm{A}}^3}{(N_{\mathrm{B}}+1)\Lambda_{\mathrm{B}}^3} \exp\left(\frac{\mu_{\mathrm{B}}-\mu_{\mathrm{A}}}{k_{\mathrm{B}}T}\right) \exp\left(-\frac{\mathcal{U}(n)-\mathcal{U}(o)}{k_{\mathrm{B}}T}\right)\right], \tag{97}$$

which depends both on the change in the chemical potential and in the potential energy.

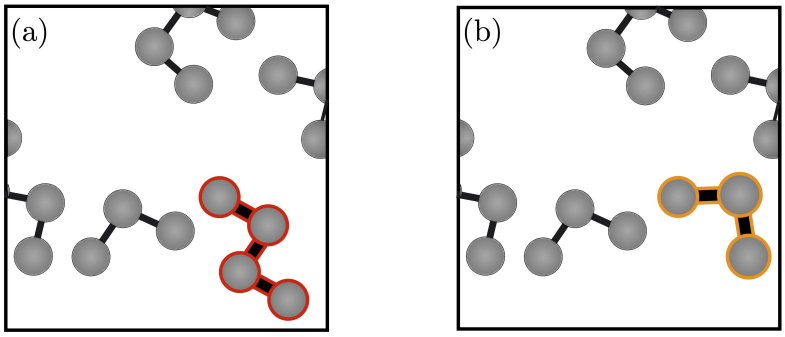

Fig. 11: Illustration of the identity change trial move. (a) One of the available n-butane molecules is chosen at random (bordered in red). (b) The chosen n-butane molecule is replaced by a propane molecule (bordered in orange). The proposed identity change is accepted or rejected according to the acceptance criterion in Eq. (97).

Trial Move 4: Configurational Bias Monte Carlo

In principle, it is possible to set up a Monte Carlo program in a way that treats molecules as rigid entities when using the translational or rotational trial move. But in this case, the *configurational bias Monte Carlo* (CBMC) trial move is often the only way to change the internal conformation of molecules with non-negligible intramolecular degrees of freedom [31]. The CBMC move is illustrated in Fig. 12. A molecule in the system is randomly selected and regrown segment by segment – either at the same location or at a different one. In the process, the regrowth is biased toward acceptable conformations. Regrowing a molecule with $\check{M}$ segments comprises three steps ([26], pp. 332–333):

1. Generate $\tilde{k}$ random trial positions for segment m. Select one of these trial positions with probability:

$$\mathrm{P}_m^{\mathrm{int}}(i) = \frac{\exp\left(-\frac{\mathcal{U}_m^{\mathrm{int}}(i)}{k_{\mathrm{B}}T}\right)}{\sum_{i=1}^{\tilde{k}} \exp\left(-\frac{\mathcal{U}_m^{\mathrm{int}}(i)}{k_{\mathrm{B}}T}\right)}, \tag{98}$$

   where $\mathcal{U}_m^{\mathrm{int}}(i)$ is the intramolecular energy of the ith trial position of segment m, which interacts with the already grown $m-1$ segments through intramolecular interactions. The denominator is referred to as the *internal Rosenbluth weight* of segment m,

$$\tilde{w}_m^{\mathrm{int}}(n) = \sum_{i=1}^{\tilde{k}} \exp\left(-\frac{\mathcal{U}_m^{\mathrm{int}}(i)}{k_{\mathrm{B}}T}\right). \tag{99}$$

2. Repeat the first step until $\tilde{l}$ trial positions are obtained. Select one of these trial

positions with probability:

$$\mathrm{P}_m^{\mathrm{ext}}(j) = \frac{\exp\left(-\frac{\mathcal{U}_m^{\mathrm{ext}}(j)}{k_{\mathrm{B}}T}\right)}{\sum\limits_{j=1}^{\tilde{l}} \exp\left(-\frac{\mathcal{U}_m^{\mathrm{ext}}(j)}{k_{\mathrm{B}}T}\right)}, \tag{100}$$

where $\mathcal{U}_m^{\mathrm{ext}}(j)$ is the intermolecular energy of the jth trial position of segment m due to intermolecular interactions. The denominator is referred to as the *external Rosenbluth weight* of segment m,

$$\tilde{w}_m^{\mathrm{ext}}(n) = \sum_{j=1}^{\tilde{l}} \exp\left(-\frac{\mathcal{U}_m^{\mathrm{ext}}(j)}{k_{\mathrm{B}}T}\right). \tag{101}$$

3. Repeat the first two steps until the molecule is fully regrown.

The probability of proposing n is provided by:

$$\alpha(o \to n) = \prod_{m=1}^{\check{M}} \mathrm{P}_m^{\mathrm{int}}(i)\mathrm{P}_m^{\mathrm{ext}}(j) = \frac{\exp\left(-\frac{\mathcal{U}(n)}{k_{\mathrm{B}}T}\right)}{\mathcal{W}(n)}, \tag{102}$$

where i and j are the chosen trial positions, and $\mathcal{W}(n)$ is the *Rosenbluth weight* of the fully regrown molecule,

$$\mathcal{W}(n) = \prod_{m=1}^{\check{M}} \tilde{w}_m^{\mathrm{int}}(n)\tilde{w}_m^{\mathrm{ext}}(n). \tag{103}$$

To find $\alpha(n \to o)$, the conformation of the molecule in o has to be retraced. This is achieved in three steps ([26], pp. 332–333):

1. Generate $\tilde{k} - 1$ random trial positions for the first segment m. The position of the segment in o is the $\tilde{k}$th trial position (denoted by Υ), which is chosen with:

$$\mathrm{P}_m^{\mathrm{int}}(\Upsilon) = \frac{\exp\left(-\frac{\mathcal{U}_m^{\mathrm{int}}(\Upsilon)}{k_{\mathrm{B}}T}\right)}{\sum\limits_{i=1}^{\tilde{k}} \exp\left(-\frac{\mathcal{U}_m^{\mathrm{int}}(i)}{k_{\mathrm{B}}T}\right)} = \frac{\exp\left(-\frac{\mathcal{U}_m^{\mathrm{int}}(\Upsilon)}{k_{\mathrm{B}}T}\right)}{\tilde{w}_m^{\mathrm{int}}(o)}. \tag{104}$$

2. Repeat the first step until $\tilde{l} - 1$ trial positions are obtained. The position of m in o is the $\tilde{l}$th trial position (denoted by Υ), which is chosen with probability

$$\mathrm{P}_m^{\mathrm{ext}}(\Upsilon) = \frac{\exp\left(-\frac{\mathcal{U}_m^{\mathrm{ext}}(\Upsilon)}{k_{\mathrm{B}}T}\right)}{\sum\limits_{j=1}^{\tilde{l}} \exp\left(-\frac{\mathcal{U}_m^{\mathrm{ext}}(j)}{k_{\mathrm{B}}T}\right)} = \frac{\exp\left(-\frac{\mathcal{U}_m^{\mathrm{ext}}(\Upsilon)}{k_{\mathrm{B}}T}\right)}{\tilde{w}_m^{\mathrm{ext}}(o)}. \tag{105}$$

3. Repeat the first two steps until the molecule is fully retraced.

The probability of retracing the old conformation is thus provided by:

$$\alpha(n \to o) = \prod_{m=1}^{\check{M}} \mathrm{P}_m^{\mathrm{int}}(\Upsilon)\mathrm{P}_m^{\mathrm{ext}}(\Upsilon) = \frac{\exp\left(-\frac{\mathcal{U}(o)}{k_{\mathrm{B}}T}\right)}{\mathcal{W}(o)}, \tag{106}$$

where $\mathcal{W}(o)$ is the *Rosenbluth weight* of the fully retraced molecule,

$$\mathcal{W}(o) = \prod_{m=1}^{\check{M}} \tilde{w}_m^{\mathrm{int}}(o)\tilde{w}_m^{\mathrm{ext}}(o). \tag{107}$$

By substituting Eq. (78), Eq. (102), and Eq. (106) in Eq. (83), we obtain:

$$\frac{\mathrm{P}_{\mathrm{acc}}(o \to n)}{\mathrm{P}_{\mathrm{acc}}(n \to o)} = \frac{\mathcal{W}(n)}{\mathcal{W}(o)}, \tag{108}$$

which any acceptance criterion must satisfy. The Metropolis-Hastings criterion is ([26], p. 334):

$$\mathrm{P}_{\mathrm{acc}}(o \to n) = \min\left[1, \frac{\mathcal{W}(n)}{\mathcal{W}(o)}\right]. \tag{109}$$

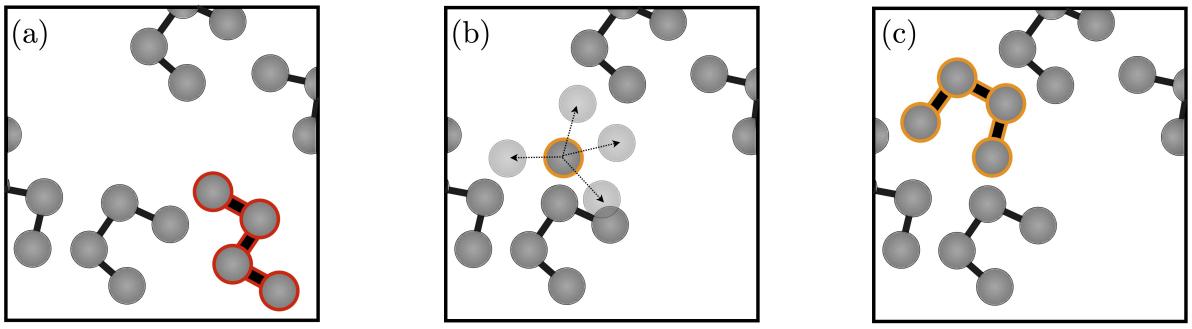

Fig. 12: Illustration of the CBMC trial move. (a) One of the available n-butane molecules is chosen at random (bordered in red). (b) The chosen n-butane molecule is removed and regrown segment by segment (bordered in orange). (c) The fully regrown molecule (bordered in orange) is accepted or rejected according to the acceptance probability in Eq. (109). Source: author's illustration inspired by Frenkel and Smit, 2002, p. 332 [26].

Trial Move 5: Insertion/Deletion of atoms or molecules

Changing the particle number in a system during simulation is important for sampling the grand canonical ensemble. This is achieved by the *insertion/deletion trial move* illustrated in Fig. 13. The system is allowed to exchange particles with an imaginary particle reservoir. To import a particle from the reservoir, first, a random position in the system is chosen with probability:

$$\alpha(o \to n) = \frac{\mathrm{d}\vec{r}}{V}, \tag{110}$$

where we recall that $\mathrm{d}\vec{r} := \mathrm{d}\vec{r}_x\mathrm{d}\vec{r}_y\mathrm{d}\vec{r}_z$.

Eq. (110) is tantamount to the volume fraction that the molecule will be occupying. The reverse trial move possesses the probability:

$$\alpha(n \to o) = \frac{1}{N+1}. \tag{111}$$

By substituting Eq. (78), Eq. (110), and Eq. (111) in Eq. (83), we obtain:

$$\frac{\mathrm{P_{acc}}(o \to n)}{\mathrm{P_{acc}}(n \to o)} = \frac{V}{(N+1)\Lambda^3} \exp\left(-\frac{\mathcal{U}(n) - \mathcal{U}(o) - \mu}{k_\mathrm{B}T}\right), \tag{112}$$

which is satisfied by the Metropolis-Hastings acceptance criterion ([26], p. 130),

$$\mathrm{P_{acc}}(o \to n) = \min\left[1, \frac{V}{(N+1)\Lambda^3} \exp\left(-\frac{\mathcal{U}(n) - \mathcal{U}(o) - \mu}{k_\mathrm{B}T}\right)\right]. \tag{113}$$

For the deletion of a particle, the Metropolis-Hastings acceptance criterion is given by ([26], p. 130):

$$\mathrm{P_{acc}}(o \to n) = \min\left[1, \frac{N\Lambda^3}{V} \exp\left(-\frac{\mathcal{U}(n) - \mathcal{U}(o) + \mu}{k_\mathrm{B}T}\right)\right]. \tag{114}$$

Both acceptance criteria in Eq. (113) and Eq. (114) depend on the change in the potential energy and in the chemical potential of the particle reservoir. An equation for evaluating μ is derived in Appendix C.

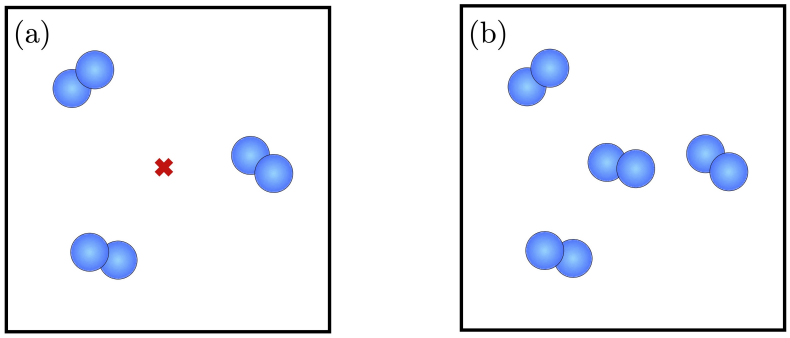

Fig. 13: Illustration of the insertion trial move. (a) A random position in the system is chosen (red cross), (b) at which a new particle is going to be inserted. The proposed insertion is accepted or rejected according to the acceptance criterion in Eq. (113).

With increasing particle density, it becomes harder and harder to successfully insert molecules at random positions in the system. More sophisticated insertion techniques include the *CBMC insertion* or the *continuous fractional component Monte Carlo* (CFCMC) method [27]. In the CBMC insertion trial move, a molecule is grown segment by segment into the system. Similarly to the conventional CBMC method, the growth is biased toward acceptable conformations. Applying the previously obtained results, we obtain the probability of proposing the insertion:

$$\alpha(o \to n) = \frac{\exp\left(-\frac{\mathcal{U}(n)}{k_\mathrm{B}T}\right)}{\mathcal{W}(n)} \frac{\mathrm{d}\vec{r}}{V}, \tag{115}$$

while the probability to propose the reverse trial move is provided in Eq. (111). By substi-

tuting Eq. (78), Eq. (111), and Eq. (115) in Eq. (83), we obtain:

$$\frac{\mathrm{P_{acc}}(o \to n)}{\mathrm{P_{acc}}(n \to o)} = \frac{\mathcal{W}(n)V}{(N+1)\Lambda^3} \exp\left(\frac{\mu}{k_\mathrm{B}T}\right), \tag{116}$$

which is satisfied by the Metropolis-Hastings acceptance criterion [32]:

$$\mathrm{P_{acc}}(o \to n) = \min\left[1, \frac{\mathcal{W}(n)V}{(N+1)\Lambda^3} \exp\left(\frac{\mu}{k_\mathrm{B}T}\right)\right]. \tag{117}$$

The CBMC insertion method is often used when modeling systems with medium density.

The CFCMC method is illustrated in Fig. 14. It expands the system with a *fractional* molecule for which intermolecular interactions are scaled [27]. The scaling factor, λ, can assume values between 0 – where the fractional molecule does not interact with other molecules – and 1 – where the molecule is fully present [32]. The CFCMC trial move attempts to change λ [27],

$$\lambda(n) = \lambda(o) + (2\zeta_\lambda - 1)\Delta_{\max}\lambda, \tag{118}$$

where ζ_λ is a random number between 0 and 1 drawn from a uniform distribution, and $\Delta_{\max}\lambda$ is the maximum possible change of λ. The CFCMC trial move results in one of the following outcomes [33]:

1. The scaling factor remains between zero and one. In analogy to the translational trial move, the new scaling factor is accepted with probability:

$$\mathrm{P_{acc}}(o \to n) = \min\left[1, \exp\left(-\frac{\mathcal{U}_\lambda(n) - \mathcal{U}_\lambda(o)}{k_\mathrm{B}T}\right)\right], \tag{119}$$

   where $\mathcal{U}_\lambda$ is the potential energy of the system including the fractional molecule.

2. The scaling factor becomes smaller than or equal to zero, i.e. $\lambda = -\delta$. The fractional molecule is removed and a molecule in the system is randomly chosen and made fractional with scaling factor $\lambda = 1 - \delta$. Similarly to the deletion of a random molecule, the trial move is accepted with probability:

$$\mathrm{P_{acc}}(o \to n) = \min\left[1, \frac{N\Lambda^3}{V} \exp\left(-\frac{\mathcal{U}_\lambda(n) - \mathcal{U}_\lambda(o) + \mu}{k_\mathrm{B}T}\right)\right]. \tag{120}$$

3. The scaling factor becomes larger than or equal to one, i.e. $\lambda = 1 + \delta$. The fractional molecule is made fully present and a new fractional molecule with scaling factor $\lambda = \delta$ is inserted at a random location in the system. Similarly to the insertion of a new molecule, the trial move is accepted with probability:

$$\mathrm{P_{acc}}(o \to n) = \min\left[1, \frac{V}{(N+1)\Lambda^3} \exp\left(-\frac{\mathcal{U}_\lambda(n) - \mathcal{U}_\lambda(o) - \mu}{k_\mathrm{B}T}\right)\right]. \tag{121}$$

In between CFCMC trial moves, conventional Monte Carlo trial moves, such as translation or rotation, are used to equilibrate the system. In this way, the system gently adapts to the gradual insertion of the new molecule [34]. The CFCMC method unfolds its full potential when used for high-density systems.

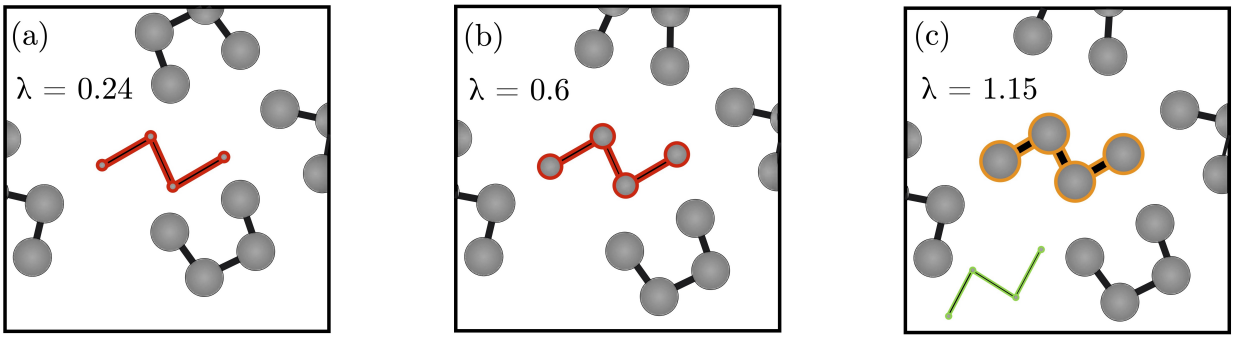

Fig. 14: Illustration of the CFCMC trial move. (a) A fractional molecule (bordered in red) with $\lambda = 0.24$ is inserted into the system. (b) The scaling factor is increased. Metaphorically speaking, the atoms in the molecule are „inflated." (c) The scaling factor becomes larger than one. The fractional molecule (bordered in orange) is made fully present and a new fractional molecule (bordered in green) is inserted with $\lambda = 0.15$. Source: author's representation inspired by Torres-Knoop et al., 2014, p. 942 [34].

2.3.2 Molecular Dynamics Simulations

Molecular dynamics simulations aim to generate a representative sample of microscopic states by following the time-evolution of the system. Newton's laws of motion form the foundation for molecular dynamics simulations:

1. A particle remains in its state of rest or continues to move in a straight line at constant velocity unless a force acts upon it.

2. The force, $\vec{F}_i$, acting on a particle, i, equals the rate of change of momentum, $\vec{P}_i$,

$$\vec{F}_i = \frac{\mathrm{d}\vec{P}_i}{\mathrm{d}t}. \tag{122}$$

3. If particle i exerts a force on particle j, then an equal but opposite force from j acts on i,

$$\vec{F}_i = -\vec{F}_j. \tag{123}$$

From Newton's laws of motion we are able to derive the *Hamiltonian equations of motion*. First, we set up the total differential of the Hamiltonian in Eq. (66),

$$\mathrm{d}\mathcal{H} = \sum_{i=1}^{N} \left(\frac{\partial \mathcal{H}}{\partial \vec{r}_i} \right)_{\substack{\vec{r}_{j\neq i}, \\ \vec{P}_{j\in N}}} \bullet \, \mathrm{d}\vec{r}_i + \sum_{i=1}^{N} \left(\frac{\partial \mathcal{H}}{\partial \vec{P}_i} \right)_{\substack{\vec{P}_{j\neq i}, \\ \vec{r}_{j\in N}}} \bullet \, \mathrm{d}\vec{P}_i, \tag{124}$$

$$= \sum_{i=1}^{N} \left(\frac{\partial \mathcal{U}}{\partial \vec{r}_i} \right)_{\vec{r}_{j\neq i}} \bullet \, \mathrm{d}\vec{r}_i + \sum_{i=1}^{N} \left(\frac{\partial \mathcal{K}}{\partial \vec{P}_i} \right)_{\vec{P}_{j\neq i}} \bullet \, \mathrm{d}\vec{P}_i, \tag{125}$$

where • denotes the dot product. Second, we use Eq. (124) and Eq. (125) to obtain the time-derivative of the Hamiltonian,

$$\dot{\mathcal{H}} = \sum_{i=1}^{N} \left(\frac{\partial \mathcal{H}}{\partial \vec{r}_i} \right)_{\substack{\vec{r}_{j\neq i}, \\ \vec{P}_{j\in N}}} \bullet \dot{r}_i + \sum_{i=1}^{N} \left(\frac{\partial \mathcal{H}}{\partial \vec{P}_i} \right)_{\substack{\vec{P}_{j\neq i}, \\ \vec{r}_{j\in N}}} \bullet \dot{P}_i, \tag{126}$$

$$= \sum_{i=1}^{N} \left(\frac{\partial \mathcal{U}}{\partial \vec{r}_i} \right)_{\vec{r}_{j\neq i}} \bullet \dot{r}_i + \sum_{i=1}^{N} \left(\frac{\partial \mathcal{K}}{\partial \vec{P}_i} \right)_{\vec{P}_{j\neq i}} \bullet \dot{P}_i, \tag{127}$$

where $\dot{r}$ and $\dot{P}$ are the time-derivatives of the position or momentum vector. Finally, comparing the coefficients in Eq. (126) and Eq. (127) provides Hamilton's equations of motion:

$$\left(\frac{\partial \mathcal{H}}{\partial \vec{r}_i} \right)_{\substack{\vec{r}_{j\neq i}, \\ \vec{P}_{j\in N}}} = \left(\frac{\partial \mathcal{U}}{\partial \vec{r}_i} \right)_{\vec{r}_{j\neq i}} = -\vec{F}_i, \tag{128}$$

$$\left(\frac{\partial \mathcal{H}}{\partial \vec{P}_i} \right)_{\substack{\vec{P}_{j\neq i}, \\ \vec{r}_{j\in N}}} = \left(\frac{\partial \mathcal{K}}{\partial \vec{P}_i} \right)_{\vec{P}_{j\neq i}} = \frac{\vec{P}_i}{m_i}. \tag{129}$$

Note that by substituting Eq. (128), Eq. (129), and Eq. (122) in Eq. (126), we obtain $\dot{\mathcal{H}} = 0$, stating that the Hamiltonian equations of motion conserve the total energy of the system. Solving the Hamiltonian equations of motion requires the evaluation of $6N$ first-order differential equations. These differential equations can often not be solved analytically, because the force acting on a particle changes whenever it or another particle changes position. As a consequence, Hamilton's equations of motion are often solved using *finite difference methods* ([29], p. 355), which are based on a Taylor expansion of the position, $\vec{r}$, velocity, $\vec{v}$, acceleration, $\vec{a}$, and/or other dynamic variables around t,

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \frac{\mathrm{d}\vec{r}(t)}{\mathrm{d}t}\Delta t + \frac{1}{2}\frac{\mathrm{d}^2\vec{r}(t)}{\mathrm{d}t^2}\Delta t^2 + \ldots, \tag{130}$$

$$\vec{v}(t + \Delta t) = \vec{v}(t) + \frac{\mathrm{d}\vec{v}(t)}{\mathrm{d}t}\Delta t + \frac{1}{2}\frac{\mathrm{d}^2\vec{v}(t)}{\mathrm{d}t^2}\Delta t^2 + \ldots, \tag{131}$$

$$\vec{a}(t + \Delta t) = \vec{a}(t) + \frac{\mathrm{d}\vec{a}(t)}{\mathrm{d}t}\Delta t + \frac{1}{2}\frac{\mathrm{d}^2\vec{a}(t)}{\mathrm{d}t^2}\Delta t^2 + \ldots. \tag{132}$$

and solved using a finite time step, Δt. Many methods were proposed to integrate this set of equations. We will discuss two of these *time-integration schemes* below:

1. the *velocity Verlet algorithm* for simulating the motion of atoms or molecules with non-negligible intramolecular degrees of freedom (i.e. flexible molecules), and

2. the *rotational-velocity-Verlet algorithm* [35] for simulating the motion of molecules with negligible intramolecular degrees of freedom (i.e. rigid molecules).

(1.) The *velocity Verlet algorithm* is based on a second-order Taylor expansion of the position about t as well as a first-order Taylor expansion of the velocity about t and $t + \frac{1}{2}\Delta t$. The algorithm consists of four steps ([29], p. 357):

1) Calculate the half-step velocity, $\vec{v}_i(t + \frac{1}{2}\Delta t)$, of each mobile particle, i, in the system from its current velocity, $\vec{v}_i(t)$, and force-induced acceleration, $\vec{a}_i(t)$,

$$\vec{v}_i(t + \tfrac{1}{2}\Delta t) = \vec{v}_i(t) + \frac{1}{2}\vec{a}_i(t)\Delta t. \tag{133}$$

2) Calculate the position, $\vec{r}_i(t + \Delta t)$, of each mobile particle, i, in the system from its current position, $\vec{r}_i(t)$, and half-step velocity, $\vec{v}_i(t + \frac{1}{2}\Delta t)$,

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i(t + \tfrac{1}{2}\Delta t)\Delta t. \tag{134}$$

3) Calculate the acceleration, $\vec{a}_i(t+\Delta t)$, of each mobile particle, i, in the system resulting from the force, $\vec{F}(\vec{r}_i(t + \Delta t))$, acting at $\vec{r}_i(t + \Delta t)$,

$$\vec{a}_i(t + \Delta t) = \frac{\vec{F}(\vec{r}_i(t + \Delta t))}{m_i}. \tag{135}$$

4) Calculate the velocity, $\vec{v}_i(t + \Delta t)$, of each mobile particle, i, in the system from its half-step velocity, $\vec{v}_i(t + \frac{1}{2}\Delta t)$, and acceleration, $\vec{a}_i(t + \Delta t)$,

$$\vec{v}_i(t + \Delta t) = \vec{v}_i(t + \tfrac{1}{2}\Delta t) + \frac{1}{2}\vec{a}_i(t + \Delta t)\Delta t. \tag{136}$$

The algorithm is started again at step 1 with the new values for $\vec{r}_i$, $\vec{v}_i$, and $\vec{a}_i$. Recall that the velocity Verlet algorithm can solely be applied to particles that are modeled with only translational degrees of freedom. However, in some instances, it is useful to constrain certain intramolecular degrees of freedom. Within the velocity Verlet method, this can be achieved by using the RATTLE algorithm [36] that is described in Appendix D.

(2.) Molecules with negligible intramolecular degrees of freedom are often modeled as rigid bodies, which possess both translational and rotational degrees of freedom. While we can use the velocity Verlet algorithm to describe the translational motion of the center of mass of a rigid body, the rotational motion requires a separate time-integration scheme. One example is the *rotational-velocity-Verlet integration scheme* developed by Rozmanov and Kusalik [35] – a modified version of the velocity Verlet algorithm that uses rotational analogs of mass, m (inertia tensor, I), position, $\vec{r}$ (orientation, q), momentum, $\vec{P}$ (angular momentum, $\vec{L}$), and force, $\vec{F}$ (torque, $\vec{T}$). Orientation is represented by normalized quaternions, $\hat{q}$, as the rotation of the local coordinate system of the rigid molecule in the global coordination frame. To emphasize the analogy to the velocity Verlet algorithm, we divide the rotational-velocity-Verlet integration scheme into four steps [35]:

1) Calculate the half-step angular momentum, $\vec{L}(t+\frac{1}{2}\Delta t)$, of the rigid molecule in the global coordination system from the angular momentum, $\vec{L}(t)$, and torque, $\vec{T}(t)$,

$$\vec{L}(t+\tfrac{1}{2}\Delta t) = \vec{L}(t) + \frac{1}{2}\vec{T}(t)\Delta t. \tag{137}$$

Transform $\vec{L}(t+\frac{1}{2}\Delta t)$ to the local coordination system,

$$\vec{L}'(t+\tfrac{1}{2}\Delta t) = \hat{q}^{-1}(t+\tfrac{1}{2}\Delta t)\vec{L}(t+\tfrac{1}{2}\Delta t)\hat{q}(t+\tfrac{1}{2}\Delta t), \tag{138}$$

using the orientation of the rigid molecule,

$$\hat{q}(t+\tfrac{1}{2}\Delta t) = \hat{q}(t) + \frac{1}{2}\dot{q}(t+\tfrac{1}{2}\Delta t)\Delta t. \tag{139}$$

Calculate the quaternion time-derivative, $\dot{q}(t+\frac{1}{2}\Delta t)$, from the local angular momentum, $\vec{L}'(t)$, its time-derivative, $\dot{L}'(t)$, and the local inertia tensor, I',

$$\dot{q}(t+\tfrac{1}{2}\Delta t) = \frac{1}{2}\hat{q}(t+\tfrac{1}{2}\Delta t)\left(\left[I'\right]^{-1}\vec{L}'(t) + \frac{1}{2}\left[I'\right]^{-1}\dot{L}'(t)\Delta t\right). \tag{140}$$

Notice: Eqs. (138) to (140) are codependent and need to be solved iteratively.

2) Calculate the rigid molecule's orientation, $\hat{q}(t+\Delta t)$, from the current orientation, $\hat{q}(t)$, and the quaternion time-derivative, $\dot{q}(t+\frac{1}{2}\Delta t)$,

$$\hat{q}(t+\Delta t) = \hat{q}(t) + \dot{q}(t+\tfrac{1}{2}\Delta t)\Delta t. \tag{141}$$

Determine the new center of mass, $\vec{r}_{\text{com}}(t+\Delta t)$, using Eq. (134) and position, $\vec{r}_i(t+\Delta t)$, of each interaction site, i, in the molecule,

$$\vec{r}_i(t+\Delta t) = \vec{r}_{\text{com}}(t+\Delta t) + \hat{q}(t+\Delta t)\left[\vec{r}_i(t) - \vec{r}_{\text{com}}(t)\right]\hat{q}^{-1}(t+\Delta t). \tag{142}$$

3) Calculate the torque, $\vec{T}(t+\Delta t)$, acting on the rigid molecule,

$$\vec{T}(t+\Delta t) = \sum_i \left[\vec{r}_i(t+\Delta t) - \vec{r}_{\text{com}}(t+\Delta t)\right] \times \vec{F}(\vec{r}_i(t+\Delta t)). \tag{143}$$

4) Calculate the global angular momentum, $\vec{L}(t+\Delta t)$, from the global half-step angular momentum, $\vec{L}(t+\frac{1}{2}\Delta t)$, and the new torque, $\vec{T}(t+\Delta t)$,

$$\vec{L}(t+\Delta t) = \vec{L}(t+\tfrac{1}{2}\Delta t) + \frac{1}{2}\vec{T}(t+\Delta t)\Delta t. \tag{144}$$

The algorithm starts again at step 1 with the new values for $\hat{q}$, $\vec{L}$, and $\vec{T}$.

Constant Temperature Simulations

The time-integration schemes discussed above conserve the energy of the system, but it is often more desirable to maintain the temperature at a fixed value. One reason is that experiments are usually carried out at constant temperature rather than at constant energy. To maintain the temperature in a molecular dynamics simulation, we apply *thermostats* that rely on the definition of temperature in the kinetic theory of gases,

$$T = \frac{2\langle \mathcal{K} \rangle}{N_\mathrm{f} k_\mathrm{B}}, \tag{145}$$

where N_f are the degrees of freedom of the system and $\langle \mathcal{K} \rangle$ is the mean kinetic energy,

$$\langle \mathcal{K} \rangle = \frac{1}{2} \sum_{i=1}^{N} m_i \langle v_i^2 \rangle, \tag{146}$$

which holds for $m_i = \text{const.}$ and where the mean-squared speed of particle i, $\langle v_i^2 \rangle = \langle \|\vec{v}_i\|^2 \rangle$, is provided by:

$$\langle v_i^2 \rangle = \int_0^\infty v_i^2 \mathrm{P}(v_i), \tag{147}$$

where $\mathrm{P}(v_i)$ is the Maxwell-Boltzmann distribution of speed (derived in Appendix E). In the following, we will discuss two thermostats:

1. the *canonical sampling through velocity rescaling* (CSVR) algorithm [37, 38], and
2. the *Nosé-Hoover thermostat* [39, 40].

(1.) Velocity rescaling thermostats aim to impose the target temperature, T_t, by scaling the instantaneous velocities of all particles in the system with the same scaling factor, κ [37]. The CSVR algorithm [38] is one of the few velocity rescaling algorithms that accurately sample the canonical ensemble [41]. In the CSVR thermostat, the system is assumed of being coupled to an external heat bath of constant temperature, T_t. The rate of change of the kinetic energy in the system is proportional to the difference between the target, $\mathcal{K}_\mathrm{t}$, and instantaneous kinetic energy, $\mathcal{K}$, plus a stochastic term [38],

$$\frac{\mathrm{d}\mathcal{K}}{\mathrm{d}t} = \frac{1}{\tau}(\mathcal{K}_\mathrm{t} - \mathcal{K}) + \frac{2}{\sqrt{\tau}} \sqrt{\frac{\mathcal{K}_\mathrm{t}\mathcal{K}}{N_\mathrm{f}}} \frac{\mathrm{d}\check{W}}{\mathrm{d}t}, \tag{148}$$

where τ is the coupling parameter and $\mathrm{d}\check{W}$ is called a Wiener noise – a stochastic process. The coupling parameter allows to distribute the velocity rescaling over several time steps. Solving Eq. (148) yields [38]:

$$\kappa = \sqrt{e^{-\frac{\Delta t}{\tau}} + \frac{\mathcal{K}_\mathrm{t}}{\mathcal{K} N_\mathrm{f}} \left(1 - e^{-\frac{\Delta t}{\tau}}\right) \sum_{i=1}^{N_\mathrm{f}} \zeta_i^2 + 2e^{-\frac{\Delta t}{2\tau}} \sqrt{\frac{\mathcal{K}_\mathrm{t}}{\mathcal{K} N_\mathrm{f}} \left(1 - e^{-\frac{\Delta t}{\tau}}\right)} \zeta_1}, \tag{149}$$

where ζ_i is an independent random number drawn from a Gaussian distribution.

(2.) In the *Nosé-Hoover thermostat*, the heat bath is made an integral part of the system by introducing an additional degree of freedom, $\bar{s}$. Each state of the extended system, $\{\vec{r}_N', \vec{P}_N'\}$, corresponds to a state of the original system, $\{\vec{r}_N, \vec{P}_N\}$. Nosé [39] used the following relationships to convert the dynamic variables from the extended to the original system:

$$\vec{r}_i = \vec{r}_i', \tag{150}$$

$$\vec{P}_i = \frac{\vec{P}_i'}{\bar{s}'}, \tag{151}$$

$$\bar{s} = \bar{s}', \tag{152}$$

$$P_{\bar{s}} = \frac{P_{\bar{s}}'}{\bar{s}'}. \tag{153}$$

To describe the motion of particles in the extended system, we need to modify Hamilton's equations of motion. The Hamiltonian of the extended system is provided by [39]:

$$\mathcal{H}_{\text{Nosé}} = \sum_{i=1}^{N} \frac{\|\vec{P}_i'\|^2}{2m_i\bar{s}'^2} + \mathcal{U}(\{\vec{r}_N'\}) + k_{\text{B}}T_{\text{t}}N_{\text{f}} \ln \bar{s}' + \frac{P_{\bar{s}}'^2}{2Q_{\bar{s}}}, \tag{154}$$

where $Q_{\bar{s}}$ is the fictitious mass of $\bar{s}'$, and the last two terms on the right-hand side correspond to the potential or kinetic energy of $\bar{s}'$. We obtain *Nosé's equations of motion* by differentiation of Eq. (154) with respect to the dynamic variables,

$$\left(\frac{\partial \mathcal{H}_{\text{Nosé}}}{\partial \vec{r}_i'}\right) = \left(\frac{\partial \mathcal{U}}{\partial \vec{r}_i'}\right) = -\frac{\mathrm{d}\vec{P}_i'}{\mathrm{d}t'}, \tag{155}$$

$$\left(\frac{\partial \mathcal{H}_{\text{Nosé}}}{\partial \vec{P}_i'}\right) = \frac{\vec{P}_i'}{m_i\bar{s}'^2} = \frac{\mathrm{d}\vec{r}_i'}{\mathrm{d}t'}, \tag{156}$$

$$\left(\frac{\partial \mathcal{H}_{\text{Nosé}}}{\partial \bar{s}'}\right) = \frac{k_{\text{B}}T_{\text{t}}N_{\text{f}}}{\bar{s}'} - \sum_{i=1}^{N} \frac{\|\vec{P}_i'\|^2}{\bar{s}'^3 m_i} = -\frac{\mathrm{d}P_{\bar{s}}'}{\mathrm{d}t'}, \tag{157}$$

$$\left(\frac{\partial \mathcal{H}_{\text{Nosé}}}{\partial P_{\bar{s}}'}\right) = \frac{P_{\bar{s}}'}{Q_{\bar{s}}} = \frac{\mathrm{d}\bar{s}'}{\mathrm{d}t'}. \tag{158}$$

Combining Eq. (151) and Eq. (156) suggests that $\bar{s}'$ is a time-scaling parameter [40],

$$\mathrm{d}t = \frac{\mathrm{d}t'}{\bar{s}'}. \tag{159}$$

which will vary during the simulation as indicated by Eq. (158). A fluctuating time step is inconvenient when seeking to determine dynamic properties of the system. As a result, Nosé's equations of motion are commonly used in the formulation of Hoover [40]. Hoover

introduced the so-called friction coefficient [40],

$$\xi = \frac{1}{\bar{s}} \frac{\mathrm{d}\bar{s}}{\mathrm{d}t}, \tag{160}$$

and derived the *Nosé-Hoover equations of motion* [40],

$$\frac{\mathrm{d}\vec{r}_i}{\mathrm{d}t} = \frac{\vec{P}_i}{m_i}, \tag{161}$$

$$\frac{\mathrm{d}\vec{P}_i}{\mathrm{d}t} = \frac{\mathrm{d}\vec{P}_i{}'}{\mathrm{d}t'} - \xi \vec{P}_i, \tag{162}$$

$$\frac{\mathrm{d}\bar{s}}{\mathrm{d}t} = \xi \bar{s}', \tag{163}$$

$$\frac{\mathrm{d}\xi}{\mathrm{d}t} = \frac{1}{Q_{\bar{s}}} \left(\sum_{i=1}^{N} \frac{\|\vec{P}_i{}'\|^2}{\bar{s}^2 m_i} - k_\mathrm{B} T_\mathrm{t} N_\mathrm{f} \right). \tag{164}$$

We are able to simplify Eq. (164),

$$\frac{\mathrm{d}\xi}{\mathrm{d}t} = \frac{N_\mathrm{f} k_\mathrm{B}}{Q_{\bar{s}}} \left[T - T_\mathrm{t} \right], \tag{165}$$

which states that the rate of change of the friction coefficient depends on the difference between the instantaneous temperature, T, and the target temperature, T_t. In addition, it indicates that $Q_{\bar{s}}$ determines the coupling strength between the system and the heat bath.

2.3.3 Force Fields in Molecular Simulations

So far, we have skipped over how to determine the potential energies or forces in classical molecular simulations. We will change that in the following. The potential energy of a (classical) system is commonly approximated by using a set of functions and corresponding interaction parameters: the *force field*. In most force fields, the potential energy is split into contributions due to bonded, $\mathcal{U}_\mathrm{intra}$, and non-bonded interactions, $\mathcal{U}_\mathrm{nb}$,

$$\mathcal{U} = \mathcal{U}_\mathrm{intra} + \mathcal{U}_\mathrm{nb}. \tag{166}$$

Bonded interactions are often described as a sum of energy contributions due to bond stretching, $\mathcal{U}_\mathrm{bond}$, bond angle bending, $\mathcal{U}_\mathrm{bend}$, and torsion, $\mathcal{U}_\mathrm{tor}$,

$$\mathcal{U}_\mathrm{intra} = \mathcal{U}_\mathrm{bond} + \mathcal{U}_\mathrm{bend} + \mathcal{U}_\mathrm{tor}, \tag{167}$$

while non-bonded interactions are commonly described by Lennard-Jones, $\mathcal{U}_\mathrm{LJ}$, and electrostatic interactions, $\mathcal{U}_\mathrm{elec}$,

$$\mathcal{U}_\mathrm{nb} = \mathcal{U}_\mathrm{LJ} + \mathcal{U}_\mathrm{elec}. \tag{168}$$

Note that more advanced force fields may use additional terms to describe bonded or non-bonded interactions. One of the most important aspects of a force field is its transferability. The functional form and set of parameters should be applicable to a variety of problems beyond those for which the force field was developed and tested ([29], pp. 231–232). The *Transferable Potentials for Phase Equilibria* (TraPPE) [42] are an example for this type of force fields. It was developed to best reproduce vapor-liquid properties of organic or small inorganic molecules – typically vapor-liquid coexistence densities, vapor pressures, and critical properties [43]. Most of the TraPPE force fields use:

- fixed bond lengths,

$$\mathcal{U}_{\text{bond}} = 0, \tag{169}$$

- a sum of harmonic potentials to describe the energy due to bond angle bending,

$$\mathcal{U}_{\text{bend}} = \sum_{\text{angles}} \frac{C_{\text{bend}}}{2}(\theta - \theta_{\text{eq}}), \tag{170}$$

  where C_{bend} is the bend constant, θ is the instantaneous bond angle, and θ_{eq} is the equilibrium bond angle,

- a sum of four-term Fourier expansions to describe the energy due to torsion,

$$\mathcal{U}_{\text{tor}} = C_{\text{tor},0} + C_{\text{tor},1}[1 + \cos(\Phi)] + C_{\text{tor},2}[1 - \cos(2\Phi)] + C_{\text{tor},3}[1 + \cos(3\Phi)], \tag{171}$$

  where $C_{\text{tor},0}, C_{\text{tor},1}, C_{\text{tor},2}$, and $C_{\text{tor},3}$ are the torsion constants and Φ is the torsion angle,

- a sum of Lennard-Jones (6,12) pair potentials to describe the energy due to short-range interactions,

$$\mathcal{U}_{\text{LJ}} = \sum_{\text{non-bonded}} 4\epsilon_{i\text{-}j}\left[\left(\frac{\bar{\sigma}_{i\text{-}j}}{d_{i\text{-}j}}\right)^{12} - \left(\frac{\bar{\sigma}_{i\text{-}j}}{d_{i\text{-}j}}\right)^{6}\right], \tag{172}$$

  where $\epsilon_{i\text{-}j}$ is the Lennard-Jones well depth, $\bar{\sigma}_{i\text{-}j}$ is the Lennard-Jones distance, and $d_{i\text{-}j}$ is the distance between interaction sites i and j, and

- a sum of Coulomb potentials to describe the energy due to electrostatic interactions,

$$\mathcal{U}_{\text{elec}} = \sum_{\text{non-bonded}} \frac{\tilde{q}_i \tilde{q}_j}{4\pi\epsilon_0 d_{i\text{-}j}}, \tag{173}$$

  where $\tilde{q}_i$ and $\tilde{q}_j$ are the partial charges of i or j, and ϵ_0 is the vacuum permittivity.

Note that TraPPE also includes all bonded atoms in the summations in Eq. (172) and Eq. (173) that are separated by more than three bonds [43]. Lennard-Jones parameters between unlike interaction sites, i and j, are determined by the Lorentz-Berthelot mixing rules [44, 45],

$$\bar{\sigma}_{i\text{-}j} = \frac{\bar{\sigma}_i + \bar{\sigma}_j}{2}, \tag{174}$$

$$\epsilon_{i\text{-}j} = \sqrt{\epsilon_i \epsilon_j}. \tag{175}$$

The TraPPE force fields often use a *united atom* (UA) representation of organic molecules, in which each carbon atom is combined with its bonded hydrogen atoms to form a single interaction site [46]. The TraPPE-UA force fields immensely reduce the number of interaction pairs for which contributions to the potential energy need to be evaluated. By design, TraPPE is able to accurately describe phase transitions and is thus often used to study adsorption in nanoporous materials ([47], pp. 105–106). In adsorption simulations, the Lennard-Jones parameters for the atoms that constitute the nanoporous material are commonly taken from generic force fields, such as the *Universal Force Field* (UFF) [48] or *DREIDING* force field [49], while partial charges need to be computed individually, because they strongly depend on the distribution of electrons and protons in the material. Partial charges are not observable properties. As a consequence, there exist many ways to compute them. We will (roughly) divide them into two groups:

1. computing partial charges from quantum mechanical computations, for example via:
   (a) Mulliken population analysis [50],
   (b) Electrostatic potential (ESP) methods ([29], pp. 189–191), or
   (c) Density derived electrostatic and chemical (DDEC) charge methods [51], and
2. computing partial charges based on information about atoms and the way in which these atoms are connected, for example via:
   (a) charge equilibration (Qeq) [52], or
   (b) extended charge equilibration (EQeq) [53].

(1.) A short introduction to quantum mechanics and density functional theory (DFT) is provided in Appendix F. These techniques form the foundation for the methods used to determine partial charges discussed in the following. (1a) The *Mulliken population analysis* will be illustrated using the example of a two-atomic molecule, assuming that the molecular orbital, ψ, can be represented by a linear combination of two atomic orbitals, φ_A and φ_B,

$$\psi(\vec{r}) = w_\mathrm{A}\varphi_\mathrm{A}(\vec{r}) + w_\mathrm{B}\varphi_\mathrm{B}(\vec{r}), \tag{176}$$

where w_A and w_B are the weights of the atomic orbitals. The electron distribution, $\rho_{e^-}(\vec{r})$, can be determined based on *Born's rule*,

$$\rho_{e^-}(\vec{r}) = \|\psi(\vec{r})\|^2 = w_\mathrm{A}^2\varphi_\mathrm{A}^2(\vec{r}) + w_\mathrm{B}^2\varphi_\mathrm{B}^2(\vec{r}) + 2w_\mathrm{A}w_\mathrm{B}\varphi_\mathrm{A}(\vec{r})\varphi_\mathrm{B}(\vec{r}). \tag{177}$$

Integration of this equation yields the number of electrons in the molecule,

$$N_{e^-} = w_\mathrm{A}^2 \int_{-\infty}^{\infty} \varphi_\mathrm{A}^2(\vec{r})\mathrm{d}\vec{r} + w_\mathrm{B}^2 \int_{-\infty}^{\infty} \varphi_\mathrm{B}^2(\vec{r})\mathrm{d}\vec{r} + 2w_\mathrm{A}w_\mathrm{B} \int_{-\infty}^{\infty} \varphi_\mathrm{A}(\vec{r})\varphi_\mathrm{B}(\vec{r})\mathrm{d}\vec{r}. \tag{178}$$

Mulliken interpreted the result in Eq. (178) as follows: the first two integrals on the right-hand side contain the number of electrons that belong to A or B, while the third integral – the *overlap integral* – contains the number of electrons that are shared by A and B [50]. Mulliken evenly distributed the electrons in the overlap integral between atoms A and B. Thus, the partial charge of A, $\tilde{q}_{\mathrm{A}}$, is calculated by:

$$\tilde{q}_{\mathrm{A}} = q_{\mathrm{A}}^{+} - \left[w_{\mathrm{A}}^{2} \int \varphi_{\mathrm{A}}^{2}(\vec{r}) + w_{\mathrm{A}} w_{\mathrm{B}} \int \varphi_{\mathrm{A}}(\vec{r}) \varphi_{\mathrm{B}}(\vec{r}) \mathrm{d}\vec{r} \right], \tag{179}$$

where q_{A}^{+} is the nucleus' charge for atom A. Mulliken's approach of evenly distributing the electrons in the overlap integral is arbitrary and less chemical meaningful for atoms with different electron affinities. In addition, Mulliken charges depend on the constitution of the molecule rather than on reproducing electrostatic interaction properties. For this reason, Mulliken charges possess little applications in molecular simulations. (1b) Often, partial charges are used that reproduce the electrostatic potential (*ESP*) surrounding the molecule. The electrostatic potential, $\phi(\vec{r})$, is a continuous, observable property defined as the force acting on a unit positive charge at any position in the system and can be calculated by ([29], p. 189):

$$\phi(\vec{r}) = \frac{1}{4\pi\epsilon_0} \left[\sum_{A=1}^{\check{M}} \frac{q_{\mathrm{A}}^{+}}{\|\vec{r} - \vec{r}_{\mathrm{A}}\|} - \int \frac{\rho_{e^-}(\vec{r})}{\|\vec{r} - \vec{r}_{e^-}\|} \mathrm{d}\vec{r}_{e^-} \right], \tag{180}$$

where $\check{M}$ is the number of atoms in the molecule, $\vec{r}_{\mathrm{A}}$ is the position of atom A, and $\vec{r}_{e^-}$ is the electron position. ESP methods calculate the electrostatic potential surrounding the molecule at discrete points in space and fit partial charges that best reproduce ϕ at these points,

$$\tilde{f}_{\mathrm{t}} = \sum_{\mathrm{p}_i} \omega_{\mathrm{p}_i} (\phi_{\mathrm{p}_i}^{*} - \phi_{\mathrm{p}_i}^{\mathrm{calc}}), \tag{181}$$

where $\tilde{f}_{\mathrm{t}}$ is target function to be minimized, p_i is a grid point for which $\phi_{\mathrm{p}_i}^{*}$ is calculated according to Eq. (180), ω_{p_i} is the weight for p_i, and $\phi_{\mathrm{p}_i}^{\mathrm{calc}}$ is the electrostatic potential at p_i calculated for the set of point charges. In essence, ESP methods differ in the way on how Eq. (181) is solved and which points are used for fitting. ESP charges are well-suited for describing interactions, but may produce chemically less reasonable partial charges for embedded atoms ([29], p. 191). (1c) The *DDEC* charge method developed by Manz and coworkers [51] is among the most recent approaches to determine partial charges, and provides several benefits. DDEC charges are [51]:

- able to accurately describe electrostatic interactions,
- applicable to periodic or non-periodic, as well as porous or non-porous systems, and
- chemically meaningful.

Manz and coworkers point out that the DDEC charge method is not so much the result of a mathematical approach as it is the outcome of trial and error [54]. Despite its novelty,

DDEC charges are often used to study adsorption in nanoporous materials ([47], p. 110).

(2.) Methods that determine partial charges based on information about atoms and the way in which they are bonded are very fast, but often suffer from lower accuracy ([29], pp. 192–195). (2a) One example is the *Qeq* method developed by Rappé and Goddard [52]. Starting point for the derivation of the Qeq method is a second-order Taylor expansion of the electrostatic energy of atom A, $\mathcal{U}_{\text{elec,A}}$, around $\tilde{q}_{\text{A}}^{*} = 0$,

$$\mathcal{U}_{\text{elec,A}}(\tilde{q}_{\text{A}}) = \mathcal{U}_{\text{elec,A}}(0) + \left(\frac{\partial \mathcal{U}_{\text{elec,A}}(\tilde{q}_{\text{A}})}{\partial \tilde{q}_{\text{A}}}\right)_{\tilde{q}_{\text{A}}^{*}=0} \tilde{q}_{\text{A}} + \frac{1}{2}\left(\frac{\partial^2 \mathcal{U}_{\text{elec,A}}(\tilde{q}_{\text{A}})}{\partial \tilde{q}_{\text{A}}^2}\right)_{\tilde{q}_{\text{A}}^{*}=0} \tilde{q}_{\text{A}}^2. \tag{182}$$

Assuming a positively or negatively charged atom, we obtain:

$$\mathcal{U}_{\text{elec,A}}(+1) = \mathcal{U}_{\text{elec,A}}(0) + \left(\frac{\partial \mathcal{U}_{\text{elec,A}}(\tilde{q}_{\text{A}})}{\partial \tilde{q}_{\text{A}}}\right)_{\tilde{q}_{\text{A}}^{*}=0} + \frac{1}{2}\left(\frac{\partial^2 \mathcal{U}_{\text{elec,A}}(\tilde{q}_{\text{A}})}{\partial \tilde{q}_{\text{A}}^2}\right)_{\tilde{q}_{\text{A}}^{*}=0}, \tag{183}$$

or

$$\mathcal{U}_{\text{elec,A}}(-1) = \mathcal{U}_{\text{elec,A}}(0) - \left(\frac{\partial \mathcal{U}_{\text{elec,A}}(\tilde{q}_{\text{A}})}{\partial \tilde{q}_{\text{A}}}\right)_{\tilde{q}_{\text{A}}^{*}=0} + \frac{1}{2}\left(\frac{\partial^2 \mathcal{U}_{\text{elec,A}}(\tilde{q}_{\text{A}})}{\partial \tilde{q}_{\text{A}}^2}\right)_{\tilde{q}_{\text{A}}^{*}=0}. \tag{184}$$

Subtracting Eq. (184) from Eq. (183) yields the electronegativity of A,

$$\chi_{\text{A}} = \left(\frac{\partial \mathcal{U}_{\text{elec,A}}(\tilde{q}_{\text{A}})}{\partial \tilde{q}_{\text{A}}}\right)_{\tilde{q}_{\text{A}}^{*}=0} = \frac{\mathcal{U}_{\text{elec,A}}(+1) - \mathcal{U}_{\text{elec,A}}(-1)}{2}, \tag{185}$$

while adding Eq. (184) to Eq. (183) yields the idempotential of A,

$$J_{\text{A}} = \left(\frac{\partial^2 \mathcal{U}_{\text{elec,A}}(\tilde{q}_{\text{A}})}{\partial \tilde{q}_{\text{A}}^2}\right)_{\tilde{q}_{\text{A}}^{*}=0} = \mathcal{U}_{\text{elec,A}}(+1) + \mathcal{U}_{\text{elec,A}}(-1). \tag{186}$$

Values for χ and J are provided in the literature. By substituting Eq. (185) and Eq. (186) in Eq. (182), we obtain [52]:

$$\mathcal{U}_{\text{elec,A}}(\tilde{q}_{\text{A}}) = \mathcal{U}_{\text{elec,A}}(0) + \chi_{\text{A}}\tilde{q}_{\text{A}} + \frac{1}{2}J_{\text{A}}\tilde{q}_{\text{A}}^2, \tag{187}$$

while the total electrostatic energy of the system is provided by:

$$\mathcal{U}_{\text{elec}}(\tilde{q}_1, \cdots, \tilde{q}_{\text{N}}) = \sum_{A=1}^{N} \mathcal{U}_{\text{elec,A}}(\tilde{q}_{\text{A}}) + \sum_{A=1}^{N}\sum_{B>A}^{N} \frac{\tilde{q}_{\text{A}}\tilde{q}_{\text{B}}}{4\pi\epsilon_0 d_{i\text{-}j}}, \tag{188}$$

where the first term on the right-hand side is the sum of individual electrostatic energies, and the second term on the right-hand side is the sum of intermolecular electrostatic energies. At equilibrium, the electrostatic energy of the system will be minimal, and the partial derivatives of $\mathcal{U}_{\text{elec}}(q_1, \cdots, q_{\text{N}})$ with respect to each partial charge will be equal,

$$\frac{\partial \mathcal{U}_{\text{elec}}(\tilde{q}_1, \cdots, \tilde{q}_{\text{N}})}{\partial \tilde{q}_1} = \frac{\partial \mathcal{U}_{\text{elec}}(\tilde{q}_1, \cdots, \tilde{q}_{\text{N}})}{\partial \tilde{q}_2} = \cdots = \frac{\partial \mathcal{U}_{\text{elec}}(\tilde{q}_1, \cdots, \tilde{q}_{\text{N}})}{\partial \tilde{q}_{\text{N}}}, \tag{189}$$

leading to $N-1$ linear equations. In combination with:

$$\tilde{q}_{\text{tot}} = \sum_{A=1}^{N} \tilde{q}_{\text{A}}, \tag{190}$$

where $\tilde{q}_{\text{tot}}$ is the total charge of the system, we obtain N equations for the N unknown partial charges we wish to determine. The Qeq method solves these equations by iteration [52]. (2b) Snurr and coworkers [53] improved the Qeq method by avoiding iteration through directly solving the N equations via Matrix calculus and a Taylor expansion around $\tilde{q}^* \neq 0$, which increases the accuracy of the method for strongly charged atoms. They dubbed this method as *extended charge equilibration* (EQeq) method.

Evaluating Interactions

Molecular simulations aim to provide information about the properties of a macroscopic system, often comprising more than 10^{20} particles. Even with today's computation power, it is not possible to store such a large number of particle coordinates in memory. For this reason, molecular simulations are often performed for a representative sample of the macroscopic system with a relatively small number of particles. When designing this *model system*, it is important to choose boundaries that mimic the presence of the surrounding infinite bulk in the macroscopic system. One way is by using *periodic boundary conditions* (see Fig. 15).

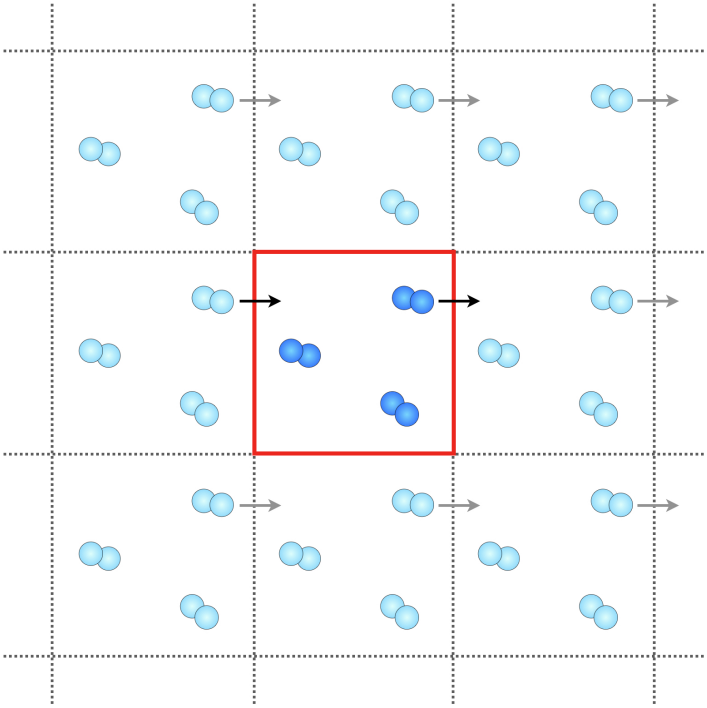

Fig. 15: Two-dimensional illustration of periodic boundary conditions. An infinite bulk is mimicked by replicating the model system (red-framed) in every direction. Every particle in the model system interacts with every particle in the periodic image systems (faint). When a particle leaves the model system on one side, one of its images enters from the opposite side. Source: author's illustration inspired by Frenkel and Smit, 2001, p. 34 [26].

When using periodic boundary conditions, the model system is replicated in every direction to mimic the macroscopic system. Each particle in the replica cells moves exactly like its original in the model system. When a particle leaves the model system on one side, one of its images enters from the opposite side. Only the coordinates of the particles in the model system need to be stored in memory, because the coordinates of the particles in the replica cell can be calculated by translational vectors,

$$\vec{n} = l_x\vec{n}_x + l_y\vec{n}_y + l_z\vec{n}_z, \tag{191}$$

where l_x, l_y, and l_z are integers, and $\vec{n}_x$, $\vec{n}_y$, and $\vec{n}_z$ are the box vectors of the model system. Each particle in the model system interacts with each (image) particle in the replica cells. Without further simplifications, we would still need to determine the interactions between the particles of the model system and each particle of the (now mimicked) macroscopic system. Short-range interactions – such as Lennard-Jones interactions – can be truncated at a certain cutoff distance, d_{cut}, without distorting the results,

$$\mathcal{U}_{\text{LJ}}(d_{i\text{-}j}) = \begin{cases} \mathcal{U}_{\text{LJ}}(d_{i\text{-}j}) & \text{for } d_{i\text{-}j} \leq d_{\text{cut}}, \\ 0 & \text{otherwise.} \end{cases} \tag{192}$$

Electrostatic interactions – such as Coulomb interactions – can not be efficiently truncated, one reason is that Coulomb energies decay with $1/d_{i\text{-}j}$ and are thus considered long-range interactions ([26], p. 291). Electrostatic interactions are often evaluated using the *Ewald summation* technique [55] illustrated in Fig. 16. The idea behind the Ewald summation is to surround each point charge, $\tilde{q}_i$, by a smoothly varying compensation charge distribution, ρ_i, of equal magnitude but opposite sign (see Fig. 16b). The compensation charge cloud is often chosen to be Gaussian,

$$\rho_i = \tilde{q}_i \frac{\varpi^3}{\sqrt{\pi^3}} e^{-\varpi^2 d_{\text{c}}^2}, \tag{193}$$

where ϖ is the Ewald splitting parameter and d_{c} the distance to the center of the distribution. The Gaussian distribution in Eq. (193) has a standard deviation of $\sigma_{d_{\text{c}}} = \sqrt{2}/\varpi$ and a zero mean. The interactions of an ion with an electrostatic potential due to compensated charges are short-ranged and can thus be truncated by a spherical cutoff. However, our aim is to determine electrostatic interactions of ions with *point* charges, not *compensated* charges ([26], pp. 293–294). Thus, the electrostatic interaction of ions with the compensation charge distribution has to be subtracted (see Fig. 16c). Since the compensation charge cloud is a periodic, smoothly varying function, its contribution to the electrostatic potential can be solved by a rapidly converging sum in Fourier space. The Coulomb potential becomes ([29], p. 338):

$$\mathcal{U}_{\text{elec}} = \sum_{\vec{n}} \sum_{i} \sum_{j>i} \frac{\tilde{q}_i \tilde{q}_j}{4\pi\epsilon_0} \frac{\operatorname{erfc}(\varpi \|\vec{d}_{i\text{-}j} + \vec{n}\|)}{\|\vec{d}_{i\text{-}j} + \vec{n}\|} \tag{194}$$

$$+\frac{1}{2V}\sum_{\vec{k}\neq 0}\sum_{i}\sum_{j}\frac{\tilde{q}_i\tilde{q}_j}{\epsilon_0}\frac{1}{\|\vec{k}\|^2}e^{-\frac{\|\vec{k}\|^2}{4\varpi^2}}\cos(\vec{k}\bullet\vec{d}_{i\text{-}j}) \tag{195}$$

$$-\frac{\varpi}{\sqrt{\pi}}\sum_{i}\frac{\tilde{q}_i^2}{4\pi\epsilon_0}, \tag{196}$$

where $\vec{k}$ are the reciprocal space vectors,

$$\vec{k}=\frac{2\pi}{V}\left[\check{l}_x(\vec{n}_y\times\vec{n}_z)+\check{l}_y(\vec{n}_z\times\vec{n}_x)+\check{l}_z(\vec{n}_x\times\vec{n}_y)\right]. \tag{197}$$

The three terms on the right-hand side of Eq. 196 are (1) the interaction energy between ions and the electrostatic potential due to compensated charges (truncated by a real-space cutoff d_{cut} with $d_{i\text{-}j}\leq d_{\text{cut}}$), (2) the interaction energy between ions and the electrostatic potential due to compensation charge clouds (truncated by maximum number of simulation box replica $\check{l}_{\text{cut}}$ with $\check{l}_x+\check{l}_y+\check{l}_z\leq\check{l}_{\text{cut}}$), and (3) the correction for ion self-interactions implicit in the second term to ensure periodicity of the compensation charge distribution. By using the formulation of the Ewald summation in Eq. (196), the sphere of simulation boxes being built up is assumed to be surrounded by a conducting medium ([28], p. 157). Appropriate values for the Ewald splitting parameter, ϖ, and maximum number of simulation box replica, $\check{l}_{\text{cut}}$, can be estimated based on the desired accuracy, ϵ_{acc}, of the Ewald Coulombic energy ([26], pp. 304–306):

$$\epsilon_{\text{acc}}=\exp(-\tilde{s}^2)/\tilde{s}^2, \tag{198}$$

$$\varpi=\frac{\tilde{s}}{d_{\text{cut}}}, \tag{199}$$

$$\check{l}_{\text{cut}}=\frac{\varpi\tilde{s}}{\pi}\min(\vec{n}_x,\vec{n}_y,\vec{n}_z), \tag{200}$$

where $\tilde{s}$ is an integer.

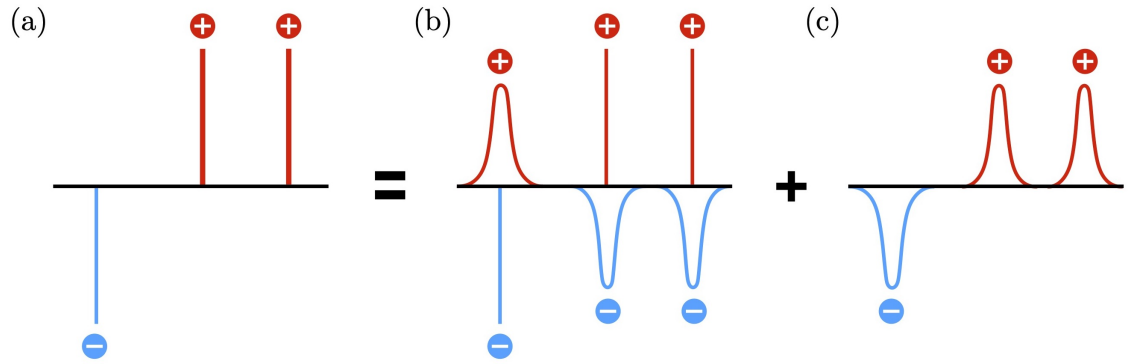

Fig. 16: The Ewald summation splits (a) a set of point charges into (b) a set of screened point charges and (c) the canceling screening background. Source: reproduction of Author's illustration [56] inspired by Frenkel and Smit, 2002, p. 294 [26].

3 Separating Alcohols and Alcohol-Water Mixtures by Exploiting Entropic Effects in CAU-10

3.1 Motivation

Short-chain alcohols obtained by fermentation play a key role in the transformation toward green chemistry, because of their use as fuel additives, fuels, or for their conversion into olefins. The fermentation broth is often a highly diluted aqueous solution that requires separation. Common separation strategies exploit differences of the components' boiling points (distillation) or diffusion rates through porous membranes (pervaporation) [57]. Both methods have their limitations. Simple distillation is often incapable of providing high-purity alcohols, because most of them form azeotropic mixtures with water. Moreover, distillation is inefficient, both in terms of environmental compatibility and economic profitability. Pervaporation, on the other hand, is more energy efficient than distillation but less versatile. The choice of membrane is based on the composition of the mixture to be separated, which can vary greatly in fermentation processes [58, 59]. Adsorption is an alternative to distillation and pervaporation – even if its application has so far often been limited to the drying of alcohol distillates. One of the several advantages of adsorption is its ability to separate alcohol mixtures from the liquid phase, making it well suited for separating fermentation broths. In the following, we will study the separation of aqueous alcohol mixtures by liquid phase adsorption in CAU-10 – an ultramicroporous MOF with internal hydrophobicity. The prime objective of this study is to understand the impact of water on the separation of alcohols from the liquid phase. To this end, we perform grand canonical Monte Carlo simulations to predict both the unary gas adsorption isotherms of linear short-chain alcohols or water, as well as the multicomponent liquid phase adsorption isotherms of water-alcohol mixtures.

3.2 State of the Research

The separation of alcohol mixtures by adsorption has been intensively studied. For instance, Krishna and van Baten [60] showed that the all-silica CHA zeolite is capable of efficiently separating linear short-chain alcohols based on differences in their saturation capacity. The authors performed CBMC simulations at 300 K to obtain the unary vapor adsorption isotherms of methanol, ethanol, n-propanol, n-butanol, n-pentanol, or n-hexanol, as well as the binary vapor adsorption isotherms for the equimolar mixtures: methanol/ethanol, ethanol/n-propanol, n-butanol/n-pentanol, n-pentanol/n-hexanol, or ethanol/n-hexanol. The unary vapor adsorption isotherms exhibited a sigmoid shape for a logarithmic scale on the pressure axis. As the chain length of the alcohols increases, the inflection point of the sigmoid curve shifts to lower pressures (due to increasing host-guest interactions), while the saturation capacity decreases (due to the decreasing packing efficiencies of the components in the pore channels). The saturation capacities are shown in Tab. 2, and their course can be described

by the following sequence:

$$\text{methanol} > \text{ethanol} > \text{n-propanol} = \text{n-butanol} > \text{n-pentanol} = \text{n-hexanol}.$$

The vapor adsorption isotherms for the equimolar mixtures: methanol/ethanol, ethanol/n-propanol, n-butanol/n-pentanol, or ethanol/n-hexanol exhibit three characteristics:

1. At low loading, the longer alcohol is preferentially adsorbed, because the separation is governed by differences in the adsorption strength.

2. As the loading increases, the differences in packing efficiency come into play. A *selectivity reversal* occurs, resulting in the longer alcohol being desorbed in favor of the shorter one.

3. At high loading, the mole fraction of the shorter alcohol in the adsorbed phase is significantly higher than that of the longer alcohol. For some mixtures, the longer alcohol is completely excluded from the adsorbed phase.

Krishna and van Baten concluded that selectivity reversal is guaranteed under saturation conditions, such as present in adsorption from liquids. On a side note, no selectivity reversal was observed for mixtures of components that possess the same saturation capacity [60].

Tab. 2: Saturation capacities for linear short-chain alcohols in all-silica CHA zeolite at 300 K [60].

Species	**Saturation capacities** [mmol/g]
methanol	1.27
ethanol	0.92
n-propanol, n-butanol	0.46
n-pentanol, n-hexanol	0.23

The study of Krishna and van Baten implies that differences in the saturation capacity of the components will mainly be responsible for the separation of aqueous alcohol mixtures by liquid phase adsorption. Differences in the saturation capacity result from varying packing efficiencies of the guest species, for example due to [61]:

- **Size exclusion**, which favors the adsorption of whichever guest species can occupy the highest number of distinct adsorption sites at saturation of the adsorbent. An example is the separation of n-hexane and 3-methylpentane in silicalite [62] – a zeolite which consists of straight ellipsoidal and zigzag channels that cross at intersections. While n-hexane can adsorb in both the straight and zigzag channels, 3-methylpentane can only occupy the intersections. As a consequence, the saturation capacity of n-hexane is significantly higher than that of 3-methylpentane, leading to the almost

complete exclusion of 3-methylpentane from the adsorbed layer when adsorbing from an equimolar binary mixture of the two components at saturation.

- **Commensurate stacking**, which favors the adsorption of whichever guest species has a stacking arrangement commensurate with the dimensions of the one-dimensional channel. An example is the separation of styrene and ethylbenzene in MIL-47 [61] – a MOF which consists of diamond-shaped one-dimensional channels. While styrene (planar) can adsorb flat on the channel walls, ethylbenzene (non-planar) can not. This results in a more efficient packing of styrene compared to ethylbenzene, and the exclusion of ethylbenzene from the adsorbed layer when adsorbing from an equimolar binary mixture at saturation.

- **Face-to-face stacking**, which favors the adsorption of whichever guest species significantly decreases its footprint when reoriented. An example is the separation of styrene and ethylbenzene in AFI zeolite with one-dimensional channels [61]. At low loading, styrene and ethylbenzene adsorb flat on the channel walls. As the loading increases, both styrene and ethylbenzene undergo reorientation. While styrene (planar) can significantly reduce its footprint by stacking in piles, ethylbenzene (non-planar) can not. As a result, adsorption of styrene is strongly favored over that of ethylbenzene when adsorption occurs under saturation conditions.

Separations that exploit differences in the saturation capacity are driven by maximizing the configurational entropy. Krishna [63] provided a quantitative explanation for these *entropic separations* based on the statistical-mechanical Boltzmann expression for entropy,

$$S = k_{\mathrm{B}} \ln w, \tag{201}$$

where k_{B} is Boltzmann's constant and w is the number of configurations (in Sec. 2.2 dubbed as *number of realizations*). Eq. (201) applies to a macroscopic system with only equally probable microscopic states, i.e. one in which every state has the same energy. This is guaranteed if: (1) the channels of the porous material consist of one-dimensional lattices of $\tilde{M}$ adsorption sites where molecules can only adsorb flatly, (2) the adsorption energy per lattice site is independent of the adsorbed guest species, and (3) no adsorbate-adsorbate interactions are present. Qualitatively, Eq. (201) states that the guest species with the largest number of configurations has the highest configurational entropy. Krishna used the model of Dávila et al. [64],

$$w = \frac{(\tilde{M} - (k-1)N_k - (l-1)N_l)!}{(\tilde{M} - kN_k - lN_l)!N_k!N_l!}, \tag{202}$$

to express the number of configurations for adsorbing N_k k-mers and N_l l-mers – occupying k or l adsorptions sites per molecule. The author obtained the following quantitative expression

for the description of entropic separations by maximizing Eq. (201):

$$x_k = \frac{1}{1 + \frac{1}{\Omega}}, \tag{203}$$

where Ω was dubbed as *entropy factor*,

$$\Omega = \frac{\left(1 - (k-1)x_k\tilde{\Theta}_{\text{mix}} - (l-1)(1-x_k)\tilde{\Theta}_{\text{mix}}\right)^{l-k}}{\left(1 - kx_k\tilde{\Theta}_{\text{mix}} - l(1-x_k)\tilde{\Theta}_{\text{mix}}\right)^{l-k}}, \tag{204}$$

x_k is the adsorbed phase mole fraction of the k-mer, and $\tilde{\Theta}_{\text{mix}}$ is the mixture occupancy of adsorption sites,

$$\tilde{\Theta}_{\text{mix}} = \frac{kN_k + lN_l}{\tilde{M}}. \tag{205}$$

Eqs. (203) and (204) need to be solved simultaneously. In doing so, Krishna recognized that:

1. independent of the mixture occupancy, no entropic separation occurs for $k = l$, because $\Omega = 1$ and $x_k = 0.5$,
2. for $k < l$ and vanishing mixture occupancies ($\tilde{\Theta}_{\text{mix}} \to 0$), separation is not driven by entropy, because $\Omega \to 1$ and $x_k \to 0.5$, and
3. for $k < l$ and high mixture occupancies ($\tilde{\Theta}_{\text{mix}} \to 1$), separation is completely driven by entropy, because $\Omega \to \infty$ and $x_k \to 1$.

Krishna's model confirms the assumption of Krishna and van Baten that differences in the saturation capacity of the components will govern the separation of alcohol mixtures by liquid phase adsorption (i.e. for $\tilde{\Theta}_{\text{mix}} \to 1$) [63]. For aqueous mixtures, the entropic separation would probably result in the accumulation of water in the adsorbed layer – even for adsorbents with internal hydrophobicity. Evidence is provided by Zhang et al. [65], who studied the vapor phase separation of short-chain alcohols and water in ZIF-90. The authors measured the unary adsorption isotherms of methanol, ethanol, n-propanol, isopropanol, or n-butanol at 308 K and predicted the separation selectivity of liquid binary alcohol-water mixtures containing up to 5 mol% alcohol using the *ideal adsorbed solution theory* (IAST) [66], described in more detail in Appendix G. The unary alcohol isotherms exhibit type I characteristics according to the IUPAC classification, with saturation capacities of 10 mmol/g (methanol), 6 mmol/g (ethanol), or 4 mmol/g (n-propanol, isopropanol, or n-butanol), while the unary water isotherm confirms the internal hydrophobicity of ZIF-90 (type V isotherm according to the IUPAC classification) and indicates a saturation capacity of ca. 18 mmol/g. The significantly higher saturation capacity of water compared to the alcohols results in the preferential adsorption of water over the whole range of liquid phase mole fractions of water considered, as predicted by IAST. Based on the findings of Zhang et al., this chapter aims to investigate the entropic impact on the separation of alcohols or alcohol-water mixtures in CAU-10, an ultramicroporous MOF with internal hydrophobicity.

3.3 Methodology

Grand canonical Monte Carlo simulations were performed to predict:

1. the vapor phase adsorption of ethanol, n-butanol, n-hexanol, or water, and
2. the liquid phase adsorption for the mixtures: ethanol/n-butanol, n-butanol/n-hexanol, water/ethanol, water/n-butanol, water/n-hexanol, or water/ethanol/n-butanol

in CAU-10 at 298.15 K. CAU-10, [Al(OH)(benzene-1,3-dicarboxylate)], was first synthesized in 2012 by Stock and coworkers [5] by linking chains of cis-connected corner-sharing $AlO_4(OH)_2$ polyhedra via 1,3-benzenedicarboxylate (isophthalic acid). The pore geometry is shown in Fig. 17 and is best described as being a sequence of rectangular cavities (*cuboids*) that are rotated 90° to each other. Previous studies on isostructural MIL-160, [Al(OH)(furan-2,5-dicarboxylate)], revealed that this pore geometry causes highly efficient entropy-based separations of hexane isomers at room temperature [67, 68]. In contrast to MIL-160, CAU-10 has a larger pore size – which is commensurate with the size of short-chain alcohols – and a more pronounced internal hydrophobicity – confirmed by type V water adsorption isotherms with inflection points at $p/p_0 \approx 0.08$ (MIL-160) or $p/p_0 \approx 0.18$ (CAU-10) [7]. CAU-10 is thus a good candidate for studying the entropic separation of aqueous alcohol mixtures. Tab. 3 provides a brief summary of the geometric properties of CAU-10 and MIL-160.

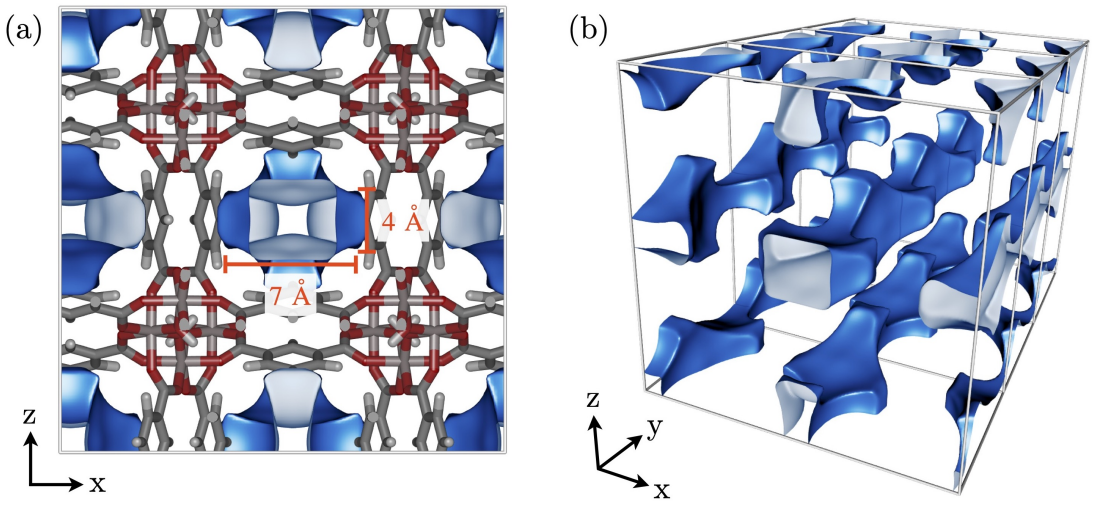

Fig. 17: Pore geometry of CAU-10 (colored blue). (a) Front view with shown pore dimensions. Coloring of atoms following the Jmol coloring scheme – i.e. aluminum - pink swan, carbon - gray, hydrogen - white, oxygen - red. (b) Diagonal view. The pore geometry was automatically calculated with iRASPA [69] using nitrogen as probe molecule with Lennard-Jones interactions only.

Tab. 3: Geometric properties of CAU-10 [5] and MIL-160 [70].

	CAU-10	**MIL-160**
Pore Dimensions [Å]	ca. 4×7	ca. 2.5×5
Pore Volume [ml/g]	0.25	0.45
Surface Area [m^2/g]	564 – 656	776

GCMC simulations were carried out with the RASPA software package [31]. Five types of trial moves were used: (1) insertion/deletion, (2) rigid body translation, (3) rigid body rotation, (4) CBMC reinsertion, and (5) identity change of guest molecules. Unary vapor adsorption isotherms were calculated up to the vapor pressure, with gas phase fugacities automatically computed by RASPA using the Peng-Robinson equation of state [71]. Adsorption from binary or ternary liquid mixtures was simulated following the procedure described by Low and coworkers [72] (see Appendix H for details on the mixture compositions). The procedure mimics the conditions in the experiment, where the liquid mixture and adsorbent are kept in a sealed tube at constant temperature. The closed system consists of three phases: vapor, liquid, and solid. At thermodynamic equilibrium, the components in the system have the same chemical potential in all three phases. There is thus no need to explicitly simulate the liquid phase in GCMC simulations, because the chemical potential can be determined from the fugacity and composition of the vapor phase. We used the modified *Universal Quasichemical Functional Group Activity Coefficients (Dortmund)* (UNIFAC (Do)) model [73] to determine the vapor phase fugacities and compositions for every liquid mixture.[2] For binary or ternary mixtures containing water, 1000 water molecules were created and equilibrated in the canonical ensemble before running GCMC simulations with CFCMC to accelerate equilibration. Details on the number of Monte Carlo cycles are given in Tab. I1 of Appendix I. Non-bonded interactions were described by an interatomic potential consisting of Lennard-Jones and Coulomb interactions using periodic boundaries. Lennard-Jones interactions were truncated by means of a 12 Å cutoff, while the Ewald summation [55] with an accuracy of 10^{-6} was used to handle long-range Coulomb interactions. The simulation box consisted of $2 \times 2 \times 4$ unit cells of CAU-10, whose atomic positions were maintained fixed during the simulations. The atomic coordinates of CAU-10 were taken from previous work [74]. The Lennard-Jones parameters for each atom of the MOF were taken from the DREIDING force field [49] (organic linker) and the Universal force field [48] (metal atoms and their bridging ions). Following the work of Cadiau et al. [7], we neglected the Lennard-Jones contributions of the aluminum atoms and the hydrogen atoms of their bridging hydroxyl groups, as well as those of the hydrogen atoms of the benzene ring. These modifications to the force field allow water to form hydrogen bonds with the otherwise shielded hydroxyl groups that bridge the aluminum atoms. Similar interactions are observed if the flexibility of the isophthalate linker is accounted for in the force field, as shown by Grenev et al. [75]. Partial charges were determined by performing single point energy calculations followed by a Mulliken analysis [50] in DMol3 as implemented in Materials Studio 7.0 (BIOVIA) using the Perdew-Burke-Ernzerhof (PBE) generalized gradient approximation (GGA) functional [76] combined with a double-numeric quality basis set with polarization functions (DNP). Alcohols were modeled according to the TraPPE-UA model [77], while water was modeled according to the TIP4P/2005 model [78]. Lennard-Jones parameters between unlike atomic species were computed by means of the Lorentz-Berthelot mixing rules [44, 45].

[2] UNIFAC (Do) calculations were performed by Dr. Mandy Klauck; ORCID iD: 0000-0001-8922-8295.

Adsorption isotherms, snapshots, adsorption heats, and radial distribution functions (RDFs) give insight into the adsorption mechanism. RDFs were determined over 10^5 cycles of the production run. The simulated isotherms were fitted to the dual-site Langmuir-Sips model [60],

$$\Gamma_i = \Gamma_{i,1}^{\mathrm{m}} \frac{b_{i,1} p^{\nu_{i,1}}}{1 + b_{i,1} p^{\nu_{i,1}}} + \Gamma_{i,2}^{\mathrm{m}} \frac{b_{i,2} p^{\nu_{i,2}}}{1 + b_{i,2} p^{\nu_{i,2}}}, \tag{206}$$

where Γ_i is the loading of adsorbate i, $\Gamma_{i,1}^{\mathrm{m}}$ and $\Gamma_{i,2}^{\mathrm{m}}$ are the saturation capacities for adsorption of i on site 1 or 2, $b_{i,1}$ and $b_{i,2}$ are the Langmuir constants for adsorption of i on site 1 or 2, and $\nu_{i,1}$ and $\nu_{i,2}$ are the number of adsorbate molecules that can simultaneously adsorb on site 1 or 2. The simulation results suggest $\nu_{i,1} = \nu_{i,2} = 1$ for n-butanol or n-hexanol, and $\nu_{i,1} = \nu_{i,2} = 2$ for ethanol. The *curve_fit* function of the scipy python software package [79] was used to fit the constants in Eq. (206) to match the simulated isotherms by means of the Levenberg-Marquardt algorithm [80]. The fitted parameters are shown in Tab. K1 of Appendix K. Experimental isotherms for the unary adsorption of ethanol, n-butanol, and water at 298.15 K on CAU-10 were measured in collaboration with the research group of Dr. Jens Möllmer of the Institut für Nichtklassische Chemie e.V. (INC) in Leipzig. The experimental method used by the INC is described in Appendix J. The n-hexanol isotherm proved impossible to measure, because of the long time that is required to attain equilibrium. For comparison with the simulated unary isotherms, Eq. (21) was used to convert the specific relative surface excess measured in the experiment to the absolute loading determined in the simulations. The density of the gas phase, ρ_{g}, in Eq. (21) was estimated using the van-der-Waals equation of state,

$$\left(p + \tilde{a}\rho_{\mathrm{g}}^2\right)\left(\frac{1}{\rho_{\mathrm{g}}} - \tilde{b}\right) = RT, \tag{207}$$

where $\tilde{a}$ and $\tilde{b}$ are the van-der-Waals parameters. The van-der-Waals parameters were determined from the critical data taken from the CRC Handbook of Chemistry and Physics ([81], pp. 6-67–6-82). Eq. (207) was solved for ρ_{g} using the *roots* function of the numpy python package [82]. Eq. (31) was used to calculate the reduced surface excess for the liquid multicomponent mixtures. The simulation results for the liquid binary mixtures are compared to IAST calculations. IAST is based on the assumption of an ideal adsorbed phase [66]. This assumption is valid for chemically similar guest species (e.g. short-chain alcohols, as shown by Krishna and van Baten [60]), but probably not for alcohol/water mixtures. However, the deviation of the IAST predictions from the simulation results helps to identify cooperative adsorption phenomena. IAST calculations were performed with the pyIAST software package [83] using the unary adsorption isotherms obtained in simulations or experiments (simulated isotherm for n-hexanol only) and vapor phase fugacities of the liquid mixtures as input. The UNIFAC (Do) model was used to determine the vapor phase fugacities for every liquid mixture. Spreading pressures in pyIAST were determined by numerical quadrature using Eq. (G12). The first integral in Eq. (G12) was solved by assuming that the adsorption isotherm follows Henry's law for pressures lower than the lowest observed pressure in the adsorption isotherm. The second term in Eq. (G12) was solved using the trapezoid rule, and

the third integral was solved by linear interpolation of the adsorption isotherm. The adsorption capacity was assumed constant for pressures beyond the highest pressure observed in the adsorption experiment, because adsorption isotherms are sufficiently saturated. Hydrogen bonding between water and alcohol might lead to *co-adsorption*, meaning that the loading of a component of the mixture exceeds the loading observed for the pure component under similar thermodynamic conditions [84]. Co-adsorption is examined by comparing the component loadings obtained from the binary liquid simulations or IAST calculations with the component loadings occurring in unary adsorption at fugacities corresponding to the partial fugacities arising for the binary liquid mixtures. Eq. (206) was used in conjunction with the fitted parameters in Tab. K1 to determine the unary component loadings. The reduced surface excesses for the components of the ternary liquid mixture are presented in ternary contour plots. Contour lines were determined based on an interpolation that uses a marching squares algorithm in conjunction with centripetal Catmull–Rom splines [85] as implemented in the *create_ternary_contour* function of the plotly python package. Isometric log-ratio transformation of the composition data was applied to improve the accuracy of the interpolation. Lines parallel to the triangle edges have a constant amount of one component. By determining the reduced surface excess at different points on several of these parallel lines, we are able to evaluate the influence of the increase of one component on the separation of the two remaining components. The reduced surface excess obtained for each parallel line was fitted to the Bi-Langmuir function,

$$\Gamma_i^{\sigma(n)} = \check{h}_2 \left(\check{h}_1 \frac{K_1 x_i'}{1 + (K_1 - 1)x_i'} + (1 - \check{h}_1) \frac{K_2 x_i'}{1 + (K_2 - 1)x_i'} - x_i' \right), \tag{208}$$

where $\check{h}_1$, $\check{h}_2$, K_1, or K_2 are the fitting parameters and x_i' is the liquid phase mole fraction of component i [86]. The *curve_fit* function of the scipy python software package [79] was used to fit all the four constants in Eq. (208) to match the simulated reduced surface excess data by means of the Levenberg-Marquardt algorithm [79]. The fitted parameters are shown in Tab. K2 – Tab. K4 of Appendix K.

3.4 Results and Discussion

3.4.1 Single-Component Adsorption

Fig. 18 shows the simulated and experimental unary adsorption isotherms for water, ethanol, n-butanol, and n-hexanol in CAU-10 at 298.15 K using a linear scale on the abscissa for all adsorptives (Fig. 18a) or a logarithmic one (Fig. 18b) for alcohols only. The secondary ordinate indicates the number of molecules per CAU-10 unit cell ([molec/uc]). The simulated adsorption isotherms are able to describe the course of the experimental isotherm qualitatively (water) or quantitatively (alcohols). The water isotherms in Fig. 18a show no adsorption at low relative pressure and then a sudden increase in water adsorption up to 15 mmol/g (simulation) or 17.5 mmol/g (experiment) at a relative pressure of about 0.25

(simulation) or 0.18 (experiment) followed by low water uptake up to a saturation capacity of 17.5 mmol/g (simulation) or 21.5 mmol/g (experiment). The results of Grenev et al. [75] suggest that the increased hydrophilicity and higher saturation capacity in the experiment are caused by defects in CAU-10 that were not accounted for in our force field. The alcohol isotherms in Fig. 18a show a sharp increase in alcohol adsorption at a low relative pressure and quickly attain a plateau. The course of the alcohol isotherms is best described using the semi-logarithmic representation in Fig. 18b.

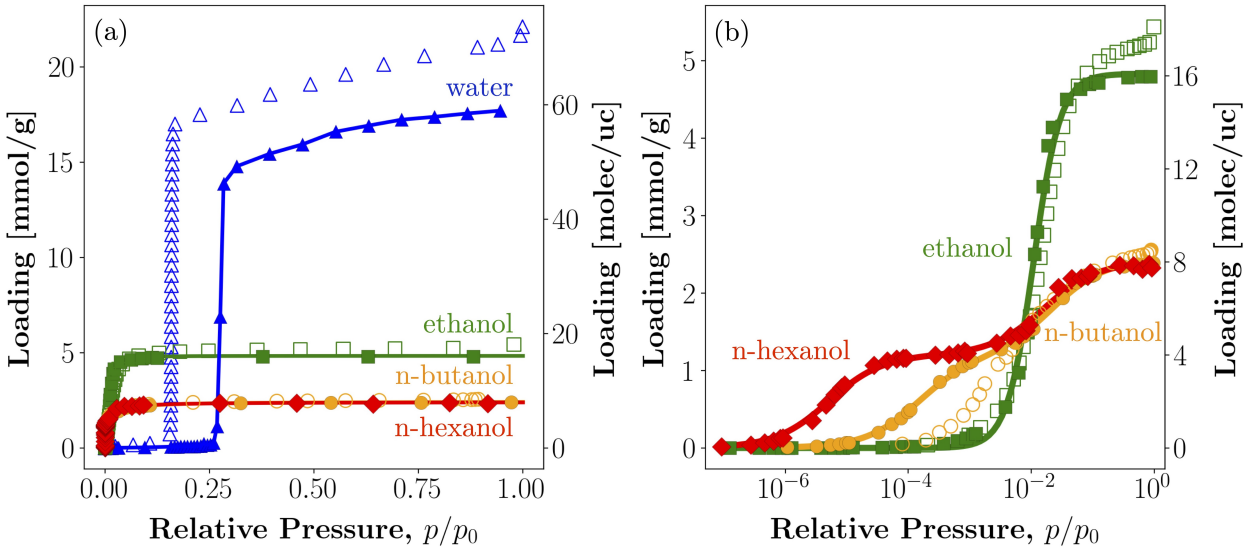

Fig. 18: Unary isotherms for water (blue triangles), ethanol (green squares), n-butanol (orange circles), and n-hexanol (red diamonds) in CAU-10 at 298.15 K: simulation (solid symbols) and experiment (outlined symbols). (a) Linear plot for all adsorptives or (b) a semi-logarithmic one for alcohols. The continuous solid lines are fits using linear interpolation (water) or the dual-site Langmuir-Sips model (alcohols). The following vapor pressures, obtained from the UNIFAC (Do) calculations, were used: $p_0 = 3173.2$ Pa (water), $p_0 = 7940.9$ Pa (ethanol), $p_0 = 926.2$ Pa (n-butanol), and $p_0 = 109.2$ Pa (n-hexanol).

The ethanol isotherms in Fig. 18b show no adsorption at low relative pressure and then a sudden increase in ethanol adsorption up to 4.8 mmol/g at a relative pressure of about 0.01. The simulated and experimental ethanol isotherms agree very well. The only relevant difference is the slight increase in ethanol uptake for $p/p_0 > 2 \cdot 10^{-2}$ observed in the experiment, which might be caused by adsorption/condensation on the external surface of CAU-10. The n-butanol isotherms in Fig. 18b show no adsorption at low relative pressure and then two elongated increases in n-butanol adsorption up to: (1) 1.2 mmol/g at a relative pressure of about 10^{-4} (simulation) or 10^{-3} (experiment) and (2) 2.4 mmol/g at a relative pressure of about 10^{-2}. The agreement between the simulated and experimental n-butanol isotherm is good. However, the first increase in n-butanol adsorption in the simulated isotherm occurs at a relative pressure which is an order of magnitude lower than that which is observed in the experiment, while the second increase is well described. A possible explanation for the difference between the simulated and experimental n-butanol isotherm will be presented when discussing the water/n-butanol mixture adsorption on p. 66 ff. The simulated n-hexanol adsorption in Fig. 18b isotherm shows no adsorption at low relative pressures and then two

elongated increases in n-hexanol adsorption up to: (1) 1.2 mmol/g at a relative pressure of about 10^{-5} and (2) 2.4 mmol/g at a relative pressure of about 10^{-2}. Both n-butanol and n-hexanol show an inflection in the isotherm at a loading of 1.2 mmol/g. An inflection in the isotherm is often an indication for the reorientation of the adsorbate molecules due to the filling of adsorption sites of different energies. Short-chain organic molecules tend to fill the adsorption sites of high adsorption energy at low loading first, but are often able to significantly increase their packing efficiency (and thus configurational entropy) when adsorbing on the low-energy adsorption sites. This can generally only be achieved if the adsorbate molecules are reoriented. The adsorption uptake at low relative pressures is a measure of the adsorption strength (i.e. the degree of adsorbate-adsorbent interactions). The adsorption strengths of the investigated adsorptives in CAU-10 is given by the sequence:

$$\text{water} << \text{ethanol} < \text{n-butanol} < \text{n-hexanol},$$

which is consistent with the simulated heats of adsorption at zero coverage and 298.15 K, i.e. -31.2 ± 0.1 kJ/mol (water), -47.7 ± 1.1 kJ/mol (ethanol), -61.5 ± 0.5 kJ/mol (n-butanol), or -75.8 ± 0.5 kJ/mol (n-hexanol), and confirms the internal hydrophobicity of CAU-10 due to the preferential adsorption of the more hydrophobic species. The saturation capacity, on the other hand, is a measure of the packing efficiency of the adsorbate molecules in CAU-10. The packing efficiency is given by the sequence:

$$\text{water} >> \text{ethanol} > \text{n-butanol} \approx \text{n-hexanol}.$$

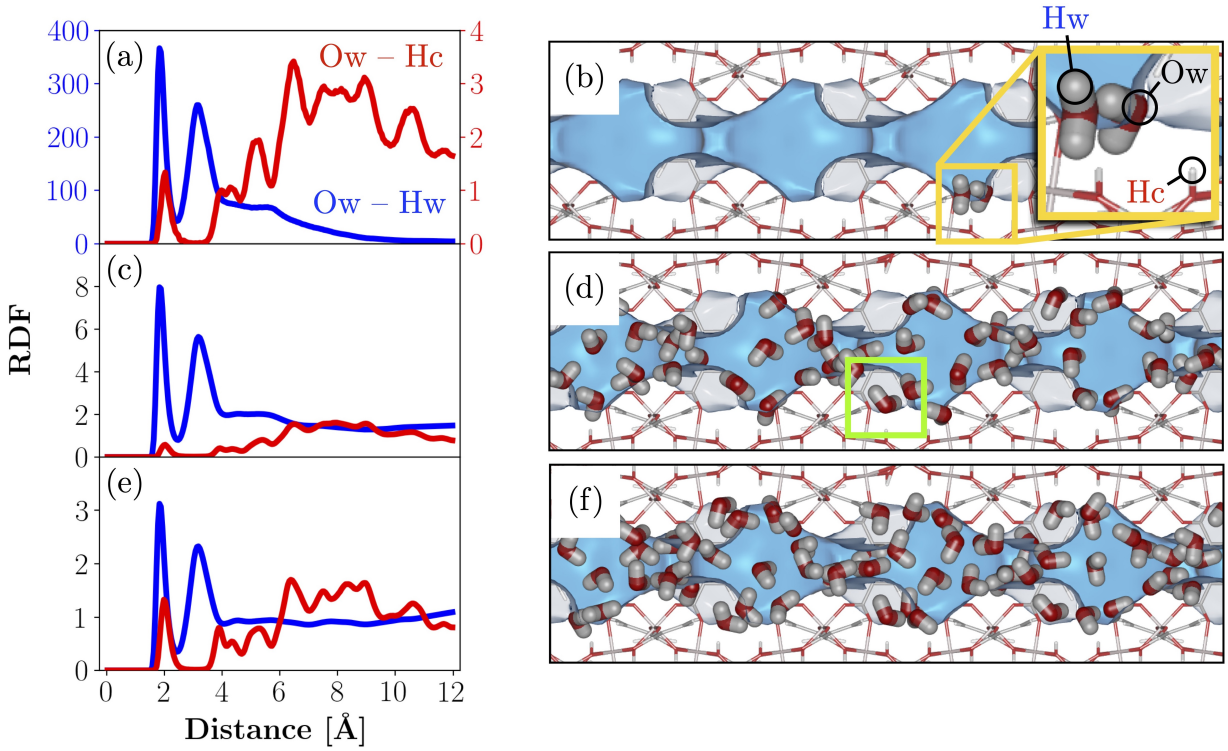

Fig. 19: RDFs of distances for oxygen in water (Ow) *and* hydrogen in the hydroxyl groups that bridge the aluminum atoms of CAU-10 (Hc, red) or hydrogen in water (Hw, blue) and corresponding snapshots of water in CAU-10 at 298.15 K and (a, b) low loading at $p/p_0 = 0.2$, (c, d) medium loading at $p/p_0 = 0.28$, or (e, f) saturation at $p/p_0 = 0.95$. Note the secondary ordinate for atom pair Ow – Hc in (a).

Insights into these sequences and the mechanisms of adsorption can be gained from snapshots and RDFs. Fig. 19 shows the RDFs of distances for selected atom pairs and the corresponding snapshots for water in CAU-10 at 298.15 K and low loading at $p/p_0 = 0.2$ (Fig. 19a, b), medium loading at $p/p_0 = 0.28$ (Fig. 19c, d), or saturation at $p/p_0 = 0.95$ (Fig. 19e, f). These loading regimes are analyzed to elucidate the simulated adsorption isotherm of water in Fig. 18a, namely the reason for the sudden water uptake at $p/p_0 = 0.25$ and the following low increase in water adsorption. Atom pairs for determining the RDFs were chosen to illustrate the most dominant interactions in the system. For the adsorption of water in CAU-10, dominant interactions are those between oxygen atoms in water (Ow) *and* hydrogen atoms in water (Hw) or hydrogen atoms in the hydroxyl groups that bridge the aluminum atoms in CAU-10 (Hc). RDFs for the remaining atom pairs can be found in Fig. L2 – L4 of Appendix L. The results in Fig. 19 are interpreted as follows:

1) **Low loading:** The first peaks in the RDFs in Fig. 19a occur at a distance that is characteristic for hydrogen bonding, i.e. 2.0 Å (Ow – Hc) or 1.8 Å (Ow – Hw). The height of the first peaks indicate the degree of hydrogen bonding, which is significantly larger for Ow – Hw (peak height ca. 380) than for Ow – Hc (peak height ca. 1.5). The extremely high peak for Ow – Hw is also a result of the very low loading, because there are fewer water molecules to average over when determining the RDF. In this light, it seems remarkable that water molecules tend to adsorb close to each other. We conclude two points from the RDFs. First, water adsorbs cooperatively. Adsorbed water molecules act as strong adsorption sites for further water adsorption – explaining the sudden uptake at $p/p_0 = 0.25$ in the isotherm in Fig. 18a. Second, CAU-10 has (more) hydrophilic adsorption sites. The snapshot in Fig. 19b indicates that the hydrophilic sites are located in the processes of the pore channels (framed in yellow). The framed area also shows the cooperative adsorption of water.

2) **Medium loading:** The first peaks in the RDFs in Fig. 19c occur at the same distances as observed at low loading, indicating hydrogen bonding for the atom pairs Ow – Hc and Ow – Hw. The degree of hydrogen bonding for Ow – Hw (peak height ca. 8) dominates relative to Ow – Hc (peak height ca. 1.0), but is significantly decreased compared to the situation observed at low loading. This does not necessarily indicate that hydrogen bonding between water molecules is less favored at medium loading, but that there are more water molecules to average over when determining the RDF due to an increased loading. The increased loading is also indicated by the snapshot in Fig. 19d. Some water molecules appear to be adsorbed „outside“ the boundaries of the pore channel (framed in lime green), but this is deceptive because iRASPA estimates the pore geometry based on Lennard-Jones interactions only, thus ignoring electrostatic interactions which are often used to describe hydrogen bonding.

3) **Saturation:** The first peaks in the RDFs in Fig. 19e occur at the same distances as observed at low loading or medium loading, again indicating hydrogen bonding for the

atom pairs Ow – Hc and Ow – Hw. The degree of hydrogen bonding for Ow – Hw (peak height ca. 3) dominates relative to Ow – Hc (peak height ca. 1.5), but is further decreased compared to the situation observed at medium loading – indicating an increased loading and thus more water molecules to average over when determining the RDF. The increased loading is also indicated by the snapshot in Fig. 19f. Many water molecules are now adsorbed „outside“ the boundaries of the pore channel. The slight increase in water uptake for $p/p_0 > 0.25$ is probably due to the filling of these adsorption sites with increasing pressure.

Fig. 20 shows the RDFs of distances for selected atom pairs and the corresponding snapshots for ethanol in CAU-10 at 298.15 K and medium loading at $p/p_0 = 6 \cdot 10^{-3}$ (Fig. 20a, b) or saturation at $p/p_0 = 0.88$ (Fig. 20c, d). These loading domains are analyzed to elucidate the simulated adsorption isotherm of ethanol in Fig. 18b. Atom pairs for determining the RDFs were chosen to illustrate the most dominant interactions in the system. For the adsorption of ethanol in CAU-10, dominant interactions are those between hydrogen atoms in the hydroxyl group of ethanol (Ha) *and* oxygen atoms in the hydroxyl group of ethanol (Oa) or carboxyl oxygen atoms in CAU-10 (Oc), as well as those between the terminal alkyl group in ethanol (CH_3) *and* the terminal alkyl group of other ethanol molecules or benzene carbon in CAU-10 (Cb). RDFs for the remaining atom pairs can be found in Fig. L5 and Fig. L6 of Appendix L. The results in Fig. 20 are interpreted as follows:

1) **Medium loading:** The first peaks in the RDFs in Fig. 20a occur at a distance that either suggests hydrogen bonding, i.e. 1.8 Å (Ha – Oa) or 2.2 Å (Ha – Oc), or that which is commonly observed for carbon-carbon interactions, i.e. 4.0 Å (CH_3 – CH_3) or 4.4 Å (CH_3 – Cb). The degree of hydrogen bonding for Ha – Oa (peak height ca. 3.5) is larger than for Ha – Oc (peak height ca. 1.4), and also larger than the degree of carbon-carbon interactions for CH_3 – Cb (peak height ca. 2.4) or CH_3 – CH_3 (peak height ca. 1.8), but is still in the same order of magnitude. Ethanol is both hydrophilic (hydroxyl group) and hydrophobic (alkyl group). The RDFs imply that ethanol interacts to a similar degree via its hydroxyl group with the hydroxyl groups of other ethanol molecules *or* the more hydrophilic adsorption sites of CAU-10, and via its alkyl group with the alkyl groups of other ethanol molecules *or* the more hydrophobic adsorption sites of CAU-10. The snapshot in Fig. 20b confirms the adsorption mechanism of ethanol in CAU-10. Ethanol molecules adsorb with their hydroxyl groups toward the hydrophilic adsorption sites in the processes of the pore channel (framed in blue), and with their alkyl groups toward the hydrophobic adsorption sites along the benzene ring (framed in brown). The snapshot also confirms the high degree of hydrogen bonding between ethanol molecules (see the four ethanol molecules on the right-hand side). We conclude that ethanol molecules adsorb cooperatively: the hydroxyl groups of adsorbed ethanol molecules act as strong adsorption sites for further ethanol adsorption – explaining the sudden increase in ethanol adsorption at $p/p_0 = 0.01$ observed in the isotherm in Fig. 18b.

2) **Saturation:** The first peaks in the RDFs in Fig. 20c occur at the same distances as those observed at medium loading, again indicating hydrogen bonding for the atom pairs Ha – Oa or Ha – Oc, as well as carbon-carbon interactions for atom pairs CH_3 – CH_3 or CH_3 – Cb. However, the degree of hydrogen bonding for Ha – Oa (peak height ca. 12) is now significantly larger than for Ha – Oc (peak height ca. 0.3), and also dominates compared to the atom pairs CH_3 – CH_3 (peak heights ca. 1.5) and CH_3 – Cb (peak heights ca. 2.1). We conclude that the adsorption at saturation is mainly governed by hydrogen bonding between ethanol molecules – confirming the cooperative adsorption of ethanol. The high degree of hydrogen bonding for Ha – Oa is also indicated by the snapshot in Fig. 20d. The dimensions of the pore channels in CAU-10 allow to host two ethanol molecules per cuboid (framed in red). Most of these ethanol pairs align their hydroxyl groups with each other (and often with the ethanol pair of adjacent cuboids), while only few ethanol molecules adsorb with their hydroxyl groups toward the hydrophilic adsorption sites of CAU-10.

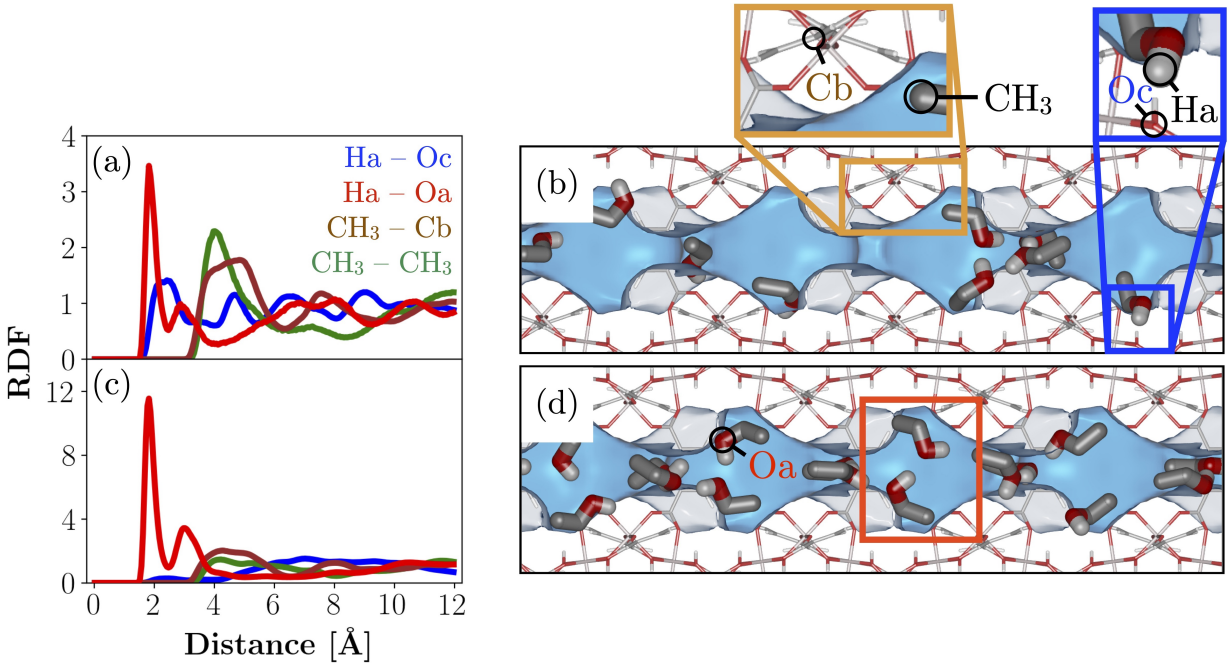

Fig. 20: RDFs of distances for hydrogen in ethanol (Ha) *and* carboxyl oxygen in CAU-10 (Oc, blue) or oxygen in ethanol (Oa, red), as well as for the terminal alkyl groups in ethanol (CH_3) *and* benzene carbon in CAU-10 (Cb, brown) or CH_3 (green), and corresponding snapshots of ethanol in CAU-10 at 298.15 K and (a, b) medium loading at $p/p_0 = 6 \cdot 10^{-3}$ or (c, d) saturation at $p/p_0 = 0.88$.

Fig. 21 shows the RDFs of distances for selected atom pairs and the corresponding snapshots of n-butanol in CAU-10 at 298.15 K and medium loading at $p/p_0 = 1.1 \cdot 10^{-3}$ (Fig. 21a, b) or saturation at $p/p_0 = 0.54$ (Fig. 21c, d). These loading domains are analyzed to elucidate the simulated adsorption isotherm of n-butanol in Fig. 18b, namely the reason for the occurrence of the inflection at 1.2 mmol/g as well as for the decreased saturation capacity compared to ethanol. Atom pairs for determining the RDFs were chosen to illustrate the most dominant interactions in the system. For the adsorption of n-butanol in CAU-10, dominant interactions

are those between hydrogen atoms in the hydroxyl group of n-butanol (Ha) *and* oxygen atoms in the hydroxyl group of n-butanol (Oa) or carboxyl oxygen atoms in CAU-10 (Oc), as well as those between the internal alkyl groups in n-butanol (CH_2) *and* the internal alkyl groups of other n-butanol molecules or benzene carbon in CAU-10 (Cb). RDFs for the remaining atom pairs can be found in Fig. L7 and Fig. L8 of Appendix L. The results in Fig. 21 are interpreted as follows:

1) **Medium loading:** The first peaks in the RDFs in Fig. 21a occur at a distance that either suggests hydrogen bonding, i.e. 1.9 Å (Ha – Oa) or 2.5 Å (Ha – Oc), or that which is commonly observed for carbon-carbon interactions, i.e. 4.4 Å (CH_2 – Cb), or 4.6 Å (CH_2 – CH_2). These distances are similar to those observed for ethanol at medium loading. But for n-butanol, the degree of adsorbate-adsorbate interactions is significantly lower than the degree of adsorbent-adsorbate interactions, as indicated by peak heights of 0.16 (Ha – Oa) or 0.17 (CH_2 – CH_2) versus 1.6 (Ha – Oc) or 2.0 (CH_2 – Cb). We conclude that the adsorption of n-butanol at medium loading is mainly governed by the interaction with CAU-10 – involving both the interaction of the hydroxyl group with the more hydrophilic adsorption sites of CAU-10 *and* the interaction of the alkyl group with the more hydrophobic adsorption sites of CAU-10. The snapshot in Fig. 21b confirms the adsorption mechanism. n-Butanol molecules adsorb with their hydroxyl groups toward the hydrophilic adsorption sites (framed in blue), and with their alkyl groups toward the hydrophobic adsorption sites (framed in brown). This adsorption mechanism is similar to that observed for ethanol at medium loading, but the longer alkyl group of n-butanol allows a stronger interaction with the more hydrophobic adsorption sites – explaining the increased adsorption heat of n-butanol relative to ethanol.

2) **Saturation:** The first peaks in the RDFs in Fig. 21c occur for the same distances as those observed at medium loading. The degree of hydrogen bonding for Ha – Oa (peak height ca. 0.07) is negligible compared to the degree of alkyl-alkyl interaction (peak height ca. 1.7 for CH_2 – CH_2), which is now similar to the degree of adsorbent-adsorbate interactions for CH_2 – Cb (peak height ca. 1.8) and Ha – Oc (peak height ca. 1.6). Fig. 21d confirms the increased degree of CH_2 – CH_2 interaction. n-Butanol molecules undergo reorientation to move from the configuration at medium loading to that at saturation. During this reorientation, most n-butanol molecules align their alkyl groups along the longitudinal axis of the cuboid – increasing both the packing efficiency of n-butanol as well as the degree of CH_2 – CH_2 interaction with the n-butanol molecules in adjacent cuboids –, while their hydroxyl groups remain aligned with the hydrophilic adsorption sites of CAU-10. We conclude two points regarding the simulated n-butanol isotherm in Fig. 18b. First, the inflection at 1.2 mmol/g is caused by the reorientation of n-butanol in CAU-10. One reason why the experimental n-butanol isotherm does not show a similarly pronounced inflection might be that the flexibility of the isophthalate linker facilitates reorientation. Recall that CAU-10 was

modeled rigid. Second, the different saturation capacities for ethanol (4.8 mmol/g or 16 molec/uc) and n-butanol (2.4 mmol/g or 8 molec/uc) are caused by the different packing efficiencies. The pore geometry of CAU-10 allows each cuboid of the pore volume to host two ethanol molecules but only one n-butanol molecule.

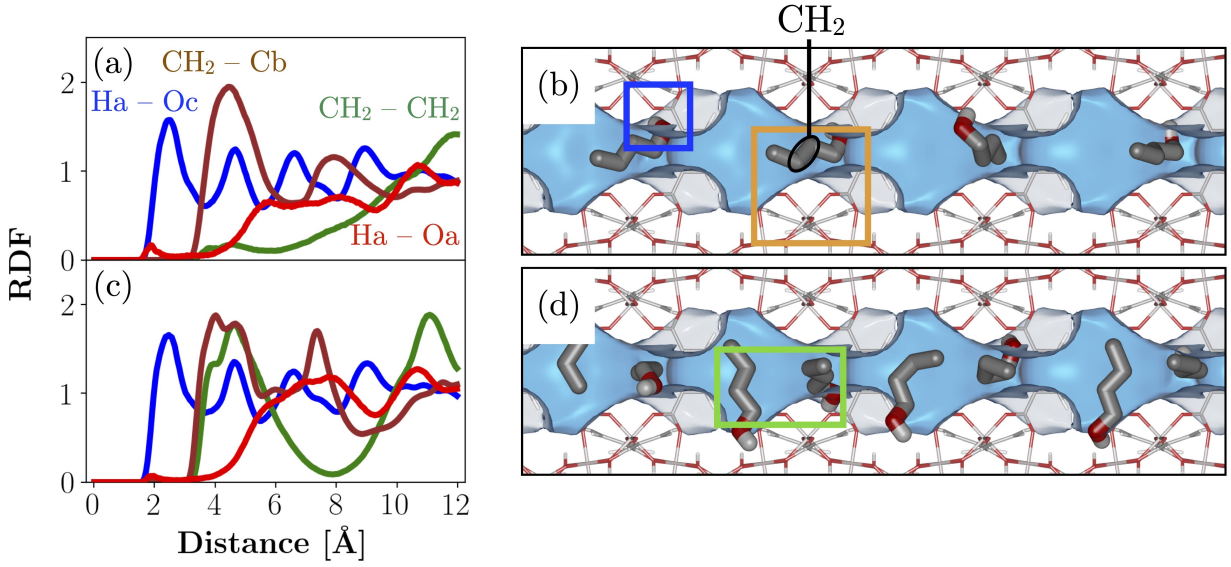

Fig. 21: RDFs of distances for hydrogen in n-butanol (Ha) *and* carboxyl oxygen in CAU-10 (Oc, blue) or oxygen in n-butanol (Oa, red), as well as for the internal alkyl groups in n-butanol (CH_2) *and* benzene carbon in CAU-10 (Cb, brown) or CH_2 (green), and corresponding snapshots of n-butanol in CAU-10 at 298.15 K and (a, b) medium loading at $p/p_0 = 1.1 \cdot 10^{-3}$ or (c, d) saturation at $p/p_0 = 0.54$.

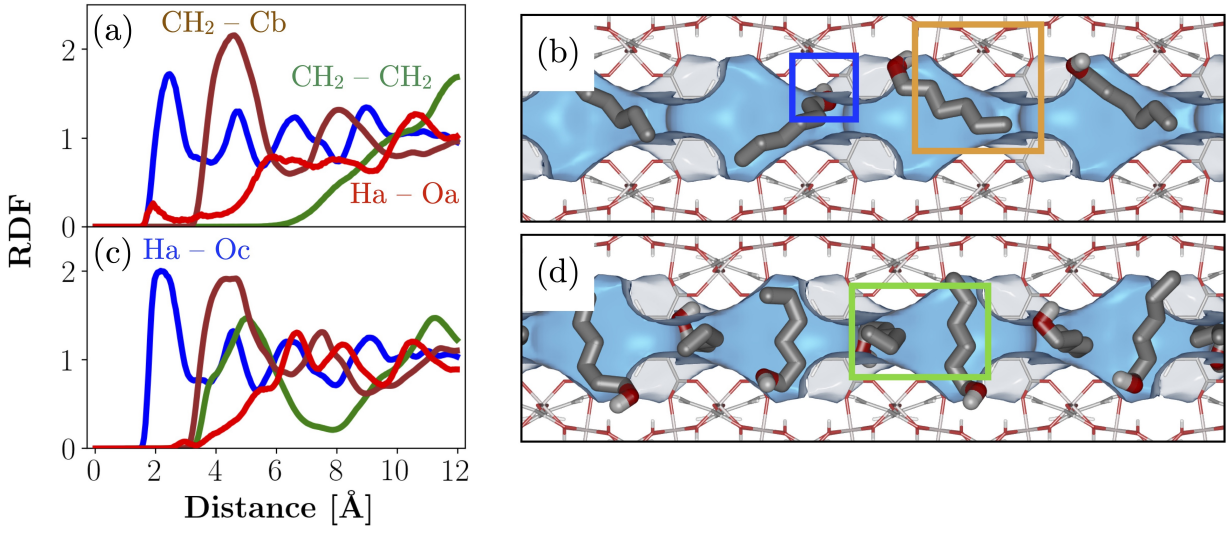

Fig. 22: RDFs of distances for hydrogen in n-hexanol (Ha) *and* carboxyl oxygen in CAU-10 (Oc, blue) or oxygen in n-hexanol (Oa, red), as well as for the internal alkyl groups in n-hexanol (CH_2) *and* benzene carbon in CAU-10 (Cb, brown) or CH_2 (green), and corresponding snapshots of n-hexanol in CAU-10 at 298.15 K and (a, b) medium loading at $p/p_0 = 9.2 \cdot 10^{-5}$ or (c, d) saturation at $p/p_0 = 0.92$.

Fig. 22 shows the RDFs of distances for selected atom pairs and the corresponding snapshots of n-hexanol in CAU-10 at 298.15 K and medium loading at $p/p_0 = 9.2 \cdot 10^{-5}$ (Fig. 22a,

b) or saturation at $p/p_0 = 0.92$ (Fig. 22c, d). These loading domains are analyzed to elucidate the simulated adsorption isotherm of n-hexanol in Fig. 18b, namely the reason for the occurrence of the inflection at 1.2 mmol/g and the similar saturation capacity compared to n-butanol. Atom pairs for determining the RDFs were chosen to illustrate the most dominant interactions in the system. For the adsorption of n-hexanol in CAU-10, dominant interactions are the same as those for n-butanol, i.e. Ha – Oa, Ha – Oc, CH_2 – Cb, and CH_2 – CH_2. RDFs for the remaining atom pairs can be found in Fig. L9 and Fig. L10 of Appendix L. The results in Fig. 22 are interpreted as follows:

1) **Medium loading:** The first peaks in the RDFs occur at 1.9 Å (Ha – Oa), 2.5 Å (Ha – Oc), 4.5 Å (CH_2 – Cb), or ca. 12 Å (CH_2 – CH_2). These distances indicate hydrogen bonding for Ha – Oa and Ha – Oc, carbon-carbon interaction for CH_2 – Cb, or negligible interaction for CH_2 – CH_2. The degree of adsorbate-adsorbate interaction is significantly lower than the degree of adsorbent-adsorbate interaction, as indicated by a peak height of 0.25 (Ha – Oa) or the late CH_2 – CH_2 peak versus peak heights of 1.7 (Ha – Oc) or 2.2 (CH_2 – Cb). The RDFs for the atom pairs Ha – Oa, Ha – Oc, and CH_2 – Cb are almost identical to those observed for n-butanol at medium loading, while the RDF for CH_2 – CH_2 differs mainly by the missing peak around 4.6 Å for n-hexanol. As with n-butanol, we conclude that the adsorption of n-hexanol at medium loading is mainly governed by the interaction with CAU-10 – involving both the interaction of the hydroxyl group with the more hydrophilic adsorption sites of CAU-10 *and* that of the alkyl group with the more hydrophobic adsorption sites of CAU-10. The snapshot in Fig. 22b confirms the conclusions drawn from the RDFs. n-Hexanol molecules adsorb with their hydroxyl groups toward the hydrophilic adsorption sites (framed in blue), and with their alkyl groups toward the hydrophobic adsorption sites (framed in brown). This adsorption mechanism is similar to that observed for ethanol and n-butanol at medium loading, but the longer alkyl group of n-hexanol allows a stronger interaction with the more hydrophobic adsorption sites – explaining the higher adsorption energy compared to n-butanol.

2) **Saturation:** The first peaks in the RDFs occur at 3.0 Å (Ha – Oa), 2.2 Å (Ha – Oc), 4.5 Å (CH_2 – Cb), or ca. 4.6 Å (CH_2 – CH_2). Compared to the situation observed at medium loading, the hydrogen bond lengths are shifted to higher distances for Ha – Oa and to lower distances for Ha – Oc, and the „missing“ CH_2 – CH_2 peak emerges. The degree of hydrogen bonding for Ha – Oa (peak height ca. 0.08) is negligible compared to the degree of alkyl-alkyl interaction (peak height ca. 1.5 for CH_2 – CH_2), which is now similar to the degree of adsorbent-adsorbate interactions for CH_2 – Cb (peak height ca. 1.9) and Ha – Oc (peak height ca. 2.0). The results are similar to the situation observed for n-butanol at saturation. The snapshot in Fig. 22d confirms that, much like n-butanol, n-hexanol molecules undergo reorientation to move from the configuration at medium loading to that at saturation. During this reorientation, most n-hexanol molecules align their alkyl groups along the longitudinal axis of the

cuboid – increasing both the packing efficiency of n-hexanol as well as the degree of CH_2 – CH_2 interaction with the n-hexanol molecules in adjacent cuboids –, while their hydroxyl groups remain aligned with the hydrophilic adsorption sites of CAU-10. Their chain lengths often force n-hexanol molecules to push their hydroxyl groups toward the more hydrophilic adsorption sites of CAU-10 – explaining the shift of the Ha – Oc peak toward lower distances in Fig. 22c. As with n-butanol, we conclude two points regarding the simulated n-hexanol isotherm in Fig. 18b. First, the inflection at 1.2 mmol/g is caused by the reorientation of n-hexanol in CAU-10. The increased adsorbent-adsorbate interaction of n-hexanol compared to n-butanol explains the more pronounced inflection observed for the n-hexanol isotherm in Fig. 18b. Second, the similar saturation capacities for n-butanol and n-hexanol (both 2.4 mmol/g or 8 molec/uc) are caused by the similar packing efficiency. The pore geometry of CAU-10 allows any cuboid of the pore volume to host one n-butanol molecule or one n-hexanol molecule.

We can draw the following conclusions by comparing the results from the RDFs and snapshots of water, ethanol, n-butanol, or n-hexanol:

1. The significance of hydrogen bonding between adsorbate molecules decreases with increasing length of the alkyl group.

2. The significance of adsorbent-adsorbate interaction increases with increasing length of the alkyl group – once again confirming the internal hydrophobicity of CAU-10.

3.4.2 Two-Component Adsorption

Fig. 23 compares the simulated reduced surface excess (Fig. 23a) and associated component loadings (Fig. 23b) observed for water/ethanol mixture adsorption in CAU-10 at 298.15 K with the corresponding IAST calculations based on the simulated or experimental unary isotherms. The U-shape of the simulated reduced surface excess in Fig. 23a confirms the preferred adsorption of water over ethanol over the whole range of liquid phase mole fractions of water (type I isotherm according to Schay and Nagy). The IAST calculations based on the simulated unary isotherms provide qualitatively wrong results and even predict a negative excess adsorption of water at high liquid phase mole fractions of water (S-shaped isotherm), while the IAST calculations based on the experimental unary isotherms give qualitatively good results, but differ quantitatively. The associated component loadings of water and ethanol shown in Fig. 23b help to explain the gap between the simulated and IAST predicted surface excesses. The simulated component loadings show a continuous increase in water adsorption with a simultaneous decrease in ethanol uptake with increasing liquid phase mole fraction of water. The IAST calculations based on the simulated unary isotherms give qualitatively similar results, but underestimate the water uptake and overestimate the ethanol uptake – explaining the S-shaped excess isotherm in Fig. 23a. The decreased water uptake in the IAST calculations presumably results from the erroneous assumption of IAST that water and ethanol form an ideal adsorbed layer. Instead, the increased water adsorption

in the simulation relative to the IAST prediction suggests that water adsorbs cooperatively with ethanol. Co-adsorption is also confirmed by comparing the simulated component loadings observed for the binary mixtures with those occurring in unary adsorption at fugacities corresponding to the partial fugacities arising for the binary liquid mixtures shown in Fig. 23b. For $x'_{\text{water}} < 0.13$, the comparison reveals that water adsorption is significantly higher for water/etanol adsorption than for unary water adsorption. Remarkably, IAST is able to predict the occurring co-adsorption to some extent as well. This phenomenon was first described by Claessens et al. [84], who studied the separation of ethanol and n-butanol in ZIF-8. What appears contradictory at first glance is in fact the result of IAST's assumption of an ideal adsorbed phase – implying some degree of water/ethanol interaction (compared to the negligible water/water interaction for $x'_{\text{water}} < 0.13$ in Fig. 23b). The IAST calculations based on the experimental unary isotherms overestimate the water uptake and predict similar ethanol loadings compared with the simulation results. The decreased water uptake observed in the simulation is probably due to the increased hydrophilicity of CAU-10 and an increased saturation capacity of water observed in the experiment.

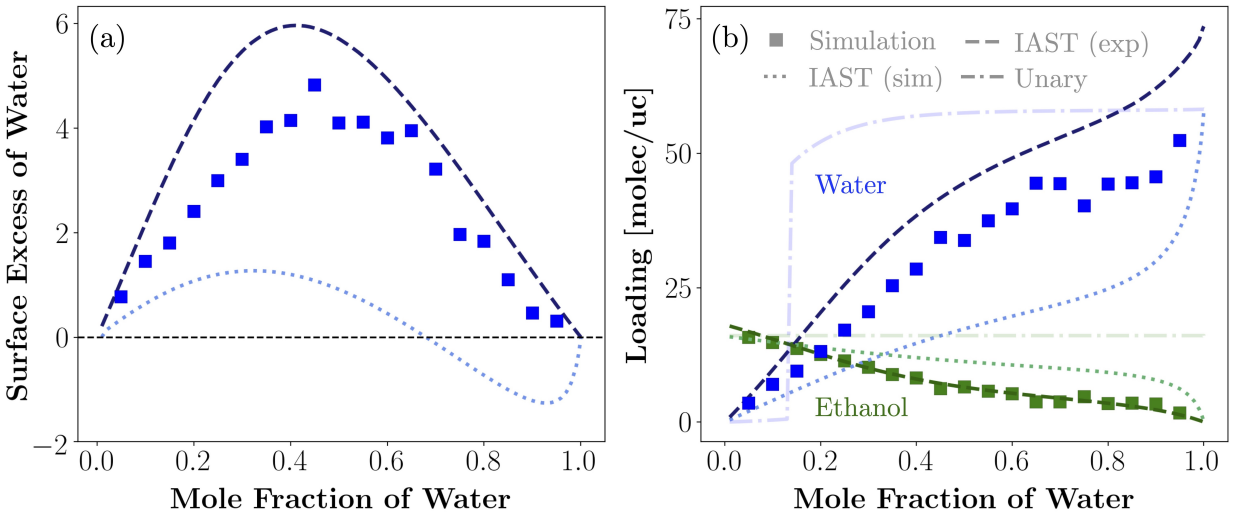

Fig. 23: (a) Reduced surface excess of water for water/ethanol adsorption in CAU-10 at 298.15 K: simulation (blue squares) and IAST calculations based on the simulated unary isotherms (cornflower blue, dotted line) or on the experimental unary isotherms (dark blue, dashed line). (b) Corresponding component loadings of water (shades of blue) or ethanol (shades of green): simulation (blue squares for water or green squares for ethanol), IAST calculations based on the simulated unary isotherms (dotted lines, cornflower blue for water or seagreen for ethanol) or on the experimental unary isotherms (dashed lines, dark blue for water or dark green for ethanol), and unary adsorption isotherms (dash-dotted lines, pale blue for water or pale green for ethanol).

Insights into the adsorption mechanism can be gained from the RDFs of distances for the atom pairs of the most dominant interactions and the corresponding snapshots in Fig. 24. RDFs and snapshots are shown for low water loading at $x'_{\text{water}} = 0.1$ (Fig. 24a, b) or medium water loading at $x'_{\text{water}} = 0.5$ (Fig. 24c, d). Adsorption is governed by hydrogen bonding between the guest species, and dominant interactions are those between hydrogen atoms in water (Hw) and oxygen atoms in water (Ow), hydroxyl hydrogen atoms in ethanol (Ha) and Ow, or Ha and hydroxyl oxygen atoms in ethanol (Oa). RDFs for the remaining atom pairs

can be found in Fig. L11 and Fig. L12 of Appendix L.

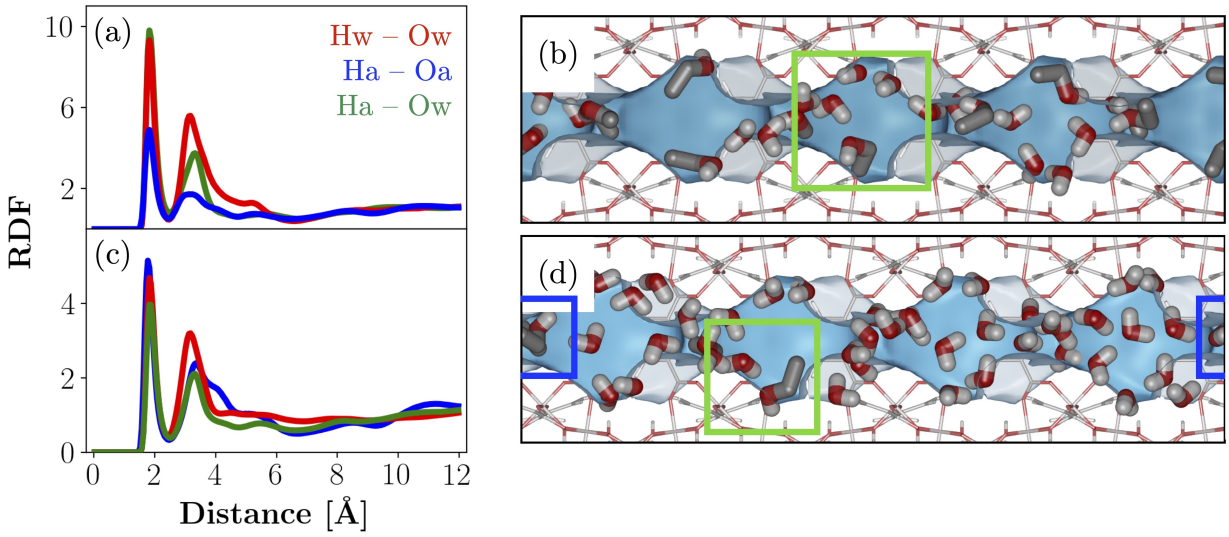

Fig. 24: RDFs of distances for hydrogen atoms in water (Hw) and oxygen atoms in water (Ow, red), as well as for hydroxyl hydrogen atoms in ethanol (Ha) *and* hydroxyl oxygen atoms in ethanol (Oa, blue) or Ow (green), and corresponding snapshots at (a, b) low water loading at $x'_{\text{water}} = 0.1$ or (c, d) medium water loading at $x'_{\text{water}} = 0.5$.

The results in Fig. 24 are interpreted as follows:

1) **Low water loading:** The first peaks in the RDFs in Fig. 24a occur at distances that are characteristic for hydrogen bonding, i.e. 1.8 Å (all atom pairs). The height of the first peaks indicate the degree of hydrogen bonding, which is largest for water/ethanol pairs (peak height ca. 9.8 for Ha – Ow), followed by water/water pairs (peak height ca. 9.3 for Hw – Ow), and lowest for ethanol/ethanol pairs (peak height ca. 4.9 for Ha – Oa). The high degree of hydrogen bonding between water/ethanol pairs supports the previous assumption of co-adsorption. The snapshot in Fig. 24b confirms the co-adsorption of water and ethanol as well. Small clusters of water are commonly found in the vicinity of the hydroxyl groups of ethanol molecules (framed in green) – explaining the high degree of hydrogen bonding between water/water or water/ethanol pairs.

2) **Medium water loading:** The first peaks in Fig. 24c occur at the same distances as those observed at low water loading, indicating hydrogen bonding between water/water, water/ethanol, and ethanol/ethanol pairs. The degree of hydrogen bonding is now strongest for the interaction between similar guest species, i.e. ethanol/ethanol (peak height ca. 5.2 for Ha – Oa) or water/water (peak height ca. 4.7 for Hw – Ow), and weakest for water/ethanol pairs (peak height ca. 4.0 for Ha – Ow) – implying that clusters of water molecules replace ethanol molecules with increasing liquid phase mole fraction of water. This is confirmed by the snapshot in Fig. 24d, which shows that most of the pore volume is occupied by water. While water molecules adsorb in a configuration that is similar to that observed at medium loading in unary water adsorption, ethanol molecules adsorb in pairs (framed in blue), or with their hydroxyl groups toward the more hydrophilic adsorption sites of CAU-10 (framed in green).

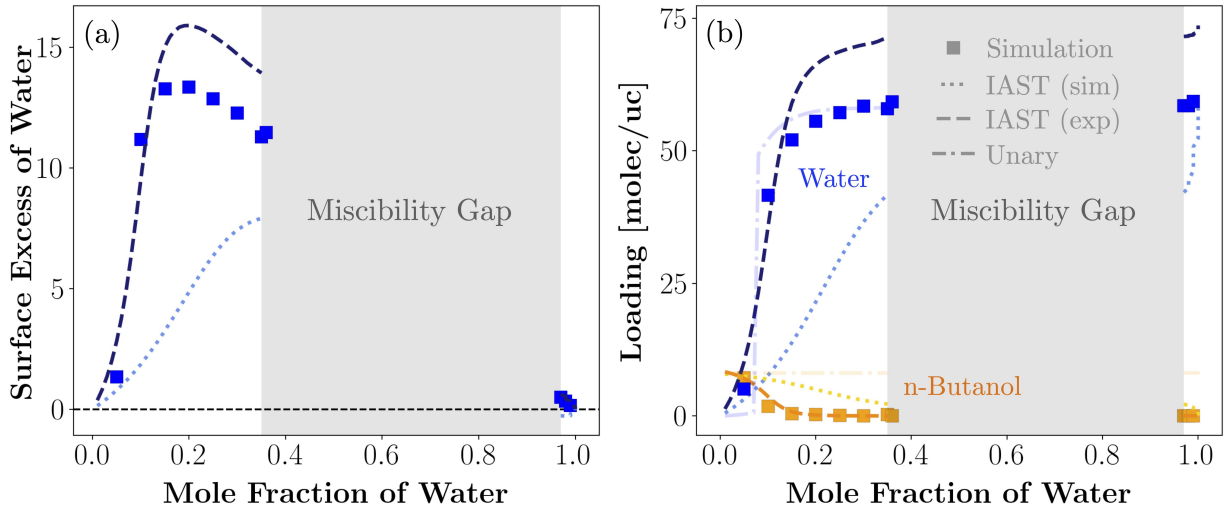

Fig. 25: (a) Reduced surface excess of water for water/n-butanol adsorption in CAU-10 at 298.15 K: simulation (blue squares) and IAST calculations based on the simulated unary isotherms (cornflower blue, dotted line) or on the experimental unary isotherms (dark blue, dashed line). (b) Corresponding component loadings of water (shades of blue) or n-butanol (shades of orange): simulation (blue squares for water or orange squares for n-butanol), IAST calculations based on the simulated unary isotherms (dotted lines, cornflower blue for water or light orange for n-butanol) or on the experimental unary isotherms (dashed lines, dark blue for water or dark orange for n-butanol), and unary adsorption isotherms (dash-dotted lines, pale blue for water or pale orange for ethanol).

Fig. 25 compares the simulated reduced surface excess (Fig. 25a) and associated component loadings (Fig. 25b) observed for water/n-butanol mixture adsorption in CAU-10 at 298.15 K with the corresponding IAST calculations based on the simulated or experimental unary isotherms. The simulated reduced surface excess of water in Fig. 25a shows a steep increase for $x'_{\text{water}} < 0.1$ up to the maximum at $x'_{\text{water}} \approx 0.18$ and a subsequent linear decrease with increasing liquid phase mole fraction of water. Note the miscibility gap for $0.3555 < x'_{\text{water}} < 0.9708$. The U-shape (type II isotherm according to Schay and Nagy) indicates a strong preference for water adsorption over the whole range of liquid phase mole fractions of water. The IAST calculations based on the simulated unary isotherms provide qualitatively satisfactory results, although predicting a negative excess adsorption of water at very high liquid phase mole fractions of water. The IAST calculations based on the experimental unary isotherms agree well with the simulation results. We will elucidate the reason why the IAST calculations based on the experimental isotherms agree much better with the simulation results than the IAST calculations based on the simulated isotherms below. We achieve this by comparing the simulated component loadings of water and n-butanol with those predicted by IAST in Fig. 25b. The simulated component loadings show a steep increase in water adsorption at $x'_{\text{water}} < 0.1$ with a simultaneous, rapid decrease in n-butanol uptake up to the complete exclusion of n-butanol from the adsorbed layer for $x'_{\text{water}} > 0.2$. The IAST calculations based on the simulated isotherms, on the other hand, predict a gradual increase in water adsorption and n-butanol desorption up to the beginning of the miscibility gap and then a sudden increase in water adsorption after the miscibility gap is left, while the IAST calculations based on the experimental isotherms agree well with

the simulation results, differing mainly in the increased water uptake for $x'_{\text{water}} > 0.12$. As with water/ethanol adsorption, the deviation of the IAST calculations based on the simulated isotherms from the simulation results is a consequence of the erroneous assumption of IAST that water and n-butanol form an ideal adsorbed layer. Instead, the increased water adsorption in the simulation suggests co-adsorption of water and n-butanol. Co-adsorption is also confirmed by comparing the simulated component loadings observed for the binary mixtures with those occurring in unary adsorption at fugacities corresponding to the partial fugacities arising for the binary liquid mixtures in Fig. 25b. For $x'_{\text{water}} < 0.08$, the comparison reveals that water adsorption is significantly higher for water/n-butanol adsorption than for unary water adsorption. Again, IAST is able to predict the occurring co-adsorption to some extent. Although the IAST calculations based on the experimental isotherms suffer from the same erroneous assumption of an ideal adsorbed layer, they agree well with the simulations. There are two reasons why. First, the IAST calculations based on the experimental isotherms predict the complete exclusion of n-butanol from the adsorbed layer for $x'_{\text{water}} > 0.2$ – tantamount to an ideal adsorbed layer. Second, the increased water adsorption in the experiment might somewhat compensate for the inability of IAST to account for the non-ideality of the adsorbed phase for $x'_{\text{water}} < 0.2$.

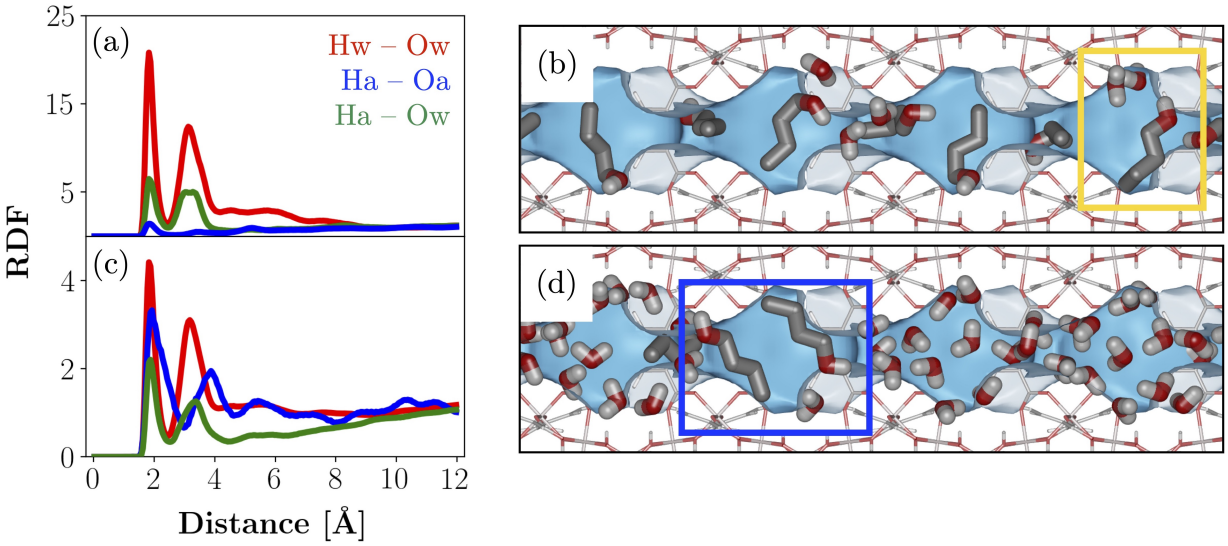

Fig. 26: RDFs of distances for hydrogen atoms in water (Hw) and oxygen atoms in water (Ow, red) as well as for hydroxyl hydrogen atoms in n-butanol (Ha) *and* hydroxyl oxygen atoms in n-butanol (Oa, blue) or Ow (green), and corresponding snapshots at (a, b) low water loading at $x'_{\text{water}} = 0.05$ or (c, d) medium water loading at $x'_{\text{water}} = 0.1$.

Insights into the adsorption mechanism can be gained from the RDFs of distances for the atom pairs of the most dominant interactions and the corresponding snapshots in Fig. 26. RDFs and snapshots are shown for low water loading at $x'_{\text{water}} = 0.05$ (Fig. 26a, b) or medium water loading at $x'_{\text{water}} = 0.1$ (Fig. 26c, d). Dominant interactions are those between hydrogen atoms in water (Hw) and oxygen atoms in water (Ow), hydrogen atoms in n-butanol (Ha) and Ow, or Ha and oxygen atoms in n-butanol (Oa). RDFs for the remaining atom pairs can be found in Fig. L13 and Fig. L14 of Appendix L. The results in Fig. 26 are interpreted

as follows:

1) **Low water loading:** The first peaks in the RDFs in Fig. 26a occur at distances that are characteristic for hydrogen bonding, i.e. 1.8 Å (all atom pairs). The height of the first peaks indicate the degree of hydrogen bonding between the atom pairs, which is significantly higher for water/water pairs (peak height ca. 20.8 for Hw – Ow) than for water/n-butanol pairs (peak height ca. 6.5 for Ha – Ow) or n-butanol/n-butanol pairs (peak height ca. 1.4 for Ha – Oa). The high degree of hydrogen bonding between water/water or water/n-butanol pairs indicates the preference of water to adsorb in pairs or to co-adsorb with n-butanol. The snapshot in Fig. 26b confirms the pairwise adsorption of water molecules and the co-adsorption of water with n-butanol (framed in yellow). The snapshot further indicates that n-butanol often adsorbs in a configuration that is similar to that observed in unary adsorption at saturation – explaining the low degree of hydrogen bonding between n-butanol/n-butanol pairs.

2) **Medium water loading:** The first peaks in Fig. 26c occur at the same distances as those observed at low water loading, indicating hydrogen bonding between water/water, water/n-butanol, and n-butanol/n-butanol pairs. The degree of hydrogen bonding between water/water pairs (peak height ca. 4.4 for Hw – Ow) is still dominant compared to n-butanol/n-butanol pairs (peak height ca. 3.8 for Ha – Oa) or water/n-butanol pairs (peak height ca. 2.2 for Ha – Ow), but significantly decreased compared to the situation observed at low water loading. Moreover, the degree of interaction between n-butanol/n-butanol pairs is now significantly higher than for water/n-butanol pairs. The snapshot in Fig. 26d shows the reason. The snapshot suggests that water molecules adsorbed „outside“ the pore channel constitute strong adsorption sites for the hydroxyl group of n-butanol molecules – allowing n-butanol to adsorb pairwise (framed in blue). Water contamination of the n-butanol sample in the experiment (e.g. because of the hygroscopy of n-butanol) might induce the pairwise adsorption of n-butanol in CAU-10 – a configuration that was not observed in the simulation of the unary adsorption – and explain the less pronounced adsorption step at ca. 1.2 mmol/g in Fig. 18b, because the pairwise adsorption requires only minor reorientation of n-butanol relative to the reorientation that is required to achieve the saturation configuration in Fig. 21d.

Fig. 27 compares the simulated reduced surface excess (Fig. 27a) and associated component loadings (Fig. 27b) for water/n-hexanol mixture adsorption in CAU-10 at 298.15 K with the corresponding IAST calculations based on the simulated or experimental unary isotherms. The simulated reduced surface excess of water in Fig. 27a shows a steep increase for $x'_{\text{water}} < 0.1$ up to the maximum at $x'_{\text{water}} \approx 0.15$ and a subsequent linear decrease with increasing liquid phase mole fraction of water, interrupted by the miscibility gap for $0.1971 < x'_{\text{water}} < 0.9978$. The U-shape (type II isotherm according to Schay and Nagy) indicates a strong preference for water adsorption over the whole range of liquid phase mole fractions of water. The IAST calculations based on the simulated unary isotherms provide

qualitatively satisfactory results, although predicting a negative excess adsorption of water at very high liquid phase mole fractions of water. The IAST calculations based on the experimental unary isotherms agree well with the simulation results.

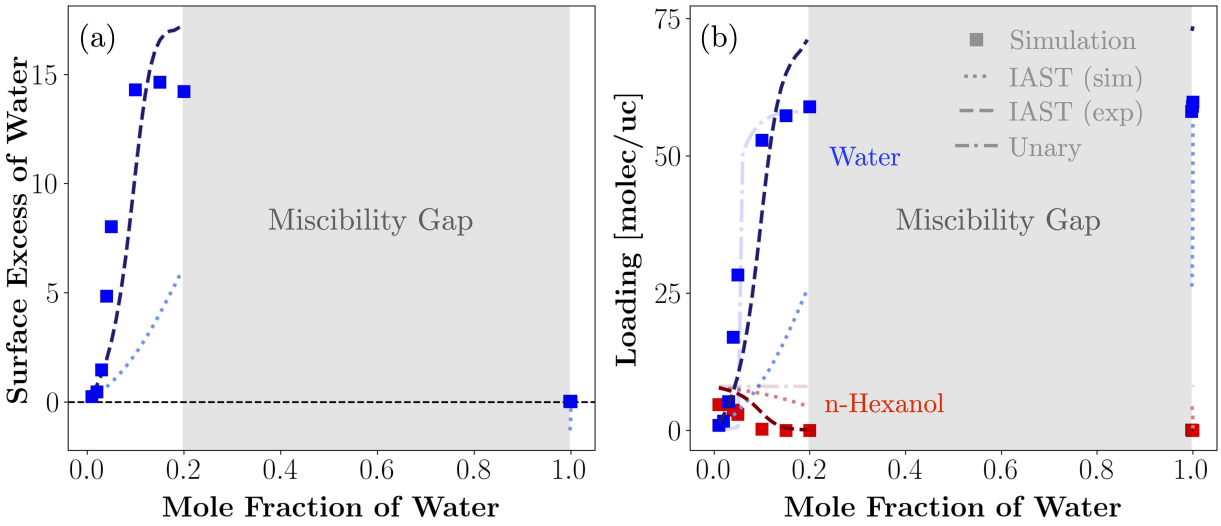

Fig. 27: (a) Reduced surface excess of water for water/n-hexanol adsorption in CAU-10 at 298.15 K: simulation (blue squares) and IAST calculations based on the simulated unary isotherms (cornflower blue, dotted line) or on the experimental unary isotherms (dark blue, dashed line). (b) Corresponding component loadings of water (shades of blue) and n-hexanol (shades of red): simulation (blue squares for water or red squares for n-hexanol), IAST calculations based on the simulated unary isotherms (dotted lines, cornflower blue for water or light red for n-hexanol) or on the experimental unary isotherms (dashed lines, dark blue for water or dark red for n-hexanol), and unary adsorption isotherms (dash-dotted lines, pale blue for water or pale red for n-hexanol).

The component loadings in Fig. 27b elucidate the reason why the IAST calculations based on the experimental isotherms agree much better with the simulation results than the IAST calculations based on the simulated isotherms. The simulated component loadings show a steep increase in water adsorption at $x'_{\text{water}} \approx 0.05$ with a simultaneous, rapid decrease in n-hexanol uptake up to the complete exclusion of n-hexanol from the adsorbed layer for $x'_{\text{water}} > 0.1$. Note that the desorption of n-hexanol begins at a loading that is significantly lower than the saturation capacity of unary n-hexanol in CAU-10. This is surprising because adsorption occurs under thermodynamic conditions similar to those present for the adsorption of unary n-hexanol at saturation. We will illuminate this phenomenon when discussing the snapshots in Fig. 28. The IAST calculations based on the simulated isotherms show a gradual increase in water adsorption and n-hexanol desorption up to the beginning of the miscibility gap and then a sudden increase in water adsorption after the miscibility gap is left, while the IAST calculations based on the experimental isotherms agree well with the simulation results, differing mainly in the increased water uptake for $x'_{\text{water}} > 0.1$. Similarly to water/ethanol and water/n-butanol adsorption, IAST erroneously assumes water and n-hexanol to form an ideal adsorbed layer. However, the increased water adsorption in the simulation suggests co-adsorption of water and n-hexanol. Co-adsorption of water and n-hexanol is also confirmed by comparing the binary with the unary component loadings that occur at fugacities corresponding to the partial fugacities arising for the binary liquid

mixtures in Fig. 27b. The comparison shows considerable water uptake for $x'_{\text{water}} < 0.05$, although water uptake is supposed to be negligible according to the unary component loadings. IAST is able to describe the occurring co-adsorption to some extent, but is unable to predict the reduced loading of n-hexanol at very low water loading. Although the IAST calculations based on the experimental isotherms suffer from the same erroneous assumption of an ideal adsorbed layer, they agree well with the simulations. There are two reasons why. First, the IAST calculations based on the experimental isotherms predict the complete exclusion of n-hexanol from the adsorbed layer for $x'_{\text{water}} > 0.1$ – tantamount to an ideal adsorbed layer. Second, the increased water adsorption in the experiment might somewhat compensate for the inability of IAST to account for the non-ideality of the adsorbed phase for $x'_{\text{water}} < 0.1$.

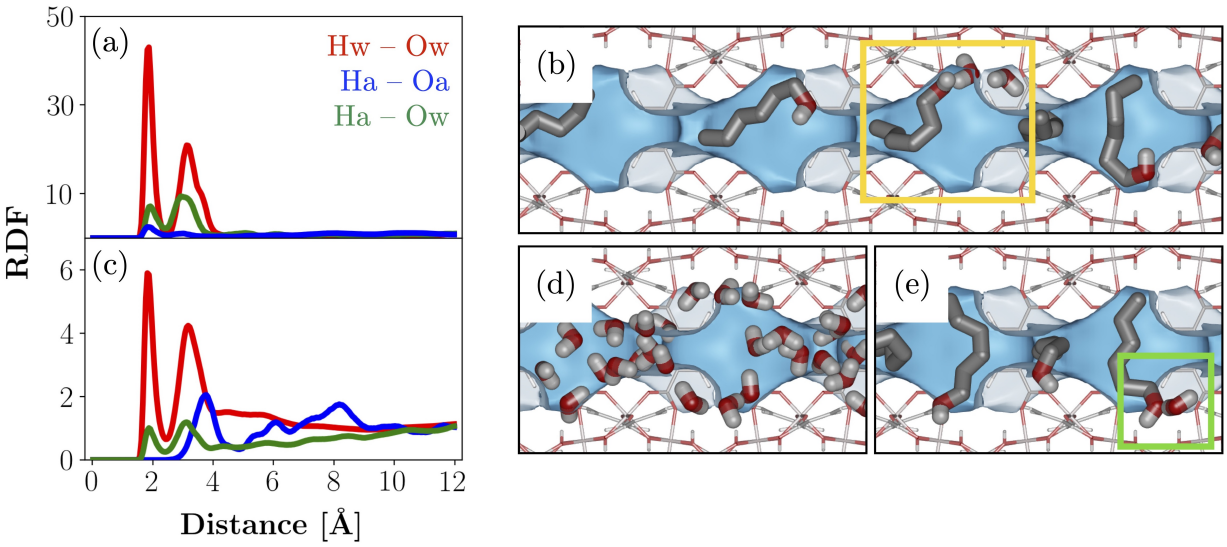

Fig. 28: RDFs of distances for hydrogen atoms in water (Hw) and oxygen atoms in water (Ow, red), as well as hydroxyl hydrogen atoms in n-hexanol (Ha) *and* hydroxyl oxygen atoms in n-hexanol (Oa, blue) or Ow (green), and corresponding snapshots at (a, b) low water loading at $x'_{\text{water}} = 0.01$ or (c, d) medium water loading at $x'_{\text{water}} = 0.05$.

Insights into the mechanisms of adsorption can be gained from the RDFs of distances for the atom pairs of the most dominant interactions and the corresponding snapshots in Fig. 28. RDFs and snapshots are shown for low water loading at $x'_{\text{water}} = 0.01$ (Fig. 28a, b) or medium water loading at $x'_{\text{water}} = 0.05$ (Fig. 28c–e). Adsorption is governed by hydrogen bonding between the guest species, and dominant interactions are those between hydrogen atoms in water (Hw) and oxygen atoms in water (Ow), hydrogen atoms in n-hexanol (Ha) and Ow, or Ha and oxygen atoms in n-hexanol (Oa). RDFs for the remaining atom pairs can be found in Fig. L15 and Fig. L16 of Appendix L. The results in Fig. 28 are interpreted as follows:

1) **Low water loading:** The first peaks in the RDFs in Fig. 28a occur at a distance that is characteristic for hydrogen bonding, i.e. 1.8 Å (all atom pairs). The height of the first peaks indicate the degree of hydrogen bonding between the atom pairs, which is significantly higher for water/water pairs (peak height ca. 43.0 for Hw – Ow) than for water/n-hexanol pairs (peak height ca. 7.2 for Ha – Ow) or n-hexanol/n-hexanol pairs (peak height ca. 2.5 for Ha – Oa). The high degree of hydrogen bonding between

water/water or water/n-hexanol pairs indicates the preference of water to adsorb in pairs or to co-adsorb with n-hexanol. This is also confirmed by the snapshot in Fig. 28b (framed in yellow). The snapshot indicates that n-hexanol often adsorbs flat on the walls of the pore channel – a configuration that is similar to that observed in unary n-hexanol adsorption at low loading. This configuration is probably induced by adsorbed water molecules, which represent strong adsorption sites for the hydroxyl group of n-hexanol. The alkyl chain of n-hexanol is consequently pushed to adsorb flat on the pore wall, which may cause adjacent n-hexanol molecules to follow suit – explaining the significantly reduced n-hexanol uptake for high liquid phase mole fractions of n-hexanol in Fig. 27 relative to the saturation capacity of n-hexanol.

2) **Medium water loading:** The first peaks in the RDFs indicate hydrogen bonding between water/water or water/n-hexanol pairs (peaks occurring at 1.8 Å), but not between n-hexanol/n-hexanol pairs (peaks occurring at 3.8 Å). The degree of hydrogen bonding between water/water pairs (peak height ca. 5.9 for Hw – Ow) dominates relative to water/n-hexanol pairs (peak height ca. 1.0 for Ha – Ow), which is similar to the situation observed at low water loading. The absence of hydrogen bonding between n-hexanol/n-hexanol pairs suggests that the configuration of n-hexanol is similar to that observed in unary adsorption at saturation (see Fig. 22c, d). The snapshots in Fig. 28d and e indicate that the pores are either almost exclusively filled with water (Fig. 28d) or n-hexanol (Fig. 28e). The strong segregation of n-hexanol and water is likely due to the hydrophobic alkyl group of n-hexanol, making it unfavorable for water clusters to adsorb along with n-hexanol molecules. The snapshots further suggest that water and n-hexanol adsorb in configurations that are similar to those observed in the unary adsorption simulations at saturation – indicating that the segregation pushes n-hexanol to increase its packing efficiency – different to the situation observed at low water loading. Even in its saturation configuration, n-hexanol molecules are able to interact with water molecules (framed in green).

Fig. 29 compares the simulated reduced surface excess (Fig. 29a) and associated component loadings (Fig. 29b) for ethanol/n-butanol mixture adsorption in CAU-10 at 298.15 K with the corresponding IAST calculations based on the simulated or experimental unary isotherms. The simulated reduced surface excess in Fig. 29a confirms the preferential adsorption of ethanol over n-butanol over the whole range of liquid phase mole fractions of ethanol, because of the U-shaped excess isotherm (in between type I and type II according to Schay and Nagy). The IAST calculations based on the simulated unary isotherms agree well with the simulation results, differing only in a decreased ethanol excess for $0.1 < x'_{\text{ethanol}} < 0.5$. The IAST calculations based on the experimental unary isotherms provide qualitatively good results, but overestimate the ethanol excess over the whole range of liquid phase mole fractions of ethanol. The associated component loadings of ethanol and n-butanol shown in Fig. 29b help to explain the deviation between simulation results and IAST calculations. The simulated component loadings show a hyperbolic increase in ethanol adsorption up to

saturation with a simultaneous decrease in n-butanol uptake up to the complete exclusion of n-butanol from the adsorbed layer at high liquid phase mole fractions of ethanol. The IAST calculations based on the simulated isotherms agree well with the simulations, differing only in a decreased ethanol uptake for $0.1 < x'_{\text{ethanol}} < 0.5$. This deviation is probably caused by cooperative adsorption of ethanol and n-butanol. The IAST calculations based on the experimental isotherms provide qualitatively good results, but overestimate the ethanol uptake and underestimate the n-butanol uptake compared to the simulation results. This deviation is probably caused by the increased saturation capacity of ethanol or the weaker adsorption strength of n-butanol observed in the experiment (see Fig. 18b or Eq. (G8)).

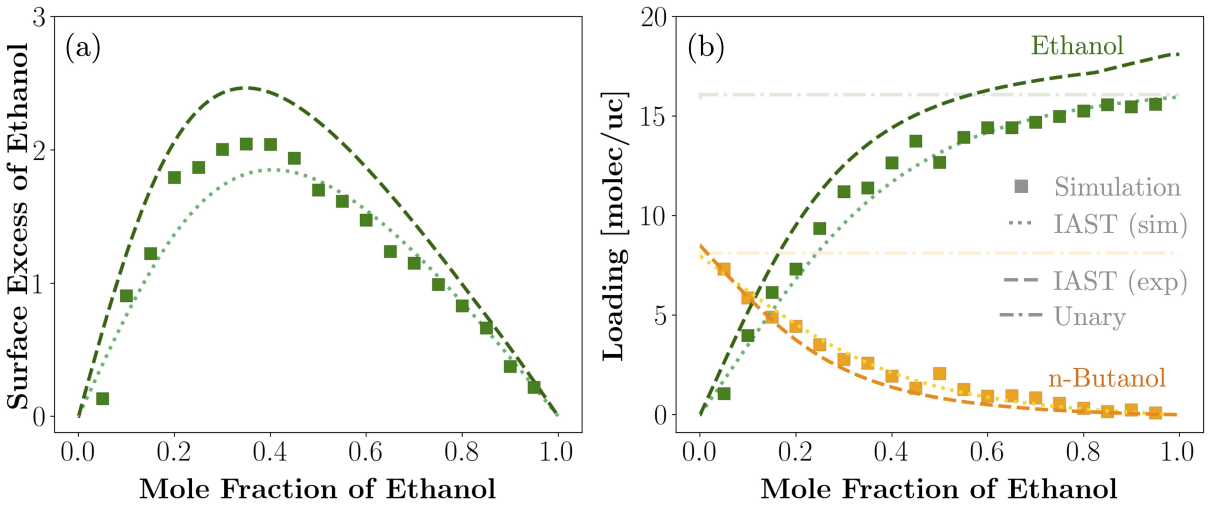

Fig. 29: (a) Reduced surface excess of ethanol for ethanol/n-butanol adsorption in CAU-10 at 298.15 K: simulation (green squares) and IAST calculations based on the simulated unary isotherms (seaweed green, dotted line) or on the experimental unary isotherms (dark green, dashed line). (b) Corresponding component loadings of ethanol (shades of green) and n-butanol (shades of orange): simulation (green squares for ethanol or orange squares for n-butanol), IAST calculations based on the simulated unary isotherms (dotted lines, seaweed green for ethanol or light orange for n-butanol) or on the experimental unary isotherms (dashed lines, dark green for ethanol or dark orange for n-butanol), and unary adsorption isotherms (dash-dotted lines, pale green for ethanol or pale orange for n-butanol).

Insights into the adsorption mechanism can be gained from the RDFs of distances for the atom pairs of the most dominant interactions and the corresponding snapshots in Fig. 30. RDFs and snapshots are shown for low ethanol loading at $x'_{\text{ethanol}} = 0.05$ (Fig. 30a, b) or medium ethanol loading at $x'_{\text{ethanol}} = 0.5$ (Fig. 30c, d). Dominant interactions are those between hydrogen atoms in ethanol (Ha(et)) and oxygen atoms in ethanol (Oa(et)), hydrogen atoms in n-butanol (Ha(but)) and Oa(et), or Ha(but) and oxygen atoms in n-butanol (Oa(but)). RDFs for the remaining atom pairs can be found in Fig. L17 and Fig. L18 of Appendix L. The results in Fig. 30 are interpreted as follows:

1) **Low ethanol loading:** The first peaks in the RDFs in Fig. 30a occur at distances that are characteristic for hydrogen bonding, i.e. 1.8 Å (all atom pairs). The height of the first peaks indicate the degree of hydrogen bonding between the atom pairs, which is significantly higher for ethanol/ethanol pairs (peak height ca. 36.1 for Ha(et) –

Oa(et)) than for ethanol/n-butanol pairs (peak height ca. 1.8 for Ha(but) – Oa(et)), and is negligible for n-butanol/n-butanol pairs (peak height ca. 0.1 for Ha(but) – Oa(but)). The RDFs suggest that the majority of ethanol and n-butanol molecules adsorb in a segregative manner, as shown in Fig. 30b. Ethanol and n-butanol are commonly found to occupy different cuboids of the pore channel. While the relatively few ethanol molecules can sometimes adsorb in pairs as suggested by the RDF in Fig. 30a – explaining the high degree of hydrogen bonding between ethanol/ethanol pairs –, n-butanol molecules rarely co-adsorb with ethanol and never with other n-butanol molecules.

2) **Medium ethanol loading:** The first peaks in Fig. 30c occur at the same distances as those observed at low ethanol loading, indicating hydrogen bonding between ethanol/ethanol, ethanol/n-butanol, or n-butanol/n-butanol pairs. The degree of hydrogen bonding for ethanol/ethanol pairs (peak height ca. 13.0 for Ha(et) – Oa(et)) dominates relative to ethanol/n-butanol pairs (peak height ca. 10.2 for Ha(but) – Oa(et)) or n-butanol/n-butanol pairs (peak height ca. 0.2 for Ha(but) – Oa(but)), but is significantly decreased compared to the situation observed at low ethanol loading. The importance of ethanol/n-butanol interactions is significantly increased. The snapshot in Fig. 30d indicates that ethanol molecules commonly adsorb in pairs (framed in red), or cooperatively with n-butanol molecules (framed in green) – explaining the increased degree of hydrogen bonding for ethanol/n-butanol pairs. It is reasonable to assume that, given the longer alkyl chain of n-butanol, ethanol/n-butanol pairs interact more strongly than ethanol/ethanol pairs – confirming the earlier assumption of co-adsorption. Still, entropic effects favor the adsorption of ethanol.

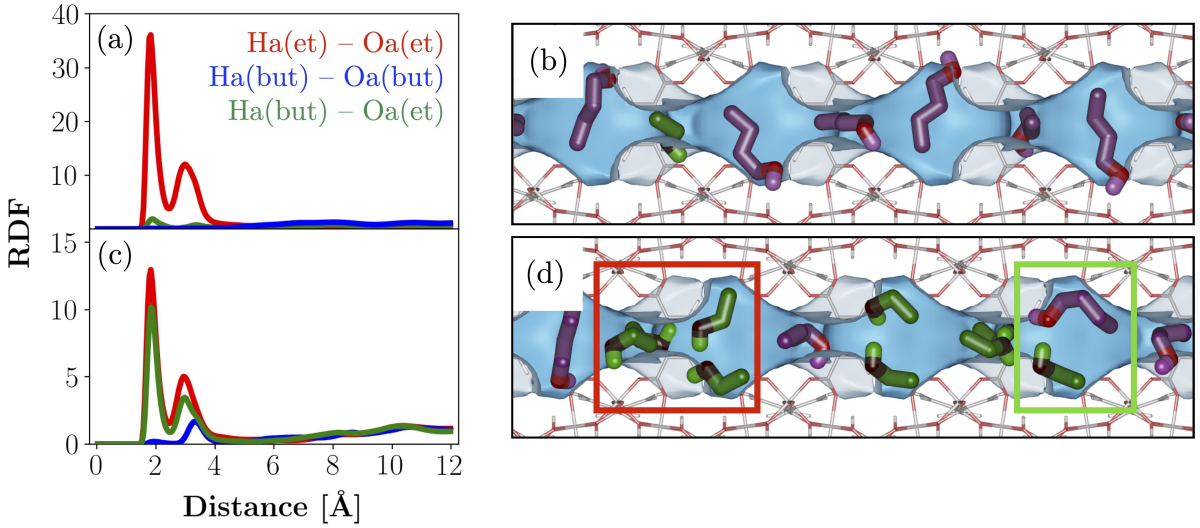

Fig. 30: RDFs of distances for hydrogen atoms in ethanol (Ha(et)) and oxygen atoms in ethanol (Oa(et), red), hydrogen atoms in n-butanol (Ha(but)) and Oa(et) (green), or Ha(but) and oxygen atoms in n-butanol (Oa(but), blue), as well as corresponding snapshots at (a, b) low ethanol loading at $x'_{\text{water}} = 0.05$ or (c, d) medium ethanol loading at $x'_{\text{water}} = 0.5$.

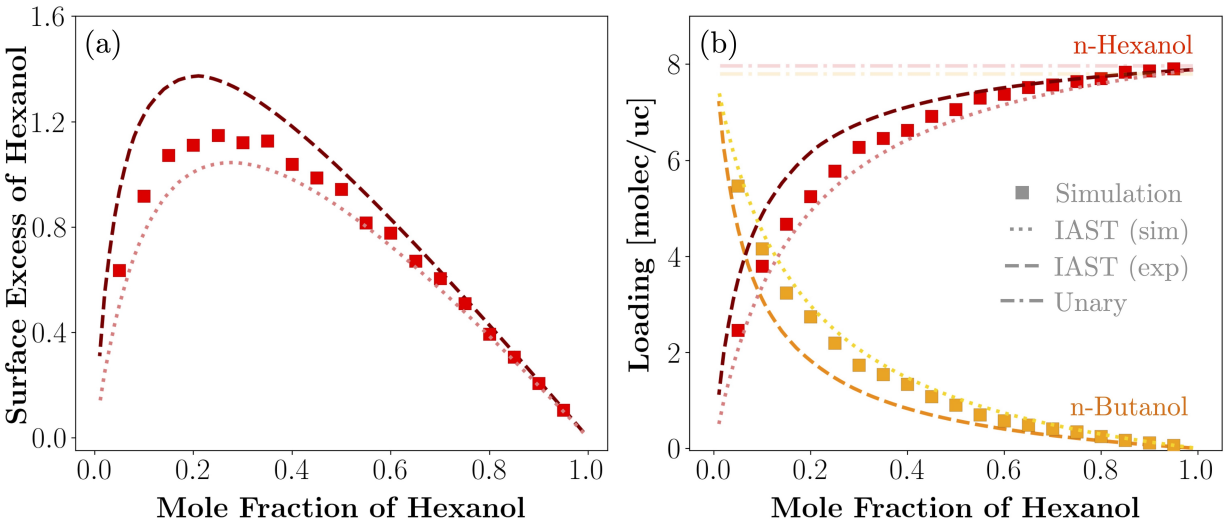

Fig. 31: (a) Reduced surface excess of n-hexanol for n-butanol/n-hexanol adsorption in CAU-10 at 298.15 K (red squares) and IAST calculations based on the simulated unary isotherms (light red, dotted line) or on the experimental unary isotherms (dark red, dashed line). (b) Corresponding component loadings of n-butanol (shades of orange) and n-hexanol (shades of red): simulation (orange squares for n-butanol or red squares for n-hexanol), IAST calculations based on the simulated unary isotherms (dotted lines, light orange for n-butanol or light red for n-hexanol) or on the experimental unary isotherms (dashed lines, dark orange for n-butanol or dark red for n-hexanol), and unary adsorption isotherms (dash-dotted lines, pale orange for n-butanol or pale red for n-hexanol).

Fig. 31 compares the simulated reduced surface excess (Fig. 31a) and associated component loadings (Fig. 31b) for n-butanol/n-hexanol mixture adsorption in CAU-10 at 298.15 K with the corresponding IAST calculations based on the simulated or experimental unary isotherms. The simulated reduced surface excess of n-hexanol in Fig. 31a shows a steep increase for $x'_{\text{hexanol}} < 0.1$ up to the maximum at $x'_{\text{hexanol}} \approx 0.3$ and a subsequent linear decrease with increasing liquid phase mole fraction of n-hexanol. The U-shape (type II isotherm according to Schay and Nagy) indicates a strong preference for n-hexanol adsorption over the whole range of liquid phase mole fractions of n-hexanol. In contrast to the previously discussed binary mixtures, separation is not driven by entropic effects, because n-butanol and n-hexanol have the same saturation capacity in CAU-10 (see Fig. 18). The IAST calculations based on the simulated unary isotherms agree well with the simulation results, differing only in a reduced n-hexanol excess. The IAST calculations based on the experimental unary isotherms provide qualitatively good results, but overestimate the n-hexanol excess over the whole range of liquid phase mole fractions of n-hexanol. The associated component loadings of n-butanol and n-hexanol shown in Fig. 31b help to explain the deviation between simulation results and IAST calculations. The simulated component loadings show a hyperbolic increase in n-hexanol adsorption up to saturation with a simultaneous decrease in n-butanol uptake up to the complete exclusion of n-butanol from the adsorbed layer at very high liquid phase mole fractions of n-hexanol. The IAST calculations based on the simulated isotherms agree qualitatively with the simulations, but slightly underestimate the n-hexanol uptake and slightly overestimate the n-butanol uptake over the whole range of liquid phase mole fractions of n-hexanol. This minor difference between the simulation results and the IAST

calculations is probably due to a deviation of the adsorbed layer from ideality. The IAST calculations based on the experimental isotherms give qualitatively good results, but underestimate the n-butanol uptake and overestimate the n-hexanol uptake. This is deviation is probably caused by the weaker interaction energy of n-butanol and CAU-10 observed in the experiment (see Fig. 18b and Eq. (G8) in Appendix G).

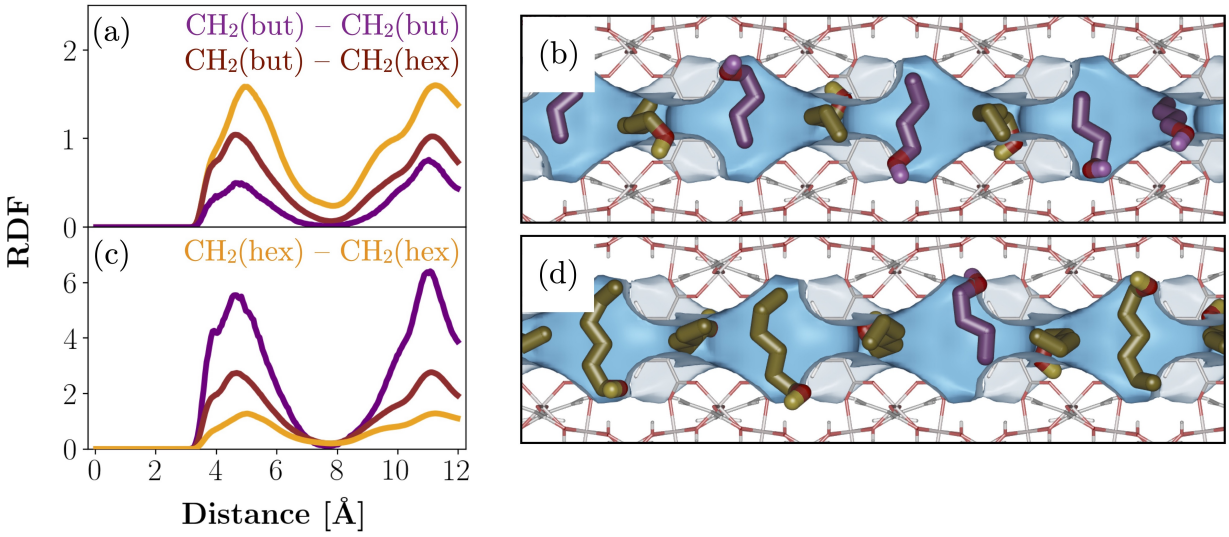

Fig. 32: RDFs of distances for alkyl groups in n-butanol (CH_2(but)) and CH_2(but) (purple), Ha(but) and alkyl groups in n-hexanol (CH_2(hex), brown), CH_2(hex) and CH_2(hex) (orange) as well as corresponding snapshots at (a, b) low n-hexanol loading at $x'_{\text{water}} = 0.1$ or (c, d) medium n-hexanol loading at $x'_{\text{water}} = 0.5$.

In Sec. 3.4.1, we already established that n-hexanol interacts more strongly with the CAU-10 framework than n-butanol – explaining the preferred adsorption of n-hexanol over n-butanol. RDFs of atom pairs representing the guest/guest interactions in the system and the corresponding snapshots in Fig. 32 will help to examine the non-ideality of the adsorbed layer. Guest/guest interactions are represented by the interactions between the alkyl groups of n-butanol or n-hexanol, i.e. CH_2(but) – CH_2(but), CH_2(but) – CH_2(hex), or CH_2(hex) – CH_2(hex). RDFs and snapshots are shown for medium n-hexanol loading at $x'_{\text{hexanol}} = 0.1$ (Fig. 32a, b) or high n-hexanol loading of at $x'_{\text{hexanol}} = 0.5$ (Fig. 32c, d). RDFs for the remaining atom pairs can be found in Fig. L19 and Fig. L20 of Appendix L. The results in Fig. 32 are interpreted as follows:

1) **Medium n-hexanol loading:** The first peaks in Fig. 32a occur at distances that are characteristic for carbon-carbon interactions, i.e. 4.7 Å (CH_2(but) – CH_2(but) or CH_2(but) – CH_2(hex)) or 5.0 Å (CH_2(hex) – CH_2(hex)). These distances are similar to those observed in unary adsorption and suggest that both n-butanol and n-hexanol adsorb in individual cuboids (recall Fig. 21d or Fig. 22d). The heights of the first peaks indicate that the degree of interaction between n-hexanol/n-hexanol pairs (peak height ca. 1.6 for CH_2(hex) – CH_2(hex)) dominates relative to n-butanol/n-hexanol pairs (peak height ca. 1.0 for CH_2(but) – CH_2(hex)) or n-butanol/n-butanol pairs (peak height ca. 0.5 for CH_2(but) – CH_2(but)). The increased degree of interaction for

n-butanol/n-hexanol compared to n-butanol/n-butanol confirms the non-ideality of the adsorbed layer – explaining the increased n-butanol uptake in the simulation relative to the IAST calculations based on the simulated isotherms. The snapshot in Fig. 32b shows that both n-butanol and n-hexanol adsorb in configurations that correspond to the ones observed at saturation in unary adsorption.

2) **High n-hexanol loading:** The first peaks in Fig. 32c occur at the same distances as those observed at low n-hexanol loading. The degree of interaction for n-butanol/n-butanol (peak height ca. 5.5 for CH_2(but) – CH_2(but)) now dominates relative to n-butanol/n-hexanol (peak height ca. 2.7 for CH_2(but) – CH_2(hex)) or n-hexanol/n-hexanol (peak height ca. 1.1 for CH_2(hex) – CH_2(hex)). The high degree of interaction between n-butanol/n-butanol pairs is probably caused by the low n-butanol loading, resulting in fewer n-butanol molecules to average over when determining the RDF. The snapshot in Fig. 32d shows a great similarity to the one observed at low n-hexanol loading, differing only in the reduced amount of n-butanol molecules, because of energetic effects that favor the adsorption of n-hexanol.

We can draw the following conclusions by comparing the results obtained for water/ethanol, water/n-butanol, water/n-hexanol, ethanol/n-butanol, or n-butanol/n-hexanol adsorption in CAU-10 discussed above:

1. Commonly, separation is driven by entropic effects that favor the adsorption of the component with the higher saturation capacity, i.e. water over ethanol, water over n-butanol, water over n-hexanol, or ethanol over n-butanol. But for n-butanol/n-hexanol, adsorption is driven by energetic effects that favor the adsorption of the component with the highest adsorption strength, i.e. n-hexanol over n-butanol.

2. Entropic effects decrease with increasing similarity of the saturation capacities of the components. Compare, for example, water/n-butanol adsorption (type II isotherm according to Schay and Nagy) with water/ethanol adsorption (type I isotherm according to Schay and Nagy).

3. Strong interactions between guest molecules, such as hydrogen bonding, reduce the efficiency of entropic separations; this is most evident in the separation of water and ethanol in CAU-10.

4. IAST is able to describe weak co-adsorption effects, such as those observed at low water loading for water/ethanol adsorption, but fails when co-adsorption becomes more significant.

3.4.3 Three-Component Adsorption

Fig. 33 shows the ternary contour plots of the reduced surface excesses for water/ethanol/n-butanol mixture adsorption at 298.15 K in CAU-10. Fig. 33a confirms the preferential

adsorption of water over the entire range of liquid phase compositions, because of the positive reduced surface excess of water over the whole range of liquid phase compositions. Water excess is strongest in the n-butanol rich corner and decreases with increasing mole fraction of ethanol in the liquid phase. This behavior is consistent with the conclusions drawn from the results obtained for water/n-butanol and water/ethanol adsorption discussed in the previous section. The strong water excess in the n-butanol rich corner is a result of the large difference in packing efficiency of water and n-butanol (see discussion of water/n-butanol separation on pp. 66ff.), while the decreasing water excess with increasing liquid phase mole fraction of ethanol is both due to a high degree of hydrogen bonding between ethanol and water, and a lower difference in packing efficiency (see discussion of water/ethanol separation on pp. 63ff.).

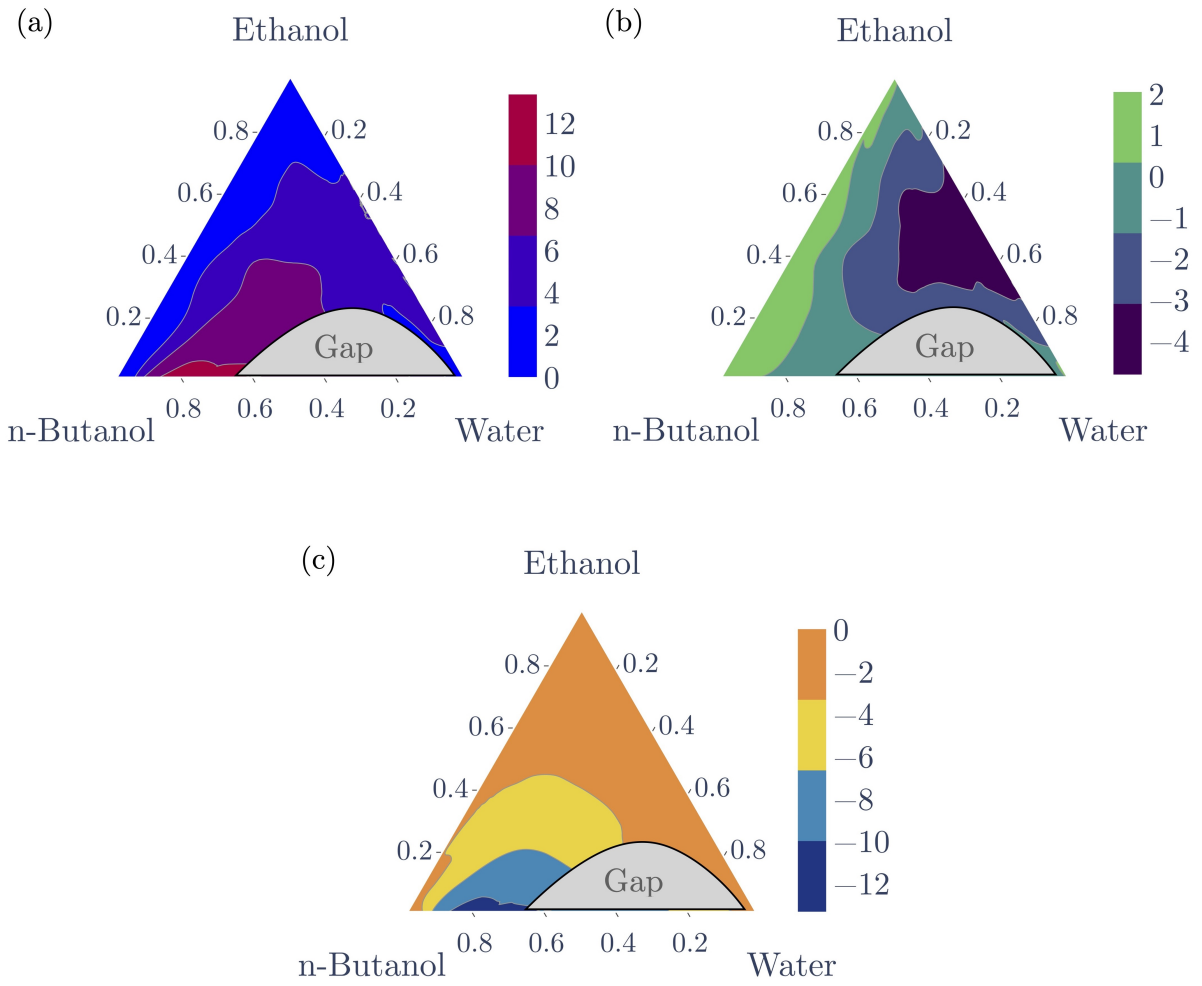

Fig. 33: Ternary contour plots showing the reduced surface excess of (a) water, (b) ethanol, or (c) n-butanol for water/ethanol/n-butanol mixture adsorption at 298.15 K in CAU-10 as a function of the liquid phase composition. Areas between contour lines are colored. Note the miscibility gaps (gray).

Fig. 33b shows a positive ethanol excess for low liquid phase mole fractions of water, which rapidly decreases and becomes negative with increasing mole fraction of water in the liquid phase. Positive and negative ethanol excess are separated by an azeotrope for $x'_{\text{n-butanol}} : x'_{\text{water}} \approx 4 : 1$. The course of the ethanol excess is consistent with the conclusions drawn

from the results obtained for adsorption from the liquid binary mixtures of water, ethanol, and n-butanol discussed in the previous section. The positive ethanol excess for $x'_{\text{n-butanol}} : x'_{\text{water}} > 4 : 1$ is both due to the different packing efficiency of ethanol and n-butanol (see discussion of water/n-butanol separation on pp. 71ff.) and the low water selectivity for water/ethanol adsorption at low liquid mole fractions of water because of the high degree of hydrogen bonding between ethanol and water (see discussion of water/ethanol separation on pp. 63ff.). The negative ethanol excess for $x'_{\text{n-butanol}} : x'_{\text{water}} < 4 : 1$, on the other hand, is due to the higher packing efficiency of water relative to ethanol or n-butanol. Fig. 33c indicates the preferential desorption of n-butanol over the entire range of liquid phase compositions, because the reduced surface excess of n-butanol is never positive. The preferential desorption of n-butanol – despite having the greatest adsorption strength of all three components in CAU-10 – confirms the entropic separation. n-Butanol excess is lowest in the n-butanol rich corner and increases with increasing mole fraction of ethanol or water in the liquid phase. This behavior is due to the same effects as those identified in the discussion of Fig. 33a.

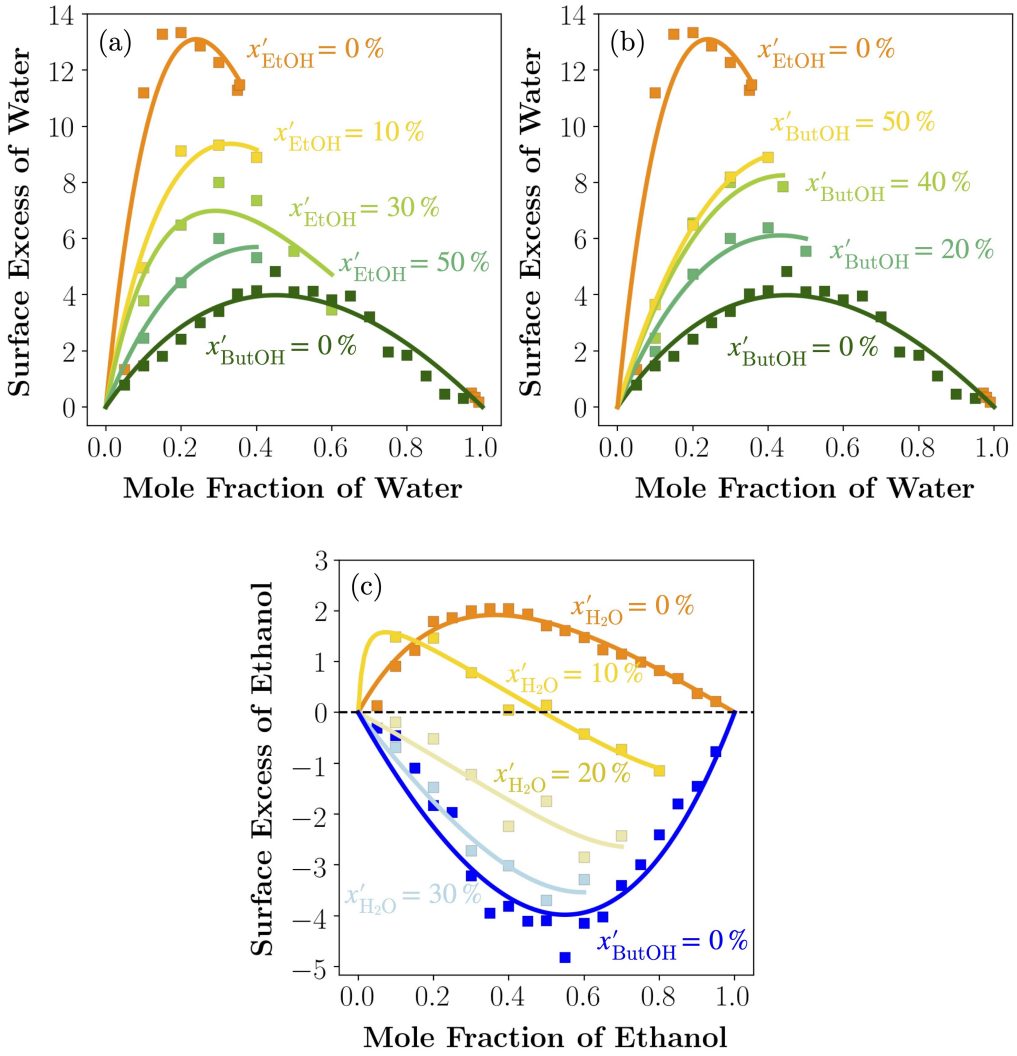

Fig. 34: Slices of the ternary contour plots showing the influence of the increment of (a) ethanol (EtOH) on the water excess, (b) n-butanol (ButOH) on the water excess, or (c) water (H_2O) on the ethanol excess. Simulation (colored squares) and Bi-Langmuir fits (solid lines). Miscibility gaps are not shown in the diagrams.

Fig. 34 shows slices of the ternary contour plots in Fig. 33 to determine the influence of the increment of one component in the ternary mixture on the excess of the remaining two components. Fig. 34a confirms the decreasing selectivity for water adsorption with increasing liquid phase mole fraction of ethanol, because the water excess isotherm changes from type II (high preference for water adsorption) to type I (moderate preference for water adsorption) according to the classification of Schay and Nagy. As pointed out above, the reduction of the water excess with increasing liquid phase mole fraction of ethanol is both due to a high degree of hydrogen bonding between ethanol and water and a lower difference in packing efficiency for water/ethanol relative to water/n-butanol. Fig. 34b confirms the increasing selectivity for water adsorption with increasing liquid phase mole fraction of n-butanol, because the water excess isotherm changes from type I (moderate preference for water adsorption) to type II (high preference for water adsorption) according to the classification of Schay and Nagy. The situation observed in Fig. 34b is reversed to what is observed in Fig. 34a. The increasing water excess with increasing liquid phase mole fraction of n-butanol is due to a higher difference in packing efficiency for water/n-butanol compared to water/ethanol. Fig. 34c confirms the rapid decrease in ethanol selectivity with increasing liquid phase mole fraction of water, because the ethanol excess changes rapidly from being positive to being strongly negative. The decrease in ethanol excess with increasing liquid phase mole fraction of water is due to a higher difference in packing efficiency for water/ethanol compared to ethanol/n-butanol. Note that for $x'_{\text{water}} = 10$ %, ethanol excess is slightly increased for low liquid phase mole fractions of ethanol – but this phenomenon is based on a single data point. The increased ethanol excess here can be explained by the conclusions drawn for water/n-butanol and water/ethanol adsorption The liquid phase in this range consists primarily of n-butanol and contains only minor amounts of ethanol and water. The results obtained for water/n-butanol adsorption suggest a significant enrichment of water in the adsorbed layer with decreasing liquid phase mole fraction of n-butanol (see discussion of water/n-butanol separation on pp. 66ff.), which in turn leads to increased ethanol adsorption due to co-adsorption of ethanol and water (see discussion of water/ethanol adsorption on pp. 63ff.).

3.5 Conclusions

Grand canonical Monte Carlo simulations were conducted to study the separation of alcohol/alcohol or alcohol/water mixtures in a nanoporous material with internal hydrophobicity – a separation process conceivable for the extraction of alcohols from fermentation broths. The results indicate that separation from the liquid phase is commonly driven by entropic effects – favoring adsorption of the component with the highest packing efficiency –, while differences in the adsorption strength only have a significant impact on the separation of components with similar packing efficiencies. However, strong interactions between guest molecules (e.g. between water and the hydroxyl groups of alcohols) reduce the efficiency of entropic separations. In conclusion, entropic effects complicate the separation of fermentation product by adsorption, because of competing water adsorption.

4 Studying Microscopic Water Droplets on the Surface of Metal-Organic Frameworks[3]

4.1 Motivation

Although metal-organic frameworks (MOFs) are among the most recent classes of nanoporous materials, they have already received a high level of attention. Composed of inorganic nodes connected by organic linkers, MOFs present almost unlimited opportunities to design the pore environment. MOFs are thus promising candidates for application in fluid separation, energy storage, and catalysis [87]. However, competing adsorption of water often reduces this potential, for example in CO_2 capture [88], Lewis acid catalysis [89], or hydrocarbon adsorption [90]. For industrial application, sound knowledge of the hydrophobicity of MOFs is therefore crucial. The hydrophobicity of any nanoporous material is usually determined by means of adsorption or contact angle measurements, which differ in their approach of bringing the nanoporous material into contact with water. Whereas the amount of (vapor) water is increased gradually in adsorption experiments or simulations, the solid surface is faced with a whole (liquid) droplet for contact angle measurements. In addition, contact angle experiments provide information about the external surface, while adsorption measurements (mainly) focus on the internal surface. If the external surface structure differs significantly from the internal one, the experiments should provide different results. But frequently no distinction is made between the internal and external hydrophobicity of MOFs based on the unstated assumption that both methods provide similar results [87, 91].

In what follows, we will study the difference between the external and internal hydrophobicity of MOFs. To this end, we conduct molecular dynamics simulations to examine how the design of the external surface influences the spreading of microscopic water droplets on MOFs, and grand canonical Monte Carlo simulations to predict the adsorption of water on the internal MOF surface.

4.2 State of the Research

Contact angle experiments are easy to perform and have thus become a popular technique for characterizing solid surfaces as hydrophilic ($\Theta_{\text{app}} < 90°$) or hydrophobic ($\Theta_{\text{app}} > 90°$). They are also frequently used to make statements about the hydrophobicity of nanoporous materials. For example:

- Lv et al. [92] reported the synthesis of a novel iron-based MOF, {[$Fe_3(\mu 3\text{-O})$](2,2-bis(4-carboxyphenyl)-hexafluoropropane)$_3 \cdot$ 2DMF}, and showed that it is capable of separating hexane isomers with remarkable selectivity. In an attempt to prove that

[3] The content of Section 4 is reproduced in part with permission from A. von Wedelstedt, H. Chen, G. Kalies, and R. Q. Snurr, *Langmuir*, 2020, **36**, 13070–13078. Copyright 2020 American Chemical Society.

separation efficiency will not be affected by competing water adsorption, the authors measured the contact angle of water and obtained $\Theta_{app} = 110°$.

- Hu et al. [93] studied the separation of propane and propene in node-functionalized versions of UiO-66 (i.e. $[Zr_6O_4(OH)_4$(benzene-1,4-dicarboxylate)$_6]$). Functionalization was achieved by adding *ortho*-, *meta*-, or *para*-(trifluoromethyl)benzoic acid during synthesis. Accordingly, we will refer to the resulting MOFs as ortho-UiO-66, meta-UiO-66, or para-UiO-66. The authors showed that their node-functionalized MOFs are able to separate propane and propene with remarkable selectivity. To prove that competing water adsorption will not affect the separation efficiency, they measured the contact angle of water and obtained $\Theta_{app} = 52.1°$ (UiO-66), $\Theta_{app} = 128.2°$ (ortho-UiO-66), $\Theta_{app} = 92.6°$ (meta-UiO-66), and $\Theta_{app} = 93.2°$ (para-UiO-66).

- Gan et al. [94] reported a novel Cu(II)-based MOF, {Cu_4(1,7-di(4-carboxyphenyl)-1,7-dicarba-*closo*-dodecaborane)$_4$(1.4-diazabicyclo[2.2.2]octane)(H_2O)}, and showed that it is capable of separating n-butanol from mixtures of acetone, n-butanol, and ethanol. Aqueous acetone/n-butanol/ethanol (ABE) mixtures are formed during fermentation and contain almost exclusively water. In an attempt to prove that competing water adsorption will not affect the separation performance, the authors measured the contact angle of water ($\Theta_{app} = 144°$) and the adsorption isotherm of vapor water (type II isotherm according to the IUPAC classification). Gan et al. concluded that both results confirm the hydrophobicity of their MOF, which is incorrect, because Type II isotherms indicate favorable adsorbate-adsorbent interactions due to the formation of an adsorbate monolayer at low pressures.

All authors make the unstated assumption that contact angle measurements and water adsorption experiments provide similar results. But both methods differ in the surface they examine. Contact angles provide information on the external surface, while adsorption studies focus (mainly) on the internal surface. If the external surface structure differs significantly from the internal one, the experiments should provide different results. This assumption is supported in particular by the findings of Gan et al. [94] above, who measured a hydrophobic external MOF surface (contact angle) and a hydrophilic internal MOF surface (water adsorption isotherm) [94]. This chapter aims to investigate the difference between the external and internal hydrophobicity of MOFs.

Before we move one, we should point out that there is strong evidence that droplets – whether microscopic or macroscopic ones – on real surfaces are usually trapped in metastable states. For example, Liu and Choi [95] conducted studies on the condensation of microscopic water droplets on lotus leaves and found that droplets can be transitioned from the intermediate Wenzel/Cassie-Baxter state to the pure Cassie-Baxter state by supplying a slight amount of kinetic energy through vibration. Similarly, Decker and Gareff [96] studied the influence of vibrations on the contact angle of water by means of a (macroscopic) capillary rise exper-

iment and showed that the apparent contact angle decreases with increasing input energy until an equilibrium contact angle is reached at higher input energies. Marmur [14] concluded that real solid surfaces have many energy barriers that a droplet must overcome. The higher the kinetic energy initially supplied to the droplet, the higher the number of energy barriers it can overcome before all the initial kinetic energy is dissipated by friction (resulting in mechanical instead of thermodynamic equilibrium) ([11], p. 43).

4.3 Methodology

Molecular dynamics simulations were conducted to examine how the design of the external surface influences the spreading of microscopic water droplets on MOFs, and grand canonical Monte Carlo simulations were performed to predict the adsorption of water on the internal MOF surface. The true nature of MOF external surfaces is normally not known, but Jung et al. [97] confirmed the existence of a certain amount of carboxylate groups (as present in organic linkers) on the external surface of a novel In(III)-based MOF, [In(1,4-phenylenediacetate)$_2$($(C_2H_5)_2NH_2$)]$_n$, and two zinc-based MOFs, IRMOF-3 and [Zn(2,2′-bipyridine-5,5′-dicarboxylate)$(H_2O)_2$]$_n$, by activation with carbodiimides and conjugation of an enhanced green fluorescent protein. Based on the results of Jung et al., we decided to model the external surfaces of MOF-5 and CAU-10 completely with protruding organic linkers and to control the hydrophobicity of the surface model by the selection of the protruding linkers. This approach allows us to investigate the relationship between surface nature and contact angle but could provide different results than in experiments.

4.3.1 Surface Modeling

MOF-5 and CAU-10 were modeled with external surfaces of different hydrophobicities, while the internal surfaces were maintained. MOF-5, [Zn_4O(1,4-benzenedicarboxylate)$_3$], was first synthesized by Yaghi and coworkers [98] in 1999 by linking Zn_4O tetrahedra via 1,4-benzenedicarboxylate (H_2BDC) linkers to yield a microporous MOF with cubic symmetry (space group $Fm\bar{3}m$). Depending on the orientation of the benzene ring in the BDC linker, MOF-5 possesses two pore types with different diameters (see Tab. 4) [99]. Several studies concern the hydrophobicity of MOF-5, for instance contact angle experiments conducted by Nguyen and Cohen [100] or water adsorption experiments performed by Ming et al. [101]. At first glance the results are contradictory: while Nguyen and Cohen report an immediate absorption of the droplet in their contact angle measurements, which is characteristic for superhydrophilic materials, Ming and coworkers obtained a type V isotherm (according to the IUPAC classification), which is characteristic for a hydrophobic material [100, 101]. It is now known that the MOF-5 framework is degraded upon water adsorption connecting the results of Nguyen et al. and Ming et al. [100, 102]. Since the degradation of MOF-5 interferes with the aims of this study, we decided to ignore the degradation of the framework during the construction of our force field. The results for MOF-5 reported in the following can hence be understood to illustrate the behavior of microscopic droplets on MOF-5 and

the influence of the external surface design on the contact angle if no degradation of MOF-5 would occur. CAU-10, [Al(OH)(benzene-1,3-dicarboxylate)], was first synthesized in 2012 by Stock and coworkers [5] by linking chains of cis-connected corner-sharing AlO_6 polyhedra via 1,3-benzenedicarboxylate to yield a tetragonal lattice (space group $I/4_1$). Thanks to its high pH and solvothermal stability, CAU-10 has already received a remarkable level of attention, in particular for the application in adsorption-driven chillers [103]. The experimental and simulated water adsorption isotherms of CAU-10 reported by several groups confirm the internal hydrophobicity of CAU-10 [7, 104]. A brief compilation of important geometric properties of MOF-5 and CAU-10 can be found in Tab. 4.

Tab. 4: Geometric properties of MOF-5 and CAU-10.

	MOF-5	**CAU-10**
Pore Diameters [Å]	7.8, 15.2 [99]	4, 7 [5]
Pore Volume [ml/g]	0.6 [105]	0.25 [5]
Surface Area [m^2/g]	3320 [105]	564 – 656 [5]

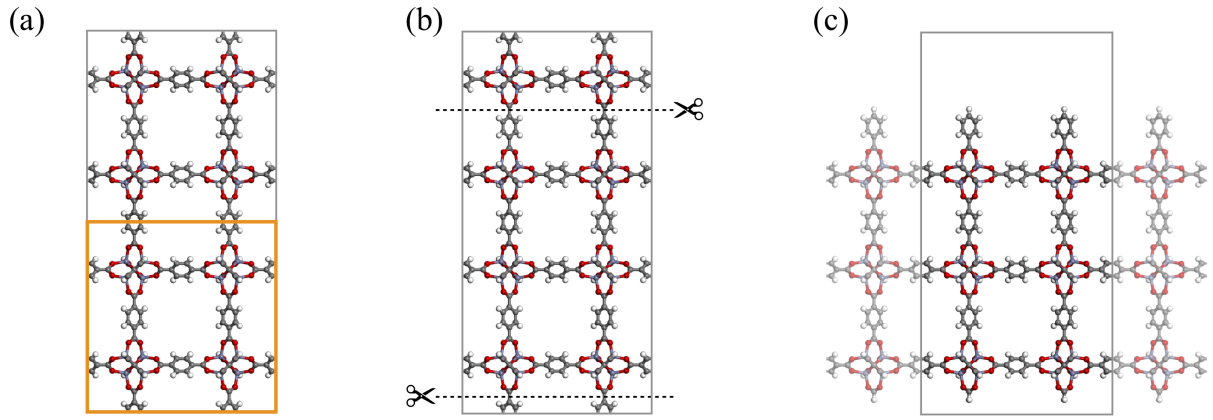

Fig. 35: Building of surface models using the example of MOF-5-phenyl. (a) The unit cell of MOF-5 (orange framed) is vertically replicated, and (b) then terminated at a desired desired upper and lower plane. (c) The resulting lower and upper cuts are capped with hydrogen atoms to form a surface module of protruding phenyl linkers, which is replicated in both directions of the interfacial plane to form the external surface model.

The investigated MOF surfaces were modeled using the crystal structure data of MOF-5 and CAU-10 available in the Cambridge Structural Database [5, 98]. Fig. 35 illustrates the general procedure for building the external surfaces. First, the unit cell of the investigated MOF (orange frame in Fig. 35a) was vertically replicated, and the obtained supercell was terminated at a desired upper and lower plane (Fig. 35b). The resulting lower cuts were then capped with hydrogen atoms and the upper cuts by either carboxylic acid groups or hydrogen atoms to form external surfaces of different hydrophobicity (Fig. 35b). The thus obtained surface layer was geometrically optimized by means of the DMol3 software package implemented in Materials Studio 7.0 (BIOVIA) using the Perdew-Burke-Ernzerhof

(PBE) generalized gradient approximation (GGA) functional [76] combined with a double-numeric quality basis set with polarization functions (DNP). The calculations comprised only the relaxation of the atomic positions of the metal-organic framework, whereas the cell parameters and angles between the cell vectors were maintained fixed at the initial values. To form the surface models of MOF-5 and CAU-10, each geometrically optimized surface layer module was replicated in both directions of the interfacial plane as shown in Fig. 35c.

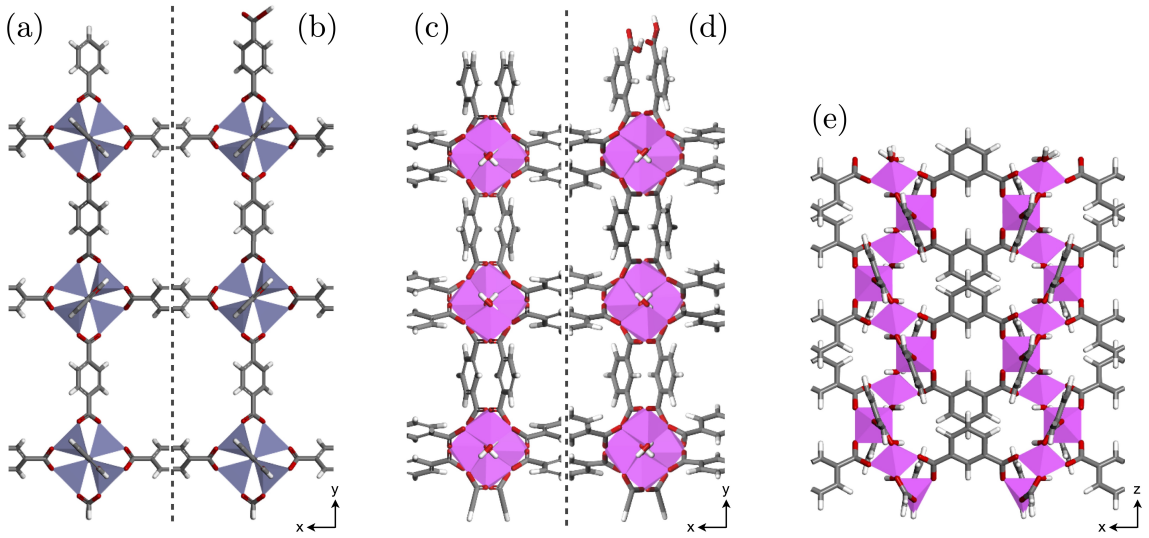

Fig. 36: Side views of surface layer modules of (a, b) MOF-5 and (c-e) CAU-10 with either protruding (a, c) phenyl or (b, d) benzoic acid linkers, and (e) with accessible pores for CAU-10 based on an alternative orientation of its unit cell.

The five different external surface models used in this study are illustrated in Fig. 36. In the following, MOF surfaces with protruding phenyl or benzoic acid linkers will be identified by the suffix *phenyl* or *benzoic*, and the surface model of CAU-10 with accessible pores on the surface in Fig. 36e will be denoted *CAU-10-open*. For CAU-10-open, electric charges of the exposed inorganic nodes are neutralized by means of one water molecule and one hydroxyl group. Although phenyl linkers will not be present after synthesizing MOF-5, as this would require the dissociation of the BDC linker during synthesis, capping reagents or modulators are often used to control the crystal size and hence the nature of the external surface [106]. The modification of the external surface of MOF-5 in the literature comprises for instance the use of capping reagents like triphenylacetic acid, diphenylacetic acid, trimethylacetic acid, acetic acid, diisopropylethylamine and triethylamine by Homan et al. [107] or 4-decylbenzoic acid by Zacher et al. [106]. Benzoic acid is another common capping reagent [108–110] and while it has, to the best of our knowledge, not yet been used to modify the external surface of MOF-5, this approach allows the formation of surface models with very different structures for contact angle simulations.

4.3.2 Simulation Details

The freely movable droplet used in each simulation comprised 3301 water molecules, and was formed by initially arranging 3375 water molecules in a cubic block within an otherwise

empty simulation box of dimensions 300 Å× 300 Å× 300 Å and performing molecular dynamics simulations for 10^6 time steps in the canonical (NVT) ensemble. Molecular dynamics simulations were performed using the LAMMPS software package developed at Sandia National Laboratories [111]. Newton's equations of motion were integrated numerically using the velocity Verlet algorithm [26] with a time step of 1 fs. The Nosé-Hoover thermostat [40] with a damping factor of 100 time steps was used to maintain the temperature of the system at 298 K. Water molecules were represented by the TIP4P-Ew model, which consists of a single Lennard-Jones interaction site (oxygen) and three charged sites, i.e. two hydrogen atoms and a massless unit in between which bears the charge of oxygen [78]. Non-bonded interactions were described by an interatomic potential consisting of Lennard-Jones and Coulomb interactions, while the SHAKE algorithm was used to maintain the internal bond lengths and bond angle [112]. Lennard-Jones parameters between unlike atomic species i and j are computed by means of the Lorentz-Berthelot mixing rules [44, 45]. Lennard-Jones interactions were truncated by means of a 12.8 Å cutoff, while the particle-particle particle-mesh (PPPM) [113] with an accuracy of 10^{-6} was used to handle long-range Coulombic interactions. Once a spherical droplets was formed, all water molecules that were not in its liquid phase were discarded.

The droplet, now comprising 3301 water molecules, was placed in the vicinity of each of the five MOF surfaces illustrated in Fig. 36, and molecular dynamics simulations in the canonical (NVT) ensemble were conducted for (usually) $5 \cdot 10^6$ time steps to reach equilibrium. The atomic positions of the metal-organic framework were maintained fixed during the simulation, and interactions between the atoms of the framework were neglected to reduce computation time. The Lennard-Jones parameters for each atom of the metal-organic frameworks were taken from the DREIDING force field [49] (organic linker) and the Universal force field [48] (metal atoms and their bridging ions), while their partial charges were determined by performing single point energy calculations followed by a Mulliken analysis [50] in DMol3 implemented in Materials Studio 7.0 (BIOVIA) using the PBE GGA functional [76] combined with a DNP basis set. The thus obtained partial charges for the periodic unit cells agree well with those reported in the literature (see Appendix M). To use the PPPM long-range interaction solver, periodic boundary conditions had to be applied, defining a series of stacked plates of solid material separated by empty space. This vacuum space was enlarged by using a vertical cell parameter of 300 Å to prevent interactions between the droplet and the periodic image of the MOF surface adjacent to the upper boundary of the simulation box. Apart from that, molecular dynamics simulations were carried out using the same settings as for the formation of the spherical droplets. Simulations were repeated for a system starting from a perturbed starting configuration if the results indicated the possibility of a trajectory dependence. For the repetitions of the simulations, the initial position of the droplet was slightly altered to perturb the system, and such instances are mentioned in the discussion. During the simulation, some water molecules will leave the liquid water

droplet and form the vapor phase of the system. The pressure of the vapor phase above the curved droplet interface will exceed the vapor pressure of water, as indicated by Kelvin's equation ([114], pp. 5–6),

$$\ln\left(\frac{p_{0,\mathrm{Dc}}}{p_0}\right) = \frac{2\sigma V_\mathrm{m}}{r_\mathrm{Dc} RT}, \tag{209}$$

where $p_{0,\mathrm{Dc}}$ is the vapor pressure above the curved droplet interface, p_0 is the vapor pressure of water above a flat interface, r_Dc is the radius of the (convex) droplet curvature, and V_m is the molar volume of water. Kelvin's equation obeys thermodynamics and provides thus a criterion to determine whether the system is *not* in thermodynamic equilibrium, i.e. if $p_0 > p_{0,\mathrm{Dc}}$. The pressure of the vapor phase in the simulation, $p_{0,\mathrm{Dc}}$, was determined by dividing the simulation box into three different regions: solid phase, liquid (droplet) phase, and vapor phase, and summing and averaging the per-atom stress tensor for each water molecule in the vapor phase over 10^6 time steps.

Grand canonical Monte Carlo (GCMC) simulations were performed to predict the adsorption of water in MOF-5 and CAU-10 at 298 K using the same force field as stated above. The simulation boxes consisted of a single unit cell for MOF-5 or $2 \times 2 \times 4$ unit cells for CAU-10. The van-der-Waals interactions were truncated by means of a 12 Å cutoff, and the Ewald summation [55] was used to handle long-range Coulombic interactions. Water molecules were represented by the TIP4P-Ew model [78]. For MOF-5, 50 water molecules were created and equilibrated in the canonical ensemble before running the GCMC simulations to accelerate the uptake. All GCMC simulations were performed using the RASPA simulation package [31] with automatically computed gas-phase fugacity values by means of the Peng-Robinson equation of state [71]. The simulations comprised on average 1.7×10^6 cycles for equilibration and 10^5 cycles for the production run. Further simulation details can be found in Appendix M.

4.3.3 Contact Angle Determination

The extraction of the contact angle, Θ_app, from the simulated data is explained and illustrated in Fig. 37. After equilibration, a production run using the above mentioned simulation settings and comprising 10^6 time steps is performed. Over the course of this production run, a three-dimensional lattice consisting of cubes with edge lengths of 3 Å (sketched in Fig. 37a) is overlaid on the droplet, which rests on the external surface of the MOF. For each cube, the number of enclosed water molecules and hence the density is calculated every 20th time step and subsequently averaged over all computed values to yield a three-dimensional water density map. Two axial slices of this density map (green cubes in Fig. 37b) and for MOF-5 in addition two diagonal slices (grey cubes in Fig. 37b) are selected and plotted as tile images (see Fig. 37c). Subsequently, a least-square circle fit is used to define the contour line of the droplet [115]. Following the work of Isaiev et al. [116], the contour line is based on points of the periphery of the water droplet with a density value of about half of the bulk

density of water (to be specific: 0.4 g/ml – 0.6 g/ml), and points of the density map within 5 Å of the tip of the protruding linkers are not taken into account to exclude the influence of the MOF surface irregularities and density fluctuations occurring in the vicinity of the solid-liquid interface. Knowing the droplet radius, r_{Drop}, the center of the fitted circle, and the surface height, we are able to calculate the apparent contact angle,

$$\Theta_{\text{app}} = \begin{cases} 180° - \arccos\left(\frac{d_{\text{sf-cc}}}{r_{\text{Drop}}}\right), & \text{if } d_{\text{sf-cc}} \geq 0; \\ \arccos\left(\frac{d_{\text{sf-cc}}}{r_{\text{Drop}}}\right), & \text{otherwise,} \end{cases} \tag{210}$$

where $d_{\text{sf-cc}}$ is the distance between the surface level (sf) and the circle center (cc).

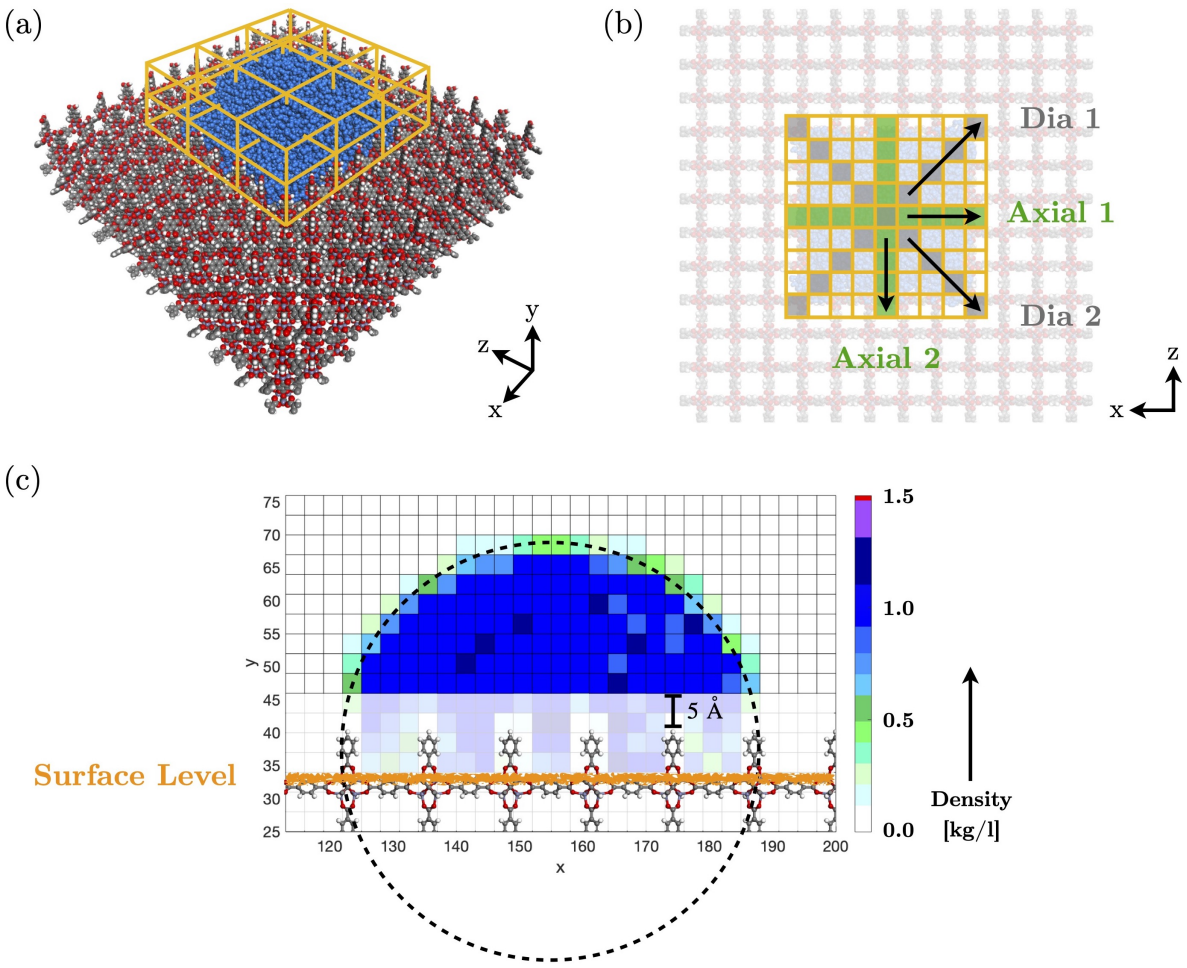

Fig. 37: Calculating the contact angle. (a) During production run, a three-dimensional grid of cubes is overlaid on the droplet. The average density of water is calculated for each cube, resulting in a three-dimensional water density map. (b) Four slices are selected from the density map to calculate the contact angle, designated Axial 1, Axial 2, Dia 1, and Dia 2. (c) For each of these slices, the contour line of the droplet is determined using a least-square circle fit.

4.4 Results and Discussion

Insights into the wetting of the MOF surface models can be gained from the movie files on the CD-ROM that accompanies this dissertation or the snapshots shown below. Snapshots

of the droplet resting on the surface of MOF-5-phenyl and MOF-5-benzoic are shown in Fig. 38. The side views in Fig. 38a and b for MOF-5-phenyl as well as in Fig. 38d and e for MOF-5-benzoic indicate the internal hydrophobicity of MOF-5, because no water enters the framework of MOF-5. The internal hydrophobicity of MOF-5 is also confirmed by the computed water adsorption isotherm of MOF-5 in Fig. 39a, which shows no adsorption at low pressure and then a sudden increase in water adsorption up to 130 wt % at a relative pressure of about 0.4. The top views in Fig. 38c and f show that the droplets adapt their spreading to the nature of the external surface and form angular (MOF-5-phenyl) or square contact areas (MOF-5-benzoic). Since MOF-5-phenyl and MOF-5-benzoic have the same internal hydrophobicity, the different spreading of the droplet is mainly determined by the different external surfaces. As a result, the hydrophobicity and orientation of the protruding linkers affect the size of the propagation area. In the case of MOF-5-phenyl, the protruding phenyl linkers (red colored in the enlarged section of Fig. 38c) impede the spreading of the droplet and lead to a compression in the z-direction, which is also visible in the side views in Fig. 38a and b, and has been reproduced in a repetition of the simulation starting from a perturbed starting configuration. Because of the lower hydrophobicity of the protruding benzoic acid linkers, the contact area of the droplet on MOF-5-benzoic is increased and the propagation in a certain direction is not favored. By employing a force field that allows for flexibility of the protruding organic linkers, the limitation of the droplet spreading might be lifted. But, due to the long simulation time of approximately 4 weeks for a single system, no internal degrees of freedom were considered.

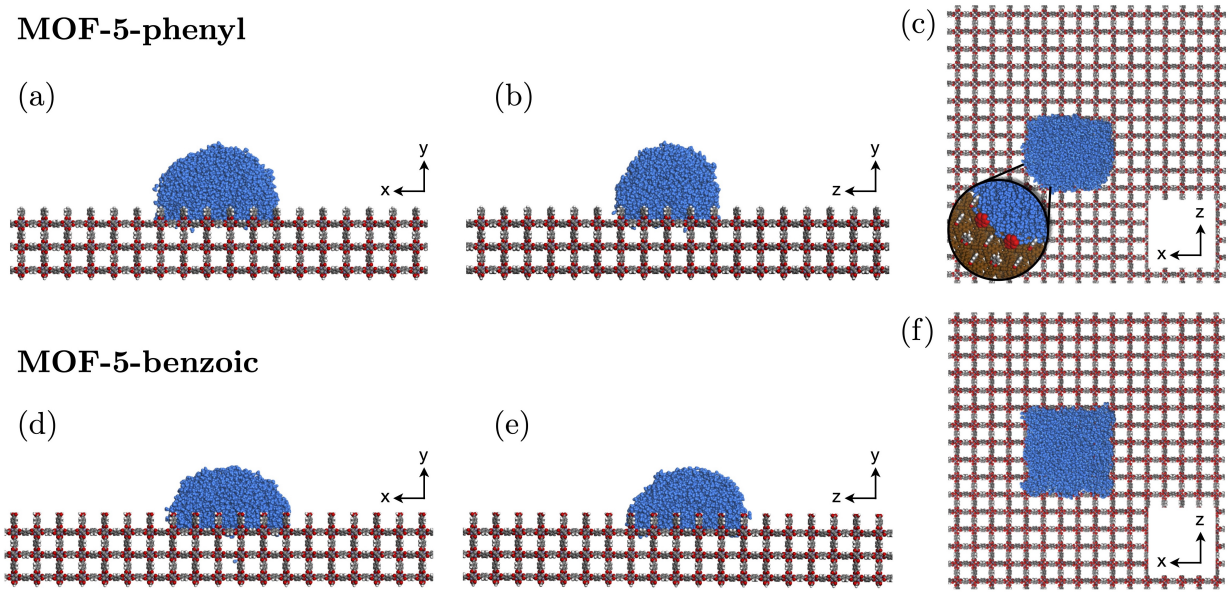

Fig. 38: Snapshots of the droplet on (a-c) MOF-5-phenyl and (d-f) MOF-5-benzoic from different perspectives.

The corresponding contact angles can be found in Tab. 5. The contact angles of the first column of Tab. 5 confirm the hydrophobic nature of the external surface of MOF-5-phenyl and account for the limited spreading in the z-direction by means of an increased contact angle for the slice Axial 2. Furthermore, the contact angles calculated from the diagonal

slices Dia 1 and Dia 2 provide similar results, representing the approximate mean of the axial contact angles. In the case of MOF-5-benzoic, the contact angles shown in the second column of Tab. 5 can be assigned to the threshold between hydrophobicity and hydrophilicity. Considering that benzoic acid linkers consist of a carboxylic acid group (hydrophilic) and a benzene ring (hydrophobic), the obtained contact angles of about 90° are reasonable. The slightly different contact angles for the axial and diagonal slices in case of MOF-5-benzoic are likely caused by the quadratic shape of the contact area, inducing a not fully hemispherical droplet.

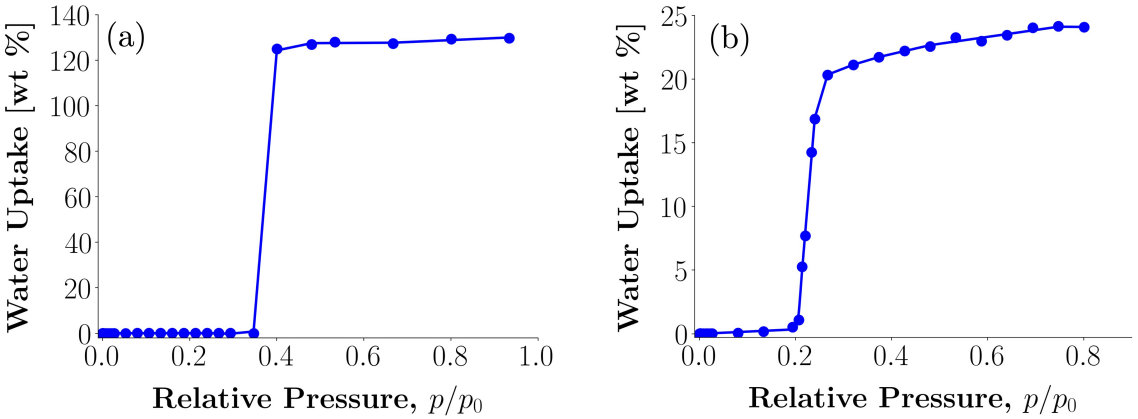

Fig. 39: Simulated water adsorption isotherms at 298 K of (a) MOF-5 and (b) CAU-10. The water uptake in weight percent (wt %) is given by [mass(water)/mass(MOF)] × 100, and the vapor pressure, p_0, of the TIP4P-Ew model at 298 K was considered to be 3749 Pa [117]. A comparison of the shown simulated adsorption isotherms with the experimental ones by Ming et al. [101] or Cadiau et al. [7] can be found in Fig. P1 in Appendix P.

Tab. 5: Contact angles for MOF-5-phenyl and MOF-5-benzoic. Compare Fig. 37b for the naming of the slices.

Slice	MOF-5-phenyl	MOF-5-benzoic
Axial 1	101°	91°
Axial 2	115°	91°
Dia 1	106°	85°
Dia 2	105°	89°

The results obtained from the simulations are logically consistent, but do not agree with the contact angle experiments conducted by Nguyen and coworkers [100], in which the absorption of the droplet into MOF-5 was immediate, leading to a $\Theta_{app} \approx 0°$ contact angle. The discrepancy between the experimental and simulated contact angles can be clarified by the adsorption isotherms measured by Ming et al. [101] which indicate an alteration of MOF-5 upon water adsorption. The adsorption isotherms of Ming et al. [101] show a sudden increase in water adsorption up to 12 wt % at a relative humidity of about 45 % as well as a large hysteresis (type V isotherm according to the IUPAC classification), along with a negligible desorption of adsorbed water (see Fig. P1 in Appendix P). The significantly

reduced adsorption capacities in subsequent adsorption/desorption cycles indicate an alteration of MOF-5 [101]. It is now known that the framework of MOF-5 is degraded at a water content of 4 wt % or more [102], providing an explanation for the difference between simulated and experimental contact angles. The computed adsorption isotherm in Fig. 39a does not consider the degradation of the framework upon water adsorption. In accordance with the experimental adsorption isotherm of Ming and coworkers, the computed isotherm shows a turning point at a relative pressure of about 0.4, but a significantly higher saturation capacity of up to 130 wt %. The reduced adsorption capacity in the experiment is probably due to the degradation of the MOF-5 framework. At the first glance, it is not immediately apparent why the droplets in the molecular dynamics simulations do not enter the framework of MOF-5-phenyl or MOF-5-benzoic although the simulated water adsorption isotherm in Fig. 39a indicates the saturation of MOF-5 at relative pressures close to 1 (i.e. at a condition corresponding to the presence of liquid water) and the pore openings are sufficiently large for water molecules to enter (see Fig. P2 in Appendix P). It is conceivable that the droplet on MOF-5-phenyl or MOF-5-benzoic is trapped in a metastable state, which the system cannot escape within the typical time scale of a molecular dynamics simulation. The idea of metastability is confirmed by the vapor phase pressures, $p_{0,\mathrm{Dc}}$, calculated for MOF-5-phenyl (2514 Pa) or MOF-5-benzoic (1728 Pa), which are significantly lower than the vapor pressure of water ($p_0 = 3749$ Pa for TIP4P-Ew model) at 298 K (recall Eq. (209)). Moreover, there is experimental evidence that an adsorption isotherm that indicates the saturation with water at relative pressures close to 1 does not necessarily correspond to an absorption of a droplet in contact angle experiments. For instance, the water adsorption isotherm and contact angle of CAU-1 or UiO-66 were measured at 298 K by Zheng et al. [118] or He et al. [119]. While the water adsorption isotherms show high uptake of water, the contact angle experiments indicate *no* absorption of the droplet [118, 119]. These experiments also support the idea of a metastability.

Snapshots of the microscopic droplet resting on the external surface of the hydrothermally stable metal-organic framework CAU-10 are shown in Fig. 40. The top views of the droplet on CAU-10-phenyl and CAU-10-benzoic in Fig. 40c and f indicate a preferred propagation in the z-direction caused by the orientation of the protruding linkers that form obstacles for the droplet propagation in the x-direction (enlarged sections in Fig. 40c and f). It can be assumed that the contact angle and the shape of the solid-liquid contact area depend on both size and shape of the droplet. As a result, the hydrophobicity of the protruding linkers has a greater influence on the droplet propagation in the z-direction than in the x-direction. However, in a repetition of the simulations, it was observed that in case of CAU-10-phenyl the droplet can surpass a row of protruding linkers, which impede the propagation in the x-direction. The movie file named „Re_CAU-10-phenyl_xz.mp4“ on the CD-ROM accompanying this dissertation shows that the propagation in the x-direction is caused by a collapse of the droplet flank shortly after its contact with the MOF surface. A repetition of the

simulation of the droplet on CAU-10-benzoic confirmed the results in Fig. 40d-f. The vapor phase pressures calculated for CAU-10-phenyl (2037 Pa) or CAU-10-benzoic (861 Pa) are significantly lower than the vapor pressure of water (3749 Pa for the TIP4P-Ew model) at the same temperature, and indicate that the systems are metastable. In the case of CAU-10-open, Fig. 40g and h show an absorption of the droplet into the MOF interior. Recalling the look of the pore in Fig. 36e, it seems that the inorganic nodes form a *guidance*, directing the droplet into the framework.

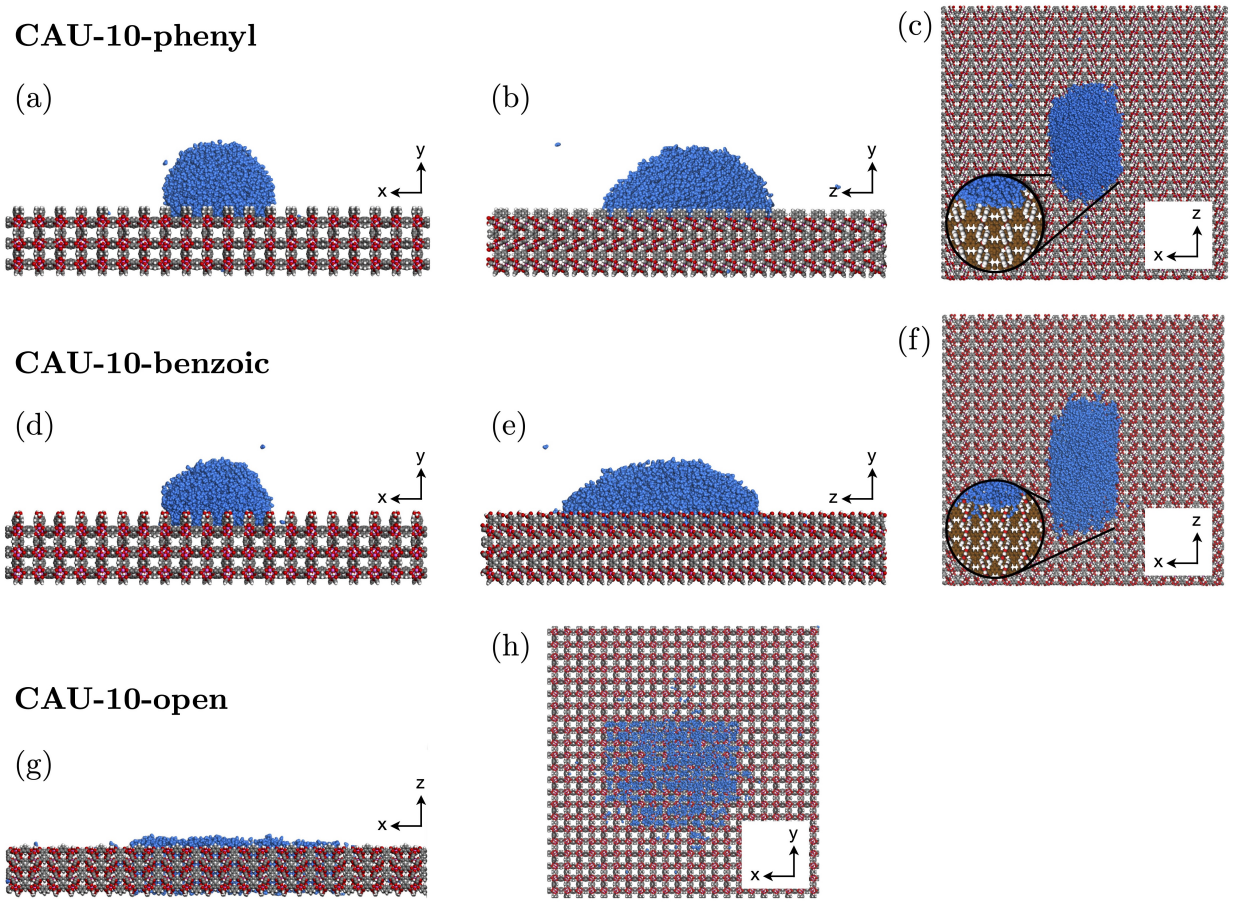

Fig. 40: Snapshots of the droplet on (a–c) CAU-10-phenyl, (d–f) CAU-10-benzoic and (g, h) CAU-10-open from different perspectives.

The contact angles calculated by means of Eq. (210) and found in Tab. 6 agree with the explanation of the snapshots in Fig. 40. The large difference in contact angles for the slices Axial 1 and Axial 2 is due to the preferred propagation of the droplet in the z-direction based on the orientation of the protruding organic linkers as explained above. Because of the restricted droplet spreading in the x-direction no diagonal slices were analyzed. The contact angles of the slice Axial 1 for CAU-10-phenyl and CAU-10-benzoic confirm the assumption that the propagation in the x-direction is largely predetermined by the size of the droplet and less dependent on the hydrophobicity of the protruding organic linkers than the propagation along the z-axis (Axial 2). In case of CAU-10-open, contact angles were not calculated explicitly, since the absorption of the droplet indicates a superhydrophilic character of the external surface corresponding to a contact angle of zero degrees. However,

this superhydrophilicity is contradictory to the hydrophobic character of CAU-10 from the simulated water adsorption isotherm in Fig. 39b. This isotherm shows very low adsorption at low pressure and then a steep increase in water adsorption up to 25 wt % at a relative pressure of about 0.2 and is in good agreement with the experimental isotherm measured by Cadiau et al. [7], which indicates only small hysteresis (type V isotherm according to the IUPAC classification). Due to the lack of experimental contact angles for CAU-10, no statement can be made about the validity of the droplet simulations. The different adsorption energies of water, i.e. -30.16 ± 0.65 kJ/mol for CAU-10 versus -18.38 ± 0.56 kJ/mol for MOF-5, support the preferred absorption of the droplet into CAU-10-open.

Tab. 6: Contact angles for CAU-10-phenyl, CAU-10-benzoic and CAU-10-open. Compare Fig. 37b for the naming of the slices.

Slice	CAU-10-phenyl	CAU-10-benzoic	CAU-10-open
Axial 1	110°	103°	$\approx 0°$
Axial 2	76°	60°	$\approx 0°$

The presented results for MOF-5 and CAU-10 indicate that in droplet simulations the behavior of a droplet sensitively depends on the nature of the external MOF surface. An analogous situation, in which the nature of the external surface significantly influences the results of the simulations, can be observed when comparing simulated and experimental transport resistances inside zeolites and during the passage through the external zeolite surface. While the simulated and experimental transport resistances inside the zeolites are essentially identical (analogous to the adsorption isotherms in this study), the transport resistance during passage through the external zeolite surface depends on the chosen surface structure in the simulations and deviates from the experimental ones (analogous to the contact angle simulations in this study) [120–122].

4.5 Conclusions

Molecular dynamics simulations of microscopic droplets on the external surfaces of metal-organic frameworks were conducted to study to what extent the nature of the external surface influences the contact angle, a quantity that is commonly considered to make general statements about the hydrophobicity of porous materials. The external surfaces of the metal-organic frameworks MOF-5 and CAU-10 were designed with either protruding phenyl or benzoic acid linkers, whereas the internal surface was maintained. A third surface model of CAU-10 was studied to account for the crystallographic asymmetry in the z-direction. We can draw two conclusions from the results:

1. The external surface nature of a metal-organic framework has a major impact on the contact angle and spreading, suggesting the necessity of a clear distinction between internal and external hydrophobicity. The hydrophobicity determined from a contact angle and from an adsorption isotherm experiment do not necessarily have to correlate.

2. Comparison of the vapor phase pressures above the curved droplet interface with the vapor phase of water above a flat interface support the idea of Marmur [14] that droplets on real surfaces are trapped in metastable states.

5 MOF-VR: A Virtual Reality Tool to Study Diffusion in Metal-Organic Frameworks[4]

5.1 Motivation

MOFs are promising emerging materials due to their wide range of potential applications, including gas storage, carbon capture, fluid separation, catalysis, drug delivery, or electrocatalysis. For use in adsorption processes, nanoporous MOFs are of particular interest, because of the high surface-to-volume ratio. The modular structure – MOFs consist of inorganic nodal building blocks that are connected by organic linking building blocks –, allows the prediction of yet unknown nanoporous MOFs in computer experiments [56, 123]. These hypothetical MOFs can be used in molecular dynamics simulations to study the diffusion of guest molecules. The emergence of immersive visualization techniques such as virtual reality (VR) greatly simplifies the interpretation of atomic or molecular trajectories in molecular dynamics simulations. VR tools offer a number of advantages in research and education:

- Enhanced Immersion: Head-mounted displays and haptic feedbacks give the user the impression of being fully immersed in a virtual reality [124].
- Simplified Interpretation: Three-dimensional visualization of atomic or molecular trajectories facilitates the identification of subtle processes in a dynamic system [125].
- Improved Understanding: Virtual reality tools can complement education to provide a deeper understanding of what is being taught [126].

In the following, MOF-VR is presented: a virtual reality program for performing and visualizing interactive molecular dynamics simulations in MOFs. It is further shown that consumer technology is capable of performing state-of-the-art molecular dynamics simulations in virtual reality and that the molecular dynamics routine implemented in MOF-VR is able of providing reliable simulation results.

5.2 State of the Research

In recent years, virtual reality applications have become popular in industry [127–129], research [130–132], and education [133–135]. When we speak of virtual reality applications, we mean technologies that place users in a completely virtual world with which they can interact in a realistic way ([136], p. 2). This virtual world is often displayed using cave automatic virtual environments (CAVEs) or head-mounted displays (HMDs). CAVEs are room-sized boxes in which projectors are used to project the virtual world onto the walls, while HMDs are headsets or goggles with displays instead of lenses to show the virtual world [137, 138]. In the field of molecular chemistry, virtual reality applications are used either for:

[4] The content of Section 5 is reproduced in part with permission from A. von Wedelstedt, G. Goebel, and G. Kalies, *J. Chem. Inf. Model.*, 2022, **62**, 1154–1159. Copyright 2021 American Chemical Society.

I) visualizing molecular models or trajectories of atoms and molecules, or

II) performing and visualizing interactive molecular dynamics simulations.

(I) Some virtual reality applications for visualizing molecular models or trajectories of atoms and molecules have been developed. For example, Salvadori et al. [139] introduced *Caffeine* in 2016: a molecular viewer designed for virtual reality systems that allows to visualize static or dynamic molecular structures, molecular orbitals, and simulation data (using the desktop version of Caffeine) by reading Protein Databank (PDB), XYZ, or Gaussian Cube files. Caffeine was developed for CAVEs with three to six projection walls. Users can interact with the virtual system via a mobile device to translate, rotate, or scale molecular structures. Another example of a molecular viewer for virtual reality systems is *ProteinVR* which was developed by Cassidy et al. [140] in 2020. *ProteinVR* is a web application that allows to visualize static molecular models of proteins in virtual reality using HMDs. Due to the web-based approach of ProteinVR, no installation of programs or plugins other than the web browser itself is required. However, ProteinVR does not support the visualization of trajectories of completed molecular dynamics simulations.

(II) Virtual reality techniques have also been used to develop applications that allow performing and visualizing interactive molecular dynamics simulations in real-time in virtual reality. In 1998, Ai and Fröhlich [137] have introduced *RealMol*: a virtual environment for rational drug design. RealMol was specifically developed for the CAVE installed at Fraunhofer-IGD, which consisted of a 2.4 m $\times$ 2.4 m box with 5 projection areas (i.e. ceiling, floor, and three walls). In RealMol, a small number of users is able to enter the CAVE and examine or discuss molecular models that were read from PDB files. Active shutter glasses are used to generate three-dimensional images of the projected virtual environment, and cybergloves allow to interact with or modify molecular models. RealMol uses NAMD [141] as back-end simulation engine to allow the use of a (modified) molecular model in an energy minimization or molecular dynamics simulation. The molecular model is then updated and visualized with the information provided by NAMD. In addition, physical properties of the dynamic system are displayed, such as temperature or forces acting on each atom, and changes in the potential energy are indicated by sound effects. The pitch of these sound effects depends on the magnitude of the energy change. RealMol was one of the first programs to allow interactive molecular dynamics simulations in virtual reality. But the interactivity is restricted, because simulations need to be restarted after each modification of the molecular model. Also, CAVEs are cost-expensive fixed installations, strongly limiting the target audience of RealMol.

HMDs such as the HTC Vive or Occulus Rift have become popular in research and industry thanks to their increasing affordability. Even the unique selling point of CAVEs – i.e. the possibility of several people being simultaneously present in virtual reality – remains no longer untouched. In 2019, O'Connor et al. [125] introduced *Narupa*: a multiperson vir-

tual reality program for performing interactive molecular dynamics simulations in real-time. Narupa calculates and visualizes the trajectory of a system based on forces it receives by sending atom positions to a back-end simulation engine. A broad range of back-end simulation engines is supported, such as LAMMPS [111], GROMACS [142], or NAMD [141]. In addition, users can define biasing forces to interactively steer the trajectories of subunits of proteins or large molecules. Biasing forces are defined by selecting a subunit that will feel the biasing force and the position to which the subunit will be pulled, plus a scaling factor that controls the strength of the biasing force [143]. Among a broad range of scientific research applications, O'Connor et al. used Narupa to study molecular transport of small molecules in zeolites using DL_POLY [144] as back-end simulation engine. During the molecular dynamics simulation, Narupa allows to navigate the guest molecule through the channels of the zeolite or to inspect the zeolite by translation or rotation. Another program for performing interactive molecular dynamics simulations in virtual reality was developed by Juárez-Jiménez et al. [143] in 2020. Similar to Narupa, this program calculates and visualizes the trajectory of a system by receiving forces from a back-end simulation engine (i.e. GROMACS [142]) and allows to steer the trajectories of subunits of proteins or large molecules by applying biasing forces. In contrast to Narupa, the program developed by Juárez-Jiménez et al. allows to initiate molecular dynamics simulations (with or without biasing forces) from snapshots of trajectories of previously conducted molecular simulations. This approach helps to study protein folding, which can have time scales of several microseconds or longer. In combination with statistical methods such as Markov state modeling, Juárez-Jiménez et al. find that the initiation of molecular dynamics simulations from trajectory snapshots is well suited for the systematic description of a system.

The overview of the state of research on the application of virtual reality techniques in the field of molecular chemistry shows that there is still room for improvement or further application possibilities:

- Practically all applications for performing interactive molecular dynamics simulations in virtual reality depend on back-end simulation engines. But the use of back-end simulation engines complicates the widespread use of virtual reality applications, because it often requires access to high-performance computer clusters or a local installation of a simulation engine. Virtual reality applications would be accessible to a broader audience if molecular dynamics routines were directly implemented into the software.
- Most virtual reality applications have been developed specifically with the study of protein folding in mind, such as ProteinVR [140], RealMol [137], or Narupa [125]. No virtual reality application has yet been developed that allows for hypothetical MOFs to be constructed and tested in molecular dynamics simulations of guest molecules.

To fill these gaps, we developed MOF-VR: a virtual reality program for performing and visualizing interactive molecular dynamics simulations in MOFs. MOF-VR provides a number of features to support research, education, and science outreach, such as:

1. facilitating the understanding of the modular structure of MOFs by allowing to create hypothetical MOFs from MOF building blocks,

2. giving insights into the dynamics of guest molecule diffusion in MOFs by allowing to perform interactive molecular dynamics simulations, or

3. providing three dimensional representations of pre-computed or real-time simulated atomic or molecular trajectories.

5.3 Implementation

MOF-VR is an open-source code released under the FreeBSD license. It is developed for Microsoft Windows and HTC Vive and written in object-oriented C# using the Unity game engine (version 2020.1.7) and SteamVR Unity plugin (version 2.6.1). The source code, manual, and a compiled version of MOF-VR are provided on the CD-ROM that accompanies this dissertation. The reader is advised to test MOF-VR or consult the accompanying manual to complement the following descriptions. MOF-VR consists of three subroutines: a construction routine to create hypothetical MOFs by hand, a molecular dynamics routine to simulate guest molecule adsorption in nanoporous materials, and a visualization routine to visualize molecular trajectories. Before discussing these subroutines in detail, we will set the scene by explaining the basic terms and concepts of a game engine.

5.3.1 Basics: Game Engine

Game engines are at the heart of any computer game and provide core functions required to run a game, such as graphics rendering, collision detection, or physics calculations. Most game engines are component-based: objects in a game (GameObjects) acquire functionalities according to the components they own. For example, collider components enable collision detection (i.e. trigger colliders) or represent the physical shape of a GameObject, mesh renderer components contain information about the three-dimensional mesh of vertices and triangle-arrays to visualize the shape and coloring of a GameObject, and transform components store or modify the position, orientation, and scale of a GameObject. In addition, GameObjects can be parents of other GameObjects, and in this case, child GameObjects will move, rotate, and scale just like their parent GameObjects.

5.3.2 Subroutine 1: Construction Routine

MOF-VR's construction routine allows the creation of hypothetical MOFs that can be used in molecular dynamics simulations. For a hands-on demonstration, the interested reader is referred to the video trailer on the CD-ROM that accompanies this dissertation. The construction routine consists of two parts:

I) combining MOF building blocks to form non-periodic crystal structures, and

II) using these non-periodic crystal structures to form periodic MOF crystals.

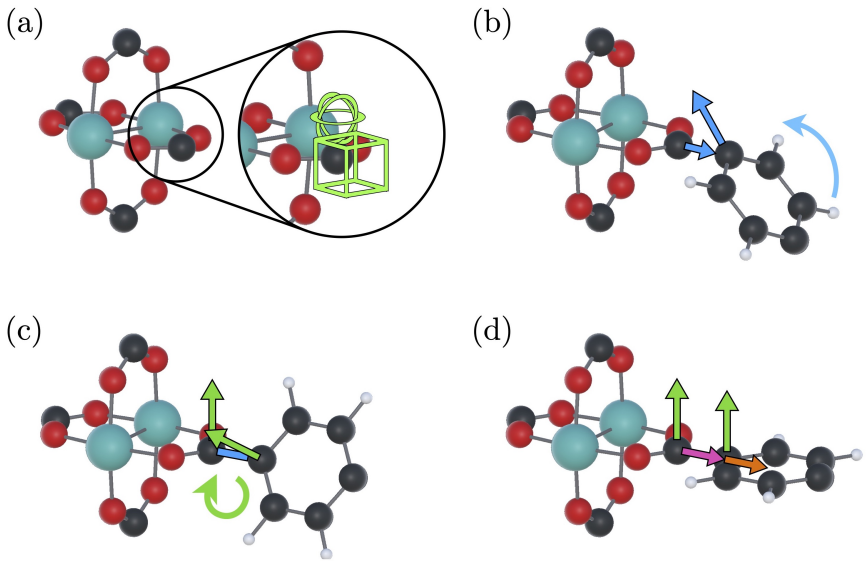

Fig. 41: Algorithm for recombining nodal and linking building blocks in MOF-VR. (a) Each connection site of a building block is equipped with two trigger colliders (green): a box(-shaped) collider to store information about the bond vector and a capsule(-shaped) collider to store information about the torsional orientation of the building block. (b) When the connection site of a nodal building block is approached by a connection site of a linking building block, the building block released from hand (here, the linker) is rotated so that the (blue) bond vectors are anti-parallel. (c) A torsional rotation around the bond vector is performed so that the (green) torsion vectors are either parallel or anti-parallel. (d) The new position of the released building block is determined by two vectors: a (purple) vector starting from the connection site of the non-released building block, parallel to its bond vector and having a magnitude equal to the typical (single) bond length between the connection site elements, and an (orange) vector starting from the connection site of the released building block and ending at its geometrical center.

(I) The algorithm implemented in MOF-VR for combining MOF building blocks into non-periodic crystal structures is illustrated in Fig. 41 and was inspired by the procedure for generating hypothetical MOFs described by Wilmer et al. [145], who decomposed experimental crystal structures of existing MOFs into nodal and linking building blocks and recombined them to form hypothetical MOFs using geometric rules. Nodal building blocks were defined as the inorganic nodes of MOFs including the terminal carboxylate groups of connected organic linkers, while linking building blocks were defined as organic molecules with terminal unsaturated nitrogen atoms or organic dicarboxylic acids without carboxylate groups. MOF-VR follows the concepts proposed by Wilmer et al. and uses similar geometric rules to allow recombination of nodal and linking building blocks in virtual reality. Several nodal and linking building blocks are available in MOF-VR (see Appendix Q) that were read from Crystallographic Information Files (CIF) provided in the Topologically Based Crystal Constructor (ToBaCCo) of Colón et al. [123, 146] or created from mfpx structure files provided at https://www.mofplus.org from Schmid and coworkers [147–149]. In MOF-VR, nodal and linking building blocks possess connection sites to which other building blocks can be connected. For nodal linking blocks, these connections site are either open

metal sites or carbon atoms in carboxylate groups, while for linking building blocks these connection sites are terminal unsaturated carbon or nitrogen atoms. Each connection site is equipped with a box(-shaped) and a capsule(-shaped) trigger collider (see Fig. 41a). While the vector from the connection site to the box collider (i.e. bond vector) stores information about the direction of the chemical bond formed by the connection with another building block, the vector from the connection site to the capsule collider (torsion vector) stores information about the torsional rotation of the connected building block about the (mutual) bond vector. Box colliders are furthermore responsible for detecting approaching connection sites of other building blocks. Two building blocks are combined if the connection site of a nodal building block is approached by the suitable connection site of a linking building block (and vice versa) and one of the building blocks is released from hand. The new position and orientation of the released building block is calculated and set in three steps:

1) The detached building block is rotated so that the bond vectors of the triggered connection sites are anti-parallel (see Fig 41b).

2) The released building block is rotated about the bond vector so that the torsion vectors of the triggered connection sites are either parallel or anti-parallel to ensure the correct torsional orientation of the released building block (see Fig 41c).

3) The position of the detached building block is determined with the help of two vectors: a vector starting from the triggered connection site of the non-detached building block, which is parallel to its bond vector and has a magnitude equal to the typical (single) bond length between the connection site elements (purple vector in Fig. 41d), and a vector starting from the connection site of the detached building block and ending at its geometrical center (orange vector in Fig 41d). The typical (single) bond length is calculated as the sum of the covalent radii of the connection site elements, following the work of Cordero et al. [150].

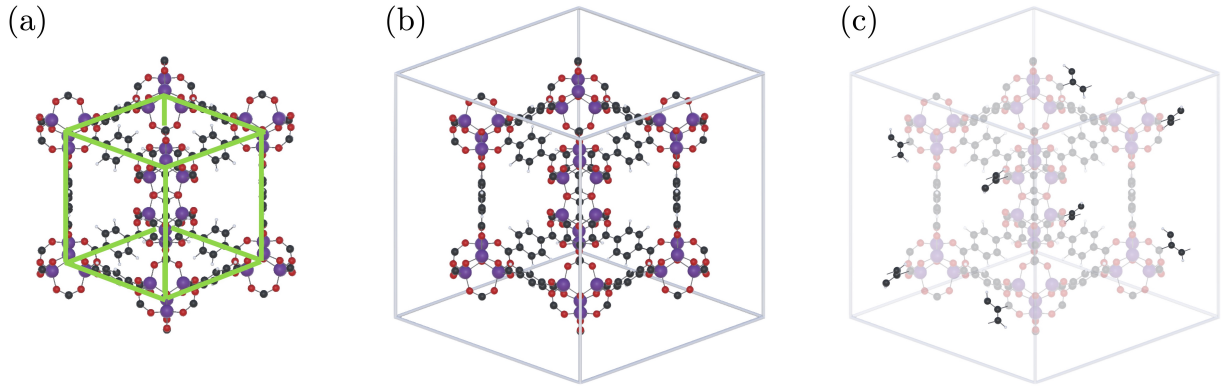

Fig. 42: Algorithm for creating periodic MOF crystals: (a) Closed non-periodic MOF crystals can be made periodic. The outermost nodal building blocks represent the corners of a parallelepiped (green lines) that has the same shape but a smaller scale than the parallelepiped that represents the periodic MOF crystal cell. (b) The cell faces and lattice vectors of the periodic MOF crystal are determined based on the parallelepiped in (a), and (c) nodal building blocks are connected across the periodic cell boundary.

The resulting MOF fragment can be further developed by adding nodal and linking building blocks, and thus created closed non-periodic crystal structures (see Fig. 42a) can in turn be used to form periodic MOF crystals. (II) The algorithm to convert a non-periodic MOF crystal to a periodic one is illustrated in Fig. 42 and consists in essence of three steps:

1) The cell faces of the periodic MOF crystal are determined. Geometrically, these cell faces form a parallelepiped, i.e. a three-dimensional figure that is limited by three pairs of parallel planes. These three pairs of planes are calculated in three steps:

    a) The eight outermost nodal building blocks of the non-periodic crystal structure are identified. These nodal building blocks represent the corners of a parallelepiped that has the same shape but a smaller scale than the parallelepiped which is formed by the cell faces of the periodic MOF crystal (see Fig. 42a).

    b) The faces of the scaled parallelepiped are determined from the positions of the eight outermost nodal building blocks (see Fig. 42a). These faces are represented by three pairs of parallel planes.

    c) The distance between the planes in each pair of planes is doubled. This enlarges the scaled parallelepiped to obtain the cell faces of the periodic MOF crystal (see Fig. 42b).

2) The lattice vectors of the periodic MOF crystal are determined (see Fig. 43). This is achieved in four steps:

    a) One of the three plane pairs is chosen.

    b) The normal vector of any of the two planes in the selected pair of planes is determined (see Fig. 43a).

    c) The determined normal vector is first projected onto any plane of the first remaining plane pair (see Fig. 43b) and then onto any plane of the second remaining plane pair (see Fig. 43c). The twice projected plane normal represents one of the three lattice vectors of the periodic MOF crystal.

    d) Steps a) to c) are repeated until all three lattice vectors are obtained. Note that no pair of planes is chosen twice in step a).

3) Unoccupied connection sites are filled with linking building blocks. Periodic boundary conditions are used to determine the correct positions of the atoms in the linking building blocks within the periodic MOF crystal cell (see Fig. 42c).

The final MOF is stored in an in-game cache and can be used in molecular dynamics simulations later on.

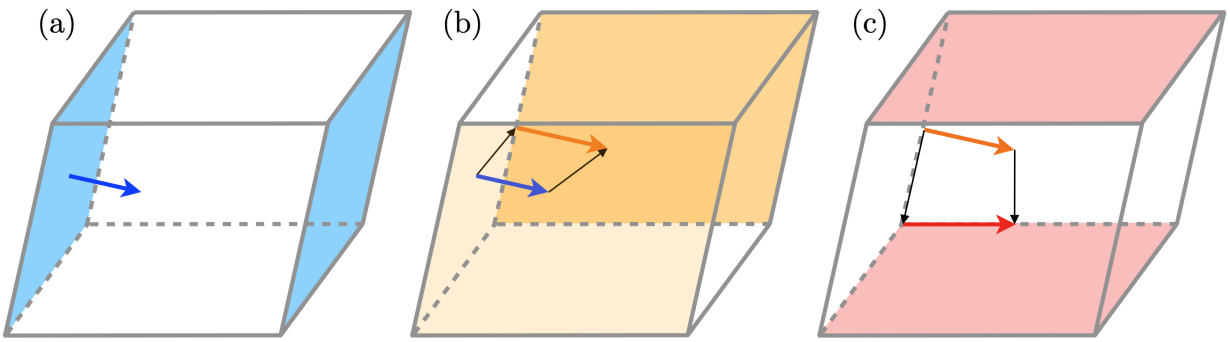

Fig. 43: Algorithm for determining the lattice vectors of the periodic MOF crystal. (a) One of the planes of the first pair of planes (blue faces) is chosen and its normal vector is determined (blue arrow). (b) The normal vector is projected onto one of the planes of the second pair of planes (orange faces), and (c) the projected (orange) vector is projected onto one of the planes of the third pair of planes (red faces). The resulting (red) vector corresponds to one of the three lattice vectors of the periodic MOF crystal.

5.3.3 Subroutine 2: Molecular Dynamics Routine

The molecular dynamics routine implemented in MOF-VR is designed to be intuitive and easily accessible for users with little to no knowledge of the theory of molecular dynamics simulations. For instance, the creation of guest definition or input files is optional, interaction parameters are automatically assigned to the interaction sites, and periodic boundary conditions are used in conjunction with the minimum image convention to mimic an infinite system. Guest molecules can be placed at any desired position or simply tossed into a MOF to start a molecular dynamics run. This design philosophy makes MOF-VR accessible to users who have little or no knowledge of the theory of molecular dynamics simulations. The molecular dynamics routine numerically integrates Newton's equations of motion using the velocity-Verlet algorithm (for fully and partly flexible guest molecules) or the rotational-velocity-Verlet algorithm [35] (for rigid guest molecules). The interatomic potential applied in the molecular dynamics routine consists of non-bonded, $\mathcal{U}_{\text{nb}}$, and bonded interactions, $\mathcal{U}_{\text{intra}}$,

$$\mathcal{U} = \mathcal{U}_{\text{nb}} + \mathcal{U}_{\text{intra}}. \tag{211}$$

Non-bonded interactions are represented by a sum of van-der-Waals, $\mathcal{U}_{\text{vdW}}$, and electrostatic interactions, $\mathcal{U}_{\text{elec}}$,

$$\mathcal{U}_{\text{nb}} = \mathcal{U}_{\text{vdW}} + \mathcal{U}_{\text{elec}}, \tag{212}$$

while bonded interactions are split into bond, $\mathcal{U}_{\text{bond}}$, bend, $\mathcal{U}_{\text{bend}}$, torsion, $\mathcal{U}_{\text{tor}}$, intramolecular van-der-Waals, $\mathcal{U}_{\text{intra,vdW}}$, and intramolecular electrostatic interactions, $\mathcal{U}_{\text{intra,elec}}$,

$$\mathcal{U}_{\text{intra}} = \mathcal{U}_{\text{bond}} + \mathcal{U}_{\text{bend}} + \mathcal{U}_{\text{tor}} + \mathcal{U}_{\text{intra,vdW}} + \mathcal{U}_{\text{intra,elec}}. \tag{213}$$

Van-der-Waals interactions are described by a Lennard-Jones potential. Lennard-Jones parameters for interaction sites of the crystal framework are automatically assigned from the Universal force field [48], while Lennard-Jones parameters for interaction sites in guest

molecules are read from the guest definition file. Lennard-Jones parameters between unlike atomic species i and j are computed by means of the Lorentz-Berthelot mixing rules [44, 45]. MOF-VR comes with a variety of guest molecule models, such as alkanes, alcohols, or small molecules. In addition, users can write their own guest definition files (see manual on the CD-ROM accompanying this work). Electrostatic interactions are described by a Coulomb potential. Coulomb forces are calculated using the Ewald summation [55] to account for the long-range nature of electrostatic interactions. Partial charges for the interaction sites of crystal structures or guest molecules are assigned in one of three ways:

1) partial charges are read from guest definition or CIF files,

2) partial charges for crystals read from CIF files are computed by the EQeq method of Wilmer et al. [53], or

3) partial charges for MOFs created via the construction routine are assigned using molecular building block-based (MBBB) charges as described by Argueta et al. [151] (see Appendix R for details).

Bond and bend interactions, $\mathcal{U}_{\text{bond}}$ or $\mathcal{U}_{\text{bend}}$, are described by harmonic potentials [152, 153], while torsion is described by the TraPPE dihedral potential [46]. Intramolecular van-der-Waals and intramolecular electrostatic interactions are computed using a scaled Lennard-Jones or Coulomb potential. Scaling factors are defined in the guest definition file (see manual on the CD-ROM that accompanies this dissertation). MOF-VR allows to keep bond lengths, bend angles, and/or dihedral angles fixed by using the RATTLE algorithm [36], or to simulate rigid guest molecules using the rotational-velocity-Verlet integration method developed by Rozmanov and Kusalik [35]. In addition to performing molecular dynamics simulations in the microcanonical ensemble, MOF-VR allows constant-temperature simulations using the CSVR thermostat [37, 38]. Moreover, MOF-VR prints the trajectories of guest molecules during simulation, calculates thermodynamic properties of the system (i.e. simulation time, kinetic energy, temperature, or mean-squared displacements (MSDs) of guest species), and provides auxiliary functions to increase intuitiveness and immersion (see Appendix S).

5.3.4 Subroutine 3: Visualization Routine

MOF-VR also offers the possibility of immersively watching molecular dynamics trajectories by reading PDB files (see Fig. 44). To ensure high graphics performance when visualizing molecular dynamics trajectories with tens of thousands of atoms in each frame, MOF-VR uses a technique called *Mesh Combining*. Meshes contain information about the vertices that define the triangles used to represent the shape of a visualized GameObject ([136], pp. 60–61). Each triangle must be rendered by the graphics engine, making visualization of parent GameObjects with many child GameObjects (such as MOFs with many framework atoms) CPU intensive. By combining mesh information from several child objects, MOF-VR reduces rendering time, thus ensuring frictionless visualization of molecular dynamics trajectories

with tens of thousands of GameObjects. Details on how to use MOF-VR's visualization routine can be found in the manual on the CD-ROM accompanying this dissertation.

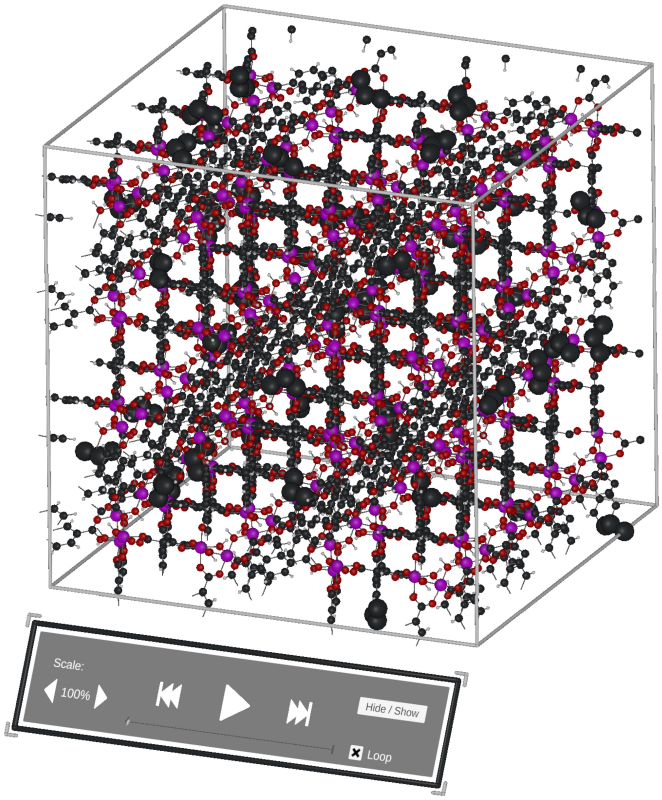

Fig. 44: Molecular dynamics trajectory of the adsorption of propane and propene in CAU-10 with over 10,000 GameObjects visualized in MOF-VR. The menu bar allows to start/stop playback, frame-by-frame visualization, or to hide certain GameObjects (e.g. MOF atoms).

5.4 Testing

Certain algorithms are at the heart of every molecular dynamics program, such as time integration schemes, force evaluation techniques, and thermostats. MOF-VR implements two time integration schemes (i.e. velocity-Verlet or rotational-velocity-Verlet algorithm [35]), four force evaluation routines (i.e. Lennard-Jones routine, Ewald summation [55], binding forces routine, or RATTLE algorithm [36]), and a thermostat (i.e. CSVR thermostat [37, 38]). The accurate implementation of these seven algorithms was tested.

5.4.1 Methods and Models

The accuracy of the molecular dynamics implementation was tested against the LAMMPS software package [111] developed at Sandia National Laboratories using verification simulations to individually confirm the functionality of the algorithms implemented in MOF-VR. The following verification simulations were used:

1) Velocity-Verlet Routine: Molecular dynamics simulation in the microcanonical ensemble of three methane molecules in MOF-5. Intermolecular forces were calculated based on Lennard-Jones interactions.

2) Rigid Body Routine: Molecular dynamics simulation in the microcanonical ensemble of a 2,3-dimethylbutane molecule in MOF-5. Intermolecular forces were calculated based on Lennard-Jones interactions, while intramolecular interactions were rigidified using the rotational-velocity-Verlet algorithm of Rozmanov and Kusalik [35] (MOF-VR) or the symplectic quaternion scheme of Miller et al. [154] (LAMMPS).

3) Electrostatics Routine: Molecular dynamics simulation in the microcanonical ensemble of a methanol molecule in MOF-5. Intermolecular forces were calculated based on Lennard-Jones and Coulomb interactions. Coulomb forces were calculated using the Ewald summation [55] with an Ewald splitting parameter of 0.25 $Å^{-1}$, a maximum of 6 Fourier space vectors in each spatial direction, and a real space cutoff of 12 Å. These Ewald parameters were calculated automatically by MOF-VR based on the desired relative cutoff error in the energy of about $1.37 \cdot 10^{-5}$. Intramolecular interactions were rigidified using the rotational-velocity-Verlet algorithm of Rozmanov and Kusalik [35] (MOF-VR) or the symplectic quaternion scheme of Miller et al. [154] (LAMMPS).

4) Constraints Routine: Molecular dynamics simulation in the microcanonical ensemble of a propane molecule in MOF-5. Intermolecular forces were calculated based on Lennard-Jones interactions, while intramolecular interactions were rigidified using the RATTLE algorithm [36] with a relative tolerance of the bond length of 10^{-4} and a maximum of 50 iterations.

5) Binding Forces Routine: Molecular dynamics simulation in the microcanonical ensemble of an n-butane molecule in MOF-5. Intermolecular forces were calculated based on Lennard-Jones interactions, while intramolecular interactions were described by harmonic potentials (for bond stretching and angle bending) and the TraPPE dihedral potential (for torsion). Bond constants were taken from the CHARMM22 force field [153] and bend or torsion constants were taken from the TraPPE-UA force field [152].

6) Thermostatting Routine: Molecular dynamics simulation in the canonical ensemble of a 2,3-dimethylbutane molecule in MOF-5 using the CSVR thermostat [37, 38] with a damping time of 100 fs to maintain the temperature of the system at 298.15 K. Intermolecular forces were calculated based on Lennard-Jones interactions, while intramolecular interactions were rigidified using the rotational-velocity-Verlet algorithm of Rozmanov and Kusalik [35] (MOF-VR) or the symplectic quaternion scheme of Miller et al. [154] (LAMMPS).

Molecular dynamics simulations were conducted for 3,000,000 time steps à 1 fs in a single MOF-5 unit cell, keeping the positions of the lattice atoms in MOF-5 fixed. Lennard-Jones interactions were calculated for all guest-MOF atom pairs in the simulation box (MOF-VR) or truncated by a spherical 12 Å cutoff (LAMMPS) using periodic boundary conditions in conjunction with the minimum image convention. Different truncation scheme are used, because the combination of truncation errors and single point precision floating-point numbers

used in Unity lead to strong error accumulation. Guest molecules were modeled according to the TraPPE force fields developed by Siepmann and coworkers [46, 77, 152] (modifications are stated above). The crystal structure of MOF-5 was taken from previous work [74]. Lennard-Jones parameters for each atom of MOF-5 were taken from the Universal force field [48], while partial charges were determined using the EQeq method of Wilmer et al. [53]. In each verification simulation, the initial state of the system was created in the following way: the game was paused, the orientation of MOF-5 and the guest molecules were set to their identity rotation (corresponding to no rotation), and the guest molecules were placed in the center of MOF-5. When the game continued, the positions of the interaction sites were written to a file that formed the basis for the LAMMPS data file, and the velocities of the interaction sites were set to zero. According to classical mechanics, the motion of particles is clearly defined by their positions and momenta. For this reason, the functionality of the simulation routines implemented in MOF-VR is tested by calculating kinetic energies (which contain information on the interaction site momenta) and radial distribution functions (which contain information on the interaction site positions) for each of the above verification simulations [155]. Average kinetic energies were calculated every 1,000 time steps based on the kinetic energies of the previous 1,000 time steps, while radial distribution functions were calculated at the end of each verification simulation and averaged over the previous 100,000 time steps.

5.4.2 Results and Discussion

Fig. 45 shows the average kinetic energies calculated for the verification simulations performed with MOF-VR or LAMMPS. In addition, the mean and the statistical error (SE) of the kinetic energies were computed using the block-averaging technique ([26], pp. 525–527). For this, the average kinetic energies of the final 1,000,000 time steps were divided into 5 blocks à 200,000 time steps. The statistical error was calculated as the standard deviation of the block averages. The results can be summarized as follows:

1) Velocity-Verlet Routine: MOF-VR calculates an average kinetic energy of 11.29 kJ/mol (SE: 0.48 kJ/mol) and LAMMPS of 11.73 kJ/mol (SE: 0.54 kJ/mol). The average kinetic energy calculated with MOF-VR is within the statistical error of the results obtained with LAMMPS (see Fig. 45a).

2) Rigid Body Routine: MOF-VR calculates an average kinetic energy of 13.42 kJ/mol (SE: 0.20 kJ/mol) and LAMMPS of 13.38 kJ/mol (SE: 0.35 kJ/mol). The average kinetic energy calculated with MOF-VR is within the statistical error of the results obtained with LAMMPS (see Fig. 45b).

3) Electrostatics Routine: MOF-VR calculates an average kinetic energy of 9.21 kJ/mol (SE: 0.31 kJ/mol) and LAMMPS of 9.72 kJ/mol (SE: 0.38 kJ/mol). The average kinetic energy calculated with MOF-VR is not within the statistical error of the results obtained with LAMMPS (see Fig. 45c). Further investigations show that this deviation is due to a perturbation of the system caused by round-off errors in the position of the

guest molecule when exporting the system from MOF-VR to LAMMPS (see Method section for details). To prove this, we have repeated the simulation in LAMMPS starting from a perturbed starting configuration by shifting the geometric center of methanol by -10^{-8} Å in every spatial direction. This shift corresponds to the round-off error. For the repeated simulation, LAMMPS computes an average kinetic energy of 10.09 kJ/mol (SE: 0.48 kJ/mol). Apparently, the kinetic energy of the system is sensitive to small deviations in the initial configuration.

4) Constraints Routine: MOF-VR calculates an average kinetic energy of 9.24 kJ/mol (SE: 0.28 kJ/mol) and LAMMPS of 9.69 kJ/mol (SE: 0.21 kJ/mol). The average kinetic energy calculated with MOF-VR is not within the statistical error of the results obtained with LAMMPS (see Fig. 45d). Again, this deviation can be attributed to a perturbation of the system caused by round-off errors in the position of the guest molecule when exporting the system from MOF-VR to LAMMPS. To prove this, we have repeated the simulation in LAMMPS starting from a perturbed starting configuration by shifting the geometric center of propane by -10^{-8} Å in every spatial direction. For the repeated simulation, LAMMPS computes an average kinetic energy of 9.94 kJ/mol (SE: 0.15 kJ/mol). Apparently, the kinetic energy of the system is sensitive to small deviations in the initial configuration.

5) Binding Forces Routine: MOF-VR calculates an average kinetic energy of 11.12 kJ/mol (SE: 0.10 kJ/mol) and LAMMPS of 11.11 kJ/mol (SE: 0.15 kJ/mol). The average kinetic energy calculated with MOF-VR is within the statistical error of the results obtained with LAMMPS (see Fig. 45e). It should be recalled, that the compared kinetic energies are averaged over the final 1 ns of the simulation. The deviation of the kinetic energies calculated by MOF-VR and LAMMPS for the simulation time between about 0.5 ns and 2 ns can be attributed to a perturbation of the system caused by round-off errors in the position of the guest molecule when exporting the system from MOF-VR to LAMMPS. To prove this, we have repeated the simulation in LAMMPS starting from a perturbed starting configuration by by shifting the geometric center of butane by -10^{-8} Å in every spatial direction. The results are shown in Fig. T1 of Appendix T and indicate similar temporary fluctuations. Apparently, the system has several preferred kinetic energy states between which it alternates.

6) Thermostatting Routine: In contrast to MOF-VR, LAMMPS only thermostats translational degrees of freedom (e.g. see: `docs.lammps.org/fix_temp_csvr.html`). That's why Fig. 45f compares the kinetic energy calculated by MOF-VR with the *translational* kinetic energy calculated by LAMMPS. MOF-VR calculates an average kinetic energy of 3.71 kJ/mol (SE: 0.08 kJ/mol) corresponding to an average temperature of 297.98 K and LAMMPS of 3.74 kJ/mol (SE: 0.04 kJ/mol) corresponding to an average temperature of 299.61 K. For LAMMPS, the average temperature is calculated from the average *translational* kinetic energy. The average kinetic energy calculated with

MOF-VR is within the statistical error of the results obtained with LAMMPS.

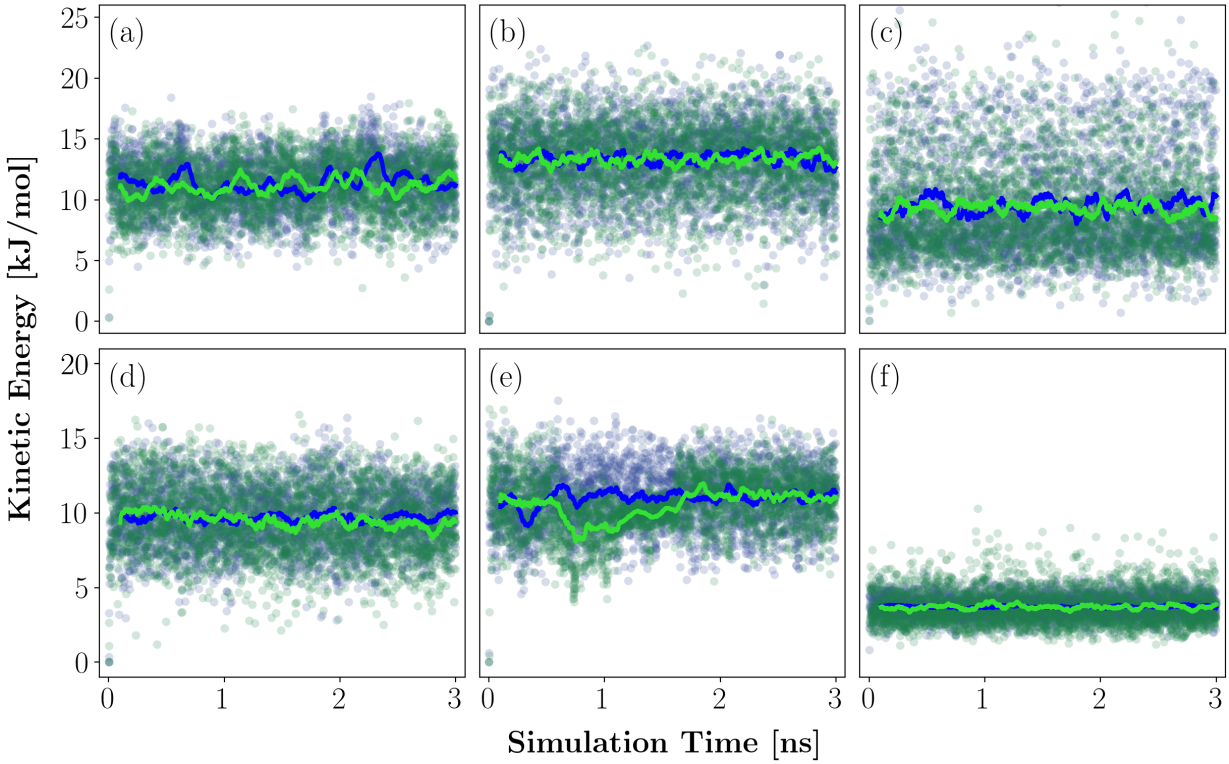

Fig. 45: Kinetic energies averaged over 1,000 time steps (dots) and their moving averages over 200 time steps (lines) calculated by MOF-VR (green) or LAMMPS (blue) for the verification simulations to test (a) the velocity-Verlet routine, (b) the rigid body routine, (c) the electrostatics routine, (d) the constraints routine, (e) the binding forces routine, and (f) the thermostatting routine.

Fig. 46 shows radial distribution functions (RDFs) calculated by MOF-VR or LAMMPS for atom pairs consisting of the carbon atoms of MOF-5 and interaction sites in guest molecules that are used in each verification simulation. RDFs for the remaining atom pairs can be found in Appendix U. The calculated RDFs can be summarize as follows:

1) **Velocity-Verlet Routine:** The RDFs shown in Fig. 46a and calculated for atom pairs consisting of carbon atoms in MOF-5 and CH_4 interaction sites in methane agree very well. The maximum probability of finding a CH_4 interaction site is at a distance of 4.5 Å or 11.5 Å to a carbon atom in MOF-5.

2) **Rigid Body Routine:** The RDFs shown in Fig. 46b and calculated for atom pairs consisting of carbon atoms in MOF-5 and CH_3 interaction sites in 2,3-dimethylbutane agree very well. The maximum probability of finding a CH_3 interaction site is at a distance of 6 Å or 12 Å to a carbon atom in MOF-5.

3) **Electrostatics Routine:** The RDFs shown in Fig. 46c and calculated for atom pairs consisting of carbon atoms in MOF-5 and CH_3 interaction sites in methanol agree very well. The maximum probability of finding a CH_3 interaction site is at a distance of 4.7 Å or 11.8 Å to a carbon atom in MOF-5.

4) **Constraints Routine:** The RDFs shown in Fig. 46d and calculated for atom pairs consisting of carbon atoms in MOF-5 and CH_3 interaction sites in propane agree very well. The maximum probability of finding a CH_3 interaction site is at a distance of 5.3 Å or 11.6 Å to a carbon atom in MOF-5.

5) **Binding Forces Routine:** The RDFs shown in Fig. 46e and calculated for atom pairs consisting of carbon atoms in MOF-5 and CH_3 interaction sites in n-butane agree very well. The maximum probability of finding a CH_3 interaction site is at a distance of 4.4 Å or 11.4 Å to a carbon atom in MOF-5.

6) **Thermostatting Routine:** The RDFs shown in Fig. 46f and calculated for atom pairs consisting of carbon atoms in MOF-5 and CH_3 interaction sites in rigid 2,3-dimethylbutane differ significantly. According to the RDF calculated by MOF-VR, the maximum probability of finding a CH_3 interaction site is at a distance of 4.2 Å or 12 Å to a carbon atom in MOF-5, while LAMMPS suggests that the maximum probability is at a distance of 6 Å or 12 Å. These differences are caused by the fact that LAMMPS only thermostats the translational degrees of freedom of rigid bodies. When using a fully flexible model for 2,3-dimethylbutane in LAMMPS, the calculated RDFs agree well with the one calculated for the rigid model by MOF-VR (see Fig. U7 of Appendix U).

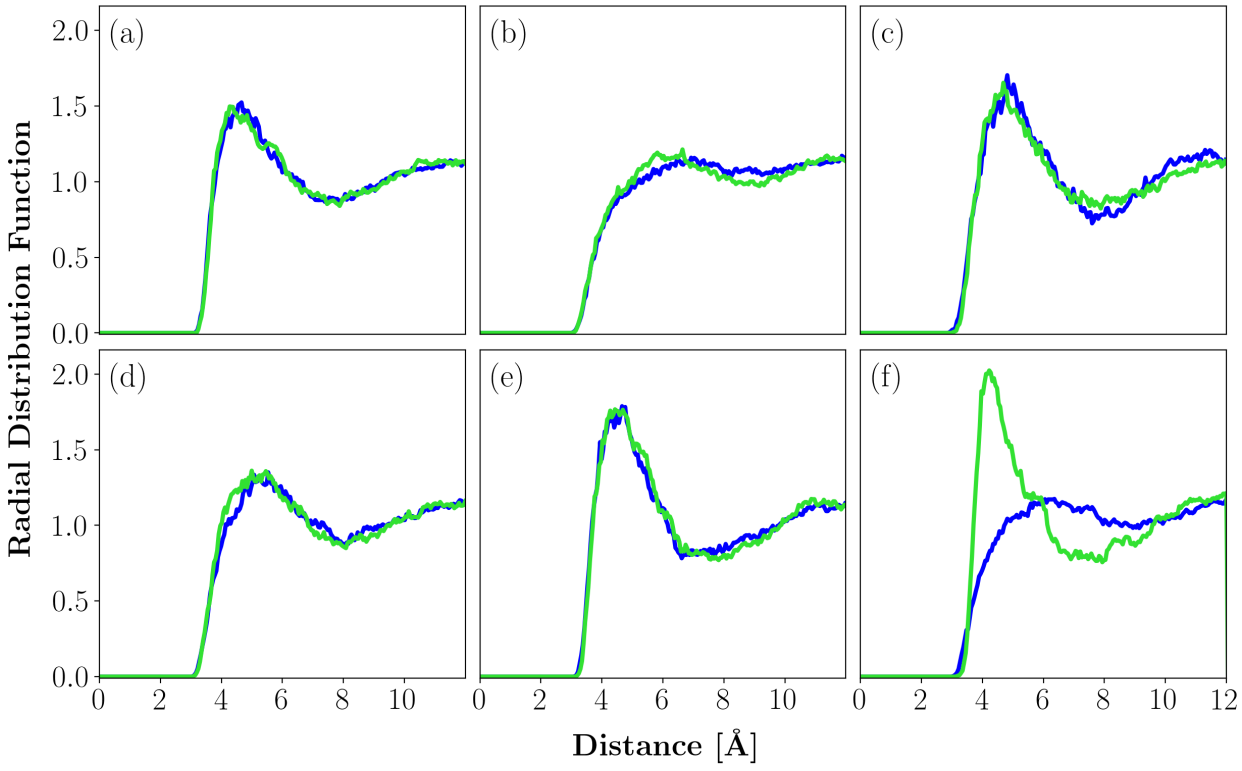

Fig. 46: Comparison of radial distribution functions generated by MOF-VR (green) or LAMMPS (blue) between carbon atoms in MOF-5 and (a) CH_4 in methane for the verification of the velocity-Verlet routine, (b) CH_3 in 2,3-dimethylbutane for the verification of the rigid body routine, (c) CH_3 in methanol for the verification of the electrostatics routine, (d) CH_3 in propane for the verification of the constraints routine, (e) CH_3 in n-butane for the verification of the binding forces routine, and (f) CH_3 in 2,3-dimethylbutane for the verification of the thermostatting routine.

5.5 Conclusions

MOF-VR was introduced: the first virtual reality program that allows hypothetical metal-organic frameworks to be created and tested in molecular dynamics simulations of guest molecules. We eliminated the need of a back-end simulation engine to perform molecular dynamics simulations in virtual reality by directly implementing molecular dynamics routines into MOF-VR using the Unity game engine and SteamVR system. Tests against the LAMMPS software package confirmed that MOF-VR is capable of performing molecular dynamics simulations with long-term energy stability. The computing resources of the average user will only allow the simulation of a small number of guest molecules in metal-organic frameworks in MOF-VR. But technological progress could give rise to virtual reality applications that efficiently implement molecular dynamics routines to allow users with little to no knowledge of the theory of molecular dynamics simulations to perform large-scale molecular dynamics simulations in virtual reality.

6 Summary

The aim of this dissertation was to shed light on a few surface phenomena occurring in or on metal-organic frameworks. In Section 3, we studied the separation of aqueous alcohol mixtures in CAU-10 – a metal-organic framework with internal hydrophobicity – using grand canonical Monte Carlo simulations. Separating alcohols from aqueous mixtures, such as fermentation broths, is an important cornerstone in the industrial transformation toward green chemistry. However, separation is often driven by entropic effects that favor the adsorption of the component with the higher saturation capacity (i.e. water), because adsorption from the liquid phase commonly occurs under saturation conditions. The aim of our study was to investigate whether entropic effects also favor the adsorption of water in metal-organic frameworks with internal hydrophobicity. The most important conclusion to be drawn from our study is that the adsorption selectivity is in favor of water for all investigated mixtures and liquid phase compositions, suggesting that the internal hydrophobicity of CAU-10 is of little importance. The preferential adsorption of water is due to energetic as well as entropic effects. For high alcohol concentrations, the hydroxyl groups of adsorbed alcohol molecules constitute strong hydrophilic adsorption sites for water adsorption (energetic effect), while with decreasing liquid phase mole fractions of the alcohol, the adsorption of water is preferred due to the higher packing efficiency of water in the channels of CAU-10 (entropic effect). The results indicate that competitive water adsorption significantly reduces the efficiency with which alcohols can be separated from the aqueous liquid phase. A viable solution may be to design the external surface of metal-organic frameworks in a way that it prevents the adsorption of water, for instance by providing an external ultrahydrophobicity.

The outlook obtained from the study in Section 3 emphasizes the importance of a clear distinction between the internal and external surface of metal-organic frameworks. However, this distinction is rarely made in the determination of the hydrophobicity of metal-organic frameworks based on the unstated assumption that the two commonly used methods, water adsorption experiments (internal surface) and contact angle measurement (external surface), provide similar results. In Section 4, we conducted molecular dynamics simulations to study how the design of the external surface influences the droplet spreading, and grand canonical Monte Carlo simulations were performed to predict water adsorption on the internal metal-organic framework surface. We can draw two conclusions from the simulation results. First, the external surface nature of metal-organic frameworks has a major impact on the droplet spreading – suggesting the necessity to clearly distinguish between the internal and external surface or hydrophobicity of metal-organic frameworks. Second, comparison of the vapor phase pressures above the curved droplet interface obtained in the simulation and the vapor pressure of water above a flat interface indicates that droplets on the surfaces of metal-organic frameworks are commonly trapped in a metastable state – a circumstance for which experimental evidence exists (see discussion on p. 90).

The studies conducted in Section 3 and 4 have demonstrated that molecular simulations are helpful tools for studying surface phenomena. However, the interpretation of the generated three-dimensional trajectories is often difficult, because most visualization tools only allow two-dimensional projections. The emergence of immersive visualization techniques such as virtual reality could greatly simplify the interpretation of atomic or molecular trajectories in molecular dynamics simulations. In Section 5, we introduced MOF-VR, a virtual reality tool developed to study guest molecules in metal-organic frameworks. MOF-VR consists of three subroutines: a construction routine to create hypothetical metal-organic frameworks by hand, a molecular dynamics suite, and a trajectory visualizer. Tests against the LAMMPS software package have shown that MOF-VR is capable of performing state-of-the-art molecular dynamics simulations of guest molecules in rigid metal-organic frameworks in virtual reality and provides reliable simulation results.

References

(1) O. Coussy, *Mechanics and physics of porous solids*, Wiley, Chichester, 2010.

(2) F. Rouquerol, J. Rouquerol and K. S. W. Sing, *Adsorption by powders and porous solids: principles, methodology, and applications*, Academic Press, San Diego, 1999.

(3) M. Thommes, K. Kaneko, A. V. Neimark, J. P. Olivier, F. Rodriguez-Reinoso, J. Rouquerol and K. S. Sing, *Pure Appl. Chem.*, 2015, **87**, 1051–1069.

(4) M. J. Schwuger and G. H. Findenegg, *Lehrbuch der Grenzflächenchemie: 10 Tabellen*, Thieme, Stuttgart, 1996.

(5) H. Reinsch, M. A. van der Veen, B. Gil, B. Marszalek, T. Verbiest, D. de Vos and N. Stock, *Chem. Mater.*, 2013, **25**, 17–26.

(6) L.-L. Li, H. Sun, C.-J. Fang, J. Xu, J.-Y. Jin and C.-H. Yan, *J. Mater. Chem.*, 2007, **17**, 4492–4498.

(7) A. Cadiau, J. S. Lee, D. Damasceno Borges, P. Fabry, T. Devic, M. T. Wharmby, C. Martineau, D. Foucher, F. Taulelle, C.-H. Jun, Y. K. Hwang, N. Stock, M. F. De Lange, F. Kapteijn, J. Gascon, G. Maurin, J.-S. Chang and C. Serre, *Adv. Mater.*, 2015, **27**, 4775–4780.

(8) D. H. Everett, *Pure and Applied Chemistry*, 1986, **58**, 967–984.

(9) L. G. Nagy and G. Schay, *Acta Chim. Hung. Tomus*, 1963, **39**, 366–390.

(10) G. Kalies, P. Bräuer and R. Rockmann, *Adsorpt. Sci. Technol.*, 2007, **25**, 493–501.

(11) K.-Y. Law and H. Zhao, *Surface wetting: characterization, contact angle, and fundamentals*, Springer, Cham, 2016.

(12) S. Safran, *Statistical thermodynamics of surfaces, interfaces, and membranes*, CRC Press, Boca Raton, 2019.

(13) T. Young, *Phil. Trans. R. Soc.*, 1805, **95**, 65–87.

(14) A. Marmur, *Soft Matter*, 2006, **2**, 12–17.

(15) R. N. Wenzel, *Ind. Eng. Chem.*, 1936, **28**, 988–994.

(16) A. B. D. Cassie, *Discuss. Faraday Soc.*, 1948, **3**, 11–16.

(17) A. B. D. Cassie and S. Baxter, *Trans. Faraday Soc.*, 1944, **40**, 546–551.

(18) G. H. Findenegg and T. Hellweg, *Statistische Thermodynamik*, Springer Spektrum, Berlin, 2nd edition, 2015.

(19) K. A. Dill and S. Bromberg, *Molecular driving forces: statistical thermodynamics in biology, chemistry, physics, and nanoscience*, Garland Science, London, 2nd edition, 2011.

(20) J. W. Gibbs, *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics*, Charles Scribner's Sons, London, 1902.

(21) G. Raabe, *Molecular Simulation Studies on Thermophysical Properties*, Springer Singapore, Singapore, 2017.

(22) D. A. McQuarrie, *Statistical mechanics*, Harper & Row, New York, 1975.

(23) D. Chandler, *Introduction to modern statistical mechanics*, Oxford University Press, New York, 1987.

(24) N. Metropolis, *Los Alamos Science (Special Issue)*, 1987, **15**, 125–130.

(25) W. K. Hastings, *Biometrika*, 1970, **57**, 97–109.

(26) D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, Academic Press, San Diego, 2nd edition, 2002.

(27) W. Shi and E. J. Maginn, *J. Chem. Theory Comput.*, 2007, **3**, 1451–1463.

(28) M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*, Clarendon Press; Oxford University Press, Oxford, 1987.

(29) A. R. Leach, *Molecular modelling: principles and applications*, Prentice Hall, Harlow, 2nd edition, 2001.

(30) M. Soroush Barhaghi, K. Torabi, Y. Nejahi, L. Schwiebert and J. J. Potoff, *J. Chem. Phys.*, 2018, **149**, 072318.

(31) D. Dubbeldam, S. Calero, D. E. Ellis and R. Q. Snurr, *Mol. Simul.*, 2016, **42**, 81–101.

(32) D. Dubbeldam, A. Torres-Knoop and K. S. Walton, *Mol. Simul.*, 2013, **39**, 1253–1292.

(33) R. Hens, A. Rahbari, S. Caro-Ortiz, N. Dawass, M. Erdős, A. Poursaeidesfahani, H. S. Salehi, A. T. Celebi, M. Ramdin, O. A. Moultos, D. Dubbeldam and T. J. H. Vlugt, *J. Chem. Inf. Model.*, 2020, **60**, 2678–2682.

(34) A. Torres-Knoop, S. P. Balaji, T. J. H. Vlugt and D. Dubbeldam, *J. Chem. Theory Comput.*, 2014, **10**, 942–952.

(35) D. Rozmanov and P. G. Kusalik, *Phys. Rev. E*, 2010, **81**, 056706.

(36) H. C. Andersen, *J. Comput. Phys.*, 1983, **52**, 24–34.

(37) G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.

(38) G. Bussi and M. Parrinello, *Comput. Phys. Commun.*, 2008, **179**, 26–29.

(39) S. Nosé, *J. Chem. Phys.*, 1984, **81**, 511–519.

(40) W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.

(41) E. Braun, S. M. Moosavi and B. Smit, *J. Chem. Theory Comput.*, 2018, **14**, 5262–5272.

(42) B. L. Eggimann, A. J. Sunnarborg, H. D. Stern, A. P. Bliss and J. I. Siepmann, *Mol. Simul.*, 2014, **40**, 101–105.

(43) B. L. Eggimann, Y. Sun, R. F. DeJaco, R. Singh, M. Ahsan, T. R. Josephson and J. I. Siepmann, *J. Chem. Eng. Data*, 2020, **65**, 1330–1344.

(44) H. A. Lorentz, *Ann. Phys.*, 1881, **248**, 127–136.

(45) D. Berthelot, *C. R. Acad. Sci.*, 1898, **126**, 1703–1855.

(46) M. G. Martin and J. I. Siepmann, *J. Phys. Chem. B*, 1999, **103**, 4508–4517.

(47) H. Zhang, B. J. Bucior and R. Q. Snurr, in *Modelling and Simulation in the Science of Micro- and Meso-Porous Materials*, ed. C. R. A. Catlow, V. V. Speybroeck and R. A. v. Santen, Elsevier, 2018, pp. 99–149.

(48) A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.

(49) S. L. Mayo, B. D. Olafson and W. A. Goddard, *J. Phys. Chem.*, 1990, **94**, 8897–8909.

(50) R. S. Mulliken, *J. Chem. Phys.*, 1955, **23**, 1833–1840.

(51) T. A. Manz and D. S. Sholl, *J. Chem. Theory Comput.*, 2010, **6**, 2455–2468.

(52) A. K. Rappe and W. A. Goddard, *J. Phys. Chem.*, 1991, **95**, 3358–3363.

(53) C. E. Wilmer, K. C. Kim and R. Q. Snurr, *J. Phys. Chem. Lett.*, 2012, **3**, 2506–2511.

(54) T. A. Manz and N. G. Limas, *arXiv:1512.08270 [cond-mat, physics:physics]*, 2015, arXiv: 1512.08270.

(55) P. P. Ewald, *Ann. Phys.*, 1921, **369**, 253–287.

(56) A. von Wedelstedt, G. Goebel and G. Kalies, *J. Chem. Inf. Model.*, 2022, **62**, 1154–1159.

(57) I. Patraşcu, C. S. Bîldea and A. A. Kiss, *Sep. Purif. Technol.*, 2017, **177**, 49–61.

(58) N. Abdehagh, F. H. Tezel and J. Thibault, *Biomass and Bioenergy*, 2014, **60**, 222–246.

(59) D. Shah, K. Kissick, A. Ghorpade, R. Hannah and D. Bhattacharyya, *J. Membr. Sci.*, 2000, **179**, 185–205.

(60) R. Krishna and J. M. van Baten, *Sep. Purif. Technol.*, 2011, **76**, 325–330.

(61) A. Torres-Knoop, J. Heinen, R. Krishna and D. Dubbeldam, *Langmuir*, 2015, **31**, 3771–3778.

(62) M. Schenk, S. L. Vidal, T. J. H. Vlugt, B. Smit and R. Krishna, *Langmuir*, 2001, **17**, 1558–1570.

(63) R. Krishna, *Sep. Purif. Technol.*, 2019, **215**, 227–241.

(64) M. Dávila, J. Riccardo and A. Ramirez-Pastor, *Chem. Phys. Lett.*, 2009, **477**, 402–405.

(65) K. Zhang, R. P. Lively, M. E. Dose, A. J. Brown, C. Zhang, J. Chung, S. Nair, W. J. Koros and R. R. Chance, *Chem. Commun.*, 2013, **49**, 3245–3247.

(66) A. L. Myers and J. M. Prausnitz, *AIChE J.*, 1965, **11**, 121–127.

(67) A. von Wedelstedt, *Investigation of the Separation of Hexane Isomers in a Novel Aluminum-based Metal-Organic Framework using Molecular Monte Carlo Methods*, Master Thesis, HTW University of Applied Sciences Dresden, 2018.

(68) P. F. Brântuas, A. Henrique, M. Wahiduzzaman, A. Wedelstedt, T. Maity, A. E. Rodrigues, F. Nouar, U.-H. Lee, K.-H. Cho, G. Maurin, J. A. C. Silva and C. Serre, *Advanced Science*, 2022, 2201494.

(69) D. Dubbeldam, S. Calero and T. J. Vlugt, *Mol. Simul.*, 2018, **44**, 653–676.

(70) D. Damasceno Borges, P. Normand, A. Permiakova, R. Babarao, N. Heymans, D. S. Galvao, C. Serre, G. De Weireld and G. Maurin, *J. Phys. Chem. C*, 2017, **121**, 26822–26832.

(71) D.-Y. Peng and D. B. Robinson, *Ind. Eng. Chem. Fund.*, 1976, **15**, 59–64.

(72) S. Chempath, R. Q. Snurr and J. J. Low, *AIChE J.*, 2004, **50**, 463–469.

(73) U. Weidlich and J. Gmehling, *Ind. Eng. Chem. Res.*, 1987, **26**, 1372–1381.

(74) A. von Wedelstedt, H. Chen, G. Kalies and R. Q. Snurr, *Langmuir*, 2020, **36**, 13070–13078.

(75) I. V. Grenev, A. A. Shubin, M. V. Solovyeva and L. G. Gordeeva, *Phys. Chem. Chem. Phys.*, 2021, **23**, 21329–21337.

(76) J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.

(77) B. Chen, J. J. Potoff and J. I. Siepmann, *J. Phys. Chem. B*, 2001, **105**, 3093–3104.

(78) C. Vega, J. L. F. Abascal and I. Nezbeda, *J. Chem. Phys.*, 2006, **125**, 034503.

(79) P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt and SciPy 1.0 Contributors, *Nat. Methods*, 2020, **17**, 261–272.

(80) J. J. Moré, in *Numerical Analysis*, ed. G. A. Watson, Series Title: Lecture Notes in Mathematics, Springer, Berlin, 1978, vol. 630, pp. 105–116.

(81) W. M. Haynes, D. R. Lide and T. J. Bruno, *CRC handbook of chemistry and physics: a ready-reference book of chemical and physical data.* CRC Press, Boca Raton, 97th edition, 2017.

(82) C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke and T. E. Oliphant, *Nature*, 2020, **585**, 357–362.

(83) C. M. Simon, B. Smit and M. Haranczyk, *Comput. Phys. Commun.*, 2016, **200**, 364–380.

(84) B. Claessens, A. Martin-Calvo, J. J. Gutiérrez-Sevillano, N. Dubois, J. F. Denayer and J. Cousin-Saint-Remi, *Langmuir*, 2019, **35**, 3887–3896.

(85) E. Catmull and R. Rom, in *Computer Aided Geometric Design*, ed. R. E. Barnhill and R. F. Riesenfeld, Academic Press, New York, 1974, pp. 317–326.

(86) M. Klauck, J. Guhlmann, T. Hähnel, M. Hauser and G. Kalies, *Adsorption*, 2022.

(87) H. Furukawa, K. E. Cordova, M. O'Keeffe and O. M. Yaghi, *Science*, 2013, **341**, 1230444.

(88) J. Canivet, A. Fateeva, Y. Guo, B. Coasne and D. Farrusseng, *Chem. Soc. Rev.*, 2014, **43**, 5594–5617.

(89) A. Dhakshinamoorthy, A. M. Asiri and H. Garcia, *Chem. Commun.*, 2014, **50**, 12800–12814.

(90) L.-H. Xie, X.-M. Liu, T. He and J.-R. Li, *Chem*, 2018, **4**, 1911–1927.

(91) K. Jayaramulu, F. Geyer, A. Schneemann, Š. Kment, M. Otyepka, R. Zboril, D. Vollmer and R. A. Fischer, *Adv. Mater.*, 2019, 1900820.

(92) D. Lv, H. Wang, Y. Chen, F. Xu, R. Shi, Z. Liu, X. Wang, S. J. Teat, Q. Xia, Z. Li and J. Li, *ACS Appl. Mater. Interfaces*, 2018, **10**, 6031–6038.

(93) P. Hu, J. Han, J. Zhou, H. Wang, C. Xiong, H. Liu, X. Zhou, Y. Wang and H. Ji, *Chem. Eng. J.*, 2021, **426**, 131302.

(94) L. Gan, A. Chidambaram, P. G. Fonquernie, M. E. Light, D. Choquesillo-Lazarte, H. Huang, E. Solano, J. Fraile, C. Viñas, F. Teixidor, J. A. R. Navarro, K. C. Stylianou and J. G. Planas, *J. Am. Chem. Soc.*, 2020, **142**, 8299–8311.

(95) Y. Liu and C.-H. Choi, *Colloid Polym. Sci.*, 2013, **291**, 437–445.

(96) E. L. Decker and S. Garoff, *Langmuir*, 1996, **12**, 2100–2110.

(97) S. Jung, Y. Kim, S.-J. Kim, T.-H. Kwon, S. Huh and S. Park, *Chem. Commun.*, 2011, **47**, 2904.

(98) H. Li, M. Eddaoudi, M. O'Keeffe and O. M. Yaghi, *Nature*, 1999, **402**, 276–279.

(99) J. L. C. Rowsell and O. M. Yaghi, *Angew. Chem. Int. Ed.*, 2005, **44**, 4670–4679.

(100) J. G. Nguyen and S. M. Cohen, *J. Am. Chem. Soc.*, 2010, **132**, 4560–4561.

(101) Y. Ming, J. Purewal, J. Yang, C. Xu, R. Soltis, J. Warner, M. Veenstra, M. Gaab, U. Müller and D. J. Siegel, *Langmuir*, 2015, **31**, 4988–4995.

(102) L. Huang, H. Wang, J. Chen, Z. Wang, J. Sun, D. Zhao and Y. Yan, *Microporous and Mesoporous Mat.*, 2003, **58**, 105–114.

(103) D. Lenzen, P. Bendix, H. Reinsch, D. Fröhlich, H. Kummer, M. Möllers, P. P. C. Hügenell, R. Gläser, S. Henninger and N. Stock, *Adv. Mater.*, 2018, **30**, 1705869.

(104) D. Fröhlich, E. Pantatosaki, P. D. Kolokathis, K. Markey, H. Reinsch, M. Baumgartner, M. A. van der Veen, D. E. De Vos, N. Stock, G. K. Papadopoulos, S. K. Henninger and C. Janiak, *J. Mater. Chem. A*, 2016, **4**, 11859–11869.

(105) T. N. Tu, H. T. D. Nguyen and N. T. Tran, *Inorg. Chem. Front.*, 2019, **6**, 2441–2447.

(106) D. Zacher, R. Nayuk, R. Schweins, R. A. Fischer and K. Huber, *Cryst. Growth Des.*, 2014, **14**, 4859–4863.

(107) R. A. Homan, D. S. Hendricks, T. M. Rayder, U. S. Thein, K. J. Fossum, A. P. Claudio Vázquez, J. Yan, R. L. Grimm, S. C. Burdette and J. C. MacDonald, *Cryst. Growth Des*, 2019, **19**, 6331–6338.

(108) L. Esrafili, M. Gharib, A. Morsali and P. Retailleau, *Ultrason. Sonochemistry*, 2020, **66**, 105110.

(109) J. Park, Q. Jiang, D. Feng, L. Mao and H.-C. Zhou, *J. Am. Chem. Soc.*, 2016, **138**, 3518–3525.

(110) T. Li, P. Hu, J. Li, P. Huang, W. Tong and C. Gao, *Colloids Surf. A: Physicochem. Eng. Asp.*, 2019, **577**, 456–463.

(111) S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1–19.

(112) J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.

(113) R. W. Hockney and J. W. Eastwood, *Computer simulation using particles*, A. Hilger, Bristol, 1988.

(114) V. M. Starov and M. G. Velarde, *Wetting and spreading dynamics*, CRC Press, Boca Raton, 2nd edition, 2019.

(115) V. Pratt, *SIGGRAPH Comput. Graph.*, 1987, **21**, 145–152.

(116) M. Isaiev, S. Burian, L. Bulavin, M. Gradeck, F. Lemoine and K. Termentzidis, *Mol. Simul.*, 2016, **42**, 910–915.

(117) H. W. Horn, W. C. Swope and J. W. Pitera, *J. Chem. Phys.*, 2005, **123**, 194504.

(118) X. Zheng, S. Liu, S. Rehman, Z. Li and P. Zhang, *Chem. Eng. J.*, 2020, **389**, 123424.

(119) S. He, H. Wang, C. Zhang, S. Zhang, Y. Yu, Y. Lee and T. Li, *Chem. Sci.*, 2019, **10**, 1816–1822.

(120) A. F. Combariza and G. Sastre, *J. Phys. Chem. C*, 2011, **115**, 13751–13758.

(121) N. E. R. Zimmermann, B. Smit and F. J. Keil, *J. Phys. Chem. C*, 2010, **114**, 300–310.

(122) J. C. S. Remi, A. Lauerer, C. Chmelik, I. Vandendael, H. Terryn, G. V. Baron, J. F. M. Denayer and J. Kärger, *Nature Mater.*, 2016, **15**, 401–406.

(123) Y. J. Colón, D. A. Gómez-Gualdrón and R. Q. Snurr, *Cryst. Growth Des.*, 2017, **17**, 5801–5810.

(124) P. Ramsamy, A. Haffegee, R. Jamieson and V. Alexandrov, in *Computational Science – ICCS 2006*, ed. D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, V. N. Alexandrov, G. D. van Albada, P. M. A. Sloot and J. Dongarra, Series Title: Lecture Notes in Computer Science, Springer, Berlin, 2006, vol. 3992, pp. 603–609.

(125) M. B. O'Connor, S. J. Bennie, H. M. Deeks, A. Jamieson-Binnie, A. J. Jones, R. J. Shannon, R. Walters, T. J. Mitchell, A. J. Mulholland and D. R. Glowacki, *J. Chem. Phys.*, 2019, **150**, 220901.

(126) D. Hamilton, J. McKechnie, E. Edgerton and C. Wilson, *J. Comput. Educ.*, 2021, **8**, 1–32.

(127) S. Ottosson, *J. Eng. Des.*, 2002, **13**, 159–172.

(128) S. Choi, K. Jung and S. D. Noh, *Concurr. Eng. Res. Appl.*, 2015, **23**, 40–63.

(129) M. J. Kim, C.-K. Lee and T. Jung, *J. Travel Res.*, 2020, **59**, 69–89.

(130) M. Secci, C. Beltrame, S. Manfio and F. Guerra, *J. Cult. Herit.*, 2019, **40**, 169–176.

(131) C.-R. Lin, R. B. Loftin and T. Stark, Proceedings. 3rd Asia Pacific Computer Human Interaction, IEEE Comput. Soc, Shonan Village Center, Japan, 1998, pp. 196–201.

(132) A. Ferreira and C. Mavroidis, *IEEE Robot. Automat. Mag.*, 2006, **13**, 78–92.

(133) M. Virvou and G. Katsionis, *Comput. Educ.*, 2008, **50**, 154–178.

(134) M. Alfadil, *Comput. Educ.*, 2020, **153**, 103893.

(135) A. Cheng, L. Yang and E. Andersen, Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, ACM, Denver Colorado USA, 2017, pp. 541–549.

(136) D. Korgel, *Virtual Reality-Spiele entwickeln mit Unity: Grundlagen, Beispielprojekte, Tipps & Tricks*, Hanser, München, 2018.

(137) Z. Ai and T. Fröhlich, *Comput. Graph Forum*, 1998, **17**, 267–273.

(138) C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon and J. C. Hart, *Commun. ACM*, 1992, **35**, 64–72.

(139) A. Salvadori, G. Del Frate, M. Pagliai, G. Mancini and V. Barone, *Int. J. Quantum Chem.*, 2016, **116**, 1731–1746.

(140) K. C. Cassidy, J. Šefčík, Y. Raghav, A. Chang and J. D. Durrant, *PLoS Comput. Biol.*, 2020, **16**, e1007747.

(141) J. C. Phillips, D. J. Hardy, J. D. C. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M. C. R. Melo, B. K. Radak, R. D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L. V. Kalé, K. Schulten, C. Chipot and E. Tajkhorshid, *J. Chem. Phys.*, 2020, **153**, 044130.

(142) H. Berendsen, D. van der Spoel and R. van Drunen, *Comput. Phys. Commun.*, 1995, **91**, 43–56.

(143) J. Juárez-Jiménez, P. Tew, M. O'Connor, S. Llabrés, R. Sage, D. Glowacki and J. Michel, *J. Chem. Inf. Model.*, 2020, **60**, 6344–6354.

(144) W. Smith and T. Forester, *J. Mol. Graph.*, 1996, **14**, 136–141.

(145) C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp and R. Q. Snurr, *Nature Chem.*, 2012, **4**, 83–89.

(146) R. Anderson and D. A. Gómez-Gualdrón, *CrystEngComm*, 2019, **21**, 1653–1665.

(147) S. Amirjalayer, M. Tafipolsky and R. Schmid, *J. Phys. Chem. C*, 2011, **115**, 15133–15139.

(148) S. Impeng, R. Cedeno, J. P. Dürholt, R. Schmid and S. Bureekaew, *Cryst. Growth Des*, 2018, **18**, 2699–2706.

(149) S. Bureekaew, V. Balwani, S. Amirjalayer and R. Schmid, *CrystEngComm*, 2015, **17**, 344–352.

(150) B. Cordero, V. Gómez, A. E. Platero-Prats, M. Revés, J. Echeverría, E. Cremades, F. Barragán and S. Alvarez, *Dalton Trans.*, 2008, **37**, 2832–2838.

(151) E. Argueta, J. Shaji, A. Gopalan, P. Liao, R. Q. Snurr and D. A. Gómez-Gualdrón, *J. Chem. Theory Comput.*, 2018, **14**, 365–376.

(152) M. G. Martin and J. I. Siepmann, *J. Phys. Chem. B*, 1998, **102**, 2569–2577.

(153) B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187–217.

(154) T. F. Miller, M. Eleftheriou, P. Pattnaik, A. Ndirango, D. Newns and G. J. Martyna, *J. Chem. Phys.*, 2002, **116**, 8649–8659.

(155) J. S. Rowlinson and F. L. Swinton, *Liquids and liquid mixtures*, Butterworth Scientific, London, 3rd edition, 1982.

(156) B. Widom, *J. Chem. Phys.*, 1963, **39**, 2808–2812.

(157) F. W. Sears and G. L. Salinger, *Thermodynamics, kinetic theory, and statistical thermodynamics*, Addison-Wesley Pub. Co, Reading, 3d edition, 1975.

(158) R. Bowley and M. Sánchez, *Introductory statistical mechanics*, Clarendon Press, Oxford, 2nd edition, 1999.

(159) D. S. Sholl and J. A. Steckel, *Density functional theory: a practical introduction*, Wiley, Hoboken, 2009.

(160) J. Kohanoff, *Electronic Structure Calculations for Solids and Molecules: Theory and Computational Methods*, Cambridge University Press, Cambridge, 2006.

(161) P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, B864–B871.

(162) W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133–A1138.

(163) I. Langmuir, *J. Am. Chem. Soc.*, 1918, **40**, 1361–1403.

(164) A. Erto, A. Lancia and D. Musmarra, *Microporous and Mesoporous Mat.*, 2012, **154**, 45–50.

(165) T. Hähnel, J. Möllmer, M. Klauck and G. Kalies, *Adsorption*, 2020, **26**, 361–373.

(166) K. Sladekova, C. Campbell, C. Grant, A. J. Fletcher, J. R. B. Gomes and M. Jorge, *Adsorption*, 2020, **26**, 663–685.

(167) R. B. Corey and L. Pauling, *Rev. Sci. Instrum.*, 1953, **24**, 621–627.

(168) W. L. Koltun, *pat.*, U. S. Patent 3170246, 1962.

(169) A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.

(170) C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.

(171) S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 1980, **58**, 1200–1211.

(172) S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

(173) S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.

(174) M. Guidon, J. Hutter and J. VandeVondele, *J. Chem. Theory Comput.*, 2010, **6**, 2348–2364.

(175) J. Hutter, M. Iannuzzi, F. Schiffmann and J. VandeVondele, *WIREs Comput. Mol. Sci.*, 2014, **4**, 15–25.

(176) U. Borštnik, J. VandeVondele, V. Weber and J. Hutter, *Parallel Comput.*, 2014, **40**, 47–58.

(177) J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing and J. Hutter, *Comput. Phys. Commun.*, 2005, **167**, 103–128.

(178) G. Lippert, J. Hutter and M. Parrinello, *Mol. Phys.*, 1997, **92**, 477–487.

(179) M. Guidon, F. Schiffmann, J. Hutter and J. VandeVondele, *J. Chem. Phys.*, 2008, **128**, 214104.

(180) M. Guidon, J. Hutter and J. VandeVondele, *J. Chem. Theory Comput.*, 2009, **5**, 3010–3021.

(181) J. VandeVondele and J. Hutter, *J. Chem. Phys.*, 2003, **118**, 4365–4369.

(182) J. Heyd and G. E. Scuseria, *J. Chem. Phys.*, 2004, **120**, 7274–7280.

(183) J. VandeVondele and J. Hutter, *J. Chem. Phys.*, 2007, **127**, 114105.

(184) S. Goedecker, M. Teter and J. Hutter, *Phys. Rev. B*, 1996, **54**, 1703–1710.

(185) C. Hartwigsen, S. Goedecker and J. Hutter, *Phys. Rev. B*, 1998, **58**, 3641–3662.

(186) M. Krack, *Theor. Chem. Acc.*, 2005, **114**, 145–152.

(187) M. G. Martin and J. I. Siepmann, *J. Phys. Chem. B*, 1999, **103**, 4508–4517.

Appendices

A Section 2: Lagrangian Multipliers

In the following, the grand canonical partition function,

$$Z(\mu, V, T) = \sum_N \sum_S \exp(-\beta_{\mathcal{L}} E_{N,S}) \exp(-\gamma_{\mathcal{L}} N), \tag{A1}$$

is used to determine the Lagrangian multipliers $\beta_{\mathcal{L}}$ and $\gamma_{\mathcal{L}}$. First, the total differential of Eq. (A1) is formed,

$$\mathrm{d} \ln Z = \left(\frac{\partial \ln Z}{\partial \gamma_{\mathcal{L}}} \right)_{V,\beta_{\mathcal{L}}} \mathrm{d}\gamma_{\mathcal{L}} + \left(\frac{\partial \ln Z}{\partial V} \right)_{\gamma_{\mathcal{L}},\beta_{\mathcal{L}}} \mathrm{d}V + \left(\frac{\partial \ln Z}{\partial \beta_{\mathcal{L}}} \right)_{\gamma_{\mathcal{L}},V} \mathrm{d}\beta_{\mathcal{L}}, \tag{A2}$$

where $Z := Z(\mu, V, T)$. The derivatives in Eq. (A2) are ([22], p. 53):

$$\left(\frac{\partial \ln Z}{\partial \gamma_{\mathcal{L}}} \right)_{V,\beta_{\mathcal{L}}} = - \frac{\sum_N \sum_S N \exp(-\beta_{\mathcal{L}} E_{N,S}) \exp(-\gamma_{\mathcal{L}} N)}{Z} = -\langle N \rangle, \tag{A3}$$

$$\left(\frac{\partial \ln Z}{\partial V} \right)_{\gamma_{\mathcal{L}},\beta_{\mathcal{L}}} = - \frac{\sum_N \sum_S \beta_{\mathcal{L}} \left(\frac{\partial E_{N,S}}{\partial V} \right) \exp(-\beta_{\mathcal{L}} E_{N,S}) \exp(-\gamma_{\mathcal{L}} N)}{Z} = -\beta_{\mathcal{L}} \langle p \rangle, \tag{A4}$$

and

$$\left(\frac{\partial \ln Z}{\partial \beta_{\mathcal{L}}} \right)_{\gamma_{\mathcal{L}},V} = - \frac{\sum_N \sum_S E_{N,S} \exp(-\beta_{\mathcal{L}} E_{N,S}) \exp(-\gamma_{\mathcal{L}} N)}{Z} = -\langle E \rangle, \tag{A5}$$

where we made use of the following expression ([22], p. 53):

$$\left(\frac{\partial E_{N,S}}{\partial V} \right) = p_{N,S}. \tag{A6}$$

Substituting Eqs. (A3) to (A5) in Eq. (A2) yields:

$$\mathrm{d} \ln Z = -\langle N \rangle \mathrm{d}\gamma_{\mathcal{L}} - \beta_{\mathcal{L}} \langle p \rangle \mathrm{d}V - \langle E \rangle \mathrm{d}\beta_{\mathcal{L}}, \tag{A7}$$

and becomes:

$$\frac{1}{\beta_{\mathcal{L}}} \mathrm{d}(\ln Z + \langle N \rangle \gamma_{\mathcal{L}} + \langle E \rangle \beta_{\mathcal{L}}) = \frac{\gamma_{\mathcal{L}}}{\beta_{\mathcal{L}}} \mathrm{d}\langle N \rangle - \langle p \rangle \mathrm{d}V + \mathrm{d}\langle E \rangle \tag{A8}$$

by applying the Legendre transformation. Equating the coefficients of Eq. (A8) with those of the purely thermodynamic expression,

$$T\mathrm{d}S = -\mu \mathrm{d}N - p\mathrm{d}V + \mathrm{d}U, \tag{A9}$$

where $p = \langle p \rangle$, $N = \langle N \rangle$, or $U = \langle E \rangle$, results in:

$$\beta_{\mathcal{L}} = \frac{1}{CT}, \tag{A10}$$

and

$$\gamma_{\mathcal{L}} = -\frac{\mu}{CT}, \tag{A11}$$

where C ensures that Eq. (A8) is expressed in the unit of energy. We obtain C by comparing the equations of state for an ideal gas in a closed system that are derived from statistical thermodynamics or classical thermodynamics. For a closed system in thermodynamic equilibrium, the free energy, $\tilde{A}$, is minimal, and the total differential of the free energy is:

$$\mathrm{d}\tilde{A} = \left(\frac{\partial \tilde{A}}{\partial N}\right)_{V,T} \mathrm{d}N + \left(\frac{\partial \tilde{A}}{\partial V}\right)_{N,T} \mathrm{d}V + \left(\frac{\partial \tilde{A}}{\partial T}\right)_{N,V} \mathrm{d}T. \tag{A12}$$

By equating the coefficients in Eq. (A12) with those of the purely thermodynamic expression:

$$\mathrm{d}\tilde{A} = \mu \mathrm{d}N - p\mathrm{d}V - S\mathrm{d}T, \tag{A13}$$

we obtain the derivative of free energy with respect to the volume,

$$\left(\frac{\partial \tilde{A}}{\partial V}\right)_{N,T} = -p = -\frac{N k_{\mathrm{B}} T}{V}, \tag{A14}$$

where the last identity on the right-hand side holds for an ideal gas. The expression of the free energy in statistical thermodynamics is given by:

$$\tilde{A} = -\frac{1}{\beta_{\mathcal{L}}} \ln Z(N, V, T), \tag{A15}$$

where $Z(N, V, T)$ is the canonical partition function. The derivative of the free energy with respect to the volume is equal to:

$$\left(\frac{\partial \tilde{A}}{\partial V}\right)_{N,T} = -\frac{1}{\beta_{\mathcal{L}}} \left(\frac{\partial \ln Z(N, V, T)}{\partial V}\right)_{N,T}. \tag{A16}$$

For the identical and indistinguishable particles in an ideal gas, the canonical partition function reduces to:

$$Z(N, V, T) = \frac{(z^{\dagger})^N}{N!}, \tag{A17}$$

where $z^{\dagger}$ is the canonical particle partition function for identical particles. Using Eq. (A17) in conjunction with Stirling's approximation, we transform Eq. (A16) into the following

expression ([18], pp. 87–88):

$$\left(\frac{\partial \tilde{A}}{\partial V}\right)_{N,T} = -\frac{N}{\beta_{\mathcal{L}}}\left(\frac{\partial \ln\left[\frac{z^{\dagger}}{N}\right]}{\partial V}\right)_{N,T}. \tag{A18}$$

The particles in an ideal gas possess only translational degrees of freedom, and the canonical partition function of a particle reduces to:

$$z^{\dagger} = \frac{V}{\Lambda^3}, \tag{A19}$$

where Λ is the de-Broglie wavelength (see Appendix B for a derivation). Substituting Eq. (A19) in Eq. (A18) yields:

$$\left(\frac{\partial \tilde{A}}{\partial V}\right)_{N,T} = -\frac{N}{\beta_{\mathcal{L}}}\left(\frac{\partial \ln\left[\frac{V}{\Lambda^3 N}\right]}{\partial V}\right)_{N,T} = -\frac{N}{\beta_{\mathcal{L}} V}, \tag{A20}$$

and by comparing the expressions in Eq. (A20) and Eq. (A14), the Lagrangian parameter,

$$\beta_{\mathcal{L}} = \frac{1}{k_{\mathrm{B}} T}, \tag{A21}$$

stating that C in Eq. (A10) and Eq. (A11) is equal to the Boltzmann factor, k_{B}.

B Section 2: Transition to Classical Partition Functions

In the following, the proportionality constant for transitioning from quantum mechanical partition functions to classical partition functions is evaluated. For this, the translational partition function, z_{trans}, of a particle in a box of side lengths $d_{\text{sys}} \times d_{\text{sys}} \times d_{\text{sys}}$ is determined based on a quantum mechanical and classical mechanical description of translational states. According to quantum mechanics, the energy, $\varepsilon_{\tilde{n}}$, of any translational quantum state of a particle is given by:

$$\varepsilon_{\tilde{n}} = \frac{\tilde{h}^2}{8md_{\text{sys}}^2}(\tilde{n}_1^2 + \tilde{n}_2^2 + \tilde{n}_3^2), \tag{B1}$$

where $\tilde{h}$ is Planck's constant and $\tilde{n}_1, \tilde{n}_2$, or $\tilde{n}_3$ are the independent quantum numbers for describing the translational quantum state. The translational partition function for a particle in the microcanonical ensemble is:

$$z_{\text{trans}} = \sum_{\tilde{n}_1=1}^{\infty}\sum_{\tilde{n}_2=1}^{\infty}\sum_{\tilde{n}_3=1}^{\infty} \exp\left(-\frac{\varepsilon_{\tilde{n}}}{k_{\text{B}}T}\right) = \sum_{\tilde{n}_1=1}^{\infty}\sum_{\tilde{n}_2=1}^{\infty}\sum_{\tilde{n}_3=1}^{\infty} \exp\left(-\frac{h^2(\tilde{n}_1^2 + \tilde{n}_2^2 + \tilde{n}_3^2)}{8md_{\text{sys}}^2 k_{\text{B}}T}\right). \tag{B2}$$

Since $\tilde{n}_1, \tilde{n}_2$, or $\tilde{n}_3$ are independent, Eq. (B2) can be factorized,

$$z_{\text{trans}} = \left[\sum_{\tilde{n}=1}^{\infty} \exp\left(-\frac{h^2\tilde{n}^2}{8md_{\text{sys}}^2 k_{\text{B}}T}\right)\right]^3. \tag{B3}$$

Under normal conditions, the consecutive terms of the sum in Eq. (B3) are close. Thus, the sum in Eq. (B3) can be approximated by an integral,

$$z_{\text{trans}} \approx \left[\int_{\tilde{n}=0}^{\infty} \exp\left(-\frac{h^2\tilde{n}^2}{8md_{\text{sys}}^2 k_{\text{B}}T}\right) \mathrm{d}\tilde{n}\right]^3. \tag{B4}$$

Eq. (B4) is the standard integral of a Gaussian and can be solved analytically to yield the quantum mechanical expression of the translational partition function of a particle ([18], p. 81),

$$z_{\text{trans}} = \left(\frac{2\pi m k_{\text{B}}T}{h^2}\right)^{3/2} d_{\text{sys}}^3. \tag{B5}$$

In classical mechanics, the energy of any translational state is described by the Hamiltonian:

$$\mathcal{H} = \frac{1}{2m}\|\vec{P}\|^2, \tag{B6}$$

which leads to:

$$z_{\text{trans}} = \frac{1}{C^3}\int_0^{d_{\text{sys}}}\int_0^{d_{\text{sys}}}\int_0^{d_{\text{sys}}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\frac{\|\vec{P}\|^2}{2mk_{\text{B}}T}\right) \mathrm{d}\vec{r}_x\, \mathrm{d}\vec{r}_y\, \mathrm{d}\vec{r}_z\, \mathrm{d}\vec{P}_x\, \mathrm{d}\vec{P}_y\, \mathrm{d}\vec{P}_z, \tag{B7}$$

where C is a proportionality constant of unit: momentum $\times$ length to ensure that z_{trans} is unitless. Since the components of the momentum vector, $\vec{P}_x, \vec{P}_y$, and $\vec{P}_z$, are independent from one another as well as from the components of the position vector, $\vec{r}_x, \vec{r}_y$, and $\vec{r}_z$, the sixfold integral in Eq. (B7) can be simplified into the following expression:

$$z_{\text{trans}} = \frac{1}{C^3} \left[\int_0^{d_{\text{sys}}} \mathrm{d}\vec{r}_x \int_0^{d_{\text{sys}}} \mathrm{d}\vec{r}_y \int_0^{d_{\text{sys}}} \mathrm{d}\vec{r}_z \right] \left[\int_{-\infty}^{\infty} \exp\left(-\frac{P^2}{2mk_{\text{B}}T} \right) \mathrm{d}P \right]^3 , \tag{B8}$$

where P is any of the three spatial components of $\vec{P}$. Eq. (B8) can be solved analytically to yield the classical mechanical expression of the translational partition function of a particle:

$$z_{\text{trans}} = \left(\frac{2\pi mk_{\text{B}}T}{C^2} \right)^{3/2} {d_{\text{sys}}}^3. \tag{B9}$$

Comparing the expressions for the translational partition function in Eq. (B5) and Eq. (B9) suggests that:

$$C = \tilde{h}. \tag{B10}$$

Hence, Planck's constant is the proportionality constant for the transition from quantum mechanical partition functions to classical partition functions.

C Section 2: Evaluation of the Chemical Potential

Monte Carlo simulations in the grand canonical ensemble allow a system to exchange particles with a particle reservoir of constant chemical potential. The chemical potential can be determined by:

$$\mu = \left(\frac{\partial \tilde{A}}{\partial N}\right)_{V,T} \approx \tilde{A}_{N+1} - \tilde{A}_N, \tag{C1}$$

where $\tilde{A}_N$ and $\tilde{A}_{N+1}$ are the Helmholtz free energies for a system with N or $N+1$ particles. Using Eq. (A15), Eq. (C1) can be transformed into:

$$\mu = -k_\mathrm{B}T \ln Z_{N+1} + k_\mathrm{B}T \ln Z_N = -k_\mathrm{B}T \ln \frac{Z_{N+1}}{Z_N}, \tag{C2}$$

where Z_N and Z_{N+1} are the canonical partition functions for a system with N or $N+1$ distinguishable particles. However, identical particles in an ideal gas are not distinguishable. Thus, the partition function has to be divided by the permutation of the N or $N+1$ particles,

$$\mu = -k_\mathrm{B}T \ln \frac{Z^\dagger_{N+1} N!}{Z^\dagger_N (N+1)!} = -k_\mathrm{B}T \ln \frac{Z^\dagger_{N+1}}{Z^\dagger_N (N+1)}, \tag{C3}$$

where $Z^\dagger_N$ and $Z^\dagger_{N+1}$ are the canonical partition functions for a system with N or $N+1$ indistinguishable particles. Substituting the expression for the classical canonical partition function in Eq. (C3) yields:

$$\mu = -k_\mathrm{B}T \ln \left(\frac{1}{(N+1)\Lambda^3} \frac{\int \cdots \int \exp\left(-\frac{\mathcal{U}(\vec{r}_1,\cdots,\vec{r}_{N+1})}{k_\mathrm{B}T} \right) \mathrm{d}\vec{r}_1 \cdots \mathrm{d}\vec{r}_{N+1}}{\int \cdots \int \exp\left(-\frac{\mathcal{U}(\vec{r}_1,\cdots,\vec{r}_N)}{k_\mathrm{B}T} \right) \mathrm{d}\vec{r}_1 \cdots \mathrm{d}\vec{r}_N} \right), \tag{C4}$$

which reduces to ([26], pp. 173–174):

$$\mu = -k_\mathrm{B}T \ln \left(\frac{\left\langle \int \int \int \exp\left(-\frac{\mathcal{U}(\vec{r}_1,\cdots,\vec{r}_N|\vec{r}_{N+1})}{k_\mathrm{B}T} \right) \mathrm{d}\vec{r}_{N+1} \right\rangle}{(N+1)\Lambda^3} \right), \tag{C5}$$

where

$$\mathcal{U}(\vec{r}_1,\cdots,\vec{r}_N|\vec{r}_{N+1}) = \mathcal{U}(\vec{r}_1,\cdots,\vec{r}_{N+1}) - \mathcal{U}(\vec{r}_1,\cdots,\vec{r}_N) \tag{C6}$$

and the numerator in Eq. (C5) is given by:

$$\langle \cdots \rangle = \int \cdots \int \mathrm{P}(\{\vec{r}_N\}) \left[\int \int \int \exp\left(-\frac{\mathcal{U}(\vec{r}_1,\cdots,\vec{r}_N|\vec{r}_{N+1})}{k_\mathrm{B}T} \right) \mathrm{d}\vec{r}_{N+1} \right], \tag{C7}$$

where $\langle \cdots \rangle$ is a short-hand notation for the canonical ensemble average in Eq. (C5) and $\mathrm{P}(\{\vec{r}_N\})$ is the probability of any configuration $\{\vec{r}_N\}$ in the canonical ensemble. The ensem-

ble average in Eq. (C7) can be evaluated by the Metropolis-Hastings algorithm, while the three-dimensional integral in the ensemble average can be solved by a classical Monte Carlo simulation. This is achieved by using the *Widom insertion* trial move [156]. In this trial move, the Metropolis-Hastings algorithm is used to generate configurations of the system from $\mathrm{P}(\{\vec{r}_N\})$. At frequent intervals in the simulation, an additional particle, $N+1$, is generated at a random position, $\vec{r}_{N+1}$, in the system. The potential energy of the additional particle is calculated, the Boltzmann factor in Eq. (C7) is determined, and the particle is removed from the system. Averaging the thus determined Boltzmann factors over all generated positions $\vec{r}_{N+1}$ yields an approximation of the ensemble average in Eq. (C5).

Often, the particle reservoir is assumed to consist of an ideal gas or ideal chain molecules. An ideal gas possesses no potential energy,

$$\mathcal{U} = 0. \tag{C8}$$

Thus, Eq. (C4) reduces to:

$$\mu_{\mathrm{id}} = -k_{\mathrm{B}}T \ln\left(\frac{\int\int\int \mathrm{d}\vec{r}_{N+1}}{(N+1)\Lambda^3}\right) = -k_{\mathrm{B}}T \ln\left(\frac{V}{(N+1)\Lambda^3}\right) = k_{\mathrm{B}}T \ln\left(\frac{p}{k_{\mathrm{B}}T}\Lambda^3\right), \tag{C9}$$

where the last identity holds because of the ideal gas law,

$$pV = (N+1)k_{\mathrm{B}}T. \tag{C10}$$

On the other hand, an ideal chain possesses only internal degrees of freedom, i.e.:

$$\mathcal{U} = \mathcal{U}_{\mathrm{intra}}. \tag{C11}$$

The chemical potential, μ_{ic}, can be computed in analogy to Eq. (C5), i.e.:

$$\mu_{\mathrm{ic}} = -k_{\mathrm{B}}T \ln\left(\frac{\left\langle \int\int\int \exp\left(-\frac{\mathcal{U}_{\mathrm{intra}}(\vec{r}_1,\cdots,\vec{r}_N|\vec{r}_{N+1})}{k_{\mathrm{B}}T}\right) \mathrm{d}\vec{r}_{N+1}\right\rangle}{(N+1)\Lambda^3}\right). \tag{C12}$$

Eq. (C12) can be transformed into ([26], p. 177):

$$\begin{aligned}\mu_{\mathrm{ic}} &= -k_{\mathrm{B}}T \ln\left(\frac{V}{(N+1)\Lambda^3}\right) \\ &\quad -k_{\mathrm{B}}T \ln\left(\frac{1}{V}\left\langle \int\int\int \exp\left(-\frac{\mathcal{U}_{\mathrm{intra}}(\vec{r}_1,\cdots,\vec{r}_N|\vec{r}_{N+1})}{k_{\mathrm{B}}T}\right) \mathrm{d}\vec{r}_{N+1}\right\rangle\right),\end{aligned} \tag{C13}$$

which is tantamount to:

$$\mu_{\mathrm{ic}} = \mu_{\mathrm{id}} + \mu_{\mathrm{ex}}, \tag{C14}$$

where

$$\mu_{\mathrm{ex}} = -k_{\mathrm{B}}T \ln\left(\frac{1}{V}\left\langle \int\int\int \exp\left(-\frac{\mathcal{U}_{\mathrm{intra}}(\vec{r}_1,\cdots,\vec{r}_N|\vec{r}_{N+1})}{k_{\mathrm{B}}T}\right) \mathrm{d}\vec{r}_{N+1}\right\rangle\right) \tag{C15}$$

is the excess chemical potential that accounts for the deviation of an ideal chain from an ideal gas. Eq. (C15) only depends on the intramolecular interaction of particle $N+1$ and the temperature. It can be determined by a Monte Carlo simulation of an empty system in the canonical ensemble using Widom insertion via CBMC [32].

D Section 2: RATTLE Algorithm

In addition to simulating the motion of fully flexible guest molecules by evaluating bond stretching, bond angle bending, and dihedral forces or rigid guest molecules by using the rotational-velocity-Verlet algorithm, MOF-VR offers to constrain certain intramolecular degrees of freedom by maintaining the distance between specified interaction sites using the RATTLE algorithm [36]. The RATTLE algorithm consists of four steps [36]:

1) Advance the positions of each mobile interaction site i in the system due to intermolecular and unconstrained intramolecular forces by applying Eqs. (133) and (134) of the velocity Verlet algorithm.

2) Correct the unconstrained positions, $\vec{r}_i\,'(t+\Delta t)$, and half-step velocities, $\vec{v}_i\,'(t+\frac{1}{2}\Delta t)$, for constraining intramolecular forces, $\vec{g}_i(t)$, acting on i at $\vec{r}_i(t)$,

$$\vec{v}_i(t+\tfrac{1}{2}\Delta t) = \vec{v}_i\,'(t+\tfrac{1}{2}\Delta t) + \frac{1}{2m_i}\vec{g}_i(t)\Delta t, \tag{D1}$$

$$\vec{r}_i(t+\Delta t) = \vec{r}_i\,'(t+\Delta t) + \frac{1}{m_i}\vec{g}_i(t)\Delta t^2. \tag{D2}$$

3) Compute the velocities, $\vec{v}_i\,'(t+\Delta t)$, due to intermolecular and unconstrained intramolecular forces acting on i at $\vec{r}_i(t+\Delta t)$ by applying Eqs. (135) and (136) of the velocity Verlet algorithm.

4) Correct the unconstrained velocities, $\vec{v}_i\,'(t+\Delta t)$, for constraining intramolecular forces, $\vec{g}_i(t+\Delta t)$, acting on i at $\vec{r}_i(t+\Delta t)$,

$$\vec{v}_i(t+\Delta t) = \vec{v}_i\,'(t+\Delta t) + \frac{1}{2m_i}\vec{g}_i(t+\Delta t)\Delta t. \tag{D3}$$

Constraining intramolecular forces are determined by using Lagrange multipliers,

$$\vec{g}_i(t) = \frac{1}{2}\sum_{\mathcal{L}(i\text{-}j)} \lambda_{\mathcal{L}}(t)\left(\frac{\partial\bar{\sigma}}{\partial\vec{r}_i}\right), \tag{D4}$$

$$\vec{g}_i(t+\Delta t) = \frac{1}{2}\sum_{\mathcal{L}(i\text{-}j)} \lambda_{\mathcal{L}}(t+\Delta t)\left(\frac{\partial\bar{\sigma}}{\partial t}\right), \tag{D5}$$

where $\lambda_{\mathcal{L}}(t)$ and $\lambda_{\mathcal{L}}(+\Delta t)$ are Lagrangian multipliers and $\bar{\sigma}(t)$ and $\bar{\sigma}(t+\Delta t)$ are holonomic constrains, such as

$$\sigma_{c(ij)} = \|\vec{d}_{i\text{-}j}\|^2 - d_{i\text{-}j}^2, \tag{D6}$$

where $d_{i\text{-}j}$ is the desired distance between i and j. Lagrangian multipliers are determined iteratively by looping through the constraints [36]. We will refrain from a detailed description of the iterative procedure at this point and instead refer to the Supporting Information of earlier work [56].

E Section 2: Maxwell-Boltzmann Distribution of Speed

The Maxwell-Boltzmann distribution of speed describes the probability of finding a particle with a certain speed in an ideal gas. Starting point for the derivation is the Boltzmann distribution. In its classical formulation, the Boltzmann distribution, $\mathrm{P}(\vec{r},\vec{v})$, can be split into:

$$\mathrm{P}(\vec{r},\vec{v}) = \mathrm{P}(\vec{r})\mathrm{P}(\vec{v}), \tag{E1}$$

where

$$\mathrm{P}(\vec{r}) = \frac{\exp\left(-\frac{\mathcal{U}(\vec{r})}{k_\mathrm{B}T}\right)\mathrm{d}\vec{r}_\mathrm{x}\mathrm{d}\vec{r}_\mathrm{y}\mathrm{d}\vec{r}_\mathrm{z}}{\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}\exp\left(-\frac{\mathcal{U}(\vec{r})}{k_\mathrm{B}T}\right)\mathrm{d}\vec{r}_\mathrm{x}\mathrm{d}\vec{r}_\mathrm{y}\mathrm{d}\vec{r}_\mathrm{z}} \tag{E2}$$

is the probability that a particle is at a certain position, $\vec{r}$, and

$$\mathrm{P}(\vec{v}) = \frac{\exp\left(-\frac{m\|\vec{v}\|^2}{2k_\mathrm{B}T}\right)\mathrm{d}\vec{v}_\mathrm{x}\mathrm{d}\vec{v}_\mathrm{y}\mathrm{d}\vec{v}_\mathrm{z}}{\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}\exp\left(-\frac{m\|\vec{v}\|^2}{2k_\mathrm{B}T}\right)\mathrm{d}\vec{v}_\mathrm{x}\mathrm{d}\vec{v}_\mathrm{y}\mathrm{d}\vec{v}_\mathrm{z}}, \tag{E3}$$

is the probability that a particle has a certain velocity, $\vec{v}$. By solving the denominator of Eq. (E3) analytically (see Sec. B), we obtain the *Maxwell distribution of velocity*,

$$\mathrm{P}(\vec{v}) = \sqrt{\left(\frac{m}{2\pi k_\mathrm{B}T}\right)^3}\exp\left(-\frac{m\|\vec{v}\|^2}{2k_\mathrm{B}T}\right)\mathrm{d}\vec{v}_\mathrm{x}\mathrm{d}\vec{v}_\mathrm{y}\mathrm{d}\vec{v}_\mathrm{z}. \tag{E4}$$

The Maxwell-Boltzmann distribution of speed, $\mathrm{P}(\|\vec{v}\|)$, can be derived based on the realization that there is an infinite number of vectors that possess the same speed, $\|\vec{v}\|$ ([157], pp. 354–357). These vectors form a sphere in velocity space, so that Eq. (E4) needs to be integrated over the spherical coordinates, namely the azimuthal angle, ω_azi, and the polar angle, ω_pol ([158], pp. 146–150). By transforming Eq. (E4) to spherical coordinates and integration, we obtain:

$$\mathrm{P}(\|\vec{v}\|) = \sqrt{\left(\frac{m}{2\pi k_\mathrm{B}T}\right)^3}\exp\left(-\frac{m\|\vec{v}\|^2}{2k_\mathrm{B}T}\right)\|\vec{v}\|^2\mathrm{d}\|\vec{v}\|\int_0^\infty \sin\omega_\mathrm{azi}\mathrm{d}\omega_\mathrm{azi}\int_0^{2\pi}\mathrm{d}\omega_\mathrm{pol}, \tag{E5}$$

which yields the Maxwell-Boltzmann distribution of speed ([157], p. 355):

$$\mathrm{P}(\|\vec{v}\|) = 4\pi\sqrt{\left(\frac{m}{2\pi k_\mathrm{B}T}\right)^3}\exp\left(-\frac{m\|\vec{v}\|^2}{2k_\mathrm{B}T}\right)\|\vec{v}\|^2\mathrm{d}\|\vec{v}\|. \tag{E6}$$

F Section 2: Basics of Quantum Mechanics

For a better understanding of the methods for determining partial charges described in Sec. 2.3.3, it seems beneficial to cover the basics of quantum mechanics. As most of this work is concerned with the application of classical simulation methods, the following explanations are intended to provide a rough idea about the concepts of quantum mechanics. Starting point for our descriptions is the time-independent Schrödinger equation,

$$\hat{\mathcal{H}}\Psi = E\Psi, \tag{F1}$$

where $\hat{\mathcal{H}}$ is the Hamilton operator, Ψ is the wave function of the system, and E its energy. The Hamilton operator is the sum of the kinetic energy operator, $\hat{\mathcal{K}}$, and the potential energy operator $\hat{\mathcal{U}}$,

$$\hat{\mathcal{H}} = \hat{\mathcal{K}} + \hat{\mathcal{U}} \tag{F2}$$

where the kinetic energy operator contains the contributions due to the motion of electrons and nuclei,

$$\hat{\mathcal{K}} = -\sum_{i=1}^{N_{\mathrm{e}^-}} \frac{\tilde{h}^2}{2m_{\mathrm{e}^-}} \nabla_{\mathrm{e}^-}^2 - \sum_{j=1}^{N_{\mathrm{n}^+}} \frac{\tilde{h}^2}{2m_{\mathrm{n}^+}} \nabla_{\mathrm{n}^+}^2, \tag{F3}$$

where n_{e^-} and n_{n^+} are the number of electrons or nuclei, $\tilde{h}$ is Planck's constant, and m_{e^-} and m_{n^+} are the mass of an electron or a nucleus. The potential energy operator contains the contributions due to nucleus-electron interactions, electron-electron interactions, and nucleus-nucleus interactions,

$$\hat{\mathcal{U}} = -\frac{1}{4\pi\epsilon_0} \left[\sum_{i=1}^{N_{\mathrm{e}^-}} \sum_{j=1}^{N_{\mathrm{n}^+}} \frac{\tilde{q}_j^+ e^2}{d_{i\text{-}j}} - \sum_{i=1}^{N_{\mathrm{e}^-}} \sum_{j=i+1}^{N_{\mathrm{e}^-}} \frac{e^2}{d_{i\text{-}j}} - \sum_{i=1}^{N_{\mathrm{n}^+}} \sum_{j=i+1}^{N_{\mathrm{n}^+}} \frac{\tilde{q}_i^+ \tilde{q}_j^+ e^2}{d_{i\text{-}j}} \right], \tag{F4}$$

where ϵ_0 is the vacuum permittivity, $\tilde{q}^+$ is the nucleus' charge, e is the elementary charge, and $d_{i\text{-}j}$ is the distance between the species. The wave function is searched that solves the Schrödinger equation in Eq. (F2) for a given system. Commonly this is done using the following simplifications ([159], pp. 8–9):

1) the positions of the nuclei are fixed, because the much larger mass of a nucleus relative to an electron results in the velocity of an nucleus being much smaller than the velocity of an electron (Born-Oppenheimer approximation), and

2) the nucleus-nucleus interactions are ignored, because the fixed nucleus positions result in these interactions being constant.

These simplifications lead to the electronic Hamilton operator ([160], p. 15),

$$\hat{\mathcal{H}}_{\mathrm{el}} = -\sum_{i=1}^{N_{\mathrm{e}^-}} \frac{h^2}{2m_{\mathrm{e}^-}} \nabla_{\mathrm{e}^-}^2 - \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{N_{\mathrm{e}^-}} \sum_{j=1}^{N_{\mathrm{n}^+}} \frac{\tilde{q}_j^+ e^2}{d_{i\text{-}j}} + \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{N_{\mathrm{e}^-}} \sum_{j=i+1}^{N_{\mathrm{e}^-}} \frac{e^2}{d_{i\text{-}j}}. \tag{F5}$$

The wave function of the system is commonly constructed by using two approximations:

1) The *Hartree product* ([29], p. 38) states that the many-electron wave function, Ψ, can be approximated by:

$$\Psi(\vec{r}_1, \cdots, \vec{r}_{N_{e^-}}) \approx \psi_1(\vec{r})\psi_2(\vec{r}) \cdots \psi_{N_{e^-}}(\vec{r}), \tag{F6}$$

where $\Psi(\vec{r}_1, \cdots, \vec{r}_{N_{e^-}})$ is the many-body wave function in Eq. (F2), and $\psi_i(\vec{r})$ are one-electron wave functions, commonly called *molecular orbitals*. The Hartree product is a rather harsh approximation and suffers from two general disadvantages:

   a) it wrongly assumes that electrons are independent, and

   b) it does not satisfy the anti-symmetry principle for fermions, stating that a wave function must change sign under mutual exchange of electrons (i.e. $\Psi(\vec{r}_1, \vec{r}_2) = -\Psi(\vec{r}_2, \vec{r}_1)$).

2) The *linear combination of atomic orbitals* (LCAO) approach ([29], p. 41), assumes that any molecular orbital, $\psi(\vec{r})$, can be approximated by a linear combination of basis functions, φ,

$$\psi = \sum_i w_i \varphi_i, \tag{F7}$$

where w_i are the weights for the atomic orbitals, φ_i.

To find the molecular orbitals that make up the wave function of the system, the electronic Hamiltonian needs to be broken down into n_{e^-} single-electron Hamiltonians, h_i, yielding single-electron Schrödinger equations of the type ([160], p. 61):

$$h_i \psi_i = \varepsilon_i \psi_i, \tag{F8}$$

but this is not readily possible due to the electron-electron interactions involved in the electronic Hamiltonian, which implies that any molecule orbital, ψ_i, can only be determined with simultaneously considering all the other electrons in the system. Thus, h_i can only be approximated. For example, the *Hartree approximation* effectively reduces the many-body electron-electron interactions by a sum of interactions of individual electrons with the electrostatic potential built up by all the other electrons. The contribution of the electron-electron interactions in Eq. (F5) is replaced by the Hartree potential ([160], p. 56),

$$\mathcal{U}_{\mathrm{H}} = \sum_{i=1}^{N_{e^-}} u_{\mathrm{H},i} = \sum_{i=1}^{N_{e^-}} \frac{e^2}{4\pi\epsilon_0} \int \frac{\rho_{e^-}(\vec{r}_j) - \rho_i(\vec{r}_j)}{\|\vec{r}_i - \vec{r}_j\|} \mathrm{d}\vec{r}_j, \tag{F9}$$

where $u_{\mathrm{H},i}$ are the effective one-electron potentials, $\rho(\vec{r}_j)$ is the total electron density of the molecule, and $\rho_i(\vec{r}_j)$ is the electron density of electron i, which is subtracted from $\rho_{e^-}(\vec{r}_j)$ to avoid unphysical self-interactions. The Hartree potential introduces the *electron density*;

an observable property of the system that represents the probability density distribution of finding an electron at any position, $\vec{r}_j$, and is defined by:

$$\rho_{e^-}(\vec{r}) = \sum_{i=1}^{N_{e^-}} \psi_i \psi_i^*, \tag{F10}$$

where ψ_i^* is the complex conjugate of ψ_i, because the wave function is complex. The Hartree approximation is based on the assumption that electrons are statistically independent from one another. This imposes some limitations to the method, namely the absence of *correlation* and *exchange* interactions in the Hartree potential. Correlation interactions arise from Coulomb repulsion: electrons avoid each other and will stay as far apart as possible – they will never come as close as the statistically independent electrons assumed by Hartree –, while exchange interactions are a consequence of Pauli's exclusion principle, which states that fermions (such as electrons) are not allowed to occupy the same quantum state. As a results, spin-like electrons will never be too close to each other – a fact ignored in the Hartree approximation, where electrons are uncorrelated.

Density functional theory (DFT) provides one way of determining the electron density and constructing single-electron Hamiltonians. It is based on two mathematical theorems that were proven by Walter Kohn and Pierre Hohenberg [161]:

1) The ground state energy, E_0, of a system is a unique functional of the electron density,

$$E_0 = E[\rho_{e^-}(\vec{r})]. \tag{F11}$$

2) The electron density that minimizes the functional in Eq. (F11) is the electron density that corresponds to the solution of the Schrödinger equation,

$$E_0 < E[\tilde{\rho}(\vec{r})], \tag{F12}$$

for $\tilde{\rho}(\vec{r}) \neq \rho_{e^-}(\vec{r})$.

The second theorem of Kohn and Hohenberg summarizes the goal of density functional theory as finding the electron density that minimizes the density functional. The density functional can be split into a term $E_{\text{known}}[\rho_{e^-}(\vec{r})]$, which contains all contributions that can be written down in a simple and closed form, and the exchange and correlation functional $E_{\text{XC}}[\rho_{e^-}(\vec{r})]$, i.e.:

$$E[\rho_{e^-}(\vec{r})] = E_{\text{known}}[\rho_{e^-}(\vec{r})] + E_{\text{XC}}[\rho_{e^-}(\vec{r})]. \tag{F13}$$

The known term includes the contributions due to the motion of the electrons, the nuclei-electron interactions, the electron-electron interactions, and the nucleus-nucleus interac-

tions ([159], p. 12),

$$E_{\text{known}}[\rho_{e^-}(\vec{r})] = -\sum_{i=1}^{N_{e^-}} \frac{\tilde{h}^2}{2m_{e^-}} \int \psi_i(\vec{r}) \nabla^2 \psi_i^*(\vec{r}) \mathrm{d}\vec{r} - \frac{1}{4\pi\epsilon_0} \sum_{j=1}^{N_{n^+}} \int \frac{\tilde{q}_j^+ e^2}{d_{i\text{-}j}} \rho_{e^-}(\vec{r}) \mathrm{d}\vec{r} + \frac{1}{4\pi\epsilon_0} \int\int \frac{\rho_i(\vec{r})\rho_j(\vec{r})e^2}{d_{i\text{-}j}} \mathrm{d}\vec{r}\mathrm{d}\vec{r} + \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{N_{n^+}} \sum_{j=i+1}^{N_{n^+}} \frac{\tilde{q}_i^+ \tilde{q}_j^+ e^2}{d_{i\text{-}j}}. \tag{F14}$$

Kohn and Sham suggested a set of single-electron functions that possess the form of the single-electron Schrödinger equations in Eq. (F8) – the so-called Kohn-Sham equations [162],

$$\left[-\frac{\tilde{h}^2}{2m_{e^-}} \nabla^2 + u(\vec{r}) + u_{\text{H}}(\vec{r}) + u_{\text{XC}}(\vec{r}) \right] \psi_i = \varepsilon_i \psi_i, \tag{F15}$$

where $u(\vec{r})$ describes the energy contribution due to the interaction of the electron with the nuclei, $u_{\text{H}}(\vec{r})$ is the single-electron Hartree potential, and $u_{\text{XC}}(\vec{r})$ is the energy contribution due to exchange and correlation. In order to determine the electron distribution, Eq. (F10) requires to evaluate the molecular orbitals, which can be obtained by solving the Kohn-Sham equations in Eq. (F15), which in turn require knowledge about the electron density to determine the Hartree potential in Eq. (F9). Thus, an iterative solution strategy is required ([159], pp. 13–14):

1. Suggest an initial electron density, $\rho_{e^-}(\vec{r})$.
2. Determine the molecular orbitals, ψ_i, by solving the Kohn-Sham equations in Eq. (F15).
3. Use the thus determined molecular orbitals to calculate the electron density, $\tilde{\rho}(\vec{r})$, according to Eq. (F10).
4. Compare the new electron density, $\tilde{\rho}(\vec{r})$, with the previous one, $\rho_{e^-}(\vec{r})$. If $\tilde{\rho}(\vec{r}) \approx \rho_{e^-}(\vec{r})$, stop the iteration, otherwise continue from step 2 using (a modified version of) $\tilde{\rho}(\vec{r})$.

So far, we have neglected to mention how to determine the exchange-correlation interactions in Eq. (F15). The idea behind DFT is that it treats large energy terms (such as the kinetic energy) as accurately as possible, while approximating the smaller ones, such as the exchange-correlation energy. There exist many ways to approximate the exchange-correlation energy. The most prominent ones are the *local density approximation* (LDA), which approximates the exchange-correlation energy of the system under study from a system for which the exact value of the exchange-correlation energy is known – i.e. the uniform electron gas ([159], p. 14),

$$u_{\text{XC}}(\vec{r}) = u_{\text{XC}}^{\text{e-gas}}[\rho_{e^-}(\vec{r})], \tag{F16}$$

or the *generalized gradient approximation* (GGA), which uses additional information about the local gradient of the electron density. But there are also so-called *hybrid functionals* that incorporate portions of the exact exchange energy (often from Hartree-Fock calculations) and

(sometimes more sophisticated) approximations of the correlation energy. The interested reader is referred to the numerous literature on this subject, as a more detailed explanation would go beyond the scope of this overview.

G Section 3: Adsorbed Solution Theories

Adsorbed solution theories aim at predicting the composition of the adsorbed phase, $\{x_i\}$, in multicomponent adsorption from single-component adsorption isotherms. The *ideal adsorbed solution theory* (IAST) is the most popular approach. IAST was first described by Myers and Prausnitz [66], who considered a multicomponent gas at constant p and T. In thermodynamic equilibrium, the chemical potential of every component, i, in the gas and adsorbed phase are equal,

$$\mu_i^{\mathrm{g}}(p_i, T) = \mu_i^{\gamma}(\tilde{\pi}_i, x_i^{\gamma}, T), \tag{G1}$$

where $\mu_i^{\mathrm{g}}(p_i, T)$ is the chemical potential of i in the gas phase (superscript g), $\mu_i^{\gamma}(\tilde{\pi}, x_i^{\gamma}, T)$ is the chemical potential of i in the adsorbed phase (superscript γ), p_i is the partial pressure of i in the gas phase, x_i^{γ} is the mole fraction of i in the adsorbed phase, and $\tilde{\pi}_i$ is the spreading pressure of i. The spreading pressure of any pure species, i, is given by the Gibbs adsorption isotherm in Eq. (9),

$$\mathrm{d}\tilde{\pi}_i = [n_i^{\gamma}]^0 \, \mathrm{d}\mu_i = -\mathrm{d}\sigma, \tag{G2}$$

where $[n_i^{\gamma}]^0$ is the unary adsorption isotherm. The dependency of $\mu_i^{\mathrm{g}}(p_i, T)$ on p_i is determined by the Lewis form of the chemical potential for an ideal gas mixture,

$$\mu_i^{\mathrm{g}}(p_i, T) = \mu_i^0(T) + RT \ln\left(\frac{p_i}{p_0}\right), \tag{G3}$$

where $\mu_i^0(T)$ is the standard chemical potential of the pure component i. Myers and Prausnitz [66] describe the dependency of $\mu_i^{\gamma}(\tilde{\pi}_i, x_i^{\gamma}, T)$ on $\tilde{\pi}_i$ and x_i^{γ} in a similar way:

$$\mu_i^{\gamma}(\tilde{\pi}_i, x_i^{\gamma}, T) = \mu_i^0(T) + RT \ln x_i^{\gamma} + RT \ln\left(\frac{p_i^*(\tilde{\pi}_i)}{p_0}\right), \tag{G4}$$

where $p_i^*(\tilde{\pi}_i)$ is the pressure for adsorption of the pure component, i, which yields the same spreading pressure as that for the mixture,

$$\tilde{\pi}_{\mathrm{mix}} = \tilde{\pi}_1(p_1^*) = \tilde{\pi}_2(p_2^*) = \cdots = \tilde{\pi}_N(p_N^*). \tag{G5}$$

Substituting Eq. (G3) and Eq. (G4) in Eq. (G2) yields the basic equation of IAST,

$$p_i = x_i^{\gamma} p_i^*, \tag{G6}$$

which is analogous to Raoult's law for the equilibrium of an ideal liquid mixture with its vapor phase. Eq. (G5) and Eq. (G6) in conjunction with:

$$\sum_{i=1}^{N} x_i^{\gamma} = 1 \tag{G7}$$

yield $2N$ equations for $2N$ unknowns (i.e. $\{x_i^\gamma\}$ and $\{p_i^*\}$). The set of equations is commonly solved by iteration for a specified number of partial pressures. Solving these equations requires determining the spreading pressure of every component, i, by integration of Eq. (G2) from $p = 0$ to $p = p_i^*$,

$$\tilde{\pi}_i = RT \int_0^{p_i^*} [n_i^\gamma]^0 \, \mathrm{d} \ln p = RT \int_0^{p_i^*} \frac{[n_i^\gamma]^0}{p} \mathrm{d}p, \tag{G8}$$

where $\mathrm{d}\mu = RT \mathrm{d} \ln p$. In practice, the isotherm will be given by a set of discrete data points $(p_j, [n_i^\gamma]_j^0)$. These data points are used to solve Eq. (G8) either analytically – by assuming that the data points follow a continuous function – or by numerical quadrature. We will illustrate these approaches in the following examples. For the analytical solution, we will assume that the data points follow *Langmuir's adsorption equation* [163],

$$[n_i^\gamma]^0 = [n_i^\gamma]^\mathrm{m} \frac{K_\mathrm{L} p}{1 + K_\mathrm{L} p}, \tag{G9}$$

where $[n_i^\gamma]^\mathrm{m}$ is the saturation capacity of i and K_L is Langmuir's constant. Substituting the Langmuir adsorption equation is Eq. (G8) yields [63]:

$$\tilde{\pi}_i = RT \int_0^{p_i^*} [n_i^\gamma]^\mathrm{m} \frac{K_\mathrm{L}}{1 + K_\mathrm{L} p} \mathrm{d}p, \tag{G10}$$

$$= RT \, [n_i^\gamma]^\mathrm{m} \ln(1 + K_\mathrm{L} p_i^*). \tag{G11}$$

For numerical quadrature, we split the integral in Eq. (G8) into three terms [83],

$$\tilde{\pi}_i = RT \int_0^{p_1} \frac{[n_i^\gamma]^0}{p} \mathrm{d}p + RT \sum_{j=1}^{k-1} \int_{p_j}^{p_{j+1}} \frac{[n_i^\gamma]^0}{p} \mathrm{d}p + RT \int_{p_k}^{p_i^*} \frac{[n_i^\gamma]^0}{p} \mathrm{d}p, \tag{G12}$$

where the first integral extrapolates the data points for pressures below p_1, the second term contains the integrals that go from data point $j = 1$ to $j = k$ with $p_k \leq p_i^* \leq p_{k+1}$, and the third integral extrapolates from p_k to p_i^*. These integrals can be solved in various ways. For example, the first integral can be solved by approximating the adsorption isotherm to follow Henry's law for $p < p_1$,

$$[n_i^\gamma]^0 = K_\mathrm{H} p, \tag{G13}$$

where K_H is Henry's adsorption constant. In doing so, the first integral on the right-hand side of Eq. (G12) reduces to:

$$RT \int_0^{p_1} \frac{[n_i^\gamma]^0}{p} \mathrm{d}p = K_\mathrm{H} p_1. \tag{G14}$$

The second term can be solved by standard numerical quadrature, e.g. by the trapezoid rule, and the third integral can be solved by linear interpolation of the adsorption isotherm

between p_k and p_{k+1}.

A disadvantage of IAST is the assumption of an ideal adsorbed phase. This assumption might be valid if the adsorbed phase is almost pure or consists of components with similar chemistry, such as 2-methylpropane and 3-methylpropane, but is often inaccurate for mixtures of components with strong interactions, such as water and benzene. In these cases, the *real adsorbed solution theory* (RAST) [66] is often applied. RAST introduces an activity coefficient, $\breve{a}_i^\gamma$, in Eq. (G6) to account for the nonideality of the adsorbed phase,

$$\breve{f}_i = \breve{a}_i^\gamma x_i^\gamma p_i^*, \tag{G15}$$

where $\breve{f}_i$ is the fugacity of i. In contrast to IAST, RAST is not capable of predicting the composition of the adsorbed phase, because experimental or simulated adsorption data is required to determine $\breve{a}_i^\gamma$ [164].

H Section 3: Compositions of the Liquid Mixtures

Tab. H1: Vapor-liquid equilibrium data according to the UNIFAC (Do) calculations for water/ethanol. Presented are the liquid phase mole fraction of water (x'_{water}), and the corresponding partial fugacities of water ($\check{f}_{\text{water}}$) or ethanol ($\check{f}_{\text{ethanol}}$).

x'_{water}	$\check{f}_{\text{water}}$ [Pa]	$\check{f}_{\text{ethanol}}$ [Pa]
0.05	390.1	7554.9
0.1	733.2	7193.8
0.15	1035.0	6854.0
0.2	1298.7	6534.3
0.25	1528.3	6233.7
0.3	1727.0	5952.0
0.35	1899.0	5688.0
0.4	2047.1	5437.9
0.45	2174.2	5200.9
0.5	2284.4	4974.6
0.55	2379.9	4754.1
0.6	2464.4	4536.6
0.65	2540.4	4312.6
0.7	2611.0	4072.0
0.75	2680.4	3798.6
0.8	2752.1	3465.9
0.85	2830.9	3030.1
0.9	2922.0	2415.0
0.95	3032.7	1488.3

Tab. H2: Vapor-liquid equilibrium data according to the UNIFAC (Do) calculations for water/n-butanol. Presented are the liquid phase mole fraction of water (x'_{water}), and the corresponding partial fugacities of water ($\check{f}_{\mathrm{water}}$) or n-butanol ($\check{f}_{\mathrm{butanol}}$).

x'_{water}	$\check{f}_{\mathrm{water}}$ [Pa]	$\check{f}_{\mathrm{butanol}}$ [Pa]
0.05	1293.5	839.5
0.1	1806.6	801.4
0.15	2235.9	766.1
0.2	2587.5	734.5
0.25	2869.4	706.6
0.3	3083.6	682.4
0.35	3103.6	680.4
0.3555	3103.6	680.4
0.9708	3120.0	542.0
0.98	3145.1	322.9
0.99	3173.2	0.0

Tab. H3: Vapor-liquid equilibrium data according to the UNIFAC (Do) calculations for water/n-hexanol. Presented are the liquid phase mole fraction of water (x'_{water}), and the corresponding partial fugacities of water ($\check{f}_{\mathrm{water}}$) or n-hexanol ($\check{f}_{\mathrm{hexanol}}$).

x'_{water}	$\check{f}_{\mathrm{water}}$ [Pa]	$\check{f}_{\mathrm{hexanol}}$ [Pa]
0.01	207.6	108.4
0.02	409.0	107.0
0.03	606.1	105.9
0.04	799.0	105.0
0.05	985.1	103.9
0.1	1846.0	99.0
0.15	2582.5	94.5
0.1971	3166.5	90.5
0.9978	3166.5	90.5
0.998	3166.8	81.2
0.999	3169.6	42.4

Tab. H4: Vapor-liquid equilibrium data according to the UNIFAC (Do) calculations for ethanol/n-butanol. Presented are the liquid phase mole fraction of ethanol (x'_{ethanol}), and the corresponding partial fugacities of ethanol ($\check{f}_{\mathrm{ethanol}}$) or n-butanol ($\check{f}_{\mathrm{butanol}}$).

x'_{ethanol}	$\check{f}_{\mathrm{ethanol}}$ [Pa]	$\check{f}_{\mathrm{butanol}}$ [Pa]
0.05	416.9	883.1
0.1	827.3	832.7
0.15	1240.7	789.3
0.2	1647.4	742.6
0.25	2053.2	696.9
0.3	2458.1	651.9
0.35	2854.8	605.2
0.4	3250.7	559.3
0.45	3646.2	513.8
0.5	4041.4	468.6
0.55	4436.7	423.3
0.6	4823.0	377.0
0.65	5208.7	331.3
0.7	5604.3	285.7
0.75	5990.8	239.2
0.8	6377.5	192.5
0.85	6764.2	145.8
0.9	7152.1	97.9
0.95	7550.6	49.4

Tab. H5: Vapor-liquid equilibrium data according to the UNIFAC (Do) calculations for n-butanol/n-hexanol. Presented are the liquid phase mole fraction of n-butanol (x'_{butanol}), and the corresponding partial fugacities of n-butanol ($\check{f}_{\mathrm{butanol}}$) or n-hexanol ($\check{f}_{\mathrm{hexanol}}$).

x'_{butanol}	$\check{f}_{\mathrm{butanol}}$ [Pa]	$\check{f}_{\mathrm{hexanol}}$ [Pa]
0.05	47.2	103.8
0.1	94.0	98.0
0.15	141.8	93.2
0.2	188.4	87.6
0.25	234.9	82.1
0.3	280.6	76.4
0.35	327.8	71.2
0.4	374.2	65.8
0.45	419.7	60.3
0.5	466.1	54.9
0.55	511.6	49.4
0.6	558.9	44.1
0.65	604.4	38.6
0.7	649.9	33.1
0.75	696.3	27.7
0.8	741.8	22.2
0.85	787.4	16.6
0.9	833.8	11.2
0.95	879.4	5.6

Tab. H6: Vapor-liquid equilibrium data according to the UNIFAC (Do) calculations for water/ethanol/n-butanol. Presented are the liquid phase mole fraction of water (x'_{water}) and ethanol (x'_{ethanol}), and the corresponding partial fugacities of water ($\check{f}_{\mathrm{water}}$), ethanol ($\check{f}_{\mathrm{ethanol}}$), or n-butanol ($\check{f}_{\mathrm{butanol}}$).

x'_{water}	x'_{ethanol}	$\check{f}_{\mathrm{water}}$ [Pa]	$\check{f}_{\mathrm{ethanol}}$ [Pa]	$\check{f}_{\mathrm{butanol}}$ [Pa]
0.1	0.1	1231.5	797.3	751.2
0.2	0.1	2116.3	773.6	680.4
0.1	0.2	1168.3	1587.7	662.1
0.3	0.1	2699.0	760.5	622.5
0.2	0.2	1996.6	1548.0	593.4
0.1	0.3	1105.9	2374.8	573.2
0.3	0.2	2531.5	1533.5	536.0
0.2	0.3	1877.9	2326.7	505.4
0.1	0.4	1042.9	3158.6	483.5
0.4	0.2	2824.4	1558.4	489.7
0.3	0.3	2365.6	2325.5	445.9
0.2	0.4	1759.1	3114.7	414.7
0.1	0.5	980.3	3942.9	392.9
0.4	0.3	2622.6	2397.4	393.0
0.3	0.4	2201.1	3146.4	350.4
0.2	0.5	1642.3	3918.6	321.2
0.1	0.6	918.1	4732.0	299.9
0.5	0.3	2707.2	2595.5	337.3
0.4	0.4	2425.0	3301.3	283.7
0.3	0.5	2039.6	4009.9	247.5
0.2	0.6	1525.7	4747.2	222.1
0.1	0.7	855.5	5530.5	205.0
0.7	0.2	2821.8	2319.3	350.9
0.6	0.3	2679.6	3040.9	251.4
0.5	0.4	2491.3	3680.9	192.9
0.4	0.5	2232.8	4299.7	156.5
0.3	0.6	1881.4	4935.9	132.7
0.2	0.7	1411.1	5612.8	116.4
0.1	0.8	794.3	6347.6	105.1

I Section 3: Monte Carlo cycles

Tab. I1: Number of Monte Carlo cycles for equilibrium and production run for the different simulation campaigns.

Campaign	**Equilibrium cycles**	**Production cycles**
water (1 Pa – 300 Pa)	54,641,00	3,000,000
water (500 Pa – 875 Pa)	99,854,400	5,000,000
water (900 Pa – 3500 Pa)	5,286,300	70,000
ethanol	2,500,000	400,000
n-butanol	68,010,000	600,000
n-hexanol	61,660,000	850,000
water/ethanol	1,282,800	40,000
water/n-butanol	569,200	25,000
water/n-hexanol	528,000	30,000
ethanol/n-butanol	15,664,000	140,000
n-butanol/n-hexanol	17,000,000	180,000
water/ethanol/n-butanol	724,800	15,000

J Section 3: Method for Measuring Vapor Adsorption Isotherms for CAU-10

CAU-10 was synthesized following the procedure described by Fröhlich et al. [104]. All chemicals were obtained from Carl Roth and used without further purification. The X-ray powder diffractogram of the CAU-10 sample used in the vapor adsorption experiments is shown in Fig. J1 and compared with the diffractogram obtained with the RASPA software package from the CAU-10 model used in the GCMC simulations. The X-ray diffractogram in the experiment was measured with an XRD 3003 TT diffractometer in conjunction with a Meteor0D energy dispersive detector (GE Inspection Technologies) using monochromatic Cu-K_α radiation. The measurement was performed at ambient temperature in the range $6° \leq 2\ \text{Theta} \leq 40°$ with a step size of 0.02°.

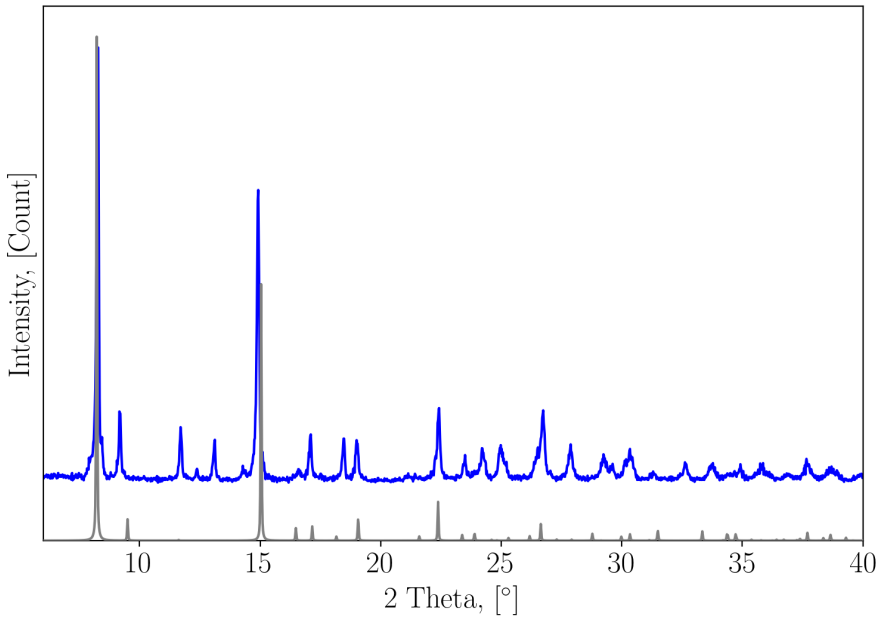

Fig. J1: X-ray powder diffractogram of the CAU-10 sample used in the vapor adsorption experiments (blue) or X-ray powder diffractogram calculated from CAU-10 model used in the GCMC simulations (gray).

The characteristic Brunauer-Emmett-Teller (BET) surface area was determined by Dr.-Ing. Mandy Klauck (HTW Dresden) by measuring the nitrogen adsorption isotherms with a 3Flex (Micromeretics) at 77.3 K. Nitrogen isotherms were measured according to the methodology described in published work [165]. The calculation of the BET surface area was based on isotherm points in the range of $0.0025 \leq p/p_0 \leq 0.02$. Three nitrogen isotherms were measured, and a mean BET surface area of 680.1 ± 1.7 m^2/g was obtained (Reinsch et al. measured BET surface areas of 564 – 656 m^2/g [5]). Vapor adsorption isotherms for water, ethanol, and n-butanol were measured by Dr. Jens Möllmer of the INC Leipzig with a Belsorp-max (MicrotracBEL) at 298.15 K. The methodology followed published work [165].

K Section 3: Fitted Parameters

Tab. K1: Parameters fitted to the Dual-Site Langmuir-Sips adsorption model.

	Ethanol	**n-Butanol**	**n-Hexanol**
$\Gamma^{\mathrm{m}}_{i,1}$ [mmol/g]	4.83	1.18	1.21
$\Gamma^{\mathrm{m}}_{i,2}$ [mmol/g]	0.0	1.26	1.20
$b_{i,1}$ [1/Pa]	8300.34	6822.87	61.52
$b_{i,2}$ [1/Pa]	0.0	45.52	195947.18

Tab. K2: Parameters fitted to the Bi-Langmuir adsorption model for increasing liquid phase mole fraction of ethanol in the ternary mixture. Result in unit mmol/g.

	$x'_{\mathrm{EtOH}} = 0$ %	$x'_{\mathrm{EtOH}} = 10$ %	$x'_{\mathrm{EtOH}} = 30$ %	$x'_{\mathrm{EtOH}} = 50$ %	$x'_{\mathrm{ButOH}} = 0$ %
$\check{h}_1$	$4.29 \cdot 10^{-3}$	487.05	956.81	243.22	202.58
$\check{h}_2$	44,836.69	0.75	0.26	1.26	1.81
K_1	2.36	2.14	2.58	1.53	1.25
K_2	1.00	1.94	2.31	1.42	1.19

Tab. K3: Parameters fitted to the Bi-Langmuir adsorption model for increasing liquid phase mole fraction of n-butanol in the ternary mixture. Result in unit mmol/g.

	$x'_{\mathrm{ButOH}} = 0$ %	$x'_{\mathrm{ButOH}} = 20$ %	$x'_{\mathrm{ButOH}} = 40$ %	$x'_{\mathrm{ButOH}} = 50$ %	$x'_{\mathrm{EtOH}} = 0$ %
$\check{h}_1$	202.58	234.81	199.80	295.28	$4.29 \cdot 10^{-3}$
$\check{h}_2$	1.81	0.75	0.26	1.26	44,836.69
K_1	1.25	1.35	1.28	1.22	2.36
K_2	1.19	1.28	1.22	1.18	1.00

Tab. K4: Parameters fitted to the Bi-Langmuir adsorption model for increasing liquid phase mole fraction of water in the ternary mixture. Result in unit mmol/g.

	$x'_{\mathrm{H_2O}} = 0$ %	$x'_{\mathrm{H_2O}} = 10$ %	$x'_{\mathrm{H_2O}} = 20$ %	$x'_{\mathrm{H_2O}} = 30$ %	$x'_{\mathrm{ButOH}} = 0$ %
$\check{h}_1$	537.78	0.47	470.24	282.94	215.89
$\check{h}_2$	0.17	4.94	0.12	0.62	1.74
K_1	1.83	63.35	0.35	0.65	0.8
K_2	1.68	0.07	0.42	0.70	0.84

L Section 3: Radial Distribution Functions

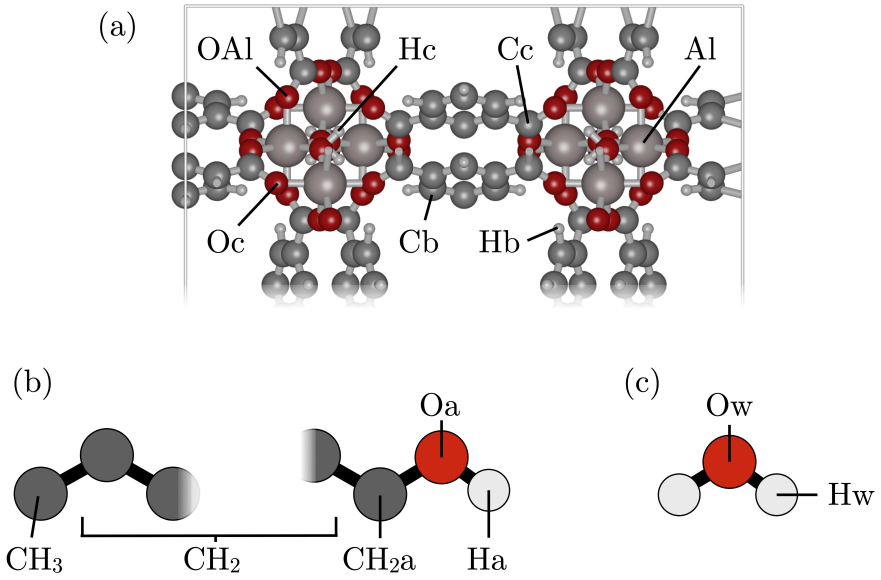

Fig. L1: Interaction sites in (a) CAU-10, (b) alcohols, or (c) water for which RDFs were determined. Coloring of atoms following the Jmol coloring scheme – i.e. aluminum - pink swan, carbon - gray, hydrogen - white, oxygen - red.

L.1 Single-Component Adsorption

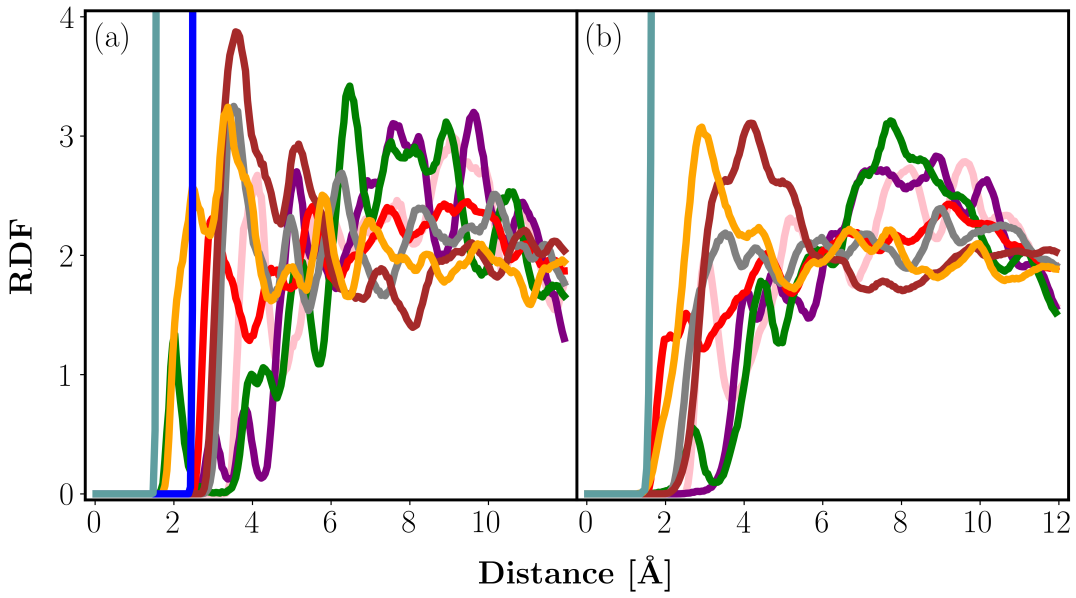

Fig. L2: RDFs of distances for (a) Ow or (b) Hw in water *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), Ow (blue), or Hw (cadet blue) in CAU-10 at 298.15 K and $p/p_0 = 0.2$ (800 Pa).

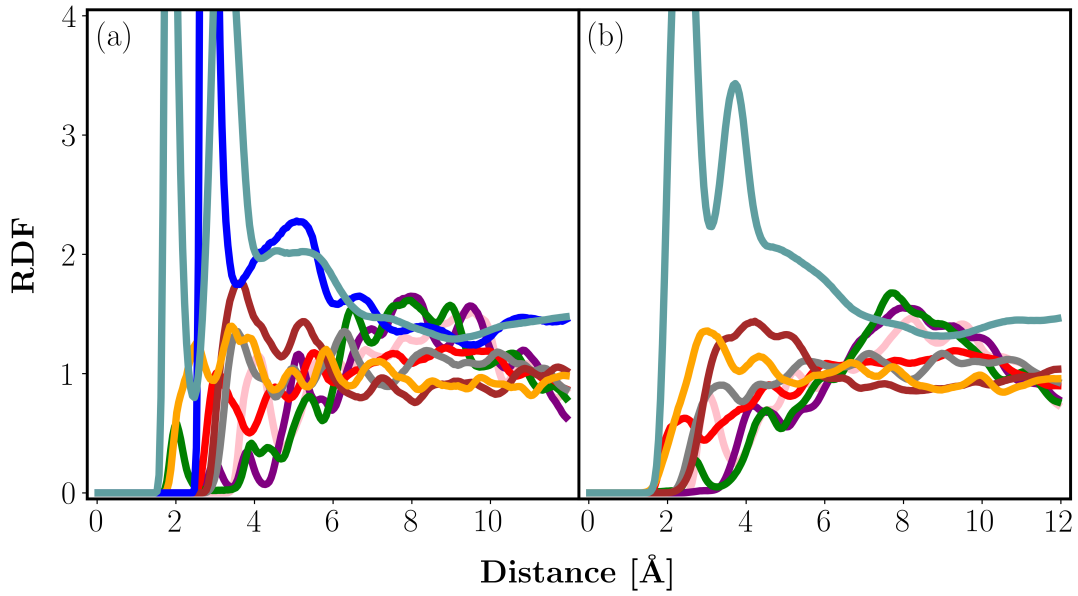

Fig. L3: RDFs of distances for (a) Ow or (b) Hw in water *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), Ow (blue), or Hw (cadet blue) in CAU-10 at 298.15 K and $p/p_0 = 0.28$ (875 Pa).

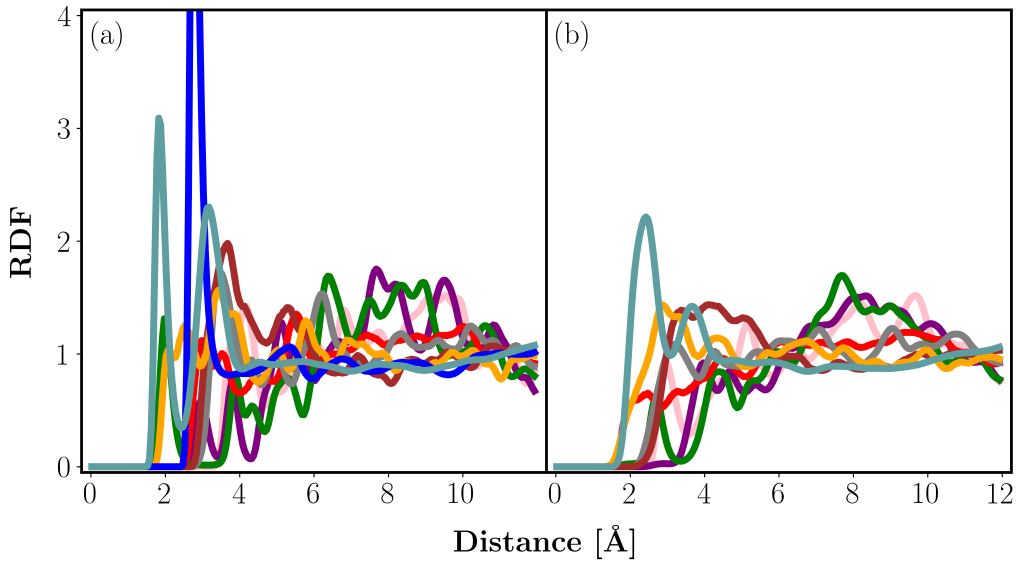

Fig. L4: RDFs of distances for (a) Ow or (b) Hw in water *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), Ow (blue), or Hw (cadet blue) in CAU-10 at 298.15 K and $p/p_0 = 0.95$ (3000 Pa).

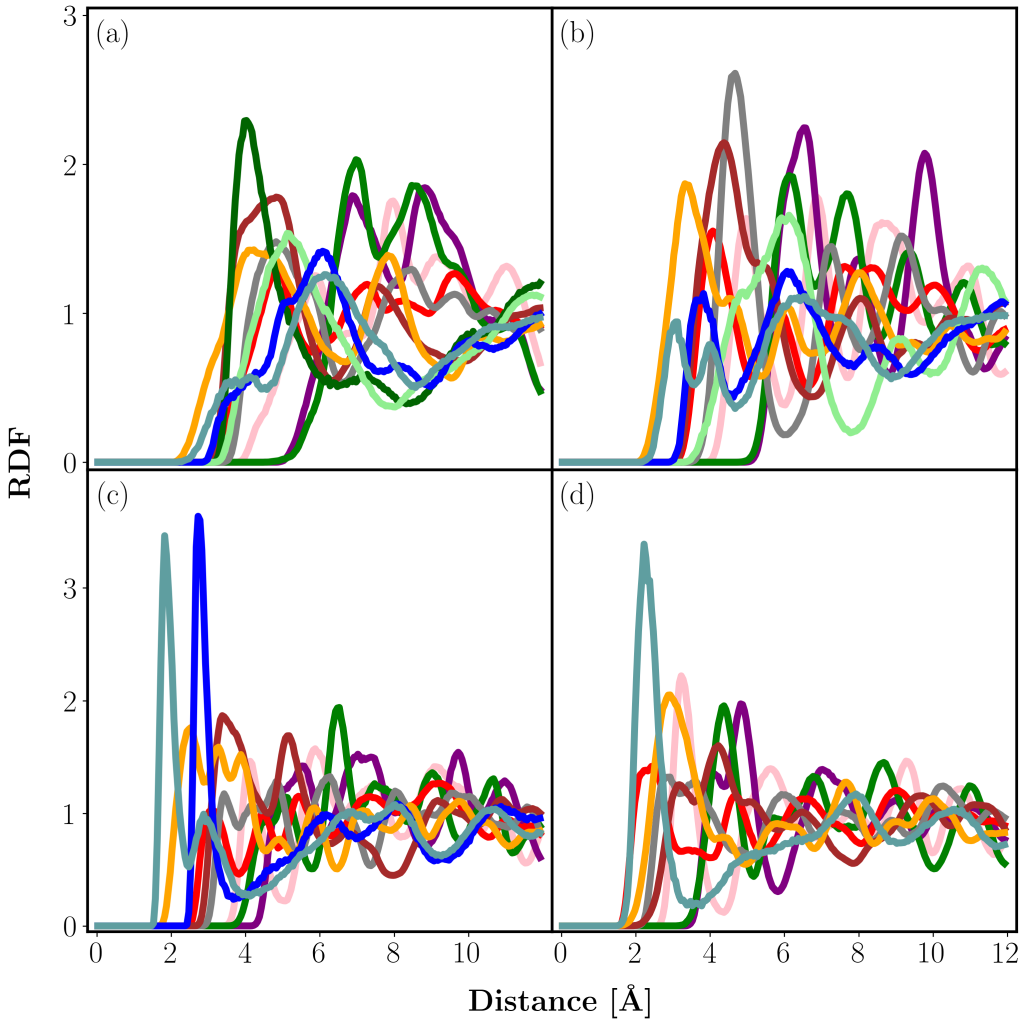

Fig. L5: RDFs of distances for (a) CH_3, (b) CH_2, (c) Oa, or (d) Ha in ethanol *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), CH_3 (dark green), CH_2 (light green), Oa (blue), or Ha (cadet blue) in CAU-10 at 298.15 K and $p/p_0 = 6 \cdot 10^{-3}$ (50 Pa).

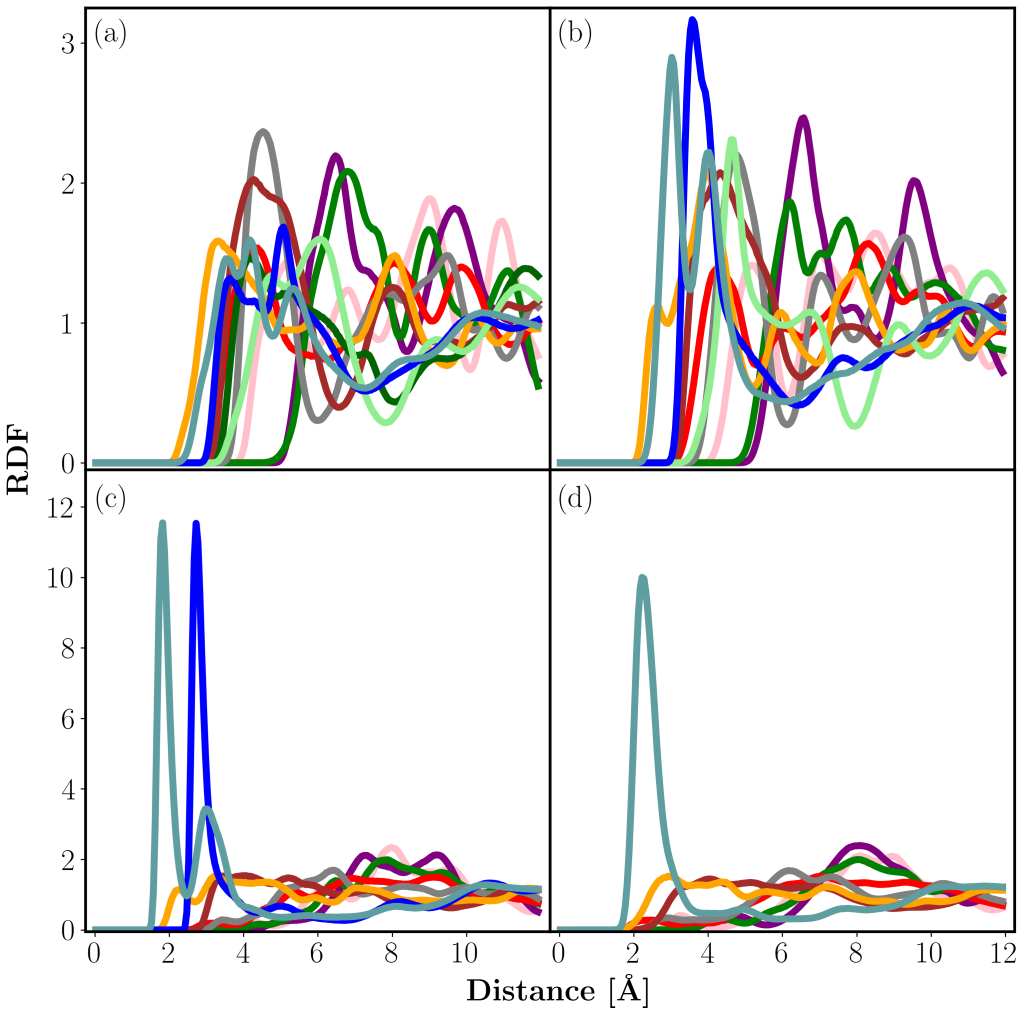

Fig. L6: RDFs of distances for (a) CH_3, (b) CH_2a, (c) Oa, or (d) Ha in ethanol *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), CH_3 (dark green), CH_2a (light green), Oa (blue), or Ha (cadet blue) in CAU-10 at 298.15 K and $p/p_0 = 0.88$ (7000 Pa).

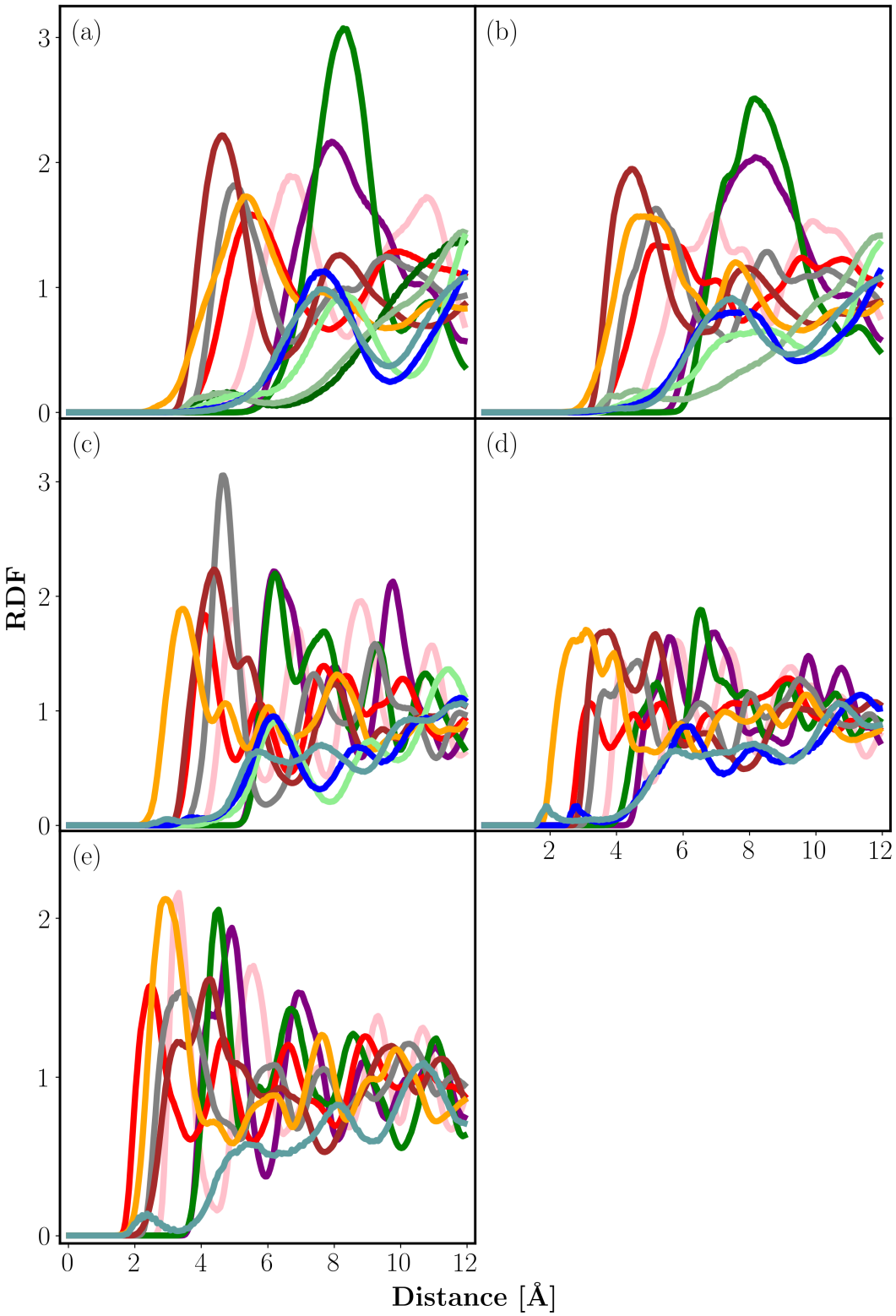

Fig. L7: RDFs of distances for (a) CH_3, (b) CH_2, (c) CH_2a, (d) Oa, or (e) Ha in n-butanol *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), CH_3 (dark green), CH_2 (dark sea green), CH_2a (light green), Oa (blue), or Ha (cadet blue) in CAU-10 at 298.15 K and $p/p_0 = 1.1 \cdot 10^{-3}$ (1 Pa).

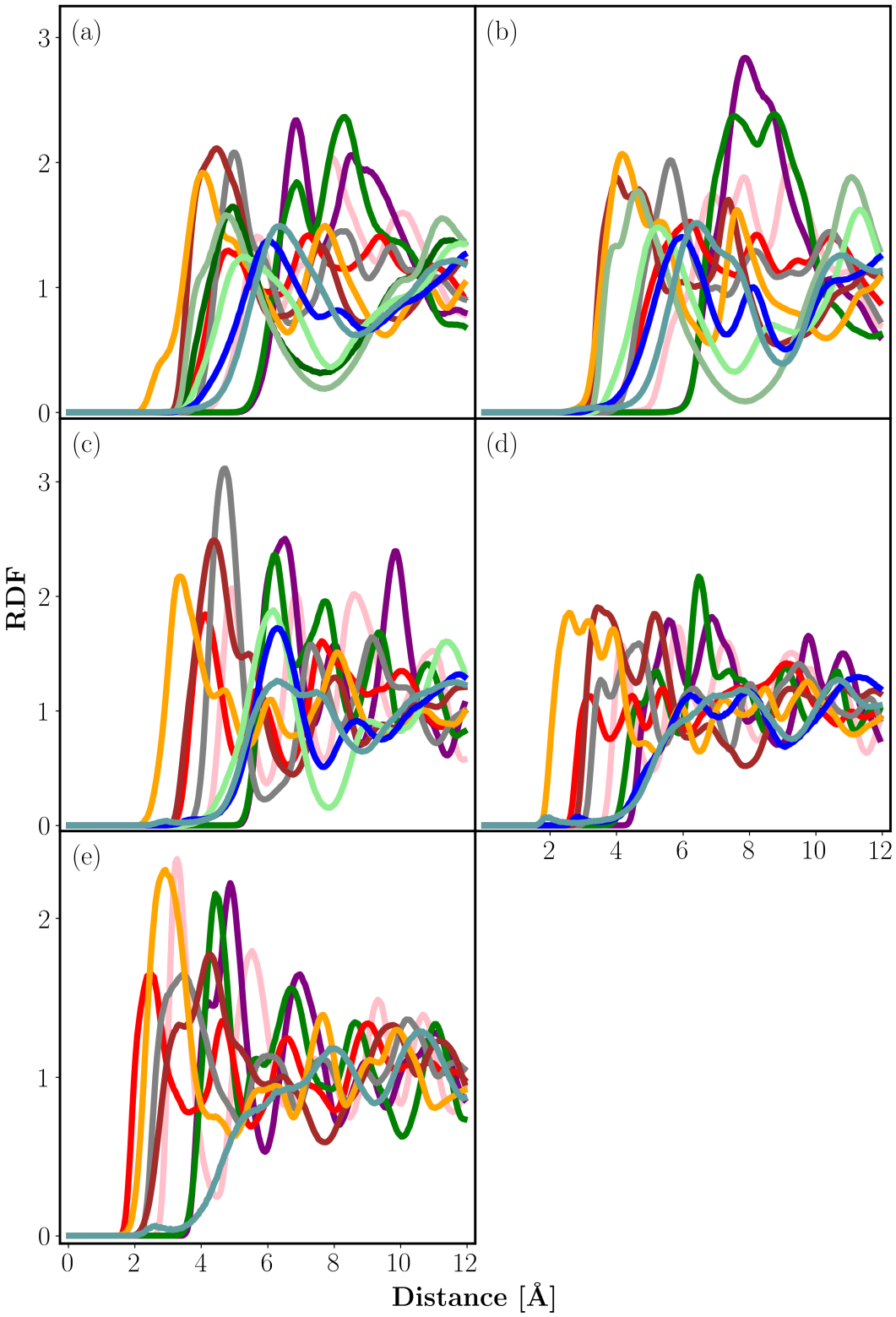

Fig. L8: RDFs of distances for (a) CH_3, (b) CH_2, (c) CH_2a, (d) Oa, or (e) Ha in n-butanol *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), CH_3 (dark green), CH_2 (dark sea green), CH_2a (light green), Oa (blue), or Ha (cadet blue) in CAU-10 at 298.15 K and $p/p_0 = 0.54$ (500 Pa).

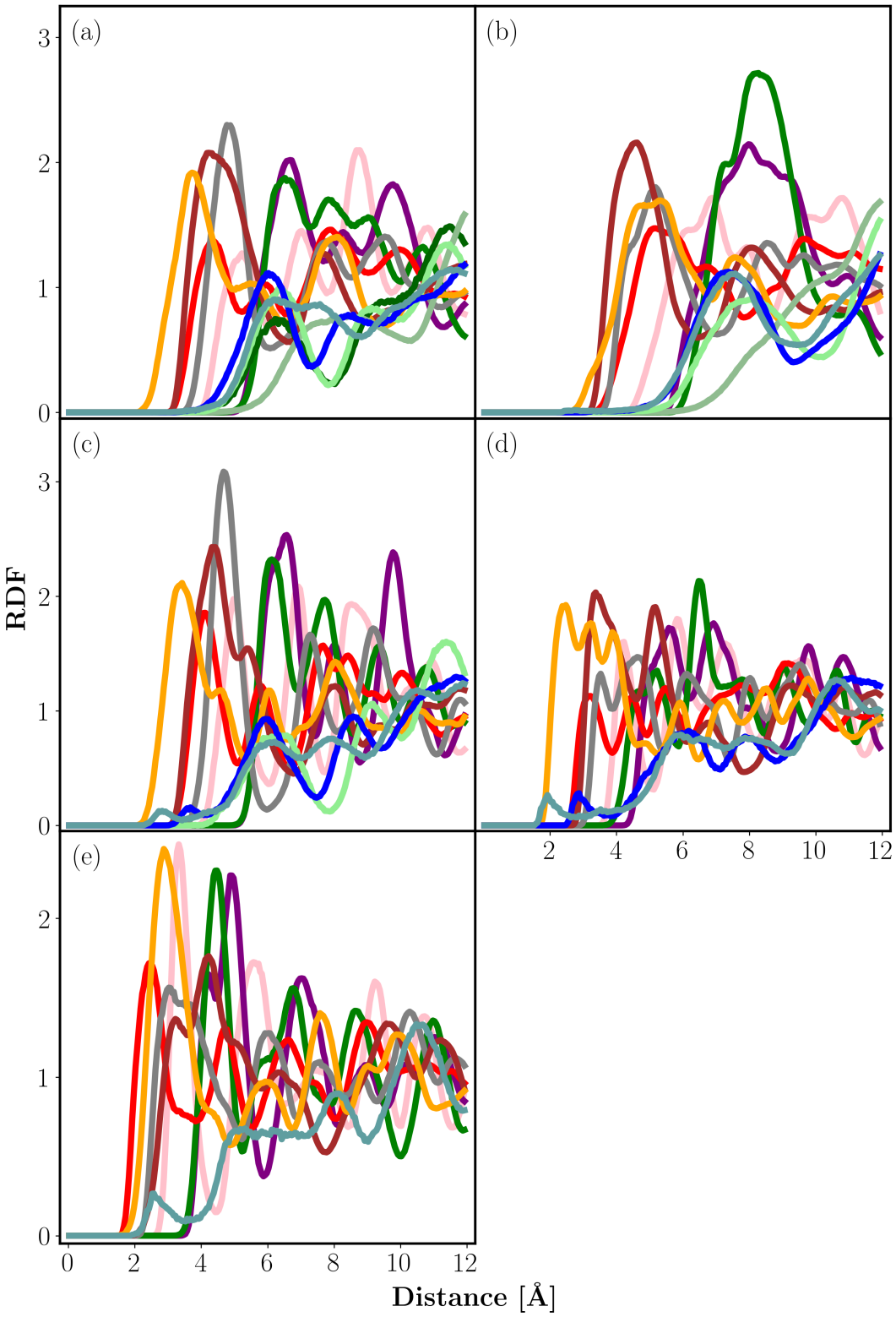

Fig. L9: RDFs of distances for (a) CH_3, (b) CH_2, (c) CH_2a, (d) Oa, or (e) Ha in n-hexanol *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), CH_3 (dark green), CH_2 (dark sea green), CH_2a (light green), Oa (blue), or Ha (cadet blue) in CAU-10 at 298.15 K and $p/p_0 = 9.2 \cdot 10^{-5}$ (0.01 Pa).

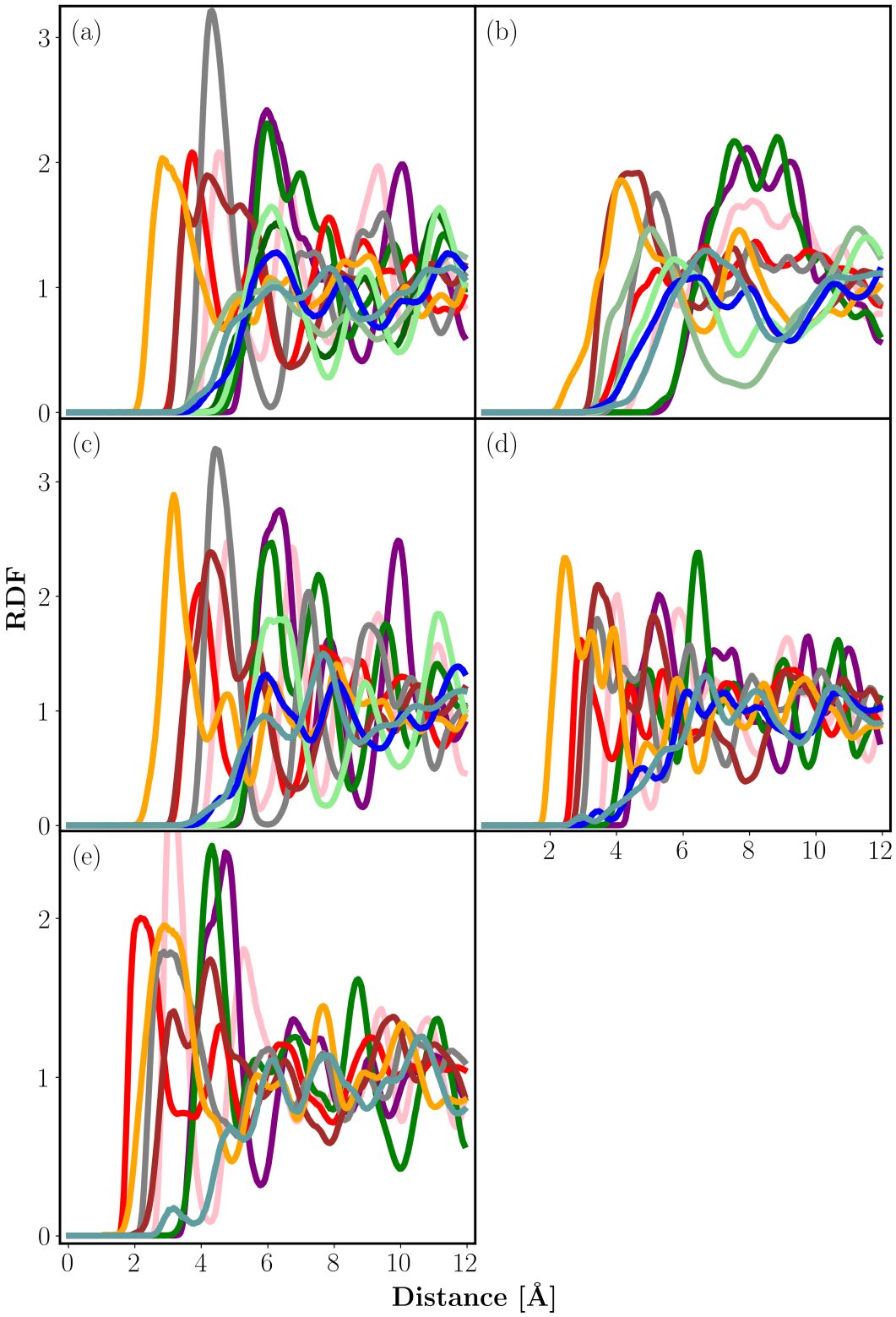

Fig. L10: RDFs of distances for (a) CH_3, (b) CH_2, (c) CH_2a, (d) Oa, or (e) Ha in n-hexanol *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), CH_3 (dark green), CH_2 (dark sea green), CH_2a (light green), Oa (blue), or Ha (cadet blue) in CAU-10 at 298.15 K and $p/p_0 = 0.92$ (500 Pa).

L.2 Two-Component Adsorption

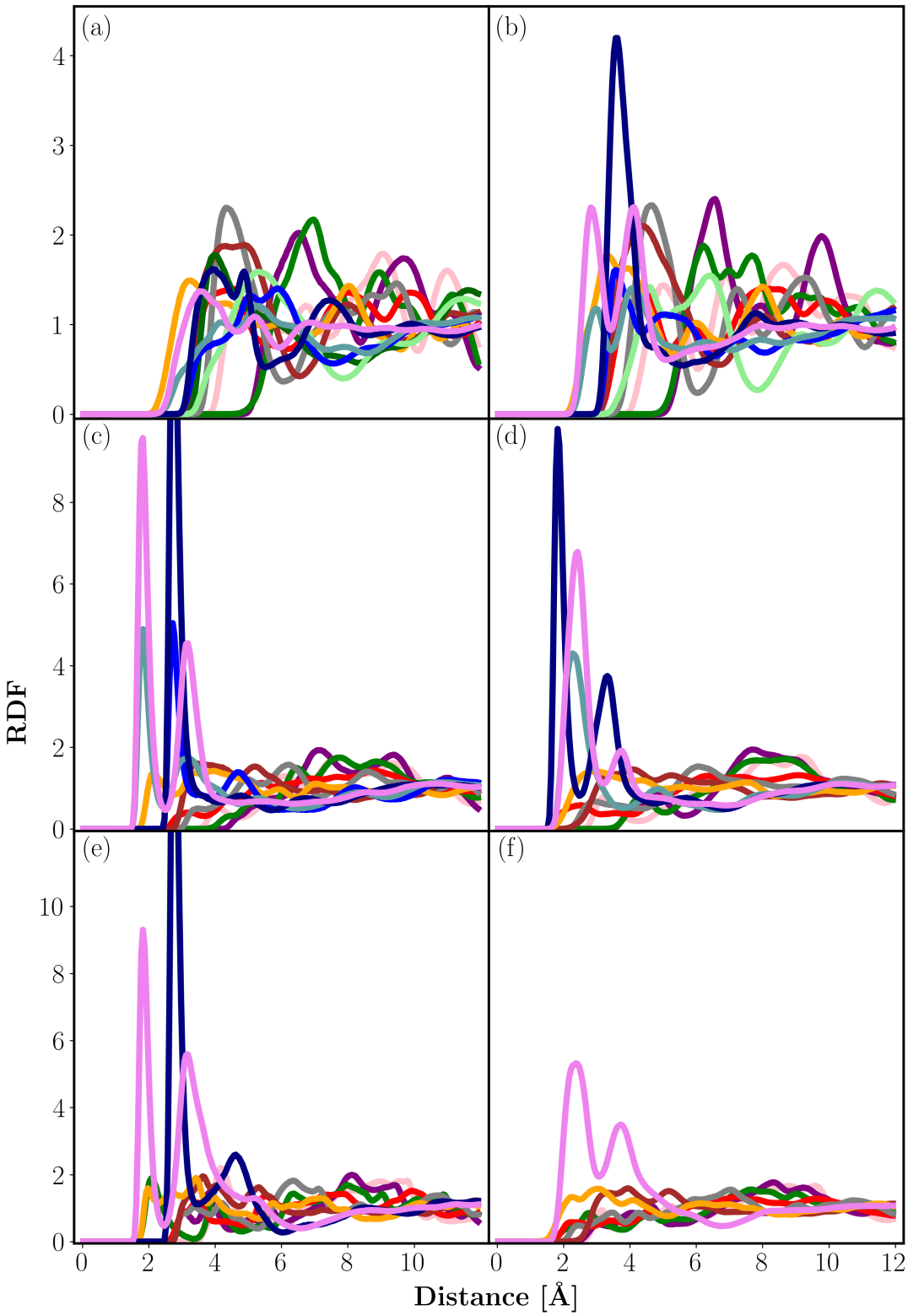

Fig. L11: RDFs of distances for interaction pairs observed for water/ethanol adsorption in CAU-10 at 298.15 K and $x'_{\text{water}} = 0.1$. RDFs are shown for (a) CH_3, (b) CH_2a, (c) Oa, (d) Ha, (e) Ow, or (f) Hw *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), CH_3 (dark green), CH_2a (light green), Oa (blue), Ha (cadet blue), Ow (navy), or Hw (violet).

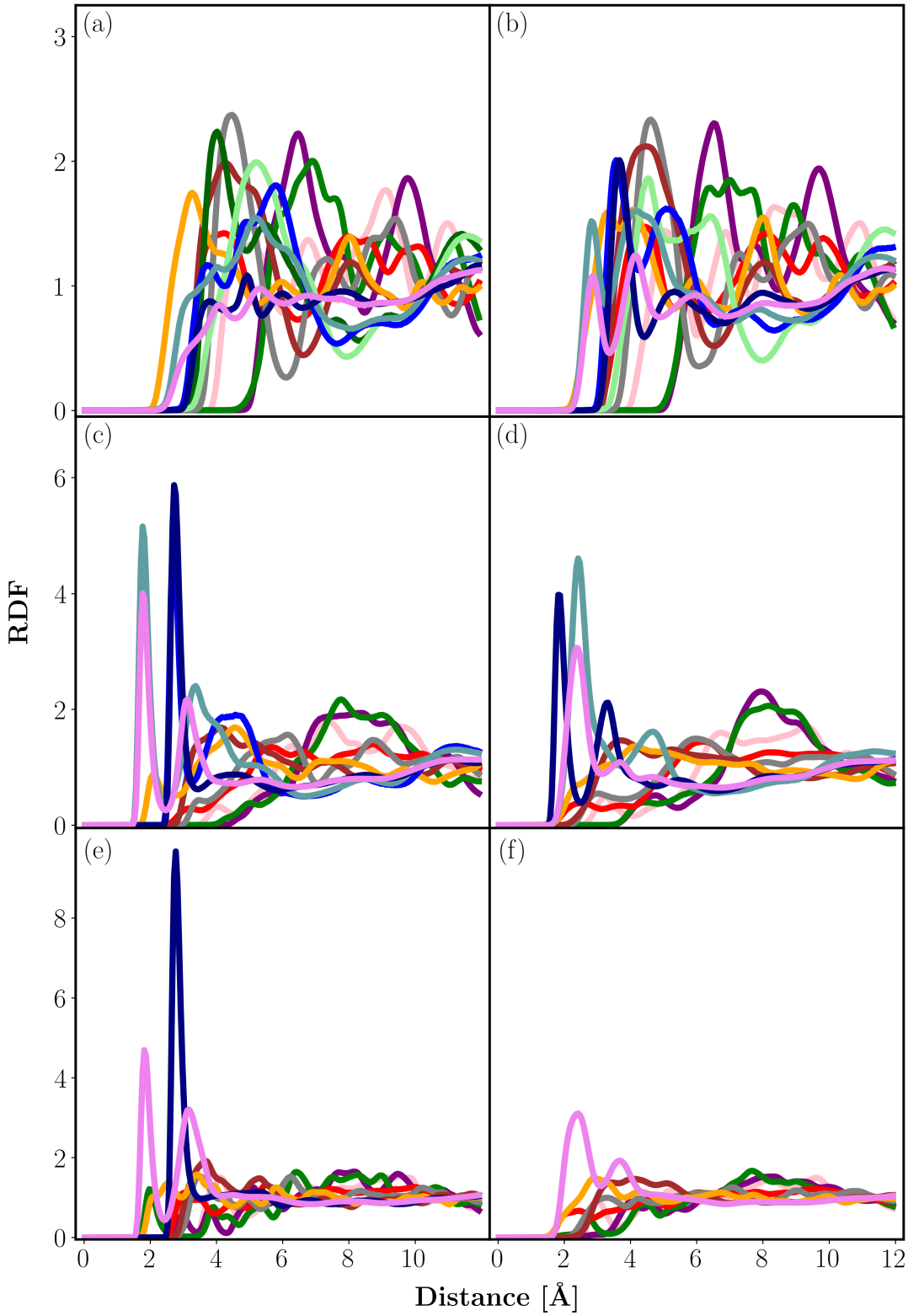

Fig. L12: RDFs of distances for interaction pairs observed for water/ethanol adsorption in CAU-10 at 298.15 K and $x'_{\text{water}} = 0.5$. RDFs are shown for (a) CH_3, (b) CH_2a, (c) Oa, (d) Ha, (e) Ow, or (f) Hw *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), CH_3 (dark green), CH_2a (light green), Oa (blue), Ha (cadet blue), Ow (navy), or Hw (violet).

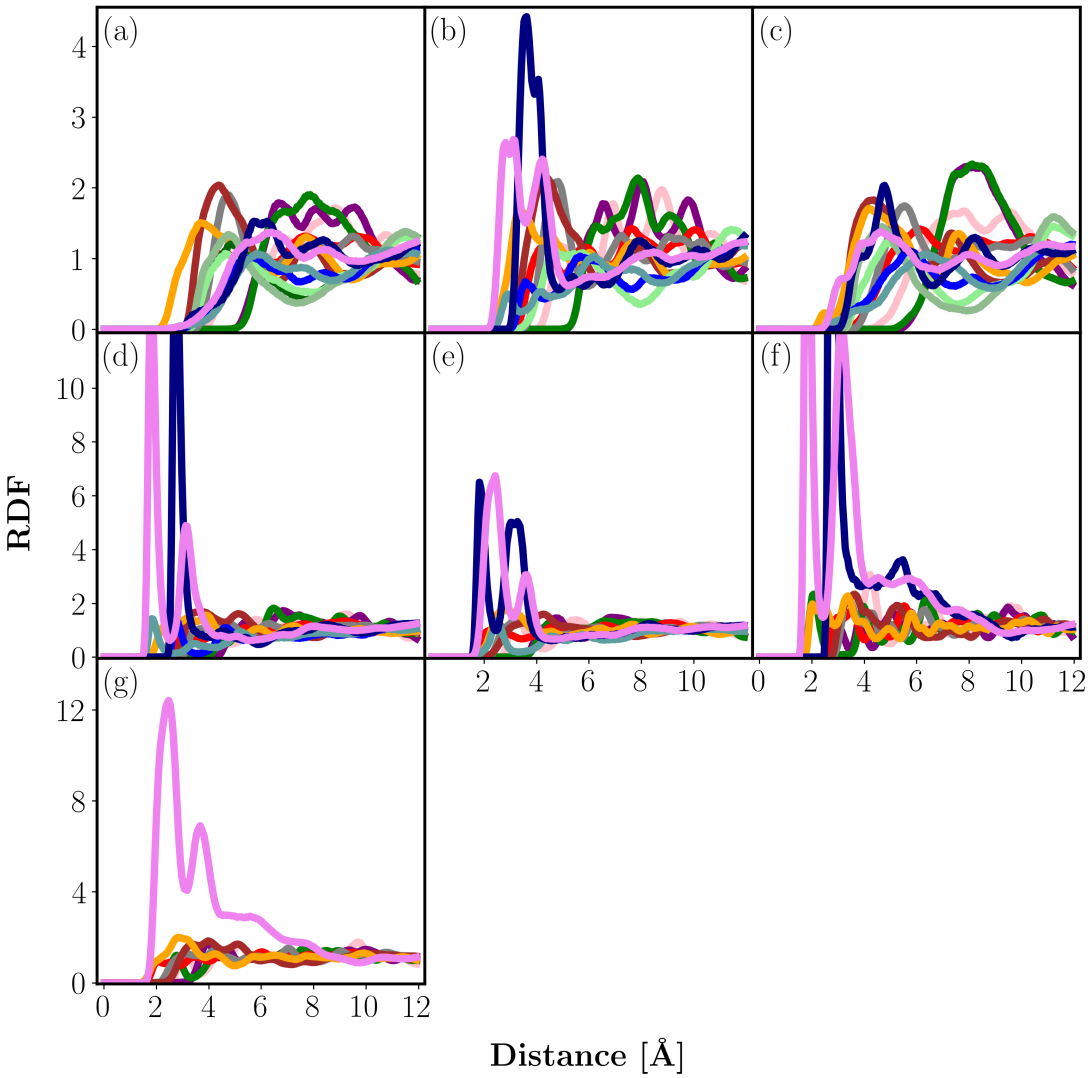

Fig. L13: RDFs of distances for interaction pairs observed for water/n-butanol adsorption in CAU-10 at 298.15 K and $x'_{\text{water}} = 0.05$. RDFs are shown for (a) CH_3, (b) CH_2, (c) CH_2a, (d) Oa, (e) Ha, (f) Ow, or (g) Hw *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), CH_3 (dark green), CH_2 (dark sea green), CH_2a (light green), Oa (blue), Ha (cadet blue), Ow (navy), or Hw (violet).

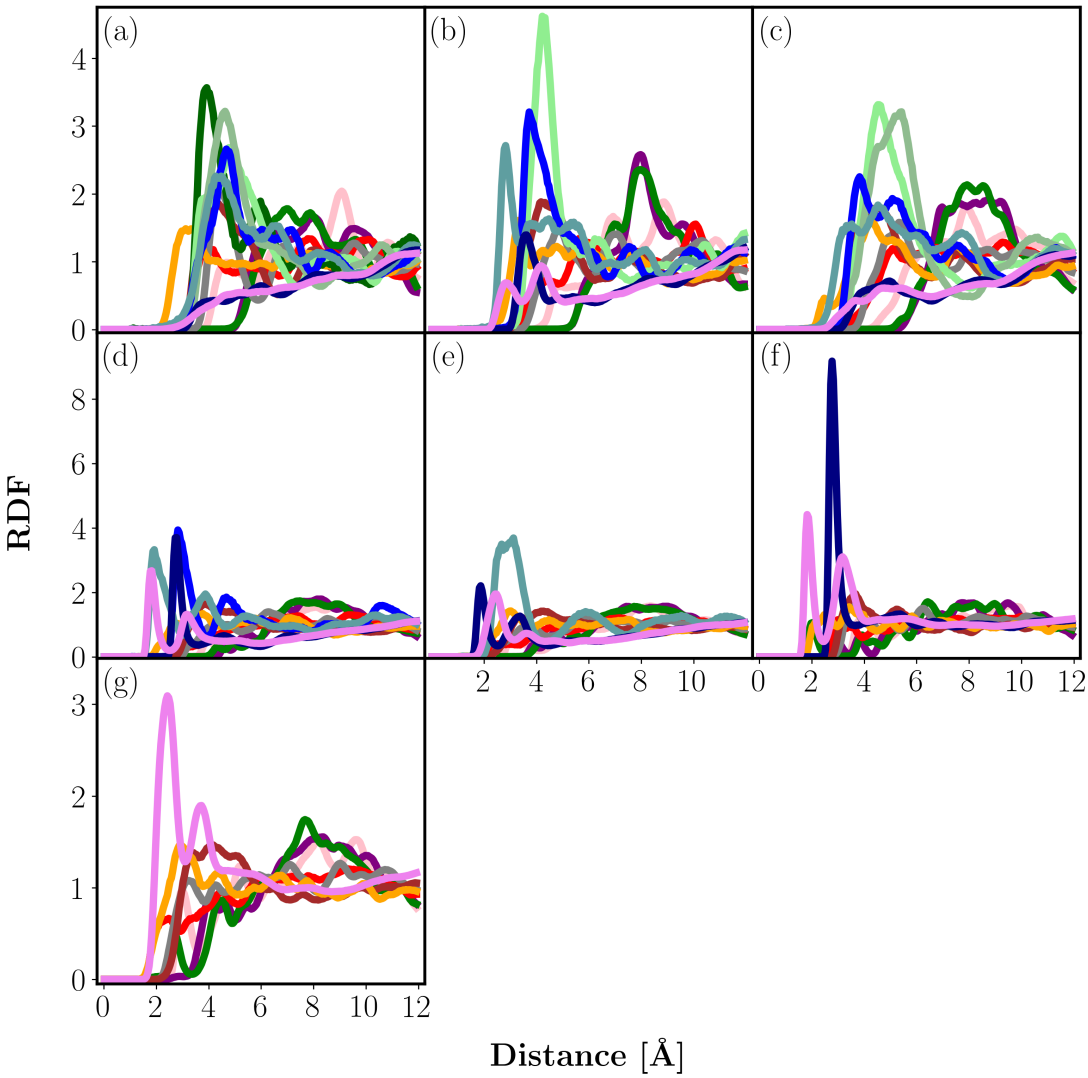

Fig. L14: RDFs of distances for interaction pairs observed for water/n-butanol adsorption in CAU-10 at 298.15 K and $x'_{\text{water}} = 0.1$. RDFs are shown for (a) CH_3, (b) CH_2, (c) CH_2a, (d) Oa, (e) Ha, (f) Ow, or (g) Hw *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), CH_3 (dark green), CH_2 (dark sea green), CH_2a (light green), Oa (blue), Ha (cadet blue), Ow (navy), or Hw (violet).

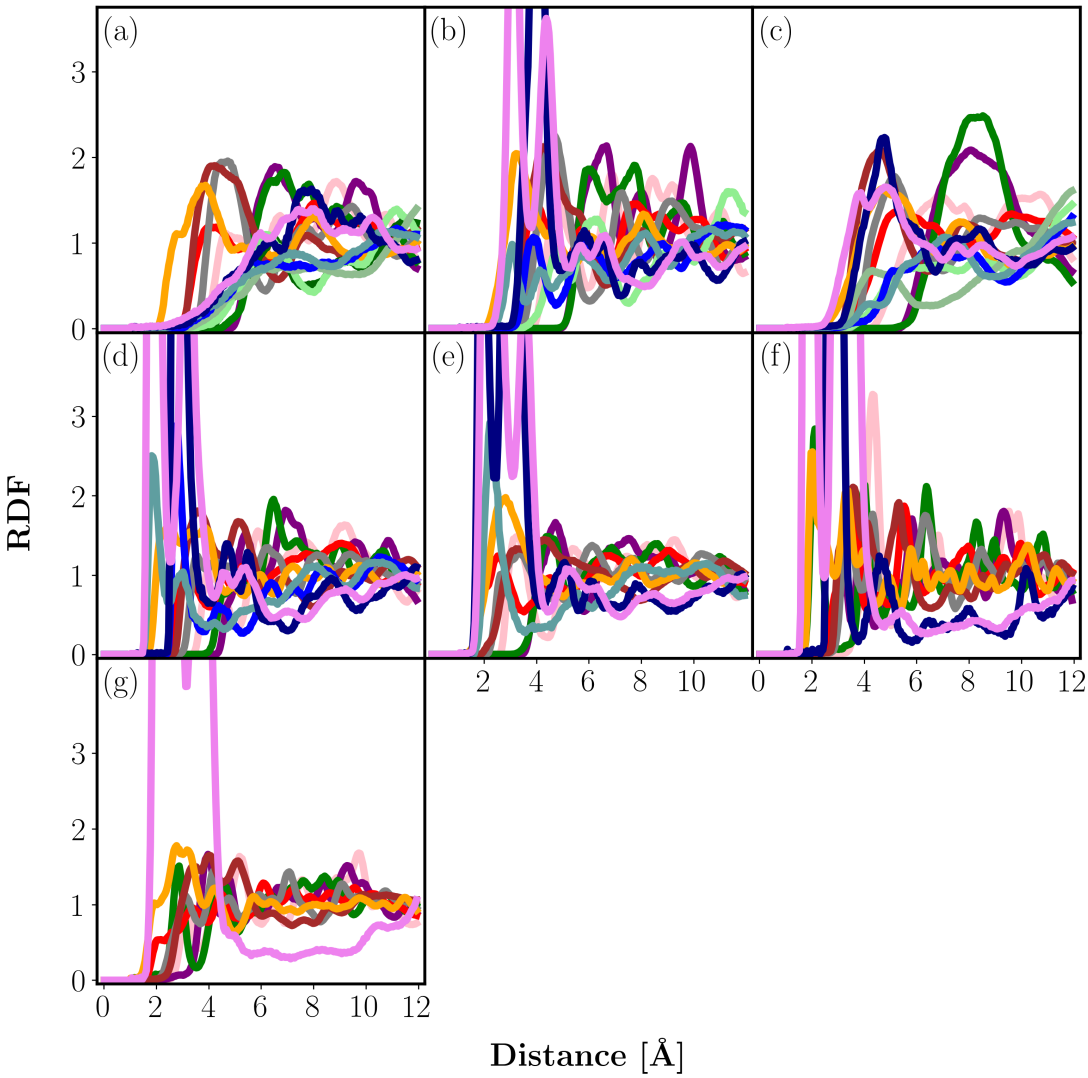

Fig. L15: RDFs of distances for interaction pairs observed for water/n-hexanol adsorption in CAU-10 at 298.15 K and $x'_{\text{water}} = 0.01$. RDFs are shown for (a) CH_3, (b) CH_2, (c) CH_2a, (d) Oa, (e) Ha, (f) Ow, or (g) Hw *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), CH_3 (dark green), CH_2 (dark sea green), CH_2a (light green), Oa (blue), Ha (cadet blue), Ow (navy), or Hw (violet).

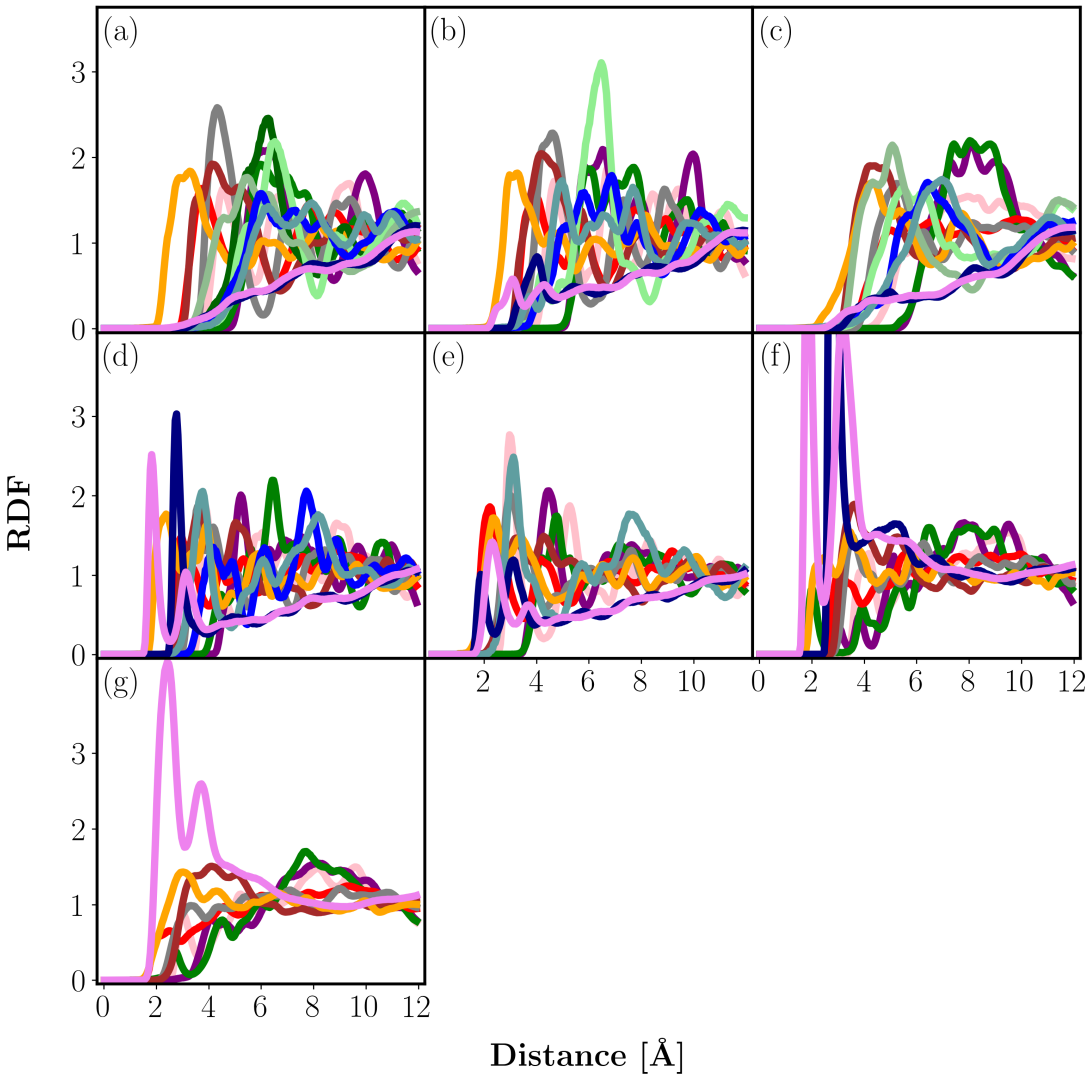

Fig. L16: RDFs of distances for interaction pairs observed for water/n-hexanol adsorption in CAU-10 at 298.15 K and $x'_{\text{water}} = 0.05$. RDFs are shown for (a) CH_3, (b) CH_2, (c) CH_2a, (d) Oa, (e) Ha, (f) Ow, or (g) Hw *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), CH_3 (dark green), CH_2 (dark sea green), CH_2a (light green), Oa (blue), Ha (cadet blue), Ow (navy), or Hw (violet).

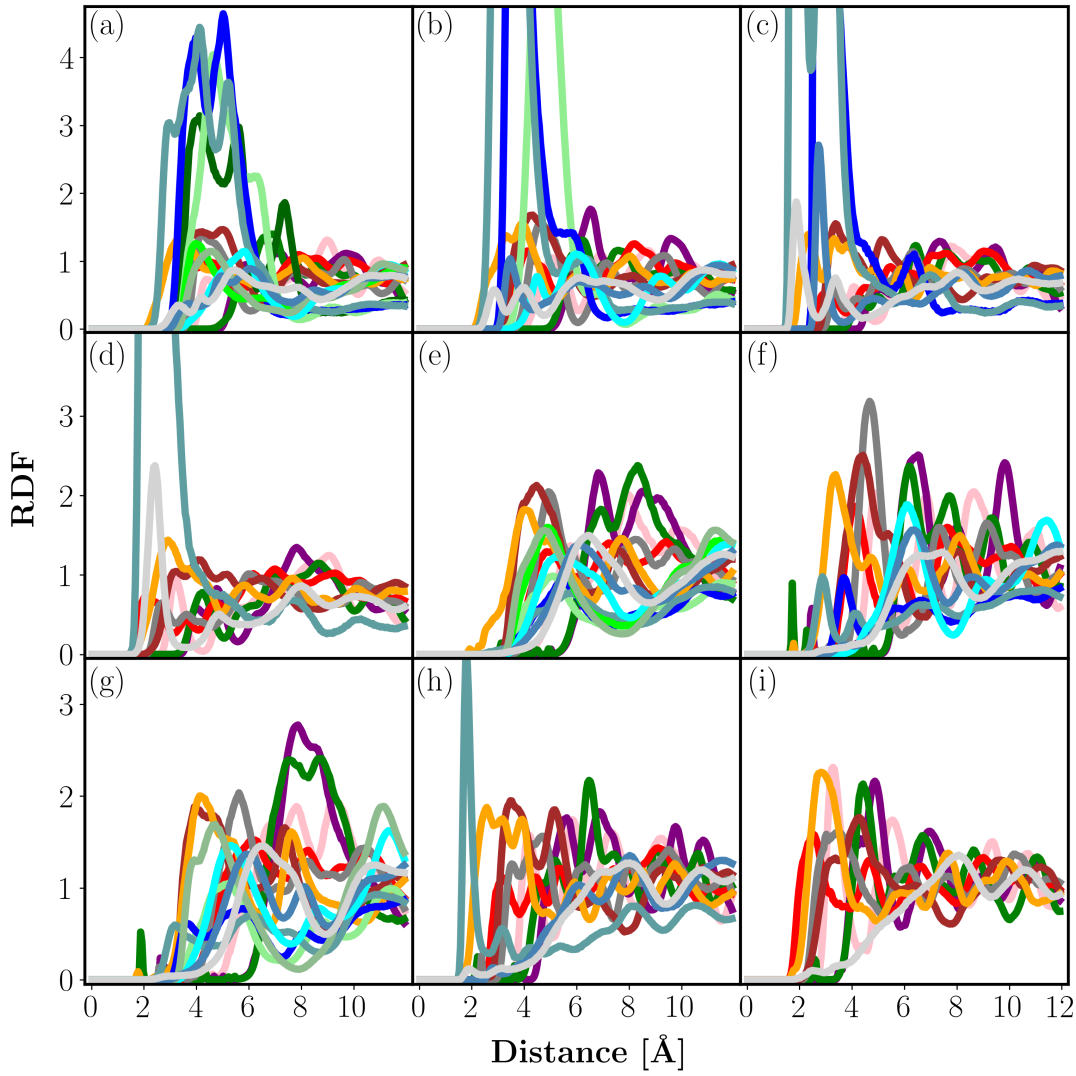

Fig. L17: RDFs of distances for interaction pairs observed for ethanol/n-butanol adsorption in CAU-10 at 298.15 K and $x'_{\text{ethanol}} = 0.05$. RDFs are shown for (a) CH_3, (b) CH_2a, (c) Oa, or (d) Ha in ethanol as well as (e) CH_3, (f) CH_2a, (g) CH_2, (h) Oa, or (i) Ha in n-butanol *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), CH_3 (ethanol: dark green, n-butanol: lime green), CH_2 (dark sea green), CH_2a (ethanol: light green, n-butanol: cyan), Oa (ethanol: blue, n-butanol: steel blue), or Ha (ethanol: cadet blue, n-butanol: gray).

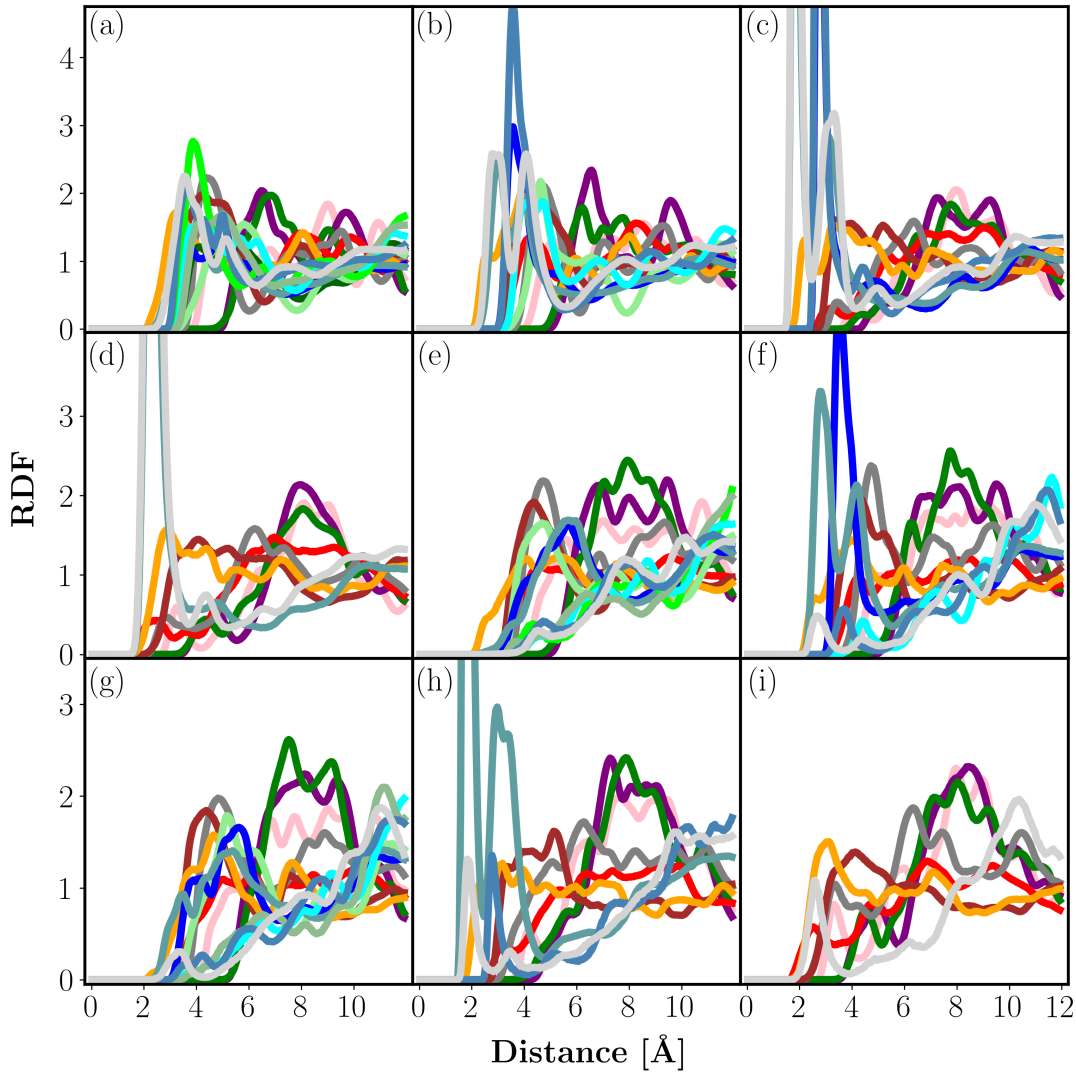

Fig. L18: RDFs of distances for interaction pairs observed for ethanol/n-butanol adsorption in CAU-10 at 298.15 K and $x'_{\text{ethanol}} = 0.5$. RDFs are shown for (a) CH_3, (b) CH_2a, (c) Oa, or (d) Ha in ethanol as well as (e) CH_3, (f) CH_2a, (g) CH_2, (h) Oa, or (i) Ha in n-butanol *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), CH_3 (ethanol: dark green, n-butanol: lime green), CH_2 (dark sea green), CH_2a (ethanol: light green, n-butanol: cyan), Oa (ethanol: blue, n-butanol: steel blue), or Ha (ethanol: cadet blue, n-butanol: gray).

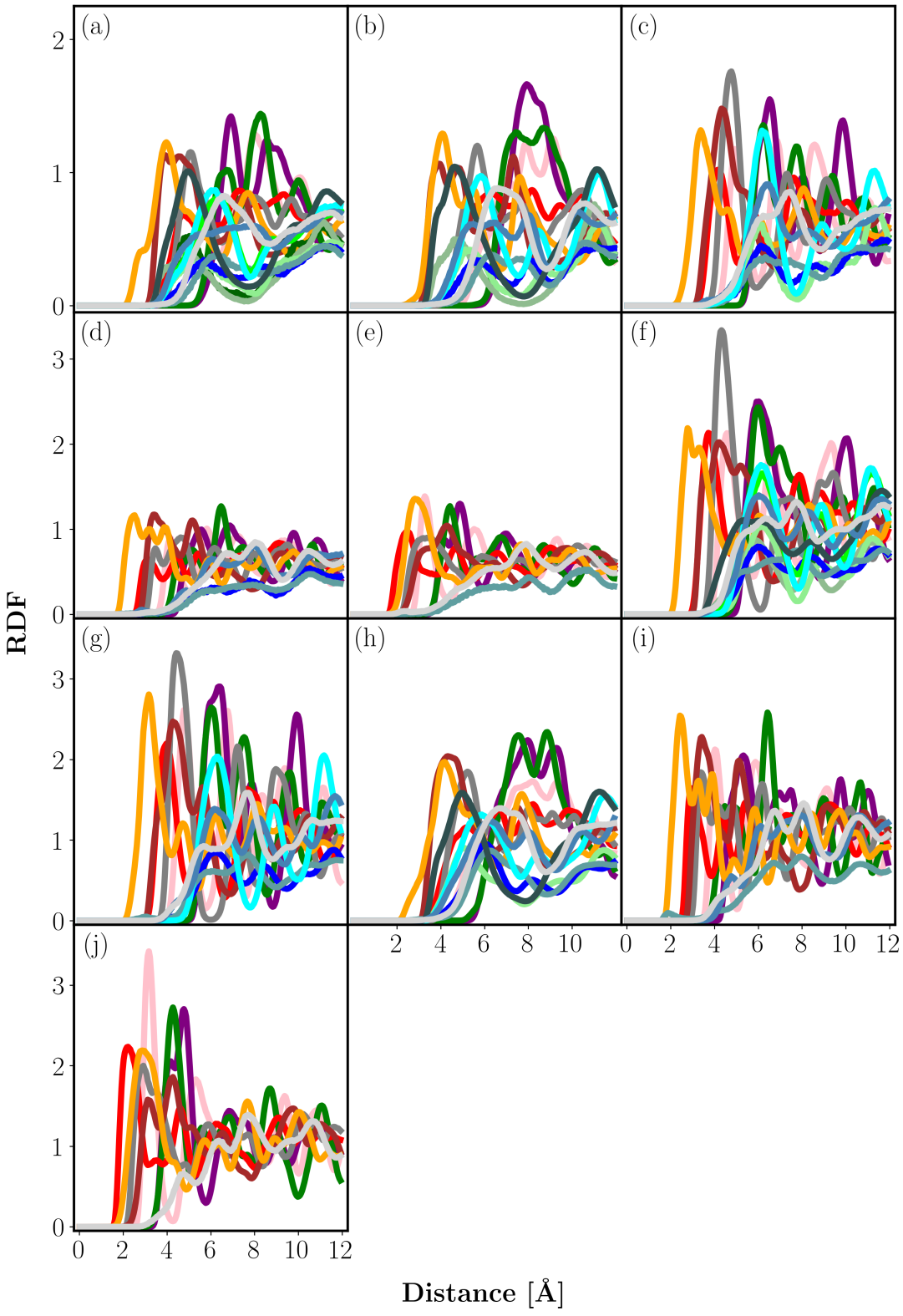

Fig. L19: RDFs of distances for interaction pairs observed for n-butanol/n-hexanol adsorption in CAU-10 at 298.15 K and $x'_{\text{hexanol}} = 0.1$. RDFs are shown for (a) CH_3, (b) CH_2, (c) CH_2a, (d) Oa, or (e) Ha in n-butanol as well as (f) CH_3, (g) CH_2a, (h) CH_2, (i) Oa, or (j) Ha in n-hexanol *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), CH_3 (n-butanol: dark green, n-hexanol: lime green), CH_2 (n-butanol: dark sea green, n-hexanol: dark slate gray), CH_2a (n-butanol: light green, n-hexanol: cyan), Oa (n-butanol: blue, n-hexanol: steel blue), or Ha (n-butanol: cadet blue, n-hexanol: gray).

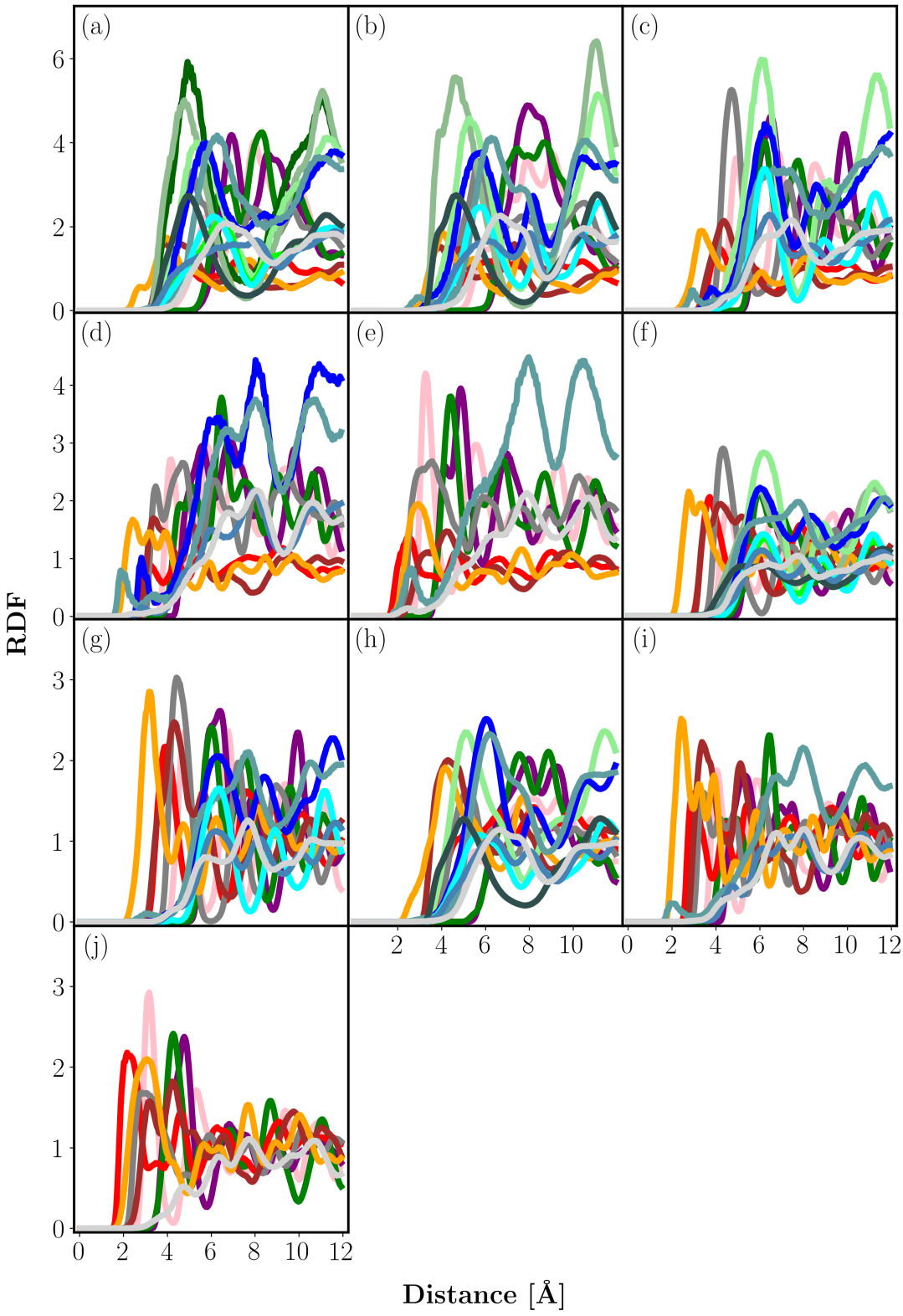

Fig. L20: RDFs of distances for interaction pairs observed for n-butanol/n-hexanol adsorption in CAU-10 at 298.15 K and $x'_{\text{hexanol}} = 0.5$. RDFs are shown for (a) CH_3, (b) CH_2, (c) CH_2a, (d) Oa, or (e) Ha in n-butanol as well as (f) CH_3, (g) CH_2a, (h) CH_2, (i) Oa, or (j) Ha in n-hexanol *and* Al (pink), OAl (purple), Hc (green), Oc (red), Cc (gray), Cb (brown), Hb (orange), CH_3 (n-butanol: dark green, n-hexanol: lime green), CH_2 (n-butanol: dark sea green, n-hexanol: dark slate gray), CH_2a (n-butanol: light green, n-hexanol: cyan), Oa (n-butanol: blue, n-hexanol: steel blue), or Ha (n-butanol: cadet blue, n-hexanol: gray).

M Section 4: General Simulation Information

M.1 Surface Models

Tab. M1: Information on the surface models.

Surface Model	Simulation Box Size (x y z)	Atom No.	Thickness	Equi. Time
MOF-5-phenyl	310 Å × 300 Å × 310 Å	92 736	33.8 Å	4 ns
MOF-5-benzoic	310 Å × 300 Å × 310 Å	94 464	33.8 Å	5 ns
CAU-10-phenyl	215 Å × 300 Å × 206 Å	93 600	30.0 Å	5 ns
CAU-10-benzoic	215 Å × 300 Å × 206 Å	96 000	30.0 Å	5 ns
CAU-10-open	236 Å × 300 Å × 236 Å	91 960	25.4 Å	4 ns

M.2 Molecular Dynamics: Interatomic Potential Parameters

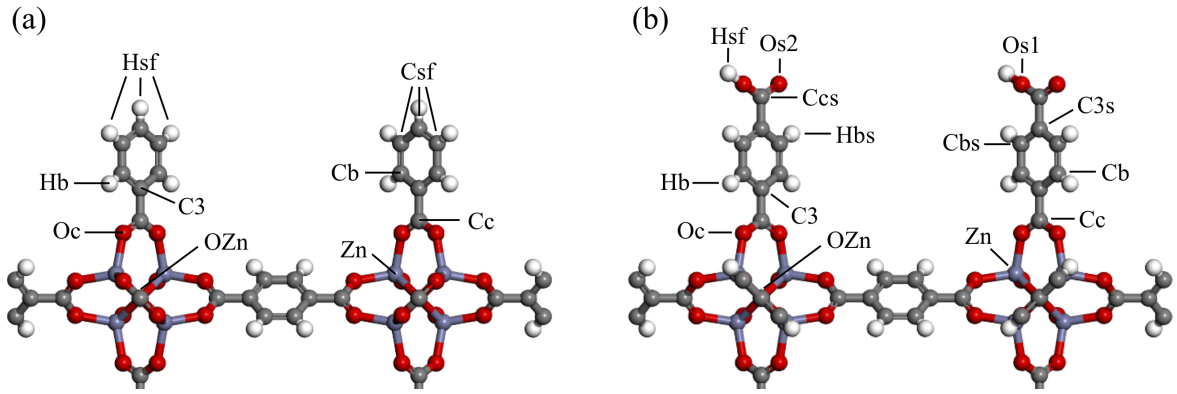

Fig. M1: Atom types used to describe (a) MOF-5-phenyl and (b) MOF-5-benzoic.

Tab. M2: Interaction parameters for the atom types in MOF-5-phenyl. See Fig. M1 for naming of interaction site labels.

Atom Type	L-J well depth [kcal/mol]	L-J distance [Å]	partial charge
Cb	0.0951	3.473	-0.102
Cc	0.0951	3.473	0.373
C3	0.0951	3.473	0.051
Csf	0.0951	3.473	-0.09
Hb	0.0152	2.846	0.149
Hsf	0.0152	2.846	0.105
Oc	0.0957	3.033	-0.499
OZn	0.06	3.118	-0.87
Zn	0.124	2.462	0.942

Tab. M3: Interaction parameters for the atom types in MOF-5-benzoic. See Fig. M1 for naming of interaction site labels.

Atom Type	L-J well depth [kcal/mol]	L-J distance [Å]	partial charge
Cb	0.0951	3.473	-0.084
Cc	0.0951	3.473	0.407
C3	0.0951	3.473	0.044
Cbs	0.0951	3.473	-0.095
Ccs	0.0951	3.473	0.417
C3s	0.0951	3.473	0.048
Hb	0.0152	2.846	0.14
Hbs	0.0152	2.846	0.108
Hsf	0.0152	2.846	0.503
Oc	0.0957	3.033	-0.518
Os1	0.0957	3.033	-0.631
Os2	0.0957	3.033	-0.398
OZn	0.06	3.118	-0.877
Zn	0.124	2.462	0.929

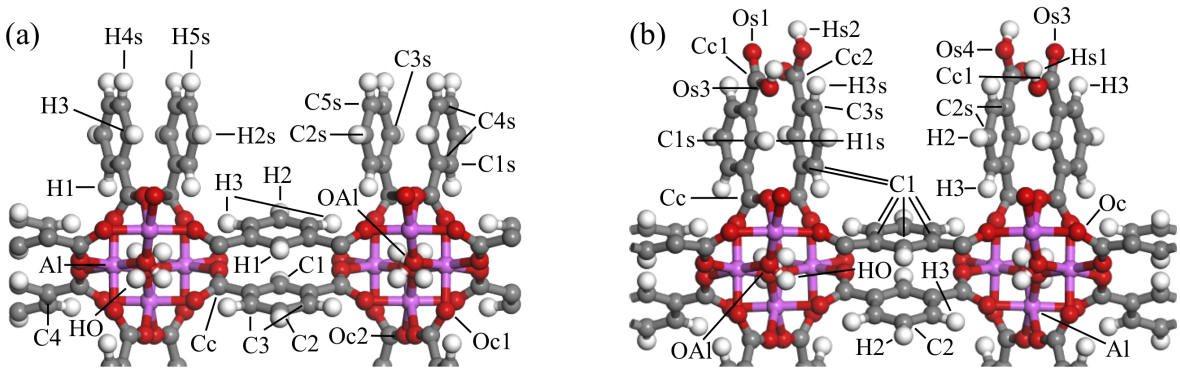

Fig. M2: Atom types used to describe (a) CAU-10-phenyl and (b) CAU-10-benzoic.

Tab. M4: Interaction parameters for the atom types in CAU-10-phenyl. See Fig. M2 for naming of interaction site labels.

Atom Type	L-J well depth [kcal/mol]	L-J distance [Å]	partial charge
Al	0.505	4.008	1.362
C1	0.0951	3.473	-0.08
C1s	0.0951	3.473	-0.078
C2	0.0951	3.473	-0.096

C2s	0.0951	3.473	-0.089
C3	0.0951	3.473	-0.073
C3s	0.0951	3.473	-0.083
C4	0.0951	3.473	-0.061
C4s	0.0951	3.473	-0.122
C5s	0.0951	3.473	-0.085
Cc	0.0951	3.473	0.552
H1	0.0152	2.846	0.165
H2	0.0152	2.846	0.115
H2s	0.0152	2.846	0.109
H3	0.0152	2.846	0.133
H4s	0.0152	2.846	0.104
H5s	0.0152	2.846	0.097
HO	0.044	2.571	0.296
OAl	0.06	3.118	-0.713
Oc1	0.0957	3.033	-0.559
Oc2	0.0957	3.033	-0.516

Tab. M5: Interaction parameters for the atom types in CAU-10-benzoic. See Fig. M2 for naming of interaction site labels.

Atom Type	L-J well depth [kcal/mol]	L-J distance [Å]	partial charge
Al	0.505	4.008	1.410
C1	0.0951	3.473	-0.068
C1s	0.0951	3.473	-0.051
C2	0.0951	3.473	-0.099
C2s	0.0951	3.473	-0.094
C3s	0.0951	3.473	-0.059
Cc	0.0951	3.473	0.575
Cc1	0.0951	3.473	0.459
Cc2	0.0951	3.473	0.472
H1	0.0152	2.846	0.188
H2	0.0152	2.846	0.115
H3	0.0152	2.846	0.142

H3s	0.0152	2.846	0.163
HO	0.044	2.571	0.299
Hs1	0.0152	2.846	0.331
Hs2	0.0152	2.846	0.301
Oc	0.0957	3.033	-0.570
OAl	0.06	3.118	-0.723
Os1	0.0957	3.033	-0.425
Os2	0.0957	3.033	-0.466
Os3	0.0957	3.033	-0.45
Os4	0.0957	3.033	-0.427

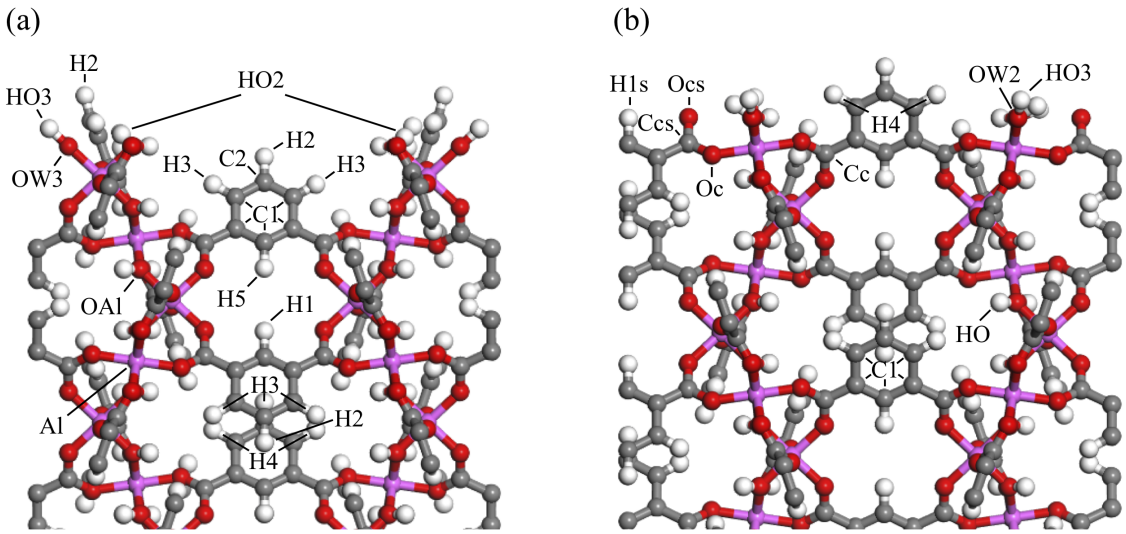

Fig. M3: Atom types used to describe CAU-10-open (a) front view (b) side view.

Tab. M6: Interaction parameters for the atom types in CAU-10-open. See Fig. M3 for naming of interaction site labels.

Atom Type	L-J well depth [kcal/mol]	L-J distance [Å]	partial charge
Al	0.505	4.008	1.362
C1	0.0951	3.473	-0.069
C2	0.0951	3.473	-0.095
Cc	0.0951	3.473	0.553
Ccs	0.0951	3.473	0.493
H1	0.0152	2.846	0.173
H1s	0.0152	2.846	0.154
H2	0.0152	2.846	0.103

H3	0.0152	2.846	0.127
H4	0.0152	2.846	0.139
H5	0.0152	2.846	0.158
HO	0.044	2.571	0.296
HO2	0.044	2.571	0.387
HO3	0.044	2.571	0.312
OAl	0.06	3.118	-0.713
Oc	0.0957	3.033	-0.55
Ocs	0.0957	3.033	-0.445
OW2	0.06	3.118	-0.535
OW3	0.06	3.118	-0.27

M.3 Trial Move Probabilities for Monte Carlo Simulations

Grand canonical Ensemble		**Canonical Ensemble**	
Translation:	0.15	Translation:	0.3
Rotation:	0.15	Rotation:	0.3
Full regrowth:	0.05	Full regrowth:	0.1
Partial regrowth	0.05	Partial regrowth	0.1
Swap:	0.5	Swap:	0.0
Widom Insertion	0.1	Widom Insertion:	0.2

M.4 Monte Carlo: Interatomic Potential Parameters

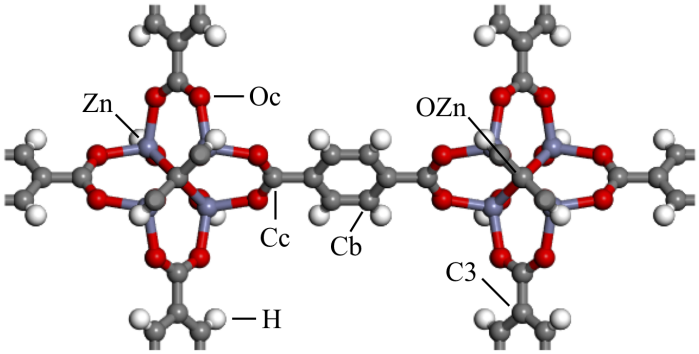

Fig. M4: Atom types used to describe MOF-5.

Tab. M7: Interaction parameters for the atom types in MOF-5 and comparison of the used partial charges with DDEC charges taken from Sladekova et al. [166]. See Fig. M4 for naming of interaction site labels.

Atom Type	L-J well depth [K]	L-J distance [Å]	partial charge	DDEC charge (Sladekova et al. [166])
Cb	47.856	3.473	-0.180	-0.1237
Cc	47.856	3.473	0.529	0.5786
C3	47.856	3.473	0.023	-0.0139
H	7.649	2.846	0.223	0.1490
Oc	48.158	3.033	-0.546	-0.5512
OZn	30.193	3.118	-0.946	-1.0220
Zn	62.139	2.462	0.918	0.9864

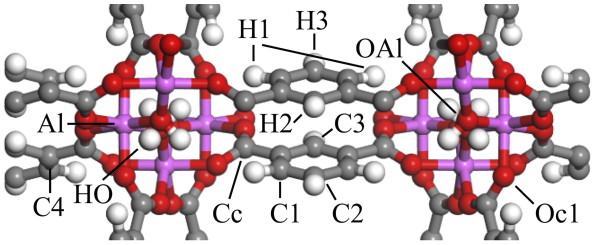

Fig. M5: Atom types used to describe CAU-10.

Tab. M8: Interaction parameters for the atom types in CAU-10 and comparison of the used partial charges with values from Cadiau et al. [7] provided by Guillaume Maurin in a private correspondence. See Fig. M5 for naming of interaction site labels.

Atom Type	L-J well depth [K]	L-J distance [Å]	partial charge	Mulliken charge (Cadiau et al. [7])
C1	47.856	3.473	-0.073	-0.0721
C2	47.856	3.473	-0.096	-0.0721
C3	47.856	3.473	-0.080	-0.0721
C4	47.856	3.473	-0.061	-0.0721
Cc	47.856	3.473	0.551	0.5697
Oc	48.158	3.033	-0.538	-0.5554
OAl	30.213	3.118	-0.713	-0.7426
H1	7.648	2.846	0.133	0.1371
H2	7.648	2.846	0.115	0.1371
H3	7.648	2.846	0.165	0.1371
HO	22.142	2.571	0.297	0.3140
Al	254.127	4.008	1.362	1.3950

M.5 Monte Carlo Cycles

Tab. M9: Monte Carlo cycles used for the water adsorption simulation in MOF-5

Pressure Range [Pa]	Equilibration	Production
0.01-1000	1,700 000	100,000
1100	1,850 000	100,000
1300	4,850 000	100,000
1500-3500	1,600 000	100,000

Tab. M10: Monte Carlo cycles used for the water adsorption simulation in CAU-10

Pressure Range [Pa]	Equilibration	Production
0.1-700	2,750,000	100,000
725-900	2,500,000	100,000
1000	1,850,000	100,000
1200-3000	1,650,000	100,000

N Section 4: Density Plots

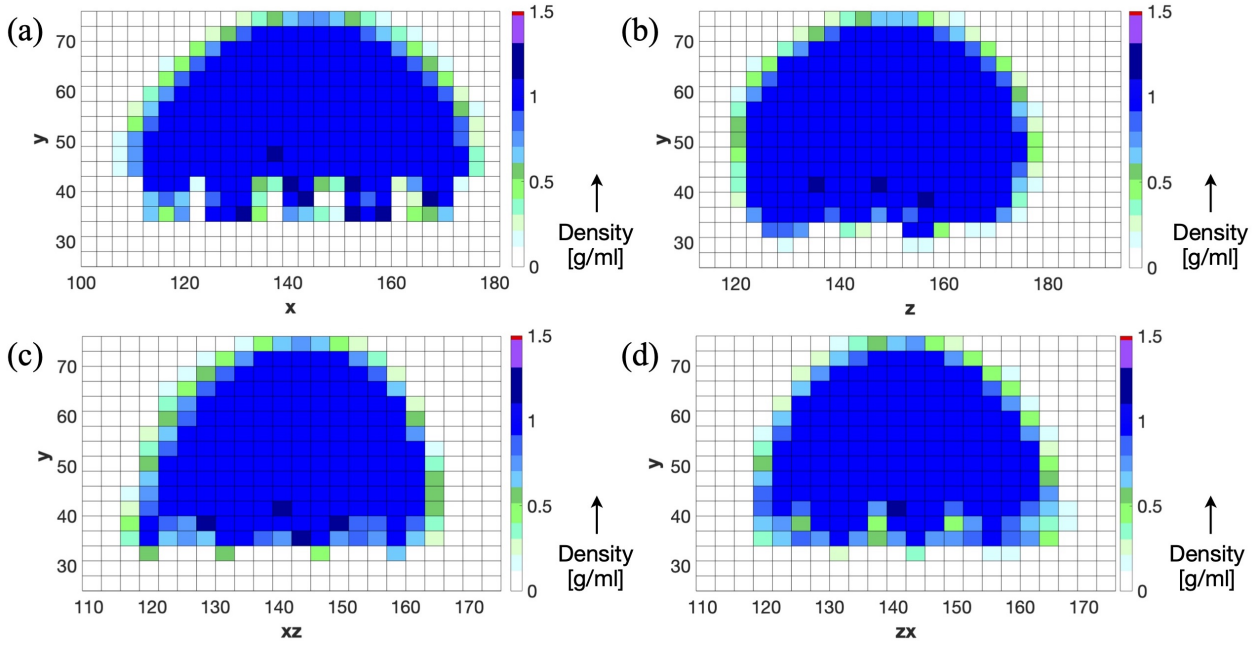

Fig. N1: Density plots for the slices (a) Axial 1, (b) Axial 2, (c) Dia1 and (d) Dia2 of MOF-5-phenyl.

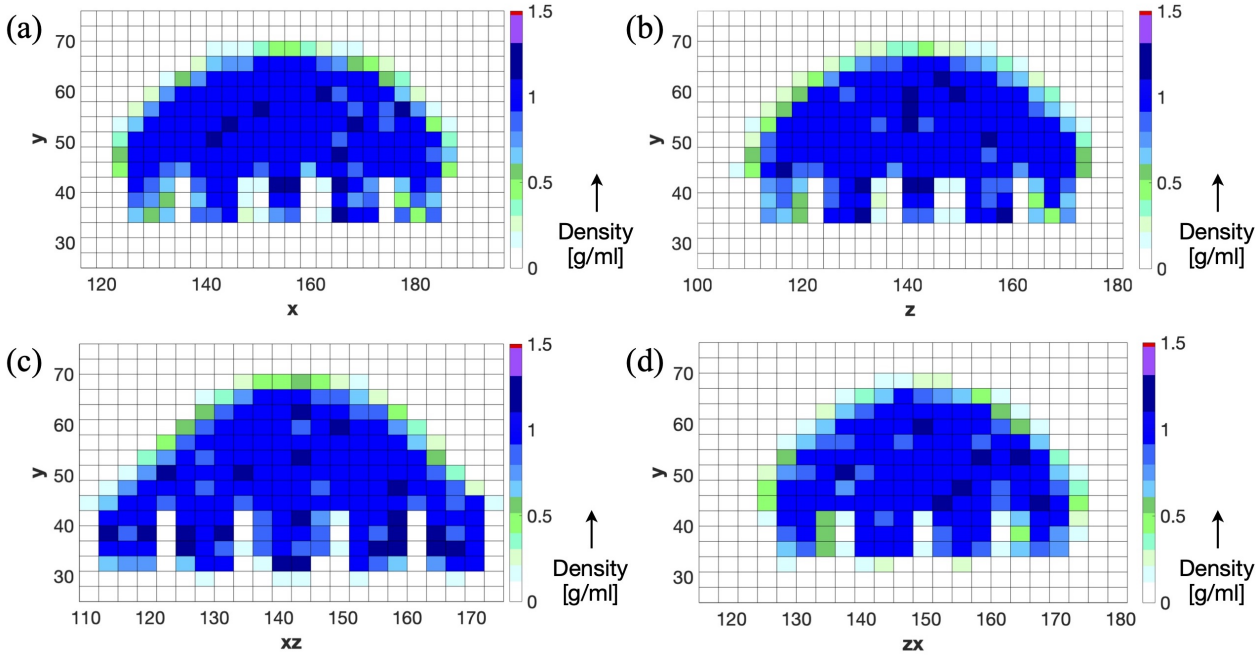

Fig. N2: Density plots for the slices (a) Axial 1, (b) Axial 2, (c) Dia1 and (d) Dia2 of MOF-5-benzoic.

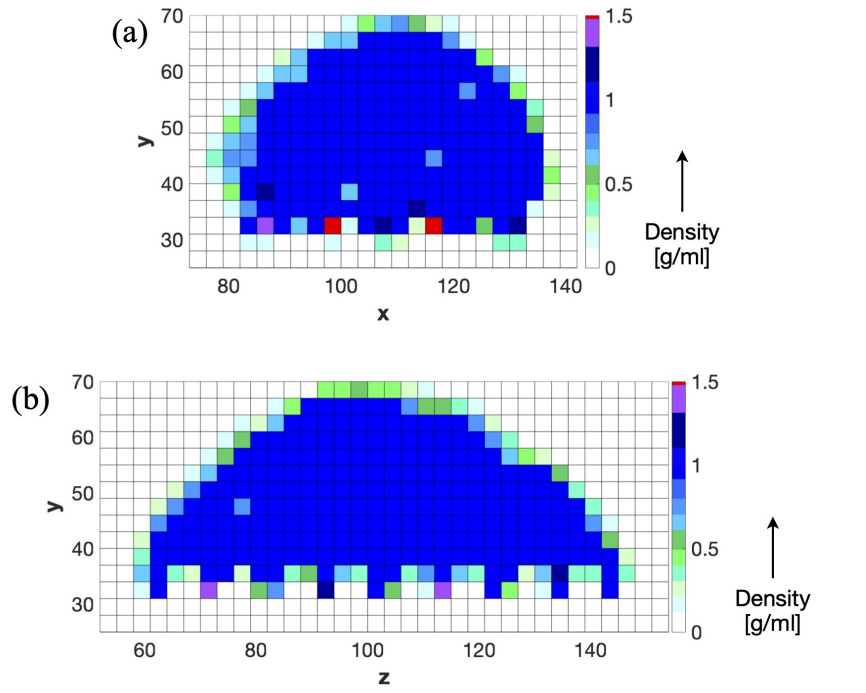

Fig. N3: Density plots for the slices (a) Axial 1 and (b) Axial 2 of CAU-10-phenyl

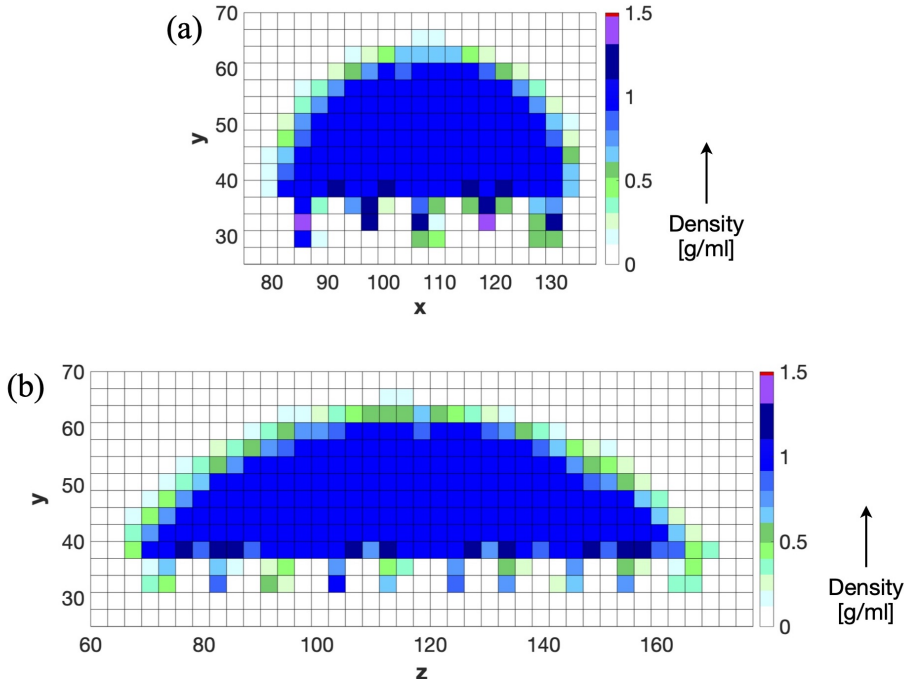

Fig. N4: Density plots for the slices (a) Axial 1 and (b) Axial 2 of CAU-10-benzoic

O Section 4: Calculation Data

Tab. O1: Center and radius of the fitted circle for MOF-5-phenyl and MOF-5-benzoic. See Fig. 37 for naming of slices.

MOF-5-phenyl	Center	Radius	**MOF-5-benzoic**	Center	Radius
Axial 1	40.4 Å	34.9 Å	Axial 1	34.3 Å	34.1 Å
Axial 2	46.3 Å	29.2 Å	Axial 2	34.4 Å	34.3 Å
Dia 1	42.9 Å	32.4 Å	Dia 1	30.2 Å	38.7 Å
Dia 2	42.3 Å	32.8 Å	Dia 2	33.2 Å	35.0 Å

Tab. O2: Center and radius of the fitted circle for CAU-10-phenyl and CAU-10-benzoic. See Fig. 37 for naming of slices.

CAU-10 (Phenyl)	Center	Radius	**CAU-10 (Benzoic)**	Center	Radius
Axial 1	27.1 Å	38.8 Å	Axial 1	-3.7 Å	66.5 Å
Axial 2	34.1 Å	32.9 Å	Axial 2	35.9 Å	27.4 Å

P Section 4: Adsorption Properties

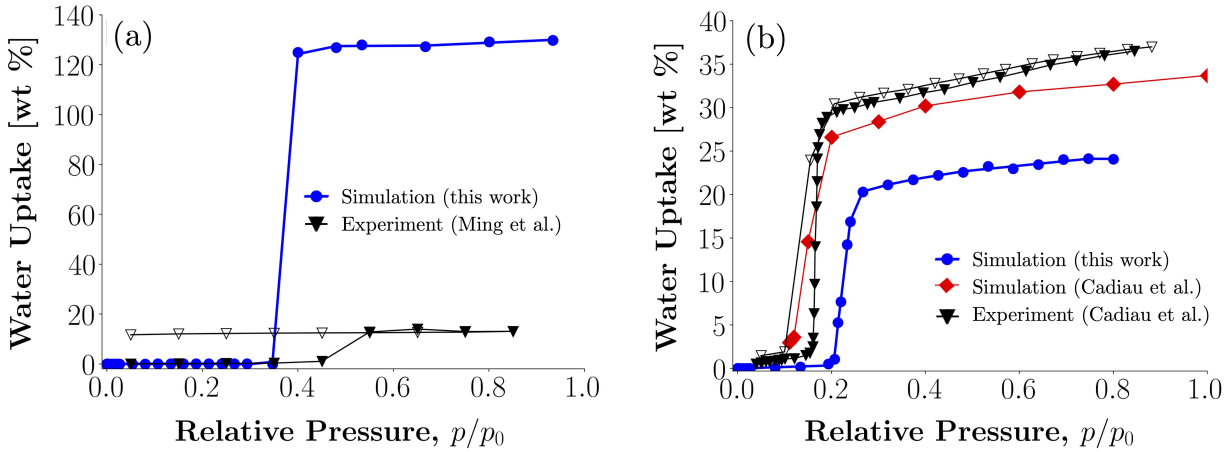

Fig. P1: (a) Comparison of the simulated adsorption isotherm of MOF-5 with the experimental adsorption isotherm measured by Ming et al. [101]. (b) Comparison of the simulated adsorption isotherm of CAU-10 with the simulated and experimental adsorption isotherm by Cadiau et al. [7]. The isotherm points were determined from the figures in the respective publications. The difference in the adsorption isotherms of Cadiau et al. and this work is due to the different force fields used. Full symbols indicate the adsorption branch, while empty symbols indicate the desorption branch of the isotherm.

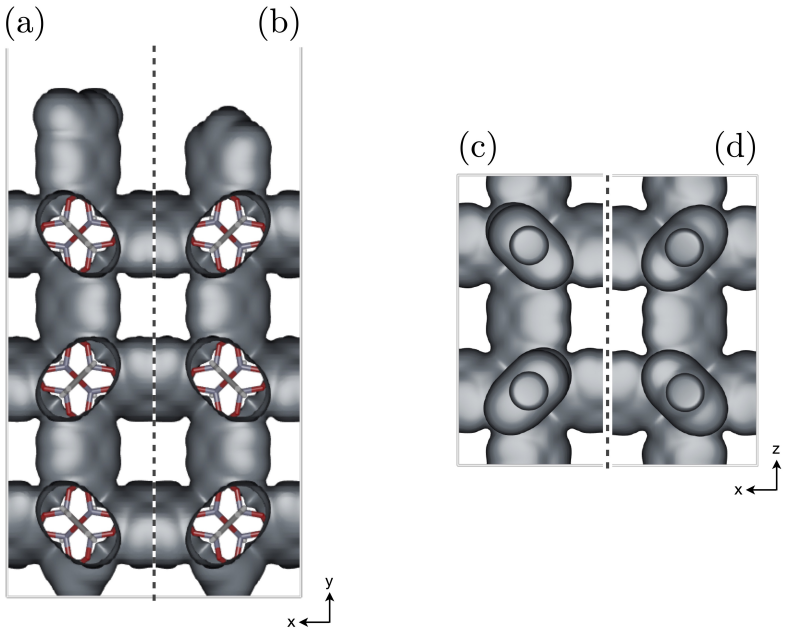

Fig. P2: Water adsorption surfaces of the surface modules of (a, c) MOF-5-benzoic and (b, d) MOF-5-phenyl from different perspectives. Generated with iRASPA [69].

Q Section 5: MOF Building Blocks

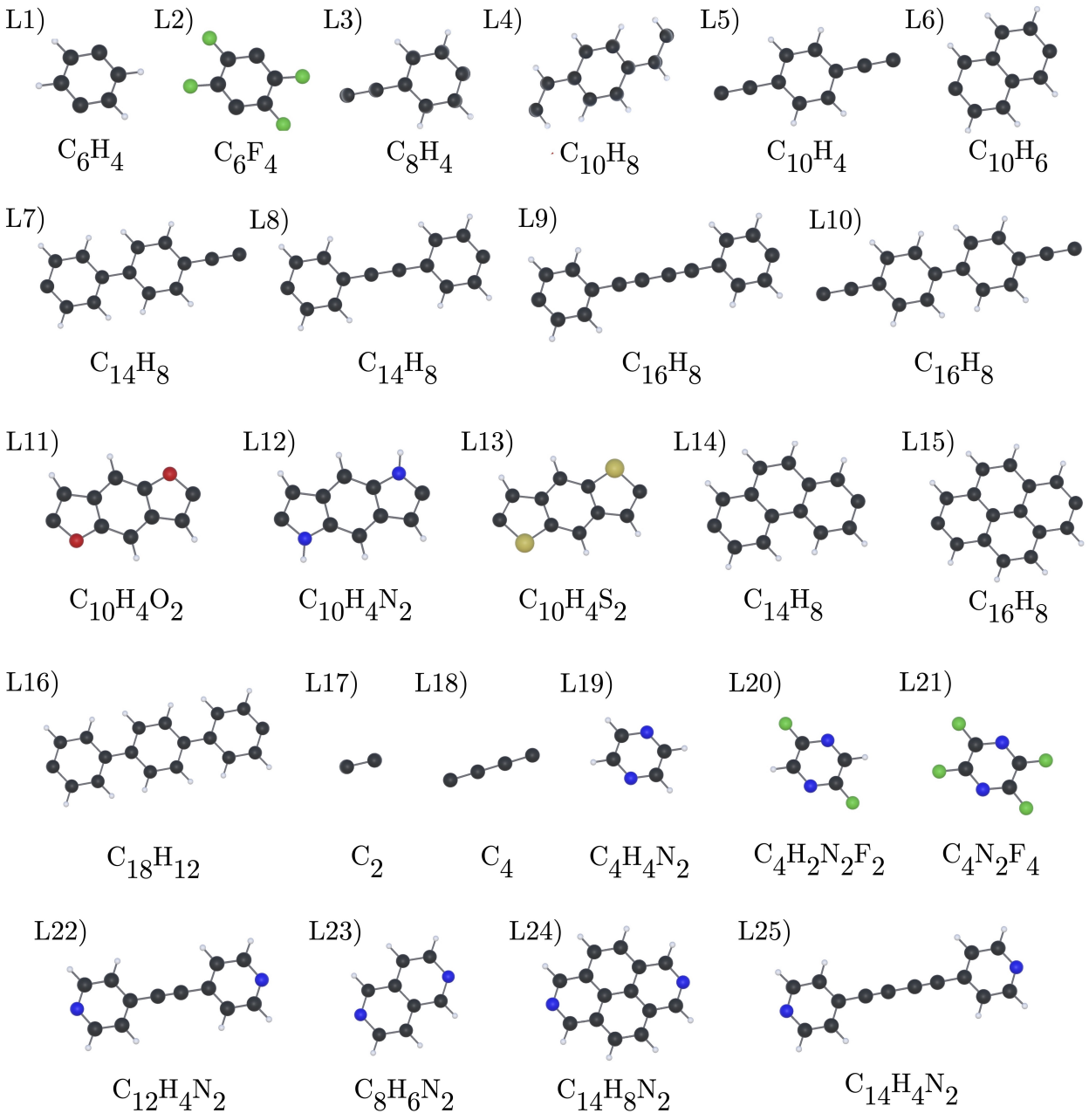

Fig. Q1: Linking building blocks implemented in MOF-VR. Atomic positions of linking building blocks were taken from the Crystallographic Information Files (CIF) accompanying ToBaCCo 3.0 and slightly modified to ensure planarity [123, 146]. Coloring of atoms following the Corey-Pauling-Koltun (CPK) coloring scheme [167, 168] – i.e. carbon - black, hydrogen - white, fluorine - green, oxygen - red, nitrogen - blue, sulfur - yellow. Connection sites are unsaturated carbon atoms (L1 – L18) or nitrogen atoms (L19 – L25).

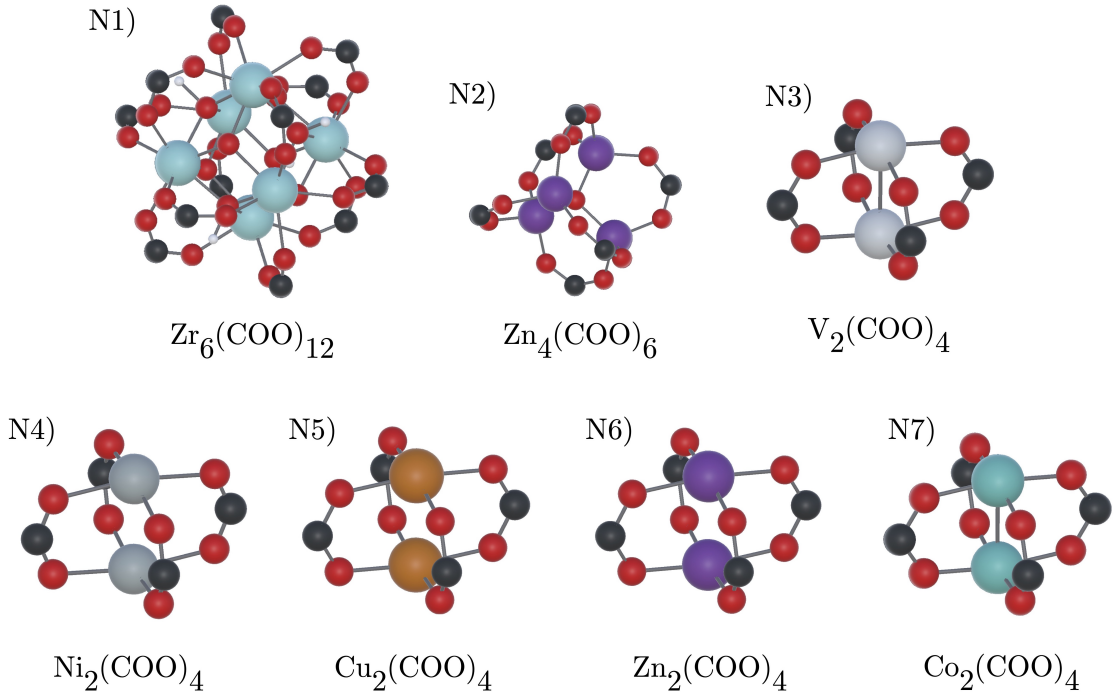

Fig. Q2: Nodal building blocks implemented in MOF-VR. Atomic positions of nodal building blocks were taken from CIF files accompanying ToBaCCo 3.0 [123, 146] (N1-N5) or mofplus.org [147–149] (N6 & N7). Coloring of atoms following the CPK coloring scheme [167, 168] – i.e. carbon - black, oxygen - red, iron - (light) gray, nickel - (deep) gray –, except for: zirconium - light blue, zinc - purple, copper - brown, cobalt - turquoise. Connection sites are unsaturated carbon atoms (N1 – N7) and/or metal atoms (N3 – N7).

R Section 5: Molecular Building Block Charges

Partial charges for the interaction sites of the MOF building blocks available in MOF-VR were determined following the molecular building block-based (MBBB) charges approach developed by Argueta et al. [151]:

1. Connection sites of MOF building blocks were capped "to emulate the typical bonding environment of the [MOF] building blocks in MOFs" ([151], p. 367). Open metal sites and carbon connection sites in nodal building blocks were capped by coordinatively bonded ammonia or methyl groups, while carbon connection sites and nitrogen connection sites in linking building blocks were capped by carboxylic acid groups or hydrogen atoms [151]. Capping groups were geometrically optimized using the Clean tool in Materials Studio 7.0 (BIOVIA).

2. The total electron densities of the properly capped, isolated MOF building blocks were determined by Γ-point density functional theory (DFT) calculations using the Becke-3-Parameter-Lee-Yang-Parr (B3LYP) [169–171] hybrid generalized gradient approximation (GGA) [76] with parametrized dispersion interaction (DFT-D3) [172, 173] in conjunction with the auxiliary density matrix method (ADMM) [174] as implemented in the Quickstep module of CP2K [175–182]. For non-metals, triple-ζ polarization quality Gaussian basis sets (TZVP-MOLOPT) were used along with auxiliary basis sets with polarization functions (pFIT3), while for metals, double-ζ polarization quality Gaussian basis sets (DZVP-MOLOPT) were used along with uncontracted auxiliary basis sets (FIT11) [183]. Plane wave basis sets were treated with a relative cutoff of 80 Ry and truncated with a planewave cutoff in the range of 1200 – 1400 Ry (see convergence procedure below). Goedecker-Teter-Hutter (GTH) pseudopotentials [184–186] were used and the Coulomb potential was truncated with a cutoff of 8 Å.

3. Partial charges for each MOF building block were determined using the density-derived electrostatic and chemical (DDEC6) charge method as implemented in the CHARGEMOL program based on the calculated total electron densities [54]. The calculated partial charges for the MOF building blocks available in MOF-VR can be viewed in the CIF files available under MOF-VR/Program/SourceCode/CoordFiles/ on the CD-ROM accompanying this work.

Plane wave cutoff and relative cutoff for each properly capped MOF building block were chosen to ensure the desired accuracy of the total energy of the system. For this, a series of single point energy calculations were performed using the Becke-Lee-Yang-Parr (BLYP) [169, 170] generalized gradient approximation (GGA) [76] with parametrized dispersion interaction (DFT-D3) [172, 173] as implemented in the Quickstep module of CP2K [175–178]. For non-metals, triple-ζ polarization quality and for metals, double-ζ polarization quality Gaussian basis sets (DZVP-MOLOPT or TZVP-MOLOPT) [183] were used along with GTH

pseudopotentials [184–186]. Planewave cutoff and relative cutoff were chosen in the following way:

1. Γ-point calculations were performed using a relative cutoff of 60 Ry and planewave cutoffs of 600 Ry, 800 Ry, 1000 Ry, 1200 Ry, 1300 Ry, 1400 Ry, and (for nodal building blocks only) 1600 Ry. The total energies for the properly capped MOF building blocks as a function of the planewave cutoff are shown in Tab. R1 to R7. Here, the properly capped MOF building blocks are labeled according to Fig. Q1 and Fig. Q2, and the capping is denoted by a prime $'$ (e.g., the capped version of linker L1 is L1$'$). In Tab. R1 to R7, planewave cutoffs used in subsequent calculations are underlined and were chosen to ensure an absolute accuracy of the total energy of at least 10^{-2} Ry (for nodal building blocks) or 10^{-3} Ry (for linking building blocks). However, this convergence criterion is not satisfied by linker L2$'$ and L17$'$ as the total energy seems to oscillate between two values. In general, the coarse convergence of all calculations is probably due to an inadequate relative cutoff of 60 Ry as indicated by the convergence of the relative cutoff described thereafter.

2. Γ-point calculations with relative cutoffs ranging from 70 Ry to 90 Ry in steps of 10 Ry were performed using the underlined planewave cutoffs in Tab. R1 to R7. The results of the performed single point energy calculations are shown in Tabs. R8 – R14 and suggest that a relative cutoff of 80 Ry is sufficient to ensure an absolute accuracy of the total energy of 10^{-6} Ry for all MOF building blocks.

Tab. R1: Total energy of properly capped MOF building blocks N1$'$, N2$'$, or N3$'$ (see Fig. Q2) as function of the planewave cutoff.

	Total Energy in [Ry]		
Cutoff in [Ry]	N1$'$	N2$'$	N3$'$
600	-956.067103	-528.683022	-450.983419
800	-956.059692	-528.680556	-450.983044
1000	-956.064422	-528.680419	-450.982941
1200	-956.066652	-528.680361	-450.982735
1300	-956.067089	-528.679661	-450.982986
1400	<u>-956.066576</u>	<u>-528.680414</u>	<u>-450.983137</u>
1600	-956.066280	-528.680468	-450.983082

Tab. R2: Total energy of properly capped MOF building blocks N4′, N5′, N6′, or N7′ (see Fig. Q2) as function of the planewave cutoff.

	Total Energy in [Ry]			
Cutoff in [Ry]	N4′	N5′	N6′	N7′
600	-542.543754	-299.991844	-324.919604	-494.276095
800	-542.543105	-299.992387	-324.920330	-494.275314
1000	-542.543431	-299.992009	-324.920008	-494.274433
1200	-542.542349	-299.992629	-324.919405	-494.275130
1300	-542.542325	-299.991193	-324.919495	-494.274707
1400	-542.542556	-299.991803	-324.919568	-494.274940
1600	-542.543341	-299.991813	-324.919638	-494.275666

Tab. R3: Total energy of properly capped MOF building blocks L1′, L2′, L3′, L4′, or L5′ (see Fig. Q1) as function of the planewave cutoff.

	Total Energy in [Ry]				
Cutoff in [Ry]	L1′	L2′	L3′	L4′	L5′
600	-112.787135	-207.144439	-123.860008	-137.552332	-135.004327
800	-112.787119	-207.143944	-123.859631	-137.55254	-135.004550
1000	-112.786950	-207.323998	-123.859537	-137.552432	-135.004427
1200	-112.787062	-207.324662	-123.859922	-137.552338	-135.004597
1300	-112.787096	-207.143725	-123.859864	-137.552414	-135.004600
1400	-112.787034	-207.324147	-123.859784	-137.552422	-135.004525

Tab. R4: Total energy of properly capped MOF building blocks L6′, L7′, L8′, L9′, or L10′ (see Fig. Q1) as function of the planewave cutoff.

	Total Energy in [Ry]				
Cutoff in [Ry]	L6′	L7′	L8′	L9′	L10′
600	-136.517517	-160.146577	-160.044395	-171.32509	-171.291754
800	-136.517570	-160.146934	-160.043954	-171.325042	-171.291909
1000	-136.517738	-160.146982	-160.044035	-171.324822	-171.291595
1200	-136.517640	-160.146953	-160.044297	-171.325046	-171.291731
1300	-136.517745	-160.146999	-160.043915	-171.324957	-171.291876
1400	-136.517800	-160.146887	-160.044009	-171.324948	-171.291856

Tab. R5: Total energy of properly capped MOF building blocks L11′, L12′, L13′, L14′, or L15′ (see Fig. Q1) as function of the planewave cutoff.

	Total Energy in [Ry]				
Cutoff in [Ry]	L11′	L12′	L13′	L14′	L15′
600	-167.137422	-156.319052	-155.576857	-170.446629	-171.642380
800	-167.137295	-156.318844	-155.576578	-170.446645	-171.642218
1000	-167.137356	-156.318748	-155.576700	-170.446610	-171.642461
1200	-167.137687	-156.318951	-155.576792	-170.446867	-171.642320
1300	-167.137610	-156.318734	-155.576745	-170.446992	-171.642431
1400	-167.137586	-156.318886	-155.576686	-170.446887	-171.642421

Tab. R6: Total energy of properly capped MOF building blocks L16′, L17′, L18′, L19′, or L20′ (see Fig. Q1) as function of the planewave cutoff.

	Total Energy in [Ry]				
Cutoff in [Ry]	L16′	L17′	L18′	L19′	L20′
600	-185.292796	-87.545831	-98.707553	-45.982321	-93.266521
800	-185.292936	-87.545493	-98.707248	-45.982272	-93.265986
1000	-185.292619	-87.422980	-98.707290	-45.982316	-93.266163
1200	-185.292784	-87.545551	-98.707475	-45.982299	-93.265897
1300	-185.292878	-87.423061	-98.707504	-45.982302	-93.266122
1400	-185.292878	-87.545544	-98.707438	-45.982280	-93.266163

Tab. R7: Total energy of properly capped MOF building blocks L21′, L22′, L23′, L24′, or L25′ (see Fig. Q1) as function of the planewave cutoff.

	Total Energy in [Ry]				
Cutoff in [Ry]	L21′	L22′	L23′	L24′	L25′
600	-140.536148	-93.622249	-69.717917	-105.008927	-104.834707
800	-140.535299	-93.622155	-69.718094	-105.008862	-104.834724
1000	-140.535810	-93.622157	-69.718023	-105.008769	-104.834756
1200	-140.535204	-93.622144	-69.718089	-105.008837	-104.834730
1300	-140.535164	-93.622119	-69.718107	-105.008810	-104.834706
1400	-140.535620	-93.622135	-69.718044	-105.008800	-104.834743

Tab. R8: Total energy of properly capped MOF building blocks N1′, N2′, or N3′ (see Fig. Q2) as function of the relative cutoff.

	Total Energy in [Ry]		
Rel. Cut. in [Ry]	N1′	N2′	N3′
70	-956.066575	-528.680408	-450.982984
80	-956.066574	-528.680408	-450.982984
90	-956.066574	-528.680408	-450.982984

Tab. R9: Total energy of properly capped MOF building blocks N4′, N5′, N6′, or N7′ (see Fig. Q2) as function of the relative cutoff.

	Total Energy in [Ry]			
Rel. Cut. in [Ry]	N4′	N5′	N6′	N7′
70	-542.542323	-299.991190	-324.919492	-494.274706
80	-542.542323	-299.991189	-324.919492	-494.274706
90	-542.542323	-299.991189	-324.919492	-494.274706

Tab. R10: Total energy of properly capped MOF building blocks L1′, L2′, L3′, L4′, or L5′ (see Fig. Q1) as function of the relative cutoff.

	Total Energy in [Ry]				
Rel. Cut. in [Ry]	L1′	L2′	L3′	L4′	L5′
70	-112.787095	-207.324146	-123.859864	-137.552413	-135.004599
80	-112.787095	-207.324146	-123.859863	-137.552413	-135.004599
90	-112.787095	-207.324146	-123.859863	-137.552413	-135.004599

Tab. R11: Total energy of properly capped MOF building blocks L6′, L7′, L8′, L9′, or L10′ (see Fig. Q1) as function of the relative cutoff.

	Total Energy in [Ry]				
Rel. Cut. in [Ry]	L6′	L7′	L8′	L9′	L10′
70	-136.517744	-160.146952	-160.044035	-171.324956	-171.291875
80	-136.517744	-160.146952	-160.044035	-171.324956	-171.291875
90	-136.517744	-160.146952	-160.044035	-171.324956	-171.291875

Tab. R12: Total energy of properly capped MOF building blocks L11′, L12′, L13′, L14′, or L15′ (see Fig. Q1) as function of the relative cutoff.

	Total Energy in [Ry]				
Rel. Cut. in [Ry]	L11′	L12′	L13′	L14′	L15′
70	-167.137612	-156.318733	-155.576744	-170.446887	-171.642430
80	-167.137609	-156.318733	-155.576744	-170.446887	-171.642430
90	-167.137609	-156.318733	-155.576744	-170.446887	-171.642430

Tab. R13: Total energy of properly capped MOF building blocks L16′, L17′, L18′, L19′, or L20′ (see Fig. Q1) as function of the relative cutoff.

	Total Energy in [Ry]				
Rel. Cut. in [Ry]	L16′	L17′	L18′	L19′	L20′
70	-185.292877	-87.545551	-98.707438	-45.982302	-93.266121
80	-185.292877	-87.545551	-98.707438	-45.982302	-93.266121
90	-185.292877	-87.545551	-98.707438	-45.982302	-93.266121

Tab. R14: Total energy of properly capped MOF building blocks L21′, L22′, L23′, L24′, or L25′ (see Fig. Q1) as function of the relative cutoff.

	Total Energy in [Ry]				
Rel. Cut. in [Ry]	L21′	L22′	L23′	L24′	L25′
70	-140.535162	-93.622118	-69.718107	-105.008810	-104.834706
80	-140.535162	-93.622118	-69.718107	-105.008810	-104.834706
90	-140.535162	-93.622118	-69.718107	-105.008810	-104.834706

S Section 5: Auxiliaries

MOF-VR provides auxiliary functions to increase intuitiveness and immersion, such as routines for handling (I) a reorientation of the system or (II) computing the correct rotation of dummy atoms of interaction sites during a molecular dynamics run.

(I) Reorientation of a system occurs whenever a system (consisting of a crystal and the enclosed guest molecules) is picked up and displaced. If reorientation is not corrected for, the changed positions of the crystal atoms will lead to a perturbation of the system. To avoid perturbation, MOF-VR handles reorientation as follows:

1. At the beginning of each time step, MOF-VR determines whether a system is picked up. In this case, the simulation is paused and all guest molecules in the system are made children of the system's crystal, causing the guest molecules to move and rotate just like the crystal.

2. If the system is released from hand, the rotation necessary to reorient the system, q_{rot}, from the orientation before pick-up, q_{a}, to the orientation after release, q_{e}, is calculated using quaternion multiplication,

$$q_{\mathrm{rot}} = q_{\mathrm{e}} \cdot q_{\mathrm{a}}^{-1}. \tag{S1}$$

3. Interaction site velocities, $\vec{v}_i$, and forces, $\vec{F}_i$, of flexible or semi-flexible guest molecules or torque, $\vec{T}$, angular velocity, $\vec{\omega}$, angular momentum, $\vec{L}$, and orientation, q, of rigid guest molecules are rotated using q_{rot} to correct for the changed orientation of the system. In addition, k vectors for electrostatics are recomputed.

4. The molecular dynamics simulation is continued.

(II) MOF-VR uses dummy atoms for the visualization of atoms that are implicitly modeled in interaction sites. These dummy atoms are children of the interactions sites they are bonded to, so that they translate and rotate just like their parent interaction sites. But in molecular dynamics simulations, interaction sites are treated as points in space that have no spatial extent and thus do not rotate due to forces resulting from an interatomic potential. As a result, the orientation of dummy atoms remains unchanged during a molecular dynamics simulation. To avoid unphysical orientations of dummy atoms, the rotation of the parent interaction side needs to be corrected. When using rigid-body time integration schemes, the rotation required for correcting the orientation of the dummy atoms is equivalent to the rotation the rigid guest molecule experiences. Quaternion multiplication is used to describe the rotation, q_{rot}, of a rigid guest molecule from $q(t - \Delta t)$ to $q(t)$, i.e.:

$$q_{\mathrm{rot}} = q(t) \cdot q^{-1}(t - \Delta t). \tag{S2}$$

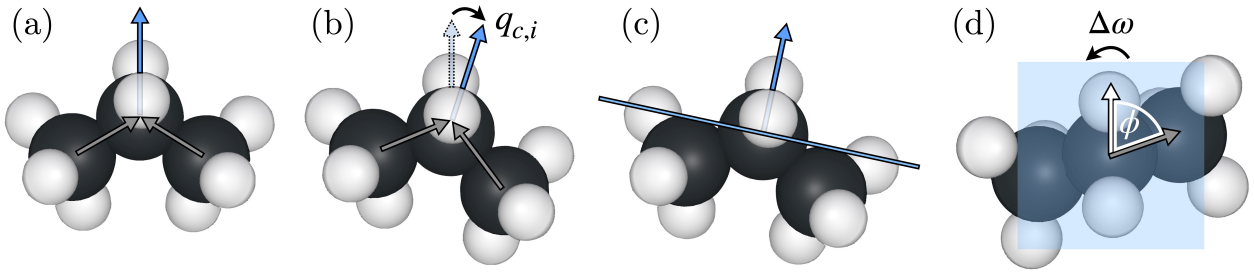

Fig. S1: Illustration of the method for correcting the orientation of dummy atoms in propane. (a) Each time step, the resulting bond vector (blue) is calculated as sum of all bond vectors (gray) originating at neighboring interaction sites j and ending at interaction site i. (b) After new positions of the interaction sites are set in the time integration scheme used, a rotation $q_{c,i}$ is determined to rotate from the resulting bond vector of the previous time step (pale blue) to the resulting bond vector of the current time step (solid blue). (c) Vectors are invariant to rotations about their axes. For this reason, $q_{c,i}$ needs to be corrected by an additional rotation about the resulting bond vector to yield the correct orientation of the dummy atoms. First, a plane (blue line) is defined that passes through the position of the interaction point and has a normal that is parallel to the resulting bond vector (blue arrow). (d) The vectors from one interaction site to a neighboring interaction site or dummy atom (gray or white arrows) are projected onto the plane (pale blue). The angle $\omega(t)$ between the projected vectors is calculated and compared to the (correct) angle $\omega(t - \Delta t)$ of the previous time step. The interaction site is rotated about the resulting bond vector by $\Delta\omega = \omega(t - \Delta t) - \omega(t)$.

The thus calculated quaternion, q_{rot}, can be used to correct the rotation of the interaction site's dummy atoms. When using non-rigid-body time integration schemes, the orientation of guest molecules is not explicitly calculated. But information about the rotation of the guest molecule can be determined from the changing orientation of the bond vectors between neighbor atoms. Fig. S1 illustrates the procedure MOF-VR uses to correct the orientation of dummy atoms in flexible or semi-flexible guest molecules described in the following:

1. Each interaction site i has an orientation vector, $\vec{r}_i$, that is calculated at the end of each time step and results from adding up the vectors that originate at neighboring interaction sites j and end at interaction site i (see Fig. S1a).

2. Each time step, parent interaction sites are rotated by a quaternion to correct the orientation of its child dummy atoms. This rotation is equivalent to the rotation required to rotate from $\vec{r}_i(t - \Delta t)$ to $\vec{r}_i(t)$ (see Fig. S1b).

3. However, vectors are invariant to rotations about their own axes. That's why the correction rotation of the previous step needs to be corrected by an additional rotation about $\vec{r}_i$ to yield the correct orientation of the dummy atom. The additional rotation is determined in four steps:

    i. Define a plane with a normal, $\vec{r}_i(t)$, passing through the position of i (see Fig. S1c).

    ii. Project the bond vector, $\vec{d}_{i\text{-}j}$, between interaction sites i and j onto the plane. If $\vec{d}_{i\text{-}j}$ is parallel to $\vec{r}_i(t)$, project $\vec{r}_j(t)$ onto the plane instead. Interaction sites j are chosen for each interaction site i at the beginning of the simulation (see Fig. S1d).

iii. Project the dummy vector, $\vec{r}_{i\text{-}d}$, between interaction site i and one of its dummy atoms, d, onto the plane. Dummy atoms d are chosen for each interaction site i at the beginning of the simulation (see Fig. S1d).

iv. Calculate the angle between the projected vectors. Compare this angle with the one for the previous time step and rotate the interaction site i about $\vec{r}_i(t)$ by the angular difference (see Fig. S1d).

T Section 5: Kinetic Energy

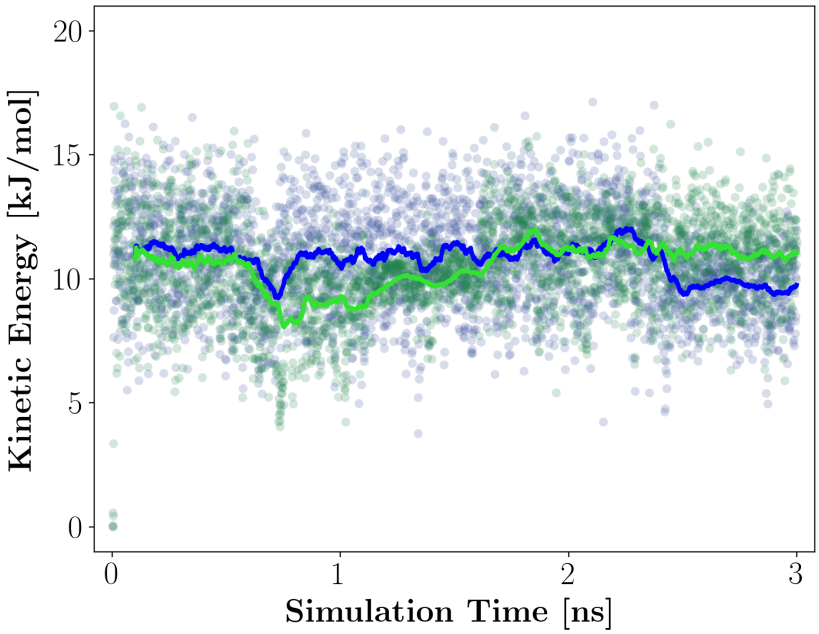

Fig. T1: Average kinetic energies over 1000 time steps (dots) and their moving averages (lines) over 200 time steps calculated by MOF-VR (green) or LAMMPS (blue) for the verification simulations to test the binding forces routine. The starting configuration in LAMMPS was perturbed by shifting the geometric center of butane by -10^{-8} Å in every spatial direction.

U Section 5: Radial Distribution Functions

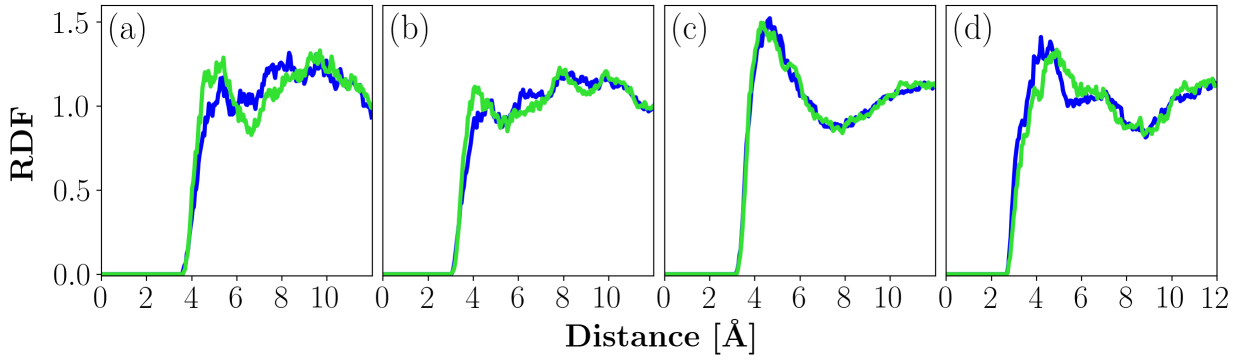

Fig. U1: Radial distribution functions (RDFs) calculated by MOF-VR (green) or LAMMPS (blue) for the verification simulation to test the velocity-Verlet routine. RDFs are shown for interaction site pairs: CH_4 in methane *and* (a) Zn in MOF-5, (b) O in MOF-5, (c) C in MOF-5, (d) H in MOF-5.

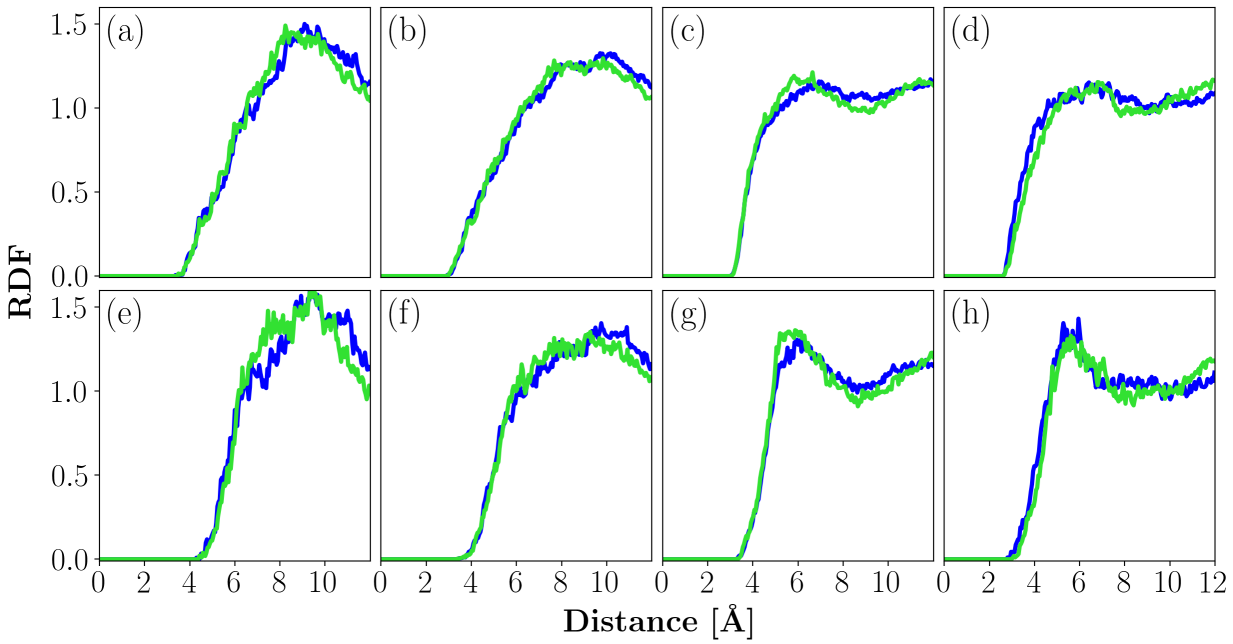

Fig. U2: RDFs calculated by MOF-VR (green) or LAMMPS (blue) for the verification simulation to test the rigid body routine. RDFs are shown for interaction site pairs: CH_3 in 2,3-dimethylbutane *and* (a) Zn in MOF-5, (b) O in MOF-5, (c) C in MOF-5, (d) H in MOF-5, or CH in 2,3-dimethylbutane *and* (e) Zn in MOF-5, (f) O in MOF-5, (g) C in MOF-5, or (h) H in MOF-5.

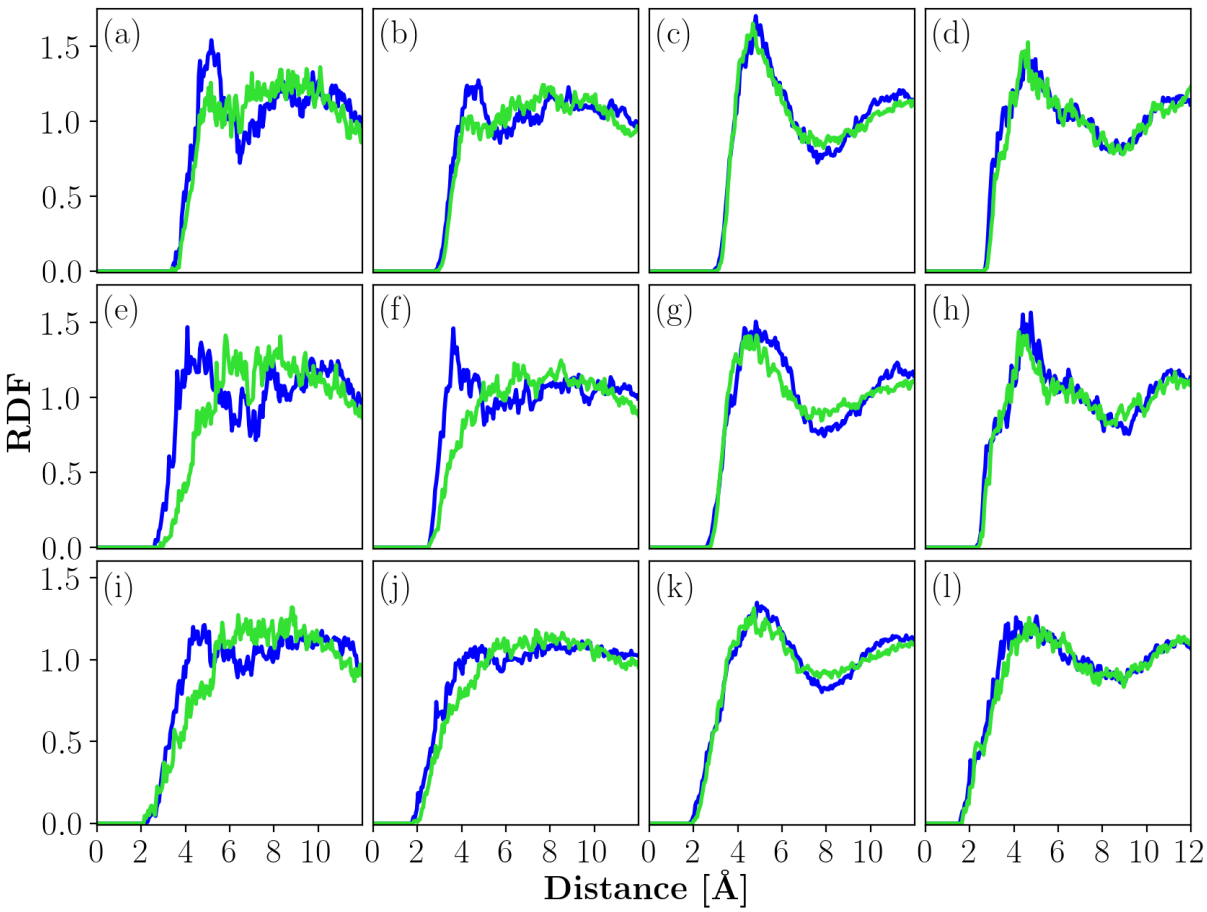

Fig. U3: RDFs calculated by MOF-VR (green) or LAMMPS (blue) for the verification simulation to test the electrostatics routine. RDFs are shown for interaction site pairs: CH_3 in methanol *and* (a) Zn in MOF-5, (b) O in MOF-5, (c) C in MOF-5, (d) H in MOF-5, O in methanol *and* (e) Zn in MOF-5, (f) O in MOF-5, (g) C in MOF-5, (h) H in MOF-5, or H in methanol *and* (i) Zn in MOF-5, (j) O in MOF-5, (k) C in MOF-5, (l) H in MOF-5.

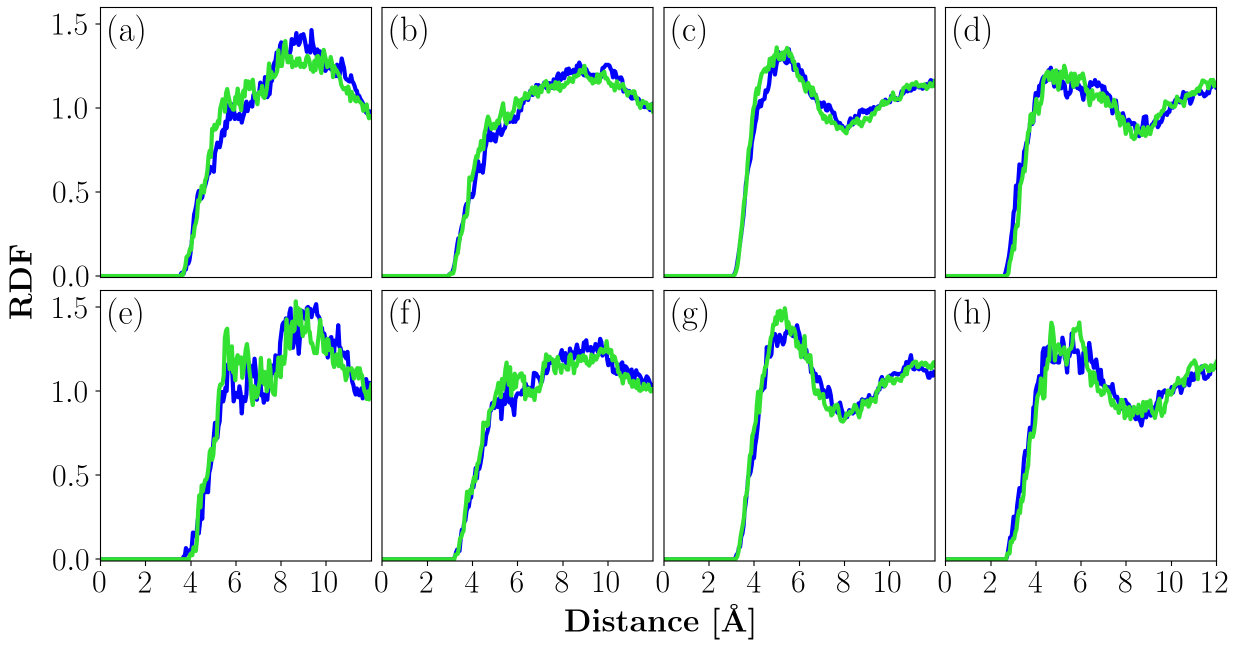

Fig. U4: RDFs calculated by MOF-VR (green) or LAMMPS (blue) for the verification simulation to test the constraints routine. RDFs are shown for interaction site pairs: CH_3 in propane *and* (a) Zn in MOF-5, (b) O in MOF-5, (c) C in MOF-5, (d) H in MOF-5, or CH_2 in propane *and* (e) Zn in MOF-5, (f) O in MOF-5, (g) C in MOF-5, (h) H in MOF-5.

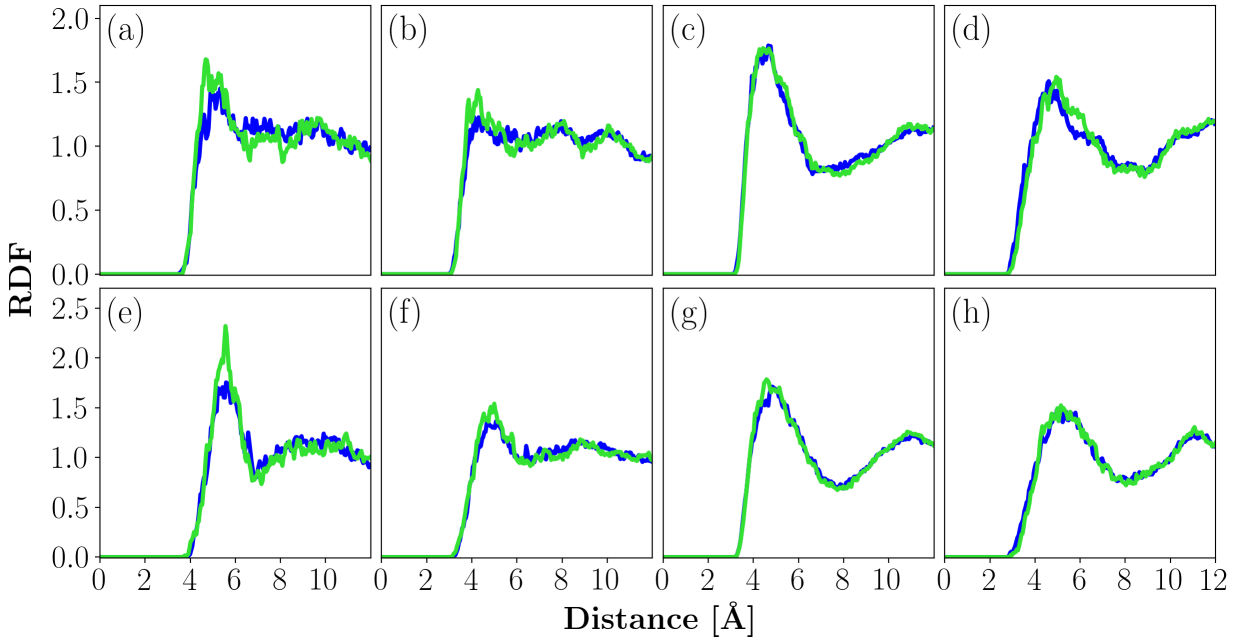

Fig. U5: RDFs calculated by MOF-VR (green) or LAMMPS (blue) for the verification simulation to test the binding forces routine. RDFs are shown for interaction site pairs: CH_3 in n-butane *and* (a) Zn in MOF-5, (b) O in MOF-5, (c) C in MOF-5, (d) H in MOF-5, or CH_2 in n-butane *and* (e) Zn in MOF-5, (f) O in MOF-5, (g) C in MOF-5, (h) H in MOF-5.

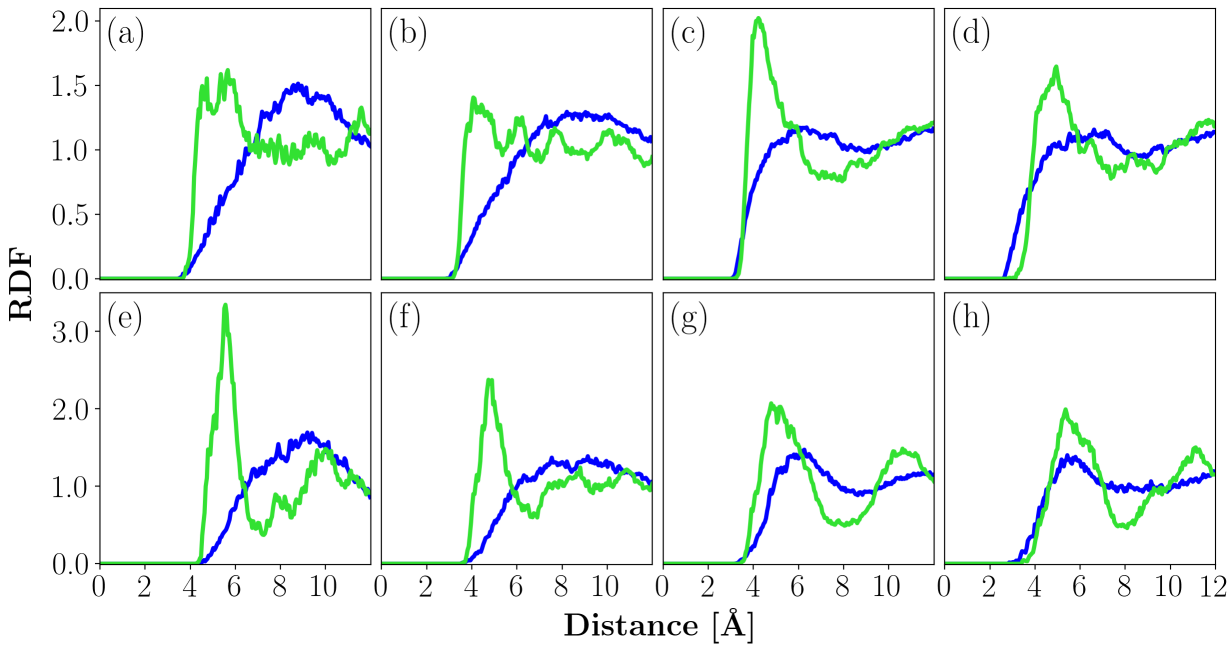

Fig. U6: RDFs calculated by MOF-VR (green) or LAMMPS (blue) for the verification simulation to test the thermostatting routine. RDFs are shown for interaction site pairs: CH_3 in 2,3-dimethylbutane *and* (a) Zn in MOF-5, (b) O in MOF-5, (c) C in MOF-5, (d) H in MOF-5, or CH_2 in 2,3-dimethylbutane *and* (e) Zn in MOF-5, (f) O in MOF-5, (g) C in MOF-5, (h) H in MOF-5.

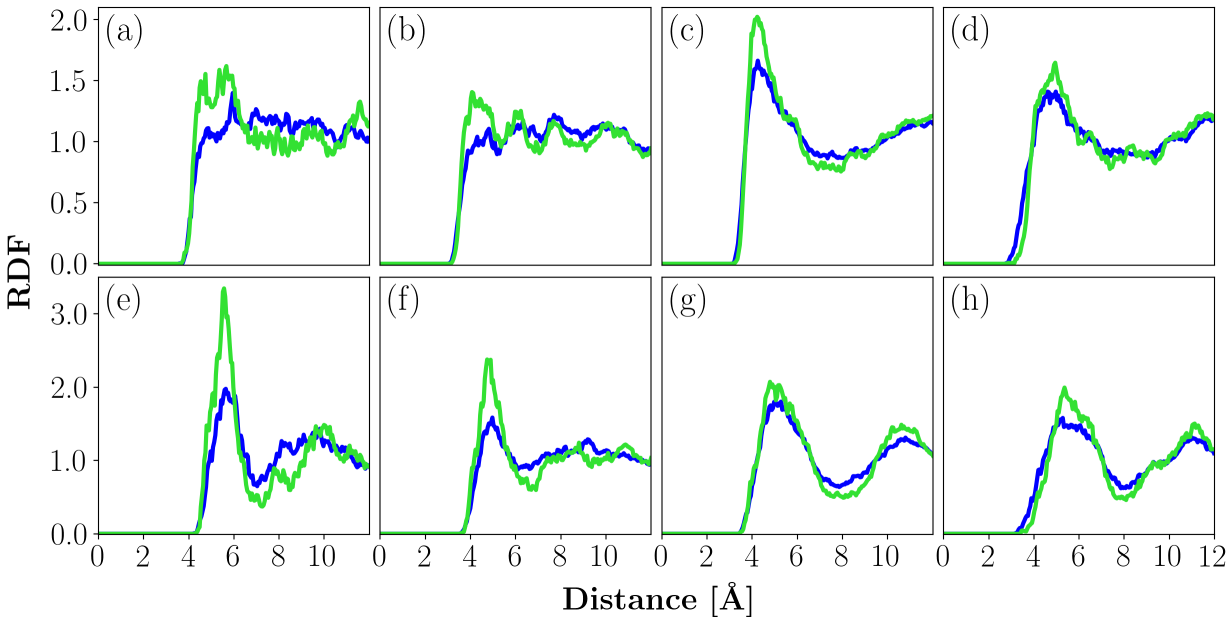

Fig. U7: RDFs calculated by MOF-VR (green) or LAMMPS (blue) for the repeated verification simulation to test the thermostatting routine. In MOF-VR, internal degrees of freedom of 2,3-dimethylbutane were rigidified, while in LAMMPS, a fully flexible 2,3-dimethylbutane molecule was used. The flexible model of 2,3-dimethylbutane was created according to the TraPPE-UA force field [187], except for bond interactions, for which the bond constants were taken from the CHARMM22 [153] force field. RDFs are shown for interaction site pairs: CH_3 in 2,3-dimethylbutane *and* (a) Zn in MOF-5, (b) O in MOF-5, (c) C in MOF-5, (d) H in MOF-5, or CH_2 in 2,3-dimethylbutane *and* (e) Zn in MOF-5, (f) O in MOF-5, (g) C in MOF-5, (h) H in MOF-5.