FACULTY OF ENGINEERING
AND ARCHITECTURE

High-Throughput Screening of Covalent Organic Frameworks for Clean Energy Applications

Juul De Vos

Doctoral dissertation submitted to obtain the academic degree of
Doctor of Engineering Physics

Supervisors
Prof. Veronique Van Speybroeck, PhD* - Prof. Sven Rogge, PhD* - Prof. Pascal Van Der Voort, PhD**

* Department of Applied Physics
Faculty of Engineering and Architecture, Ghent University

** Department of Chemistry
Faculty of Sciences, Ghent University

August 2024

GHENT
UNIVERSITY

ISBN 978-94-6355-863-1
NUR 928, 961
Wettelijk depot: D/2024/10.500/68

Members of the Examination Board

Chair

Prof. Em. Luc Taerwe, PhD, Ghent University

Other members entitled to vote

Prof. Stefaan Cottenier, PhD, Ghent University
Karen Leus, PhD, Ghent University
Prof. Rochus Schmid, PhD, Ruhr-Universität Bochum, Germany
Prof. Toon Verstraelen, PhD, Ghent University

Supervisors

Prof. Veronique Van Speybroeck, PhD, Ghent University
Prof. Sven Rogge, PhD, Ghent University
Prof. Pascal Van Der Voort, PhD, Ghent University

This research has been conducted at the **Center for Molecular Modeling (CMM)** in the framework of a Concerted Research Action, in collaboration with:

- The Centre for Ordered Materials, Organometallics and Catalysis (COMOC) at Ghent University (Prof. P. Van Der Voort).
- The Synthesis, Bioresources and Bioorganic Chemistry Research Group (SynBioC) at Ghent University (Prof. C.V. Stevens).

Preface

Met het schrijven en verdedigen van deze doctoraatsthesis komt er een einde aan mijn tijd aan de universiteit. Een rit die meer dan tien jaar geleden begon in de eerste Bachelor in de Ingenieurswetenschappen. De keuze voor de afstudeerrichting Engineering Physics was een evidentie, aangezien ik al sinds het middelbaar gefascineerd was door het begrijpen van de mechanismen achter natuurfenomenen en het (wiskundig) beschrijven ervan. Tegen de tijd dat ik een masterthesis moest kiezen, was het mij ook duidelijk geworden dat computationele simulaties cruciaal zijn om inzicht te krijgen in deze complexe fenomenen en hun gedrag te voorspellen. De combinatie met een sterk implementatie-aspect en de duidelijke link met maatschappelijk relevante problematieken overtuigde me om voor het Centrum voor Moleculaire Modellering (CMM) te kiezen. Het fundamentele onderzoek maakte me zo enthousiast dat ik uiteindelijk ervoor koos om het onderwerp van mijn masterthesis voort te zetten in een doctoraatsonderzoek, waarvan deze scriptie het finale resultaat is.

Uiteraard was het bereiken van deze mijlpaal onmogelijk zonder de ondersteuning van een heleboel mensen. Van elk van hen heb ik heel veel geleerd en ze hebben me geholpen om mezelf te ontplooien, zowel op professioneel als op persoonlijk vlak.

Eerst en vooral wil ik mijn promotoren, Veronique, Sven, en Pascal, bedanken. Veronique, ik wil je bedanken voor alle kansen die je me gegeven hebt en de voortdurende steun die ik heb mogen ervaren. Je hebt het CMM uitgebouwd tot een stabiele en enthousiaste onderzoeksgroep, met een uitstekende omkadering die ons toelaat om ons op onderzoek te concentreren. Ook heel hard bedankt voor de inzet om mijn interne verdediging nog in mei te laten plaatsvinden, zodat ik me in de zomermaanden optimaal kon focussen op het verwelkomen van onze tweede dochter. Ik vond het een eer om te mogen werken onder jouw supervisie, waardoor ik nu kan zeggen dat ik een doctoraat behaald heb bij de Belgische Nobelprijswinnares. Ik apprecieer het ook enorm dat je de tijd nam om naar het afscheid van papa te komen, ondanks de drukke agenda en vele deadlines in die periode. Dat toont dat je niet alleen een fantastische wetenschapster bent, maar ook een groot mens.

Sven, ik wil je bedanken voor de dagelijkse begeleiding en het aanscherpen van mijn soft skills. Het was een geruststelling te weten dat ik zonder een meeting in te plannen je bureau kon binnenlopen wanneer er een wetenschappelijk probleem opdook of om een discussie te houden. Je hebt ervoor gezorgd dat mijn schrijf- en presenteerskills sterk verbeterd zijn, zeker tijdens de intensieve trainingsperiode om mijn SB-FWO beurs binnen te halen. Ik waardeer de eerlijke en open gesprekken die we konden voeren. Ook bedankt om mij af en toe uit mijn comfortzone te pushen, waardoor ik grenzen kon verleggen en nieuwe dingen kon bijleren.

Pascal, het was een plezier om samen te werken met zo'n sterke experimentele partner en ik wil je graag bedanken voor de kritische experimentele feedback die je hebt gegeven tijdens mijn werk. Ik wil graag jou en prof. Stevens bedanken voor de experimentele samenwerking binnen het GOA-project.

Het is al een tijdje geleden, maar tijdens het binnenhalen van mijn SB-FWO beurs heb ik heel waardevolle input gekregen van Bart Hommez en Lennart Joos. Bart, bedankt voor de coaching over patenten en om het valorisatiepotentieel van de aanvraag te belichten. Lennart, bedankt voor jouw inzicht in carbon capture systemen en het enthousiasme dat je toonde voor het project. Ik ben er oprecht van overtuigd dat jullie bijdrage het project naar een hoger niveau hebben getild.

Graag zou ik ook Sander B. bedanken voor de nauwe samenwerking tijdens mijn onderzoek. Sander, het was een plezier om met je samen te werken. Jouw expertise, inzet en helpende hand hebben een enorme impact gehad op de kwaliteit van mijn werk.

De bovenstaande mensen hebben stuk voor stuk bijgedragen aan mijn wetenschappelijk onderzoek. Maar ook de rest van het CMM wil ik graag bedanken om van de werkvloer een aangename plaats te maken waar het altijd een plezier was om naartoe te komen. Ik zie jullie ondertussen niet als collega's, maar eerder als een groep vrienden. Niet in het minst mijn kantoorgenoten Tom, Massimo, en Kuber, oftwel the Boys. Jullie maakten van ons kantoor een tweede thuis en stonden altijd open om een gezellig babbeltje te slaan of om mijn wetenschappelijke of minder wetenschappelijke frustraties aan te horen. Ik hoop dat de tekening op ons bord ergens een mooi plaatsje krijgt in de CMM-legacy. Ook de groep van generatiegenoten op het CMM (Liesbeth, Ruben, Karen, Sander V., Siebe, en nog eens Tom en Massimo), was het nu de Kindergarten of eerder de OKRA, hebben heel hard geholpen om het groepsgevoel te versterken, met hier en daar eens een drankske, een etentje, of een danske. Ik zou ook graag Aran en Siebe willen bedanken om de FMLs in goeie banen te leiden en Sander V. en Siebe om samen de MaChT-quiz tot een goed einde te brengen. Leen en Mieke wil ik graag bedanken voor de uitstekende technische en administratieve hulp doorheen de jaren.

Ook naast het werk kon de boog niet altijd gespannen staan. Daarvoor wil ik graag Jarl en Mathijs bedanken, voor onze tweewekelijkse (of vaak zelfs frequentere) afspraak in het Jan Breydelstadion om stoom af te laten, en ook Steve, voor de gezellige etentjes en ontspanningsmomenten. Daarnaast wil ik de giants Alexander, Emiel, Noah, en Aragorn bedanken voor de epische Machiavelli-battles en bijhorende babbels.

Twee mensen zonder wie dit alles helemaal onmogelijk zou zijn geweest zijn mijn ouders. Bedankt voor al jullie goeie zorgen, jullie liefde, en om me de ruimte te geven om mijn vleugels uit te slaan en te ontdekken wat ik graag doe. Mama, jij bent sinds kinds af aan een onvoorwaardelijke steun geweest en bent altijd blijven geloven in wat ik deed, ook toen het wat lastiger werd. Ook bedankt om een fantastische omie te zijn en voor alle leuke momenten die je al aan Odette gegeven hebt. Papa, jij hebt mij de schoonheid van de wereld getoond en mijn nieuwsgierigheid om alles te ontdekken aangewakkerd. Ik ben heel dankbaar dat ik in jullie warme nest heb mogen opgroeien en voor de diepe wortels die jullie mij gegeven hebben.

Ook wil ik graag Els en Danny bedanken. In de eerste plaats om een prachtige dochter op te voeden, maar ook om hier en daar te helpen bij een klusje in ons huis, waar ik me zelf te onzeker over voelde om met mijn twee linkerhanden aan te pakken. Ook bedankt om een oma en opa te zijn die steeds met veel liefde voor Odette paraat staan.

Tot slot zou ik ook mijn gezin willen bedanken. Odette, hopelijk heb jij niet teveel gemerkt van dit hele avontuur en heb ik steeds mijn papa-rol kunnen blijven uitvoeren. Bedankt om je vrolijke zelve te zijn en altijd een lach op mijn gezicht te toveren wanneer je om een knuffel kwam vragen. Ook bedankt om mij via de babyfoon bij te staan tijdens de late uren en nachtjes waarin deze thesis afgerond moest worden. Giselle, jij bent pas helemaal op het einde van deze periode in beeld gekomen. Jij hebt me de motivatie gegeven om deze thesis op tijd af te krijgen, opdat ik in jouw eerste maanden de ruimte had om te genieten van je aanwezigheid. Ik wil dat jullie weten hoe ongelooflijk trots ik ben dat ik jullie papa mag zijn. Ik zal proberen om jullie alle liefde van de wereld te geven en er altijd voor jullie te zijn als jullie mij nodig hebben. Ik kijk ernaar uit om langs de zijlijn te staan om te zien hoe jullie jezelf ontplooien en de wereld ontdekken.

Caro, mijn liefste mormeltje, jij was de veilige thuishaven die mij de tijd en ruimte heeft gegeven om dit onderzoek tot een goed einde te brengen.

Bedankt om mij altijd te blijven steunen doorheen de jaren, voor al je geduld, en voor het luisterende oor wanneer ik het nodig had. Jij kon de druk bij naderende deadlines verlichten en hielp om alles te relativeren, waardoor de afgelopen jaren veel gemakkelijker werden. Ik ben zo dankbaar en trots dat ik naast je mag staan en dat we alle avonturen met elkaar kunnen delen. Bedankt om elke dag een lach op mijn gezicht te toveren en voor het plezier in mijn leven te zorgen. Je bent een pracht van een vrouw en ik hou zo ongelooflijk veel van jou.

Juul De Vos
Gent, augustus 2024

Contents

Preface iii

Contents vii

List of Symbols xi

List of Abbreviations xv

Samenvatting xix

Summary xxiii

I High-throughput screening of covalent organic frameworks for clean energy applications 1

1 Introduction 3

- 1.1 Nanoporous materials 5
- 1.2 Covalent organic frameworks and selected applications 7
  - 1.2.1 Carbon capture 8
  - 1.2.2 Vehicular methane storage 11
  - 1.2.3 Heterogeneous catalysis 12
- 1.3 The haystack of reticular chemistry 12
- 1.4 Computational high-throughput screenings 16
- 1.5 Goal and outline 19

2 Molecular modeling tools 21

2.1 From quantum mechanics to force fields 21

2.1.1 The Schrödinger equation and Born-Oppenheimer approximation 22

2.1.2 Force fields 26

I. Energy expression 27

II. Parametrization of force fields 32

III. From cluster force fields to periodic force fields 33

2.2 Sampling of the potential energy surface 36

2.2.1 Static exploration of the PES 36

I. Geometry optimization 37

II. Vibrational analysis 37

2.2.2 Dynamic exploration of the PES 38

I. Molecular dynamics 38

II. Monte Carlo 39

3 *In silico* design of covalent organic frameworks 41

3.1 Automated structure assembly 42

3.2 Implementation in Python 48

3.2.1 Core objects 48

3.2.2 Workflow 51

3.3 Case studies 52

4 ReDD-COFFEE: a diverse database of covalent organic framework structures and force fields 63

4.1 Generation of the ReDD-COFFEE database 64

4.1.1 Initialization of the SBUs and topological embeddings 65

4.1.2 Enumeration of trial combinations 67

4.1.3 Thresholds to filter out unphysical structures 69

4.1.4 Derivation of the system-specific force fields 70

4.2 Investigating the accuracy of the system-specific force fields . 71

4.2.1 Comparison with *ab initio* clusters 71

4.2.2 Prediction of experimental, structural characteristics . 72
4.3 Analyzing the diversity of the database 74
4.3.1 Distribution of linkage types in COF databases 75
4.3.2 Quantitative diversity metrics 76
4.3.3 Selection of a diverse subset 78

5 High-throughput screenings of covalent organic frameworks 81
5.1 Textural features of nanoporous materials and COF property-property relations 82
5.2 Determining the performance limits of COFs for vehicular methane storage 85
5.3 Identifying promising COF candidates for carbon capture . . 87
5.3.1 Step 1: screening of the diverse subset 90
5.3.2 Step 2: machine learning prediction of the working capacity and selectivity 91
5.3.3 Step 3: mixture selectivity of the most promising COFs 95
5.3.4 Design strategies for top-performing COFs 96

6 Conclusions and perspectives 99
6.1 Conclusion 99
6.2 Perspectives 102

II Published Papers 107

A Selection of Publications in International Peer-Reviewed Journals 109
Paper I: Quantifying the Likelihood of Structural Models through a Dynamically Enhanced Powder X-Ray Diffraction Protocol . 111
Paper II: Exploring the phase stability in interpenetrated diamondoid covalent organic frameworks 123
Paper III: ReDD-COFFEE: a ready-to-use database of covalent organic framework structures and accurate force fields to enable high-throughput screenings 133

Paper IV: Engineering of Phenylpyridine- and Bipyridine-Based Covalent Organic Frameworks for Photocatalytic Tandem Aerobic Oxidation/Povarov Cyclization 155
Paper V: High-Throughput Screening of Covalent Organic Frameworks for Carbon Capture using Machine Learning 171

B List of Publications 189

Publications in international peer–reviewed journals 189
Conference contributions 191
  Oral presentations 191
  Poster presentations 191
Master's thesis 192

C List of Software Packages 193

D Acknowledgements 197

Bibliography 199

List of Symbols

Within this thesis, italicized variables indicate scalars and italicized, bold–faced variables indicate vectors.

Alphanumeric symbols	
$\boldsymbol{a}$	First unit cell vector
A_{acc}	Accessible surface area
A_{grav}	Gravimetric accessible surface area
A_{vol}	Volumetric accessible surface area
Å	Ångström
A_{ϕ}	Amplitude of a torsional barrier
$\boldsymbol{b}$	Second unit cell vector
B	Balance
$\boldsymbol{c}$	Third unit cell vector
d	Out-of-plane distance or RAC depth
D	Disparity
d^0	Rest out-of-plane distance
E	Energy
E^{clust}	Force field energy of a cluster
E_{cov}	Covalent force field energy
E_{cross}	Force field energy of the cross terms
E_{bend}	Force field energy of the bend terms
E_{bond}	Force field energy of the bond terms
E_{def}	Deformation energy
E_{ei}	Electrostatic force field energy
E_{FF}	Force field energy
E^{inter}	Force field energy of the interactions between SBUs
E^{intra}	Force field energy of the interactions within SBUs
$E_{\text{non-cov}}$	Non-covalent force field energy

$E_{\text{out-of-plane}}$	Force field energy of the out-of-plane terms
E^{per}	Energy of the periodic material
E_{torsion}	Force field energy of the torsion terms
E_{vdW}	Van der Waals force field energy
E_n^e	n^{th} Born-Oppenheimer surface
E_{ij}	j^{th} edge belonging to Wyckoff set W_i
E_{nm}	Energy eigenvalue
f	Rescaling factor
$\boldsymbol{F}_i$	Force acting on atom i
f_s	Normal mode frequency
h	Planck constant
$\hbar$	Reduced Planck constant, with $\hbar = h/2\pi$
$\hat{\mathcal{H}}$	Hamiltonian operator
i	Imaginary unit
I	Atom identity
K	Force constant
K_{oop}	Force constant associated with an out-of-plane distance
K_r	Force constant associated with a bond distance
K_θ	Force constant associated with a bending angle
m_i	Mass of atom i
M	Unit cell mass or multiplicity of a torsional barrier
N	Number of particles
N_c	Upper limit for the number of (topology, SBUs) combinations
N_x^{ads}	Uptake capacity of guest species x at adsorption conditions
N_x^{des}	Uptake capacity of guest species x at desorption conditions
p	Bond charge increment
P	Pressure
p^{clust}	Cluster force field parameter
p^{per}	Periodic force field parameter
$p(\boldsymbol{s})$	Probability of a configuration $\boldsymbol{s}$
q	Average energy of adsorption
q_i	Partial charge of atom i
Q_{st}	Isosteric heat of adsorption
r	Bond distance
r^0	Rest bond length
R	Gas constant
$\boldsymbol{R}$	Nuclear coordinates
R_{wp}	Weighted profile residual
$\boldsymbol{s}$	Configuration of molecular system
S	Covalent radius
S_{ideal}	Ideal selectivity

S_{mix}	Mixture selectivity
t	Time
T	Temperature or connectivity
$\hat{\mathcal{T}}_e$	Kinetic energy operator associated with the electrons
$\hat{\mathcal{T}}_N$	Kinetic energy operator associated with the nuclei
V	Unit cell volume or variety
V_{acc}	Accessible volume
$\hat{\mathcal{V}}_{ee}$	Coulomb potential energy operator between the electrons
V_{grav}	Gravimetric accessible volume
V_{ij}	j^{th} vertex belonging to Wyckoff set W_i
$\hat{\mathcal{V}}_{Ne}$	Coulomb potential energy operator between the electrons
$\hat{\mathcal{V}}_{NN}$	Coulomb potential energy operator between the electrons and the nuclei
V_{vol}	Volumetric accessible volume or pore fraction
W_i	Wyckoff set i
y_x	Mole fraction of guest species x in a multicomponent gas
Y	Intensity of the PXRD profile
Z	Nuclear charge

Greek symbols	
α	Scaling factor or polarizability
β	Thermodynamic beta, with $\beta = 1/k_B T$
δ	Partial charge or Kronecker delta
ΔN_x	CH_4 deliverable capacity (x = CH_4) or CO_2 working capacity (x = CO_2)
Δt	Timestep
ϵ_i	Force field parameter for the van der Waals interactions
θ	Bending angle or Bragg angle
θ^0	Rest bending angle
μ	Chemical potential
$\rho(\boldsymbol{r})$	Electron density
σ_f	Standard deviation of rescaling factors f
$\sigma_{f,\text{max}}$	Threshold for the standard deviation of rescaling factors f
σ_i	Force field parameter for the van der Waals interactions
ϕ	Dihedral angle
ϕ^0	Rest dihedral angle
χ	Pauling electronegativity or out-of-plane angle
$\lvert\psi_{nm}\rangle$	Stationary eigenstates of the time-independent Schrödinger equation
$\lvert\psi^e\rangle$	Electronic wavefunction

$\lvert\psi_n^e\rangle$	n^{th} electronic wavefunction
$\lvert\psi^N\rangle$	Nuclear wavefunction
$\lvert\psi_{nm}^N\rangle$	nm^{th} nuclear wavefunction
$\lvert\Psi(t)\rangle$	Total wavefunction of a molecular system

Mathematical symbols

$\langle x_i\rangle$	Mean value of scalars x_i
$N!$	Factorial operator
$\forall$	For all
$\boldsymbol{\nabla}$	Vector differential operator

List of Abbreviations

AASBU	Automated assembly of secondary building units
ANG	Adsorbed natural gas
API	Application programming interface
AP-RDF	Atomic-property weighted radial distribution function
$\mathbf{APS_{ideal}}$	Ideal adsorbent performance score
$\mathbf{APS_{mix}}$	Mixture adsorbent performance score
ARPA-E	Advanced Research Projects Agency-Energy
AuToGraFS	Automated topological generator for framework structures
BFGS	Broyden–Fletcher–Goldfarb–Shanno optimization
CCS	Carbon capture and sequestration
CCU	Carbon capture and utilization
CMM	Center for Molecular Modeling
CNG	Compressed natural gas
COF	Covalent organic framework
COMOC	Centre for Ordered Materials, Organometallics and Catalysis
config	Configuration
CoRE	Computation-Ready Experimental
CSD	Cambridge Structural Database
CURATED	Clean, Uniform, and Refined with Automatic Tracking from Experimental Database
DFT	Density functional theory
FF	Force field
GAN	Generative adversial network
GB	Gradient boosting regression
GCMC	Grand-canonical Monte Carlo
GGA	Generalized gradient approximation
HF	Hartree-Fock
hMOF	Hypothetical MOF database of Wilmer *et al.*
HPC	High-performance computing

IC	Internal coordinate
I.E.	Individual assignment of the edge Wyckoff sets
IUPAC	International Union of Pure and Applied Chemistry
I.V.	Individual assignment of the vertex Wyckoff sets
IZA	International Zeolite Association
KRR	Kernel ridge regression
LDA	Local density approximation
LNG	Liquid natural gas
l.u.	Length unit
MAE	Mean absolute error
MBIS	Minimal Basis Iterative Stockholder
MC	Monte Carlo
MD	Molecular dynamics
ML	Machine learning
MLP	Multi-layer perceptron regression or machine learning potential
MOF	Metal-organic framework
MOVE	Methane opportunities for vehicular energy
MSD	Mean signed deviation
oop	Out-of-plane distance
PCC	Pearson correlation coefficient
PES	Potential energy surface
PXRD	Powder X-ray diffraction
QMOF	Quantum MOF database of Rosen *et al.*
RAC	Revised autocorrelation function
RCSR	Reticular Chemistry Structure Resource
ReDD-COFFEE	Ready-to-use and Diverse Database of Covalent Organic Frameworks with Force field based Energy Evaluation
RF	Random forest regression
RMSD	Root mean square deviation
$\mathbf{RMSD_{max}}$	Threshold for the root mean square deviation
SBU	Secondary building unit
STP	Standard temperature and pressure
SRCC	Spearman rank correlation coefficient
SVR	Support vector regression
SynBioC	Synthesis, Bioresources and Bioorganic Chemistry Research Group
ToBaCCo	Topologically-Based Crystal Constructor
TraPPE	Transferable Potentials for Phase Equilibria
t-SNE	T-distributed Stochastic Neighbor Embedding
UCiC	Unit cell information capacity

U.E.	Unified assignment of the edge Wyckoff sets
UFF	Universal force field
US	United States
U.V.	Unified assignment of the vertex Wyckoff sets

Samenvatting

Materiaalkunde is een interdisciplinair onderzoeksgebied op het grensvlak van chemie, fysica, en ingenieurswetenschappen, met als doel het ontwerpen van nieuwe functionele materialen met uitstekende eigenschappen voor verschillende toepassingen. Onder de onderzochte materialen vormt de klasse van nanoporeuze materialen, zoals zeolieten, metaal-organische roosters (MOFs, van het Engelse *metal-organic frameworks*), en covalente organische roosters (COFs, van het Engelse *covalent organic frameworks*), een veelbelovend platform omwille van de grote experimentele controle over hun porositeit op de nanoschaal. Hun grote interne oppervlakte stelt gasmoleculen in staat om de interactie met de poriewanden te maximaliseren, resulterend in unieke eigenschappen voor gasopslag, separatie, en katalyse. Meer specifiek worden duurzame-energietoepassingen, zoals CO_2-afvang na verbranding en opslag van methaan voor voertuigen, intensief onderzocht om de impact van de huidige industrie op het klimaat te verminderen en minder afhankelijk te worden van de eindige hoeveelheid fossiele brandstoffen.

De recente ontwikkeling van COFs heeft veel aandacht gekregen binnen de wetenschappelijke en industiële wereld. Deze nanoporeuze materialen bestaan uitsluitend uit organische, lichte bouwstenen die verbonden zijn door sterke covalente bindingen, waardoor ze een hoge mechanische, thermische, en chemische stabiliteit hebben. Bovendien kunnen hun fysische en chemische eigenschappen op het moleculair niveau geregeld worden door de bouwstenen of de manier waarop ze zijn verbonden, d.w.z. hun topologie, te veranderen in een concept genaamd reticulaire chemie. Hierdoor zijn COFs de ideale constructiematerialen, aangezien hun structuren op zeer modulaire manier kunnen worden aangepast om een specifieke eigenschap te verkrijgen.

Reticulaire chemie stelt ons echter ook voor de uitdaging om het geschikte materiaal te vinden voor de juiste toepassingen, aangezien er in principe een ongelimiteerd aantal van deze materialen bedacht kunnen worden, vertrekkende van een zeer groot aantal bouwblokken en topologieën. Experimentele screening van deze uitgestrekte materiaalruimte om de beste COF voor een bepaalde toepassing te identificeren zou een tijdrovende, zo niet onmogelijke,

taak zijn. De recente ontwikkeling van supercomputers heeft uitgebreide computationele screenings mogelijk gemaakt om deze zoektocht naar materialen te versnellen. In dergelijke screenings wordt meestal gestart van een database van materialen, die in twee klassen ingedeeld kunnen worden. Aan de ene kant zijn er de experimentele databases met structuren die verkregen zijn uit experimentele metingen zoals röntgendiffractie. Vaak focussen deze databases zich te veel op subklassen die veel voorkomen in de literatuur, en benadrukken screenings die hiervan gebruikmaken onvoldoende de minder onderzochte subklassen. Aan de andere kant zijn er de hypothetische databases, die materialen bevatten die computationeel gegenereerd zijn en mogelijk nog niet gesynthetiseerd werden. Alhoewel deze databases een brede en evenwichtige representatie van de materiaalruimte bieden, bestrijken ze er vaak slechts een beperkt deel van.

Om de ontwikkeling van COFs te versnellen en experimentele onderzoekers naar veelbelovende materialen te leiden, hebben we in deze thesis een diverse en direct bruikbare database van COFs gegenereerd. Vervolgens werd een computationele screening uitgevoerd voor twee duurzame-energietoepassingen: CO_2-afvang na verbranding en opslag van methaan voor voertuigen. Zoals geïntroduceerd in Hoofdstuk 3, wordt de database gegenereerd met behulp van een geautomatiseerd ontwerpalgoritme, dat geïmplementeerd is in een Python-pakket. De methodologie is gebaseerd op een bouwblokconcept waarin de topologische inbedding dient als fundering die vervolgens wordt gedecoreerd met een set van bouwblokken. In elke stap van de iteratie wordt weloverwogen een gunstige configuratie van een bouwblok geselecteerd op basis van geometrische en energetische overwegingen om uiteindelijk een plausibel periodiek materiaal te verkrijgen. In principe hangt de voorkeur voor een specifieke bouwblokconfiguratie af van de configuraties van de aangrenzende bouwstenen, en moeten daarom alle materiaalconfiguraties geëvalueerd worden. Aangezien dit aantal zeer groot kan worden, verhoogden we de computationele efficiëntie van ons algoritme door een additieve benadering te implementeren, waarbij de bouwblokken één voor één worden toegevoegd. Het gebruik van ons algoritme wordt gedemonstreerd in drie toepassingen.

Ons geautomatiseerd ontwerpalgoritme wordt toegepast op een groot aantal combinaties van topologieën en bouwblokken in Hoofdstuk 4 om de ReDD-COFFEE (van het Engelse *Ready-to-use and Diverse Database of Covalent Organic Frameworks with Force field based Energy Evaluation*) database te genereren, bestaande uit 268 687 COF-structuren en systeemspecifieke krachtvelden. Deze database heeft een grote diversiteit qua poriegeometrie, linkers, en verbindingen, maar omvat slechts een beperkt aantal functionele groepen. De systeemspecifieke krachtvelden bieden een efficiënte maar nauwkeurige beschrijving van de atomaire interacties. Ze zijn afgeleid van de clusterkrachtvelden van de respectievelijke bouwblokken waaruit elke COF opgebouwd is en die op hun beurt gefit zijn aan kwantummechanische referentiedata. Hiervoor werd gebruik gemaakt van het QuickFF-concept dat eerder binnen het Centrum voor Moleculaire Modellering werd ontwikkeld. Door deze aanpak kunnen een groot aantal systeemspecifieke krachtvelden worden afgeleid uit een beperkt aantal *ab initio* berekeningen. De verbeterde nauwkeurigheid van deze krachtvelden ten opzichte van de generieke krachtvelden die meestal worden gebruikt in uitgebreide screenings is aangetoond door structurele en vibrationele materiaaleigenschappen te voorspellen. Door de ReDD-COFFEE-database vrij beschikbaar te maken via de Materials Cloud, hopen we dat het andere onderzoekers zal inspireren om de mogelijkheden van deze intrigerende functionele materialen verder te verkennen.

In Hoofdstuk 5 wordt de ReDD-COFFEE-database gebruikt in drie uitgebreide screeningstudies. Ten eerste worden de texturele eigenschappen van COFs gekarakteriseerd en worden relaties tussen deze eigenschappen vastgesteld. Deze worden bovendien vergeleken met andere klassen van nanoporeuze materialen, zoals MOFs en zeolieten. Vergeleken met deze materiaalklassen bezitten COFs de laagste massadichtheden en de hoogste gravimetrisch toegankelijke interne oppervlakten en poriefracties. Ten tweede worden de prestatiegrenzen van COFs voor de opslag van methaan voor voertuigen onderzocht. Hoewel COFs aan de ARPA-E-doelstelling voor de gravimetrische leverbare capaciteit van aardgas kunnen voldoen, wordt de doelstelling voor de volumetrische leverbare capaciteit niet gehaald. Dit is vergelijkbaar met de huidige best presterende nanoporeuze materialen, zoals Co(bdp) en NJU-Bai 43. Het is dus nog te vroeg voor de commerciële ontwikkeling van COFs en andere nanoporeuze materialen voor geadsorbeerd aardgas (ANG, van het Engelse *adsorbed natural gas*) in het wegtransport.

Ten derde worden veelbelovende kandidaten voor de afvang van CO_2 uit puntbronnen na verbranding geïdentificeerd. Tijdens deze screening maken we gebruik van machinaal leren om de CO_2-werkcapaciteit en ideale CO_2/N_2-selectiviteit van alle COFs in de database te voorspellen. Op deze manier kunnen we 3305 materialen met een hoog potentieel selecteren in slechts een fractie van de benodigde rekenkracht voor de karakterisering van de volledige database. De mengselselectiviteit van deze materialen werd vervolgens bepaald om competitieve adsorptie en co-adsorptie in rekening te brengen. De resultaten van onze studie tonen aan dat COFs met een poriediameter van ongeveer 1.0 nm een relatief hoge werkcapaciteit en ideale selectiviteit vertonen. Deze materialen hebben adsorptieplaatsen met een hoge CO_2-affiniteit in poriën die omgeven worden door aromatische ringen, die vaak voorkomen in de bestpresterende 3D COFs. Deze sites zijn meestal afwezig in de gelaagde structuur van 2D COFs. In deze materialen is de chemische samenstelling van het rooster een cruciale factor om de CO_2-afvang te verbeteren. Onder de bestudeerde functionele groepen vertonen fluorsubstituenten een superieure prestatie. Bovendien vertonen COFs met een imide-, (acyl)hydrazone-, of (keto)enamine-verbinding een hogere prestatie voor CO_2-afvang. Uit de top-20 van 2D en 3D COFs werden drie kandidaten geselecteerd voor experimentele synthese op basis van de beschikbaarheid van een synthetische route.

We hopen dat deze scriptie kan bijdragen aan het snel evoluerende onderzoek naar COFs met op maat gemaakte eigenschappen en dat we hun ontwikkeling in duurzame-energietoepassingen hebben kunnen versnellen. Dit werk verlegt niet alleen de grenzen van het onderzoek naar COFs, maar benadrukt ook het potentieel van computationele studies bij het sturen van experimentele inspanningen naar de synthese van geavanceerde en functionele materialen. De inzichten verworven in deze thesis zijn niet alleen van toepassing op het gebied van COFs, maar kunnen ook dienen als blauwdruk voor het systematisch ontwerpen van nieuwe materialen binnen het bredere domein van de materiaalkunde.

Summary

Materials science is an interdisciplinary research field at the intersection of chemistry, physics, and engineering with the goal of designing new functional materials with outstanding properties for various applications. Among the investigated materials, the class of nanoporous materials, such as zeolites, metal-organic frameworks (MOFs), and covalent organic frameworks (COFs), forms a promising platform due to their tailored porosity at the nanoscale. Their large internal surface area enables guest molecules to maximize the interaction with the pore walls, resulting in unique properties for gas separation, energy storage, catalysis, and many more applications. Of particular interest are clean energy applications, such as carbon capture and vehicular methane storage, which are heavily investigated to reduce the impact of our current industry on the climate and to become less dependent on the finite amount of fossil fuels.

The recent development of COFs has sparked a lot of attention within the scientific and industrial community. These nanoporous materials are solely composed of organic and lightweight building blocks, that are connected by strong covalent bonds, which endows them with a large mechanical, thermal, and chemical stability. Furthermore, their physicochemical properties can be tuned on the molecular level by changing the building blocks or the way in which they are connected, *i.e.*, their topology, a concept coined reticular chemistry. This makes COFs the ideal engineering materials, since their structures can be tuned in a highly modular approach to obtain a targeted property.

However, reticular chemistry also comes with the challenge to identify the ideal material for a specific application, since the number of materials that can be synthesized is almost unlimited, starting from a large pool of possible building blocks and topologies. Experimental screening of this vast material space to identify the best COF for a certain application would be a cumbersome, if not unfeasible, task. The recent development of high-performance computing (HPC) facilities has enabled computational high-throughput screenings to tremendously accelerate this quest for materials. In such high-throughput screenings a large set of materials is computationally

characterized to find promising candidates, establish property-property relationships, or predict performance limits for a targeted application. Typically, the starting point of such screening study is a materials database, which can be divided in two classes. On the one hand, there are experimental databases containing structures that have been extracted from experimental measurements such as single-crystal X-ray diffraction. Often, these databases are biased towards subclasses that are omnipresent in literature, and screenings using them do not sufficiently emphasize the less researched subclasses. On the other hand, there are hypothetical databases, containing materials that are generated *in silico* and are possibly not-yet-synthesized. Whereas they provide a broad and well-balanced representation of material space, they often cover only part of it.

To accelerate the development of COFs and guide experimental researchers towards promising materials, we generated in this thesis a diverse and ready-to-use database of COFs. Subsequently, we performed a computational high-throughput screening on two clean energy applications, *i.e.*, post-combustion carbon capture and vehicular methane storage. As introduced in Chapter 3, the database is generated using an automated *in silico* assembly algorithm, which is implemented in a Python package. This methodology is based on a building block concept in which the topological embedding acts as blueprint which is decorated with a set of building blocks. In each step of the iteration, the favorable building block configuration is selected based on geometric and energetical considerations, to generate a plausible periodic material. In principle, the favor for a specific building block configuration also depends on the configurations of the neighboring building blocks, and therefore all materials configurations should be evaluated. Since this number can become very large, the computational efficiency of our algorithm is improved by implementing an additive approach, in which the building blocks are added one-by-one. The applicability of our top-down approach is demonstrated by three case studies.

Our automated assembly algorithm is adopted on a large number of combinations of topologies and building blocks in Chapter 4 to generate the Ready-to-use and Diverse Database of Covalent Organic Frameworks with Force field based Energy Evaluation (ReDD-COFFEE) of 268 687 COF structures and system-specific force fields. The database has a large diversity in terms of pore geometry, linkers, and linkages, but only includes a limited number of functional groups. The system-specific force fields provide an efficient, yet accurate, description of the atomic interactions. They are derived from the cluster force fields of the respective building blocks that constitute each COF. These cluster force fields are fitted to quantum mechanical reference data using the QuickFF concept developed earlier at the Center for Molecular

Modeling. With this approach, a large number of system-specific force fields can be derived with only a limited number of *ab initio* calculations. The improved accuracy of these force fields over generic force fields often used in high-throughput screenings is illustrated by predicting structural and vibrational characteristics. By making the ReDD-COFFEE database freely available via the Materials Cloud, we hope it may inspire fellow researchers to continue exploring the potential of these intriguing functional materials.

In Chapter 5, the ReDD-COFFEE database is employed in three high-throughput screening studies. Firstly, the textural properties of COFs are characterized and property-property relations between them are established. Furthermore, these are compared with other classes of nanoporous materials, *i.e.*, MOFs and zeolites. Compared to these material classes, COFs exhibit the lowest mass densities and highest gravimetric accessible surface areas and pore fractions. Secondly, the performance limits of COFs for vehicular methane storage are investigated. Whereas COFs are able to meet the ARPA-E target for gravimetric deliverable capacity of natural gas, the target for the volumetric deliverable capacity is not reached, similar to the current top-performing nanoporous materials, such as Co(bdp) and NJU-Bai 43. Therefore, COFs and other nanoporous materials are too premature for commercial development of storage materials for adsorbed natural gas (ANG) in vehicular transport.

Thirdly, promising candidates for carbon capture from post-combustion point sources are identified. During this screening, we adopt machine learning to predict the CO_2 working capacity and ideal CO_2/N_2 selectivity of all COFs in the database. As such, we could select a total of 3305 materials with high potential in only a fraction of the computational time required for a full database characterization. These materials are subsequently characterized with mixture GCMC calculations to take competitive adsorption and coadsorption into account. The results of our study demonstrate that a relatively high working capacity and ideal selectivity are found in COFs with a pore diameter around 1.0 nm. These materials exhibit adsorption sites with high CO_2 affinity in pores enclosed by aromatic rings, which are commonly found in top-performing 3D COFs. In contrast, the layered nature of 2D COFs results in an absence of such high performing adsorption sites. In these COFs, the framework chemistry emerges as a crucial factor to enhance the carbon capture performance. Among the studied functional groups, fluor substituents show superior performance. Furthermore, COFs with imide, (acyl)hydrazone, or (keto)enamine linkages demonstrate a higher potential for post-combustion carbon capture. Among the top-20 2D and 3D COFs, we identified three three candidates for experimental synthesis based on the availability of a synthetic route.

We hope that this thesis may contribute to the developing research on COFs with tailored properties and that we have accelerated their adoption in clean energy applications. This work not only pushes the boundaries of COF research, but also underscores the potential of computational studies in steering experimental endeavors towards the synthesis of advanced and functional materials. The insights obtained from this thesis are not only applicable to the field of COFs, but can also serve as a blueprint for the systematic design of novel materials within the broader realm of materials science and engineering.

Part I

High-throughput screening of covalent organic frameworks for clean energy applications

1

Introduction

Our history is categorized either by the prevalent materials (think of the Bronze Age and the Iron Age) or by their revolutionary manufacturing processes (such as the Industrial Revolution). Therefore, it is reasonable to say that we are currently in the age of Materials Science, defined by the wealth of materials that are discovered by a scientific method.[1] To arrive at this point, materials science, and science in general, has evolved through four major eras, going from the experimental one, over theoretical science, through the computational era, until we are nowadays on the cusp of the data-driven era. Each of these fields can be distinguished by a totally different scientific approach.

The origin of materials science can be traced back to the era of experimental science. In this era, starting at the dawn of humankind, scientists laid the foundations by systematically probing and manipulating materials to understand their properties. Great breakthroughs were the development of metallurgy,[2] ceramics,[3] and glass.[4] Due to the empirical nature of these experiments, this era relied heavily on trial-and-error approaches and intuition, as illustrated in 1878 when Thomas Edison invented the light bulb. As filament, he required a long-lasting, low-heat, and low-power lighting element. After 14 months of heavy testing thousands of materials, going from boxwood over coconut shells to hairs cut from his assistant's beard, he discovered that a carbonized cotton thread suited his needs most. Later on, this was perfected to the tungsten filament still used in current incandescent light bulbs.[5]

In the 17th and 18th century, *i.e.*, the Age of Enlightenment, scientists started to understand how a material was composed at a fundamental level and what determined its properties. Dalton first discovered that materials were composed of atoms,[6] which were later on described by Thomson and Rutherford to consist of negatively charged electrons[7] and a positively charged nucleus.[8] This theoretical era culminated in 1926 with the dawn of quantum mechanics and the formulation of the Schrödinger equation, which elegantly describes the interactions between these subatomic particles.[9] With these discoveries, scientists were finally able to not only observe a material's properties, but also to understand and explain why it possessed these characteristics. These insights enabled them to guide their experiments to deliberately design new materials with tailored functionalities.

The advent of powerful computers midway through the 20th century enabled researchers to solve the quantum mechanical equations much faster and gain unprecedented insight into the behavior of materials at various length and time scales.[10] The rapid growth of computational resources has made it possible to screen large portions of material space in a high-throughput fashion. This has resulted in various large material database, such as the Materials Project, which tries to collect the fundamental thermodynamic and electronic properties of all known inorganic materials.[11] So far, this database has already been adopted to design new materials for solar cells,[12] batteries,[13] and other technologies.[14] By generating materials *in silico*, *i.e.*, existing solely in a digital environment, their properties could be determined even before their synthesis. Since these calculations strongly guided experimental researchers, the need for extensive trial-and-error experiments was drastically reduced and the discovery of new materials was remarkably accelerated.[15, 16] This computational era has brought us on the verge of a golden era for materials science.[5]

Besides providing valuable insight in a material's behavior, computational calculations also generate an enormous amount of data. Coupled with the advances in machine learning and data analysis of the decade,[17, 18] this marked the beginning of the data-driven era of materials science. Large datasets, such as that of the Materials Project, are mined to uncover hidden patterns and predictive relationships, guiding researchers even more to identify promising materials for specific applications.[19] Moreover, the range of properties observed in a database can be expanded even more by generating new materials with targeted properties using generative adversarial networks (GANs).[20] Furthermore, machine learning potentials (MLPs) allow for a tremendous increase in the efficiency and accuracy of the interaction models employed in molecular simulations.[21–24] As such, materials can be characterized with quantum mechanical accuracy much faster than ever before.

These four eras of science have not evolved individually, but are closely intertwined. Combining experimental efforts, theoretical insights, computational simulations, and data-driven approaches, results in a deeper understanding of materials than ever before. This holistic approach pushes the boundaries of materials design to accelerate the synthesis of new and functional materials for the development of innovative solutions across diverse industries. One of the remarkable results of this method is the design of materials with tailored porosity at the nanoscale.[25, 26] These nanoporous materials, thoroughly introduced in Section 1.1, exhibit unique properties that make them ideal for applications such as gas separation, energy storage, and catalysis. Before stating the goal of this thesis in Section 1.5, we will introduce in Section 1.2 the specific materials subclass on which this thesis focuses, *i.e.*, covalent organic frameworks, with a specific focus on some potential industrial applications. Due to their modular nature, an almost unlimited number of these materials can be imagined, as explained in Section 1.3. To navigate in this haystack, the role of computational high-throughput screenings in identifying promising materials and finding performance trends within this extended material space is explained in Section 1.4.

1.1 Nanoporous materials

The most famous example of porous materials are probably sponges, due to their presence in everyone's kitchen. The term "porous" refers to the existence of pores or voids in their structure, offering them a unique advantage in aquatic ecosystems. Natural sponges extract their nutrients from the water that flows through their pores. Specialized cells in their bodies allow the sponge to take in the water, transport it, and filter out microscopic particles. Within the environment of a regular kitchen, the pores of sponges can trap and remove particles, grease, and food residues from surfaces. Furthermore, they can take up large amounts of liquids, such as water. When applying an external pressure, for example by squeezing the sponge, the absorbed liquid can be forced out of the material again.[27]

Whereas the typical size of the pores in sponges ranges from micrometer to milimeters, nanoporous materials demonstrate even smaller pores. According to the IUPAC convention, these nanopores can be divided in micropores (pore width smaller than 2nm), mesopores (pore width between 2 nm and 50 nm), and macropores (pore width larger than 50 nm).[28–30] Therefore, they possess an exceptionally high internal surface area that provides interaction sites for other species to interact with.[31] Furthermore, nanoporous materials

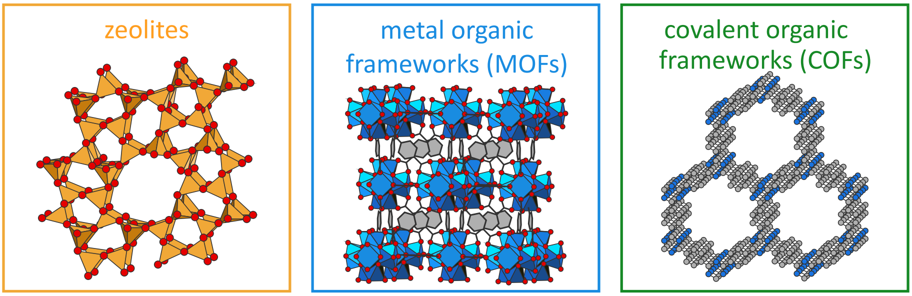


Figure 1.1: A visualization of three classes of nanoporous materials, *i.e.*, zeolites (left), metal-organic frameworks (MOFs), and covalent organic frameworks (COFs).

provide a level of control and customization that is not achievable in biological systems.[32] For example, chemically active species can be introduced at predefined sites in the material to allow for specific functionalization.[33] Moreover, nanoporous materials exhibit a broad structural diversity, leading to a range of interaction strengths between the material and the adsorbed or absorbed guests. These interactions span from non-bonding van der Waals and hydrogen bond interactions to bonding interactions, such as metal-ligand coordination bonds and covalent bonds.[34] An overview of the three classes of nanoporous materials mentioned in this thesis, is provided in Fig. 1.1.

Back in 1756, the Swedish mineralogist Axel F. Cronstedt discovered a mineral with the ability to boil water. It was the first report of a naturally occuring zeolite, a term originating from the Greek words for “to boil” (ζεω) and “stone” (λιϑοσ).[35] They are three-dimensional aluminosilicates, composed of tetrahedrally coordinated TO_4 sites, with T being a cation such as silicon or aluminum. Only in 1948, the path toward the adoption of zeolites in industry was carved with the synthesis of the first artificial zeolites,[36–38] quickly followed by their commercialization by Union Carbide.[39, 40] Due to their remarkable molecular sieve properties, zeolites have emerged in various industrial applications, such as adsorption and ion-exchange,[41, 42] and have become the workhorses of today’s petrochemical industry.[43–45] The functionality of zeolites can be expanded to a certain extent by changing the atoms occupying the T-sites or the charge-compensating anions inside the pores.[41, 42, 46] However, due to the fixed alternating pattern of tetrahedrally coordinated atoms and oxygen bridges, only a limited number of structural configurations are available for zeolites, limiting their tunability.[47]

With the invention of metal-organic frameworks (MOFs) in the 1990s, a new class of nanoporous materials with a much higher versatility than zeolites was discovered.[25, 48–53] MOFs are framework materials composed of inorganic and organic building blocks. These building blocks can be coordinated more freely compared with the tetrahedral T-sites of zeolites, resulting in many more structural configurations that are available to MOFs.[54, 55] Their structural tunability presents an enormous potential for this material class as functional materials. Therefore, they have been proposed for applications such as heterogeneous catalysis, gas storage and separation, liquid phase applications, drug delivery, sensing, and (opto)electronic devices. However, since the inorganic and organic building blocks are connected with relatively weak coordination bonds, their mechanical, thermal, and chemical stability often remains low, hampering their commercialization.[46, 56] It was only in 2005, with the synthesis of the first covalent organic framework (COF), that a new class of nanoporous materials was discovered with a tremendously higher stability.[26]

1.2 Covalent organic frameworks and selected applications

Between 1916 and 1919, Lewis, Kossel, and Langmuir laid the basis for the concept of the covalent bond between atoms.[57–59] Through this bond, electrons are shared between atoms, resulting in exceptionally strong and directed bonds. Since then, chemists have mastered the craft of synthesizing extended systems in zero dimensions, *i.e.*, isolated molecules, and one dimension, *i.e.*, linear polymers, with increased stability based on this covalent bond. However, for a long time, the synthesis of covalently bonded systems in two and three dimensions remained unfeasible.[60] It was only in the beginning of the 21st century that this challenge was addressed with the synthesis of COF-1 and COF-5, the first 2D COFs, in 2005,[26] and COF-102, COF-103, COF-105, and COF-108, the first 3D COFs, in 2007.[61]

Covalent organic frameworks (COFs) are nanoporous framework materials that are assembled by connecting organic building blocks through covalent bonds.[26, 61–66] Since they only contain organic atoms, they are the lowest mass density crystals known to date, with COF-105 and COF-108 being typical examples with a mass density of 0.18 g/cm^3 and 0.17 g/cm^3, respectively.[61] The strong covalent bond results in materials with a high mechanical, thermal, and chemical stability, which makes them of particular interest for a multitude of applications, such as gas storage and separation,[67–69] catalysis,[70–73] (opto)electronic[74–76] and sensing devices,[77] and drug delivery.[78]

In experimental COF synthesis, several precursors are mixed and react with each other. In Fig. 1.2, two regions are distinguished in such a precursor:

a linker core and reactive groups. Whereas the former remains unaffected during the reaction, the reactive groups of multiple precursors interact to form a covalently bound linkage that holds the structure together. Once the COF is formed, it can be partitioned into so-called secondary building units (SBUs). These contain a linker core and part of the linkage, which we call the SBU's linkage section throughout this thesis. A point of extension of the linkage section is defined as the geometric center of the bond that connects it with the linkage section of another SBU. These definitions are illustrated in Fig. 1.2 for COF-5. By abstracting each SBU into a zero-dimensional vertex and representing the connectivity between them as edges, the topology of a framework is determined as a periodic graph. A large collection of experimentally observed topologies is provided in the Reticular Chemistry Structure Resource (RCSR).[55] Some 3D COFs are observed to have interpenetrated topologies, in which multiple frameworks are interwoven with each other.[79, 80] Whereas a topology in fact is a periodic graph that only contains vertices and edges between them, the RCSR database provides an embedding for each of them. Such topological embedding is a realization of the topology in 3D Euclidean space, where the vertices are assigned a position in a periodically repeating unit cell.

The first COFs were synthesized based on boron chemistry, *i.e.*, they assembled forming a boronate ester, boroxine, or borosilicate linkage. Whereas this resulted in highly crystalline materials, they appeared to be unstable in water.[81, 82] Later on, new linkage classes were proposed with a large variety of functionalities beyond crystallinity. For example, imine-linked COFs, which form the largest class of synthesized COFs to date and are illustrated by COF-300 as a prototypical example, provide materials with an enhanced stability,[79, 83, 84] which is even surpassed by enamine COFs.[85] Triazine and hydrazone COFs are often proposed to be adopted in catalysis since they provide coordination centers for transition metals.[86, 87] An overview of frequently observed COF linkages is provided in Fig. 1.3.

Below, we will focus on three specific applications for which COFs are promising materials: carbon capture, vehicular methane storage, and heterogeneous catalysis. The choice for these applications is informed by their relevance in the current societal endeavor to achieve cleaner and more sustainable energy production and utilization.

1.2.1 Carbon capture

Anthropogenic emission of greenhouse gases is one of the most urgent and pressing threats to today's society,[89] resulting in a sharp increase of natural disasters caused by climate change, such as extreme temperatures, heavy

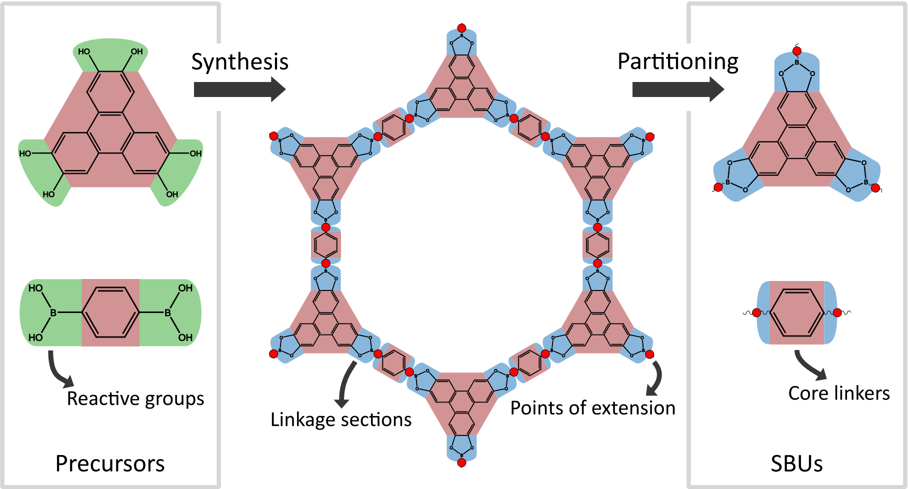


Figure 1.2: Illustration of the nomenclature of SBUs employed throughout this thesis. Whereas the linker cores (red) remain unaltered during synthesis, the reactive groups (green) of two precursors react to form a linkage. Each linkage consists of multiple linkage sections (blue) that are assigned to an SBU. A point of extension (red dot) forms the connection between two linkage sections. Figure adapted from Ref. 88 with permission of the Royal Chemical Society (©2023).

precipitation and pluvial floods, river floods, droughts, and storms.[90] The intensive industrial development and explosive population growth of the past century have disturbed the natural equilibrium between the emission and capture of greenhouse gases.[91] Most efforts to curb the effects of climate change focus on reducing anthropogenic CO_2 emissions, since CO_2 has the most significant long-term impact on the Earth's radiative balance due to its abundance and long lifespan.[89] In the long-term, the only sustainable solution to climate change is the evolution towards a CO_2-neutral economy, which, however, requires the development of new decarbonized industrial processes.[91] In anticipation of this goal, it is essential to uncover materials that can efficiently and inexpensively capture CO_2 from the flue gases of stationary point sources, such as power plants and industrial production sites, which are responsible for about half of the CO_2 emissions.[91] As visualized in Fig. 1.4, the separated CO_2 molecules can either be safely stored in empty oil reservoirs (carbon capture and sequestration, CCS),[92–95] or they can be reused in industrial processes, thereby closing the carbon loop (carbon capture and utilization, CCU).[96, 97]

Current carbon capture technologies used in industry rely on chemisorption in a process termed amine scrubbing.[98] However, the high regeneration

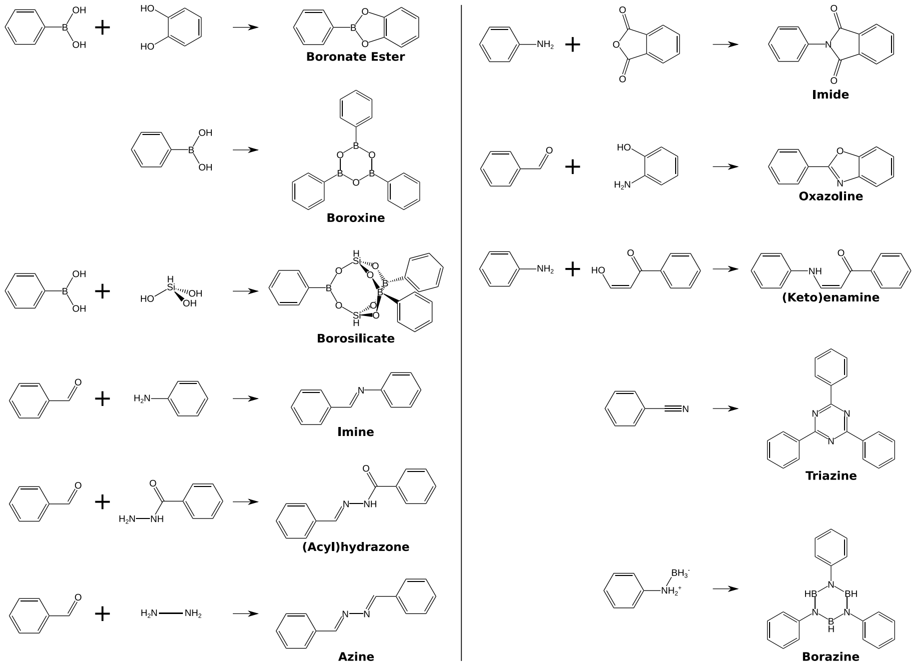


Figure 1.3: Model reactions for the formation of some frequently occuring COF linkages. Figure adapted from Ref. 88 with permission of the Royal Chemical Society (©2023).

energy, fast equipment corrosion, and solvent decomposition cause this technology to be highly expensive and prevent its widespread adoption.[99] As illustrated in Fig. 1.4, a more viable approach is to convert the flue gas through fixed beds of nanoporous materials, which pores are especially designed to selectively capture CO_2 from the flue gas. This process requires far less regeneration energy, and thus financial cost, since the molecules are only physisorbed to the framework rather than chemisorbed as in amine scrubbing.[100] Activated carbon, zeolites, MOFs, and COFs are all proposed as adsorbent materials for carbon capture.[68, 101–105] However, the adoption of activated carbon, MOFs, and zeolites in industrial environments has been hampered by several challenges, such as reduced CO_2 selectivity or strong performance deterioration in the presence of water, often present in flue gases.[68, 103] The high mechanical, thermal, and chemical stability of COFs, in combination with the large internal surface area on which CO_2 molecules can be adsorbed and the enormous structural versatility, makes these materials of particular interest for industrial carbon capture applications.[68, 106–108] In **Paper V**, an extended high-throughput screening is performed to establish performance trends and to identify promising COFs for carbon capture.

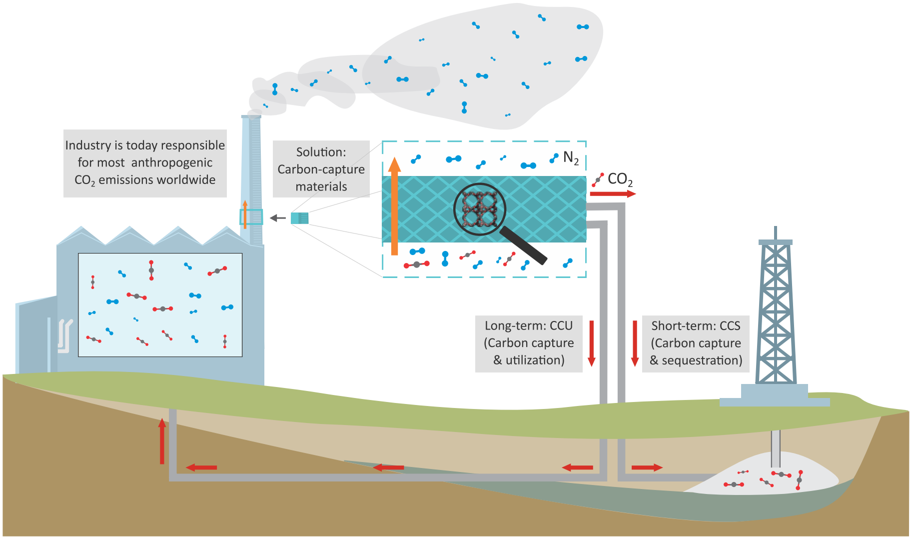


Figure 1.4: The carbon pathway, illustrating the proposed mechanism to separate CO_2 from industrial flue gases using COFs. Once captured, the recovered CO_2 is either sequestered (CCS) or reutilized (CCU).

1.2.2 Vehicular methane storage

Another application for which COFs are highly promising due to their low mass density and high internal surface area is vehicular storage of natural gas, which mainly consists of methane.[109, 110] This is a green alternative to traditional petroleum and gasoline based fuels.[111, 112] However, the low energy density of natural gas prevents its adoption in practical applications. Several approaches, such as compressed natural gas (CNG),[113] liquid natural gas (LNG),[114] and adsorbed natural gas (ANG),[115, 116] are proposed to densify the natural gas. Nanoporous materials are used in the latter approach to store the gas at a higher storage pressure. The methane is then gradually released until a lower depletion pressure is reached, at which point the tank should be refilled. The deliverable capacity of a nanoporous material is defined as the amount of gas that can be released every cycle. The Methane Opportunities for Vehicular Energy (MOVE) program of the Advanced Research Projects Agency-Energy (ARPA-E) of the US Department of Energy has set targets on this deliverable capacity for nanoporous materials to compete with a 250-bar CNG tank.[117] Promising materials should exhibit a gravimetric deliverable capacity higher than 0.5 g/g, whereas the volumetric deliverable capacity should exceed 315 vSTP/v.[117] The unit vSTP/v determines how much volume the delivered methane molecules in one volume unit COF would occupy if they would be placed under standard temperature and pressure (STP), *i.e.*

273.15 K and 1 atm, in the absence of the COF. The performance limits of COFs for this application are investigated in **Paper III**.

1.2.3 Heterogeneous catalysis

At the hearth of the chemical industry are chemical reactions, in which reactants transform to produce products. However, often the rate at which these reactions occur is too slow. In catalytic processes, a substance, *i.e.*, the catalyst, increases the reaction rate without being consumed by the overall reaction. The catalyst facilitates the reaction by providing an alternative pathway with lower activation energy. In homogeneous catalysis, the catalyst occurs in the same phase as the reactants and products, which is typically the gas phase. However, this type of catalysis suffers from difficult product separation, low recyclability, and high costs.[118, 119] In heterogeneous catalysis, the catalyst is in another phase as the reactants and products. Typically, this is a solid material, such as graphitic carbon nitride,[120] MOFs,[121] and COFs.[122] A particular type of catalysis is photocatalysis, in which the chemical reaction is activated by absorption of light.[73, 123–125] In **Paper IV**, our experimental partners synthesized new COFs that can be adopted in a photocatalytically enhanced Povarov cyclization reaction, in which glycine esters are oxidized into quinolines and α-dicarbonyl compounds.

1.3 The haystack of reticular chemistry

Similar to MOFs, COFs are composed of SBUs, assembled in a topological net. This principle, coined reticular chemistry,[126] provides a design recipe for the targeted synthesis of crystalline nanoporous materials. Firstly, a topological embedding should be chosen as blueprint for the resulting material. Secondly, for every family of vertices in the topology, *i.e.*, a so-called Wyckoff set, an SBU should be selected that matches their local symmetry. Lastly, the modular building blocks should be reticulated into the actual COF structure.[65] Whereas there is a large diversity in symmetry of MOF building blocks, this remains limited for COFs, due to the inherent organic nature of their SBUs. Therefore, most building blocks employed in COF synthesis to date are trigonal, square, or tetrahedral. These are especially suited to assemble in high-symmetry topologies, such as **hcb**,[26] **sql**,[127] **kgm**,[128] **dia**,[79] **bor**,[61] **ctn**,[61] and **pts**.[129] However, extensive efforts are being made to increase the diversity of SBUs, already resulting in COFs containing hexagonal,[130] cubic,[131] or even one-dimensional building blocks.[131] Therefore, with the further development of the field of COF synthesis, the number of available topologies and their complexity can increase significantly.[132]

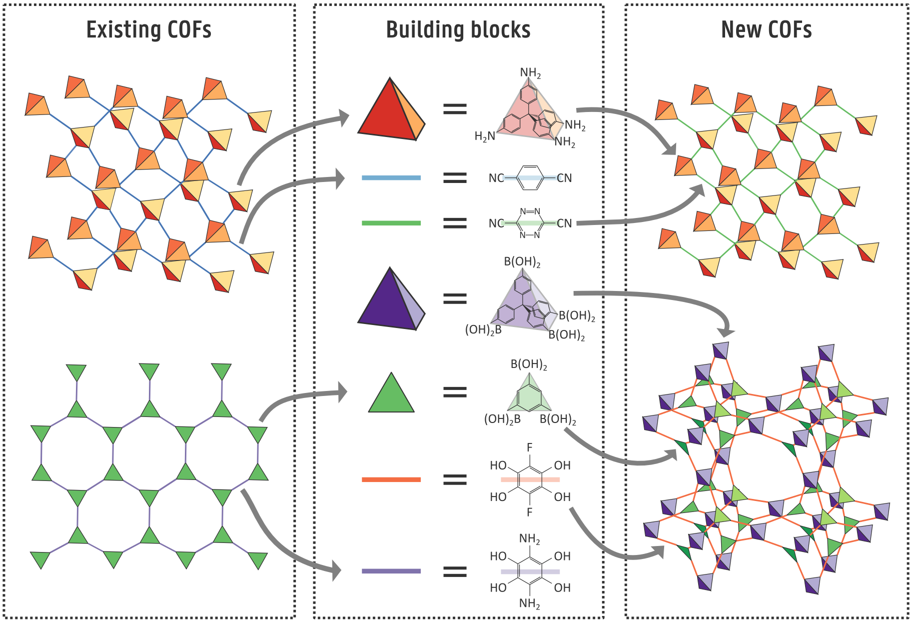


Figure 1.5: According to the principle of reticular chemistry, the structure of COFs can be easily decomposed into SBUs. Combining these building blocks in alternative ways allows one to construct an enormous number of possible not-yet-synthesized structures, which is unfeasible to screen experimentally.

The true power of reticular chemistry lies in the large freedom of available SBUs for a given topology, as exemplified in Fig. 1.5. Therefore, the physicochemical properties of the framework can be largely tuned by selecting the appropriate composition of building blocks, thereby making them the ideal engineering materials. For example, already in 2008 Tilford *et al.* managed to tailor the pore size of a COF family between 11 Å and 18 Å by functionalizing the linker with alkyl groups (see Fig. 1.6).[133] In another example, the conductivity of a porphyrin-based COF could be switched from hole-dominated to electron-dominated or ambipolar by introducing a different metal in the porphyrin ring.[134] Furthermore, as again illustrated in Fig. 1.5, the same SBU can be observed in multiple topologies by combining them with different other building blocks, or altering the synthesis route. As such, reticular chemistry lies at the origin of the enormous structural and functional versatility of these material classes.

However, since a huge number of SBUs can be adopted, and since the number of possible topologies that can still be discovered for COFs is very large, the number of hypothetical COFs is, analogous to MOFs, almost unlimited. Therefore, several attempts have been undertaken to collect all possible

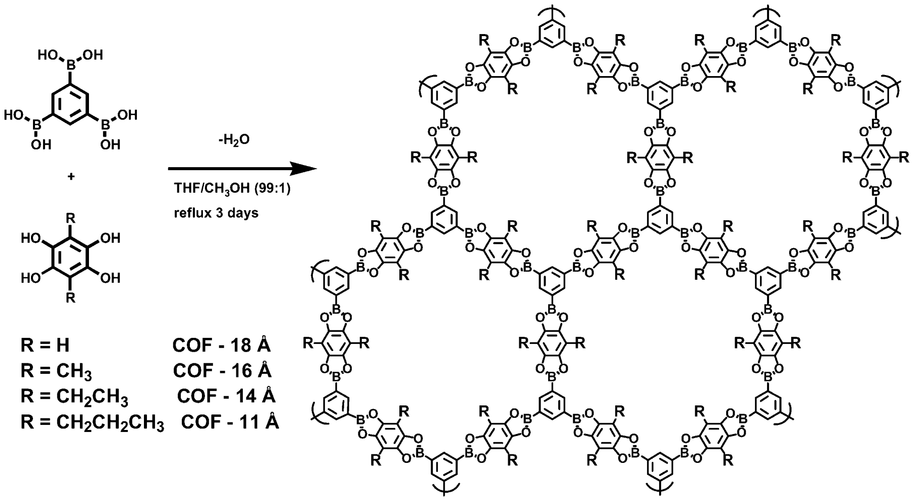


Figure 1.6: In reticular chemistry, new materials can be designed by changing the SBUs involved in the synthesis process. By carefully changing the alkyl functionality of the linkers, Tilford *et al.* managed to tune the pore diameter of these COFs with great precision. Figure reproduced from Ref. 133 with permission of John Wiley and Sons (©2008).

MOFs and COFs in large material databases. Depending on the origin of the involved structures, these databases can be divided into experimental and hypothetical databases. The experimental ones exist of materials that are already synthesized, containing either the structure as synthesized, or after guest removal and/or computational structure optimization. In one of the first efforts to generate a large experimental database, Chung *et al.* identified a subset of 5109 MOF structures from the Cambridge Structural Database (CSD)[135] and collected them in the Computation-Ready Experimental (CoRE) MOF database[136]. Later, they expanded this database to include a total of 14 142 structures.[137] An automated screening algorithm was implemented by Moghadam *et al.* in the CSD Python API to directly identify a MOF structure when it is added to the CSD. Their subset contained a total of 69 666 materials at the time of its first publication.[138] Recently, 15 713 MOFs for which efficient density functional theory (DFT) calculations are feasible were extracted from the previously mentioned databases and collected in the QMOF database.[139] Currently, there are two experimental COF databases available. Encouraged by the success of the CoRE MOF database, Tong *et al.* developed a CoRE COF database, containing experimental COF structures that are ready for computational simulations.[140] Later on, Ongari *et al.* published the Clean, Uniform, and Refined with Automatic Tracking from Experimental Database (CURATED), where the focus is on a uniform, consistent and reproducible protocol to make the structures clean and error-free.[141]

Hypothetical databases contain possibly not-yet-synthesized materials that are generated *in silico*. Opposed to experimental databases, which naturally focus on cheap and easy-to-synthesize materials, they provide a broad and diverse subset of material space. Moreover, Lan *et al.* demonstrated that promising hypothetical materials possibly can be synthesized by experimentally creating four new materials selected from an *in silico* generated COF database, among which two materials in a not-yet-observed topology.[142] Whereas energy minimization approaches exist, such as the Automated Assembly of Secondary Building Units (AASBU),[143] that design reasonable structures, most hypothetical databases are constructed with geometric procedures. Such a geometric procedure forms a periodic material by connecting SBUs with one another, after which a geometry optimization relaxes the structure. Depending on the SBU assembly process, two classes of geometric procedures are discerned, *i.e.*, bottom-up and top-down approaches. In the bottom-up approach, building blocks are automatically assembled without prior knowledge of the underlying topology. In their pioneering work, Wilmer *et al.* constructed a database of 137 953 MOFs adopting a bottom-up approach, starting from a set of 102 SBUs.[144] Later on, Fernandez *et al.* generated a database of 324 500 MOFs from 66 SBUs and 19 functional groups.[145] However, it turned out that the topological diversity in these databases is limited. For example, the database of Wilmer *et al.* recognized only six different topologies.[146] To increase the number of topologies utilized in a database, several software packages, such as Zeo++,[147] AuToGraFS,[148] Weaver,[149] TOBASCCO,[150] and ToBaCCo[151, 152] have proposed a top-down approach to generate materials *in silico*. In such a top-down method, a topological embedding is selected *a priori*, after which suitable SBUs are placed on its nodes. For example, the ToBaCCo software package has been adopted to generate a database of 13 512 MOFs, starting from 78 SBUs and 41 topologies.[151] For COFs, two hypothetical databases have been constructed by Martin *et al.*[153] and Mercado *et al.*[154] The first database contains a total of 4147 interpenetrated 3D COFs derived from 620 non-interpenetrated geometries. They used commercially available precursors and limited themselves to the experimentally observed **dia**, **bor**, and **ctn** topologies to derive experimentally feasible materials.[153] Mercado *et al.* generated 69 840 COFs from a total of 666 SBUs and 839 2D and 3D topologies.[154]

As illustrated by this literature overview, the number of (hypothetical) MOFs and COFs is overwhelming. Therefore, finding the most optimal material for a given application from this vast material space is like searching for a needle in a haystack. A full experimental characterization of all materials is not only cumbersome, but also very expensive, both in terms of time and financial cost. Because simulations are significantly faster and cheaper than experiments, computational high-throughput screenings have attracted much attention in the past decade to quickly find performance trends and identify promising materials in the haystack of reticular material classes.

1.4 Computational high-throughput screenings

The field of computational modeling aims to develop reliable methods to accurately characterize a given material. However, the most accurate methods typically require a large amount of computational resources, even for a single material. Applying these algorithms to many materials in a high-throughput fashion would be unfeasible. Therefore, as illustrated in Fig. 1.7, high-throughput screenings frequently adopt a multi-step procedure to systematically filter the most promising materials. These are subsequently exposed to more expensive computations to characterize them more accurately.[1] In general, high-throughput screenings on large material sets have three major advantages.[155] Firstly, promising materials can be identified quickly. By adopting a database containing experimental materials, the synthetic feasibility can be ensured. However, starting from experimental materials is no prerequisite, as illustrated by Boyd *et al.*, who synthesized Al-PMOF and Al-PyrMOF after they demonstrated an exceptional performance for CO_2 capture during a computational high-throughput screening.[156] Secondly, useful property-property relationships can be established. When only a handful data points are available, it might be difficult to distinguish whether a relationship exists. However, as demonstrated in Fig. 1.8, clear trends emerge when increasing the sample size. Thirdly, the performance limits of a material class for a specific application can be predicted. Therefore, the selected dataset should represent the targeted material space correctly. Databases of hypothetical materials are ideally suited for this purpose since their *in silico* assembly algorithm provides a large degree of freedom in the material's geometry.

Within the field of nanoporous materials, such high-throughput screenings have primarily concentrated on gas storage and separation properties.[159] In their pioneering study, Wilmer *et al.* explored the uptake of CH_4, the main component of natural gas, in MOFs,[144] after which several others followed and investigated both MOFs[136, 151, 152, 160] and COFs.[88, 153, 154, 160, 161] Similarly, H_2 adsorption has also attracted a lot of attention due to its high potential as an energy carrier to replace fossil fuels.[151, 162–166] Several examples of gas separation processes have been investigated, including the purification

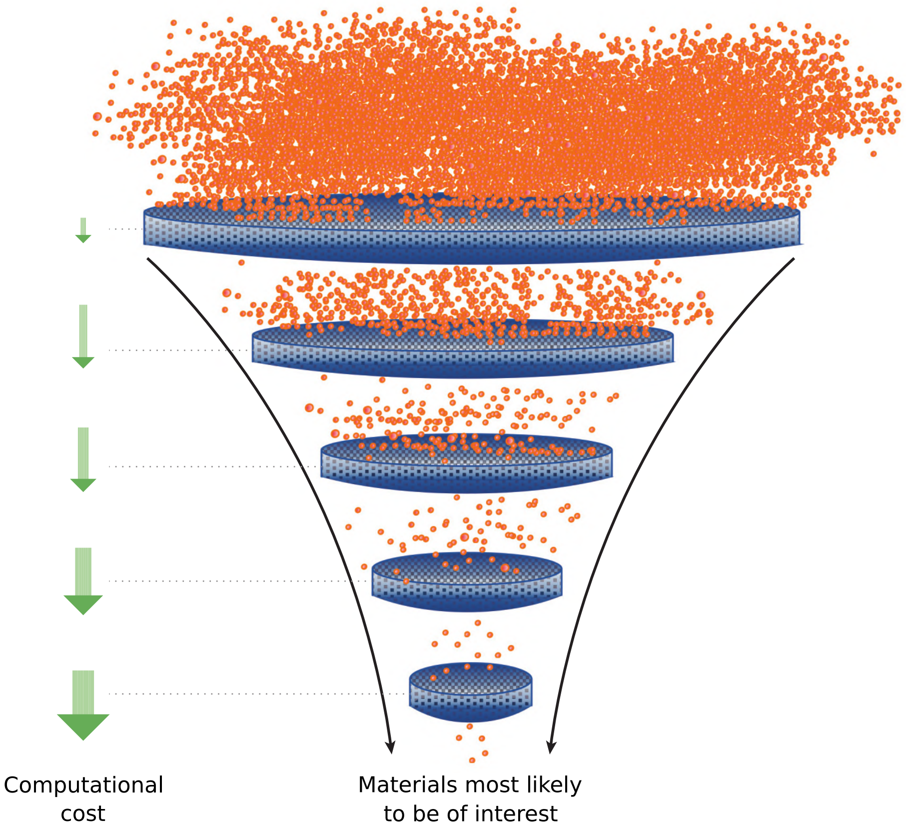


Figure 1.7: Computational high-throughput screenings frequently adopt a multi-step procedure to find promising materials. In each step, a new descriptor is calculated that requires an increasing amount of computational resources. Based on the calculated descriptor, the sampled material space is narrowed down to a more interesting region. Figure adapted from Ref. 1 with permission of Annual Reviews (©2015).

of natural gas from undesired trace gases, such as CO_2,[158, 167–171] N_2,[167, 172] and H_2,[170, 172–174] and pre- and post-combustion carbon capture. The latter focuses on CO_2/H_2[170, 175–177] and CO_2/N_2[141, 156, 158, 168, 169, 178–186] separations, respectively. Recently, more complex properties have been calculated in high-throughput screenings, such as the thermal conductivity,[187] mechanical stability,[152, 188] electronic band gap,[139, 189, 190] and catalytic activity.[191]

Traditionally, high-throughput screenings relied on brute-force screenings within an existing database. However, the growing number of available materials has made this approach unfeasible, even with current computational resources.[142, 192] To accelerate high-throughput screenings, researchers have

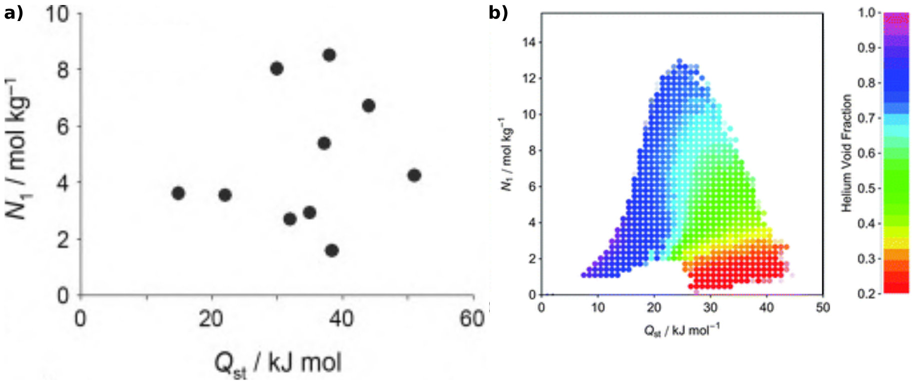


Figure 1.8: High-throughput screenings are able to establish performance trends that can not be distinguished with limited data. Both figures visualize the relation between CO_2 uptake at 2.5 bar and the heat of adsorption. With only 10 data points, no clear trend can be distinguished in a). However, when the results of more than 130 000 MOFs are combined in b), a well-defined structure-property relation emerges. Figure adapted from Refs. 157 and 158 with permission of John Wiley and Sons (©2011) and the Royal Society of Chemistry (©2012).

started turning to machine learning (ML) algorithms.[19, 193] Such workflow typically starts with featurizing each material in the database. Most commonly, textural features, such as the pore diameter and internal surface, are adopted. Chemical descriptors, describing the chemical environment of a framework, drastically increase the accuracy of a ML algorithm for most applications.[194, 195] Examples of such descriptors are compositional descriptors derived from the chemical composition,[190, 194, 196] revised autocorrelation functions (RACs),[197, 198] and atomic-property weighted radial distribution functions (AP-RDFs).[145, 199, 200] Once all materials are featurized, a ML algorithm learns trends from a subset of the database and leverages this knowledge to predict the characteristics of previously unseen materials more rapidly than traditional calculations. ML has been employed to predict the uptake and separation of various gases and their combinations,[201] such as the storage of H_2,[202–205] CO_2,[145, 196] N_2,[199] and CH_4,[192, 194, 195, 206, 207] and the separation of CO_2/N_2,[196, 208, 209] CO_2/H_2,[196, 200, 210] CO_2/CH_4,[209, 211] CH_4/H_2,[212] and noble gases.[213, 214] Furthermore, other applications of ML can be found in the prediction of ideal synthesis conditions,[215] electronic properties,[139, 189, 190] and mechanical stability.[188]

1.5 Goal and outline

As outlined in Section 1.4, a lot of high-throughput screenings have already been performed on MOFs, and they start to find their way in the field of COF research. However, until now, the performance limits of COFs for several applications are not yet clear. Furthermore, screening studies crucially depend on the existence of diverse databases and the accurate description of the interatomic interactions, which is currently lacking in the field of COFs. Therefore, the goal of this thesis is to characterize the COF material class for clean energy applications, such as vehicular methane storage and carbon capture, by means of a computational high-throughput screening starting from an extended and diverse, *in silico* generated database with accurate interatomic interactions. As depicted in Fig. 1.9, this will be accomplished via several intermediate goals, each discussed in the following chapters.

In **Chapter 2**, the necessary tools to accurately model crystalline materials and predict their macroscopic properties will be introduced. Starting from the Schrödinger equation, the potential energy surface (PES) will be defined. A force field description of the PES provides a good balance between efficiency and accuracy, which is required for reliable high-throughput screenings. We demonstrate that a force field for a periodic material can be efficiently derived from the cluster force fields of its underlying building blocks. Several sampling techniques are introduced to explore relevant regions in a material's phase space with molecular simulations.

Before being able to start such simulations, the geometry of a material, *i.e.*, the positions of all atoms and a unit cell to describe the periodic boundary conditions, needs to be defined. However, since our goal is to generate a large database containing a huge amount of materials, a manual procedure is not feasible. In **Chapter 3**, an automated procedure is introduced and implemented to *in silico* generate nanoporous materials for which the reticular chemistry principle applies. This protocol follows a top-down approach, starting from the SBUs and the topology of a material, using both geometric and energetic criteria. The algorithm is illustrated by three case studies.

Adopting this procedure to various combinations of SBUs and topologies allows us to automatically generate an extensive number of COF geometries. In this way, the ReDD-COFFEE database is generated in **Chapter 4**, containing a total of 268 687 COFs with an *ab initio* derived force field for each of them. The diversity of the database is analyzed and compared with already existing COF databases. Furthermore, we demonstrate that the system-specific force fields provided in the database indeed reproduce experimental properties more accurately than the generic force fields typically used in high-throughput screenings.

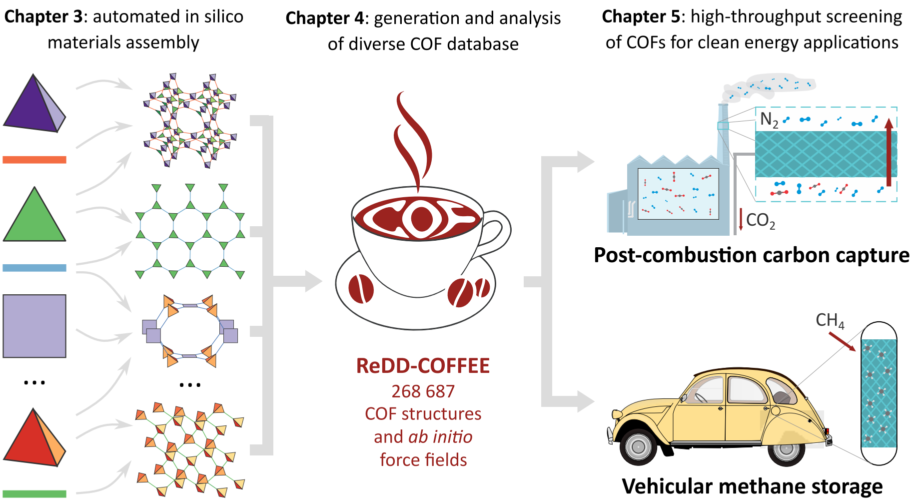


Figure 1.9: Schematic overview of the workflow to perform high-throughput screenings followed in this thesis. In Chapter 3, an automated protocol is introduced to derive hypothetical nanoporous materials *in silico*. This procedure is adopted in Chapter 4 to generate a diverse COF database, which is subsequently screened for clean energy applications in Chapter 5.

Finally, this thesis culminates in **Chapter 5**, where several high-throughput screenings on the ReDD-COFFEE database are performed. First, we start with a fast screening of the textural properties of COFs and compare them with other classes of nanoporous materials, *i.e.*, MOFs and zeolites. Secondly, the performance limits of COFs for vehicular methane storage are investigated. Third, an extended high-throughput screening of COFs for carbon capture is performed, using machine learning algorithms to greatly accelerate the screening. We establish performance trends, formulate design rules, and identify promising candidates for experimental synthesis.

In **Chapter 6**, a concluding overview of the key findings of this thesis is provided, together with a personal perspective on possible avenues for high-throughput screenings in nanoporous materials.

2

Molecular modeling tools

A material's macroscopic properties are determined by its behavior on the microscopic level. As illustrated in Fig. 2.1, the field of molecular modeling provides insight in this behavior with four typical steps. After the atomic structure of a material is defined, the interatomic interactions are determined by accurate, yet efficient descriptions. The interactions between the atoms are detailed by quantum mechanics, which is described in the Schrödinger equation. However, even with several approximations, determining the electronic structure of a material requires an enormous amount of computational power. Therefore, several approximate methods to describe the interatomic interactions are introduced, as explained in Section 2.1. The next step in a molecular modeling exercise is the sampling of the multidimensional phase space, which is defined by the positions and momenta of the atoms in the material. The goal of these sampling algorithms, which are rooted in statistical physics and thermodynamics, is to provide and characterize a set of representative material configurations, from which the macroscopic properties can be derived in the last step of Fig. 2.1. In Section 2.2, the two main categories to divide phase space sampling techniques, *i.e.*, static and dynamic exploration methods, and their application in this thesis are discussed.

2.1 From quantum mechanics to force fields

Section 2.1.1 introduces the concepts of quantum mechanics to determine the interatomic interactions. Via the Born-Oppenheimer approximation, the

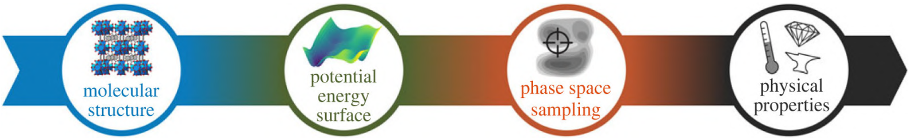


Figure 2.1: An overview of the four typical steps in a molecular modeling exercise. After the structure of a material is defined on the molecular level, the interatomic interactions have to be determined to provide an accurate description of the potential energy surface (PES, see Section 2.1). Once the relevant regions in phase space are samples, the physical properties of a material can be derived. Figure reproduced from Ref. 24 with permission of the Royal Society (©2023).

potential energy surface (PES) of a material is introduced, which describes the potential energy associated with each configuration of nuclear coordinates. Fig. 2.2 illustrates that there is a wide variety of methods to determine the PES. Whereas quantum mechanical methods are the most accurate ones, they can only describe small systems for a limited time scale. Since knowledge of the electronic structure is not strictly required to calculate the structural and adsorption properties focused on in this thesis, a more efficient description provided by force fields is adopted throughout this work, as introduced in Section 2.1.2.

2.1.1 The Schrödinger equation and Born-Oppenheimer approximation

Quantum mechanics postulates that all physicochemical properties of a molecular system can be derived from its total wavefunction $|\Psi(t)\rangle$, which characterizes the probabilistic nature of all elements and the wave-particle duality. In 1926, Edwin Schödinger formulated his famous Schrödinger equation, describing the dynamics of this wavefunction through time:[9]

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

where $\hbar$ is the reduced Planck constant. The operator $\hat{\mathcal{H}}$ is the molecular Hamiltonian representing the interactions between the different particles:

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

with $\hat{\mathcal{T}}$ and $\hat{\mathcal{V}}$ being the kinetic and potential energy operators, respectively, associated with the nuclei (N) and the electrons (e). In the context of

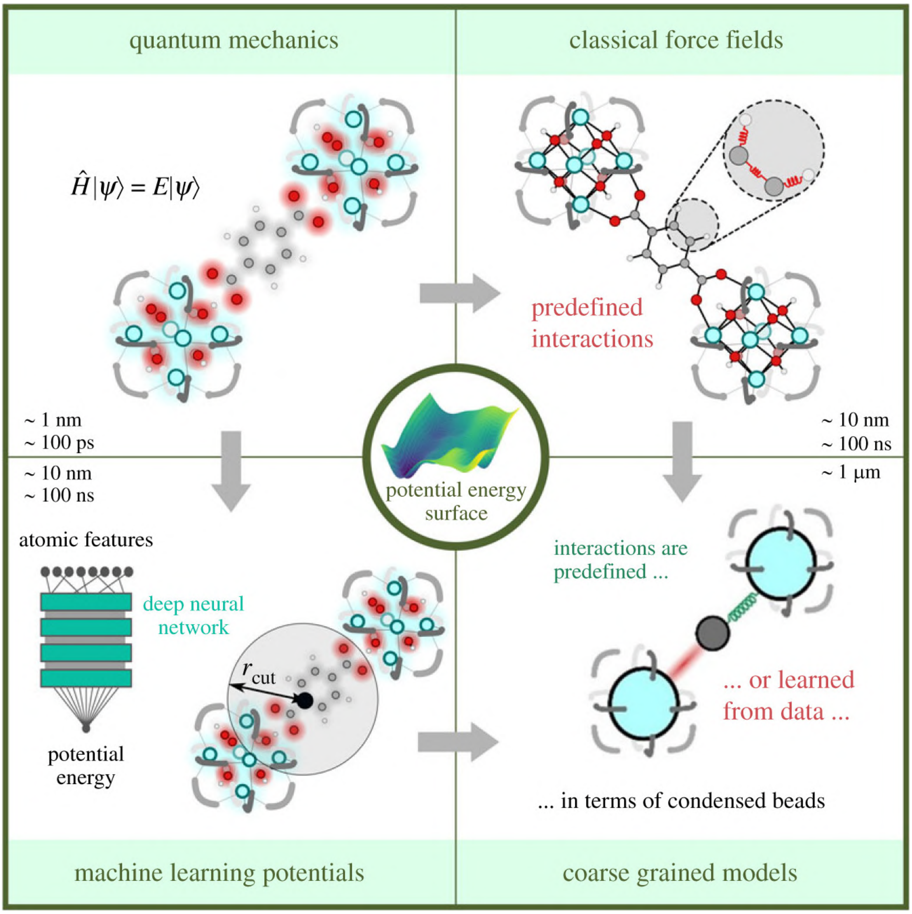


Figure 2.2: An overview of several methods to determine the potential energy surface (PES). The most accurate methods are those based on quantum mechanics. However, these are very expensive, limiting the length and time scales for which they can be employed. Classical force fields and machine learning potentials[21] allow to perform simulations on a larger length and timescale. Coarse-grained models can describe even larger systems by combining groups of atoms in beads.[216] Figure adapted from Ref. 23 with permission of Elsevier (©2021).

molecular systems, only the Coulomb potential energy provides a significant contribution to $\hat{\mathcal{V}}$ since the forces are dominated by electromagnetic interactions on the relevant length scales.

In the absence of external time-dependent fields, the Hamiltonian of Eq. 2.1 contains no time dependency and a separation of variables can be applied to the Schrödinger equation. This results in the time-independent Schrödinger equation of Eq. 2.3, which formulates the task of finding the wavefunction and its associated energy as a stationary eigenvalue problem:

$$\hat{\mathcal{H}}\left|\psi_n\right\rangle = E_n\left|\psi_n\right\rangle , \tag{2.3}$$

In this expression, $\left|\psi_n\right\rangle$ are the stationary eigenstates with corresponding eigenenergies E_n. The total wavefunction $\left|\Psi\right\rangle$ can now be expressed as a linear combination of the stationary eigenstates $\left|\psi_n\right\rangle$.

However, the outlined formalism is not of practical use for most systems since it becomes increasingly difficult to determine the wavefunction of molecular systems with a large number of electrons. Therefore, the Born-Oppenheimer approximation is frequently adopted to solve the Schrödinger equation for realistic materials.[217] This approximation consists of two parts. Firstly, the clamped-nuclei approximation observes that the wavefunction $\left|\psi\right\rangle$ can be separated in a nuclear part, $\left|\psi^N\right\rangle$, and an electronic part, $\left|\psi^e\right\rangle$, since electrons are more than three orders of magnitude lighter than nuclei and therefore equilibrate on much shorter timescales. Neglecting the nuclear motion results in the following electronic eigenvalue problem:

$$\left(\hat{\mathcal{T}}_e + \hat{\mathcal{V}}_{NN} + \hat{\mathcal{V}}_{Ne} + \hat{\mathcal{V}}_{ee}\right)\left|\psi_n^e\right\rangle = E_n^e(\boldsymbol{R})\left|\psi_n^e\right\rangle \tag{2.4}$$

with $\left|\psi_n^e\right\rangle$ and $E_n^e(\boldsymbol{R})$ being the electronic wavefunction and corresponding eigenenergy, respectively, given a specific fixed configuration of the atomic nuclei, $\boldsymbol{R}$. Repeatedly solving the electronic Schrödinger equation for different atomic configurations, as illustrated in Fig. 2.3, results in high-dimensional surfaces as a function of the nuclear coordinates, $\boldsymbol{R}$, also known as Born-Oppenheimer surfaces. Secondly, when the electronic energy levels are well separated, the adiabatic approximation can be used, which assumes that the system remains in the same electronic eigenstate upon a slight change of the nuclear configuration. Within this approximation, the nuclear wavefunction $\left|\psi^N\right\rangle$ can be determined by solving the remaining nuclear Schrödinger equation:

$$\left(\hat{\mathcal{T}}_N + E_n^e(\boldsymbol{R})\right)\left|\psi_{nm}^N\right\rangle = E_{nm}\left|\psi_{nm}^N\right\rangle \tag{2.5}$$

This yields the energy levels associated with the vibrational and rotational motion of the nuclei, indicated in Fig. 2.3 in blue and green, respectively. Since these are mostly separated by a much smaller energy difference than the electronic energy level, the adiabatic approximation is valid for the majority of molecular systems. Furthermore, when the electronic energy levels are well separated, only the Born-Oppenheimer surface of the electronic ground

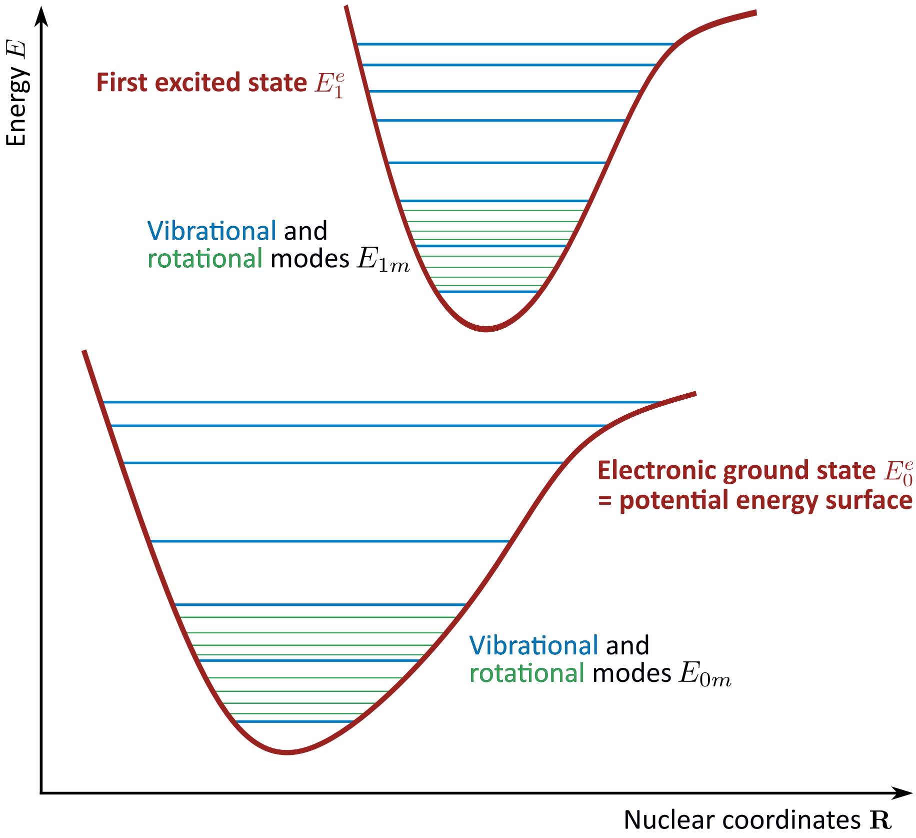


Figure 2.3: Illustration of the two lowest Born-Oppenheimer surfaces (red), *i.e.*, the electronic ground state ($n = 0$) and the first excited state ($n = 1$). The vibrational (blue) and rotational (green) energy levels determined by the nuclear Schrödinger equation (Eq. 2.5) are also visualized.

state ($n = 0$) is of interest, which will be referred to as the potential energy surface (PES). It describes the variations of the ground state energy as a function of the nuclear coordinates, such that the different extrema of this surface represent the relevant atomic configurations, *i.e.*, the (meta)stable configurations and the transition states separating them. Due to the large mass of the nuclei, these particles can be treated as classical particles, eliminating the need to solve the nuclear Schrödinger equation of Eq. 2.5. Therefore, calculating the PES from Eq. 2.4 is a key concept in molecular modeling for which a broad range of methods exist.

The most accurate methods to solve the electronic Schrödinger equation originate from quantum mechanics and can be roughly categorized into two classes. Wavefunction-based methods, such as (post-)Hartree-Fock

(HF),[218–222] solve the electronic Schrödinger equation using an approximated wavefunction. Since the wavefunction is $4n$-dimensional, with n being the number of electrons in the molecular system, this is a computationally very expensive task, however. Therefore, electron-density based methods, such as density functional theory (DFT),[10] are centered around the calculation of the electron density. Whereas the electron density, $\rho(\boldsymbol{r})$, is only three-dimensional, and thus requires less computational resources than wavefunction-based methods, it nevertheless provides sufficient information to derive all relevant system properties of the electronic groundstate.[223] Whereas HF is an approximate method with an exact solution, DFT can be considered as an exact method with an approximate solution due to the unknown exchange-correlation functional, *i.e.*, a contribution to the total energy. These exchange-correlation functionals are categorized into different classes based on their underlying theoretical principles, such as local density approximation (LDA), generalized gradient approximation (GGA), meta-GGA, and hybrid functionals.[224] Each class offers a trade-off between computational efficiency and accuracy, with hybrid functionals, such as B3LYP,[225–227] being widely used for their balanced performance in many applications. Since the dynamic correlation of instantaneously interacting dipoles is not correctly captured in these semilocal functionals, long range dispersion is not present in DFT calculations. Therefore, dispersion corrections are added to the energy, with the most commonly adopted being the D3 scheme proposed by Grimme.[228]

While DFT increases the attainable system size and time scale in molecular simulations compared to HF, they remain limited to a few nm and hundreds of ps. Furthermore, these calculations require an enormous amount of computational resources, even when considering a single material. Given our aim to characterize thousands of COFs, for which the minimal system size easily exceeds hundreds of atoms, current high-performance computing (HPC) infrastructures are not sufficient to reach this goal. To reduce the computational cost even more, the explicit electronic degrees of freedom can be neglected. In this approach, the PES is *a priori* postulated as an analytical function, thereby introducing the concept of force fields as explained in Section 2.1.2.

2.1.2 Force fields

Whereas knowledge of the electronic structure of a molecular system is indispensable in some applications, such as the determination of the electronic band gap of a material or the excited states of a molecule, it is not always a prerequisite. Especially for the structural and adsorption properties targeted in this thesis, exact knowledge of the electronic behavior of the system is not required and the PES can be approximated by an analytical expression as a function of the nuclear coordinates, *i.e.*, a force field. Since the energy and the forces of a given configuration can be determined efficiently by evaluating a simple mathematical expression, rather than calculating the wavefunction or the electron density explicitly as in quantum mechanical methods, an enormous speed-up is obtained. Therefore, force fields are typically adopted in simulations of large systems and to study their behavior over long time periods. Also in high-throughput screenings, where a large number of materials are characterized, force fields are the favored method to describe the PES due to the limited cost to characterize a single material. In Section 2.1.2.I, the analytical expression of the force fields used in this thesis is introduced. This expression contains a number of unknown force field parameters that should be fitted according to the parametrization procedures outlined in Section 2.1.2.II. To limit the computational cost of directly deriving periodic force fields, we have employed a cluster approach throughout this thesis, which is introduced in Section 2.1.2.III.

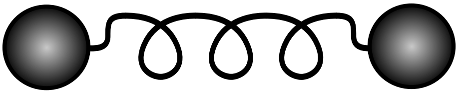
I. Energy expression

As illustrated in Eq. 2.6, a force field typically contains a number of physically inspired energy contributions, which add up to the total energy expression. We distinguish between covalent and non-covalent contributions. Whereas the non-covalent contributions involve electrostatic and van der Waals interactions that have a large spatial extent, the covalent contributions describe interactions between atoms connected through chemical bonds. Within a force field, these shorter-range bonds, which define the connectivity of the molecular system, are fixed. The covalent part is composed of several contributions corresponding to different kinds of atom patterns, such as bonds, bends, and dihedral angles. The energy of each of these contributions is represented by an analytical expression as a function of the internal coordinates of the involved atom patterns and a set of *a priori* unknown parameters such as the equilibrium bond length and its force constant. Many flavors of force fields exist, differing in the proposed analytical expressions and the procedure to determine the force field parameters. Below, the various energy terms are described, focusing on the expressions adopted in the QuickFF,[229, 230] UFF,[231] and MM3[232] force fields, frequently used throughout this thesis.

$$\begin{aligned}
E_{\text{FF}} &= E_{\text{cov}} + E_{\text{non-cov}} \\
E_{\text{covalent}} &= E_{\text{bond}} + E_{\text{bend}} + E_{\text{out-of-plane}} + E_{\text{torsion}} + E_{\text{cross}} \qquad (2.6) \\
E_{\text{non-cov}} &= E_{\text{ei}} + E_{\text{vdW}}
\end{aligned}$$

Bond distance The interaction between two chemically bonded atoms is most often described by a harmonic potential, which can be interpreted as if the bonds are replaced by mechanical springs. The internal coordinates of interest are the bond distances r_i and the unknown force fields parameters are the rest bond lengths r_i^0 and the force constants $K_{r,i}$.

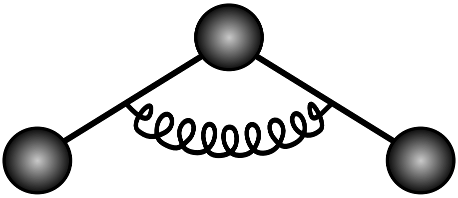
$$E_{\text{bond}}^{\text{QuickFF/UFF}} = \sum_i \frac{K_{r,i}}{2}(r_i - r_i^0)^2 \tag{2.7}$$



Bending angle Similar to the bond terms, the QuickFF force fields used in this work describe the bend terms as a harmonic potential in terms of the angles θ_i. However, in general, UFF uses a harmonic potential in terms of the cosine of the angle. When the neighbors of the central atom in the bend form a linear ($n = 2$), trigonal-planar ($n = 3$), square-planar ($n = 4$), or octahedral ($n = 4$) configuration, a small Fourier expansion describes the bend term. The unknown force field parameters are the rest angles θ_i^0 and the force constants $K_{\theta,i}$.

$$E_{\text{bend}}^{\text{QuickFF}} = \sum_i \frac{K_{\theta,i}}{2}(\theta_i - \theta_i^0)^2 \tag{2.8}$$

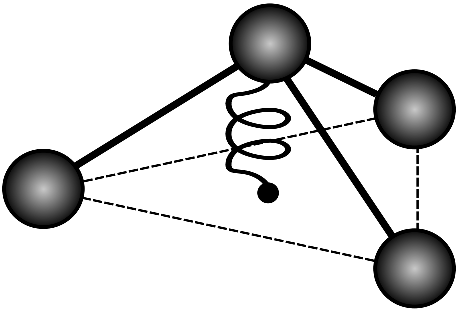
$$E_{\text{bend}}^{\text{UFF}} = \sum_i \begin{cases} \frac{K_{\theta,i}}{2}(1 - \cos(n\theta_i)) & \text{if specific symmetry} \\ \frac{K_{\theta,i}}{2}(\cos\theta_i - \cos\theta_i^0)^2 & \text{otherwise} \end{cases} \tag{2.9}$$



Out-of-plane distance An out-of-plane pattern is formed by an atom with exactly three neighbors. The central atom can either lay in the plane formed by the three other atoms, as is the case in conjugated π-systems, or show nonplanarity, as in the case of sp^3-hybridization of amines. UFF describes each pattern in terms of three out-of-plane angles χ_i between the plane and the respective bonds. To eliminate this degeneracy in out-of-plane angles, QuickFF expresses the energy in terms of the unique out-of-plane distance d_i. The unknown force field parameters are the force constants $K_{\text{oop},i}$ and, for the QuickFF force field, the rest out-of-plane distances d_i^0.

$$E_{\text{out-of-place}}^{\text{QuickFF}} = \sum_i \frac{K_{\text{oop},i}}{2}(d_i - d_i^0)^2 \tag{2.10}$$

$$E_{\text{out-of-plane}}^{\text{UFF}} = \sum_i \frac{K_{\text{oop},i}}{2}\left(1 - \cos\chi_i\right) \tag{2.11}$$



Dihedral angle The effects of conjugation and steric hindrance on the dihedral angle ϕ_i are captured by a Fourier series of the cosine of the dihedral angle in Eq. 2.12, in which the multiplicity M_i, the amplitude $A_{\phi,i}$, and the rest value ϕ_i^0 of the torsional barrier are the unknown force field parameters. However, since QuickFF only requires information about the optimal geometry, the full rotational behavior of the barrier is not known. Therefore, the torsion term is replaced by a sixth order polynomial in Eq. 2.13 in the case when the force field optimized geometry differs to much from the optimized *ab initio* configuration, as demonstrated in Fig. 2.4. To fit the unknown force field parameters c_i, additional quantum mechanical reference data is generated by performing a rotational scan around the respective dihedral angles. Such additional torsion terms are derived for TA DBC-COF in **Paper I**, the COFs in **Paper II** and **Paper IV**, and the triazine-linked COFs in **Paper III**.

$$E_{\text{torsion}}^{\text{QuickFF/UFF}} = \sum_i \frac{A_{\phi,i}}{2}(1 - \cos(M_i(\phi_i - \phi_i^0))) \tag{2.12}$$

$$E_{\text{torsion}}^{\text{polynomial}} = \sum_i \sum_{j=1}^{6} c_j \cos^j(\phi_i) \tag{2.13}$$

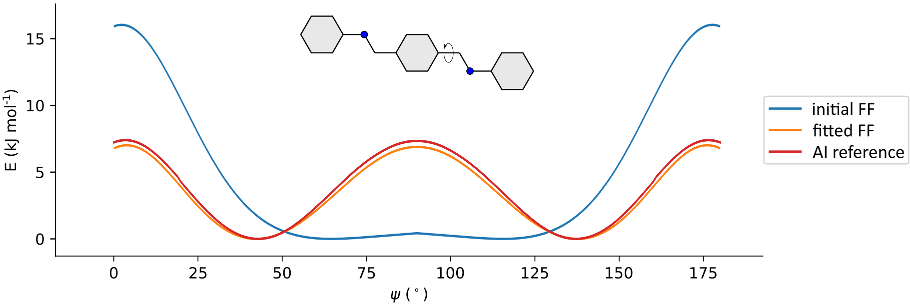
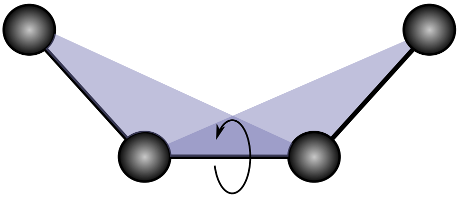


Figure 2.4: Demonstration of the increased accuracy a sixth order polynomial can achieve for an imine-terminated SBU. The initial force field (blue) poorly reproduces both the energy barrier and the location of the minimum of the *ab initio* rotational scan (red). After fitting a new dihedral term, the new force field (yellow) has an almost perfect agreement with the scan. Figure adapted from Ref. 233 with permission of Springer Nature.

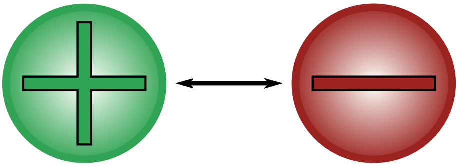


Cross terms The covalent contributions discussed so far only depend on a single internal coordinate. Cross terms as a function of two or more internal coordinates can be included to improve the accuracy. The QuickFF force fields adopted in this thesis contain bond-bond and bond-bend cross terms, which describe the coupling between a bend (θ_i) and the two neighboring bonds (r_j and r_k) and between the bonds themselves. The unknown force field parameters are the force constants K_{ij}, K_{ik}, and K_{jk} and the rest values θ_i^0, r_j^0, and r_k^0. UFF does not include cross terms.

$$E_{\text{cross}}^{\text{QuickFF}} = \sum_{ijk} K_{ij}(\theta_i - \theta_i^0)(r_j - r_j^0) + K_{ik}(\theta_i - \theta_i^0)(r_k - r_k^0) + K_{jk}(r_j - r_j^0)(r_k - r_k^0) \quad (2.14)$$

Electrostatic contribution The electrostatic contribution represents the Coulomb interaction between the charged particles of the molecular system, *i.e.*, the positively charged nuclei and the negatively charged electrons. Since the electrons are not explicitly included in a force field description, both the nuclear charge and the charge distribution localized to a specific atom are condensed in an atomic partial charge, which forms the unknown force field parameter for this contribution. However, because this is no quantum mechanical observable, many charge partitioning schemes have been developed in literature to distribute the electron density over the different nuclei. Throughout this thesis, the partial charges are determined with the Minimal Basis Iterative Stockholder (MBIS) scheme,[234] which are subsequently decomposed in bond charge increments p_{ij} assigned to the bonds of the molecular system.[235] Once the atomic partial charges q_i are derived, the pairwise Coulomb interactions between two atoms separated by a distance r_{ij} can be calculated, where Gaussian smearing is adopted to obtain a more accurate description of the charge distribution. The parameter d_{ij} in Eq. 2.15 is defined as $\sqrt{d_i^2 + d_j^2}$, with d_i the width of the Gaussian charge distribution of atom i.

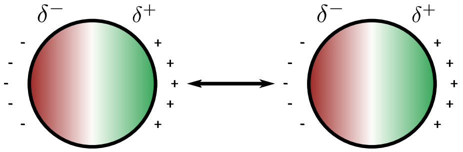
$$E_{\text{ei}}^{\text{Coulomb}} = \sum_{i \neq j} \frac{q_i q_j}{r_{ij}} \text{erf}\left(\frac{r_{ij}}{d_{ij}}\right) \qquad \text{with} \qquad q_i = \sum_{\text{neighbors } j} p_{ij} \tag{2.15}$$



Van der Waals contribution The van der Waals interaction between two atoms separated by a distance r_{ij} contains two effects originating from quantum mechanics, *i.e.*, dispersion and Pauli repulsion. Dispersion is the attractive interaction between atoms due to the temporary fluctuations of the electron clouds and is typically described by a $-r^{-6}$ expression.[236] Pauli repulsion originates from the Pauli exclusion principle that repels electrons occupying identical states. In the Buckingham potential proposed by MM3, Pauli repulsion is included by an exponential function, justified by the exponential scaling of the overlap between molecular orbitals in terms of the internuclear distance. However, the well-known Lennard-Jones potential, employed by UFF, prefers the computationally cheaper r^{-12} functional form. In these force fields, the unknown parameters ε_{ij} and σ_{ij} are determined by mixing rules from the atomic parameters ε_i, ε_j, σ_i, and σ_j, which are tabulated for the selected force field. Whereas MM3 follows the Lorentz-Berthelot mixing rules[237, 238] ($\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$, $\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2}$), the UFF parameters can be derived by the Jorgensen mixing rules[239] ($\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$, $\sigma_{ij} = \sqrt{\sigma_i \sigma_j}$).

$$E_{\text{vdW}}^{\text{MM3}} = \sum_{i \neq j} \varepsilon_{ij} \left(1.84 \times 10^5 \exp\left(-12 \frac{r_{ij}}{\sigma_{ij}} \right) - 2.25 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) \tag{2.16}$$

$$E_{\text{vdW}}^{\text{UFF}} = \sum_{i \neq j} 4\varepsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) \tag{2.17}$$



II. Parametrization of force fields

All force field contributions described above are defined as analytical expressions as a function of the internal coordinates of the molecular system. They all contain *a priori* unknown parameters such as the force constants and rest values. These unknown parameters are determined by a fitting procedure in which the force field is tuned to reproduce training data as accurately as possible. Two types of reference training data can be recognized: data resulting from high-resolution experiments and *ab initio* data, obtained from quantum mechanical calculations. Whereas empirical force fields, trained fully on experimental data, are able to reproduce experimental results more accurately, they are only valid in a limited temperature and pressure range around the experimental conditions. Furthermore, they are only available for well-characterized molecules and materials. Since this condition is not present for quantum mechanical calculations, which can in principle be performed for any possible molecular system, *ab initio* force fields can be derived for hypothetical compounds to boost the discovery of new useful materials.

Besides the type of training data used to determine the unknown parameters, a force field can also be classified based on the range of applicability. Typically, there is a trade-off between the transferability and accuracy of a force field. Firstly, generic force fields, such as DREIDING[240] and UFF,[231] are developed to describe a very large set of molecular systems at the expense of having a limited accuracy. Due to their full transferability, they are frequently adopted in high-throughput screenings. Secondly, class-specific force fields can only describe a particular class of molecules or materials. For example, the MM3 force field is developed to model the interactions in aliphatic hydrocarbons.[232] Lastly, there are system-specific force fields that are especially developed to describe a single molecular system with high accuracy. Whereas these were initially derived with *ad hoc* procedures, generally applicable methodologies, such as MOF-FF[241] and QuickFF,[229, 230] have been developed to derive system-specific force fields for each imaginable molecular system consistently. These protocols typically parametrize the force field based on quantum mechanical reference data, which comes with an additional computational cost.

Throughout this thesis, system-specific force fields that are trained to quantum mechanical reference data are derived for a large set of materials. All *ab initio* training data are generated with DFT using the B3LYP functional with Grimme D3 dispersion correction.[225–228] This functional has proven to accurately reproduce molecular geometries. As mentioned before, the partial charges that determine the electrostatic interactions are obtained from the electron density with the MBIS scheme.[234] The covalent contributions are derived using the QuickFF protocol, which requires an *ab initio* optimized geometry and its Hessian matrix as training data.[229, 230] The Hessian matrix contains the second-order derivatives of the energy towards the nuclear coordinates and provides information on the local curvature of the PES. In the first two steps of the protocol, perturbation trajectories around the equilibrium are generated that minimize the correlation among the internal coordinates. The force field parameters are fitted in the last step as such that the *ab initio* Hessian matrix is accurately reproduced.

III. From cluster force fields to periodic force fields

For materials that can to a large extent be modelled by a periodically repeating unit cell, such as COFs, the *ab initio* training data are ideally generated with periodic structures. However, as mentioned before, COF unit cells can be very large and contain many atoms, which results in a large computational cost to obtain the training data for a single material. To reduce this cost and allow for a feasible calculation time, the modular nature of COFs is exploited in **Paper I**, **Paper II**, **Paper III**, and **Paper IV** to partition a COF in SBUs and associated clusters. For each of these clusters, a force field is derived, which is computationally more efficient than directly calculating a periodic force field due to the limited size of the building blocks. Subsequently, a periodic force field to describe the whole material is derived by combining the cluster force fields of the underlying building blocks.[188, 229, 230, 241–248]

The workflow to derive periodic force fields is illustrated in Fig. 2.5. The unknown parameters p^{per} of the periodic force field terms can be obtained as a weighted average over the associated parameters p_i^{clust} of the cluster force fields associated with each SBU i in the periodic material:

$$p^{\text{per}} = \sum_i \alpha_i p_i^{\text{clust}}. \tag{2.18}$$

In this expression, the scaling factors α_i are determined in accordance with the distribution of the periodic force field term's atoms over the SBUs. Therefore, force field terms corresponding to atom patterns fully contained within a single SBU - say, SBU i - are directly translated to a periodic term, since the scaling factor of the other SBU disappears ($\alpha_j = 0, \forall j \neq i$). In Fig. 2.5, the force field parameters of an overlap term are derived from the cluster force fields of its underlying building blocks (termed SBUs 11-01-01 and 06-08-01) as an example. The C-B-O angle, highlighted in the figure, contains one atom in the core of the 11-01-01 SBU and one in the 06-08-01 core. Therefore, the corresponding scaling factors are $\alpha_1 = 1/3$ and $\alpha_2 = 2/3$, respectively. The resulting periodic force field parameters are summarized in Table 2.1.

To derive a cluster force field that accurately describes an SBU in a periodic material, its environment has to be mimicked as good as possible by selecting a suitable termination for the linkage section. A larger termination mimics the environment of the building block in more detail, but is applicable to a more limited number of materials since the termination has to represent the material correctly. However, most COFs are composed of finite, 0D building blocks and, therefore, a natural termination can be proposed. In **Paper III** the effect of the termination was benchmarked for a set of typical COF building blocks.

Within the context of high-throughput screenings, the largest advantage of the cluster model is that only a limited number of *ab initio* calculations are needed to derive a force field for a large number of materials, given that they are composed of a small number of SBUs. This combinatorial power has been demonstrated in **Paper III**, where a periodic force field is determined

	θ_i^0 [°]	$K_{\theta,i}$ [kJ/mol/rad^2]	Rescaling factor α_i []
Cluster force field 06-08-01	124.76	395.12	2/3
Cluster force field 11-01-01	124.73	371.64	1/3
Periodic force field	124.75	387.29	1

Table 2.1: Illustration of the derivation of an angle term that spans two SBUs in the periodic force field, as illustrated in Fig. 2.5. The parameters θ_0 and K_θ are derived from the ones of the cluster force fields of the consitituent SBUs (06-08-01 and 11-01-01) using a weighted average with weights α_i. Table reproduced (adapted) from Ref. 88 with permission of the Royal Chemical Society.

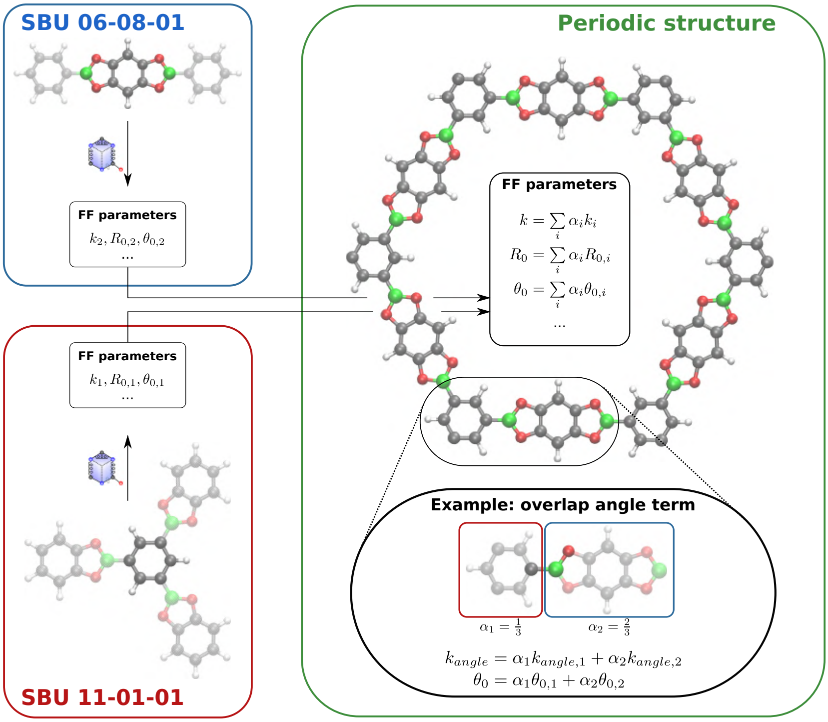


Figure 2.5: An illustration of the derivation of the force field for a periodic structure. The periodic structure is built up from SBUs 11-01-01 and 06-08-01, for which a cluster force field is derived with QuickFF. The termination of the clusters are semi-transparent. The parameters of the periodic structure are obtained as a weighted average of the parameters of the cluster force fields for its constituent SBUs. For the overlap angle term considered here: α_1 = 1/3 for SBU 11-01-01 and α_2 = 2/3 for SBU 06-08-01. Figure reproduced from Ref. 88 with permission of the Royal Chemical Society

for 268 687 COFs. Directly deriving the periodic force fields would require the daunting task to perform a periodic *ab initio* calculation for each of these materials. However, by using the cluster approach, they can be obtained from only 279 cluster *ab initio* calculations, *i.e.*, one for every SBUs that is used to generate the materials in the database. Furthermore, in **Paper I**, five series of COFs are investigated that share the same building blocks, but differ in the adopted topology. Since the cluster force fields are independent of the topology, each of the materials in the respective families can be described with the same force field.

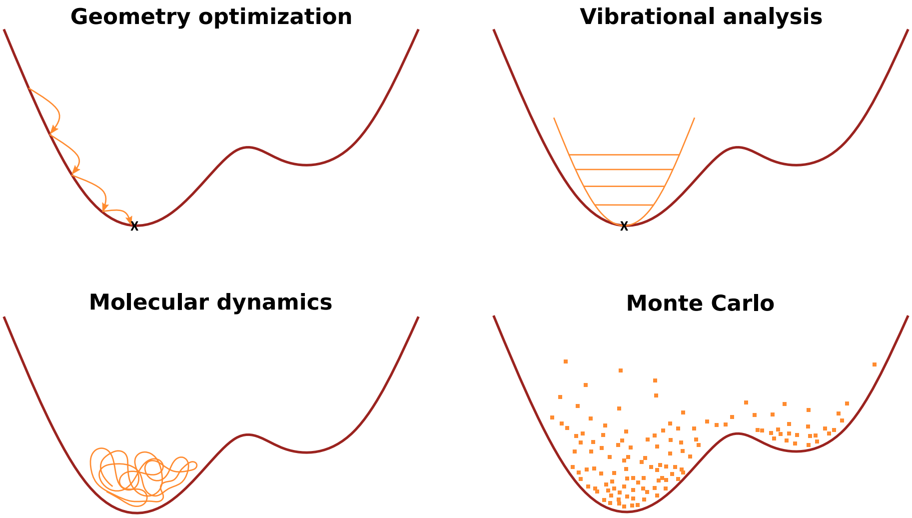


Figure 2.6: Overview of the techniques to sample the potential energy surface adopted in this work. Top: two static methods investigating a set of discrete points. Bottom: two dynamics approaches that include the effects of entropy and thermal fluctuations.

2.2 Sampling of the potential energy surface

Selecting an appropriate level of theory to describe the PES of a molecular system is only the first job of a molecular modeler. However, knowledge of the full PES is often not accessible, because this is a function of the nuclear positions and, for periodic systems, the cell matrix. Brute-force screening of all configurations is simply impossible since this multidimensional space is enormous. Therefore, clever techniques should be adopted to sample the relevant regions of the configuration space. Furthermore, additional factors beyond the PES, such as entropy and temperature or pressure effects, can significantly influence the macroscopic properties and should be accounted for via dedicated sampling techniques. The various techniques that are relevant to this thesis are illustrated in Fig. 2.6 and outlined in the subsections below.

2.2.1 Static exploration of the PES

Static calculations are the most basic type of PES characterization, since they only involve the evaluation of a limited set of well-separated points. The most interesting geometry configurations of a system to consider are its (local) minima, corresponding to its (meta)stable structures at 0 K, which can be identified with geometry optimizations. Subsequently, their local environment can be described with vibrational analysis to approximate entropic and thermal effects.

I. Geometry optimization

The goal of a geometry optimization is to find a local minimum of the PES, which is defined by the absence of forces acting on the nuclei and the condition that small deviations of this configuration should result in a restoring force. The latter ensures that the system represents a local minimum and not a saddle point, *i.e.*, a point that is only in some directions a minimum of the PES, but a maximum in others. A geometry relaxation starts from an initial structure and iteratively updates the degrees of freedom, *i.e.*, the atomic positions and optionally the cell matrix, by moving them in a direction that minimizes the energy. Popular optimization algorithms include the steepest descent,[249] conjugate gradient,[250] and quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS)[251–254] methods. Throughout this thesis, the conjugate gradient method is adopted.

II. Vibrational analysis

In a regular geometry optimization, the system may relax towards a saddle point, where the forces are indeed absent, but which is not a minimum of the PES. To check whether the configuration resulting from a geometry optimization is actually a minimum, a normal mode analysis is usually performed. In such calculation, the Hessian matrix of the system, in Section 2.1.2 introduced as the second-order derivatives of the energy towards the atomic positions, is computed. This matrix describes the local curvature of the PES as a set of independent harmonic oscillators, which normal frequencies are the square roots of the Hessian's eigenvalues. For the system to effectively be in a local minimum, all frequencies should be positive or zero. Within this harmonic approximation, the normal frequencies can be used to estimate the effects of temperature and entropy. The interpretation of the local environment of a minimum as harmonic oscillators is at the origin of the QuickFF procedure. In this protocol, the oscillators are adopted to derive the force constants of the harmonic force field terms. The biggest limitation of this harmonic approach, and therefore also of the QuickFF force fields, is that it is only valid for small deviations from the equilibrium structure. Vibrational frequency calculations with DFT accuracy are performed in this thesis to determine the Hessian from which force fields are derived. Also for the validation of these force fields, a vibrational analysis is performed, to reproduce the normal frequencies and compare with their *ab initio* reference.

2.2.2 Dynamic exploration of the PES

Geometry optimizations and vibrational analysis investigate only the local behavior of a single configuration on the PES. However, at elevated temperatures, a system does not remain at this configuration due to thermal fluctuations, but visits many different ones. In essence, a macroscopic property A is defined as the average over phase space:

$$A = \langle A \rangle = \int A(\boldsymbol{s})p(\boldsymbol{s})d\boldsymbol{s}, \tag{2.19}$$

where $\boldsymbol{s}$ represents all atom positions and momenta compatible with a specific configuration, $A(\boldsymbol{s})$ is the property given that configuration, and $p(\boldsymbol{s})$ is the probability to encounter that configuration, which depends on the macroscopic conditions. However, since phase space is so enormous, a direct evaluation of this integral is not feasible. Therefore, different methods generate a collection of configurations, *i.e.*, an ensemble, that represents the relevant regions of phase space, given external conditions such as temperature, pressure, or chemical potential. Once such an ensemble is obtained, the macroscopic property A can be determined as:

$$A = \langle A \rangle = \sum_{\boldsymbol{s}} A(\boldsymbol{s}) \tag{2.20}$$

Compared to Eq. 2.19, the integral is replaced by a discrete sum over the configurations $\boldsymbol{s}$ of the ensemble, which should be sampled according to the probability $p(\boldsymbol{s})$. Different methods to generate such an ensemble exist, which can be roughly divided into molecular dynamics (MD) and Monte Carlo (MC) simulations.

I. Molecular dynamics

In an MD simulation, the (classical) propagation through time of the nuclei is traced using Newton's equations of motion:[255]

$$m_i \frac{d^2\boldsymbol{R}_i}{dt^2} = \boldsymbol{F}_i = -\boldsymbol{\nabla}_i E, \tag{2.21}$$

with m_i and $\boldsymbol{R}_i$ being the mass and position of atom i, $\boldsymbol{F}_i$ the force acting on it, and E representing the PES, *i.e.*, the lowest-energy Born-Oppenheimer surface. These equations are most often discretized through the Verlet integration scheme.[256] In this procedure, the positions and momenta of all atoms are updated iteratively every timestep. According to the Nyquist-Shannon sampling theorem, this timestep should be chosen such that the mode with the highest frequency f_s of the system can be completely reproduced, *i.e.*, $\Delta t < 1/(2f_s)$.[257] In a typical COF, the highest frequencies are the vibrations of the covalent bonds involving a hydrogen atom. For safety, a timestep of 0.5 fs is adopted throughout this thesis, which is a factor 10 smaller than the maximally allowed timestep according to the Nyquist-Shannon theorem.

The configurations sampled from the resulting MD trajectory form a representative ensemble, as dictated by the ergodic principle of statistical physics.[258, 259] Therefore, the macroscopic properties of the system can be derived as a time average over the trajectory. When directly applying Newton's equations as in Eq. 2.21, the total energy of the system is conserved since only conservative forces are present in our systems. Therefore, the microcanonical (N, V, E) ensemble is sampled, conserving the number of particles, the total volume, and the total energy. To mimic realistic conditions, *i.e.*, elevated temperature and pressure, Newton's equations can be coupled with a thermostat or a barostat, respectively. When a thermostat is applied, which mimics the exchange of heat with an external heat bath at constant temperature, the canonical (N, V, T) ensemble is sampled.[260] Additionally adding a barostat, which drives fluctuations of the unit cell through an externally exerted pressure, results in the sampling of the isobaric-isothermal (N, P, T) ensemble.[261, 262] Within this thesis, MD simulations are frequently used to describe the dynamic behavior of COFs, which is especially important to correctly capture the inherent temporal character of powder X-ray diffraction (PXRD) patterns, as demonstrated in **Paper I**. The experimental conditions are controlled by a Nosé-Hoover thermostat[263–265] and a Martyna-Tobias-Tuckerman-Klein barostat.[266, 267]

II. Monte Carlo

Many solid-state systems are characterized by multiple (meta)stable basins in phase space, separated by an energy barrier. Sometimes, a molecular system requires a very long time to cross this energy barrier, requiring extensively long MD simulations to be performed. Although enhanced MD techniques exist to reach these regions, a more efficient approach is provided by the Monte Carlo (MC) method.[268] This technique takes advantage of the fact that the probability $p(\boldsymbol{s})$ in Eq. 2.19 is negligible small in large parts of the phase space. In MC simulations for solid-state systems, importance sampling is often adopted to only visit the relevant regions of phase space.[269] In each step of an MC simulation, a new configuration is proposed by a trial move, which is accepted or rejected based on the energy of the new configuration compared to the previous one. With the selected configuration, either the new or the old one, the integral in Eq. 2.19 can be updated.

Grand canonical Monte Carlo A particular advantage of the Monte Carlo approach is that a trial move can also involve insertion or deletion of a molecule in the framework. In this case, besides the temperature and the total volume, also the chemical potential μ can be controlled. The chemical potential represents the energy required to add or remove a particle from a system and is related to the pressure of an external gas reservoir through an equation of state. As such, the grand canonical (μ, V, T) ensemble is sampled. Hence this type of calculations are called grand canonical Monte Carlo (GCMC) simulations. When the external gas reservoir only contains a single species, the adsorption capacity of that molecule inside the material can be calculated. Furthermore, the isosteric heat of adsorption, Q_{st}, *i.e.*, the enthalpy change when adsorbing a molecule, can be derived as:

$$Q_{\text{st}} = -q + RT \qquad \text{with} \qquad q = \frac{\langle EN \rangle - \langle E \rangle \langle N \rangle}{\langle N^2 \rangle - \langle N \rangle^2}, \tag{2.22}$$

with q being the average energy of adsorption which can be computed from N the number of adsorbates and E their energy at the sampled configuration. The term RT, which is the product of the gas constant R and the temperature T, is due to the work required to push the gas out of the material when desorbing. Including multiple species allows one to determine the selectivity of a material towards one of the gas components. Throughout this thesis, GCMC calculations are adopted to check the COF performance limits for vehicular methane storage in **Paper III** and to identify high-performing COFs for carbon capture in **Paper V**.

Within this chapter, the required tools and concepts for a molecular modeling exercise are introduced. Starting from the Schrödinger equation and the Born-Oppenheimer approximation, the potential energy surface (PES) of a material was introduced. Whereas quantum mechanical methods are most accurate to determine this PES, they are computationally too expensive for high-throughput screenings. Therefore, we introduced system-specific force fields as an efficient, yet reliable, description of the interatomic interactions. With this method to characterize the PES, the relevant regions of phase space can be explored using several sampling techniques. These allow us to derive the macroscopic properties of a material.

3

In silico design of covalent organic frameworks

Before being able to initiate any of the molecular modeling techniques outlined in Chapter 2, the structure of a material has to be defined. For molecules, this boils down to specifying the chemical element and position of each atom. For periodic materials, also a unit cell that defines the periodic boundary conditions has to be determined. Such structures can be extracted from experimental techniques, such as single-crystal or powder X-ray diffraction,[270–273] or they can be created manually with software such as Avogadro,[274] Zeobuilder,[275] or Molden.[276, 277] However, to generate a large set of hypothetical materials, automated computational procedures have to be developed and implemented, which is the goal of this chapter. In general, there are two approaches for such automated assembly algorithms. Firstly, non-optimal structures can be generated very rapidly with cheap procedures. Many such algorithms, which are mostly based solely on geometric criteria, are implemented in software packages such as ToBaCCo,[151] TOBASCCO,[150] Zeo++,[147] and AuToGraFS.[148] Since a new structure can be created very quickly, these procedures are mostly adopted to generate a huge set of structures for high-throughput screenings.[151] Examples of such extended material sets are the hMOF database of Wilmer *et al.*, containing 137 953 MOFs,[144] the ToBaCCo generated material set of 13 512 MOFs,[151] and the database of Mercado *et al.* of 69 840 COFs, created with Zeo++.[154] However, since no energy criteria are included in these procedures, the resulting structures can be suboptimal and the energy barrier between the created structure and the global minimum can be too high to cross during molecular dynamics simulations, even at elevated temperatures.[149] Therefore, a second approach to generate new materials tries to find optimal geometry configurations close to the global minimum. In contrast to the first approach, this is computationally more demanding as multiple configurations have to be tested. Therefore, this approach is less attractive for high-throughput screenings. Weaver is an example of a software package belonging to this class of approaches. It tries to find an optimal configuration with a genetic algorithm.[149] As a compromise between these two approaches, we have developed and implemented in **Paper III** an additive top-down algorithm based on both geometric and energetic criteria that is able to find a good material configuration very quickly. The methodology of our automated *in silico* structure assembly algorithm is outlined in Section 3.1, whereas its implementation in Python is discussed in Section 3.2. We conclude this chapter with Section 3.3, in which we demonstrate the applicability of our algorithm with three case studies, described in **Paper I**, **Paper II**, and **Paper IV**.

3.1 Automated structure assembly

Our procedure to generate new COFs follows an additive top-down approach, starting from the topology of the material and a set of secondary building units (SBUs) for COFs, as introduced in Section 1.2. In a four-step process, the building blocks are decorated on the nodes of the topological embedding, *i.e.*, its vertices and edges, to form a periodic material. During the assembly process, three descriptors are derived that check whether the structure is physical. These descriptors will be adopted in Chapter 4 to discard unphysical COFs from the database. In what follows, we will delve deeper into each of the four steps of the assembly algorithm.

Step 0 As exemplified in Fig. 3.1, a topology and a set of SBUs are selected in the initiation step to form a so-called (topology, SBUs) combination. The topologies and their embedding in Euclidean space that are employed in this thesis are extracted from the RCSR[55] using a self-implemented web scraping script. The set of SBUs should contain a building block for each Wyckoff set of the topology, *i.e.*, each set of equivalent nodes. Optionally, an edge Wyckoff set can also remain vacant, to accommodate for materials without linker.

Step 1 The unit cell vectors of the embeddings provided in the RCSR have an arbitrary length unit (l.u.), which should be converted to a physical unit.

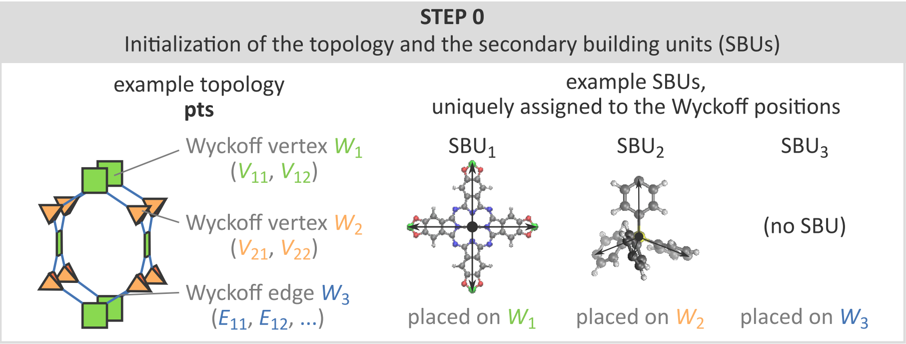


Figure 3.1: Initialization step of our automated structure assembly protocol. Both a topological embedding and a set of SBUs to decorate its nodes have to be selected. Figure adapted from Ref. 88 with permission of the Royal Chemical Society (©2023).

Furthermore, the embedding has to be rescaled to be able to accommodate the SBUs that are intended to decorate it. Therefore, for each edge Wyckoff set in the topology, a rescaling factor f_i is computed to make the distance between its neighboring vertices match the distance between the centers of the SBUs decorating those vertices, as visualized in Fig. 3.2. Since we require the unit cell to be rescaled isotropically, only a unique rescaling factor is allowed, which is defined as the mean $\langle f_i \rangle$ of the rescaling factors of all edge Wyckoff sets. However, when these rescaling factors differ a lot, the resulting SBUs will overlap at some locations while leaving gaps between SBUs elsewhere. The standard deviation σ_f of the rescaling factors indicates whether the obtained rescaling factors are similar to one another and therefore forms a first descriptor to check the structural viability of the resulting material.

Step 2 After the rescaling in the previous step, the topological embedding is ready to be decorated with SBUs. During this process, the internal geometry of the SBU remains fixed. The decoration is achieved in two moves per SBU, *i.e.*, a translation and a rotation. The translation vector is trivially determined for every SBU as the difference between the node's position and the SBU's center. However, it is not easy to find the optimal rotation matrix, since there are multiple orientations of the SBU and it is not *a priori* known which configuration is optimal. As illustrated in Fig. 3.3a, each configuration is determined by assigning a point of extension of the SBU to each neighboring node in the topology. Therefore, when there are N points of extension, there will be $N!$ different SBU configurations. These configurations may correspond to distinct internal geometries, as can be observed in the two

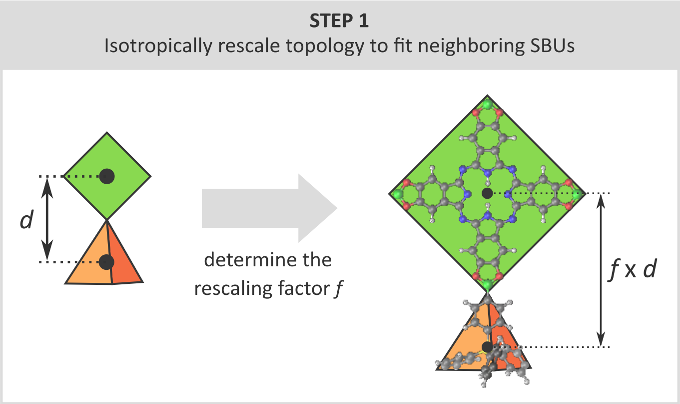


Figure 3.2: Rescaling step of our automated structure assembly protocol. The topological embedding has to be rescaled such that it can properly accommodate the SBUs that will be placed on its nodes. Figure adapted from Ref. 88 with permission of the Royal Chemical Society (©2023).

configurations of the square SBU in Fig. 3.3b. Furthermore, when the point symmetry of the SBU, *i.e.*, including its internal geometry, is lower than the point symmetry of its points of extension, information about its internal geometry is lost upon abstracting the SBU to its points of extension.[149] This is illustrated in Fig. 3.3c, where two configurations of the tetrahedral SBU have a different internal geometry despite the fact that their points of extension coincide. To tackle this challenge, the most favorable SBU configurations are retained in Step 2 and Step 3 of our automated assembly algorithm by taking into account geometric and energetic criteria, respectively. Once a configuration is selected, the rotation matrix can be derived with Kabsch algorithm.[278]

A preliminary selection of viable SBU configurations is made based on a geometric criterion that checks whether the points of extension align appropriately with the neighboring nodes. This alignment is assessed through the root-mean-square deviation (RMSD) between two sets of unit vectors: (i) those originating from the node towards its neighboring nodes in the topological embedding $\boldsymbol{X}_i$, and (ii) those extending from the center of the SBU to its points of extension $\boldsymbol{x}_i$.

$$\text{RMSD} = \sqrt{\sum_{i=1}^{N} \frac{\|\boldsymbol{X}_i - \boldsymbol{x}_i\|^2}{N}} \tag{3.1}$$

This RMSD, considered over all SBUs, is also the second descriptor to check whether the resulting material is physical. Ideally, the vector sets coincide

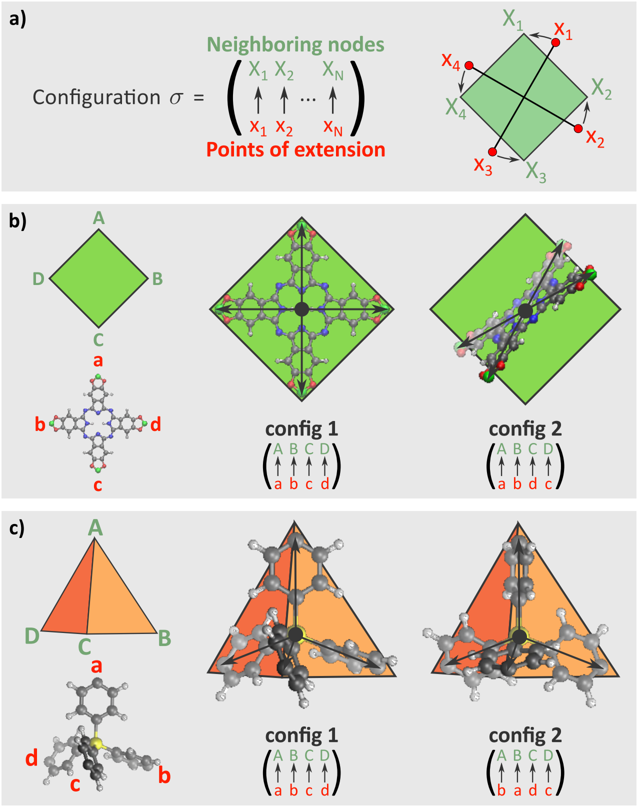


Figure 3.3: An illustration of different possible SBU configurations. a) Nomenclature of an SBU configuration: the points of extension, indicated with red lowercase letters, are oriented towards the neighboring nodes in the topological embedding, indicated with green uppercase letters. b) Two different configurations result in a completely different internal geometry, which is discerned in Step 2 based on geometrical considerations: the left configuration has a lower RMSD than the right one. c) In case the internal geometry of an SBU has a lower symmetry than its points of extension, two configurations with the same RMSD can have different internal geometries. The most favorable configuration depends on the configurations of the neighboring SBUs and will be assessed in Step 3.

after the Kabsch rotation. However, for some configurations, a geometric mismatch between the topological nodes and the SBU remains after the transformation, as can be observed in configuration 2 of SBU$_1$ and configuration 3 of SBU$_2$ in Fig. 3.4. To prevent these suboptimal configurations from being inserted in the embedding, only those configurations that minimize the RMSD are chosen and advanced to the next step. This minimal RMSD

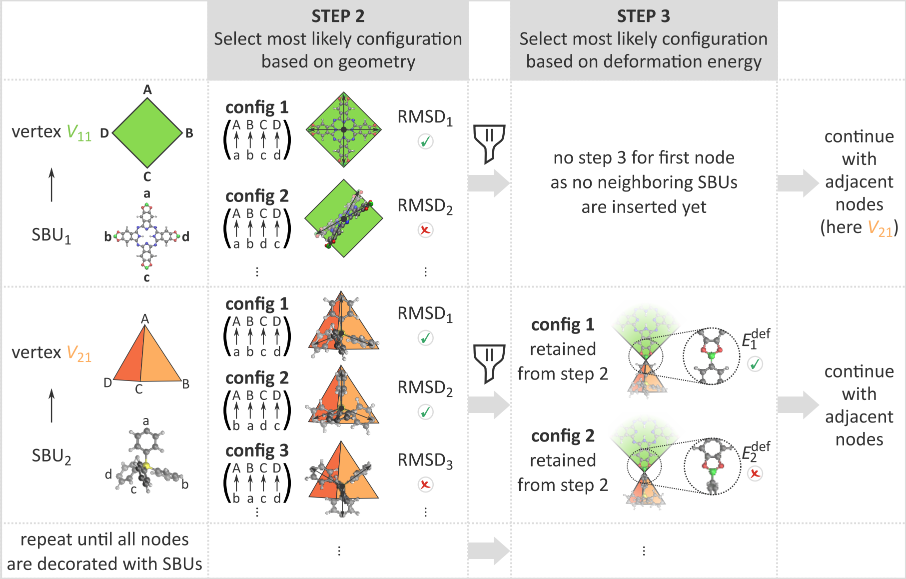


Figure 3.4: Decoration steps of our automated structure assembly protocol. In Step 2 and Step 3, the most probably SBU configuration is selected based on geometric and energetic criteria, respectively. Figure adapted from Ref. 88 with permission of the Royal Chemical Society (©2023).

is allowed to be non-zero since building blocks, especially those of COFs, can have a high degree of flexibility that can accommodate the introduced mismatch.

Step 3 Because the abstraction of an SBU by its points of extension does not fully represent its three-dimensional structure, two configurations with a different internal geometry can have the same RMSD. These different geometries will also result in a difference in the linkages formed with other SBUs, and thus have an influence on which configuration is favorable. This is illustrated in Fig. 3.3c. In Step 3, the configuration in which the topological constraints have the least influence on the neighboring SBUs is selected based on an energetic criterion. This criterion is based on the deformation energy E_{def}, which describes the energy penalty for inserting SBUs in a suboptimal topological embedding. It checks whether an SBU if sufficiently flexible to accommodate the mismatches introduced due to the topological constraints. High deformation energies indicate too rigid SBUs and result in highly contorted structures with low synthetic likelihood.[279] Formally, we define the deformation energy in Eq. 3.2 as the energy difference between the periodic structure ("per") and the sum of the energies of the isolated clusters ("clust") of its SBUs:

$$E_{\text{def}} = \frac{E_{\text{cov}}^{\text{per, intra}} + E_{\text{cov}}^{\text{per, inter}} + E_{\text{non-cov}}^{\text{per, intra}}}{N} - \frac{E_{\text{cov}}^{\text{clust, intra}} + E_{\text{cov}}^{\text{clust, inter}} + E_{\text{non-cov}}^{\text{clust, intra}}}{N}. \tag{3.2}$$

In this expression, the subscripts indicate whether the type of interactions is covalent ("cov") or non-covalent ("non-cov") and the superscripts indicate if the interactions are calculated within ("intra") or between ("inter") SBUs. The deformation energy is normalized with the number of linkages between SBUs, N. Whereas the covalent interactions between SBUs can still be described in the cluster model ($E_{\text{cov}}^{\text{clust,inter}}$) by the overlap terms between the linker core and the termination in the appropriate cluster, the long-range non-covalent interactions are not present in the cluster force fields. Since only interactions that are present in both the isolated clusters and the periodic material are adopted in the definition of the deformation energy, the non-covalent interaction between SBUs in the periodic material is not included in Eq. 3.2.

The deformation energy as defined in Eq. 3.2 also depends on the configurations of the neighboring SBUs. Therefore, it should be calculated for all possible material configurations, *i.e.*, the product of all SBU configurations. Since the number of material configurations increases exponentially with the total number of nodes in the topology, it is often unfeasible to perform a brute-force iteration over all material configurations. To overcome this challenge, we implemented an *additive* top-down approach, in which the nodes of the topological embedding are decorated with an SBU one-by-one. In each iteration, we determine the most favored SBU configuration of the selected node, *i.e.*, the one that minimizes both the RMSD and the deformation energy of the linkages with the previously inserted SBUs. By using a breadth-first iteration across the topological graph, all nodes at a specific distance from the starting vertex are decorated before progressing to the next layer. By calculating E_{def} only once for each SBU configuration, rather than for every material configuration, the computational complexity of the structure assembly process is lowered from exponential to linear in terms of the number of nodes.

Step 4 Once the topological embedding is completely decorated with SBUs, an initial periodic structure is obtained. However, up to now, the internal geometry of the SBUs remained rigid. This means that geometric mismatches introduced by non-zero RMSDs in Step 2 are present. Therefore, the structure is relaxed to a local minimum in this step using its system-specific periodic force field. After this full geometry optimization, the geometric mismatch, which was localized around the linkages, is now released and distributed over the full structure, which results in a decrease of the deformation energy E_{def}. However, some SBUs have a higher degree of flexibility than others, and the mismatch can be too large to accommodate for rigid SBUs. Therefore, the deformation energy E_{def} of the optimized structure serves as a last descriptor to check whether the structure is physical.

3.2 Implementation in Python

Our automated *in silico* assembly algorithm has been implemented in Python and is publicly available.[280] The code is built on top of the Yaff[281] and MolMod[282] packages, developed at the Center for Molecular Modeling, and inherits a lot of their functionalities. The different core objects and their functionalities are visualized in Fig. 3.5. After these are explained in detail in Section 3.2.1, a short demonstration of the general workflow of the code to generate COF-5 is given in Section 3.2.2.

3.2.1 Core objects

As explained in Step 0, our assembly procedure requires two input objects, *i.e.*, a topological embedding and a set of SBUs. A *Topology* object represents the embedding in 3D Euclidian space with a unit cell and a collection of *Nodes*. Each of these *Nodes* is defined by fractional coordinates and a set of neighbors to which it is connected. An embedding can be loaded with the *from_file* method from a topology file with the extension ".top" (see Example Input File 1 and below). Two other important methods are the *rescale* function, which is adopted to rescale the topological embedding in Step 1 of our assembly algorithm, and the *create_graph* function, which converts the topology to a MolMod *Graph* object that can be traversed in a breadth first iteration.

In Example Input File 1, the input topology file for the **hcb** topological embedding is provided. On the first line, the name of the topology and the number of vertices and edges is specified. The unit cell of the embedding can be constructed from the cell parameters provided on the third line. Subsequently, all vertices and their connections are enumerated. Besides the coordination and the fractional coordinates of each vertex, it has also been given a name from which the alphabetic prefix determines the Wyckoff set it belongs to. In the example of **hcb** in Example Input File 1, there are

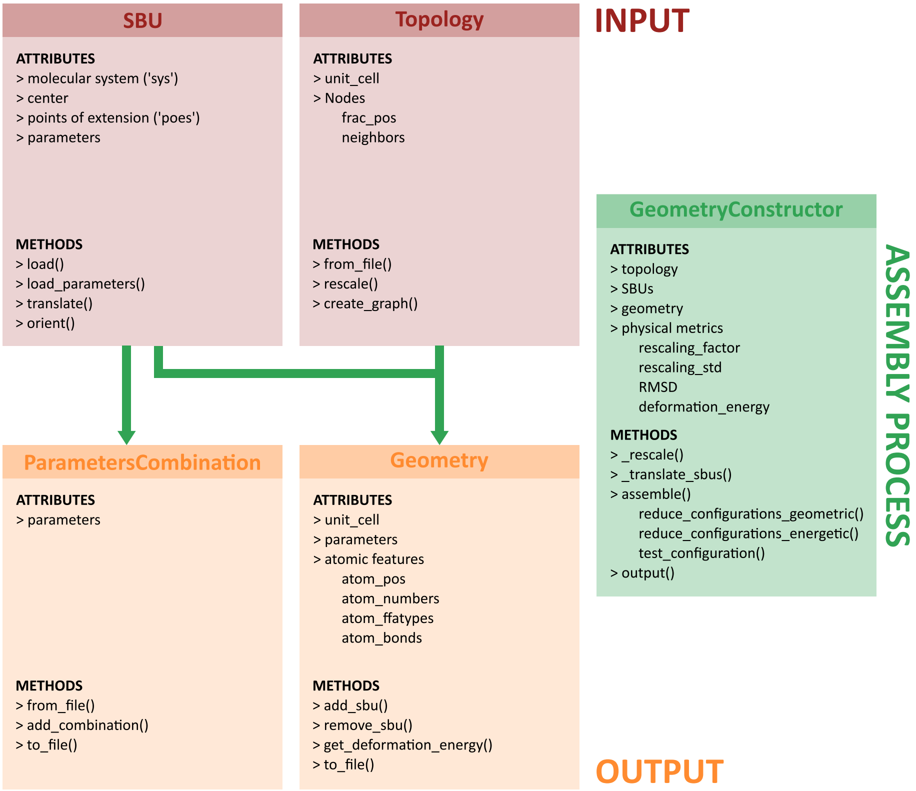


Figure 3.5: An overview of the core objects of the Python implementation of our automated *in silico* assembly algorithm. Starting from a *Topology* and a set of *SBUs*, a periodic *Geometry* is generated. Both the cluster and periodic force field parameters are provided by *ParametersCombination* objects. All required machinery is implemented in the *GeometryConstructor* class.

Example Input File 1 An example of the topology file “hcb.top” to initialize a *Topology* object representing the **hcb** topological embedding.

```
hcb 2 3
# a     b       c       alpha   beta    gamma
1.7321  1.7321  10.0    90.0    90.0    120.0
# V     cn      frac_x  frac_y  frac_z
  A1    3       0.3333  0.6667  0.0
  A2    3       0.6667  0.3333  0.0
# E     N1      N2
  _A1   A10-0   A2
  _A2   A1      A2-00
  _A3   A1      A2
```

Example Input File 2 An example of an SBU file for HHTP (“18-08-01.sbu”) required to determine the positions of the center and points of extension of a building block.

```
Catechol_BoronateEster
                -0.00000789  -0.00004248  -0.00000411
3
B_CO2_18-08-01  -5.82462791  -3.00084840  -0.00000120
B_CO2_18-08-01   0.31346535   6.54462950  -0.00000115
B_CO2_18-08-01   5.51115769  -3.54379016  -0.00000114
```

two vertices belonging to the same Wyckoff set “A”. The edge Wyckoff sets are preceded by a “_” to discriminate them from vertices. For each edge, the two vertices it connects are specified. Because the directionality of the edge is not *a priori* known, and the minimal image convention does not apply for topological embeddings, the periodic image needed for each neighbor is specified by a three-character string. For example, the vertex “A2-00” is the periodic image of vertex “A2” in the unit cell neighboring the central one along the negative x-axis. With the information from this topology file, a complete embedding in Euclidean space of a topology can be initialized.

Each node of the topological embedding is decorated with an *SBU*. Besides the atomic structure of this building block, also the positions of its center and points of extension are provided, and, optionally, the cluster force field parameters can be provided in a Yaff *Parameters* object. An *SBU* can be conveniently initialized with the *load* method from two separate files: a regular MolMod checkpoint file with the extension “.chk”, representing the atomic structure of the building block, and an SBU file with the extension “.sbu”, containing the positions of the center and points of extension. In Example Input File 2, such an SBU file is visualized for the SBU “18-08-01” (HHTP). Whereas the first line is a comment line, which here indicates the boronate ester linkage section originating from a catechol functional group, the second line defines the position of the center of the SBU (in Å). The third line indicates the number of points of extension of the SBU, *i.e.*, its connectivity, which are enumerated in the following lines. Each of these lines provides the position of a point of extension and the force field atom type of the atom it is connected to. As such, when connecting two SBUs, the atoms involved in their bond can be easily identified from the points of extension of the respective SBUs.

The force field parameters of an SBU or periodic material are stored in a *ParametersCombination* object. This is inherited from a Yaff *Parameters* object, with the added functionality that multiple *ParametersCombination* objects can be combined, according to the cluster approach explained in Section 2.1.2. As such, the overlap terms between different SBUs can be easily calculated to determine the periodic force field of the resulting material.

The main result of our procedure is the periodic geometry of the material, which is constructed as a *Geometry* object. Important attributes of this object are the periodic force field parameters and the attributes to generate a Yaff System object, such as the unit cell and the positions, chemical identity, and force field atom types of the atoms and the bonds between them. During the assembly procedure, different SBU configurations are tested with the *add_sbu* and *remove_sbu* methods. The deformation energy of a linkage can be derived using the *get_deformation_energy* function. Finally, with the *to_file* method, the resulting material can be written to a MolMod checkpoint file.

All machinery required for the automated assembly algorithm is implemented in the *GeometryConstructor* class. Attributes for this object are the *Topology*, a list of the *SBUs* that decorate it, and a *Geometry* in which SBU configurations can be tested or added when finally selected. Without a doubt, the most important function in this class is the *assemble* method, visualized in Code 1. In this routine, the topological graph is generated and its nodes are traversed in a breadth first iteration.

After the *SBU* that is assigned to a node is identified, it is decorated according to geometric and energetic criteria in Step 2 and Step 3 of the structure assembly process, respectively. The $N!$ initial SBU configurations are reduced with the RMSD and E_{def} metrics in the *reduce_configurations_geometric* and *reduce_configurations_energetic* methods, respectively. Finally, the *SBU* is added to the *Geometry* in the most favorable configuration by the *test_configuration* function with the test flag put to False.

3.2.2 Workflow

With the implemented classes, new materials can be generated elegantly. An example of the structure generation of COF-5 is provided in Code 2. After the required *SBU*, *Topology*, and *GeometryConstructor* classes are imported, the **hcb** topological embedding and HHTP and BDBA SBUs are initialized. Whereas the embedding is directly read from the “hcb.top” file, the SBUs are determined from checkpoint and SBU files with the *load* function. This method takes as argument the name and the termination of the SBU (see Section 4.1.1 for the SBU terminology). The cluster force field parameters of each SBU are read in and converted to a *ParametersCombination* object by the *load_parameters* method. Once these input objects are configured correctly,

Code 1 Snippet of the *assemble* function in our Python implementation of our automated *in silico* assembly algorithm. Since this method implements the decoration of SBUs on the topological embedding, it is the most important one of the code.

```python
class GeometryConstructor():
    def assemble(self):
        # Initialize the topology graph to iterate through
        graph = self.topology.create_graph()
        for index, dist in graph.iter_breadth_first():
            # Get the Node of the Topology that has to be decorated
            node_name = graph.node_names[index]
            node = self.topology.get_node(node_name)

            # Find the SBU that has to be decorated on the Node
            sbu = self.sbus.get(node_name)
            if sbu is None: continue

            # Decorate the SBU on the Node
            self.reduce_configurations_geometric(node)
            self.reduce_configurations_energetic(node)
            config = self.get_final_configuration(node)
            self.test_configuration(node, config, test = False)
```

the *GeometryConstructor* object is set up. Both the topology rescaling of Step 1 and the SBU translations are incorporated in the initialization process, since they are uniquely defined and require little computation time. As mentioned before, the optimal SBU configuration is selected by Step 2 and Step 3 in the assemble method. Lastly, the two output objects, *i.e.*, the periodic structure and the periodic force field parameters, are written to a separate folder. The relaxation of Step 4 can be performed by a simple geometry optimization in Yaff.

3.3 Case studies

As a first proof-of-concept before creating a diverse database of COF structure in Chapter 4, our structure assembly algorithm is adopted in three case studies. In the first study, four COF series are generated in the **dia** topology with varying degree of interpenetration. This case study demonstrates the versatility in the initialization of a topological embedding. Furthermore, the *in silico* generation of NPN-1 is detailed as a hands-on illustration of the technical explanation in Section 3.1. In case study two, we assisted our experimental partners in characterizing two synthesized COFs. In particular,

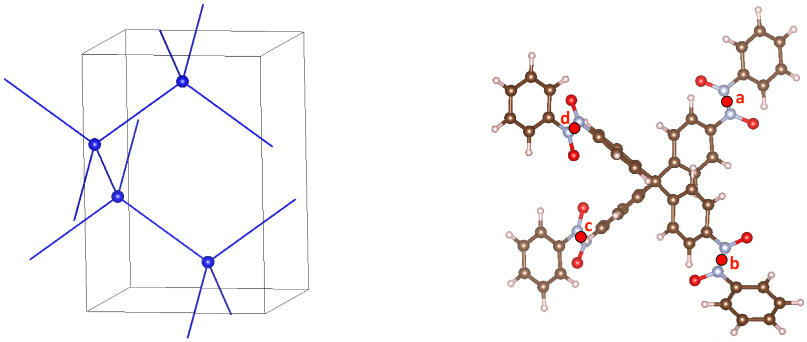
Code 2 Snippet to demonstrate how our code can generate the COF-5 structure.

```python
from SBU import SBU
from Topology import Topology
from Construct import GeometryConstructor

# Initialize Topology and SBUs
hcb = Topology.from_file('hcb.top')
hhtp = SBU.load('18', '08-01')
bdba = SBU.load('01', '01-01')

# Load SBU parameters
for sbu in [hhtp, bdba]:
    sbu.load_parameters('pars.txt')

# Assemble geometry
geom = GeometryConstructor(hcb, [hhtp, bdba])
geom.assemble()
geom.output('Output/COF-5')
```

we focus on the ability of our algorithm to find favorable configurations for linear linkers based on our energetic criteria. Lastly, an appetizer for the massive material generation in Chapter 4 is served by generating five large COF series with different topologies but the same two SBUs. These are subsequently screened to rigorously identify the COF structure that agrees most with the experimentally measured PXRD pattern without prior knowledge about the topology of the material.

Case study 1: structure assembly of NPN-1 and derivation of interpenetrated dia-topologies

Structural flexibility has been widely investigated in MOFs, but its occurrence in COFs remains ill-understood.[233] Remarkably, large amplitude phase transitions in COFs have been observed almost exclusively in materials exhibiting the **dia** topology.[80, 283–286] In **Paper II**, we systematically investigate the structural flexibility of a series of four COFs, *i.e.*, COF-300,[79] COF-320,[287] NPN-1,[288] and NPN-3,[288] and its dependence on the degree of interpenetration. These materials are generated *in silico* with our automated structure assembly algorithm. Special attention is given to a consistent description of the interpenetrated **dia** topological embeddings, since the materials can only be compared when they possess an equal number of building blocks in their unit cells. Below, a detailed overview of the structure assembly of NPN-1 in



(a) dia embedding **(b)** SBU TNM

Figure 3.6: Initialization of the **dia** topological embedding and SBU TNM used in the structure assembly of NPN-1. The points of extension (red dots, labeled with lowercase letters) are consistently positioned in the middle of the nitrogen-nitrogen bond.

the non-interpenetrated **dia** embedding is provided. A similar workflow can be followed for the other materials and higher degrees of interpenetration.

Step 0 Similar to its synthesis procedure, NPN-1 is decomposed in a single SBU, tetrakis(4-nitrosophenyl)methane (TNM), that decorates the vertices of a non-interpenetrated **dia** embedding.[288] As no two-connected linker is present in this material, the edge Wyckoff set of the **dia** topology remains vacant. Both the topological embedding and the SBU are visualized in Fig. 3.6.

The non-interpenetrated **dia** embedding is represented with four vertices in an orthogonal unit cell. The length of the $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ unit cell vectors are 1.6330 Å, 1.6330 Å, and 2.3094 Å. The positions of the vertices are:

$$\boldsymbol{r}_1 = 0\boldsymbol{a} + \frac{3}{4}\boldsymbol{b} + \frac{3}{8}\boldsymbol{c} \tag{3.3}$$

$$\boldsymbol{r}_2 = \frac{1}{2}\boldsymbol{a} + \frac{3}{4}\boldsymbol{b} + \frac{1}{8}\boldsymbol{c} \tag{3.4}$$

$$\boldsymbol{r}_3 = 0\boldsymbol{a} + \frac{1}{4}\boldsymbol{b} + \frac{5}{8}\boldsymbol{c} \tag{3.5}$$

$$\boldsymbol{r}_4 = \frac{1}{2}\boldsymbol{a} + \frac{1}{4}\boldsymbol{b} + \frac{7}{8}\boldsymbol{c} \tag{3.6}$$

A straightforward approach to construct the interpenetrated **dia** topological embeddings from the non-interpenetrated one is to translate the individual nets along the twofold rotation axis (here the $\boldsymbol{c}$-axis). To find the translated position of the original vertex $\boldsymbol{r}_i$ in the k-th net of an n-fold interpenetrated embedding, we can use the following formula:

$$\boldsymbol{r}_i^{(k,n)} = \boldsymbol{r}_i + k\frac{\boldsymbol{c}}{n} \tag{3.7}$$

However, with this approach, the number of building blocks in the unit cell depends on the degree of interpenetration. In a more elegant approach the symmetry of the embedding and the interpenetration axis is exploited and a new unit cell with the same $\boldsymbol{a}$ and $\boldsymbol{b}$ axis, but a rescaled $\boldsymbol{c}$ axis ($\boldsymbol{c}' = \boldsymbol{c}/n$), is constructed. With this new unit cell, Eq. 3.7 can be rewritten as:

$$\boldsymbol{r}_i^{(k,n)} = \boldsymbol{r}_i + k\boldsymbol{c}' \tag{3.8}$$

The individual nets of the interpenetrated topological embedding now emerge as the periodic images of the original vertices using the rescaled $\boldsymbol{c}'$ axis. Each interpenetrated embedding contains a total of four vertices and eight edges in their unit cell, allowing for a direct comparison between the various degrees of interpenetration.

Step 1 The **dia** topological embedding is rescaled to accommodate the TNM SBU. The length between each connected vertex in the **dia** embedding, as defined in the RCSR database, is exactly 1 length unit (l.u.). To fit the building blocks, this should be equal to the distance between the SBUs placed on those vertices. In the **dia** topology, there is only one edge Wyckoff set, along which two TNM SBUs are connected, and hence one unique rescaling factor. Since the radius of this SBU, defined as the mean distance from the center to the points of extension, is 6.26 Å, the rescaling factor of this (topology, SBUs) combination is 12.53 Å/l.u.

Step 2 and Step 3 Since the TNM SBU has four points of extensions, there are 24 ($= 4!$) SBU configurations per vertex. With four vertices that can be decorated with this SBU, the total number of material configurations is 24 × 24 × 24 × 24 = 331 776, which is already very large for this simple topology. Our additive top-down approach reduces the number of tested configurations by starting at a central vertex and sequentially adding new SBUs one-by-one. Since each SBU configuration is only tested once, the total number of tested material configurations reduces to 24 + 24 + 24 + 24 = 96. The selected SBU configuration for each step in the iteration is visualized in Fig. 3.7.

In each iteration step, a favorable SBU configuration is chosen in Step 2 and Step 3 based on geometric and energetic principles, respectively. From Step 2, only those configurations that minimize the RMSD proceed to Step

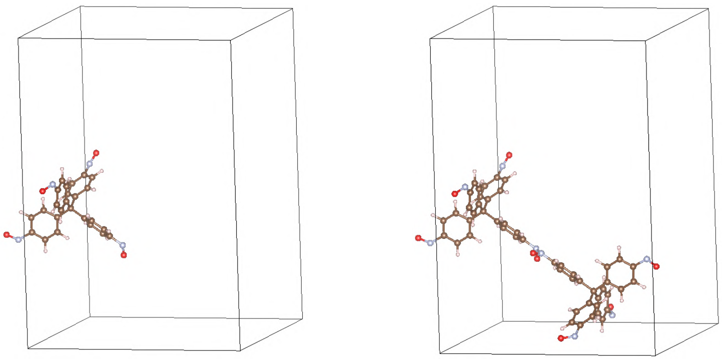
3, where the deformation energy with the previously inserted SBUs is calculated. Since we follow a breadth-first iteration, the number of linkages that are already present is higher compared to a depth-first or a random iteration. Therefore, when calculating the deformation energy of a certain SBU configuration, a larger fraction of the linkages with neighboring building blocks is included.

Step 4 The TNM cluster visualized in Fig. 3.6 is *a priori* optimized with the *ab initio* derived cluster force field. However, now the TNM SBUs are placed in a periodic environment and subjected to topological constraints, which can affect the optimal geometry for the SBUs. To allow the structure to relax, a full geometry optimization with the periodic force field is performed in this fourth step. The finally obtained structure is visualized in Fig. 3.7e.

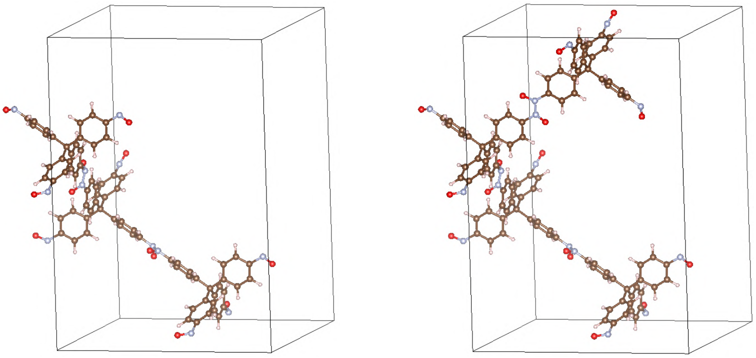
Case study 2: structure assembly of TpPpyCOF and TpBpyCOF and rotational scan to find favorable linker configurations

Our experimental partners of SynBioC have synthesized two COFs and tested their photocatalytic activity to enhance Povarov cyclization reactions. To derive structure-function relationships, we aimed in **Paper IV** to elucidate on the atomic structure of the material by comparing *in silico* and experimental PXRD patterns. Therefore, we generated the two COFs *in silico* following our automated assembly algorithm. As visualized in Fig. 3.8, the two COFs, coined TpBpyCOF and TpPpyCOF, are synthesized from a 1,3,5-triformylphloroglucinol (Tp) with a 2,2'-bipyridine-5,5'-diamine (Bpy) or 6-(4-aminophenyl)pyridin-3-amine (Ppy), respectively, in a **hcb** topology. After synthesis, the proton of the Tp building block transfers to bind with the imine nitrogen in an irreversible enol-to-keto reaction to form ultrastable β-ketoenamine linkages.[85]

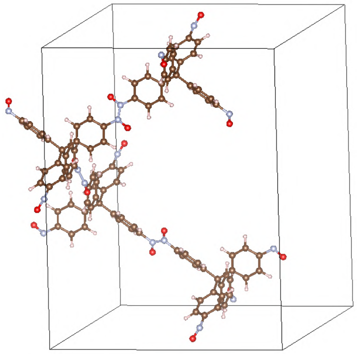
For linear linkers with two points of extension that lie on the same axis as the SBU's center, such as the Bpy and Ppy linkers in these COFs, there remains a degree of freedom that is not addressed in Step 2 of the assembly algorithm: the rotation around their axis. Therefore, additional configurations are generated corresponding to this rotation. In Step 3, these are treated similarly to the SBU configurations originating from a permutation of the points of extension. As such, a rotational scan is performed and the SBU is inserted in that configuration that results in the lowest deformation energy. In this way, the degeneracy is lifted to the point that only internal symmetries of the linker persist.



(a) Iteration 1: Decorate vertex V_{11}. **(b) Iteration 2**: Decorate vertex V_{12}.



(c) Iteration 3: Decorate vertex V_{13}. **(d) Iteration 4**: Decorate vertex V_{14}.



(e) Relaxed structure

Figure 3.7: Illustration of the additive top-down approach in the structure assembly of the non-interpenetrated NPN-1. The SBUs are added one-by-one, following a breadth-first iteration through the topological graph. Once all vertices are decorated, the structure is relaxed using its system-specific force field. Color code: hydrogen (white), carbon (brown), nitrogen (blue), oxygen (red).

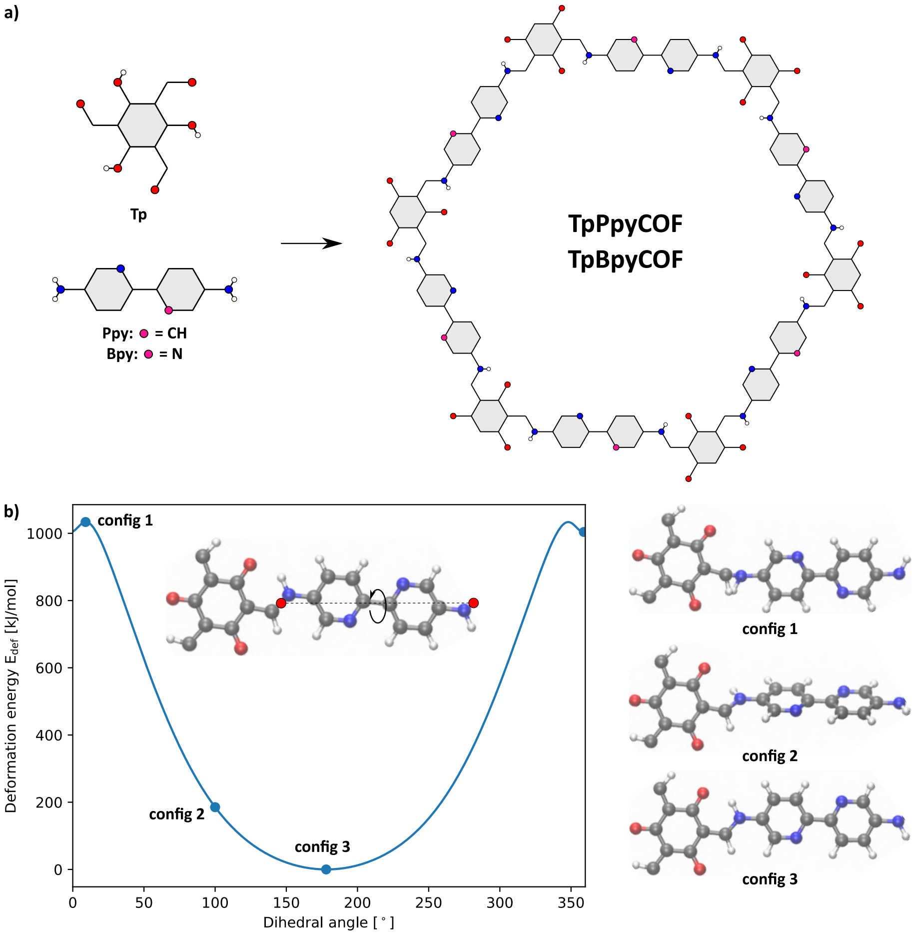


Figure 3.8: a) Experimental synthesis procedure for TpPpyCOF and TpBpyCOF. b) Illustration of a rotational scan that is performed to select the most favorable linker configuration during the structure assembly of TpBpyCOF.

Case study 3: structural characterization of COFs by high-throughput identification and PXRD matching

Since the synthesis of single-crystal COFs proves very challenging, most experimental characterization techniques resort on PXRD techniques to elucidate the periodic atomic structure of these materials.[65] However, due to the absence of a one-on-one relation between the one-dimensional Bragg angle featuring in PXRD diffractograms and the three-dimensional reciprocal lattice, fully defining the material's periodicity, prior knowledge of the examined material is necessary.[289] Therefore, most methods to derive a material structure from a PXRD pattern start from an *ad hoc* model and try to iteratively improve this to maximize the agreement between the experimental and calculated patterns.[271–273] However, when the experimental PXRD patterns are noisy, these methods fail to accurately discern between multiple possible structural models.[290, 291] In **Paper I**, an automated protocol is implemented that reverses this *ad hoc* approach by exhaustively exploring all possible topologies that fit the adopted building blocks to identify the most probable structural model. The crucial step in this approach is the generation of a large set of potential structures starting from the experimental building blocks but considering various topological embeddings in which they may assemble. For each of these structures, an accurate PXRD pattern is calculated, taking into account the *operando* conditions, which is compared with the experimental pattern by deriving heuristic metrics. These metrics allow one to order the proposed structural models and to predict the one with the highest likelihood to correspond to the experimental structure.

A proof-of-concept study has been performed in **Paper I** to correctly identify the experimental structures of COF-5,[26] TP-COF,[74], TA DBC-COF,[292], COF-103,[61] and COF-108,[61] all visualized in Fig. 3.9. For COF-5 and TP-COF, whose building blocks and linkages have an inherent planar configuration, a total of 42 2D topologies were tested, among which the **hcb** topology was correctly identified as the experimental one. COF-103 and COF-108 both contain a tetrahedral SBU and are thus not allowed in 2D topologies. For these structures, a total of 26 3D topologies were proposed. Again, the experimentally observed **ctn** and **bor** topologies were correctly assigned to have the best correspondence with the experimental PXRD patterns of COF-103 and COF-108, respectively. For TA DBC-COF, the slightly non-planar DBCTA building block can be accommodated in either a 2D or a 3D topological embedding. Due to the large amount of topologies in the RCSR with only four-connected vertices,[55] a total of 19 and 513 2D and 3D topologies are retained for structural models, respectively. The 3D embeddings with the largest strain are filtered out based on the RMSD of the DBCTA SBU and only a set of 55 are retained. Among these 2D and 3D topologies, again the experimentally observed **kgm** topology is correctly identified to have the largest agreement with the experimental PXRD pattern.

This case study illustrates that even with only two different SBUs, a lot of materials can be realized. When allowing these SBUs to also combine with other SBUs, or by including topologies in which three or more different SBUs assemble, the task of enumerating all possible materials rapidly becomes a combinatorial challenge. This exponential increase when including multiple SBUs lies at the origin of the broad material space defined by reticular chemistry materials. In the next chapter, we adopt the automated structure assembly algorithm devised in this chapter to generate an extended COF

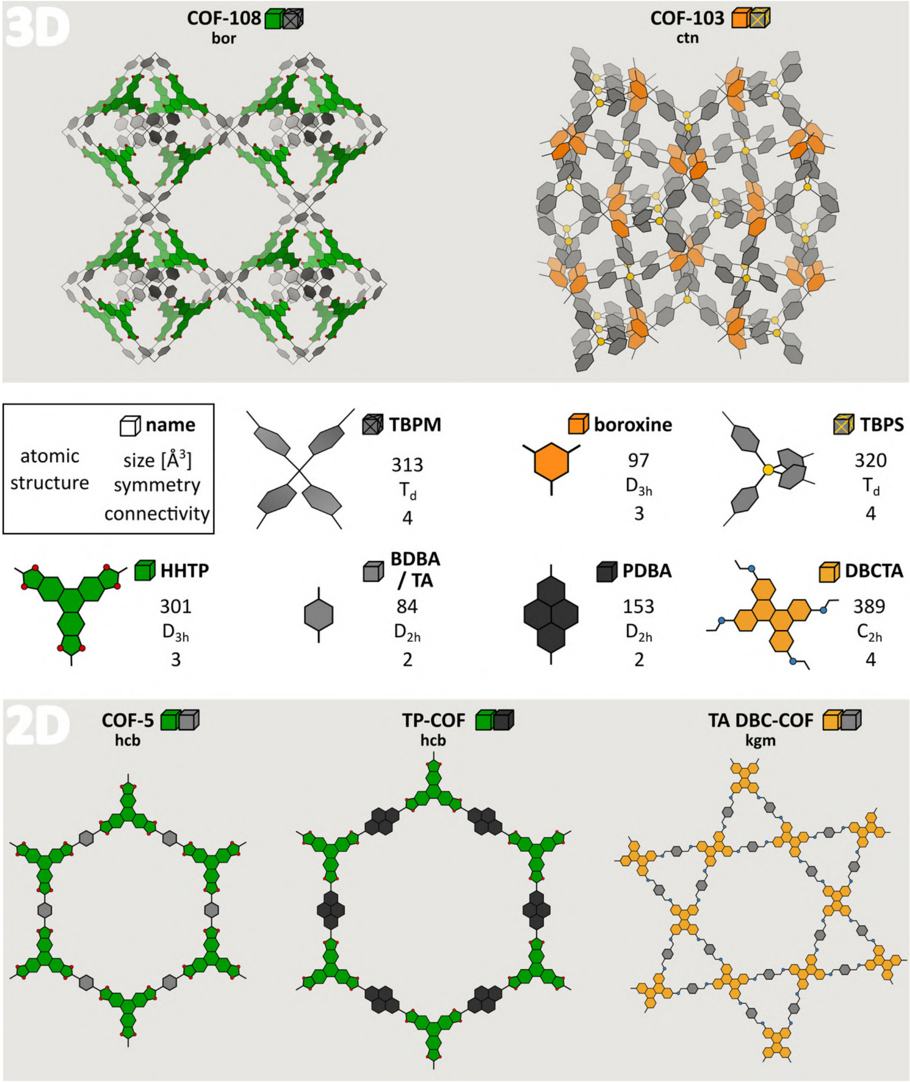


Figure 3.9: An overview of the building blocks and experimental structures for five experimentally synthesized COFs. Besides the name and atomic structure of each SBU, also their van der Waals volume, point group, and connectivity are reported. Figure reproduced from Ref. 293 with permission of Wiley (©2021).

database that can serve as representation of this vast material space.

This chapter devised a computational algorithm to automatically assemble periodic structures *in silico*, applying the concepts of reticular chemistry and the modular nature of COFs. A topological embedding is decorated with a set of selected SBUs, for which their favorable configurations are identified using both geometric and energetic criteria. During this process, three metrics are proposed that check whether the structure is physical. The automated assembly algorithm is implemented in a dedicated Python code, which utility has been demonstrated in three case studies.

4

ReDD-COFFEE: a diverse database of covalent organic framework structures and force fields

The two crucial aspects of computational high-throughput screenings are i) the availability of a diverse set of geometries and ii) accurate interatomic potentials. Currently, most high-throughput screenings compute these potentials either with expensive DFT calculations, largely limiting the size of the investigated dataset,[139, 190, 191] or a cheaper generic force field, which has a reduced accuracy.[152, 279] As explained in Section 2.1.2, system-specific force fields fitted to *ab initio* data achieve a higher accuracy compared to generic force fields, while maintaining a lower computational cost than quantum mechanical methods. As a result, quantum-mechanics based force fields have only been derived in some studies for a smaller set of materials.[188]

To establish a reliable representation of the COF material space, we developed the ReDD-COFFEE database in **Paper III**: a ready-to-use database including 268 687 COF geometries based on 279 SBUs and 1114 topologies and an *ab initio* derived force field for each of them. The acronym ReDD-COFFEE originates from Ready-to-use and Diverse Database of Covalent Organic Frameworks with Force field based Energy Evaluation. For each material, a system-specific force field is derived from the cluster force fields of its constituent SBUs using the method outlined in Section 2.1.2. Whereas this protocol was introduced earlier through our QuickFF procedure,[229, 230] the scope of involved materials is tremendously expanded. The computational

268 687
COF structures
and *ab initio*
force fields

MATERIALS CLOUD

Figure 4.1: The logo of the ReDD-COFFEE database and the link to its landing page on the Materials Cloud. Our database contains 268 687 COF geometries and an *ab initio* derived system-specific force field for each of them. ReDD-COFFEE is an acronym for Ready-to-use and Diverse Database of Covalent Organic Frameworks with Force field based Energy Evaluation.

cost of the required *ab initio* calculations is kept within a feasible amount by only deriving a cluster force field for the 279 SBUs instead of directly determining a periodic force field for all 268 687 COFs, which would be a daunting task. Since the materials are optimized with respect to their derived force field, the database is ready-to-use in follow-up high-throughput screenings. Our ReDD-COFFEE database is publicly available on the Materials Cloud (see Fig. 4.1).[294]

The full database generation procedure is detailed in Section 4.1. To validate that the derived system-specific force fields indeed obtain a higher accuracy than generic force fields, their ability to reproduce experimental and *ab initio* structural and vibrational characteristics is verified in Section 4.2. Lastly, in Section 4.3, the diversity of the ReDD-COFFEE database in terms of pore geometry, linker cores, linkages, and functional groups is determined and compared with other COF databases.

4.1 Generation of the ReDD-COFFEE database

All 268 687 structures of the ReDD-COFFEE database are generated following our additive top-down approach outlined in Section 3.1. The 279 SBUs constituting all materials, divided over 11 linkage types, are introduced in Section 4.1.1. To generate an initial list of (topology, SBUs) combinations, the procedure outlined in Section 4.1.2 is used. In Section 4.1.3, three filters are described that ensure only physical geometries are present in the ReDD-

COFFEE database. Our final database contains a total of 5856 2D COFs and 262 831 and 3D COFs, distributed over 95 and 1019 topologies, respectively. For each of these materials, a system-specific force field is derived, as explained in Section 4.1.4.

4.1.1 Initialization of the SBUs and topological embeddings

A total of 279 SBUs are adopted to generate the ReDD-COFFEE database. As depicted in Fig. 1.2, an SBU consists of a linker core and a linkage section, uniquely defined by the reactive group it originates from and the resulting linkage. Therefore, each SBU is labeled with three numbers, *i.e.*, XX-YY-ZZ, each indicating the linker core (coreXX), originating reactive group (reacYY), and resulting linkage (linkZZ), respectively. For example, SBU 01-02-04 is formed by combining a phenyl linker (core01) with linkage section 02-04, originating from an aldehyde reactive group (reac02) to form an imine linkage (link04). 275 SBUs are generated by combining 30 experimentally observed linker cores with eleven frequently reported linkages that originate from eleven reactive groups. The four remaining ones have no experimental precursor but emerge during COF synthesis as explained below. An overview of the employed linker cores and linkages is provided in Figs. 4.2 and 1.3, respectively. The ReDD-COFFEE database includes boronate ester (link01),[26] boroxine (link02),[26] borosilicate (link03),[295] imine (link04),[79] (acyl)hydrazone (link05),[296] azine (link06),[77] imide (link07),[297] oxazoline (link08),[298] (keto)enamine (link09),[85, 299] triazine (link10),[300] and borazine (link11)[301] linkages. Whereas most linkages are synthesized by the reaction of two precursors, some linkages, *i.e.*, boroxine, triazine, borazine, and borosilicate, emerge by combining three precursors. This largely limits the topologies these linkages can appear in, since they require each SBU to be connected with a three-connected vertex for their linkage. Although these cores have no experimental precursor, and are, therefore, technically speaking no linker cores, they are consistently treated as if they were. More specifically, their labels use a similar nomenclature, adopting an artificial reactive group (reac11), and they are also visualized in Fig. 4.2 as core31, core32, core33, and core34, respectively.

The Euclidean representations of the 2D and 3D topologies used to generate the ReDD-COFFEE database are extracted from the RCSR database[55] with a web scraping script. They are validated by computing the coordination sequence and vertex symbol of all vertex sets and automatically comparing them with the ones provided in the RCSR. A total of 2495 topologies with embedding type 1, *i.e.*, with all nearest-neighbor distances at the same distance connected by an edge, are extracted since these are most frequently

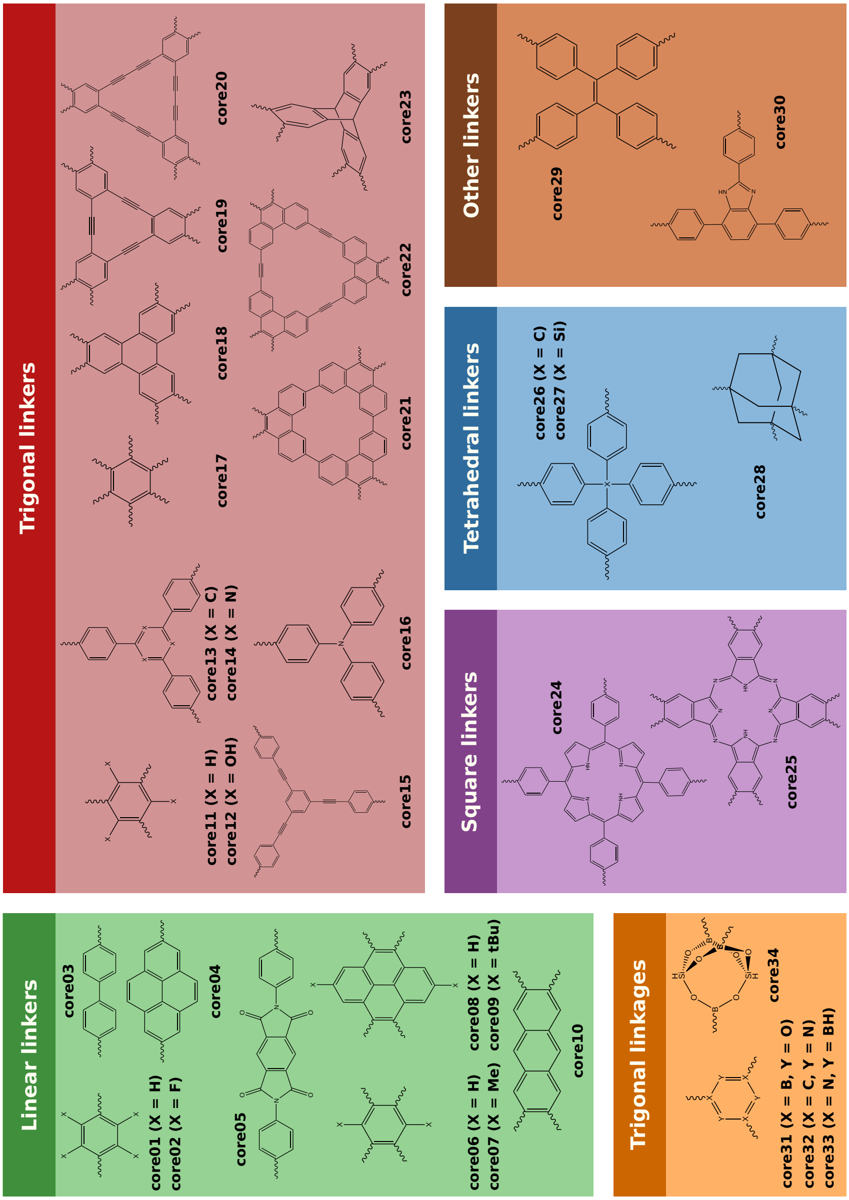


Figure 4.2: An overview of the linker cores employed to generate the ReDD-COFFEE database. They are classified in linear, trigonal, square, tetrahedral, and other linkers. Whereas the trigonal linkages that emerge during synthesis, *i.e.*, boxorine (core31), triazine (core32), borazine (core33), and borosilicate (core34), do not originate from an experimental precursor, they are nonetheless visualized here to give a concise overview. Figure reproduced from Ref. 88 with permission of the Royal Chemical Society (©2023).

observed experimentally.[302] Retaining only those with a maximum coordination number of four, the highest connectivity among the SBUs adopted in our database (see Fig. 4.2), results in 1272 topological embeddings that are used as input for the ReDD-COFFEE database.

Each (topology, SBUs) combination, and its resulting structure generated by the automated *in silico* assembly algorithm, is provided the unique label "top_SBU$_1$_SBU$_2$_..._SBU$_N$". From this name, both the name of the topology ("top") and the corresponding building blocks can be determined. SBU$_i$ is the SBU that is placed on the i-th Wyckoff set, following the order of Wyckoff sets defined in the RCSR. The label "None" defines an edge Wyckoff set that remains vacant. For example, COF-5 can be found in the ReDD-COFFEE database with the label "hcb_18-08-01_01-01-01" and COF-108 is identified as "bor_18-08-01_26-01-01_None".

4.1.2 Enumeration of trial combinations

The total amount of (topology, SBUs) combinations that can be enumerated for a specific topology is determined by the product of SBUs that can be placed on each individual Wyckoff set. Given that a topology can contain many Wyckoff sets, the total number of combinations increases rapidly. To address this, a combinatorial approach, illustrated in Fig. 4.3, is employed for every linkage type.

For each Wyckoff set, the possible SBUs are selected based on two criteria: i) the number of points of extension has to match the coordination number of the nodes and ii) the SBU needs to have a linkage section corresponding to the predefined linkage type. For edges, also the option to remain vacant is included. In the example in Fig. 4.3, the **nka** topology contains seven Wyckoff sets, *i.e.*, two with coordination number four, another two with three-fold connectivity, and three Wyckoff edges ("linkers"). We consider an illustrative SBU set that results in two options per Wyckoff set.

Without restrictions, the total number of (**nka**, SBUs) combinations would be $2^7 = 128$, symbolically visualized in the top left panel of Fig. 4.3. These can be reduced by unifying Wyckoff sets with the same coordination number, as such that these can only be occupied with the same SBU. By unifying the edge Wyckoff sets, the edges can therefore only be decorated in two different ways, instead of $2^3 = 8$ when all Wyckoff edges would be assigned individually. Similarly, when unifying the Wyckoff vertices with the same coordination number, these have to be occupied with the same SBU. This would reduce the options for both the four- and three-connected Wyckoff sets to two, instead of $2^2 = 4$.

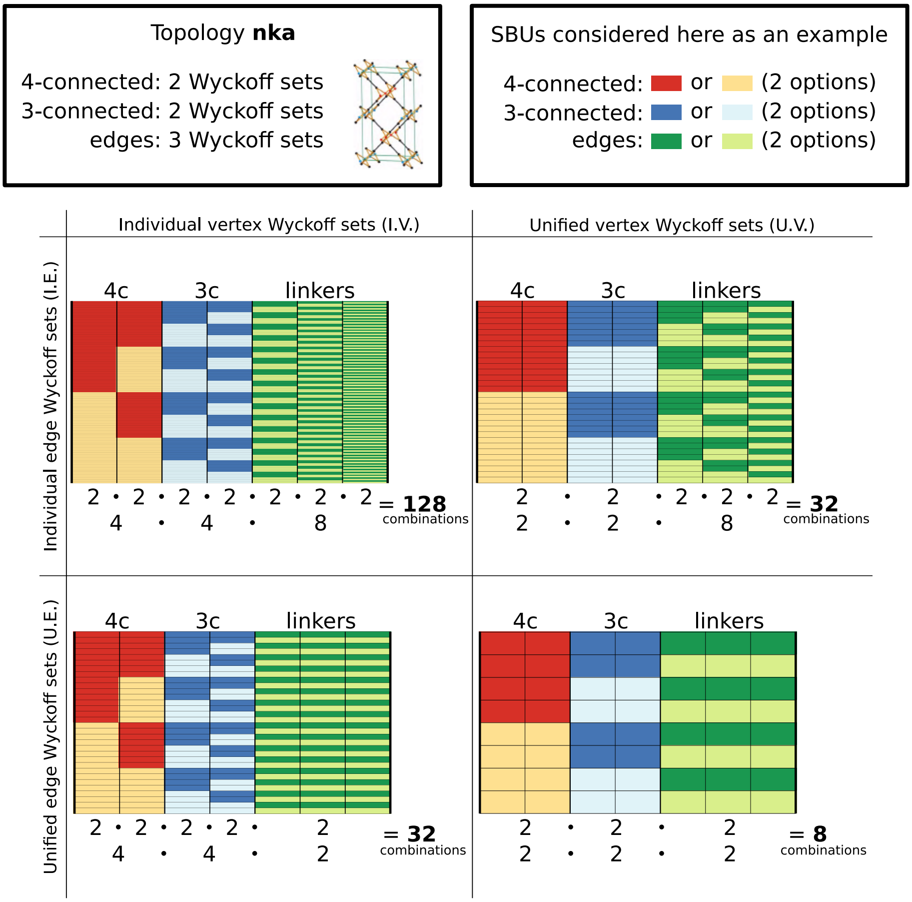


Figure 4.3: A visual illustration of the approach to enumerate trial (topology, SBUs) combinations in the **nka** topology when two options are available for each Wyckoff set. To limit the number of combinations, the Wyckoff sets with the same coordination number can be unified and allocated with the same SBU. Each of the four panels illustrate one of the combination sets. Figure reproduced from Ref. 88 with permission of the Royal Chemical Society (©2023).

In our approach, the Wyckoff edges and Wyckoff vertices with the same coordination number are either all assigned individually (I.E. and I.V., respectively), or they are unified (U.E. and U.V., respectively). To make a list of all (topology, SBUs) combinations, four combination sets are consulted: (I.V. + I.E.), (U.V. + I.E.), (I.V. + U.E.), and (U.V. + U.E.). These coincide with the four panels in Fig. 4.3. The input of our structure assembly procedure is provided by each combination set with total number of combinations lower than a certain upper limit, N_c, here chosen to be 10^4 to keep the number of structures within a feasible amount.

4.1.3 Thresholds to filter out unphysical structures

Applying the enumeration approach explained in Section 4.1.2 to our set of 279 SBUs and 1272 topologies results in a total of 5 537 951 (topology, SBUs) combinations. These are all used as input for our automated assembly algorithm, as described in Section 3.1. During this procedure, three metrics, *i.e.*, the deviation of the rescaling factors σ_f, the largest RMSD, and the deformation energy E_{def}, are calculated to check whether the structure is physical. These are adopted in three filters to identify combinations that result in unphysical structures with a low synthetic likelihood, which are discarded from the ReDD-COFFEE database. Additionally, a computational constraint is imposed to finish the optimizations within a feasible timeframe.

Since we require the topological embedding to be rescaled isotropically, the rescaling factors f_i of all edge Wyckoff sets should not differ much. If they would deviate largely, the SBUs can start to overlap or gaps between them could occur. Therefore, we demand that the standard deviation σ_f of all calculated rescaling factors remains below the threshold $\sigma_{f,\text{max}}$. To only retain those structures with almost perfectly fitting SBUs, the threshold $\sigma_{f,\text{max}}$ is chosen to be 0.22 Å/l.u. Most discarded (topology, SBUs) combinations are mixed-linker topologies with largely varying linker length. The total number of combinations that continues to Step 2 is 749 859.

The second filter is imposed on the largest allowed RMSD when introducing an SBU in the topological embedding. However, the building blocks of COFs can have an appreciable degree of flexibility and, therefore, some geometric mismatch can be tolerated. COF-108 is for example assembled by combining the HHTP and TBPM building blocks in a **bor** embedding.[61] Whereas the TBPM SBU has the perfect tetrahedral point group T_{d}, the four-connected vertices in the topological embedding possess the subgroup $D_{2\text{d}}$, resulting in a RMSD of 0.09 Å. This causes the SBUs to be slightly reshaped with respect to their relaxed equilibrium geometry when no topological constraints are enforced. To incorporate this COF, which is one of the first synthesized, but avoid structures in which the geometric mismatch is too large, the RMSD of the most favorable SBU configurations after Step 2 should be lower than the threshold RMSD_{max} = 0.11 Å. This reduces the number of allowed (topology, SBUs) combinations to 403 581.

To keep the computational resources required for the optimizations of Step 4 within a feasible amount, also a computational constraint is introduced for each generated material. Optimizations were only attempted when there are no more than 10 000 atoms in the unit cell and when the initial volume is smaller than 10 000 nm^3. Furthermore, only those structures that succesfully converged towards a minimum in the potential energy surface are included in the ReDD-COFFEE database. The total amount of optimized materials adds up to 313 909.

The third and last filter checks whether the SBUs have sufficient flexibility to accommodate the introduced mismatch in Step 2 of the automated assembly algorithm. To release the introduced strain, localized around the linkages between SBUs, the structures are optimized using their system-specific force field. However, not all SBUs have enough flexibility to find a low-energy configuration. Therefore, the deformation energy E_{def} of the optimized structure is consulted and structures with a value exceeding 14 kJ/mol are considered to be unphysical due to their largely deformed SBUs. Since these materials have a low synthetic likelihood, they are discarded from the database. After this last filter, the total number of COFs in the ReDD-COFFEE database is 268 687.

4.1.4 Derivation of the system-specific force fields

One of the unique features of the ReDD-COFFEE database is the system-specific force field that is derived for each of the materials. The periodic force field of a COF is obtained by combining the cluster force fields of its constituent SBUs, as explained in Section 2.1.2.III. Since this procedure only depends on the connection between SBUs, and not on the specific SBU configurations, the periodic force field can already be generated before the exact material configuration is known. Therefore, it can be adopted in Step 3 and Step 4 to determine the deformation energy of an SBU configuration and to relax the structure, respectively. In the following section, we demonstrate that these *ab initio* derived force fields indeed describe the atomic interaction more accurately than the generic force fields typically used in high-throughput screenings.

4.2 Investigating the accuracy of the system-specific force fields

As a first validation of the accuracy of the system-specific force fields accompanying the COFs in the ReDD-COFFEE database, we predict structural and vibrational properties with both our QuickFF force fields and the universal force field (UFF)[231] and compare them with *ab initio* and experimental data. Whereas vibrational frequencies and internal coordinates are calculated and compared with quantum mechanical reference data in Section 4.2.1 on the cluster level, Section 4.2.2 analyses the ability of the periodic force fields to predict experimental PXRD patterns and single crystal structures.

4.2.1 Comparison with *ab initio* clusters

All SBU clusters are optimized with both the QuickFF and UFF force fields, starting from the *ab initio* relaxed geometry, and the deviations of the internal coordinates are determined. As reported in Table 4.1, our cluster force field correctly predict the *ab initio* geometry, with RMSD errors on the bonds, bends, and out-of-plane distances of 4.73×10^{-3} Å, 7.18×10^{-1} °, and 4.12×10^{-2} Å, respectively. These are significantly smaller than the errors calculated for the UFF relaxed clusters, which amount to 3.56×10^{-2} Å, $2.87°$, and 4.50×10^{-2} Å, respectively. Within COFs, the most difficult internal coordinates to describe are the dihedral angles of non-planar SBUs, since these are dominated by long-range electrostatic and van der Waals interactions. Still, these internal coordinates are predicted more accurately by the QuickFF force fields than by UFF, as can be observed by the RMSD errors of $9.40°$ and $22.27°$, respectively.

Besides the internal coordinates, also the vibrational frequencies are calculated and compared with the ones derived from the *ab initio* Hessian. Also for these properties, the QuickFF force fields outperform UFF. In all frequency regions, the RMSD error is lower for QuickFF than for UFF, as can be observed in Table 4.1. While UFF is still able to get similar RMSDs in the low-frequency regime (< 500 cm^{-1}), the QuickFF force fields describe the high frequencies (> 500 cm^{-1}) an order of magnitude more accurately than the UFF ones, because the QuickFF parameters are fitted to reproduce the quantum mechanically derived Hessian.

		QuickFF		UFF	
		RMSD	MSD	RMSD	MSD
ICs	bonds [Å]	4.73×10^{-3}	1.85×10^{-3}	3.56×10^{-2}	1.80×10^{-2}
	bends [°]	7.18×10^{-1}	-6.82×10^{-3}	2.87	-1.45×10^{-2}
	dihedrals [°]	9.40	-7.94×10^{-2}	22.27	-1.43×10^{-1}
	oops [Å]	4.12×10^{-2}	-3.87×10^{-3}	4.50×10^{-2}	-6.13×10^{-4}
Frequencies	all [cm^{-1}]	18.7	-2.38	3.07×10^{2}	-1.98×10^{2}
	0-100 [cm^{-1}]	7.12	-1.95	12.5	-6.84
	100-500 [cm^{-1}]	16.7	-8.43	48.8	-38.7
	500-1000 [cm^{-1}]	14.1	-8.79×10^{-1}	1.15×10^{2}	-1.08×10^{2}
	1000-3000 [cm^{-1}]	25.5	-1.79	4.40×10^{2}	-3.68×10^{2}
	>3000 [cm^{-1}]	5.99	4.58×10^{-1}	1.61×10^{2}	-6.90

Table 4.1: Validation of the system-specific QuickFF cluster force fields. Both the internal coordinates (ICs) and the vibrational frequencies of the force field relaxed clusters are compared with those of the quantum mechanically optimized structure. The root-mean-square deviation (RMSD) and the mean signed deviation (MSD) are mentioned. Table adapted from Ref. 88 with permission of the Royal Chemical Society (©2023).

4.2.2 Prediction of experimental, structural characteristics

Whereas the comparison of internal coordinates and vibrational frequencies on the cluster level with *ab initio* data already provides a first indication of the accuracy of the QuickFF force fields, we are mostly interested in their prediction of experimentally measured, structural characteristics of the periodic structures. As mentioned in Section 3.3, the PXRD pattern is a key macroscopic descriptor of COFs, due their challenging synthesis as single crystals.[65] Similar to the workflow outlined in **Paper I**, a static and dynamic PXRD pattern are calculated for various COF structures, both with the periodic QuickFF and UFF force fields. The dynamic approach takes into account *operando* conditions and the inherent temporal character of experimental measurements.[293] Whereas the static PXRD pattern is derived from the optimized structure, the dynamic one is an ensemble average of the pattern calculated for different snapshots of an MD trajectory at operating conditions.

To illustrate the difference between both force fields, the PXRD patterns of CTF-1[303] and COF-103[61] are visualized in Fig. 4.4a and b. The experimental peaks at low values of 2θ are reproduced by all force fields. However, the peak at $27°$ and $23°$ in the patterns of CTF-1 and COF-103, respectively, is accurately predicted only by the dynamically averaged QuickFF PXRD

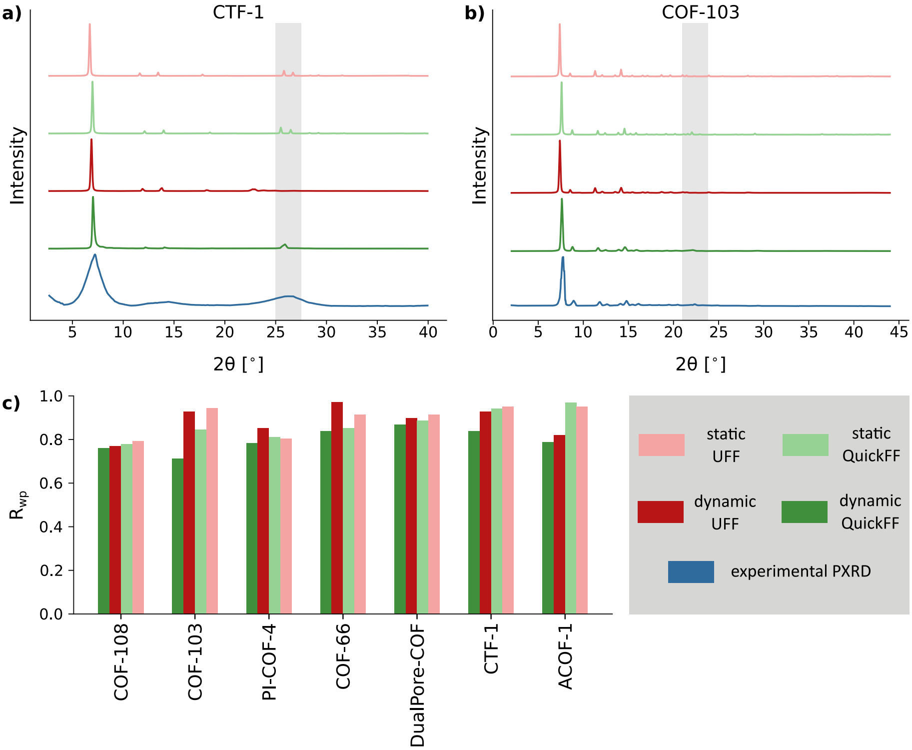


Figure 4.4: Validation of the increased accuracy of the *ab initio* derived force fields in the ReDD-COFFEE database by comparing calculated PXRD patterns with the experimental patterns of seven COFs. A total of four PXRD patterns are determined for each material, *i.e.*, the static and dynamic PXRD patterns for both the system-specific QuickFF force fields and the generic UFF force field. In the visualization of a) CTF-1 and b) COF-103, the PXRD patterns are offset vertically for clarity. The experimental peaks discussed in the text are indicated by a gray area. c) The weighted profile residual, R_{wp}, for the agreement with the experimental pattern is determined for the calculated patterns of all selected COFs. Figure adapted from Ref. 88 with permission of the Royal Chemical Society (©2023).

patterns. The dihedral angles between aromatic rings are poorly described by the UFF force field, resulting in a large unit cell and a shift of these peaks toward lower angles.

As a more quantitative descriptor to describe these observations, we adopted the weighted profile residual, R_{wp}, to compare the computed PXRD patterns of a diverse set of seven COFs, *i.e.*, CTF-1,[303] COF-103,[61] COF-108,[61] PI-COF-4,[78] COF-66,[304] DualPore-COF,[128] and ACOF-1,[305] with their experimental patterns. The weighted profile residual R_{wp} is defined as:[293]

$$R_{\mathrm{wp}} = \sqrt{\frac{\sum w_i \left(Y(\theta_i) - Y_{\mathrm{ref}}(\theta_i)\right)^2}{\sum w_i Y_{\mathrm{ref}}(\theta_i)^2}} \tag{4.1}$$

with each Bragg location θ_i having an intensity $Y(\theta_i)$ and a weight $w_i = \frac{1}{Y_{\mathrm{ref}}(\theta_i)}$. A better agreement with the experimental reference pattern, indicated by the subscript "ref", results in an R_{wp} value closer to zero. In Fig. 4.4c, the weighted profile residuals are determined for the static and dynamic PXRD patterns calculated with both QuickFF and UFF force fields for all COFs in the diverse test set. For most materials, QuickFF predicts the experimental PXRD pattern more accurately with both the static and dynamic approach. Only for PI-COF-4[78] and ACOF-1,[305] the static PXRD is predicted better by UFF. However, once the experimental conditions are taken into account by the dynamic approach, the QuickFF force fields again outperform the UFF ones.

Whereas the synthesis of single-crystal COFs proves challenging, some experimental groups have succeeded in determining their atomic structure with single crystal X-ray diffraction[80] or 3D rotation electron diffraction.[287] For these materials, the COF's structure can be determined with an even higher resolution compared to analyzing the PXRD pattern. To validate that the COF geometries are also predicted more accurately by our QuickFF force field at these higher resolutions, the crystal structures of some materials are reproduced with MD calculations at *operando* conditions. Our system-specific force fields outperform the generic UFF force fields in predicting the experimental structures of COF-300,[80] LZU-111,[80] and two distinct phases of COF-320,[287] observed at different temperatures and having a distinct pore structure and unit cell volume. Besides the bonds, for which the relative differences obtained by our QuickFF force fields are only half of those obtained by UFF, also the bends and dihedral angles are modelled at a substantially higher precision when described by our system-specific force fields. Also the unit cell lengths and unit cell volume are predicted better by the QuickFF force fields, again with the relative differences being only half the ones obtained with UFF.

4.3 Analyzing the diversity of the database

To correctly establish a representation of a material space, databases should sample all regions in it adequately and equally. On the one hand, as illustrated in Section 4.3.1, experimental databases are often biased towards cheap and easy-to-synthesize materials. On the other hand, hypothetical databases lack materials in specific regions of the material space. Such databases are described by a low variety V or a low balance B, respectively. Moreover, to maximize the information contained in a database and obtain a higher disparity D, it is better to cover strongly different subclasses that are well-separated in material space instead of including closely related subclasses. In Section 4.3.2, these diversity metrics are formally introduced and calculated for five COF databases. Since many materials in our ReDD-COFFEE database strongly resemble each other, due to the specific generation approach, in Section 4.3.3 a subset is identified with similar diversity but much fewer structures, which is ideally suited for high-throughput screenings using computationally more expensive sampling protocols.

4.3.1 Distribution of linkage types in COF databases

The COF material space is usually divided into several classes according to the formed linkage. As mentioned in Section 1.2, each of these subclasses has its unique characteristics and advantages for specific applications. In Fig. 4.5, the distributions of some frequently observed linkages of two experimental and two hypothetical databases are visualized and compared with our ReDD-COFFEE database. The linkages of a specific material are identified by scanning the material graph for linkage patterns, as explained in Section S3.2 of the SI of **Paper III**. The majority of the materials in the CoRE[140] and CURATED[141] databases are synthesized with imine and boronate ester linkages, which are indeed abundantly present in literature. However, other subclasses are less frequently detected in these experimental databases. Therefore, Martin *et al.*[153] and Mercado *et al.*[154] generated hypothetical databases with a geometric top-down approach to broaden the scope of examined COF subclasses. Both databases focused on a diverse set of linker cores but limited the number of included linkage types. The experimentally omnipresent imine linkages are present in both hypothetical databases. Martin *et al.* also included boronate ester COFs, the second most frequently observed linkage type, and to a lesser extent borosilicate COFs. In contrast, the database of Mercado *et al.* did not include boronate ester COFs, but focused on some less frequently described linkage types, *i.e.*, amide, amine, and carbon-carbon linkages. Whereas amine linked COFs are not present in any of the two experimental databases, the amide and carbon-carbon linked COFs represent only 1.79% and 1.46% of the structures in the CoRE and CURATED databases, respectively.

To generate the ReDD-COFFEE database, we chose to continue on the path of the hypothetical databases and to include a large variety of linkage types.

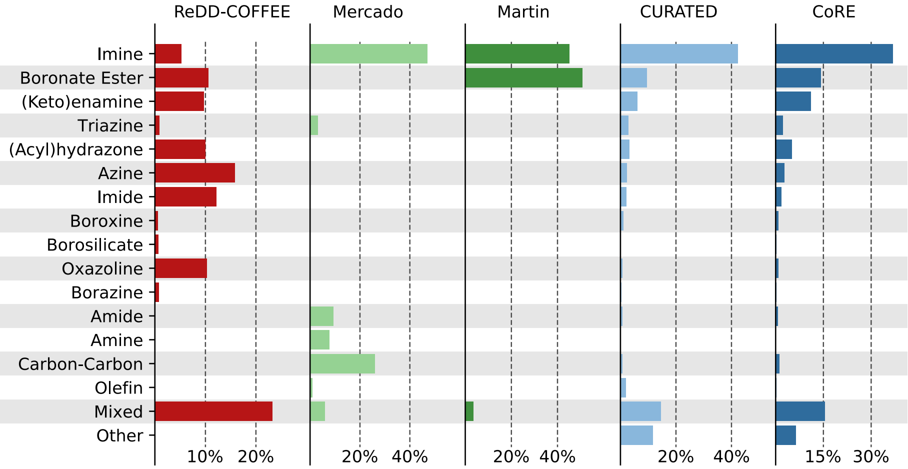


Figure 4.5: Distribution of often observed linkages in five COF databases. Whereas the experimental CoRE[140] and CURATED[141] databases are indicated in blue, the hypothetical databases of Martin *et al.*[153] and Mercado *et al.*[154] are specified in green. Our hypothetical ReDD-COFFEE database is visualized in red. Figure reproduced from Ref. 88 with permission of the Royal Chemical Society (©2023).

Besides the abundant imine and boronate ester COFs, also less frequently observed linkages, that still occupy a larger fraction of the experimental databases than amide and carbon-carbon linked COFs, are included. These subclasses are experimentally more relevant but remain until now largely unexplored. In addition, the structures are well distributed over most linkage classes since the only linkage dependent requirement in the (topology, SBUs) combination is that the linkage sections of neighboring SBUs have to match and form their predefined linkage. As mentioned in Section 4.1.1, the number of topologies available for boroxine, triazine, borazine, or borosilicate linkages is strongly limited because of the requirement of a three-connected vertex. This explains that there are relatively fewer structures assembled in these linkage subclasses compared to the other linkage types.

4.3.2 Quantitative diversity metrics

Moosavi *et al.* introduced three metrics, *i.e.*, the variety V, balance B, and disparity D, to quantitatively assess the diversity of a subset of material space for a specific domain of MOF chemistry.[198] We extended this to COF databases in **Paper III** and considered their defining chemistry to consist of four domains: i) the pore geometry, and the chemical environment of the ii) linker cores, iii) linkages, and iv) functional groups. The pore geometry can be characterized with eight structural parameters, such as the gravimetric accessible surface area, the pore fraction, and the mass density, which are calculated with Zeo++.[306] To describe the chemical environments of a COF, revised autocorrelation functions (RACs) are adopted,[197] which have been proven useful in multiple applications.[197, 198, 307–309] RACs are combinations of atomic properties for specified atom pairs, averaged over the crystal graph. Each of the three investigated chemical environments has its own set of atom pairs. These are determined by the start and scope atom lists, which focus on those atoms that are involved in the respective chemical environment (ii-iv) and are defined in Section S3.2 of the SI of **Paper III**. Once these atom lists are determined, the difference and product RACs can be computed by Eqn. 4.2 and 4.3, respectively.

$$ {}^{\text{start}}_{\text{scope}}P^{\text{diff}}_{d} = \sum_{i}^{\text{start}} \sum_{j}^{\text{scope}} (P_i - P_j)\delta(d_{ij}, d) \tag{4.2} $$

$$ {}^{\text{start}}_{\text{scope}}P^{\text{prod}}_{d} = \sum_{i}^{\text{start}} \sum_{j}^{\text{scope}} P_i P_j \delta(d_{ij}, d) \tag{4.3} $$

The atomic property P can be one of the six following: the atom identity (I), connectivity (T), nuclear charge (Z), covalent radius (S), Pauling electronegativity (χ), or polarizability (α). As indicated by the Kronecker delta δ, only atom pairs from which the number of bonds between the atoms, d_{ij}, matches the depth parameter, d, are allowed. Since there are six properties, two RAC types, and we consider a maximum depth of three bonds, a total of 48 decriptors are defined for each of the three chemical environments.

Once all descriptors for a specific domain of COF chemistry are defined, the three diversity metrics that characterize how a subset covers material space can be calculated. Here, we consider the material space as the combination of the five COF databases discussed before: the two experimental CoRE and CURATED databases, the two hypothetical databases of Martin *et al.* and Mercado *et al.*, and our ReDD-COFFEE database. To calculate the variety V and the balance B, the material space is partitioned into 1000 bins, which can be considered subclasses, with k-means clustering. The variety V is the fraction of these subclasses that are sampled by the considered material subset. The balance B checks that each bin is sampled equally by quantifying the evenness of the material distribution among the sampled subclasses. Ideally, all sampled bins contain the same number of materials. Whereas the variety adopts a clustering-approach to check the covered fraction of material space, the disparity D uses a distance-based approach. Thus, it also characterizes the spread of the bins. Together, the three diversity metrics determine the diversity from a material subset of material space for each investigated domain of COF chemistry.

The variety V, balance B, and disparity D are visualized in Fig. 4.6 for each domain and COF database. As can be observed, the hypothetical databases have a higher diversity of geometric properties than the experimental ones. This can be explained by the large freedom of *in silico* materials assembly algorithms compared to experimental techniques. Among the hypothetical databases, our ReDD-COFFEE database has the largest diversity, due to the large number of topologies that can be included by filtering the structures based on the deformation energy rather than on the RMSD. Whereas the experimental databases have a more balanced distribution of linker cores, the combinatorial freedom in hypothetical databases to select linker cores that assemble in a COF results in a higher variety. Furthermore, in contrast to the other databases, which structures mostly contain only two different SBUs, the structures of the ReDD-COFFEE database can combine multiple building blocks. While the balance of linkages is best in the CoRE database and the database of Martin *et al.*, our ReDD-COFFEE database demonstrates the highest diversity in terms of variety and disparity. The hypothetical databases of Martin *et al.* and Mercado *et al.* have the highest diversity in terms of the functional groups. Their databases are generated starting from a large set of linker cores with various functional groups, which has an even higher diversity than the experimental database. As illustrated in Fig. 4.2, we started from a limited set of linker cores, to which a small number of functional groups is attached, but focused on a large number of linkage types. However, the diversity of functional groups in our ReDD-COFFEE database can be expanded in a two-step procedure. First, the nonfunctionalized database can be screened to identify a set of materials with a high potential for the targeted application. Secondly, this material set can be easily functionalized *a posteriori* to investigate the influence of various functional groups.

4.3.3 Selection of a diverse subset

The large number of structures in the ReDD-COFFEE database might both be an advantage and a disadvantage for high-throughput screenings. Especially when the targeted property requires heavy calculations, a smaller subset with similar diversity might be very beneficial. Therefore, in **Paper III** we implemented the iterative procedure described in Ref. 198 to generate such a smaller subset, again using the features from which the diversity metrics

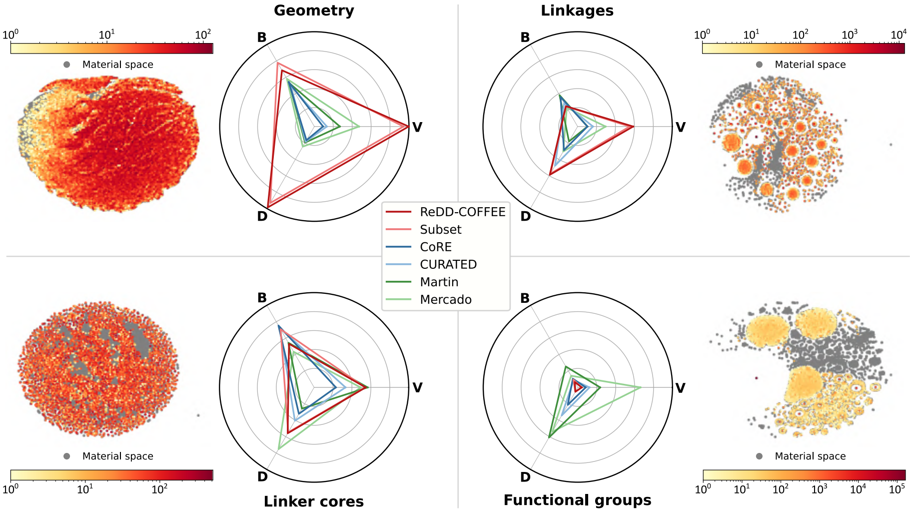


Figure 4.6: The variety V, balance B, and disparity D to describe the diversity of four domains of COF chemistry in five COF databases. Also the diversity metrics of a subset of 10 000 COFs from the ReDD-COFFEE database are calculated. For each domain, a t-SNE plot of the COF material space, *i.e.*, the union of the databases, is depicted in gray, overlaid by the density of the ReDD-COFFEE materials. Figure reproduced from Ref. 88 with permission of the Royal Chemical Society (©2023).

were derived, *i.e.*, geometric properties and RACs to describe the chemical environment of the linker cores, linkages, and functional groups. In an iterative procedure, the set of selected structures is gradually filled. The initial material in this set is randomly chosen. In each further step, the structure that has the largest minimal distance to the set of already selected structures is added. As the size of the selected set increases, the variety V and disparity D continue to grow. However, after a sharp peak of the balance B when the empty bins are filled, it will start to gradually drop when further structures are added. In Fig. 4.6, these diversity metrics are derived for a subset of 10 000 structures. Whereas the occupied region in material space is only slightly smaller, resulting in a comparable variety V and disparity D, the balance B is higher than the one of the complete ReDD-COFFEE database. In the complete ReDD-COFFEE database, many materials explore analogous regions in material space, and the information gained from one structure can usually be learned from the other structures. For computationally expensive high-throughput screenings that involve costly calculations, such as the ones in Sections 5.2 and 5.3, utilizing such subset can effectively lower computational expenses, while maintaining an accuracy comparable to screening the full database.

Concluding this chapter, we have generated the ReDD-COFFEE database of 268 687 COF structures and accompanying *ab initio* force fields. By comparing the diversity of our database with that of other COF databases, we could quantify that the ReDD-COFFEE has an excellent diversity in terms of pore geometry, linkers, and linkages, but includes only limited functional groups. The force field of each material is derived from quantum mechanical reference data of its constituent SBUs. The increased accuracy of these force fields is demonstrated by reproducing *ab initio* and experimental structural and vibrational characteristics.

5

High-throughput screenings of covalent organic frameworks

As discussed in Section 1.4, computational high-throughput screenings have three large advantages.[155] All three of them are explored by high-throughput screenings performed on our ReDD-COFFEE database generated in Chapter 4. Firstly, property-property relations can be established that are not observable with only a limited dataset. In **Paper III** and Section 5.1, such property-property relations are determined for the textural properties of COFs. Furthermore, these are compared with other classes of nanoporous materials, *i.e.*, MOFs and zeolites. Secondly, the performance limits of a material class can be demonstrated. The potential application of COFs for vehicular methane storage is investigated and their ability to meet the ARPA-E targets[117] is examined in Section 5.2 and **Paper III**. Lastly, promising materials for a targeted application can be identified or design rules to synthesize them can be proposed. This advantage is illustrated in **Paper V**, in which potential candidates for post-combustion carbon capture are determined and design rules are formulated to guide experimental researchers to synthesize COFs with exceptional performance for CO_2 capture. These results are summarized in Section 5.3.

Whereas the calculation of textural properties in the first screening study is very cheap, allowing them to be calculated for all materials in the database, the grand-canonical Monte Carlo (GCMC) simulations to determine the adsorption properties of the two following high-throughput screenings can become computationally expensive, especially for large-pore materials that fit many guest molecules. To reduce the computational cost, while maintaining a similar accuracy, the methane properties are derived for a diverse subset of 10 000 COFs, as mentioned in Section 4.3.3, in the second screening. In the third high-throughput screening, a similar approach has been followed to compute the carbon capture performance of a varied subset of 15 000 COFs. To characterize the complete ReDD-COFFEE database, a machine learning workflow has been adopted to identify the most promising materials within this full database of 268 687 materials.

5.1 Textural features of nanoporous materials and COF property-property relations

Our first high-throughput screening investigates the textural properties of the COFs in our ReDD-COFFEE database, determines property-property relations between them, and compares these with other classes of nanoporous materials. Since textural properties are relatively cheap to calculate, all 268 687 materials in the database could be characterized within a reasonable timeframe. These properties can provide an initial understanding of the COF material class and are often applied in high-throughput screenings as a first filter to reject materials that do not meet the target design criteria.[160, 169, 177, 181]

The computed properties involve the gravimetric and volumetric accessible surface areas, the gravimetric accessible volume, and the pore fraction. The accessible surface area of a unit cell, A_acc, is the area described by the center of a probe with a given radius that revolves the framework atoms. The derived quantities, *i.e.*, the gravimetric accessible surface area, A_grav, and the volumetric accessible surface area, A_vol, can be derived by normalizing with the mass, M, and volume, V, of such unit cell, respectively. Similarly, the accessible volume, V_acc, is the unit cell volume that can fit the center of such a spherical probe. The derived quantities are the gravimetric accessible volume, V_grav, and the volumetric accessible volume or pore fraction, V_vol. All derived quantities are formally defined in Eqn. 5.1 and 5.2.

$$A_\text{grav} = \frac{A_\text{acc}}{M} \qquad A_\text{vol} = \frac{A_\text{acc}}{V} \tag{5.1}$$

$$V_\text{grav} = \frac{V_\text{acc}}{M} \qquad V_\text{vol} = \frac{V_\text{acc}}{V} \tag{5.2}$$

Also the mass density and the diameters of the largest included sphere, free sphere, and included sphere along the free path are calculated. All textural properties are derived with Zeo++.[306] A probe radius of 1.82 Å, equal to the kinetic radius of nitrogen gas,[310] is adopted to compute the accessible surface area and volume.

Property-property relations between the textural properties of COFs are visualized in Fig. 5.1. As a result of their porous nature and their composition of organic, lightweight SBUs, COFs possess a very low mass density and exceptionally high gravimetric properties. As demonstrated by Fig. 5.1c, 3D COFs have a lower mass density than the more densely packed 2D COFs. Whereas the majority of the 3D COFs in our database possess a mass density lower than 200 kg/m^3, that of 2D COFs is mostly between 200 and 600 kg/m^3. The gravimetric accessible surface areas of 2D and 3D COFs reach values in the range of 1750 to 3000 m^2/g and 6000 to 10 000 m^2/g, respectively. On the one hand, as evidenced by Fig. 5.1a and b, the heaviest COFs have only very narrow pores that are too small to accept guest molecules to enter the material and, therefore, possess no accessible surface area or volume. On the other hand, the lightest COFs include the largest pores. Since their unit cell increases rapidly, while the accessible surface area only grows steadily, the volumetric accessible surface area of these materials, *i.e.*, the ratio of these two properties, also drops to zero, as can be observed in Fig. 5.1a and b. COFs with an optimal balance between accessible surface area and pore volume can be found in between these two regions of materials with a very low or very high mass density. These COFs achieve large pore volumes, coupled with high volumetric and gravimetric accessible surface areas.

Fig. 5.1d demonstrates that the pore diameter and, related, the mass density are controlled not only by the topology and the SBUs, but also by the linkage type. Linkages with a small spatial extent decrease the space between the linker cores. These include boronate ester and imide COFs, but also linkage types that emerge during COF synthesis, such as boroxine and triazine, since the linker cores of these SBUs are only separated by a single bond. Therefore, they result in materials with a lower pore diameter and higher mass density as opposed to COFs with linkages with a larger spatial extent, such as azine or (acyl)hydrazone. The property-property relations visualized in Fig. 5.1 illustrate that the porosity of COFs can be tuned by a careful selection of the topology, SBUs, and linkages.

Besides property-property relations between COFs, we have also compared them with other classes of nanoporous materials. Fig. 5.2 visualizes the property-property relations of the COFs in our ReDD-COFFEE database together with the database of zeolite structures of the International Zeolite Association (IZA)[47] and three MOF databases, *i.e.*, the hypothetical ToBaCCo[151] and hMOF[144] databases and the experimental QMOF database.[139] For each material class, the same qualitative relation is observed. As determined in Fig. 5.2a and b, nanoporous materials with a pore fraction around 0.4

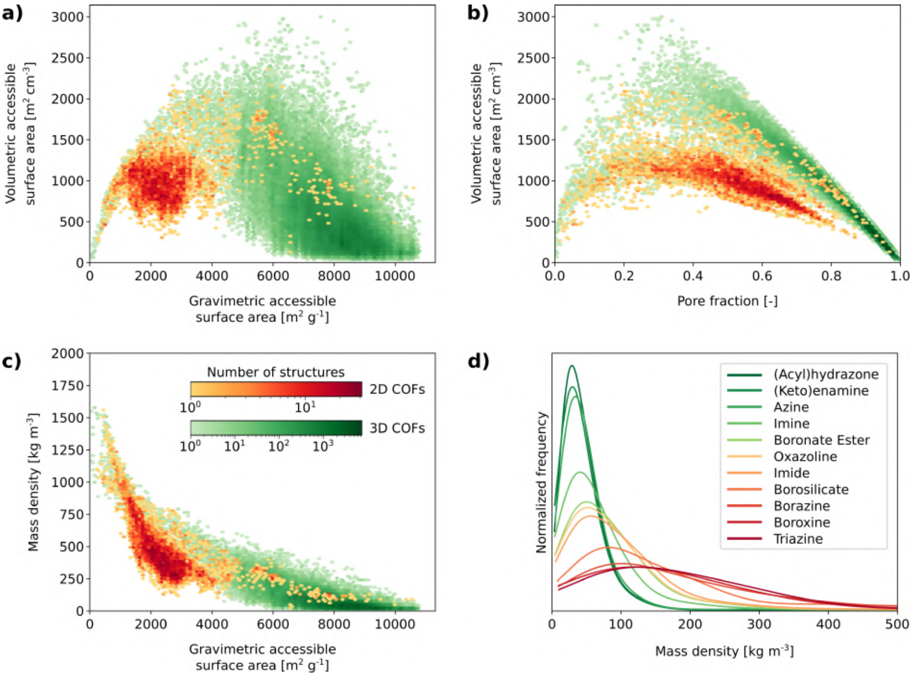


Figure 5.1: a)-c) Property-property relationships between the textural properties of the materials in our ReDD-COFFEE database. Relation between a) the gravimetric accessible surface area and b) the pore fraction and the volumetric accessible surface area. c) The mass density as a function of the gravimetric accessible surface area. d) Histograms of the mass density for each subclass of linkage types. Figure reproduced from Ref. 88 with permission of the Royal Chemical Society (©2023).

and a gravimetric accessible surface area of 4000 m^2/g result in the largest volumetric accessible surface area of almost 2500 m^2/cm^3. Their mass density is around 500 kg/m^3, as illustrated in Fig. 5.2c. For all classes of nanoporous materials, a higher pore fraction leads to materials with a larger gravimetric accessible surface area. Among the three material classes, COFs have the lowest mass density and the highest pore fraction, with the structures that approach a pore fraction of 1.0 reaching gravimetric accessible surface areas up to 9000 m^2/g. Whereas Fig. 5.1 demonstrates that COFs have a broad range of textural properties, Fig. 5.2 emphasizes their large porosities and ultrahigh gravimetric accessible surface areas and pore volumes when compared to MOFs and zeolites. This unique combination of properties makes them of particular interest for adsorption applications.

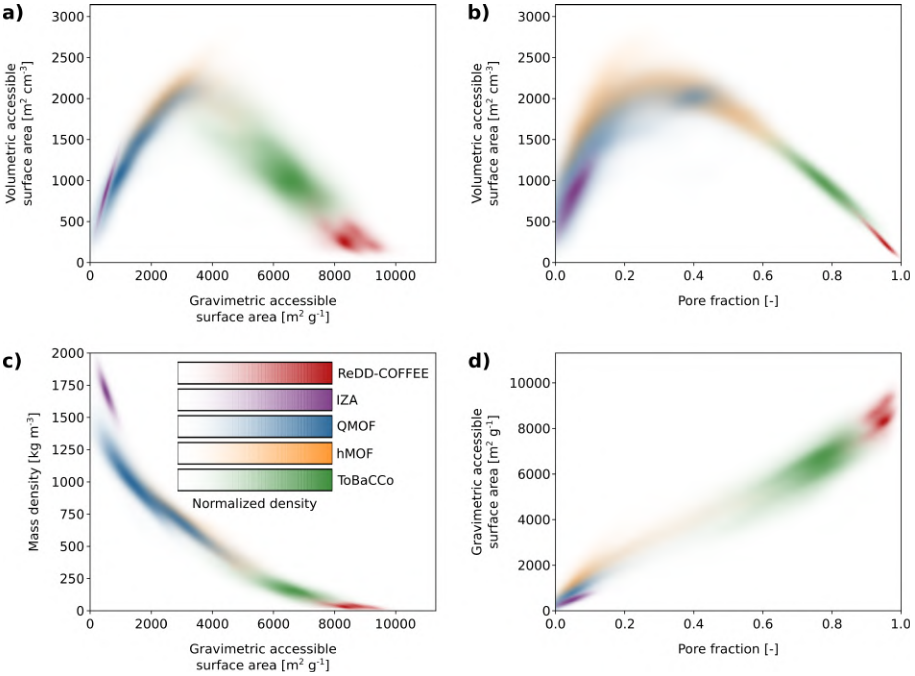


Figure 5.2: Comparison of the textural properties of the COFs in our database and other nanoporous materials, *i.e.*, MOFs and zeolites. In a)-c), the same property-property relations as Fig. 5.1 are provided. d) The relation between the pore fraction and the gravimetric accessible surface area. Figure reproduced from Ref. 88 with permission of the Royal Chemical Society (©2023).

5.2 Determining the performance limits of COFs for vehicular methane storage

As explained in Section 1.2.2, COFs offer high promise for the vehicular storage of natural gas, mainly consisting of methane, due to their high internal surface area and low mass density.[109, 110] The MOVE program of the US Department of Energy imposed two targets on the deliverable capacity, defined in Eq. 5.3, to be commercially interesting. A candidate nanoporous material should have a gravimetric deliverable capacity larger than 0.5 g/g and a volumetric deliverable capacity of 315 vSTP/v.[117] They proposed to measure the deliverable capacity between a storage pressure of 65 bar and a depletion pressure of 5.8 bar.[117, 160]

$$\Delta N_{\mathrm{CH_4}} = N_{\mathrm{CH_4}}^{\mathrm{ads}} - N_{\mathrm{CH_4}}^{\mathrm{des}} \tag{5.3}$$

To determine the performance limits of COFs for vehicular methane storage, the deliverable capacity for a diverse subset of 10 000 materials of the ReDD-COFFEE database is calculated with GCMC simulations. The subset is selected following the approach explained in Section 4.3.3. A benchmark study to reproduce the experimental methane isotherms of four COFs[67] shows that combining the DREIDING force field[240] for the host-guest interactions, the TraPPE united atom model[311] for the guest-guest interactions, and the MBIS charges[234] for the electrostatic contributions provides the best accuracy while limiting the computational cost. To reduce the computational resources even more, both the framework and the guest geometries are fixed due to the rigidity of their covalent bonds.

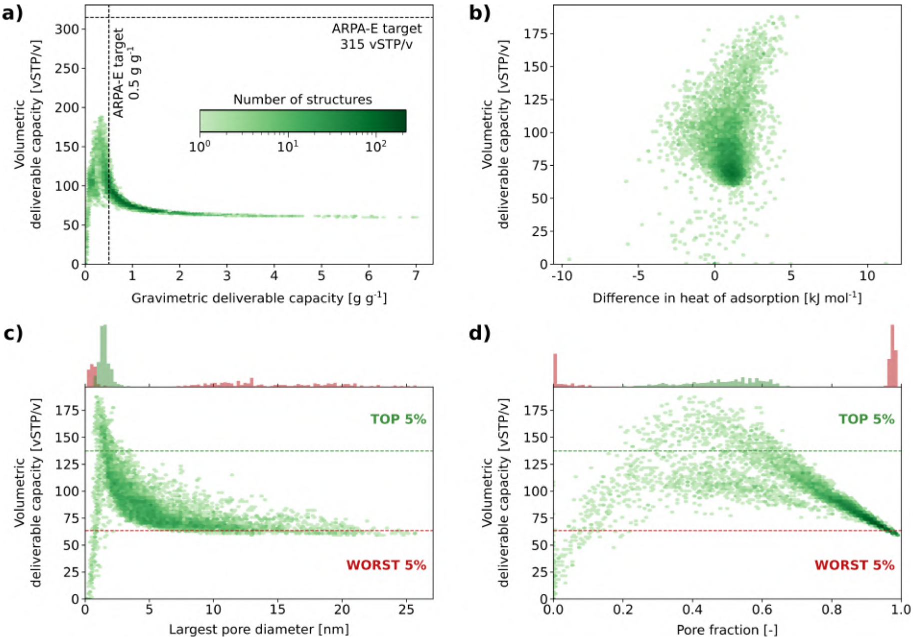


Figure 5.3: Property-property relationships between adsorption properties for the vehicular methane storage of a diverse subset of 10 000 COFs of the ReDD-COFFEE database. a) Comparison between the gravimetric and volumetric deliverable capacity. Also the ARPA-E targets are indicated. b) The relation between the volumetric deliverable capacity and the difference in heat of adsorption at storage and depletion pressure. c)-d) The volumetric deliverable capacity as a function of the largest pore diameter and the pore fraction. The histograms on top of the plots establish the range of values in which the 5% best and worst materials in terms of the volumetric deliverable capacity can be found. Figure reproduced from Ref. 88 with permission of the Royal Chemical Society (©2023).

Fig. 5.3a illustrates there is a trade-off between the gravimetric and volumetric deliverable capacity. The COFs with the lowest mass density reach the highest gravimetric deliverable capacity, up to 7.0 g/g. However, these also comprise the largest volumes, which result in lower volumetric deliverable capacities of around 60 vSTP/v. Fig. 5.3c and d visualize the property-property relationships between the volumetric deliverable capacity and the largest pore diameter and the pore fraction, respectively, together with histograms of those materials with the 5% highest or lowest volumetric deliverable capacity, as a function of the largest pore diameter and pore fraction. These demonstrate that the COFs with the largest volumetric deliverable capacity exhibit a pore fraction between 0.2 and 0.7 and a pore diameter between 0.7 nm and 3.4 nm. The materials with the highest volumetric deliverable capacity furthermore possess a large difference between the isosteric heat of adsorption at the storage and depletion pressures, as evidenced in Fig. 5.3b. In these COFs, the methane-methane interactions are more dominant than the methane-framework interactions. This ensures that the methane molecules are only weakly confined at the low depletion pressure and, therefore, increases the amount of gas that could be delivered.

As observed in Fig. 5.3a, many COFs meet the ARPA-E threshold of 0.5 g/g for the gravimetric deliverable capacity. However, there is no material that reaches the target of 315 vSTP/v for the volumetric deliverable capacity. This does not imply that COFs cannot compete with other nanoporous materials for vehicular methane storage. In fact, the material in our subset with the largest volumetric deliverable capacity of 187.4 vSTP/v, *i.e.*, the boronate ester COF ths-c3_11-01-01_06-08-01_06-08-01, performs comparable to current state-of-the-art materials, such as MOF-5 (182 vSTP/v),[312] HKUST-1 (183 vSTP/v),[313] Co(bdp) (197 vSTP/v),[314] or NJU-Bai 43 (198 vSTP/v).[315] Previous research has demonstrated that physical constraints prevent the ARPA-E targets from being reached, even by these record-holding materials.[160, 316] Despite that functionalization or anchoring of alkali metals to the COF[317] can increase the volumetric deliverable capacity, our high-throuhgput screening indicates that currently the performance limits of COFs are insufficient for the commercial development of the adsorbed natural gas (ANG) technology, similar to other classes of nanoporous materials.

5.3 Identifying promising COF candidates for carbon capture

As a last clean energy application, COFs are investigated for their adoption as nanoporous adsorbents in carbon capture, as introduced in Section

1.2.1.[68, 106–108] High-potential candidates for post-combustion carbon capture using a pressure swing adsorption process are identified and design rules are proposed to guide experimental researchers towards COFs with high CO_2 capture performance. Furthermore, we demonstrate that machine learning (ML) algorithms can significantly accelerate our high-throughput screening by identifying a set of promising materials. In post-combustion carbon capture, CO_2 is isolated from the flue gases of industrial power plants, which is a mixture of mainly N_2 and CO_2 molecules. The three key advantages of this technology are the easy installation on existing infrastructures, the stationary nature of these point sources, and the relative high concentration of CO_2 molecules in flue gases.[318]

In a typical carbon capture process, CO_2 molecules are captured at adsorption conditions and regenerated at desorption conditions. The specific adsorption and desorption conditions are determined by the type of process. In the pressure swing adsorption approach considered here, the temperature is kept fixed at 298 K and the pressure varies from 10 bar to 0.1 bar during adsorption and desorption, respectively. Similar to the deliverable capacity of methane, introduced in Section 5.2, the CO_2 working capacity, $\Delta N_{\mathrm{CO_2}}$, is the amount of molecules that can be captured every cycle, as defined in Eq. 5.4:

$$\Delta N_{\mathrm{CO_2}} = N_{\mathrm{CO_2}}^{\mathrm{ads}} - N_{\mathrm{CO_2}}^{\mathrm{des}} \tag{5.4}$$

However, in contrast to the case of vehicular methane storage, the flue gas of industrial power plants consists of a mixture of N_2 and CO_2. Therefore, the perfect adsorbent material should not only have a high working capacity, but also favor the adsorption of CO_2 molecules over N_2 molecules, *i.e.*, they should have a high CO_2/N_2 selectivity. Depending on the adopted calculations, two types of selectivity can be defined. As defined in Eq. 5.5, the ideal selectivity, S_{ideal}, can be derived from the CO_2 and N_2 uptakes as calculated with single-component GCMC simulations:

$$S_{\mathrm{ideal}} = \frac{N_{\mathrm{CO_2}}^{\mathrm{ads}}}{N_{\mathrm{N_2}}^{\mathrm{ads}}} \qquad \text{(single-component GCMC)} \tag{5.5}$$

When mixture GCMC calculations are used, mimicking the flue gas as a binary inlet gas with mole fraction $y_{\mathrm{CO_2}}$ and $y_{\mathrm{N_2}}$, the more accurate mixture selectivity, S_{mix}, can be determined by Eq. 5.6. Since this takes into account the interactions between the two gas components, this metric also characterizes competitive adsorption and coadsorption.

$$S_{\mathrm{mix}} = \frac{N_{\mathrm{CO_2}}^{\mathrm{ads}}/y_{\mathrm{CO_2}}}{N_{\mathrm{N_2}}^{\mathrm{ads}}/y_{\mathrm{N_2}}} \qquad \text{(mixture GCMC)} \tag{5.6}$$

Designing the ideal COF for CO_2/N_2 separation is a trade-off between maximizing the working capacity and selectivity of a material.[169] The working capacity is highest in materials with large pores, which increases the gas uptake at adsorption conditions. However, frameworks with relatively small pores demonstrate a higher selectivity due to strengthened interaction with the pore walls. To allow for a robust and effective ranking of the top candidates, the adsorbent performance score (APS), defined in Eq. 5.7, combines these two adsorption properties into a single parameter.[319]

$$\text{APS}_x = \Delta N_{\text{CO}_2} S_x \tag{5.7}$$

The subscript x indicates the type of the employed GCMC calculations and can either be “ideal” or “mix”. In all GCMC simulations, the host-guest and guest-guest interactions are described with the MM3[232] and UFF[231] force fields, respectively, whereas the partial charges for the electrostatic contribution are derived with MBIS.[234] This level of theory agrees best with the experimental isotherms[67] and *ab initio* Henry coefficients,[320] as determined in a benchmark study. Similar to the GCMC calculations in Section 5.2, the framework and the guests are considered rigid.

The procedure proposed in this high-throughput screening is depicted in Fig. 5.4. In the first step, detailed in Section 5.3.1, the CO_2 working capacity and ideal CO_2/N_2 selectivity are calculated with single-component GCMC calculations for a diverse subset of 15 000 COFs from the ReDD-COFFEE database. Whereas the ideal selectivity provides a poor quantitative prediction of the mixture selectivity, the ranking of these two metrics is strongly related. As such, the ideal selectivity can serve as a first qualitative prediction of the performance of a set of materials. Since this subset is selected using the method introduced in Section 4.3.3, its COFs are representative for the complete database and the working capacity and ideal selectivity of all 268 687 COFs can be easily learned from its results. This is exactly what is done by a ML model in the second step of our high-throughput screening, as explained in Section 5.3.2. With these derived results, the most promising COFs are selected and passed on to the third and last step of our screening. In Section 5.3.3, these COF candidates are characterized with mixture GCMC calculations to include the effects of competitive adsorption and coadsorption. Finally, the materials are ranked based on APS_{mix} to identify the COFs with the highest potential for post-combustion carbon capture. In Section 5.3.4, the top-performing COFs are analyzed to establish design rules to guide experimentalist researchers.

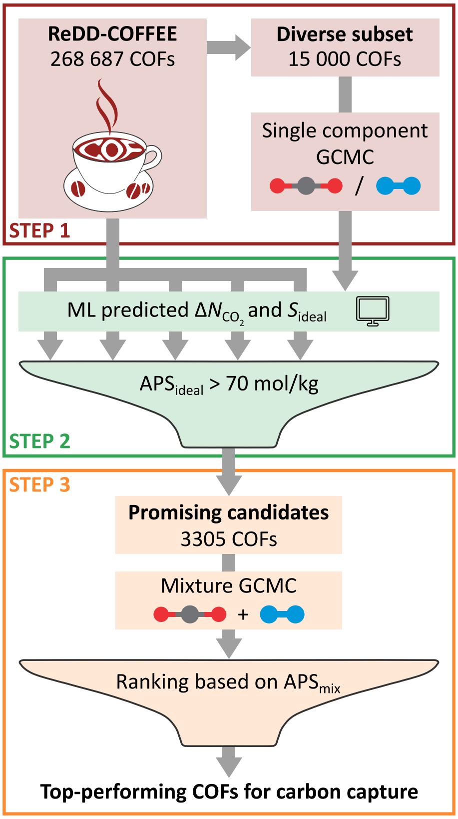


Figure 5.4: Schematic overview of the multistep procedure followed in our high-throughput screening to identify top-performing COF candidates for carbon capture. Figure reproduced from Ref. 321 with permission of the American Chemical Society (©2024).

5.3.1 Step 1: screening of the diverse subset

In the first step of our high-throughput screening, single-component GCMC calculations are performed on 15 000 COFs to determine the CO_2 working capacity, ΔN_{CO_2}, and the ideal CO_2/N_2 selectivity, S_{ideal}. These results are summarized in Fig. 5.5. The trade-off between the working capacity and the selectivity is visualized in Fig. 5.5a. On the one hand, COFs can achieve a very high CO_2 working capacity, up to 123.3 mol/kg. However, as mentioned before, this is observed for materials with very large pores ($>$ 10 nm) that have a low selectivity.

On the other hand, COFs with very small pores, *i.e.*, between 0.35 nm and 0.4 nm, have the highest selectivity, with values up to 53.5, but a very low working capacity. As illustrated by Fig. 5.5b, CO_2 molecules bind more tightly to these materials than N_2 molecules. This is evidenced by the large difference between the heat of adsorption of these guests for the materials with the highest ideal selectivity. Whereas the pores of these COFs can perfectly accommodate CO_2 molecules, which have a kinetic diameter of 0.33 nm, they have more difficulties with storing N_2, due to their larger kinetic diameter of 0.364 nm. As displayed in Fig. 5.5a, the materials with the highest APS_ideal have a pore diameter around 1.0 nm. These materials do not have the largest working capacities or selectivities, which are observed for the very large and very small pore materials, respectively, but achieve relatively high values of both properties simultaneously. The material with the highest APS_ideal of 527.8 mol/kg is the (acyl)hydrazone COF twt-b_11-04-05_11-02-05_None_None, which has a working capacity of 35.9 mol/kg and an ideal selectivity of 14.7. Whereas its ideal selectivity can be compared with that of conventional adsorbents such as zeolites 13X ($\Delta N_{\text{CO}_2} = 2.3$ mol/kg, $S_\text{ideal} = 17.45$)[322] and 5A ($\Delta N_{\text{CO}_2} = 4.18$ mol/kg, $S_\text{ideal} = 46.5$),[323] its CO_2 working capacity is much higher.

In Fig. 5.5c and d, the dependency of the carbon capture performance on the linkage types of COFs is studied. To this end, the characterized subset is divided in different classes, according to the linkage type that connects the SBUs. Fig. 5.5c demonstrates that COFs with an (acyl)hydrazone, imide, or (keto)enamine linkage (different shades of green in Fig. 5.5c and d) typically exhibit higher APS_ideal values. The distributions of APS_ideal for these classes have medians of 33.1 mol/kg, 31.8 mol/kg, and 30.6 mol/kg, respectively, significantly higher than those of other linkage classes. This can also be concluded from Fig. 5.5d, which depicts the fraction of each linkage class in a subset generated by only selecting those structures that have an APS_ideal exceeding the threshold indicated on the x-axis. The subset belonging to the lowest possible APS_ideal threshold of 0 mol/kg coincides with the full material set of 15 000 COFs. When systematically increasing this threshold, only well-performing materials are selected for the subset. The best-performing materials, *i.e.*, those with a high APS_ideal value, mostly contain an imide linkage. This linkage is observed in 69.4% of the COFs with an APS_ideal above 200 mol/kg. Also (acyl)hydrazone and ketoenamine linkages are frequently present in this subset, with a linkage share of 16.7% and 5.6%, respectively.

5.3.2 Step 2: machine learning prediction of the working capacity and selectivity

Since the characterized subset is selected to have a similar diversity as the ReDD-COFFEE database, the CO_2 working capacity and ideal CO_2/N_2 selectivity for all 268 687 materials can be derived using a ML model, as

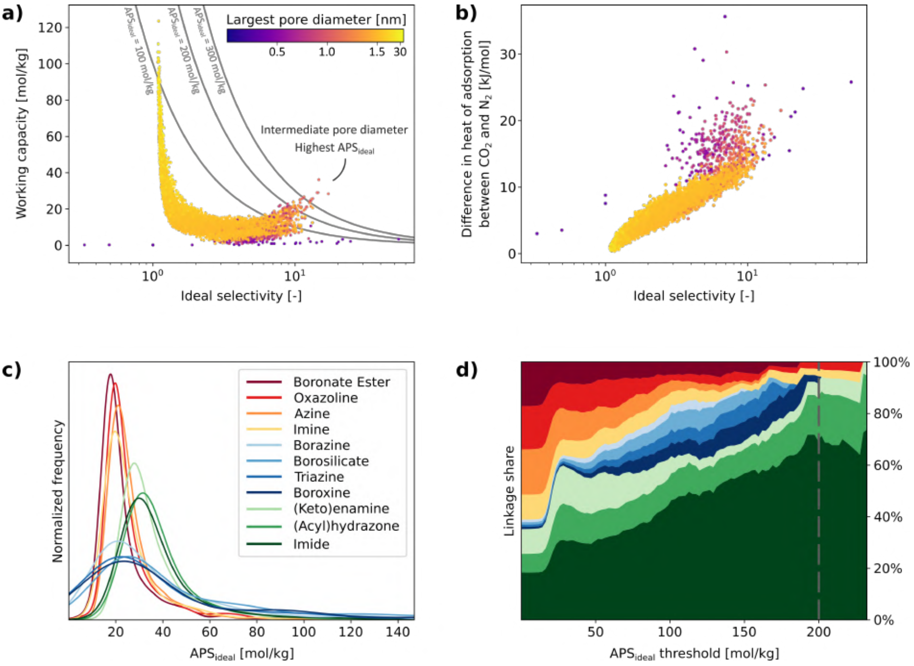


Figure 5.5: Analysis of the first step of our high-throughput screening in which the working capacity and ideal selectivity of a diverse subset of 15 000 COFs is determined. The relation between the ideal selectivity and a) the working capacity and b) the difference in heat of adsorption of CO_2 and N_2 are visualized, together with the largest pore diameter indicated by the color code. c) Histograms of APS_{ideal} for each linkage subclass. d) The share of each linkage class in a varying subset determined by all COFs with an APS_{ideal} exceeding the threshold defined on the x-axis. The color code in c) and d) are the same. Figure adapted from Ref. 321 with permission of the American Chemical Society (©2024).

is done in this second step. Five different ML algorithms are adopted that have previously succeeded in accurately predicting adsorption properties in nanoporous materials, *i.e.*, kernel ridge regression (KRR), support vector regression (SVR), multi-layer perceptron regression (MLP), random forest regression (RF), and gradient boosting regression (GB).[198, 207, 324] Each of these models is trained to predict five adsorption targets, *i.e.*, the CO_2 uptake at adsorption conditions, the CO_2 uptake at desorption conditions, the N_2 uptake at adsorption conditions, the CO_2 working capacity, and the ideal CO_2/N_2 selectivity. Besides the direct calculation of the working capacity and selectivity, they can also be derived from machine learning models trained on the individual guest uptakes. Similar to the calculation of the diversity

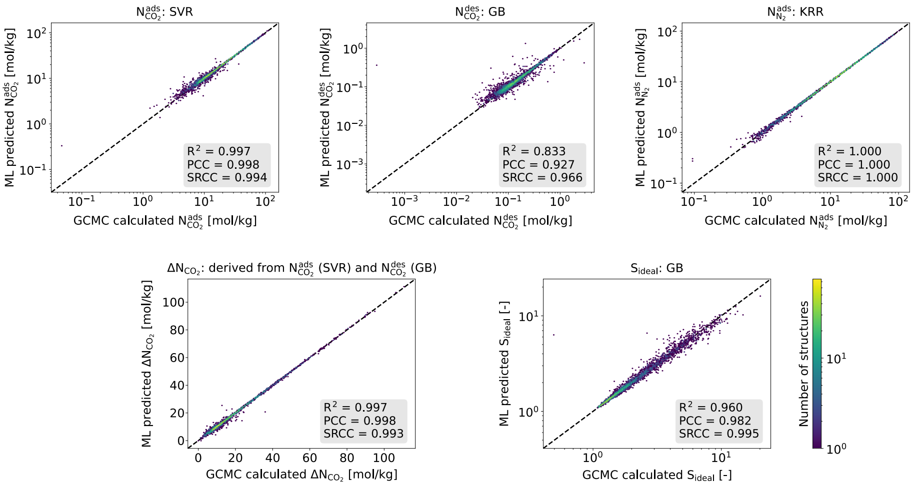


Figure 5.6: Comparison between the most accurate ML predictions of the five adsorption properties and their actual values as determined by GCMC calculations for the 5000 COFs in the test set. The best model for each target is indicated above the respective plot. Furthermore, the coefficient of determination (R^2), Pearson correlation coefficient (PCC), and Spearman rank correlation coefficient (SRCC) are reported in the insets. Figure reproduced from Ref. 321 with permission of the American Chemical Society (©2024).

metrics in Section 4.3.2, each COF is featurized using textural parameters and RACs to describe the chemical environment of the linker cores, linkages, and functional groups. To benchmark the ML models, the diverse subset is divided in a train and test set of 10 000 and 5000 materials, respectively.

The most accurate ML models and their test set predictions are visualized in Fig. 5.6. The insets report the coefficients of determination (R^2) and the Pearson and Spearman rank correlation coefficients (PCC and SRCC, respectively). These metrics indicate that among the individual uptakes, the CO_2 uptake at desorption conditions is most difficult to predict, followed by the CO_2 uptake at adsorption conditions. This can be explained since the chemical environment of COFs is challenging to characterize.[195] At the desorption pressure of 0.1 bar, this chemistry is more important than at the adsorption pressure of 10 bar, where the guest uptake is highly correlated with the textural features. The framework chemistry is also more crucial to determine the CO_2 uptakes than the N_2 adsorption, due to the higher quadrupolar moment of CO_2.

We observe that the ML algorithms directly trained on the working capacity are mainly fitted to the CO_2 uptake at adsorption conditions, which is two orders of magnitude larger than that at desorption condition. Since the

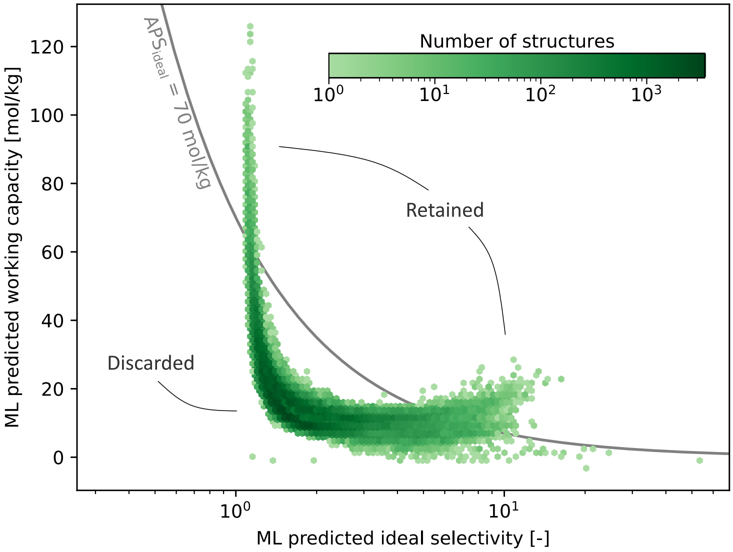


Figure 5.7: 2D histogram of the ML predicted CO_2 working capacity and ideal CO_2/N_2 selectivity for all 268 687 COFs in the ReDD-COFFEE database. The threshold of APS_{ideal} = 70 mol/kg determines the promising materials that are retained for further screening in Section 5.3.3. Figure reproduced from Ref. 321 with permission of the American Chemical Society (©2024).

small variations of the desorption CO_2 uptake can be better described by a ML model specifically trained at desorption conditions, the most accurate predictions of the working capacity are obtained for the model derived from the individual uptakes of CO_2 at adsorption (SVR) and desorption conditions (GB). On the contrary, the selectivity is largely determined by the synergy of the CO_2 and N_2 uptakes, causing the most accurate model to be a directly trained ML algorithm (GB). Since small deviations of the N_2 adsorption have a large influence on this property and it heavily depends on the chemistry of the framework,[325] the selectivity is more difficult to learn by a ML model than the working capacity, as illustrated by the smaller R^2 value and Pearson correlation coefficient of 0.960 and 0.982, respectively. However, since the model correctly describes the order of the selectivities, as verified by the large Spearman rank correlation coefficient of 0.995, our goal of selecting a set of high potential COFs for further screening is still reached.

Fig. 5.7 provides an updated version of Fig. 5.5a, with the ML predicted working capacity and ideal selectivity of all 268 687 COFs in the ReDD-COFFEE database. A promising set of 3305 materials is selected by introducing a threshold of 70 mol/kg on the APS_{ideal}. These are retained in Step 3 for further screening.

5.3.3 Step 3: mixture selectivity of the most promising COFs

In the third step of our screening approach, the set of 3305 promising COFs retained in Step 2 are characterized with mixture GCMC calculations using a binary inlet gas with a CO_2:N_2 ratio of 15:85, mimicking the typical flue gas composition of industrial power plants. With this type of simulations, the more accurate mixture selectivity, S_{mix}, can be derived. As opposed to the ideal selectivities calculated in the previous steps, this feature takes into account the effects of competitive adsorption and coadsorption. When the two gas components do not interact, as in single-component GCMC calculations, a specific adsorption site is available to both CO_2 and N_2 molecules. However, when they can interact, only one molecule can attach to this site. Furthermore, the nature of the adsorption site can change when adding other molecules. These effects can alter the guest uptakes and, thus, also the selectivity.

Due to the small interactions with the pore walls, switching from single-component to mixture GCMC calculations changes the selectivity of the large-pore COFs only negligibly, as illustrated in Fig. 5.8a. When the pore sizes become smaller and the ideal selectivities increase, larger corrections are observed. Whereas the deviations remain small when the mixture selectivity is lower than the ideal selectivity, the selectivity can increase by multiple orders of magnitude when switching from ideal to mixture selectivity. As such, the ideal selectivity behaves as a rough lower limit for the mixture selectivity and the highest APS_{mix} will be even larger than the highest $\text{APS}_{\text{ideal}}$. These findings indicate that the ideal selectivity serves as an acceptable starting point to predict the separation performance of COFs. However, the introduction of a mixture gas resembling experimental conditions has a significant impact on the specific adsorption characteristics.

In Fig. 5.8b, an updated version of Fig. 5.5a and Fig. 5.7 is provided for the set of promising COFs in which the ideal selectivity has been replaced by the mixture selectivity. As could also be observed in Fig. 5.8a, the mixture selectivity of 2D COFs is on average larger than that of 3D COFs. However, due to the larger spread of the selectivity of 3D COFs, these provide the materials with the highest APS_{mix} values. With an APS_{mix} of 2835.1 mol/kg, the best performing material is the 3D imide COF twt-b_12-03-07_17-09-07_None_None ($\Delta N_{\text{CO}_2} = 13.2$ mol/kg, $S_{\text{mix}} = 215.3$). Also the 2D COF with the highest APS_{mix} value is an imide COF, *i.e.*, hne_17-09-07_17-09-07_17-09-07_02-03-07_02-03-07_02-03-07_02-03-07_02-03-07 ($\Delta N_{\text{CO}_2} = 5.7$ mol/kg, $S_{\text{mix}} = 127.1$). Both materials are visualized in Fig. 5.9a. Despite being the top performers, these materials exhibited a notably lower ideal selectivity of 7.8 and 9.0, respectively. This demonstrates the necessity

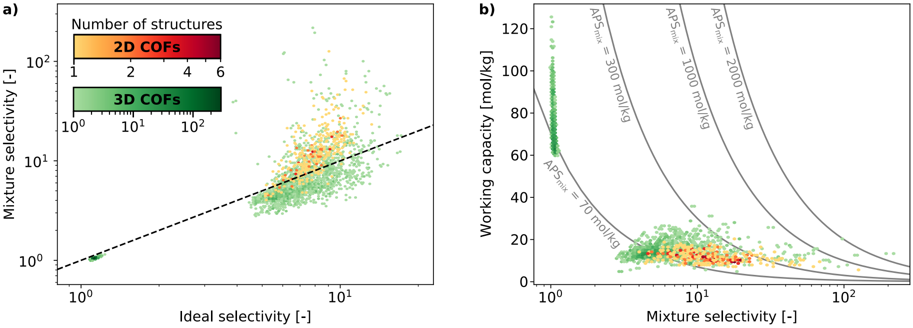


Figure 5.8: Analysis of the third step of our high-throughput screening in which the working capacity and mixture selectivity of set of 3305 promising COFs is determined. a) Comparison between the ideal selectivity and the mixture selectivities. b) The relation between the CO_2 working capacity and the mixture CO_2/N_2 selectivity of the COFs in the characterized subset. The color codes indicate the density of 2D (orange/red) and 3D (green) COFs. Figure reproduced from Ref. 321 with permission of the American Chemical Society (©2024).

of mixture GCMC calculations to effectively rank the COFs with the best carbon capture performance.

5.3.4 Design strategies for top-performing COFs

To determine design rules for experimental researchers to synthesize outstanding COFs for carbon capture, we closely investigated the top-20 of 2D and 3D COFs with the largest APS_{mix} values. Our hypothesis of Section 5.3.1 that imide linkages result in materials with an enhanced carbon capture performance is validated since 75% and 100% of the structures in the top-20 of 3D and 2D COFs exhibit this linkage type, respectively. Among the top-20 of 3D COFs, a three-fold connected phenyl ring is frequently observed, as proven by the presence of SBUs 11-03-04 and 17-09-07, which are detected in 20% and 70% of the materials, respectively. Within the top-20 of 2D COFs, the fluor-functionalized and imide-terminated phenyl rings, represented by SBUs 02-03-07 and 17-09-07, are recognized in 50% and 55% of the structures, respectively. Furthermore, the top-5 of 2D COFs contains only materials that are assembled by these two building blocks. These SBUs are visualized in Fig. 5.9c.

The distribution of CO_2 molecules for the best-performing 3D and 2D COF is visualized in Fig. 5.9a. For the 3D COFs, adsorption sites in pores with a diameter of 1.0 nm enclosed by approximately parallel aromatic rings provide

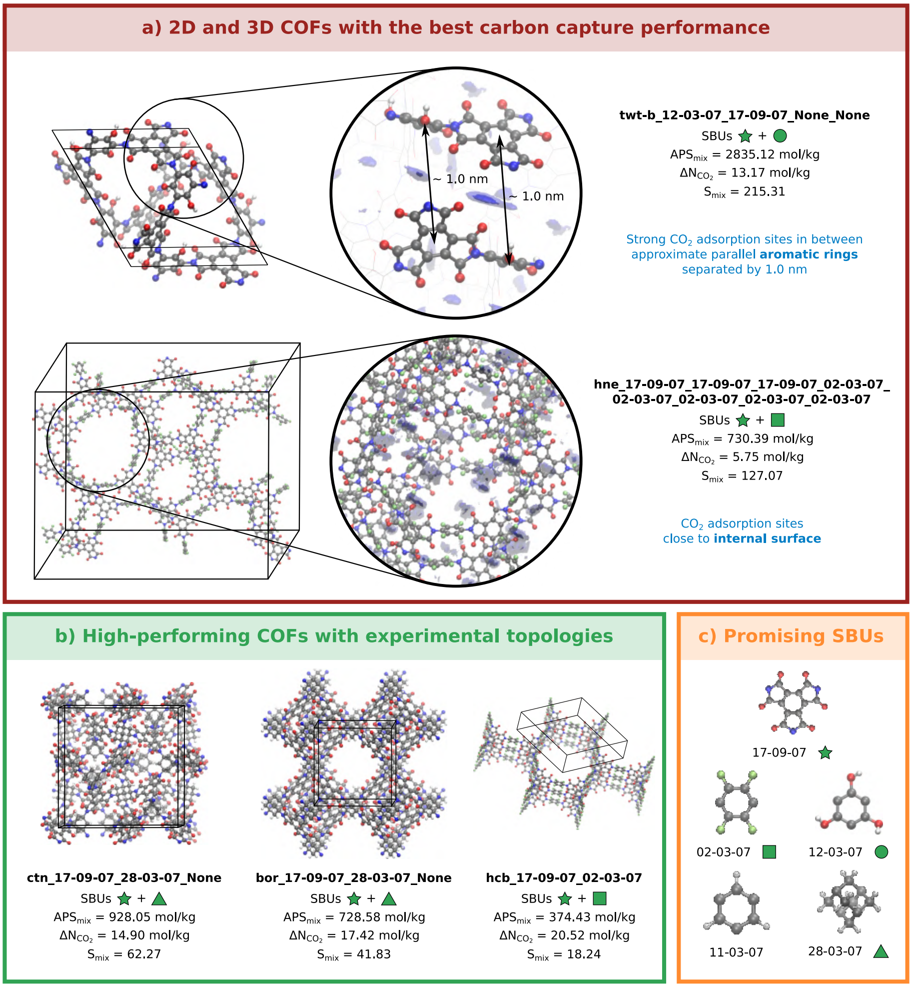


Figure 5.9: Visualization of the top-performing materials and a proposal for experimental candidates identified during our high-throughput screening. a) The 2D and 3D COFs with the highest APS_{mix} value. The insets visualize a detail of the CO_2 distribution at 10 bar. b) Candidates for experimental synthesis with high carbon capture performance. Both the topology and SBUs of these materials are observed in already synthesized COFs. c) An overview of some frequently occuring SBUs in the top-20 of best-performing 2D and 3D COFs. Figure reproduced from Ref. 321 with permission of the American Chemical Society (©2024).

the highest CO_2 densities. In a recent study, Boyd *et al.* found a similar conclusion after they screened a MOF database for carbon capture.[156] This also explains our conclusion of Section 5.3.1 that COFs with a pore diameter around 1.0 nm have an increased carbon capture performance. Since the aromatic rings in 2D COFs are mainly aligned with the layer, these high-density pockets are usually absent in these materials, explaining their lower APS_{mix} compared to 3D COFs. The large influence of the selected linkage and SBUs on the carbon capture performance of 2D COFs demonstrates that the chemical environment of the COF framework is crucial for these materials. Therefore, the CO_2 molecules in these materials are mostly located close to the pore walls, as can be observed for the best-performing 2D COF in Fig. 5.9a. The high CO_2 selectivity of fluor-functionalized COFs has been experimentally verified in triazine COFs.[326–328]

Despite the fact that hypothetical databases are very valuable to explore property-property relations and performance limits for a targeted application, synthesizing the best-performing materials can be exceptionally challenging, if not impossible. Interestingly, three COFs are identified in the top-20 of 2D and 3D COFs with experimentally observed topologies. These materials, exhibiting the **ctn**,[61] **bor**,[61] and **hcb**[26] topology, are visualized in Fig. 5.9b. Since also their SBUs have been successfully employed in COF synthesis, these materials would be excellent candidates for experimental validation, as suggested explicitely with hypothetical synthesis protocols in the SI of **Paper V**.[78, 329, 330]

In this chapter, we have performed three high-throughput screenings on the ReDD-COFFEE database. We determined the textural properties of COFs and compared them with MOFs and zeolites, demonstrating that COFs exhibit lower mass densities and higher gravimetric accessible surface areas and pore volume than the other classes of nanoporous materials. The performance limits of COFs for vehicular storage of ANG are identified. Although that the best COFs in our database can compete with the currently best performing nanoporous materials, it is still too early for their commercial development for this application. Furthermore, we also investigated the possibility to adopt COFs as nanoporous adsorbent beds in post-combustion carbon capture. Using a machine learning approach, we established several property-property relations between the macroscopic and microscopic properties of COFs, identified promising subclasses, and proposed candidates for experimental synthesis.

6

Conclusions and perspectives

6.1 Conclusion

The synthesis of the first COFs has sparked interest in this new class of functional nanoporous materials with enhanced stability and structural versatility. This large freedom in atomic geometries, enabled by the concept of reticular chemistry, allows control over the material's macroscopic physicochemical properties on the microscopic level. However, it also implies that an almost unlimited number of materials can be proposed. An experimental characterization of this vast material space is a cumbersome, if not impossible, process, due to the time-intensive synthesis of a single material, the cost of its precursors, or the harsh synthesis conditions. Therefore, computational high-throughput screenings have been developed in the past decade as an efficient and reliable toolbox to characterize material classes *in silico* prior to their experimental synthesis. Such high-throughput screenings have three major advantages: i) identifying promising candidates for a targeted application, ii) establishing property-property relationships to guide experimental researchers, and iii) exploring performance limits to determine the applicability of the material class for a certain application.[155]

The hypothetical database used in our high-throughput screenings is generated using an automated *in silico* assembly algorithm, based on an additive top-down approach, as explained in Chapter 3 and **Paper III**. The top-down approach starts from the topological embedding and deliberatily decorates its nodes with SBUs. By applying both geometrical and energetical criteria to select favorable SBU configurations, plausible periodic materials can be generated. Since the energetic criteria depend on all SBU configurations, the deformation energy should in principle be computed for each material configuration, which can become very large. To increase the computational efficiency of our algorithm, an additive approach is proposed in which the SBUs are added one-by-one by inserting them in a breadth-first iteration through the topological graph. Whereas Section 3.3, **Paper I**, **Paper II**, and **Paper IV** demonstrate our additive top-down approach in three case studies on COFs, it can also be applied to other reticular material classes, such as MOFs and zeolites.

In Chapter 4 and **Paper III**, our additive top-down approach is adopted to a large set of (topology, SBUs) combinations to generate our ReDD-COFFEE database of 268 687 COF geometries and accompanying system-specific force fields. ReDD-COFFEE is an acronym for Ready-to-use and Diverse Database of Covalent Organic Frameworks with Force field based Energy Evaluation. Using the deformation energy introduced in Section 3.1, those structures that have a low synthetic feasibility are discarded from the database. In Section 4.3, the diversity of the ReDD-COFFEE database is quantitatively compared with that of other experimental and hypothetical COF databases. Whereas the experimental databases are biased towards the often reported imine and boronate ester COFs, the hypothetical databases only include a small number of linkage types. To provide a reliable representation of material space, we included a well-balanced and diverse set of 11 linkages in our database. Furthermore, compared to the previously established COF databases, the ReDD-COFFEE database exhibits a high diversity in terms of the pore geometry and linker cores, originating from the large topological variety facilitated by the preceding top-down approach. However, the diversity in functional groups remains limited, although this can be circumvented by functionalizing the database *a posteriori*. Since all materials are provided in an energy minimum of the potential energy surface, the database is ready-to-use in molecular simulations. By providing open access to the database via the Materials Cloud, we encourage other researchers to further investigate the possibilities offered by these fascinating functional materials.[294]

Besides its reliance on the availability of a diverse database of material geometries, a high-throughput screening typically also requires an accurate, yet efficient, description of the interatomic interactions in the material. Due to the scale of our high-throughput screening, which prevents *ab initio* characterization, we have used force fields throughout this thesis. However, unlike most high-throughput screenings, which rely on generic, fully transferable force fields, we have derived system-specific force fields, fitted to quantum mechanical reference data, for each material in the database. This was possible by exploiting the reticular chemistry concept to derive force fields for periodic materials from a limited set of cluster force fields for the building blocks. This approach, introduced in Section 2.1.2, avoids the cost of deriving *ab initio* training data for all 268 687 COFs, but requires only the reference data of the 279 SBUs from which the database is generated. As such, by taking advantage of the modular nature of our database, an optimal balance between accuracy and computational feasibility in describing the interatomic interactions was obtained. The increased accuracy of these force fields is demonstrated in Section 4.2 and **Paper III** by validating their ability to reproduce experimental and *ab initio* structural characteristics such as PXRD patterns and single-crystal structures.

Besides characterizing the textural properties, we screened the ReDD-COFFEE database for two clean energy applications, *i.e.*, vehicular methane storage and post-combustion carbon capture, in **Paper III** and **Paper V**. In Section 5.1, property-property relationships between the textural properties of the COFs in our database are established and compared with MOFs and zeolites. This analysis showed that similar trends in the textural properties of multiple nanoporous material classes can be observed. Compared with the other material classes, COFs demonstrate low mass density and high gravimetric accessible surface areas and pore fractions, but low volumetric accessible surface areas. Furthermore, the performance limits of COFs as attractive storage materials for ANG in vehicular transport are checked in Section 5.2. Whereas the highest volumetric deliverable capacity approaches that of current state-of-the-art materials, the ARPA-E target for this property is not reached, preventing nanoporous materials from commercial development for this application.

Lastly, promising COFs for post-combustion carbon capture are identified in Section 5.3, together with design rules to guide experimental researchers towards high performing COFs. During this high-throughput screening, we adopted a machine learning algorithm trained on a diverse subset to predict the CO_2 working capacity and ideal CO_2/N_2 selectivity of all materials in the ReDD-COFFEE database. In this way, we could select a set of 3305 promising candidates with only a fraction of the computational cost compared to a full characterization of the database. To incorporate the effects of competitive adsorption and coadsorption, these promising COFs are subsequently characterized with mixture GCMC calculations. These demonstrate that the ideal selectivity provides a first indication of the carbon capture performance, but the mixture selectivity is required to effectively rank the best-performing COFs for carbon capture. From the first step of our screening study, it can be concluded that COFs with a pore diameter of around 1.0 nm exhibit both a high working capacity and high ideal selectivity.

These materials demonstrate adsorption sites with high CO_2 affinity in pores with this diameter that are enclosed by aromatic rings, commonly found in the top-performing 3D COFs. Unlike their 3D counterparts, 2D COFs lack these potential adsorption sites due to their layered nature. In these materials, the framework chemistry is a crucial factor to design high-potential COFs, with fluor substituents showing a superior performance among the studied functional groups. For both 2D and 3D COFs, materials featuring an imide, (acyl)hydrazone, or (keto)enamine linkage exhibit a high affinity for CO_2. Among the top-20 2D and 3D materials, three prime candidates for experimental synthesis are identified due to the availability of a synthetic route. While there are already COFs constructed with the topology and SBUs of these candidates, the current approach to COF synthesis remains largely empirical, lacking the systematic design facilitated by reticular synthesis. Whereas the deformation energy provides a first insight into the synthetic feasiblity of COFs, comprehensive descriptors to predict the stability and crystallinity of COFs are currently lacking. Nevertheless, once new synthetic pathways emerge in the rapidly evolving field of COF synthesis, the proposed materials can be quickly synthesized.

6.2 Perspectives

Of course, this thesis is only one piece in the ongoing field of COF research and high-throughput screenings on nanoporous materials. Similar as I could stand on the shoulders of the researchers that came before me and take advantage of their pioneering work, I hope that my work contributes to the fascinating science in the field of nanoporous materials. In this concluding section, I will provide a short overview of possible future avenues building on the research performed in this thesis.

Within the context of automated *in silico* materials design, substantial progress is possible in three areas:

- Firstly, the topological representation of a material is capable of describing a wide range of spatial disorder phenomena, such as missing SBUs, phase separation, and grain boundaries, and can even represent finite crystals.[331] Therefore, the top-down approach has a large potential to generate defective structures, which remains largely unexplored.
- Secondly, little attention has been devoted to the accurate prediction of low-energy structures. At the moment, Weaver is the only software package that screens multiple material configurations to find the most optimal combination of SBU configurations.[149] Whereas its developers

implemented a generic algorithm, also other optimization routines can be explored.

- Lastly, many materials, such as multivariate reticular materials, demonstrate multiple low-energy configurations. The specific configuration of a cell represents chemical or physical information, and the sequence of cells in the material can be interpreted as an array of multistate bits. The amount of information that can be stored in a single cell of a material, *i.e.*, the number of feasible material configurations, is defined as the unit cell information capacity (UCiC).[332] With a slight adaptation of the top-down scheme, these configurations can be explicitly enumerated, to enable a theoretical prediction of this metric.

Turning our attention to the generation and screening of material databases, I again discern three major directions for future development.

- In the first place, further efforts have to be made to develop descriptors similar to the deformation energy defined in Eq. 3.2 that can predict the synthetic feasibility of a material.[279, 333] Besides providing unprecedented insight in synthesis processes, such descriptors would also allow us to extract those materials from hypothetical databases that can actually be synthesized.

- Secondly, it is more important for a database to provide a balanced representation of material space than to have a large number of materials. As illustrated in Section 4.3.1 for COF databases, experimental databases have the tendency to overlook subclasses with few synthesized materials, even though they can have outstanding characteristics, and hypothetical databases represent only part of the material space due to their specific assembly algorithm. Therefore, small datasets with high diversity are preferred over large databases with limited diversity. A promising pathway to generate new materials has been proposed by Majumdar *et al.*[334] They analyzed experimental MOF databases to identify regions in material space that are underrepresented in hypothetical databases. Using this information, they generated a set of 20 000 MOFs that drastically improved the diversity of the hypothetical databases.

- Finally, it is often unnecessary to screen a whole database to identify candidates for a targeted application. Using an active learning approach, high-performing materials from an existing database can be efficiently found.[210, 335] In such iterative procedure, new structures are added to a subset in each step to maximize the predictive power of the

> characterized dataset. When a material can be featurized before its explicit construction, the database can even remain virtual and only those structures that have to be computed should be generated.[192, 209, 213]

In the past few years, machine learning potentials (MLPs) have emerged as a new level of theory between system-specific force fields and quantum mechanical calculations. These are numerical potentials generated by fitting a machine learning model to extensive quantum mechanical training data.[21–24] Since their analytical expression, which involves a huge number of parameters, allows to learn complex relationships, MLPs are able to approach *ab initio* accuracy. Whereas they are mostly adopted to describe local interactions, they have started to find their way in determining processes that are dominated by non-covalent interactions, such as adsorption and separation.[336] While the quest for an accurate universal MLP, virtually applicable to any imaginary material, is open,[337, 338] the field of high-throughput screenings would largely benefit from a database accompanied with an MLP for each material. This would tremendously increase the attained accuracy of any high-throughput screening performed on it. Similar to the cluster approach followed in this thesis, deriving an MLP for the respective SBUs instead of the complete materials, could drastically reduce the required training data, and thus computational cost, for obtaining potentials for a large material dataset.[339]

Obviously, high-throughput screenings on the ReDD-COFFEE database are not limited to the ones demonstrated in this thesis. Numerous screening studies are possible, only limited by the imagination and perseverance of the involved scientists. Whereas most screenings to date have focused on adsorption properties, due to their computational efficiency, other applications have started to emerge as well. Our ReDD-COFFEE database is especially interesting for those applications that require MD simulations, such as mechanical stability[152, 188] and thermal conductivity,[187] since these will benefit most from the higher accuracy of the interactions described by our derived *ab initio* force fields. Also high-throughput screenings on quantum mechanical properties such as the electronic band gap[139, 189, 190] and catalytic activity[191] have been performed. Since these are computationally costly, they would benefit most from a reduced subset of the ReDD-COFFEE database or an active learning approach, as it would be too expensive to characterize the whole database.

With this thesis we hope to have contributed to the further development of COFs with tailored properties for diverse applications, ranging from gas adsorption for clean energy application to catalysis. As we stand at the forefront of materials design, this work not only expands the frontiers of

COF research but also demonstrates the potential of computational research in guiding experimental efforts towards the realization of advanced and functional materials. The insights gained from this study not only advance the field of COFs but also serve as a blueprint for the rational design of novel reticular nanomaterials in the broader landscape of materials science and engineering.

Part II

Published Papers

A

Selection of Publications in International Peer-Reviewed Journals

Paper I **Quantifying the likelihood of structural models through a dynamically enhanced powder X-ray diffraction protocol**
S. Borgmans, S.M.J. Rogge, J.S. De Vos, C.V. Stevens, P. Van der Voort, V. Van Speybroeck
Angewandte Chemie International Edition, **2021**, 60 (16), 8913–8922

Paper II **Exploring the phase stability in interpenetrated diamondoid covalent organic frameworks**
S. Borgmans, S.M.J. Rogge, J.S. De Vos, P. Van Der Voort, V. Van Speybroeck
Communications Chemistry, **2023**, 6 (1), 5

Paper III **ReDD-COFFEE: A ready-to-use database of covalent organic framework structures and accurate force fields to enable high-throughput screenings**
J.S. De Vos, S. Borgmans, P. Van Der Voort, S.M.J. Rogge, V. Van Speybroeck
Journal of Materials Chemistry A, **2023**, 11 (14), 7468–7487

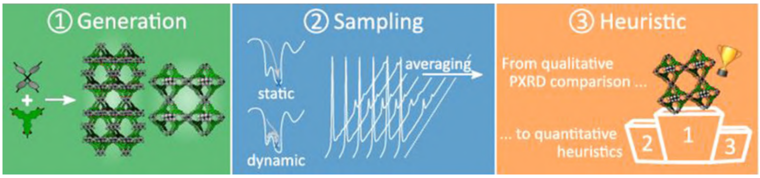
Paper IV **Engineering of Phenylpyridine- and Bipyridine-Based Covalent Organic Frameworks for Photocatalytic Tandem Aerobic Oxidation/Povarov Cyclization**
M. Debruyne, S. Borgmans, S. Radhakrishnan, E. Breynaert, H. Vrielinck, K. Leus, A. Laemont, J. De Vos, K.S. Rawat, S. Vanlommel, H. Rijckaert, H. Salemi, J. Everaert, F. Vanden Bussche, D. Poelman, R. Morent, N. De Geyter, P. Van Der Voort, V. Van Speybroeck, and C.V. Stevens
ACS Applied Materials & Interfaces, **2023**, 15 (29), 35092–35106

Paper V **High-Throughput Screening of Covalent Organic Frameworks for Carbon Capture using Machine Learning**
J.S. De Vos, S. Ravichandran, S. Borgmans, L. Vanduyfhuys, P. Van Der Voort, S.M.J. Rogge, and V. Van Speybroeck
Chemistry of Materials, **2024**, 36 (9), 4315–4330

Paper I

Quantifying the Likelihood of Structural Models through a Dynamically Enhanced Powder X-Ray Diffraction Protocol

Sander Borgmans, Sven M. J. Rogge*, Juul S. De Vos, Christian V. Stevens, Pascal Van Der Voort, and Veronique Van Speybroeck*



Angewandte Chemie International Edition, **2021**, 60 (16), 8913-8922

The Supporting Information is available at:
https://doi.org/10.1002/anie.202017153

J. S. De Vos was responsible for the first step in the protocol, *i.e.*, the automated generation of hypothetical structures. He generated *ab initio* reference data for each of the seven building blocks in this work, from which cluster force fields were derived. By combining these SBUs on various topologies, a set of hypothetical structural models was generated together with a system-specific force field for each of them. Subsequently, a geometry optimization was performed on each of these materials.

---

*Corresponding authors

GDCh
Research Articles
Check for upd
Angewandte International Edition Chemie

Operando Modeling **Hot Paper**

How to cite: *Angew. Chem. Int. Ed.* **2021**, *60*, 8913–8922
International Edition: doi.org/10.1002/anie.202017153
German Edition: doi.org/10.1002/ange.202017153

Quantifying the Likelihood of Structural Models through a Dynamically Enhanced Powder X-Ray Diffraction Protocol

Sander Borgmans, Sven M. J. Rogge, Juul S. De Vos, Christian V. Stevens, Pascal Van Der Voort, and Veronique Van Speybroeck**

Abstract: *Structurally characterizing new materials is tremendously challenging, especially when single crystal structures are hardly available which is often the case for covalent organic frameworks. Yet, knowledge of the atomic structure is key to establish structure-function relations and enable functional material design. Herein, a new protocol is proposed to unambiguously predict the structure of poorly crystalline materials through a likelihood ordering based on the X-ray diffraction (XRD) pattern. Key of the procedure is the broad set of structures generated from a limited number of building blocks and topologies, which is submitted to operando structural characterization. The dynamic averaging in the latter accounts for the operando conditions and inherent temporal character of experimental measurements, yielding unparalleled agreement with experimental powder XRD patterns. The proposed concept can hence unquestionably identify the structure of experimentally synthesized materials, a crucial step to design next generation functional materials.*

Introduction

The design of functional materials is instrumental for advancing technological solutions to tackle pressing societal problems,[1–7] such as combating climate change and creating clean energy pathways. In the last few years, new materials with particular functionalities have been proposed at a very high pace, due to an in-depth microscopic understanding of their functionality and the growing toolbox of synthetic chemists. Within this toolbox lies the concept of reticular design,[8] where judiciously chosen building blocks are assembled in a predesigned topology, giving rise to materials such as metal-organic frameworks (MOFs)[9] and covalent organic frameworks (COFs).[10] Due to their extensive tailorability and attractiveness for a vast range of applications, the structural design of MOFs and COFs has formed a very active field of research.[11–15] Thanks to a meticulous control over synthesis conditions and an ever-expanding versatile set of experimentally available building blocks, it has even become possible to synthesize combinations of building blocks in different structures. To fully explore the functionalities of the enormous amount of potential structures that can be formed from a given set of building blocks, an accurate protocol is needed to systematically derive structural models for these materials at realistic working conditions of temperature and pressure.[16,17] Such structural characterization is a conditio sine qua non for the atomically guided design of functional applications in the fields of catalysis, sorption, light harvesting and more.

Reticular synthesis states that the atomic-level structure of any framework material is defined by a structural model containing three elements:[18] (i) the material's building blocks, (ii) its topology, and (iii) the embedding of the building blocks in this topology. When designing new experimental materials, X-ray crystallography is typically adopted to construct such a structural model. While single-crystal X-ray diffraction (SCXRD) measurements allow for the direct atomic-level determination of a material's bulk structure,[19] it is notoriously difficult to obtain large single crystals for most porous materials such as MOFs and COFs and one usually resorts to powder X-ray diffraction (PXRD) measurements[20] or electron crystallography.[21,22] During a PXRD measurement, however, the 3D reciprocal space is non-injectively projected onto a 1D Bragg angle, making the determination of the true structural model impossible without prior information about the structural model. Given the need for such an ad hoc model—which is absent in SCXRD—PXRD measurements require a method to verify the validity of these proposed structural models. Therefore, as discussed in more detail in Section S1.1, whole powder pattern fitting (WPPF) methods are used to either reject or accept a possible model based on a visual analysis of the experimental and modeled PXRD patterns, and on error measures such as the weighted

---

[*] S. Borgmans, Dr. S. M. J. Rogge, J. S. De Vos,
Prof. Dr. V. Van Speybroeck
Center for Molecular Modeling (CMM), Ghent University
Technologiepark 46, 9052 Zwijnaarde (Belgium)
E-mail: Sven.Rogge@UGent.be
Veronique.VanSpeybroeck@UGent.be

Prof. Dr. C. V. Stevens
Research Group SynBioC, Department of Green Chemistry and Technology
Faculty of Bioscience Engineering, Ghent University
Campus Coupure, Coupure Links 653, 9000 Gent (Belgium)

Prof. Dr. P. Van Der Voort
Center for Ordered Materials, Organometallics and Catalysis (COMOC)
Department of Inorganic and Physical Chemistry, Ghent University
Krijgslaan 281 (S3), 9000 Gent (Belgium)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.202017153.

profile residual.[23] However, when noisy PXRD patterns are obtained, such as for structures that lack long-range order, extracting any structural information becomes increasingly difficult. As a result, classical methods that rely on the diffraction data to refine initial structural parameters, such as the WPPF methods or other popular top-down methods, fail to accurately discern between multiple possible structural models.[24–28]

To circumvent this ambiguity and ensure that no structural models are overlooked when characterizing challenging functional materials, we here present an automated protocol that fully accounts for the dynamic averaging of structures taking place at operating conditions. The procedure, schematically illustrated in Figure 1, reverses the aforementioned experimental PXRD approach by exhaustively exploring all potential structures starting from the building blocks and all possible topologies, and identifying the most likely structural model upon comparing with an experimental PXRD pattern. To that end, a large set of possible structural models is generated from the provided building blocks, which are assembled in various topologies in the first step of our protocol. In the second step, PXRD patterns are derived, using either a static or a dynamic approach in which the inherent dynamic character of an experimental measurement is introduced. In the third and final step, a heuristic is derived and applied to order the proposed structural models as a function of their likelihood to reproduce the experimental diffraction pattern. Such a systematic computational protocol has the added benefit of replacing the ad hoc visual analysis by a well defined metric, allowing quantitative statements on how well the model performs.

COFs are ideal materials to test our protocol, as their structural characterization is challenging due to often noisy PXRD patterns, owing to a lack of long-range order and crystallinity in most experimental samples. These promising materials are made up from light, covalently bound, organic constituents, combining low mass densities with a high stability. Potential applications are situated in the field of drug delivery,[29–31] (photo-)catalysis,[32–34] sensing,[35–37] (opto)-electronics[38–40] and gas/liquid storage and separation.[41–43] In general, one can distinguish between 3D COFs, which are completely connected by strong covalent bonds, and 2D COFs, which form layered structures whose stacking order is determined by much weaker forces, such as dispersion and Coulomb interactions.[44] Only a handful of SCXRD patterns are available for COFs, which are limited to the strongly connected 3D COFs.[45] Up to now, it is therefore common practice to propose initial structural COF models by choosing a topology based on the size of each of the building blocks, their point symmetry, and their connectivity with neighboring building blocks (see Figure 2 for representative examples), favoring the most symmetric topology when multiple possibilities arise.[46] However, with over 500 different COFs being synthesized,[47] accurate and systematic characterization techniques are crucial to enable rational functional material design.

Results and Discussion

Material Generation

The general applicability of the in silico protocol is illustrated for materials being built from seven different building blocks, illustrated in Figure 2, giving rise to at least five different experimentally characterized COFs. These

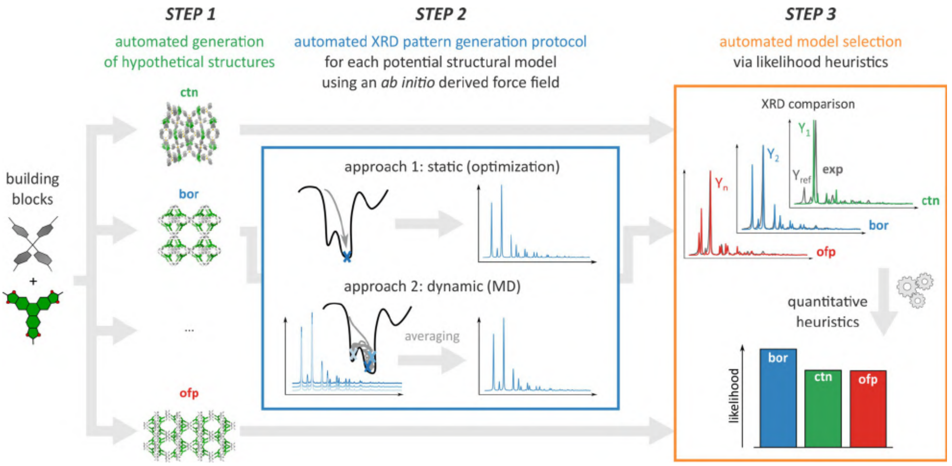


Figure 1. The proposed three-step structural characterization workflow. For each of the hypothetical structural models (generated in step 1), representative PXRD patterns are derived via either a static or a dynamic approach (step 2), which are then ranked according to their likelihood to reproduce the experimental reference PXRD pattern (step 3).

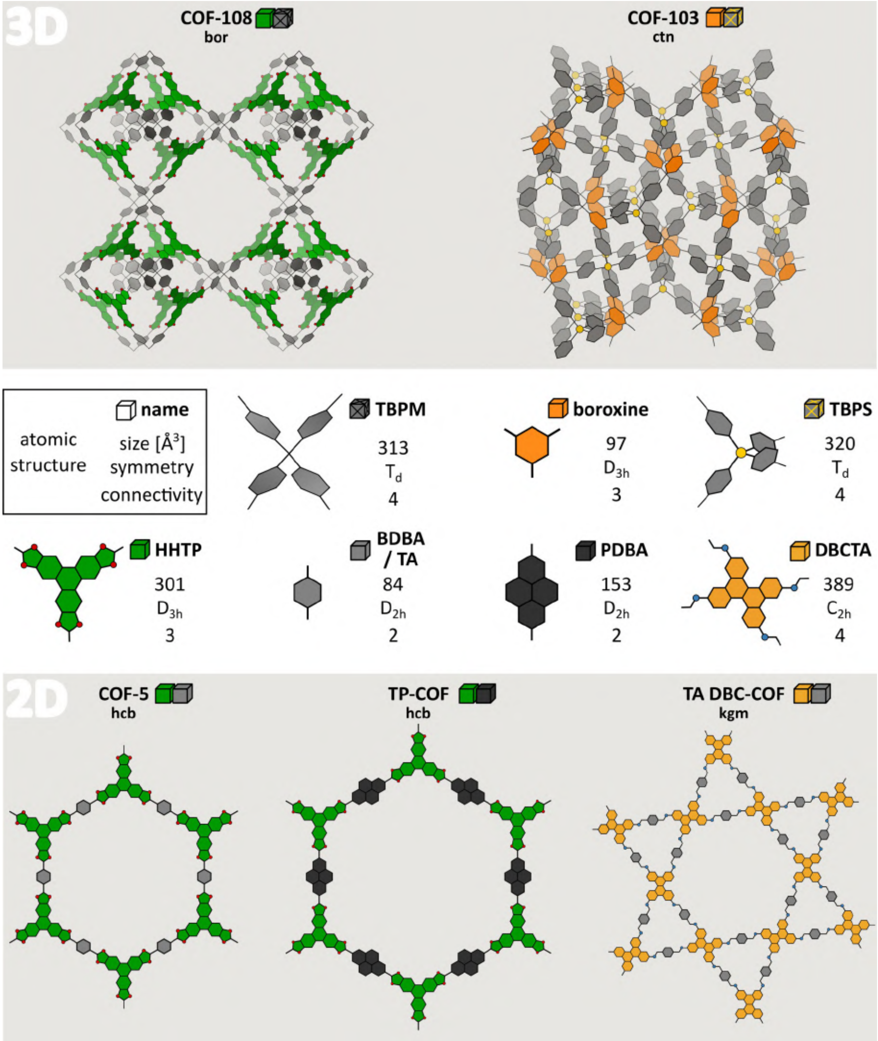
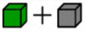
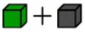
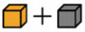
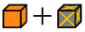
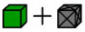


Figure 2. Building blocks and experimentally derived structural models in which they assemble for five experimentally synthesized COFs. The atomic structure and name of each building block (see Table S1 for the abbreviations) are reported, next to their size, symmetry, and connectivity, as expressed by the van der Waals volume, the point group, and number of extension points, respectively.

building blocks were chosen based on their varying sizes, symmetries and connectivities. The two 3D COFs, COF-103 and COF-108, were among the first 3D COFs to have ever been synthesized[48] and exhibit the **ctn** and **bor** topology, respectively. The bottom panel shows the three 2D COFs. Similar to the chosen 3D COFs, COF-5 was one of the first ever synthesized COFs,[10] adopting the **hcb** topology. TP-COF[49,50] was chosen as an isoreticularly expanded equivalent of COF-5, whereas TA DBC-COF[51] was chosen due to its unique **kgm** topology.

Due to the modular nature of COFs, the seven building blocks of Figure 2 can be assembled in a combinatorial amount of structural models and hence hypothetical COF structures. When considering the five different combinations of two building blocks each that give rise to the five COFs in Figure 2, almost 700 unique structural models can be con-

Table 1: Number of 2D and 3D structural models generated by our in-house protocol (step 1 in Figure 1).

		# hypothetical structural models	
blocks	*experimentally synthesized material*	2D	3D
	COF-5	42	
	TP-COF	42	
	TA DBC-COF	19	513 (55)
	COF-103		26
	COF-108		26

Block labels correspond to those in Figure 2. For TA DBC-COF the number of structural models with a 3D topology that are retained is reported between brackets.

structed by only looking at the possible topologies for the co-condensation of each building block combination. In Table 1, this number is broken down for each building block combination.

Most building block combinations give rise to only a tenfold of structural models due to the assumed rigidity of the building blocks. However, the building block combination that assembles TA DBC-COF can synthesize in 19 and 513 distinct 2D and 3D topologies, even though only a 2D structural model was resolved experimentally. This huge number of hypothetical models can be attributed to two factors. First, the slight non-planar character of the DBCTA building block accommodates both 2D and 3D topologies, in contrast to all other building blocks considered here. Second, removing the TA building block would preserve the original topology, similar as for any other twofold connected building block. Consequently, the TA DBC-COF can hypothetically synthesize in any of the large subset of topologies that contain only fourfold coordinated building blocks. By filtering out the most strained of these structural models, 19 2D and 55 3D hypothetical models are retained for the TA DBC-COF. For the five building block combinations of Figure 2, step 1 of our protocol hence generates a total of 210 structural models. For each of these models, a fully flexible ab initio based force field was derived using the QuickFF[52,53] protocol (see Section S2) to characterize and validate them with respect to the experimental PXRD patterns.

Automated Model Selection Protocol via Heuristics

An ideal PXRD pattern consists of discrete Bragg reflections, where each peak j is defined by its position θ_j, intensity I_j and shape,[23] as illustrated in Figure 3a. These components are uniquely defined for each structural model; deviations between experimental and representative computational PXRD patterns (see Section S1.2) can hence help localize model inaccuracies through these three components. To assess the correspondence among PXRD patterns, a quantitative measure is required, which, in our protocol, relies on the similarity index S_I and the weighted profile residual R_{wp}. A detailed explanation and alternative figures of merit are outlined in Section S1.4. To understand how they penalize differences between two patterns, it is instructive to consider their variations for different types of deviations in either peak or trough regions. To this end, Figure 3b visualizes four hypothetical PXRD patterns (colored) that are compared to a given reference profile (in gray). Pattern 1 has a low deviation in both the experimental peak and trough regions, whereas pattern 4 has a high deviation in both regions. Pattern 2 exhibits a high deviation in the peak regions of the experimental PXRD pattern but a low deviation in the trough regions, whereas the opposite is true for pattern 3.

From Figure 3b, it is clear that the similarity index S_I mainly focuses on the overlap of the peaks, but is largely insensitive to deviations in the trough regions. This results in comparable similarity indices for PXRD patterns 1 and 3, and 2 and 4, which differ only in the trough regions. As S_I is normalized to one (see Eq. S1.10), direct comparison between various patterns is possible, in contrast to the R_{wp} factor. However, R_{wp} considers relative instead of absolute differences, as each peak is weighted with an uncertainty factor w_i that is inversely proportional to the reference profile's height. As a result, R_{wp} penalizes the mismatch between the profile intensities, both in peak and trough regions, as is evident when considering the similar R_{wp} factors for patterns 2 and 3. Since the overall relative intensity difference between the reference and the calculated PXRD patterns conveys no physical meaning, we will globally scale each computational pattern to minimize its weighted profile residual when comparing with the full experimental pattern. This will result in a unique R_{wp} value for each pattern, whereas S_I is unaffected as it is scale invariant. Since these figures of merit can vary significantly with the quality of the reference pattern, there is no specific heuristic value at which the agreement is deemed sufficient. However, as the best of two models is defined by a lower value of R_{wp}, ($R_{wp,ideal} = 0$) and a higher value of S_I ($S_{I,ideal} = 1$), the heuristic will allow for a quantitative ordering of which models are most likely to reproduce the experimental data, resulting in an automated model selection criterion.

Finally, the credibility of any figure of merit is significantly influenced by the presence of background noise in the experimental reference patterns. Especially for poorly crystalline COFs, significant background noise is expected, which originates from amorphous regions, and should be filtered. Clearly, a non-negligible experimental background signal at Bragg angles that correspond to a reflection peak in the computational pattern would artificially improve the figure of merit. As such, eliminating the background profile is essential before employing any PXRD heuristic. This is achieved here using the robust Bayesian analysis method (see Section S1.5),[54] resulting in a protocol that can also be applied if the material contains large amorphous regions. Henceforth, only figures of merit applied on the background filtered profile will be reported, which will be indicated by the "B" subscript.

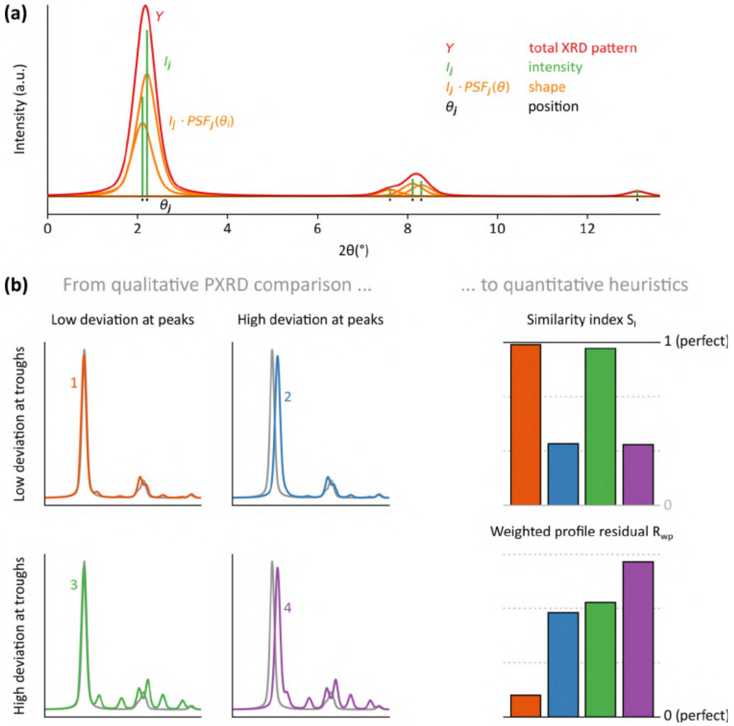


Figure 3. (a) Illustrative PXRD pattern (red) built up from six individual Bragg peaks that are each characterized by their position (black), intensity (green) and shape (orange). (b) Illustration of the figures of merits used in step 3 of the procedure for four hypothetical and representative PXRD patterns.

3D COFs

At first instance, the protocol is applied to the two building block combinations that give rise to the 3D COFs of Figure 2. In the static approach, all potential structural models of Table 1 are optimized towards their equilibrium structure at 0 K and the PXRD patterns of these optimized structures are derived following the procedure outlined in Section S1.3. The resulting likelihood ordering, based on the heuristic value, is visualized in Figure 4 a,b, where the reference experimental PXRD patterns were extracted from ref. [48]. As the heuristic is calculated based on the relative difference between two profile intensities, the 2θ range is limited to the measured experimental data. Consequently, any diffraction peak of our hypothetical models that falls outside this experimental 2θ range is not taken into account when calculating the heuristics. As this can artificially increase the heuristic results for these structural models—it is expected that the experimental pattern has no matching peaks outside the experimental 2θ range—the names of these structural models are indicated by an asterisk (*) in Figure 4 and should be considered with care.

The green regions in Figure 4 a,b correspond to successfully retrieved experimental structures, whereas the gray regions indicate other likely potential structural models based on the heuristic values. For the 3D-COFs, COF-103 and COF-108, the static approach succeeds in identifying the correct experimentally resolved structural model when taking both the $S_{I,B}$ and $R_{wp,B}$ values into account. The gray zones correspond to structures which might potentially be viable, but it is immediately clear that only one heuristic value is favorable, either $S_{I,B}$ is close to one or $R_{wp,B}$ is close to zero. For COF-103, shown in Figure 4 a, an appreciable difference is observed between the best **ctn** topology and the second best **mmm** topology, facilitating the correct model selection. While our protocol also correctly identifies the experimentally reported topology for COF-108 (Figure 4 b), the distinction between the best topology (**bor**, indicated in green) and the second best topology (**asn** when considering $R_{wp,B}$ and **ofp** when considering $S_{I,B}$, indicated in gray) is smaller than for COF-103 due to the larger residual experimental background for COF-108 (see Section S1.5). Clearly, the static approach of our protocol is able to correctly identify the most probable structural model for 3D COFs, which are connected

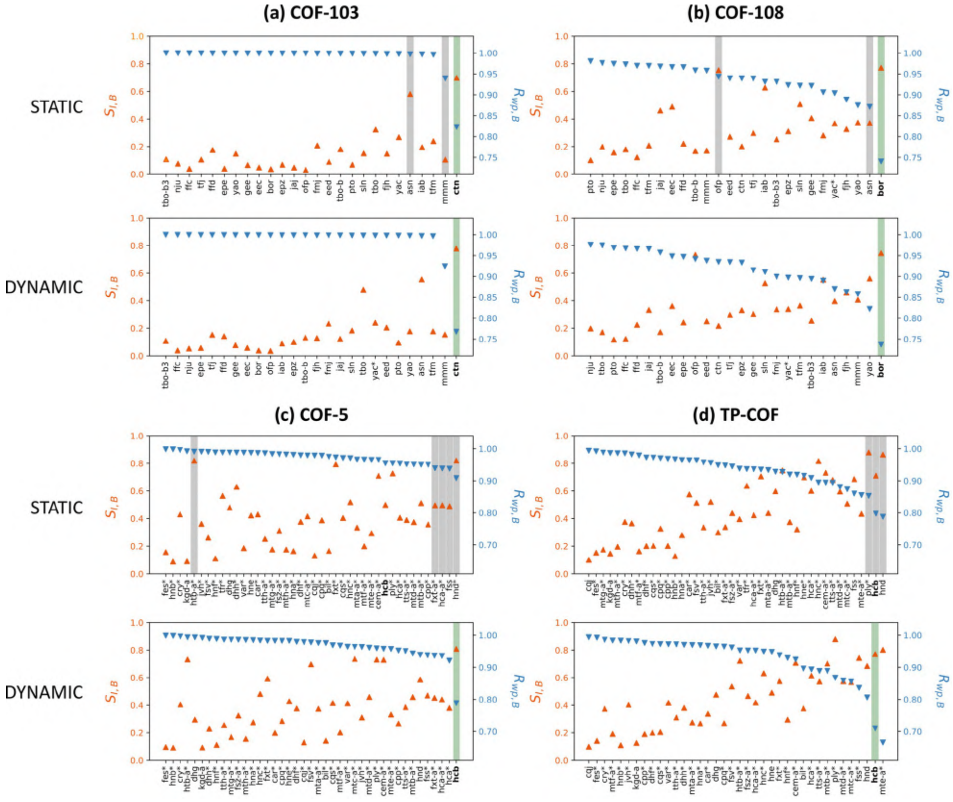


Figure 4. Similarity index $S_{I,B}$ (orange) and weighted profile index $R_{wp,B}$ (blue) ordered as a function of decreasing $R_{wp,B}$ for (a) COF-103, (b) COF-108, (c) COF-5 and (d) TP-COF, without background noise, using either a (top) static or (bottom) dynamic approach. The green regions indicate a successfully retrieved structural model corresponding to the experimentally reported topology (in bold face), whereas the gray regions correspond to favorable heuristic values that are unsuccessful. An asterisk (*) indicates that the heuristic results for that topology are artificially increased due to significant peaks outside the experimental 2θ range, which are not taken into account when comparing the PXRD patterns.

through strong covalent bonds and thus less prone to dynamic effects.

2D COFs

Turning our attention to the 2D COFs, visualized in Figure 4 c,d, our static approach is unable to identify the best structural models for COF-5 and TP-COF, with no distinct outlier, in contrast to the 3D COFs. The **htb-a**, **fxt-a**, **hca-a**, **fss** and **hnd** topologies for COF-5, and the **ply**, **hcb** and **hnd** topologies for TP-COF, indicated in gray, have similar $S_{I,B}$ and/or $R_{wp,B}$ values. However, an optimal heuristic combination is missing for the experimentally reported **hcb** topology of both materials.[10,50] This mismatch originates from the inability of the static approach to account for dynamic averaging at working temperatures, as only one low-energy structure with a fixed layer inclination is considered (see also Section S5). This limitation is expected to manifest itself for every 2D COF, as their layers are held together by weak, non-covalent forces. However, for TA DBC-COF the static approach does predict the correct **kgm** topology, as in this case the DBCTA building blocks hinder layer shearing due to steric constraints (Figure S18).

To resolve these dynamic effects in structures that have a large intrinsic freedom, characterized by shallow minima on the free energy surface due to the presence of weak interactions, it is necessary to generate PXRD patterns through MD simulations at operando conditions of temperature and pressure (see Figure 1). In this dynamic approach, the PXRD pattern of any structural model is obtained by averaging PXRD patterns collected at different snapshots from the MD trajectory of this structural model. As an example, Figure 5 b visualizes the experimental PXRD pat-

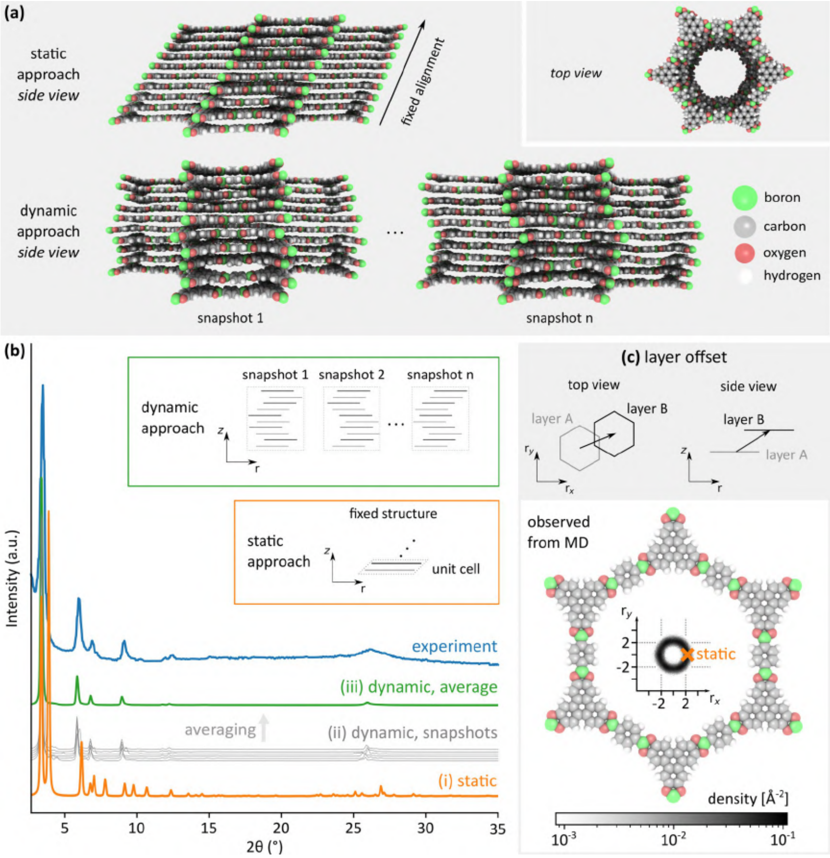


Figure 5. (a) Atomistic representation of the stacking in both the static and dynamic approach, containing two and ten layers, respectively, in their unit cell. (b) Diffraction pattern comparison for COF-5 using the **hcb** topology for the optimized structure (static, orange), six representative snapshots taken from the MD trajectory (gray) and the diffraction pattern obtained by averaging over fifty snapshots of the MD trajectory (dynamic, green). (c) Definition of the layer offset between subsequent layers, showing the variation of this offset during an MD simulation as a heatmap.

tern of COF-5 together with (i) the diffraction pattern of the optimized **hcb** structural model (static approach), (ii) the diffraction patterns corresponding with several snapshots through the MD simulation, and (iii) the diffraction pattern obtained by averaging over the patterns obtained at 50 uniformly distributed MD snapshots, where each snapshot corresponds to an embedding in the **hcb** topology (dynamic approach). Clearly, the time-averaged PXRD pattern of the dynamic approach significantly improves the representation of the true atomic structure during measurement, increasing the similarity between the simulated and experimental PXRD patterns.

Turning our attention to the differences between the computational PXRD patterns in Figure 5b, two observations can be made. First, compared to the PXRD pattern in the static approach, the reflection peaks occurring in the PXRD patterns of the MD snapshots are either shifted towards lower 2θ values or disappear completely, with the exception of the first peak at $2\theta = 3.4°$. While the shift in peak positions indicates an increase in distance between the diffraction planes at finite temperatures, the complete disappearance of peaks indicates that the static structure is too symmetric, featuring extra diffraction planes compared to the experimental structure. This artificial symmetry in optimized structures is a well-known effect,[55] which is resolved by the MD simulations at finite temperature in our dynamic approach, as demonstrated in Figure 5b. Second, different MD snapshots give slightly different PXRD patterns. These small deviations are expected, as they correspond to the natural dynamics of the atoms at operando conditions and replace the missing atomic displacement factors in Eq. S1.4. Therefore, the MD-averaged PXRD pattern obtained in the dynamic approach contains, in a natural way, the broadening effects that are absent in the static approach. As a result, Figure 5b demonstrates that our dynamic approach leads to a much better correspondence with the experimental PXRD pattern compared to the static approach.

For 2D COFs, the largest discrepancy between the static and dynamic approach finds its origin in the ability to describe the dynamic character of the layer movement, which is evident from Figure 5a. In a static simulation, the layer offset (see Figure 5c and Section S1.7) has a fixed magnitude and direction, as defined by the cell vectors. Considering two layers in the optimized structure, they are either aligned, yielding an eclipsed stacking (AA), or displaced with respect to each other, yielding serrated (AB) or inclined stacking configurations, where the latter is illustrated in Figure 5a. In contrast, in a dynamic simulation, the layer offset fluctuates in both magnitude and direction during the simulation, as shown in Figure 5a for two snapshots. Moreover, in our dynamic approach, a larger system consisting of ten layers is considered, giving rise to more complex layer behavior. Figure 5c shows the observed layer offsets both in a static and dynamic approach through a heatmap. Dynamically, a range of offsets is found, all with a magnitude around 2.2 Å without any preferential directionality, whereas in the static approach only one single offset is found with the same magnitude but a fixed direction, which is insufficient to account for the dynamics of the system.

Based on the COF-5 case study in Figure 5, one expects the dynamic approach to yield a much better heuristic value compared to the static approach for 2D COFs consisting of weakly bound layers. Figure 4c,d indeed reveals significant alterations when comparing the dynamic with the static approach for 2D COFs. Remarkably, in the dynamic approach, the most likely structural model coincides with the expected topology for COF-5 and, when ignoring the structural models that contain diffraction peaks outside the experimental 2θ range, also for TP-COF. We also adopted the dynamic approach for the 3D COFs and TA DBC-COF, finding no substantial difference with the static approach due to the strong connectivity and steric hindrance in these materials allowing only minimal deviations under finite temperatures and pressures (see Figure 4a,b and Figure S18). These results illustrate the general applicability of our dynamic protocol to predict the correct structural model.

Model Refinement

While the dynamic approach of the in silico protocol correctly predicts the topology for all COFs considered here, the precise embedding of the building blocks inside the topology can still give rise to different structural models. In turn these can lead to slight variations in the cell vectors and can significantly influence the intensities of the XRD patterns (see Section S1.2). Consider for instance TA DBC-COF, which assembles in the **kgm** topology. Even within this topology, different relative orientations of the imine linkages give rise to three different pore structures and distinct diffraction patterns, as illustrated in Figure 6. In other words, the TA building block can be embedded in three different ways, with the imine linkages of TA pointing in opposite directions (kgm1), or with the imine linkages both pointing in the same direction, which is either bent away (kgm2) or towards (kgm3) the smaller triangular pore. These different embeddings can not all be reached during a regular MD simulation as they correspond to different (meta)stable configurations that are separated by too large energetic barriers.

The heuristic values for the different embeddings reported in this figure demonstrate that the kgm3 model best reproduces the experimental PXRD pattern. Surprisingly, this structural model does not agree with the experimentally proposed model (kgm1), which is illustrated in Figure 2. This experimental model was constructed ad hoc in Material Studio and subsequently refined using the Pawley method,[51] which, however, does not fit the atomic positions, as discussed in Section S1.1. This case study illustrates that, even when the topology is correctly identified, it is instructive to consider whether different embeddings of the building blocks within the topology are possible, at least when multiple (meta)stable configurations exist. From these results it follows that our protocol not only succeeds in finding the optimal structural model for newly synthesized materials, but can also improve upon existing models through refinement of the local geometry and embedding, increasing the similarity between the experimentally observed and calculated diffraction patterns in a quantitative manner.

Conclusion

Crystal structure determination of complex nanoporous materials is key for establishing structure-function relations in material design. However, finding the atomic-level structure of experimentally synthesized materials may become a daunting task. This is certainly true for materials that lack long-range order, such as some COFs, where current techniques struggle to systematically and reproducibly identify the correct structural model as they are highly sensitive to the quality of the experimental data and are prone to confirmation bias. Herein, we presented an in silico protocol that reverses the generally applied strategy of extracting possible models from the observed PXRD peaks. Instead, by generating an exhaustive set of structural models for any given building block combination and generating representative PXRD patterns for each model through a static or a dynamic approach that accounts for the operando experimental conditions, the different structural models are ranked based on intuitive heuristics that unambiguously quantify the correspondence between the computed and experimental PXRD patterns, thereby uniquely identifying the most likely model.

The procedure was validated for several well-characterized COFs, but is equally applicable to other modular building

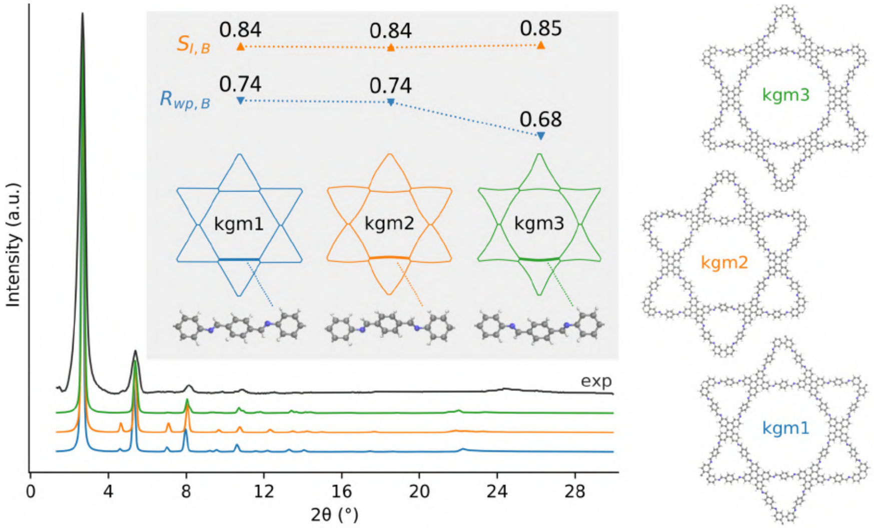


Figure 6. TA DBC-COF model refinement for three distinct structural models with the **kgm** topology but varying orientation of the imine linkage and hence varying embedding, resulting in a different pore geometry and PXRD heuristic value. The PXRD patterns are obtained experimentally[51] (gray, background-filtered) or through our dynamic approach (colored).

block materials, such as MOFs and zeolites, in any topology. Starting from seven building blocks, many structural models were proposed, through all possible topologies. For 3D COFs, a static approach without dynamic averaging proved sufficient considering their characteristic strong covalent bonds. However, for 2D COFs, which show a high degree of flexibility due to the weak interaction between the layers, it was essential to account for different dynamically generated snapshots at experimental temperatures. This approach yields a natural analog to experimental XRD measurements and corrects for an artificially too symmetric structure obtained at 0 K in a static approach. Thus, the in silico protocol proposed here can successfully detect experimental crystal structures, even for 2D COFs. Interestingly, our approach also succeeds in showing finer details of the structure. For the TA DBC-COF, our procedure can distinguish between various structural models that all exhibit the experimental topology but have different orientations of the imine linkages. Such subtleties would easily be overlooked when visually comparing PXRD peak positions.

The presented in silico workflow opens the perspective to systematically screen large databases of materials, to identify their atomic-level structural models, and establish causal structure-function relations. We hope that such detailed knowledge will help computational and experimental material scientists alike by opening the door for a more precise design of challenging modular materials for various applications.

Acknowledgements

This work is supported by the Research Board of Ghent University (BOF) through a Concerted Research Action (GOA010-17). S.M.J.R. and J.S.D.V acknowledge the Fund for Scientific Research-Flanders (FWO) for a postdoctoral fellowship (grant no. 12T3519N) and a strategic basic (SB) research fellowship (grant no. 1S94519N). V.V.S. acknowledges funding from the European Union's Horizon 2020 research and innovation program (consolidator ERC grant agreement no. 647755—DYNPOR (2015–2020)). The computational resources and services used in this work were provided by VSC (Flemish Supercomputer Center), funded by Ghent University, FWO, and the Flemish Government department EWI.

Conflict of interest

The authors declare no conflict of interest.

Keywords: covalent organic frameworks · materials characterization · operando modeling · structure prediction · X-ray diffraction

---

[1] B. M. Connolly, J. P. Mehta, P. Z. Moghadam, A. E. H. Wheatley, D. Fairen-Jimenez, *Curr. Opin. Green Sustainable Chem.* **2018**, *12*, 47–56.
[2] X. Zhao, P. Pachfule, S. Li, T. Langenhahn, M. Ye, G. Tian, J. Schmidt, A. Thomas, *Chem. Mater.* **2019**, *31*, 3274–3280.

[3] K. H. Cho, D. D. Borges, U.-H. Lee, J. S. Lee, J. W. Yoon, S. J. Cho, J. Park, W. Lombardo, D. Moon, A. Sapienza, G. Maurin, J.-S. Chang, *Nat. Commun.* **2020**, *11*, 5112.

[4] H. Zhong, M. Ghorbani-Asl, K. H. Ly, J. Zhang, J. Ge, M. Wang, Z. Liao, D. Makarov, E. Zschech, E. Brunner, I. M. Weidinger, J. Zhang, A. V. Krasheninnikov, S. Kaskel, R. Dong, X. Feng, *Nat. Commun.* **2020**, *11*, 1409.

[5] K. Gottschling, G. Savasci, H. Vignolo-González, S. Schmidt, P. Mauker, T. Banerjee, P. Rovó, C. Ochsenfeld, B. V. Lotsch, *J. Am. Chem. Soc.* **2020**, *142*, 12146 – 12156.

[6] B. M. Connolly, D. G. Madden, A. E. H. Wheatley, D. Fairen-Jimenez, *J. Am. Chem. Soc.* **2020**, *142*, 8541 – 8549.

[7] J. Zheng, D. Barpaga, B. A. Trump, M. Shetty, Y. Fan, P. Bhattacharya, J. J. Jenks, C.-Y. Su, C. M. Brown, G. Maurin, B. P. McGrail, R. K. Motkuri, *J. Am. Chem. Soc.* **2020**, *142*, 3002 – 3012.

[8] O. M. Yaghi, M. O'Keeffe, N. W. Ockwig, H. K. Chae, M. Eddaoudi, J. Kim, *Nature* **2003**, *423*, 705 – 714.

[9] H. Li, M. Eddaoudi, M. O'Keeffe, O. M. Yaghi, *Nature* **1999**, *402*, 276 – 279.

[10] A. P. Côté, A. I. Benin, N. W. Ockwig, M. O'Keeffe, A. J. Matzger, O. M. Yaghi, *Science* **2005**, *310*, 1166 – 1170.

[11] X. Zhao, P. Pachfule, S. Li, T. Langenhahn, M. Ye, C. Schlesiger, S. Praetz, J. Schmidt, A. Thomas, *J. Am. Chem. Soc.* **2019**, *141*, 6623 – 6630.

[12] K. S. Deeg, D. D. Borges, D. Ongari, N. Rampal, L. Talirz, A. V. Yakutovich, J. M. Huck, B. Smit, *ACS Appl. Mater. Interfaces* **2020**, *12*, 21559 – 21568.

[13] S. Park, Z. Liao, B. Ibarlucea, H. Qi, H.-H. Lin, D. Becker, J. Melidonie, T. Zhang, H. Sahabudeen, L. Baraban, C.-K. Baek, Z. Zheng, E. Zschech, A. Fery, T. Heine, U. Kaiser, G. Cuniberti, R. Dong, X. Feng, *Angew. Chem. Int. Ed.* **2020**, *59*, 8218 – 8224; *Angew. Chem.* **2020**, *132*, 8295 – 8301.

[14] C. Krishnaraj, H. Sekhar Jena, L. Bourda, A. Laemont, P. Pachfule, J. Roeser, C. V. Chandran, S. Borgmans, S. M. J. Rogge, K. Leus, C. V. Stevens, J. A. Martens, V. Van Speybroeck, E. Breynaert, A. Thomas, P. Van Der Voort, *J. Am. Chem. Soc.* **2020**, *142*, 20107 – 20116.

[15] R. K. Sharma, P. Yadav, M. Yadav, R. Gupta, P. Rana, A. Srivastava, R. Zbořil, R. S. Varma, M. Antonietti, M. B. Gawande, *Mater. Horiz.* **2020**, *7*, 411 – 454.

[16] J. Keupp, R. Schmid, *Faraday Discuss.* **2018**, *211*, 79 – 101.

[17] A. L. Goodwin, *Nat. Commun.* **2019**, *10*, 4461.

[18] See Ref. [8].

[19] Y.-B. Zhang, J. Su, H. Furukawa, Y. Yun, F. Gándara, A. Duong, X. Zou, O. M. Yaghi, *J. Am. Chem. Soc.* **2013**, *135*, 16336 – 16339.

[20] C. S. Diercks, O. M. Yaghi, *Science* **2017**, *355*, eeal1585.

[21] J. Li, J. Sun, *Acc. Chem. Res.* **2017**, *50*, 2737 – 2745.

[22] J.-H. Dou, M. Q. Arguilla, Y. Luo, J. Li, W. Zhang, L. Sun, J. L. Mancuso, L. Yang, T. Chen, L. R. Parent, G. Skorupskii, N. J. Libretto, C. Sun, M. C. Yang, P. V. Dip, E. J. Brignole, J. T. Miller, J. Kong, C. H. Hendon, J. Sun, M. Dincă, *Nat. Mater.* **2020**, *20*, 222 – 228.

[23] V. Pecharsky, P. Zavalij, *Fundamentals of Powder Diffraction and Structural Characterization of Materials*, Springer US, **2009**.

[24] A. Le Bail, *Powder Diffr.* **2005**, *20*, 316 – 326.

[25] G. S. Pawley, *J. Appl. Crystallogr.* **1981**, *14*, 357 – 361.

[26] A. Le Bail, H. Duroy, J. L. Fourquet, *Mater. Res. Bull.* **1988**, *23*, 447 – 452.

[27] L. B. McCusker, *Acta Crystallogr. Sect. A* **1991**, *47*, 297 – 313.

[28] M. Falcioni, M. W. Deem, *J. Chem. Phys.* **1999**, *110*, 1754 – 1766.

[29] Q. Fang, J. Wang, S. Gu, R. B. Kaspar, Z. Zhuang, J. Zheng, H. Guo, S. Qiu, Y. Yan, *J. Am. Chem. Soc.* **2015**, *137*, 8352 – 8355.

[30] V. S. Vyas, M. Vishwakarma, I. Moudrakovski, F. Haase, G. Savasci, C. Ochsenfeld, J. P. Spatz, B. V. Lotsch, *Adv. Mater.* **2016**, *28*, 8749 – 8754.

[31] L. Bai, S. Z. F. Phua, W. Q. Lim, A. Jana, Z. Luo, H. P. Tham, L. Zhao, Q. Gao, Y. Zhao, *Chem. Commun.* **2016**, *52*, 4128 – 4131.

[32] S. Lin, C. S. Diercks, Y.-B. Zhang, N. Kornienko, E. M. Nichols, Y. Zhao, A. R. Paris, D. Kim, P. Yang, O. M. Yaghi, C. J. Chang, *Science* **2015**, *349*, 1208 – 1213.

[33] H. Li, Q. Pan, Y. Ma, X. Guan, M. Xue, Q. Fang, Y. Yan, V. Valtchev, S. Qiu, *J. Am. Chem. Soc.* **2016**, *138*, 14783 – 14788.

[34] G. Lin, H. Ding, R. Chen, Z. Peng, B. Wang, C. Wang, *J. Am. Chem. Soc.* **2017**, *139*, 8705 – 8709.

[35] G. Das, B. P. Biswal, S. Kandambeth, V. Venkatesh, G. Kaur, M. Addicoat, T. Heine, S. Verma, R. Banerjee, *Chem. Sci.* **2015**, *6*, 3931 – 3939.

[36] S.-Y. Ding, M. Dong, Y.-W. Wang, Y.-T. Chen, H.-Z. Wang, C.-Y. Su, W. Wang, *J. Am. Chem. Soc.* **2016**, *138*, 3031 – 3037.

[37] S. Dalapati, E. Jin, M. Addicoat, T. Heine, D. Jiang, *J. Am. Chem. Soc.* **2016**, *138*, 5797 – 5800.

[38] S. Wan, J. Guo, J. Kim, H. Ihee, D. Jiang, *Angew. Chem. Int. Ed.* **2009**, *48*, 5439 – 5442; *Angew. Chem.* **2009**, *121*, 5547 – 5550.

[39] X. Feng, L. Liu, Y. Honsho, A. Saeki, S. Seki, S. Irle, Y. Dong, A. Nagai, D. Jiang, *Angew. Chem. Int. Ed.* **2012**, *51*, 2618 – 2622; *Angew. Chem.* **2012**, *124*, 2672 – 2676.

[40] M. Calik, F. Auras, L. M. Salonen, K. Bader, I. Grill, M. Handloser, D. D. Medina, M. Dogru, F. Löbermann, D. Trauner, A. Hartschuh, T. Bein, *J. Am. Chem. Soc.* **2014**, *136*, 17802 – 17807.

[41] C. J. Doonan, D. J. Tranchemontagne, T. G. Glover, J. R. Hunt, O. M. Yaghi, *Nat. Chem.* **2010**, *2*, 235 – 238.

[42] L. A. Baldwin, J. W. Crowe, D. A. Pyles, P. L. McGrier, *J. Am. Chem. Soc.* **2016**, *138*, 15134 – 15137.

[43] X. Zhan, Z. Chen, Q. Zhang, *J. Mater. Chem. A* **2017**, *5*, 14463 – 14479.

[44] S. B. Alahakoon, S. D. Diwakara, C. M. Thompson, R. A. Smaldone, *Chem. Soc. Rev.* **2020**, *49*, 1344 – 1356.

[45] T. Ma, E. A. Kapustin, S. X. Yin, L. Liang, Z. Zhou, J. Niu, L.-H. Li, Y. Wang, J. Su, J. Li, X. Wang, W. D. Wang, W. Wang, J. Sun, O. M. Yaghi, *Science* **2018**, *361*, 48 – 52.

[46] P. J. Waller, F. Gándara, O. M. Yaghi, *Acc. Chem. Res.* **2015**, *48*, 3053 – 3063.

[47] D. Ongari, A. V. Yakutovich, L. Talirz, B. Smit, *ACS Cent. Sci.* **2019**, *5*, 1663 – 1675.

[48] H. M. El-Kaderi, J. R. Hunt, J. L. Mendoza-Cortés, A. P. Côté, R. E. Taylor, M. O'Keeffe, O. M. Yaghi, *Science* **2007**, *316*, 268 – 272.

[49] S. Wan, J. Guo, J. Kim, H. Ihee, D. Jiang, *Angew. Chem. Int. Ed.* **2008**, *47*, 8826 – 8830; *Angew. Chem.* **2008**, *120*, 8958 – 8962.

[50] B. J. Smith, N. Hwang, A. D. Chavez, J. L. Novotney, W. R. Dichtel, *Chem. Commun.* **2015**, *51*, 7532 – 7535.

[51] N. Keller, T. Sick, N. N. Bach, A. Koszalkowski, J. M. Rotter, D. D. Medina, T. Bein, *Nanoscale* **2019**, *11*, 23338 – 23345.

[52] L. Vanduyfhuys, S. Vandenbrande, T. Verstraelen, R. Schmid, M. Waroquier, V. Van Speybroeck, *J. Comput. Chem.* **2015**, *36*, 1015 – 1027.

[53] L. Vanduyfhuys, S. Vandenbrande, J. Wieme, M. Waroquier, T. Verstraelen, V. Van Speybroeck, *J. Comput. Chem.* **2018**, *39*, 999 – 1011.

[54] W. I. F. David, D. S. Sivia, *J. Appl. Crystallogr.* **2001**, *34*, 318 – 324.

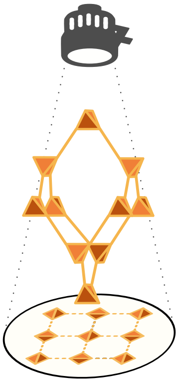
[55] A. M. Pütz, M. W. Terban, S. Bette, F. Haase, R. E. Dinnebier, B. V. Lotsch, *Chem. Sci.* **2020**, *11*, 12647 – 12654.

Manuscript received: December 25, 2020
Accepted manuscript online: January 25, 2021
Version of record online: March 8, 2021

Paper II

Exploring the phase stability in interpenetrated diamondoid covalent organic frameworks

Sander Borgmans, Sven M. J. Rogge*, Juul S. De Vos, Pascal Van Der Voort, and Veronique Van Speybroeck*



Communications Chemistry, **2023**, 6, 5

The Supporting Information is available at:
https://doi.org/10.1038/s42004-022-00808-y

J. S. De Vos generated *ab initio* reference data for each of the five building blocks in this work, from which cluster force fields were derived. By combining these SBUs on the **dia** topology, structural models were generated, together with a system-specific force field for each of them, and subsequently optimized. A consistent mathematical description of the embedding of largely interpenetrated **dia** topologies has been derived.

*Corresponding authors

communications chemistry

ARTICLE

https://doi.org/10.1038/s42004-022-00808-y OPEN

Exploring the phase stability in interpenetrated diamondoid covalent organic frameworks

Sander Borgmans[1], Sven M. J. Rogge[1✉], Juul S. De Vos[1], Pascal Van Der Voort[2] & Veronique Van Speybroeck[1✉]

Soft porous crystals, which are responsive to external stimuli such as temperature, pressure, or gas adsorption, are being extensively investigated for various technological applications. However, while substantial research has been devoted to stimuli-responsive metal-organic frameworks, structural flexibility in 3D covalent organic frameworks (COFs) remains ill-understood, and is almost exclusively found in COFs exhibiting the diamondoid (**dia**) topology. Herein, we systemically investigate how the structural decoration of these 3D **dia** COFs—their specific building blocks and degree of interpenetration—as well as external triggers such as temperature and guest adsorption may promote or suppress their phase transformations, as captured by a collection of 2D free energy landscapes. Together, these provide a comprehensive understanding of the necessary conditions to design flexible diamondoid COFs. This study reveals how their flexibility originates from the balance between steric hindrance and dispersive interactions of the structural decoration, thereby providing insight into how new flexible 3D COFs can be designed.

---

[1] Center for Molecular Modeling (CMM), Ghent University, Technologiepark-Zwijnaarde 46, 9052 Zwijnaarde, Belgium. [2] Center for Ordered Materials, Organometallics and Catalysis (COMOC), Department of Inorganic and Physical Chemistry, Ghent University, Krijgslaan 281 (S3), 9000 Gent, Belgium. ✉email: Sven.Rogge@UGent.be; Veronique.VanSpeybroeck@UGent.be

Responsive nanoporous materials can undergo structural phase transitions between different crystalline metastable phases under external stimuli such as temperature, pressure, or adsorption[1]. As phase transitions impact the internal pore architecture of these so-called soft porous crystals (SPCs), these materials are highly sought-after for applications in gas and fluid capture, gas separation, heterogeneous catalysis, nanosensing and drug delivery[2–13]. Thanks to extensive experimental efforts, the range of SPCs has expanded appreciably in recent years, especially for the class of metal-organic framework (MOF) materials[14–16]. Taking MIL-47 and MIL-53 as textbook exemplars[17,18], their well-known winerack topology potentially allows for transitions in a well-defined range of external conditions. However, the actual occurrence of phase transitions under a given set of thermodynamic conditions is uniquely determined by the relative stability of its different phases and the free energy barriers between them. This is a general observation for SPCs: while their topology dictates the potential for flexibility, the degree of interpenetration and the building blocks that decorate this topology—or structural decoration—defines whether or not this flexibility comes to expression. To expand the range of possible SPCs, it is, therefore, key to understand to which extent this structural decoration impacts the relative stability of the different metastable phases in an a priori flexible topology. While several studies have sought to investigate the effect of altering the linker or metal ion in SPCs[19–21], the impact of interpenetration, i.e., the occurrence of two or more individual networks catenated with each other[22], on the relative phase stability in SPCs remains ill-understood despite the vast amount of interpenetrated framework structures. In this work, we, therefore, aim to provide computational guidelines to alter the phase stability and hence phase transition behaviour of SPCs by controlling their degree of interpenetration. To this end, we will quantify the impact of interpenetration on the relative stability of four framework materials exhibiting the flexible diamondoid (**dia**) topology and investigate how external triggers alter this stability.

Although recent years have witnessed a strong increase in the number of soft porous MOF crystals, the number of covalent organic frameworks (COFs) with large-amplitude flexibility under external stimuli remains extremely limited[23–28]. While this may seem plausible given that, in contrast to MOFs, COFs are framework materials that are completely built up from strong covalent bonds[29], the few COFs that do exhibit flexibility shed light on how these strong directed bonds may result in 'soft' materials nonetheless. Remarkably, all flexible COFs found to date almost exclusively exhibit the **dia** topology, which is also regularly encountered in flexible MOFs[30,31]. An example of this is COF-300[32,33], for which the **dia** topology, the non-interpenetrated structure, and its building blocks are visualised in Figs. 1, 2a. Similar to the majority of diamondoid frameworks, COF-300 is interpenetrated such that the individual nets are translated along the twofold rotation axis (the [001] crystal axis in Fig. 2b), belonging to the so-called interpenetration class Ia[34]. When viewed along this axis, the material exhibits large square-shaped (SQ) channels, independent of the degree of interpenetration, as visualised in Fig. 2c. Experiments have shown that, besides these SQ channels, the channel network in diamondoid COFs can adopt various sizes and shapes (see Fig. 2c), either due to different synthesis protocols or the presence of guest molecules. This gives rise to smaller square-shaped channels (sq) or rectangular channels (rect), respectively, as reported in Supplementary Table 1 in Supplementary Note 2. Aside from these variations, other, possibly more complex, material distortions can manifest depending on the external stimuli, such as a collective shearing or individual tilting of the tetratopic nodes[35,36]. However, as these additional material distortions have not yet been associated with flexible diamondoid COFs experimentally, and accounting for their contributions to the phase stability would

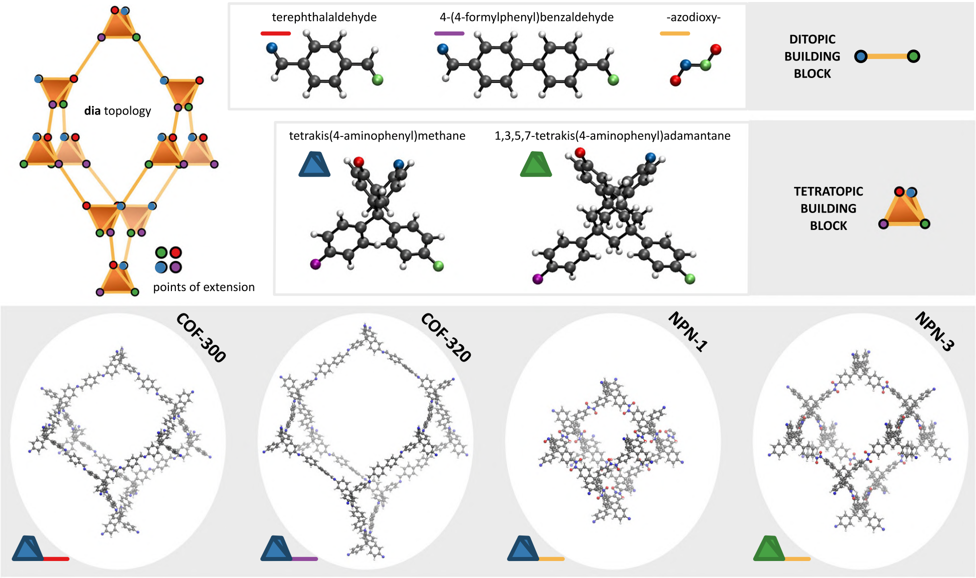


Fig. 1 Illustration of the four considered materials, and their ditopic and tetratopic building blocks. As the NPN COFs form an azodioxy linkage through selfcondensation during synthesis, the ditopic building block is labelled as such instead of its precursor. Additionally, the four unique points of extension for each tetratopic building block, coinciding with the nitrogen atoms, have been indicated by the red, green, blue and purple colours.

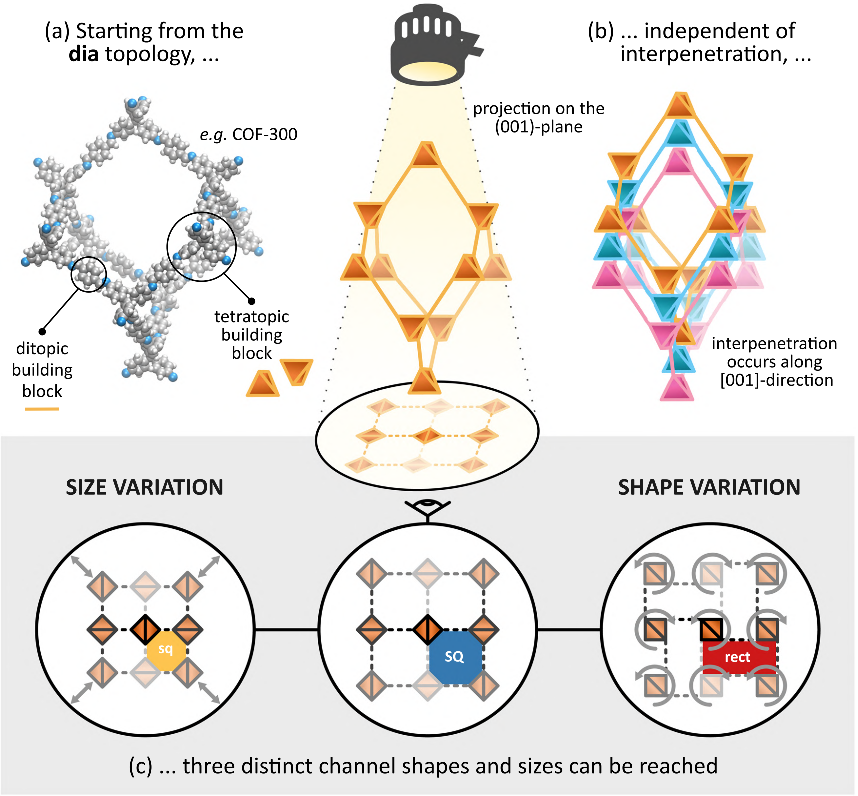


Fig. 2 Illustration of the dia topology and its experimental channel behaviour. The dia topology is elaborated by depicting the **a** non-interpenetrated atomic representation of COF-300 and **b** a threefold interpenetrated diamondoid framework. **c** When viewed along the ideal twofold rotation axis ([001]-direction), three distinct channel shapes and sizes emerge, namely, sq, SQ and rect, independent of the degree of interpenetration. These channel shapes and sizes are defined by the relative orientation and distance between adjacent tetratopic building blocks.

require substantially more complex models, they are not explicitly taken into account in this work. As such, in general, three distinct channel shapes can be identified in the diamondoid topology, namely, SQ, sq, and rect. While the observation of different phases at different degrees of interpenetration and experimental conditions alludes to a correlation between them, systematic insight into how the structural decoration and external triggers affect the phase stability in diamondoid COFs is still lacking. Moreover, as evident from the broad variety in the experimental channel shape and size information in Supplementary Table 1, a comprehensive theoretical understanding of the phase behaviour would offer a reliable reference framework for experiments. Therefore, we will herein determine the occurrence and relative stability of these different phases for four **dia** COFs. As illustrated in Fig. 1, these materials represent a versatile set with fundamentally different linkers (imine/azodioxy) and tetratopic building blocks (methane/adamantane). COF-300 and COF-320[37] differ only in the length of their linker, with COF-300 incorporating the shorter linker. Both NPN-1[38] and NPN-3[38] contain a much shorter and more rigid linker; the latter has a more rigid tetratopic building block containing an adamantane cage instead of a tetraphenylmethane moiety. While single-crystal-to-single-crystal structural transitions have been reported in the literature for COF-300 and COF-320, no such transitions have been reported for either NPN-1 or NPN-3 at the moment (see Supplementary Table 1). In this work, we aim to understand and verify this striking difference, and its connection to their fundamentally different structural decoration. Together, these four **dia** COFs have been observed for a wide range of interpenetration, from four to nine interpenetrating nets, thereby forming a varied set of materials to derive general relations between the interpenetration and flexibility in framework materials exhibiting the **dia** topology.

Results

Choice of collective variables to map the phase transition. From a fundamental point of view, identifying the occurrence of different metastable phases, their transition barrier, and the corresponding transition mechanism, requires knowledge of the free energy of these phases and the pathways connecting them, and this at the thermodynamic conditions and degree of interpenetration of interest. Such a free energy landscape is expressed as a function of so-called collective variables (CVs) that

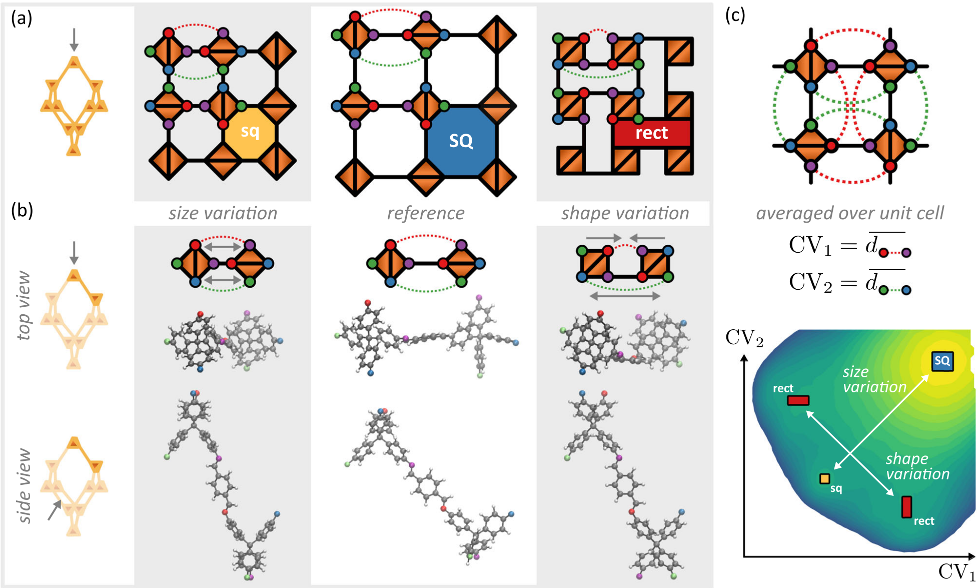


Fig. 3 Definition of the collective variables and the corresponding transition mechanism between the three distinct phases, SQ, sq and rect. a The distinct phases, where the points of extensions have been indicated by coloured circles, in correspondence with Fig. 1. **b** The variations in channel size and shape, relative to the reference, can be defined via two types of distances, indicated by the red and green dotted lines. Pure size variations only occur when both distances increase or decrease simultaneously, whereas shape variations are introduced when the distances change differently. Their atomic representations, in both top and side view, for the case of COF-300, are provided to illustrate the corresponding transition mechanism. This clearly visualises that a size decrease elongates the dia cage in the [001] direction while compressing it in the perpendicular directions, while a shape change is facilitated by rotating tetratopic building blocks. **c** Through averaging the two types of distances in panels **a** and **b** over the unit cell, two collective variables are obtained, capable of describing the simultaneous occurrence of all phases and validating the choice of CVs, as demonstrated in the free energy surface for fourfold interpenetrated COF-300.

distinguish between all potential phases. For the three **dia** phases, this means that the CVs should be able to describe both variations in shape and size of the originally square-shaped channels, so as to describe the SQ, sq, and rect phases. From the projection of these three phases on the (001) plane, shown in Fig. 2c, it is clear that such CVs can be constructed from the relative projected positions of the points of extension that connect adjacent tetratopic building blocks in the structure. To see this, Fig. 3a visualises the three **dia** phases for which the four points of extension of the tetratopic building blocks have been indicated with coloured circles, in correspondence to Fig. 1. For two adjacent vertices, shown in Fig. 3b, both shape and size variations can be distinguished based on the two distances indicated with red and green dotted lines. These distances are calculated between projected points of extensions that are not directly connected within the same tetratopic building block. When these distances either shrink or expand simultaneously, they describe a size variation defining the transition between the sq and SQ phases. In contrast, when these distances do not vary identically, they yield an additional shape variation, giving rise to transitions between the SQ and rect phases or between the sq and rect phases. From the atomic representations of each of the transition mechanisms in Fig. 3b, it is clear that size variations elongate or compress the diamondoid cage in the [001] direction, whereas shape changes are facilitated by a complementary rotation of two adjacent tetratopic building blocks. For the whole unit cell, containing four tetratopic building blocks, as elaborated in Supplementary Note 1, a phase transition requires a concerted motion of all four building units. Therefore, the set of CVs used throughout this work are obtained by considering the two aforementioned distances, averaged over all such distances present in the unit cell. These CVs correlate well with macroscopic system properties, as outlined in Supplementary Note 3 and Supplementary Figs. 1, 2. As shown in Fig. 3c, these CVs succeed in describing the relative stability of the different phases in the fourfold interpenetrated COF-300 at 300 K and will therefore be adopted in the remainder of this work.

Free energy landscape as a function of structural decoration. Having established that the two chosen collective variables can distinguish between all **dia** phases, we first aim to investigate how interpenetration alters the relative stability of these phases. Intuitively, an increasing degree of interpenetration implies a stronger steric repulsion as it increases the materials' density and limits the accessible space around the building blocks. However, as we will demonstrate later in this work, interpenetration impacts the stability of the different phases differently, thereby showing the potential to also promote the existence of multiple phases and phase transitions. Hence, interpenetration can fundamentally alter to what extent the COF is responsive to external stimuli such as temperature, pressure, or gas adsorption. Consequently, as the degree of interpenetration may be modified by varying, e.g., the synthesis conditions[39–41], interpenetration allows for control over the range of thermodynamic conditions for which the material shows flexibility. To outline this effect on the range of possible channel shapes and sizes, the degree of interpenetration is varied for the considered materials from onefold (no interpenetration) to eightfold (or elevenfold for COF-320). The maximal degree of interpenetration corresponds to the point at which the individual nets are so close together that they start to overlap, resulting in prohibitively large steric repulsion.

Focusing on the archetypal COF-300, Fig. 4 showcases that the relative stability of the distinct phases is indeed substantially impacted depending on the interpenetration, which is clear when considering the variations in the free energy landscapes at 300 K. For the non-interpenetrated COF-300, a large accessible phase space is observed with a single broad minimum for a large square-shaped channel (SQ). When introducing a second catenated net, evidently, the accessible phase space decreases due to the steric crowding, which destabilises small channel sizes. However, the dispersive interactions between the nets counteract this, resulting in the emergence of a metastable small square-shaped channel phase (sq). This becomes even more pronounced when adding a third net, where the metastable sq phase is pushed towards a slightly larger channel size, and metastable rectangular shapes (rect) emerge. Ultimately, when adding a fourth net, the balance between steric hindrance and dispersive interactions shifts. This suppresses the manifestation of small square-shaped channels and pushes the metastable state towards a rectangular shape. In this way, the attractive dispersion is maximised while avoiding steric crowding around the building blocks. Similarly, looking at a fivefold interpenetration, the metastable rect phases are pushed towards an even larger size and shape anisotropy, until the metastable phases completely disappear for all degrees of interpenetration higher than five. Further increasing the degree of interpenetration at this point only limits the accessible phase space, driven by the steric hindrance. As such, by adjusting the degree of interpenetration, the size and shape variations of each emerging phase can be governed, thus controlling the compliance of the 1D channel shape, although the SQ phase remains the most stable under these conditions.

Similar free energy landscapes for COF-300 at other temperatures (50, 100 and 400 K), as well as for the remaining materials, namely, COF-320 (at 50, 100, 300 and 400 K), NPN-1 (at 300 K) and NPN-3 (at 300 K), are reported in Supplementary Figs. 6–15 in Supplementary Note 5. In Fig. 5, the main features of these free energy landscapes are summarised for COF-300 and COF-320, as a function of the degree of interpenetration and temperature. It depicts the free energy differences between the available metastable states and the stable SQ state in the top panels, and their corresponding transition barriers in the bottom panels. Notably, no data for the NPN COFs is shown, as our simulations demonstrate that the structural decoration of these materials, especially their short and rigid linker, precludes flexibility. This gives rise to only a single stable SQ phase for NPN-1 and NPN-3 regardless of interpenetration or temperature, despite having the same **dia** topology as the flexible COF-300 and COF-320. Similarly, as the structural decoration of the highly interpenetrated (and the non-interpenetrated) frameworks inhibits the materialisation of more than one phase, this data is omitted as well.

Figure 5 demonstrates that, at low degrees of interpenetration, the metastable sq phase appears for COF-300 and COF-320, irrespective of the temperature. This sq phase later disappears for intermediate degrees of interpenetration in favour of the metastable rect phase, which finally disappears as well at higher degrees of interpenetration, leaving only the SQ phase. Moreover, while the transition process between the SQ phase to any metastable phase is highly activated with free energy barriers exceeding 100 kJ mol^{-1}, the corresponding transition barriers generally decrease with increasing degree of interpenetration, until steric crowding makes the small channels inaccessible, in line with the increasing Pauli repulsion between the catenated nets. Conversely, no clear trend is observed when considering the reverse transition barrier height, from any metastable phase to the SQ phase, as a function of the interpenetration. Interestingly, Fig. 5 does show significant reverse transition barriers to the SQ phase, going as high as 70 kJ mol^{-1} for fivefold interpenetrated COF-320 at 100 K. Evidently, Fig. 5 allows us to extract general

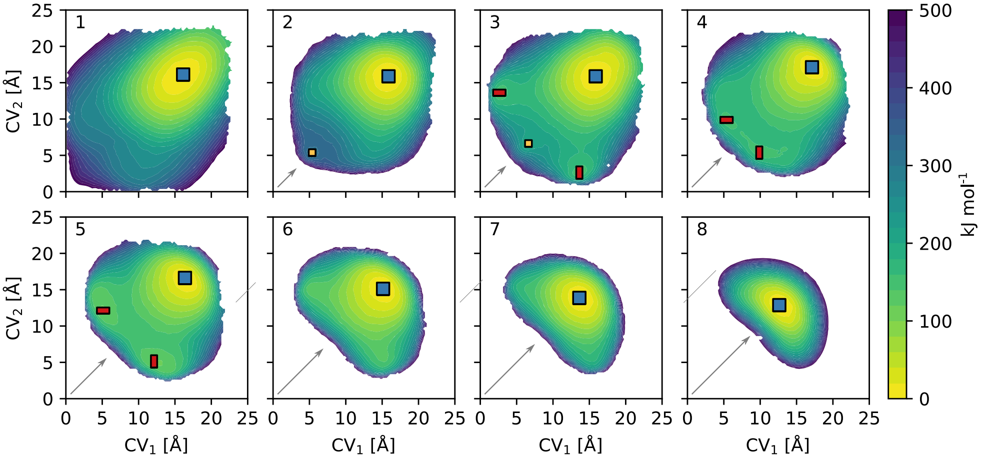


Fig. 4 Free energy surfaces for COF-300 at 300 K as a function of interpenetration. The degree of interpenetration varies from one (non-interpenetrated) to eightfold interpenetrated, as indicated in the top left corner of each subpanel. For each subpanel, the different (meta)stable phases are indicated. Regions with free energy exceeding the free energy minimum by more than 500 kJ mol^{-1} are omitted from the plot, as they are not accessible under realistic conditions. The arrows indicate the growing steric crowding with increasing interpenetration.

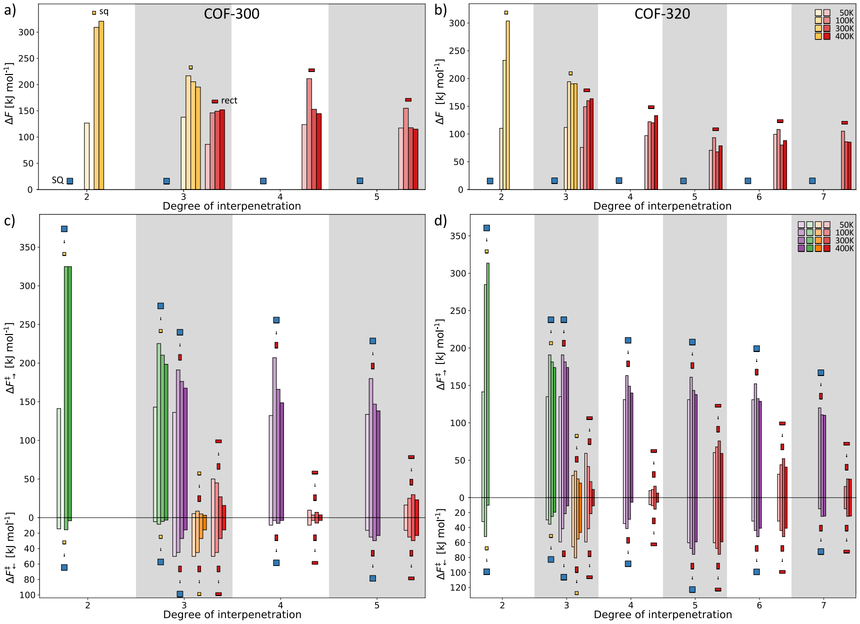


Fig. 5 Overview of the relative phase stability as a function of the degree of interpenetration and temperature. The phase stability is visualised through the free energy difference of each (meta)stable phase with respect to the most stable SQ phase for **a** COF-300 and **b** COF-320. Additionally, to illustrate the stability of each phase, the transition barrier height between them is illustrated as well, for **c** COF-300 and **d** COF-320. The temperature effect is visualised through the colour shade, going from 50 K (light) to 400 K (dark), whereas the colour relates to the relevant phase symbols. The different symbols introduced in Fig. 3 are used to distinguish between the different phases.

trends, but it abstractifies the size and shape anisotropy variations that are visible in the free energy landscapes, as depicted in Fig. 4. Importantly, when comparing these variations in Supplementary Figs. 6–15 as a function of the structural decoration, the NPN COFs do exhibit similar stabilisation and destabilisation effects as COF-300 and COF-320, with a very similar shape of the accessible free energy landscape, although the stabilisation effects do not give rise to additional (meta)stable phases.

Finally, while it is clear that the degree of interpenetration and the temperature impart significant variations to the transition barriers, the SQ phase remains the most stable phase for all considered materials. However, as reported in Supplementary Table 1, various channel shapes emerge in diamondoid COFs when exposed to different guest molecules, such as water. As such, we herein sought to explore how the inclusion of water impacts the structure and relative stability of the different phases in COF-300(7), which experimentally amounts to a volume contraction of about 5% for the SQ phase[26–28]. Similar to Fig. 4, the effect of water loading is observed by analysing the variations in the free energy landscapes, when increasing the loading from one to 16 water molecules per channel. This is illustrated in Fig. 6, where the loading is indicated in the top left corner of each pane. From the final panel, it is clear that the SQ channel first decreases in size when water is present in the pores, with a minimum volume for a loading of eight molecules per channel, after which the size increases again. Additionally, a decrease in the available phase space can be observed, due to the density increase, similar to the interpenetration effect. When comparing the most contracted water-filled framework to the guest-free framework, the contraction corresponds to a change (expressed as a function of the collective variables) from 13.9 Å × 14.1 Å to 13.1 Å × 13.6 Å, or, expressed in terms of the unit cell volume, from circa 5594 $Å^3$ to circa 5275 $Å^3$ (−6%). While these values differ somewhat from those observed experimentally, as tabulated in Supplementary Table 6, the relative change in volume upon water adsorption is very similar to the values from those references mentioned in Supplementary Table 1, succeeding in a qualitative description of the experimentally observed pore shrinkage upon water adsorption. This adsorption-induced shrinkage is substantially smaller than the potential temperature-induced flexibility between the SQ and sq phases for the empty framework predicted in Fig. 4.

Discussion

The characterisation of responsive nanoporous materials, their internal transition mechanisms, and their corresponding operating range is essential to understand, improve and apply them in an industrial context. Only through the identification of responsive structural decorations, and the relevant triggers that can potentially induce transitions, one can systematically expand the collection of functional SPCs tailored to specific applications. Sparked by a lack of such information for COFs, we herein investigated the effect of different structural decorations—the building blocks and the degree of interpenetration—in the inherently flexible **dia** topology on the potential for phase transformations. To this end, the different experimental phases (sq, SQ and rect) that emerge when viewing the structure along the channel direction were characterised, and their internal transition mechanism was exposed as the cooperative rotation

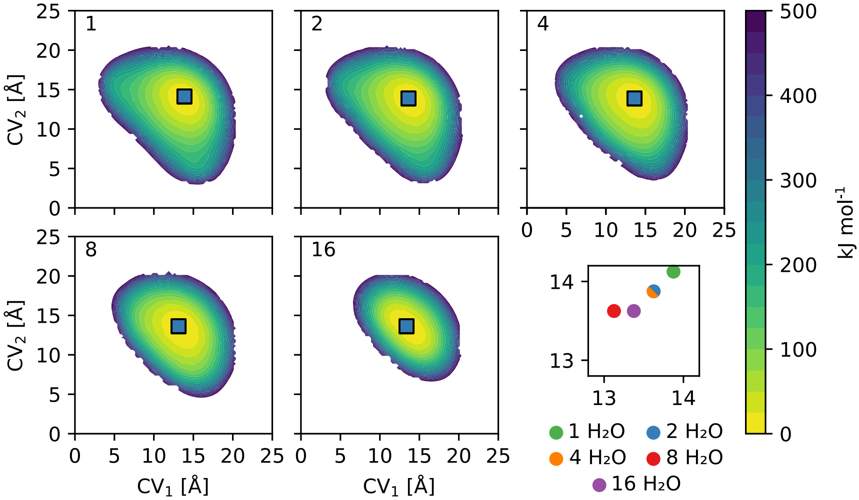


Fig. 6 Free energy surfaces for COF-300(7) at 300 K as a function of water loading. The water loading is exponentially varied from one to 16 molecules per channel, as indicated in the top left corner of each subpanel, with four channels per unit cell. The different (meta)stable phases are indicated in the figure. Regions with free energy exceeding the free energy minimum by more than 500 kJ mol^{-1} are omitted from the plot, as they are not accessible under realistic conditions. The final subpanel shows the exact position of the SQ phase for each loading in detail.

and translation of neighbouring tetratopic building blocks perpendicular to the channel direction.

The relative phase stability for four **dia** materials was characterised using two-dimensional free energy landscapes, capturing both size and shape variations of the channels, and their respective transitions. From our enhanced sampling molecular dynamics calculations, clear trends could be observed. While certain structural decorations, such as the rigid azodioxy linker and adamantane cage in the NPN COFs, preclude flexibility, others, such as the degree of interpenetration, can fundamentally shift the delicate equilibrium between the stabilisation and destabilisation effects, driving the phase equilibrium. These (de)stabilisation effects originate from the steric hindrance and dispersive interactions, which disfavour and promote closely packed phases, respectively.

Depending on the number of catenated nets, metastable phases arise in COF-300 and COF-320, stabilised by attractive dispersive interactions between the nets, that become unstable for large degrees of interpenetration due to steric crowding. This gives rise to an optimal, intermediate degree of interpenetration for each metastable phase, with significant transition barriers that could trap the structure in a certain phase. However, by varying the temperature, we observed that the temperature cannot be straightforwardly used as a driving force to completely destabilise the SQ phase. Instead, through introducing water molecules, significant size variations could be induced, shifting the SQ phase towards lower volumes. Finally, regardless of the structural decoration and the considered external triggers, a large square-shaped channel was identified as the most stable phase, warranting more investigation to explore whether building blocks can be designed that destabilise this phase. Our results predict that such a building block should be sufficiently long to allow for stabilisation of the sq or rect phase before steric hindrance destabilises these phases again. Moreover, the here-derived workflow can easily be extended to probe the effect of (anisotropic) stress as a potential trigger towards certain phases, introducing new application avenues, such as a configurable sensing device with adaptable channels.

Methods

Structure generation. All initial materials were generated in silico using our in-house structure assembly software, which is based on a top-down approach starting from the building blocks of the material and the underlying topology. A detailed procedure of this protocol can be found in Supplementary Note 1, together with a mathematical description of the interpenetrated **dia** topologies.

Computational details. The enhanced sampling molecular dynamics (MD) simulations were performed in the (N, P, $\boldsymbol{\sigma_a} = \mathbf{0}$, T) ensemble[42] using the in-house software package Yaff[43] as the MD engine, employing a velocity Verlet integration scheme with a timestep of 0.5 fs. The temperature was fixed at 300 K with a Nosé-Hoover chain thermostat[44–46] with three beads and a relaxation time of 100 fs. The pressure was controlled at 1 bar by a Martyna–Tuckerman–Tobias–Klein barostat[47,48] with a relaxation time of 1000 fs. The interatomic interactions were modelled through system-specific force fields, derived in accordance with the general QuickFF 2.2.4 protocol[49,50], using the default parametrization (see Supplementary Note 6 and Supplementary Figs. 17–21 therein for the force field derivation and validation). The electrostatic interactions were computed with an Ewald summation using a real space cut-off r_{cut} of 15 Å, a scaling factor α equal to 0.213 Å^{-1}, and a reciprocal space cut-off $k_{\max}$ of 0.320 Å^{-1}. Analogously, the van der Waals interactions were calculated with a real space cut-off of 15 Å. Both cut-offs were smoothed using a truncation model. For the water loading simulations, the TIP4P/2005f force field was used[51,52], as elaborated in Supplementary Note 6.5.

To enhance the exploration of the phase space and access all possible phases, umbrella sampling was adopted using a harmonic bias potential as a function of the collective variables as defined in Fig. 3. For the CVs, a uniform grid with a spacing of 1 Å was employed, extended with small patches where a spacing of 0.5 Å was used to accurately reproduce the free energy barriers in regions where neighbouring simulations did not overlap adequately, as reported in Supplementary Tables 2–5. The force constants for the harmonic bias potential (κ_i) of both CVs were initialised at 10 kJ mol^{-1} Å^{-2}, and increased until a maximum of 150 kJ mol^{-1} Å^{-2} for those grid points where the simulation deviated from the umbrella centre (due to a steep free energy profile). Each individual simulation had a run time of 15 ps, including an equilibration time of 5 ps. The resulting free energy from the collection of simulations was calculated with the weighted histogram analysis method (WHAM)[53], using the software package from ref. [54]. Since the WHAM code failed to converge in situations where the sampled two-dimensional phase space was not rectangular, an iterative procedure was applied to combine smaller rectangular regions, as described in Supplementary Figs. 3–5 of Supplementary Note 4. This procedure was validated through the application of our in-house WHAM code in ThermoLIB, which did not suffer from convergence issues.

The resulting free energy surfaces were symmetrized in accordance with the symmetry of the collective variables, ensuring smoother profiles for the minimal free energy path (MFEP) analysis. The MFEP analysis was performed with the Minimum Energy Path Surface Analysis (MEPSA) code[55], as elaborated in Supplementary Note 5.3. The barrier heights in Fig. 5 for COF-300 and COF-320 were derived from the one-dimensional MFEPs as the difference between the local minimum and the neighbouring local maximum, providing a clear measure of the transition probability, as illustrated in Supplementary Fig. 16.

Data availability

The relevant input files and computational data which generated the results of this work are available from the online GitHub repository at https://github.com/SanderBorgmans/SupportingInformation or upon request from the corresponding authors.

Code availability

The Gaussian code, used to perform the ab initio cluster calculations, can be licensed from Gaussian, Inc. (see https://gaussian.com/). The QuickFF (used to derive the force fields), Yaff (MD engine) and ThermoLIB (free energy evaluation engine) software packages are freely accessible via https://molmod.ugent.be/software/. The MEPSA code for the minimal free energy path analysis is freely available at http://bioweb.cbm.uam.es/software/MEPSA/. Representative input and processing scripts are available at https://github.com/SanderBorgmans/SupportingInformation.

Received: 22 July 2022; Accepted: 29 December 2022;

References

1. Horike, S., Shimomura, S. & Kitagawa, S. Soft porous crystals. *Nat. Chem.* **1**, 695–704 (2009).
2. Dybtsev, D. N., Chun, H. & Kim, K. Rigid and flexible: a highly porous metal–organic framework with unusual guest-dependent dynamic behavior. *Angew. Chem. Int. Ed.* **43**, 5033–5036 (2004).
3. Chen, B., Xiang, S. & Qian, G. Metal-organic frameworks with functional pores for recognition of small molecules. *Acc. Chem. Res.* **43**, 1115–1124 (2010).
4. Liu, Y., Xuan, W. & Cui, Y. Engineering homochiral metal-organic frameworks for heterogeneous asymmetric catalysis and enantioselective separation. *Adv. Mater.* **22**, 4112–4135 (2010).
5. Férey, G. et al. Why hybrid porous solids capture greenhouse gases? *Chem. Soc. Rev.* **40**, 550–562 (2011).
6. Mohideen, M. I. H. et al. Protecting group and switchable pore-discriminating adsorption properties of a hydrophilic–hydrophobic metal–organic framework. *Nat. Chem.* **3**, 304–310 (2011).
7. Horcajada, P. et al. Metal–organic frameworks in biomedicine. *Chem. Rev.* **112**, 1232–1268 (2012).
8. Wang, J.-H., Li, M. & Li, D. A dynamic, luminescent and entangled MOF as a qualitative sensor for volatile organic solvents and a quantitative monitor for acetonitrile vapour. *Chem. Sci.* **4**, 1793–1801 (2013).
9. Mason, J. A. et al. Methane storage in flexible metal–organic frameworks with intrinsic thermal management. *Nature* **527**, 357–361 (2015).
10. Yuan, S. et al. Flexible zirconium metal-organic frameworks as bioinspired switchable catalysts. *Angew. Chem. Int. Ed.* **55**, 10776–10780 (2016).
11. Li, L. et al. Flexible–robust metal–organic framework for efficient removal of propyne from propylene. *J. Am. Chem. Soc.* **139**, 7733–7736 (2017).
12. Taylor, M. K. et al. Near-perfect CO_2/CH_4 selectivity achieved through reversible guest templating in the flexible metal–organic framework Co(bdp). *J. Am. Chem. Soc.* **140**, 10324–10331 (2018).
13. Krause, S., Hosono, N. & Kitagawa, S. Chemistry of soft porous crystals: structural dynamics and gas adsorption properties. *Angew. Chem. Int. Ed.* **59**, 15325–15341 (2020).
14. Schneemann, A. et al. Flexible metal–organic frameworks. *Chem. Soc. Rev.* **43**, 6062–6096 (2014).
15. Fraux, G. & Coudert, F.-X. Recent advances in the computational chemistry of soft porous crystals. *Chem. Commun.* **53**, 7211–7221 (2017).
16. Rogge, S. M. J., Waroquier, M. & Van Speybroeck, V. Unraveling the thermodynamic criteria for size-dependent spontaneous phase separation in soft porous crystals. *Nat. Commun.* **10**, 1–10 (2019).
17. Barthelet, K., Marrot, J., Riou, D. & Férey, G. A breathing hybrid organic–inorganic solid with very large pores and high magnetic characteristics. *Angew. Chem. Int. Ed.* **41**, 281–284 (2002).
18. Serre, C. et al. Very large breathing effect in the first nanoporous chromium(III)-based solids: MIL-53 or CrIII(OH)·{O2C-C6H4-CO2}·{HO2C-C6H4-CO2H}x·H2Oy. *J. Am. Chem. Soc.* **124**, 13519–13526 (2002).
19. Yot, P. G. et al. Impact of the metal centre and functionalization on the mechanical behaviour of MIL-53 metal–organic frameworks. *Eur. J. Inorg. Chem.* **2016**, 4424–4429 (2016).
20. Vanduyfhuys, L. et al. Thermodynamic insight into stimuli-responsive behaviour of soft porous crystals. *Nat. Commun.* **9**, 1–9 (2018).
21. Bitzer, J., Teubnerová, M. & Kleist, W. Increasing the complexity in the MIL-53 structure: the combination of the mixed-metal and the mixed-linker concepts. *Chem. Eur. J.* **27**, 1724–1735 (2020).
22. Gong, Y.-N., Zhong, D.-C. & Lu, T.-B. Interpenetrating metal–organic frameworks. *CrystEngComm* **18**, 2596–2606 (2016).
23. Zhang, J.-P., Lin, Y.-Y., Zhang, W.-X. & Chen, X.-M. Temperature- or guest-induced drastic single-crystal-to-single-crystal transformations of a nanoporous coordination polymer. *J. Am. Chem. Soc.* **127**, 14162–14163 (2005).
24. Ma, Y.-X. et al. A dynamic three-dimensional covalent organic framework. *J. Am. Chem. Soc.* **139**, 4995–4998 (2017).
25. Yang, Q.-Y. et al. Reversible switching between highly porous and nonporous phases of an interpenetrated diamondoid coordination network that exhibits gate-opening at methane storage pressures. *Angew. Chem. Int. Ed.* **57**, 5684–5689 (2018).
26. Ma, T. et al. Single-crystal x-ray diffraction structures of covalent organic frameworks. *Science* **361**, 48–52 (2018).
27. Chen, Y. et al. Guest-dependent dynamics in a 3D covalent organic framework. *J. Am. Chem. Soc.* **141**, 3298–3303 (2019).
28. Sun, T., Wei, L., Chen, Y., Ma, Y. & Zhang, Y.-B. Atomic-level characterization of dynamics of a 3D covalent organic framework by cryo-electron diffraction tomography. *J. Am. Chem. Soc.* **141**, 10962–10966 (2019).
29. Côté, A. P. et al. Porous, crystalline, covalent organic frameworks. *Science* **310**, 1166–1170 (2005).
30. Carrington, E. J. et al. Solvent-switchable continuous-breathing behaviour in a diamondoid metal–organic framework and its influence on CO_2 versus CH_4 selectivity. *Nat. Chem.* **9**, 882–889 (2017).
31. Yang, H. et al. Visualizing structural transformation and guest binding in a flexible metal–organic framework under high pressure and room temperature. *ACS Cent. Sci.* **4**, 1194–1200 (2018).
32. Uribe-Romo, F. J. et al. A crystalline imine-linked 3-D porous covalent organic framework. *J. Am. Chem. Soc.* **131**, 4570–4571 (2009).
33. Ma, T. et al. Observation of interpenetration isomerism in covalent organic frameworks. *J. Am. Chem. Soc.* **140**, 6763–6766 (2018).
34. Blatov, V. A., Carlucci, L., Ciani, G. & Proserpio, D. M. Interpenetrating metal–organic and inorganic 3D networks: a computer-aided systematic investigation. Part I. Analysis of the Cambridge structural database. *CrystEngComm* **6**, 377–395 (2004).
35. Beanland, R. Structure of planar defects in tilted perovskites. *Acta Crystallogr. Sect. A* **67**, 191–199 (2011).
36. Boström, H. L. B. Tilts and shifts in molecular perovskites. *CrystEngComm* **22**, 961–968 (2020).
37. Zhang, Y.-B. et al. Single-crystal structure of a covalent organic framework. *J. Am. Chem. Soc.* **135**, 16336–16339 (2013).
38. Beaudoin, D., Maris, T. & Wuest, J. D. Constructing monocrystalline covalent organic networks by polymerization. *Nat. Chem.* **5**, 830–834 (2013).
39. Jiang, H.-L., Makal, T. A. & Zhou, H.-C. Interpenetration control in metal–organic frameworks for functional applications. *Coord. Chem. Rev.* **257**, 2232–2249 (2013).
40. Elsaidi, S. K. et al. Putting the squeeze on CH_4 and CO_2 through control over interpenetration in diamondoid nets. *J. Am. Chem. Soc.* **136**, 5072–5077 (2014).
41. Gupta, M. & Vittal, J. J. Control of interpenetration and structural transformations in the interpenetrated MOFs. *Coord. Chem. Rev.* **435**, 213789 (2021).
42. Rogge, S. M. J. et al. A comparison of barostats for the mechanical characterization of metal-organic frameworks. *J. Chem. Theory Comput.* **11**, 5583–5597 (2015).
43. Verstraelen, T., Vanduyfhuys, L., Vandenbrande, S. & Rogge, S. M. J. Yaff, yet another force field. http://molmod.ugent.be/software/ (2022).
44. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268 (1984).
45. Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).
46. Martyna, G. J., Klein, M. L. & Tuckerman, M. Nosé-Hoover chains: the canonical ensemble via continuous dynamics. *J. Chem. Phys.* **97**, 2635–2643 (1992).
47. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189 (1994).

48. Martyna, G. J., Tuckerman, M. E., Tobias, D. J. & Klein, M. L. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.* **87**, 1117–1157 (1996).
49. Vanduyfhuys, L. et al. QuickFF: a program for a quick and easy derivation of force fields for metal-organic frameworks from ab initio input. *J. Comput. Chem.* **36**, 1015–1027 (2015).
50. Vanduyfhuys, L. et al. Extension of the QuickFF force field protocol for an improved accuracy of structural, vibrational, mechanical and thermal properties of metal-organic frameworks. *J. Comput. Chem.* **39**, 999–1011 (2018).
51. Ghosh, P., Kim, K. C. & Snurr, R. Q. Modeling water and ammonia adsorption in hydrophobic metal–organic frameworks: single components and mixtures. *J. Phys. Chem. C.* **118**, 1102–1110 (2014).
52. Sun, Y. et al. High-rate nanofluidic energy absorption in porous zeolitic frameworks. *Nat. Mater.* **20**, 1015–1023 (2021).
53. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
54. Grossfield, A. WHAM: the weighted histogram analysis method. http://membrane.urmc.rochester.edu/content/wham (2020).
55. Marcos-Alcalde, I., Setoain, J., Mendieta-Moreno, J. I., Mendieta, J. & Gómez-Puertas, P. MEPSA: minimum energy pathway analysis for energy landscapes. *Bioinformatics* **31**, 3853–3855 (2015).

Acknowledgements

This work is supported by the Research Board of Ghent University (BOF) through a Concerted Research Action (GOA010-17). S.M.J.R. and J.S.D.V. acknowledge the Fund for Scientific Research–Flanders (FWO) for a postdoctoral fellowship (grant no. 12T3522N) and a strategic basic (SB) research fellowship (grant no. 1S94519N). V.V.S. acknowledges the Research Board of Ghent University (BOF). The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by VSC (Flemish Supercomputer Center), funded by Ghent University, FWO, and the Flemish Government—department EWI.

Author contributions

S.B., S.M.J.R. and V.V.S. initiated the discussion and designed the paper. S.B., S.M.J.R., J.S.D.V., P.V.D.V. and V.V.S. were involved in the discussion of the results. S.B., S.M.J.R. and V.V.S. wrote the manuscript with the contributions of all authors. J.S.D.V. generated the structures and initial force fields. S.B. extended the force fields where necessary, performed the enhanced sampling calculations and analysed the results.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42004-022-00808-y.

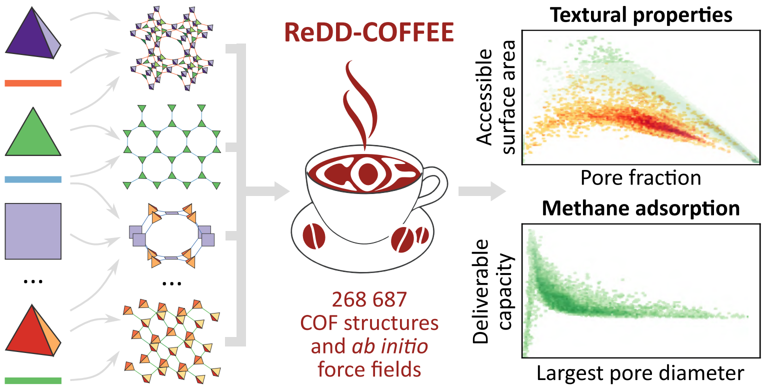
Correspondence and requests for materials should be addressed to Sven M. J. Rogge or Veronique Van Speybroeck.

Peer review information *Communication Chemistry* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Paper III

ReDD-COFFEE: a ready-to-use database of covalent organic framework structures and accurate force fields to enable high-throughput screenings

Juul S. De Vos, Sander Borgmans, Pascal Van Der Voort, Sven M. J. Rogge*, and Veronique Van Speybroeck*



Journal of Materials Chemistry A, **2023**, 11 (14), 7468–7487

The Supporting Information is available at:
https://doi.org/10.1039/D3TA00470H

J. S. De Vos generated the *ab initio* reference data of all 279 SBUs and derived their cluster force fields. He extracted all topologies from the RCSR database and adopted these in his own implementation of the automated *in silico* assembly algorithm to generate the ReDD-COFFEE database. He analyzed the diversity of the database and calculated the textural properties of this database and four other databases of nanoporous materials. J. S. De Vos performed all GCMC calculations to characterize COFs for vehicular methane storage. He analyzed all results and wrote the initial draft of the manuscript.

*Corresponding authors

Journal of
Materials Chemistry A

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *J. Mater. Chem. A*, 2023, **11**, 7468

ReDD-COFFEE: a ready-to-use database of covalent organic framework structures and accurate force fields to enable high-throughput screenings†

Juul S. De Vos,[a] Sander Borgmans,[a] Pascal Van Der Voort,[b]
Sven M. J. Rogge*[a] and Veronique Van Speybroeck*[a]

Covalent organic frameworks (COFs) are a versatile class of building block materials with outstanding properties thanks to their strong covalent bonds and low density. Given the sheer number of hypothetical COFs envisioned *via* reticular synthesis, only a fraction of all COFs have been synthesized so far. Computational high-throughput screenings offer a valuable alternative to speed-up such materials discovery. Yet, such screenings vitally depend on the availability of diverse databases and accurate interatomic potentials to efficiently predict each hypothetical COF's macroscopic behavior, which is currently lacking. Therefore, we herein present ReDD-COFFEE, the Ready-to-use and Diverse Database of Covalent Organic Frameworks with Force field based Energy Evaluation, containing 268 687 COFs and accompanying *ab initio* derived force fields that are shown to outperform generic ones. Our structure assembly approach results in a huge amount of computer-ready structures with a high diversity in terms of geometric properties, linker cores, and linkage types. Furthermore, the textural properties of the database are analyzed and the most promising COFs for vehicular methane storage are identified. By making the database freely accessible, we hope it may also inspire others to further explore the potential of these intriguing functional materials.

Received 27th January 2023
Accepted 15th March 2023

DOI: 10.1039/d3ta00470h

rsc.li/materials-a

1 Introduction

Covalent organic frameworks (COFs)[1–6] form a class of nanoporous materials that show great potential for diverse applications, including the storage and separation of gases[7–12] or from solution,[13–15] catalysis,[16–19] energy storage,[20,21] optoelectronics,[22–24] sensing,[25] and drug delivery.[26] They are built from organic linkers, which form secondary building units (SBUs) that are held together by strong, covalently bound linkages. These covalent bonds typically result in COFs with a high mechanical, thermal, and chemical stability, and their building block structure makes them highly tuneable, similar to metal–organic frameworks (MOFs).[27,28] Substituting these building blocks, or the material's topology, can greatly influence the properties of the synthesized structure, such as the pore size,[29] adsorption capacity,[30] electronic properties,[31] or flexibility.[32] This fascinating feature makes materials for which the reticular principle holds the ideal engineering materials. Yet, the almost unlimited number of COFs that can be synthesized using reticular synthesis makes an experimental exploration of the whole COF material space for a given application unfeasible.[33,34] Therefore, computational high-throughput screenings offer an efficient and cheap alternative to accelerate materials discovery.[35,36] Vital for the predictive power of any high-throughput screening is the material database on which the computational screening is performed, which should be both diverse and uniformly distributed over the material space.[37] In addition, a cheap yet accurate description of the interatomic interactions is crucial to ensure viable property predictions. In this work, we therefore present the ReDD-COFFEE database containing 268 687 COF structures, showcasing a great diversity in terms of geometric properties, linker cores, and linkages, and

---

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

[b] *Centre for Ordered Materials, Organometallics and Catalysis (COMOC), Ghent University, Krijgslaan 281 (S3), 9000 Ghent, Belgium*

† Electronic supplementary information (ESI) available: Additional details about the database construction and force field generation and validation, together with extensive analysis and illustrative case studies. Further information and analysis of the diversity metrics and subset selection. Supplementary property–property relations for textural and adsorption properties. Benchmark of the adopted parameters in the force field optimizations, Zeo++ calculations, and GCMC calculations. All 268 687 COF structures and *ab initio* derived force fields of the ReDD-COFFEE database are publicly available at **https://doi.org/10.24435/materialscloud:nw-3j**. The relevant input files and computational data which generated the results of this work are available at **https://doi.org/10.5281/zenodo.7697262**. See DOI: **https://doi.org/10.1039/d3ta00470h**

accompany each material with an *ab initio* derived force field. Whereas such system-specific force fields have been developed earlier for a limited set of materials,[38–50] we have largely extended the scale of materials for which they are derived. Our ready-to-use database ensures molecular simulations can be directly started from the provided structures and force fields. This is explicitly demonstrated by performing a high-throughput screening on our database with the goal to unveil property–property relationships for textural and adsorption properties and discover attractive vehicular methane storage materials.

Computational high-throughput screenings offer a valuable alternative to accelerate materials discovery.[35,36,51,52] Such screening studies select a database of material geometries and perform several calculations on each material to predict their macroscopic behavior. Especially for MOFs, high-throughput screenings are abundant, especially in the fields of gas adsorption[53–60] and separation[61–66] processes. Recently, also the discovery of MOFs with targeted electronic and catalytic properties has gained attention.[67–70] A limited number of screening studies has been performed to shed light on their mechanical stability.[40,56] All these high-throughput screening studies start from one of the many constructed MOF databases. These databases can be divided into two major categories, depending on the origin of the structures they contain. On the one hand, experimental databases are built from already synthesized materials, containing either the synthesized structure or the structure as obtained after computational structure optimization and/or guest removal. The CoRE MOF database,[54] for instance, contains a subset of 5109 MOF structures identified from the Cambridge Structural Database (CSD)[71] and was later expanded to include 14 142 structures.[72] Moghadam *et al.* implemented an automated screening algorithm in the CSD Python API to instantly identify a MOF when it is added to the CSD.[73] At the time of its first publication, this subset contained 69 666 MOF materials.[73] Recently, the QMOF database was established, containing a subset of 15 713 materials from the previously mentioned databases for which DFT calculations can be carried out efficiently.[68] On the other hand, hypothetical databases contain *in silico* generated structures. They broaden the material space and provide a large and diverse set of structures. Although energy minimization approaches exist to generate plausible geometries, such as the Automated Assembly of Secondary Building Units (AASBU) method,[74] hypothetical databases generally rely on geometric procedures that connect SBUs with one another to form a periodic material without any optimization. These geometric procedures can be divided into two classes: bottom-up and top-down methods, depending on how the SBUs are assembled. In the bottom-up approach, SBUs are naturally grown until a periodic framework is formed. Wilmer *et al.* applied this method to a set of 102 SBUs to generate a database of 137 953 MOFs,[53] whereas a database of 324 500 MOFs was generated from 66 SBUs and 19 functional groups by Fernandez *et al.*[75] However, later research showed that the topological diversity of the structures constructed using this bottom-up approach is limited.[76] Top-down approaches typically result in a larger variety of topologies. These methods initially define the topological net, after which the SBUs are deliberately placed on the net's nodes. Multiple software packages, such as Zeo++,[77] AuToGraFS,[78] Weaver,[79] TOBASCCO,[80] and ToBaCCo,[55,56] use such top-down approach to generate hypothetical frameworks. The latter has been used to construct a database of 13 512 MOFs starting from 78 SBUs and 41 topologies.[55]

Also for COFs, several high-throughput screenings have emerged, although they mainly focus on gas adsorption and separation processes, such as methane storage,[81–83] hydrogen storage,[84,85] and carbon capture,[86–88] among others.[89–91] These high-throughput COF screenings all rely on one of the four large COF databases constructed to date. The two experimental databases, the CoRE[89] and CURATED[86] databases, focus on the boronate ester and imine COFs, which are abundantly present in literature. They naturally underrepresent other linkage types that are less frequently observed experimentally, although these linkage types may result in materials with unique features. Therefore, Martin *et al.*[82] and Mercado *et al.*[83] created two databases of hypothetical COFs to explore different regions of material space. However, because they focused on a large diversity of linker cores, they included only a limited number of linkage types, some of which are rarely realized experimentally. Moreover, when considering those linkage types that are experimentally relevant, these hypothetical databases are significantly lacking. As each linkage type provides unique properties, a suboptimal representation of these subclasses in a database will result in a largely untapped potential for COF materials. For example, boronate ester COFs have outstanding crystallinity,[6] and, whereas imine linked COFs already possess improved stability,[6] enamine COFs have an even higher stability.[92] Furthermore, triazine and hydrazone COFs provide coordination centers for transition metals that can be adopted in catalysis.[93,94] Yet, both enamine and hydrazone COFs are absent in the current hypothetical databases and are underrepresented with respect to imine COFs in the experimental ones.

To tap into the potential of these and other COF linkage types, we present in this work a hypothetical database describing a diverse set of linkage types, including both frequently observed linkages and linkages that are not often synthesized experimentally. Furthermore, an unprecedented feature of our database is that a system-specific force field is generated for each material, starting from the cluster force fields of the underlying building blocks. These are derived using our in-house developed QuickFF routine.[38,39] As such, each structure of the database can directly be used in high-throughput studies.

Besides the material's geometry and partial atomic charges derived for the COFs in the aforementioned experimental databases, the four databases lack the necessary information to model the interatomic interactions. To model these interactions, one can resort to computationally expensive *ab initio* methods, such as density functional theory (DFT), which largely limits the length and time scales that are feasible to simulate,[67,68,70] or a less expensive generic force field with reduced accuracy.[56,95] To obtain a description of the interatomic interactions with a higher accuracy than generic force fields but a lower computational cost than *ab initio* techniques, some studies derived system-specific force fields for a smaller set of materials.[40] Such force fields are attractive to perform high-throughput screenings and to remove structures with a low synthetic likelihood from the database. For instance, many structures in the hypothetical databases of Martin *et al.*[82] and Mercado *et al.*[83] contain largely deformed SBUs and are likely unphysical. To identify the structures that are experimentally the most feasible, the synthetic likelihood of a large database of structures has been predicted with DFT energies[96] and force field free energy calculations.[95] It has been shown that while free energy calculations are necessary to predict the most favorable configuration out of a set of polymorphs, energy metrics are sufficient to predict the synthetic likelihood of hypothetical materials.[95] However, this requires one to augment the versatile COF database with a relatively inexpensive yet sufficiently accurate method to describe the interatomic interactions.

Therefore, we present in this work the ReDD-COFFEE database: a ready-to-use database of 268 687 COFs. ReDD-COFFEE is an acronym for Ready-to-use and Diverse Database of Covalent Organic Frameworks with Force field based Energy Evaluation. In our database, an optimized atomic structure and an *ab initio* derived force field is provided for each material. Essential to generate system-specific force fields for the huge amount of structures within this database, is the building block approach. In this approach, cluster force fields are fitted to quantum mechanical reference data for a limited number of smaller building blocks and then assembled to derive force fields suitable for the periodic structures. This procedure was introduced earlier through our QuickFF procedure,[38,39] but is now adopted for a much larger number of materials. By deriving cluster force fields for the underlying building blocks, the number of required *ab initio* calculations is greatly reduced, while a high accuracy of the interatomic potential is maintained.[40–50] All COFs are assembled starting from a limited set of 279 SBUs, which are selected to result in a diverse database, as elaborated in Section 4.2. In contrast to other hypothetical databases, we have explicitly included a large number of linkage types in our database. The SBUs are combined taking physical constraints into account and the structures that have the lowest synthetic likelihood are removed with a deformation energy filter. We firstly demonstrate that our system-specific force fields achieve a higher accuracy than generic force fields. Next, we show that our database has a great diversity in terms of geometric properties, linker cores, and linkages compared with the already existing COF databases. Furthermore, we establish property–property relations in between textural properties and compare them with other nanoporous materials, *i.e.*, MOFs[53,55,68] and zeolites.[97] Finally, the applicability of the database to identify COFs for vehicular methane storage is demonstrated by identifying promising candidates and determining property–property relations for adsorption properties on a diverse subset of the database containing 10 000 structures. All 268 687 optimized structures and force fields of the ReDD-COFFEE database are publicly available at **https://doi.org/10.24435/materialscloud:nw-3j** and are ready-to-use for further high-throughput screenings.

2 Methodology to construct database

2.1 Terminology to describe a COF's structure

The nomenclature of the COF building blocks used throughout this paper mimics the experimental synthesis of the material. Experimentally, a COF is synthesized by mixing several precursors, which react to form a covalently bound linkage that holds the structure together. As illustrated in Fig. 1, we herein distinguish two subregions for each precursor: a linker core, which remains unaltered during linkage formation, and reactive groups, which react with the reactive groups of other

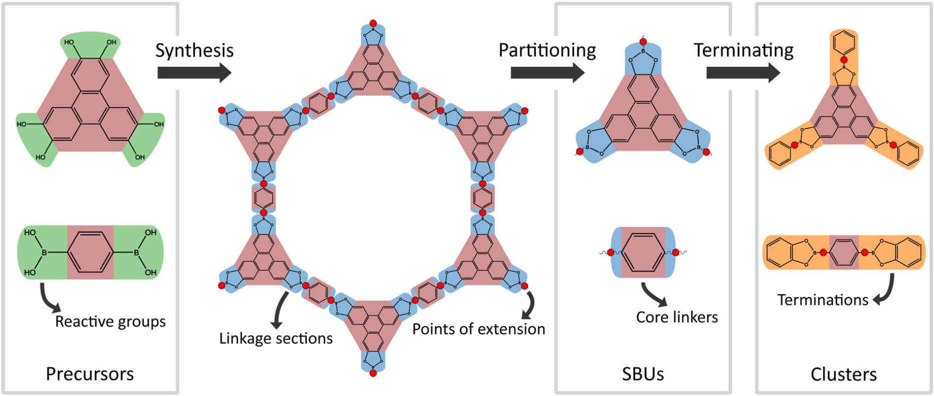


Fig. 1 Illustration of the terminology used in this paper for COF-5. Reactive groups (green) that are attached to linker cores (red) react to form linkages, consisting of linkage sections (blue) that are assigned to an SBU. Two linkage sections connect at a point of extension (red dots). Choosing an appropriate termination (orange) for the SBU defines the accuracy of the cluster model.

precursors to form the linkage. The COF itself can be partitioned into secondary building units (SBUs),[33,34] each containing the linker core and part of the linkage, which we will call the SBU's linkage section. The linkage sections of two adjacent SBUs connect at a point of extension to form a linkage. In this work, the points of extension are always located in the geometric center of the central bond of the linkage. As will be explained in Section 2.4, this partitioning of the material in SBUs enables us to derive a cluster force field for each building block, which can be combined to accurately describe the whole COF structure. To obtain an accurate cluster force field for these SBUs, their environment has to be mimicked as detailedly as possible by choosing an appropriate termination for the linkage section. The combination of the linker core with its termination is defined as the cluster for this SBU.

2.2 Structure assembly and database construction

Our top-down approach to generating a versatile COF database starts from a topological blueprint and a selection of SBUs that can be placed on its nodes (vertices and edges). A topology is a periodic graph in which the connection between vertices, *i.e.*, the connection between the SBUs of the material, is represented by edges. This is illustrated in Fig. 2 for the **pts** topology. The collection of both vertices and edges is defined as the nodes of the topology. The process of determining the topology is straightforward once the structure of the material is known and the building blocks or the linkages between them are defined. In this case, the topology directly follows by reducing each building block into a zero-dimensional vertex and storing the connectivity between building blocks as edges that connect these vertices. However, in the inverse approach, *i.e.*, when only the topology and its SBUs are known, it is more difficult to generate the material's structure at the atomic level, because it is not clear how the three-dimensional SBUs should be inserted on the nodes.[79] Since this is a vital part of our database generation, we here implement an additive top-down approach. This takes into account both geometric and energetic parameters to introduce the rigid building blocks into the topology to obtain a physical structure. As illustrated in Fig. 2, the procedure consists of a four-step process, which is explained in more detail below. Three filters check whether the structure is physical or whether it should be rejected. If the optimized material passes each filter, it is added to the final database. Additional details to construct the database are provided in Section S1 of the ESI.† In Section S1.6,† the structure assembly procedure is illustrated with a detailed case study of COF-108.

Step 0. In step 0 of Fig. 2, the topology and SBUs, *i.e.*, a so-called (topology, SBUs) combination, are selected as input. The Reticular Chemistry Structure Resource (RCSR) contains a large number of experimentally observed two- and three-

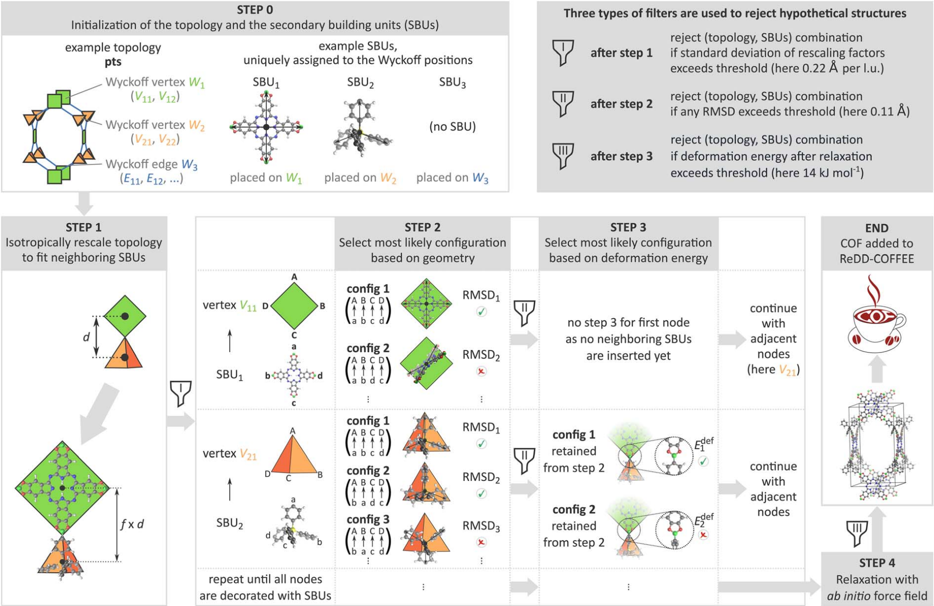


Fig. 2 The four-step process followed to assemble a COF structure. This procedure starts from a topology and a selection of SBUs for each set of equivalent nodes. In step 1, the topology is rescaled to fit the SBUs. In step 2 and step 3, the most favorable SBU configurations are chosen based on geometric and energetic considerations. By using an additive top-down approach, a low-energy structure is quickly generated. Three filters are implemented to ensure that only physical structures are added to the database.

dimensional topologies, embedded in a Euclidean representation.[98] We extracted these embeddings using a web scraping script and only retained the 2495 topologies with embedding type 1, *i.e.*, where all nearest-neighbor vertices at the same distance are connected by an edge. These topologies are most frequently observed experimentally.[99] Subsequently, an SBU is assigned to each set of equivalent nodes, *i.e.*, each Wyckoff set of the topology, based on two restrictions. Firstly, the number of points of extension of the SBU has to be the same as the coordination number of the node, as it has to be able to connect with each neighbor. Secondly, the two linkage sections of neighboring SBUs have to form one of the linkage types considered in this manuscript. For example, it is not possible to combine SBUs belonging to an imine and a boronate ester linkage. Since each Wyckoff set can be assigned a different SBU, the theoretical upper limit for the number of distinct SBUs in a material is given by the total number of Wyckoff sets of the material's topology. To also include materials without a linear linker in the database, an edge Wyckoff set may also remain vacant, in which case the SBUs that decorate the neighboring vertices are directly connected. In this work, the SBUs are chosen from a set of 279 building blocks. Four of them are SBUs that have no experimental precursor but emerge during COF synthesis as product of the reaction groups: boroxine, triazine, borazine, and borosilicate. The other 275 are generated by combining 30 experimentally observed linker cores with eleven frequently observed linkage types that correspond to eleven reactive groups. Our database contains imine,[100] boronate ester,[1] (keto)enamine,[21,92] triazine,[101] (acyl)hydrazone,[102] azine,[103] imide,[104] boroxine,[1] borosilicate,[105] oxazoline,[106] and borazine[107] COFs (see Fig. S1 and S2 of the ESI†). In this way, a total of 5 537 951 (topology, SBUs) combinations in 1272 topologies are initialized. Each of these combinations is given the unique label top_SBU_1_SBU_2_…_SBU_N, in which “top” indicates the topology and SBU_i is the SBU placed on the i-th Wyckoff set (see Section S1.3 of the ESI†).

Step 1. After choosing the (topology, SBUs) combination in step 0, the topology is isotropically rescaled to be able to accommodate the SBUs in step 1 of Fig. 2. To this end, a rescaling factor f_i is calculated for each edge Wyckoff set so that the distance between its connecting vertices is the same as the distance between the centers of the SBUs occupying those vertices. We require the unit cell to be rescaled isotropically, *i.e.*, with the same rescaling factor in every direction. The single rescaling factor is defined as the mean of the rescaling factors averaged over all Wyckoff sets. To prevent the rescaling factors from differing too much and avoid overlapping SBUs, the standard deviation of the set of rescaling factors is not allowed to surpass a threshold of 0.22 Å per l.u., with l.u. being the length unit of the topology (see Section S1.5 of the ESI†). This constraint corresponds to the first filter visualized in Fig. 2. After this first filter, 749 859 (topology, SBUs) combinations are passed on to step 2. Most discarded combinations contain mixed-linker topologies in which the lengths of the linear linkers are too different.

Step 2. Once the topology is rescaled to accommodate the SBUs, they can be inserted onto their assigned nodes. This is accomplished by orienting the points of extension of the SBU towards the neighboring nodes in the topology while keeping the SBU's internal geometry fixed. Once a point of extension is chosen for each neighboring node, the optimal transformation is found by Kabsch algorithm.[108] However, it is not *a priori* known which point of extension has to be oriented towards which neighboring node. Every permutation of the points of extension results in a different SBU configuration, with possibly a different internal geometry. This is illustrated in Fig. 2 by configurations 1 and 2 of the square SBU_1 that is placed on vertex V_{11}. The points of extension of SBU_1 are indicated with the lower case letters a, b, c, and d, whereas the neighboring nodes of vertex V_{11} are specified with the upper case letters A, B, C, and D. When points of extension a and b are oriented towards nodes A and B, respectively, the least strain configuration is found when assigning the points of extension c and d to the nodes C and D, as in the first configuration in Fig. 2. However, in the second configuration, these points are assigned to the nodes D and C, respectively, for which the Kabsch algorithm defines a transformation that results in a less favorable configuration. Furthermore, for some SBUs, the point symmetry of their internal geometry is lower than the point symmetry of their points of extension. In these cases, information about the internal geometry of the building block is lost by abstracting the SBU to its points of extension.[79] Therefore, even for two configurations that orient the points of extension towards the neighboring nodes equally well, different internal geometries may be found. This is illustrated for configurations 1 and 2 of the tetrahedral SBU_2 that is placed on vertex V_{21} in Fig. 2. Whereas the SBU perfectly fits the vertex in both configurations, two different internal geometries are observed, which differ in the orientation of the phenyl rings. To only retain the most favorable SBU configurations, geometric and energetic criteria are taken into account in step 2 and step 3 of the structure assembly process, respectively.

In step 2 of Fig. 2, a first selection of reasonable SBU configurations is made based on the geometric consideration whether the points of extension can be oriented towards the neighboring nodes correctly. This is quantified by the root-mean-square deviation (RMSD) between the unit vectors pointing (i) from the center of the SBU towards the points of extension, and (ii) from the node towards its neighbors in the topology. In the ideal case, these two sets of vectors would be overlapping once the SBU is positioned on the node using the transformation obtained from Kabsch algorithm.[108] However, for some configurations, such as configuration 2 of SBU_1 and configuration 3 of SBU_2 in Fig. 2, the transformation results in a less optimal configuration and a geometric mismatch between the SBU and the topological node is introduced. To avoid that these SBU configurations would be inserted in the topology, only those configurations that minimize the RMSD are selected and passed on to step 3. Furthermore, to avoid structures with too large a mismatch, a second filter checks if the minimal RMSD for each node is lower than the threshold $\text{RMSD}_{\text{max}} = 0.11$ Å (see Section S1.5 of the ESI† for the rationalization of this value). If this is not the case, for instance when trying to insert a tetrahedral SBU on the square vertex V_{11} in Fig. 2, the (topology, SBUs) combination is rejected. After this step, 403 581 combinations in 1196 topologies are retained.

Step 3. As explained before, an identical RMSD does not imply the same internal geometry as the representation of the SBU by its points of extension does not fully capture its three-dimensional structure. Depending on the geometry of the neighboring SBUs, the linkage formed between the SBUs in one configuration can be energetically more favorable than in another configuration, although they have the same geometric mismatch, such as configurations 1 and 2 of the tetrahedral SBU that is placed on vertex V_{21} in Fig. 2. In step 3, the deformation energy E_{def}, formally introduced in Section 2.4, is used to discriminate between such configurations and identify the one in which the topological constraints have the least influence on the SBUs. This energy depends on the selected configuration of the neighboring SBUs. Therefore, it should in principle be calculated for each possible material configuration, which is a product of all SBU configurations. However, as the number of these material configurations increases exponentially with the number of nodes in the topology, a brute-force iteration over all material configurations is unfeasible, definitely when constructing an extensive database.

This challenge motivated us to introduce our *additive* top-down approach, in which the nodes of the topology are decorated one-by-one. In each step, the optimal configuration of the SBU of the selected node, *i.e.*, the configuration that minimizes both the RMSD and the deformation energy of the linkages with the already inserted SBUs, is identified. By following a breadth-first iteration through the topological graph, all nodes at a certain distance from the starting vertex are decorated before continuing with the next layer. As such, the number of linkages with neighboring building blocks is increased for each SBU as compared to a random iteration or a depth-first iteration. As indicated in Fig. 2, step 3 can be omitted for the starting node, as no neighboring SBUs are present yet and the deformation energy E_{def} is the same for all configurations. The final SBU configuration for vertex V_{11} is thus chosen from the set of configurations that minimizes the RMSD in step 2. In the next step of the breadth-first iteration, a neighboring node to the ones already occupied is decorated with an SBU. In the example of Fig. 2, the vertex V_{21} is selected once a configuration for the SBU that is put on vertex V_{11} is determined. From the SBU configurations that minimize the RMSD in step 2, the deformation energy E_{def} of the linkage from this SBU with the already inserted SBU on vertex V_{11} is calculated. Finally, the SBU configuration that minimizes the deformation energy E_{def} is selected and inserted on the vertex V_{21}. As such, the deformation energy E_{def} has to be calculated only once for every SBU configuration, instead of for every material configuration, and the computational complexity of the structure assembly process is reduced from exponential to linear in terms of the number of nodes in the topology. For linear linkers, which only have two points of extension lying on the same axis as the center of the SBU, there is still a degree of freedom that step 2 can not describe: the rotation around their axis. By using the deformation energy E_{def} in step 3, this degeneracy is lifted to the point that only the internal symmetries of the linker remain.

Step 4. After step 3 is completed for each node in the topology and all SBUs are inserted and connected, an initial periodic structure is obtained. However, up to now, the SBUs remained rigid. As the threshold for the maximally allowed RMSD in step 2 is quite high, this means that some mismatch between SBUs is allowed, which is motivated by the fact that COF building blocks can have an appreciable degree of flexibility that can accommodate for the introduced mismatch. Therefore, after the initial material is assembled, it is relaxed in step 4 using its system-specific force field that is generated according to the procedure described in Section 2.4 to find an optimal structure. The mismatch between the SBUs, which was localized around the linkages, will now be released after a full geometry optimization. However, as some SBUs are more flexible than others, the final optimized structure can have a low synthetic likelihood. Therefore, the last filter uses the deformation energy E_{def} again to check whether the strain that was introduced between the rigid SBUs is sufficiently released. The threshold is here defined as 14 kJ mol^{-1} (see Section S1.5 of the ESI†). When the deformation energy of the optimized structure does not exceed this threshold, it is finally included in the ReDD-COFFEE database, which contains 268 687 structures, distributed over 1114 topologies. In total, there are 5856 2D structures and 262 831 3D structures. A detailed overview of the distribution of the linkages and dimensionality found in our database is given in Table S1 of the ESI.†

2.3 Diversity metrics

It is important to compare the diversity of our database with already existing COF databases to verify that a large part of material space is covered and that these regions are sampled equally to avoid focus on certain subclasses.[82,83,86,89] Inspired by a previous MOF study by Moosavi *et al.*, three metrics are defined below to assess the diversity of a subset of the material space: the variety V, the balance B, and the disparity D.[37] We will calculate these diversity metrics on four domains: (i) the pore geometry, and the chemical environment of the (ii) linkages, (iii) linker cores, and (iv) functional groups. Whereas the pore geometry can be described by eight structural parameters, the chemical environments of the COF are assessed with revised autocorrelation functions (RACs),[109] which were proven to be useful in multiple applications.[37,109–112] Further details about these diversity metrics are described in Section S3 of the ESI.†

The RACs are calculated between two sets of atoms, which are defined for each chemical environment (ii–iv). Initially, the linkages (ii) that hold together the COF are identified by scanning the material graph for linkage patterns. By removing these linkages from the material graph, the linker cores (iii) that constitute the COF are retrieved. The functional groups (iv) are detected as parts of the linker cores that are attached to its skeleton with exactly one bond and do not exist of a single hydrogen atom. Once the start and scope atom lists are determined for each chemical environment (see Section S3.2 of the ESI†), the difference and product RACs are defined in eqn (1) and (2), respectively.

$$ {}^{\text{start}}_{\text{scope}}P^{\text{diff}}_{d} = \sum_{i}^{\text{start}} \sum_{j}^{\text{scope}} (P_i - P_j)\delta(d_{ij}, d) \tag{1}$$

$$ {}^{\text{start}}_{\text{scope}}P^{\text{prod}}_{d} = \sum_{i}^{\text{start}} \sum_{j}^{\text{scope}} P_i P_j \delta(d_{ij}, d) \tag{2}$$

Each RAC investigates one out of six properties P: the atom identity (I), connectivity (T), Pauling electronegativity (χ), covalent radius (S), nuclear charge (Z), or polarizability (α). The depth d indicates the number of bonds that must be present between the considered atoms in the start and the scope list. Together with the definition of the start and scope atom lists, which depend on the chemical environment, this depth specifies the atom pairs over which to iterate. With two types of RACs, six properties, and a maximum depth of three bonds, a total of 48 descriptors are obtained for each environment (linkages, linker cores, and functional groups).

Once all materials are featurized, either with structural parameters or with RACs, the diversity metrics that determine how well a material set covers the material space can be defined. The material space is in this context described by the union of all available databases: the four existing databases and our database presented here. Before calculating the diversity metrics, the material space is subdivided into a specific number of bins, here chosen to be 1 000, defined through k-means clustering. These bins partition the material space into subclasses. The variety V of the considered database checks if each of the subclasses is sampled by measuring the number of bins that are examined. To make sure that each subclass is sampled equally, the balance B indicates the evenness of the distribution of materials among the sampled bins. In the ideal case, all covered subclasses are represented with the same number of structures. Lastly, the disparity D quantifies again the fraction of material space that is covered by the considered database, using a distance-based approach instead of the clustering into bins. Therefore, it is also a measure of the spread of the bins. The mathematical definition of these diversity metrics is given in Section S3.3 of the ESI.† Together, the variety V, the balance B, and the disparity D summarize the diversity of a subset of the material space for each investigated domain.

2.4 Force field generation

The ReDD-COFFEE database aims to report on both material geometries and system-specific force fields. When compared to *ab initio* methods, these force fields can be directly adopted to perform high-throughput simulations on a larger number of structures and for longer time scales. Moreover, it can do so with a higher accuracy than generic force fields. Most system-specific force fields are derived from periodic *ab initio* data. However, for a database containing more than 100 000 structures, such procedure becomes unfeasible. To reduce the amount of necessary *ab initio* data, we exploit the building block nature of reticular materials to partition each COF in SBUs and associated clusters. For each of these clusters, a system-specific cluster force field is derived. Following our QuickFF procedure,[38,39] a system-specific force field for the periodic structure is obtained by combining the cluster force fields to account for all SBUs in a specific material. As many materials can be generated with a limited set of SBUs, this proven approach circumvents the need for expensive periodic *ab initio* reference data and requires only a modest set of cluster calculations.[38–50] Section S2 of the ESI† contains supplementary details about this approach.

The accuracy of these cluster force fields is to a large extent defined by the choice of termination. A larger termination mimics the environment of the SBU in the material in more detail but is only applicable for a smaller set of materials as the termination has to represent the material correctly. Herein, we have chosen to define a single termination for each observed linkage section (blue in Fig. 1). The cluster termination is therefore independent of the SBU environment and can be adopted in each material containing that SBU. As such, there are as many clusters as there are combinations of linker cores and linkage sections. These terminations are depicted in Fig. S3 of the ESI.†

Once a cluster force field is obtained for each SBU in the material, the parameters of the periodic force field are derived. As the cluster approximation has the smallest impact on covalent terms that are fully embedded in the SBU, these terms are directly adopted from the cluster force field. In contrast, the overlap terms that span multiple SBUs are most accurately described in the cluster force field of the SBU with the majority of the atoms. Hence, these parameters are obtained by taking a weighted average over the respective terms in both constituent SBUs. For each term, the weights associated with each cluster involved in the term are proportional to the number of atoms embedded in the SBU core of the cluster. Also bond charge increments of bonds spanning two SBUs are obtained following this procedure. Charge neutrality is satisfied by using bond charge increments to derive the partial charges in the periodic material.[113] The averaging is only allowed when all covalent bonds involved in the term have the same bond order in both clusters involved, which is mostly the case in this paper. When this is not the case, the parameters are directly obtained from the building block that mimics the environment correctly.

Inserting the SBUs in a periodic material introduces topological constraints to its environment, which can impose strain between the SBUs that is not present when considering an isolated cluster. The equilibrium geometry of the SBU in the material is therefore not necessarily the same as in the cluster model. This is for example illustrated in COF-108,[2] where the tetrahedral TBPM building block, with point group T_d, is inserted in the **bor** topology and placed on a vertex with point group D_{2d}. Due to the geometric mismatch, which results in an RMSD of 0.09 Å before optimization, the SBUs are slightly reshaped with respect to their equilibrium structure.

To quantify the energy penalty for inserting the SBUs in a suboptimal environment, the deformation energy E_{def} is introduced as the energy difference between the periodic material and the sum of the energies of the isolated clusters it is composed of. Only interactions that are present both in the periodic material and the isolated clusters are included.

Although they differ only slightly from the cluster force field parameters, the force field parameters of the periodic structure are also used for the calculation of the energies of the different clusters to ensure a consistent comparison. More specifically, the covalent and non-covalent interactions in the periodic material are divided into interactions between atoms within the same SBU ($E_{\text{cov}}^{\text{per,intra}}$ and $E_{\text{non-cov}}^{\text{per,intra}}$) and in different SBUs ($E_{\text{cov}}^{\text{per,inter}}$ and $E_{\text{non-cov}}^{\text{per,inter}}$). In the periodic material, the interactions within the same SBU, *i.e.*, $E_{\text{cov}}^{\text{per,intra}}$ and $E_{\text{non-cov}}^{\text{per,intra}}$, are also described in the cluster model by $E_{\text{cov}}^{\text{clust,intra}}$ and $E_{\text{non-cov}}^{\text{clust,intra}}$, respectively. The covalent interactions between different SBUs in the periodic material, *i.e.*, $E_{\text{cov}}^{\text{per,inter}}$, exactly correspond to the overlap terms between the core and the terminations of the appropriate clusters $E_{\text{cov}}^{\text{clust,inter}}$. To avoid that these overlap terms would be counted in both corresponding clusters, they are weighted with the same rescaling factors as in the generation of the periodic force field. However, no cluster counterpart for the long-range non-covalent interactions between different SBUs $E_{\text{non-cov}}^{\text{per,inter}}$ can be defined, as these interactions are not integrated into the isolated cluster model. They are therefore neglected in the calculation of the deformation energy. As the topological constraints scale with the number of linkages between SBUs, N, the deformation energy is normalized with this number, resulting in the definition of E_{def} in eqn (3).

$$E_{\text{def}} = \frac{E_{\text{cov}}^{\text{per,intra}} + E_{\text{cov}}^{\text{per,inter}} + E_{\text{non-cov}}^{\text{per,intra}}}{N} - \frac{E_{\text{cov}}^{\text{clust,intra}} + E_{\text{cov}}^{\text{clust,inter}} + E_{\text{non-cov}}^{\text{clust,intra}}}{N} \quad (3)$$

Since the deformation energy indicates the energy change upon inserting an SBU in a topology, it describes how easily the SBUs can accommodate the mismatches introduced due to topological constraints, *i.e.*, if they are sufficiently flexible. High values of the deformation energy indicate that the SBUs are too rigid to accommodate the introduced mismatches and result in highly contorted structures. As described in Section 2.2, COFs with a high deformation energy have a low synthetic likelihood and are therefore discarded from the database *via* the third filter.

3 Computational details

For each SBU, a cluster force field is derived that is fitted to the *ab initio* Hessian and equilibrium structure. An initial cluster geometry is assembled with Avogadro (v1.2.0)[114] and afterwards optimized with Gaussian 16 (Revision C.01)[115] using the B3LYP functional[116–118] with Grimme D3 dispersion correction.[119] The Hessian of the relaxed structure is calculated using the same level of theory. When imaginary frequencies are observed, indicating that the cluster is optimized to a saddle point on the potential energy surface, a small perturbation is added to the geometry and the procedure is repeated until all frequencies are positive. All calculations adopt the 6-311++G(d,p) Pople basis set[120] and the NoSymm flag is used. Once the *ab initio* reference data is obtained, a cluster force field is derived comprising three main parts. The van der Waals interactions are described with a Buckingham potential using the MM3 parameters derived by Allinger *et al.*[121] The partial charges for the electrostatic interactions are obtained with the MBIS procedure[122] as implemented in HORTON (v2.0.0).[123] Bond charge increments[113] are used to divide the charges over the different bonds and Gaussian smearing is applied to obtain a more accurate description of the charge distribution. Finally, the covalent part of each cluster force field is fitted to the *ab initio* Hessian and equilibrium structure using QuickFF (v2.2.4),[38,39] where both the electrostatic and van der Waals interactions are provided as reference force fields. Each set of equivalent atoms is given a unique force field atom type. Specific details about the triazine dihedral angles and the out-of-plane terms are provided in Section S2.1 of the ESI.† Besides the QuickFF derived force field, also a UFF force field[124] is generated for each SBU to compare their accuracy with system-specific force fields. The covalent and van der Waals interactions are defined according to the rules mentioned in the original UFF paper,[124] which are implemented in our in-house force field generator software. As no partial charge fitting scheme is imposed by the UFF method, and as our focus is to describe the covalent part of the force field, the electrostatic interactions are defined identically to the QuickFF derived force field. For the validation of the cluster force fields, each cluster is optimized with both force fields, using our in-house force field engine Yaff (v1.6.0),[125] and the vibrational frequencies are calculated with a normal mode analysis (NMA) using TAMkin (v1.2.6).[126]

The initial periodic structure is generated by placing the SBUs on the nodes of the topology and its periodic force field is derived from the cluster force fields of the underlying building blocks, as described in Sections 2.2 and 2.4, respectively. For 2D COFs, a 1 × 1 × 2 supercell is used to include two layers in the simulation cell that can better describe the specific type of stacking. The initial structure is a perfect eclipsed structure with an interlayer distance of 10 Å. This interlayer distance is chosen sufficiently large to avoid overlap between the layers and decreases during the optimization. The structures are relaxed with Yaff,[125] following a three-step procedure. During the structure assembly process, nearby SBUs may contain atoms that almost overlap. The potential well of the Buckingham potential that describes the van der Waals interactions of the QuickFF derived force field would diverge to minus infinity when two such atoms approach each other. To push these atoms apart and generate a more physical geometry, the first step of the optimization procedure is to optimize the atomic positions with the UFF force field for 50 steps. The van der Waals interaction of the UFF force field is described by a Lennard-Jones potential, which is repulsive at short distances.[124] Subsequently, the system-specific QuickFF force field is applied to fully optimize the atomic positions. Finally, the unit cell parameters are added to the degrees of freedom that are relaxed. For each optimization, the conjugate gradient optimizer as implemented in Yaff is adopted. A real-space cutoff r_{cut} of 11 Å is used for the nonbonded interactions. Tail corrections are used to estimate the van der Waals interactions beyond this cutoff distance. Furthermore, the electrostatic interactions are calculated with an Ewald summation, with scaling factor $\alpha = 0.26$ Å^{-1} and reciprocal space cutoff $k_{\max} = 0.26$ Å^{-1}. A truncation model is used to smooth these interactions. The textural properties of the optimized structures are derived with Zeo++ (v0.3).[127] To calculate the accessible surface area and volume, a probe radius of 1.84 Å, coinciding with the kinetic radius of nitrogen,[128] and 3000 Monte Carlo samples are adopted. In Section S5 of the ESI,† benchmark studies for all simulation parameters are provided.

Molecular dynamics (MD) simulations are performed to compute the powder X-ray diffraction (PXRD) patterns and single crystal structures at operating conditions[50] using the Yaff software package[125] with the same force field parameters as determined for the optimizations. The MD simulations sample the $(N, P, \boldsymbol{\sigma}_a = \mathbf{0}, T)$ ensemble at operating conditions, *i.e.*, a pressure of 1 atm and a temperature of 100 K for COF-300 and LZU-111, a pressure of 1 atm and a temperature of 89 K and 298 K for the distinct phases of COF-320, respectively, and a pressure of 1 bar and a temperature of 300 K for the calculation of the PXRD patterns. They are controlled by a Nosé–Hoover chain thermostat[129–131] with three beads and a relaxation time of 100 fs, and a Martyna–Tuckerman–Tobias–Klein barostat[132,133] with a relaxation time of 1000 fs, respectively. The velocity Verlet integration scheme is used, with a timestep of 0.5 fs. For the computation of the crystal structures, 9900 snapshots are collected from an MD trajectory of 500 ps, where the first 5 ps are considered as equilibration run. An MD trajectory of 200 ps is created to compute the PXRD patterns and 50 snapshots are extracted from the last 100 ps. 3D COFs in the test set are simulated using a 2 × 2 × 2 supercell, whereas a 1 × 1 × 5 supercell is adopted for the 2D COFs, resulting in simulation cells of ten layers to account for the inherent freedom in layer movement. All PXRD patterns are computed using the *pyobjcryst* python package, based on the ObjCryst++ Object-Oriented Crystallographic Library.[134] A Cu K$_\alpha$ wavelength of 1.54056 Å is adopted. The peak shape is computed with a pseudo-Voigt shape function, where the mixing parameters are set to $\eta_0 = 0.5$, $\eta_1 = \eta_2 = 0$, and a fixed width W of 0.02° in Caglioti's formula together with the U and V parameters equal to 0°.

In Section 4.4, we perform grand-canonical Monte Carlo (GCMC) simulations[135] using RASPA[136] to calculate methane storage capacities at pressures of 5.8 bar and 65 bar. A cutoff radius of 14 Å is used for all interactions and tail corrections are applied, as required by the TraPPE model.[137] The simulation cell contains as many unit cells as needed to ensure a distance of twice the cutoff radius in every direction. The simulations are run for 10 000 cycles, from which the first 5000 are discarded for equilibration. As outlined in Section S5.3 of the ESI,† this is sufficient to let the system equilibrate. For the subset of 10 000 COFs that are contained in our test set, 1600 structures are discarded from the GCMC analysis as their pore sizes are prohibitively large, preventing the calculation of the interactions between the framework and the methane molecules at 65 bar within a reasonable timeframe. As explained in Section 4.4, it is expected that these structures have a low volumetric deliverable capacity and are therefore unviable candidates for vehicular methane storage. Six additional COFs are discarded as they require too much memory.

4 Results

4.1 Accuracy of the system-specific force fields

As explained in Section 2.4, we have derived a system-specific QuickFF force field for each material in the ReDD-COFFEE database from the cluster force fields of its constituent SBUs. In this section, we will verify that these QuickFF force fields indeed achieve a higher accuracy than universal force fields typically used in high-throughput screenings. Here, we will compare the system-specific force fields with the widely adopted UFF force field.[124] To this end, we will assess the ability of the cluster force fields in reproducing the vibrational frequencies and internal coordinates of the *ab initio* optimized clusters, and we will quantify to what extent the periodic force fields can reproduce experimental powder X-ray diffraction (PXRD) patterns and single crystal structures.

Starting from the *ab initio* optimized geometry, both the QuickFF and UFF force fields are used to relax the SBU clusters. Subsequently, the adjustments of the internal coordinates, *i.e.*, the bonds, bends, dihedral angles, and out-of-plane distances, are measured. The QuickFF force field successfully reproduces the *ab initio* optimized geometry, with RMSD errors on the bonds, bends, and out-of-plane distances as small as 4.73×10^{-3} Å, $7.18 \times 10^{-1\circ}$, and 4.12×10^{-2} Å, respectively (see Fig. S26 in Section S2.3 of the ESI†). These are substantially lower than the RMSD errors obtained for the UFF optimized structures, which amount to 3.56×10^{-2} Å, 2.87°, and 4.50×10^{-2} Å, respectively. The most challenging internal coordinates to describe with force fields are the dihedral angles of non-planar COF building blocks, as they are primarily dictated by long-range electrostatic and van der Waals interactions. The RMSD error of these internal coordinates in the QuickFF relaxed structure is relatively large, namely 9.40°, which is still improved substantially compared to the UFF relaxed structure, for which the RMSD error is 22.27°.

For these optimized clusters, the force field vibrational frequencies are derived and compared with the ones obtained from the *ab initio* Hessian. Similar to the internal coordinates, also the vibrational frequencies are described more accurately by the QuickFF force field (see Fig. S27 in Section S2.3 of the ESI†). The RMSD error on the frequencies is lower than the errors obtained for the UFF force field in all frequency regions. While they are still comparable in the low-frequency regime (<500 cm^{-1}), the QuickFF force fields outperform UFF by an order of magnitude for the higher frequencies (>500 cm^{-1}), as the QuickFF parameters are fitted to reproduce the *ab initio* Hessian.

This quantitative comparison between the force field and *ab initio* derived values of both the optimized cluster geometry and the vibrational frequencies provides an indication of the accuracy our QuickFF force fields can reach. However, we are more interested in the capacity of our periodic force fields to predict experimentally measured, structural properties. For COFs, a key macroscopic descriptor is the PXRD pattern, capturing the

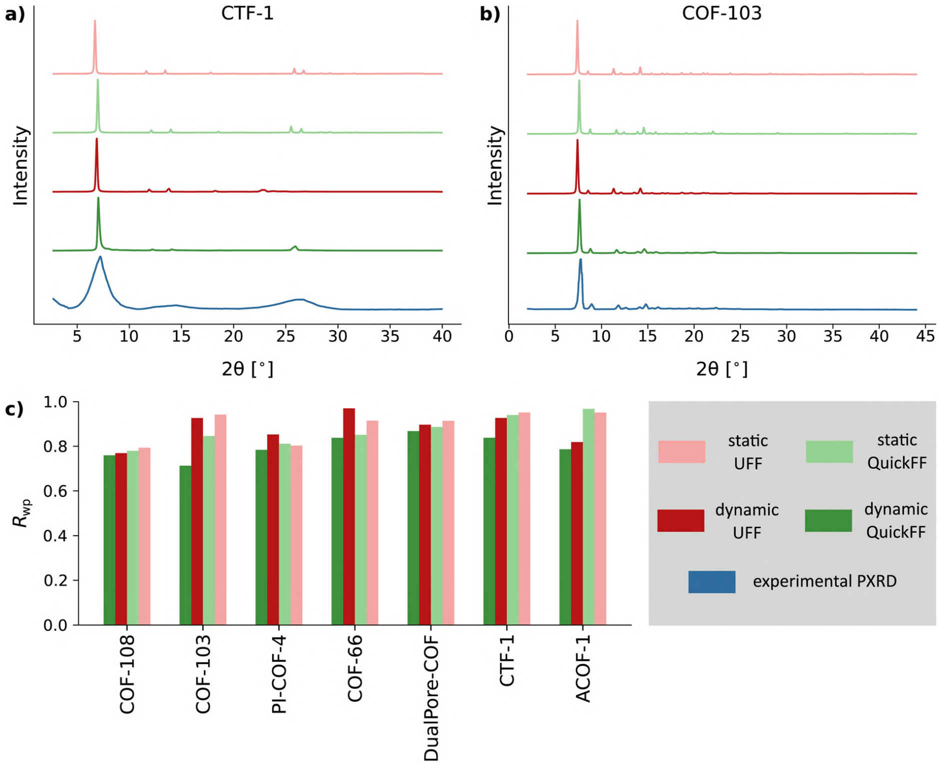


Fig. 3 Comparison of the experimental PXRD patterns of seven COFs with four computational patterns for each material. Static and dynamically averaged PXRD patterns are derived for both the system-specific QuickFF force fields derived in this work and the transferable UFF force field. Diffractograms are offset vertically for clarity. The PXRD patterns of (a) CTF-1 and (b) COF-103 are visualized. (c) Weighted profile residual R_{wp} quantifying the comparison of each of the calculated patterns with the experimental pattern for all selected COFs. The lower the value of R_{wp}, the better the agreement with experiment.

atomic structure of the material, typically used due to the challenging synthesis of single crystal COFs.[5] In Fig. 3c, the weighted profile residual R_{wp} is used as a heuristic metric to compare the calculated PXRD patterns of a diverse set of seven COFs with their experimentally measured pattern. For each material, a static and dynamically averaged PXRD pattern is derived both with the QuickFF and UFF force fields, following our procedure outlined in ref. 50. Dynamical averaging at *operando* conditions is necessary for COFs to account for the inherent temporal character of experimental measurements.[50] In contrast to the static approach, during which the PXRD pattern is calculated for the optimized structure, the dynamic approach starts from an MD trajectory performed at operating conditions. The resulting PXRD pattern is an ensemble average of the pattern calculated for different snapshots from this trajectory.

For the large majority of the examined materials, the R_{wp} metric is lower for the QuickFF than for the UFF force field for both the static and dynamically averaged patterns, indicating a better agreement of the former with the experimental pattern. Of the materials shown in Fig. 3, the static PXRD is predicted better by UFF than QuickFF only for PI-COF-4 and ACOF-1.[12,26] However, once experimental *operando* conditions are taken into account to obtain higher accuracy, our system-specific force fields again outperform the generic ones. As an example, the PXRD patterns for CTF-1 and COF-103 are provided in Fig. 3a and b. It can be observed that the peak that is detected at 27° and 23° in the experimental patterns of CTF-1 and COF-103,[2,138] respectively, is correctly reproduced in the dynamically averaged QuickFF PXRD pattern. However, the poor description of the dihedral angles between aromatic rings in the UFF force field enlarges the unit cell parameters and shifts these peaks to lower angles. PXRD patterns for the additional COFs can be found in Fig. S28 of the ESI.†

When single crystals are available, the structure of a COF can be determined with an even higher resolution compared to analyzing the PXRD pattern. While the synthesis of single crystal COFs is challenging, some studies have succeeded in experimentally determining the framework's atomic structure with single crystal X-ray diffraction (SCXRD) or 3D rotation electron diffraction (RED).[139,140] To verify that our system-specific force fields also predict the COF structure more accurately than generic ones at these higher resolutions, additional MD calculations at *operando* conditions are performed to reproduce the crystal structures of COF-300 and LZU-111. Furthermore, simulations are executed to distinguish between two distinct phases of COF-320, which are observed at different temperatures and both have a specific pore structure and unit cell volume.

As can be observed in Tables S6–S9 of the ESI,† our system-specific QuickFF force fields indeed achieve an overall better agreement with experiment than the generic UFF force field in describing the crystal structures of COF-300 and LZU-111 and the distinct phases of COF-320. Especially for the bonds, a consistently better description is achieved by our force fields, with relative differences being half of the ones obtained by UFF. Whereas some bends are more accurately described by the generic force field, most of them achieve a significantly higher precision in the simulations performed with our system-specific force fields. Also the dihedral angles, which are most difficult to reproduce due to the importance of the nonbonded interactions, are better described by our *ab initio* force fields. Finally, the unit cell volume and the unit cell lengths are reproduced more accurately by the system-specific force fields, again with the relative difference only being half the one obtained with UFF.

4.2 Diversity of the database

To faithfully describe the material space a database represents, all regions in this space must be sampled adequately and equally. Whereas experimental databases have several regions that are overrepresented, hypothetical databases often lack structures in specific areas of the material space. Such databases are characterized by a low variety V or a low balance B, respectively. Furthermore, it is better to include regions in material space that are well-separated, to obtain a higher disparity D and maximize the information contained in the database. Due to our specific database generation approach, many materials in the ReDD-COFFEE database strongly resemble each other. Therefore, it would also be interesting to identify a subset of the database that has a similar diversity and covers a comparable region of material space as the whole database but with much fewer structures.

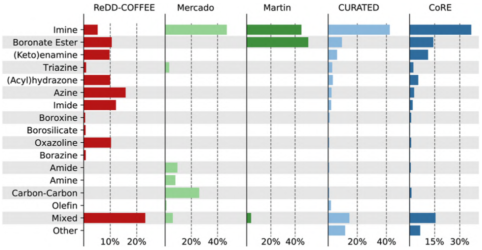


Fig. 4 Distribution of frequently occurring linkage types in five COF databases. The CoRE[89] and CURATED[86] databases are experimental ones, while the databases of Martin[82] and Mercado[83] contain hypothetical structures, similar to our ReDD-COFFEE database.

The diversity of experimental and hypothetical databases is illustrated in Fig. 4, where the distributions of selected frequently observed linkages of four COF databases are compared with the one presented in this work. Typically, the COF material class is divided into several subclasses according to the formed linkage. Each of these subclasses has its unique characteristics and advantages or disadvantages for several applications.[6,141] While the majority of materials in the CoRE and CURATED databases contain imine and boronate ester linkages, other linkage types are less frequently observed in these experimental databases, indicating that they are more difficult to synthesize. To broaden the scope of investigated COF structures, Martin *et al.*[82] and Mercado *et al.*[83] built hypothetical databases using a geometric top-down approach. They chose to focus on a diverse set of linker cores but limited the versatility of linkage types. Both hypothetical databases contain the experimentally abundant imine linkages. In addition, the database of Martin *et al.* also focuses on boronate ester COFs, the second most observed linkage type, and to a lesser extent on borosilicate COFs, while other linkages are not described. Mercado *et al.* chose not to focus on boronate ester COFs, but to describe some less frequently observed linkage types, *i.e.*, the amide, amine, and carbon–carbon linkages. The amide and carbon–carbon linked COFs represent only 1.79% and 1.46% of the experimental CoRE and CURATED databases, respectively, while the amine linked COFs are not present in either of the two databases at all. We chose to continue on this path and include a large variety of linkage types in the ReDD-COFFEE database. We also include linkages that are less frequently observed than the abundant imine and boronate ester COFs but occupy a larger fraction of the experimental databases than amide and carbon–carbon linked COFs. These linkages cover regions in material space that are experimentally more relevant but remain largely uncharted. Furthermore, the structures are well distributed over most linkage types as the only linkage dependent criterium in the (topology, SBUs) combination is that the linkage sections of neighboring SBUs have to combine to form their specified linkage. The structures that are assembled using a boroxine, triazine, borazine, or borosilicate linkage are present less frequently than the other COF subclasses, as the natural constraint that each SBU has to be connected with a three-connected vertex for their linkage largely limits the possible topologies.

Moosavi *et al.* developed a systematic approach to assess the diversity of a MOF database,[37] which we extended to COF databases in Section 2.3. The resulting diversity metrics, *i.e.*, variety V, balance B, and disparity D, for each domain and each of the five COF databases are plotted in Fig. 5. The geometric properties of the structures in hypothetical databases are more

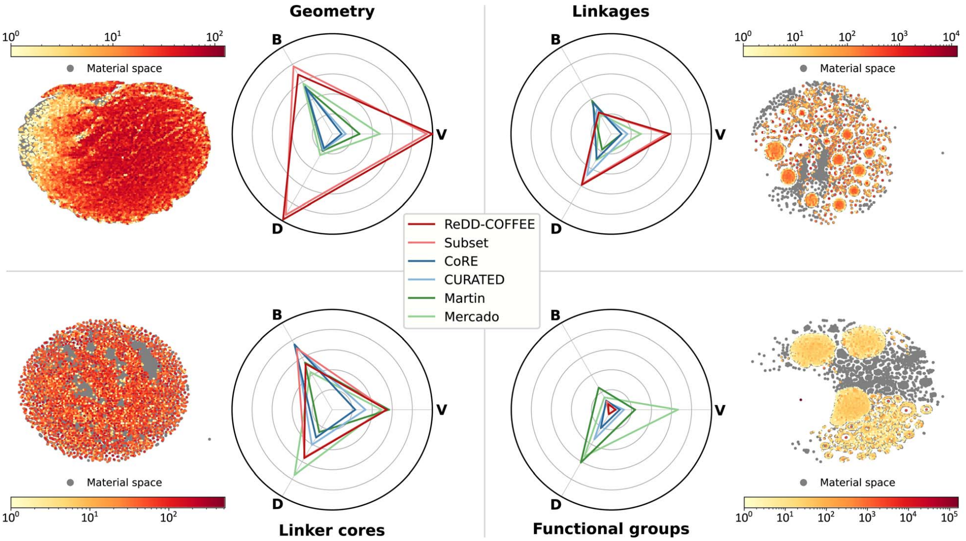


Fig. 5 Diversity metrics (variety V, balance B, and disparity D) for the five domains in COF chemistry. All five databases discussed in the text are considered, together with the diverse subset of 10 000 materials. A t-SNE plot visualizes the material space in gray. Overlaid on this plot are the density of the structures from our ReDD-COFFEE database.

diverse than the ones in experimental databases. This comes as no surprise, as *in silico* materials assembly algorithms provide more structural freedom than experimental synthesis procedures. Compared to the other hypothetical databases, the structures generated in this work can deviate more strongly from the perfect topological graph as provided in the RCSR, because the deformation filter only removes those structures for which the deformation energy is prohibitively large. Therefore, the structures in the ReDD-COFFEE database are assembled in a larger number of topologies than the ones in the databases of Martin *et al.* and Mercado *et al.*[82,83] Furthermore, while the distribution of the linker cores in experimental databases is more balanced, their variety is higher in hypothetical databases due to the combinatorial freedom in choosing the linker cores that constitute the COF. Moreover, while the experimental materials and the structures in the databases of Martin *et al.* and Mercado *et al.* are mostly restricted to containing two different SBUs at maximum, the structures generated in this work can combine more than two building blocks in one structure. The diversity of the linkages present in the ReDD-COFFEE database is better in terms of variety and disparity, but the linkages are more balanced in the CoRE database and the database of Martin *et al.* There is no clear preference between the hypothetical databases on the one hand and the experimental databases on the other hand. Lastly, the diversity of the functional groups is the largest in the hypothetical databases of Martin *et al.* and Mercado *et al.* They exploited a large database of linker cores with various functional groups, which are even more diverse than the ones used in experimental structures. In contrast, since we chose to focus on describing different linkage types, we started from a limited number of linker cores, to which a limited set of functional groups were attached (see Fig. S1 of the ESI†). However, the ReDD-COFFEE database can be used to extend the scope towards more functional groups. To this end, a two-step procedure could be used. Our nonfunctionalized database could first be reduced to a set of frameworks with a high potential for the targeted application, after which they can be easily functionalized *a posteriori*.

While our ReDD-COFFEE database combines a high diversity in terms of geometric properties and linker core and linkage chemistry with accurate system-specific force fields, the large number of structures can also form a barrier to its adoption in high-throughput screenings. Therefore, we created a smaller set of 10 000 structures, using the same descriptors that were used to derive the diversity metrics, *i.e.*, geometric properties and RACs to characterize the linker core, linkage, and functional group chemical environments. Starting from a random initial structure, an iterative procedure is followed that selects the structure that has the largest minimal distance to the set of already selected structures in each step. As materials can only be added to the selected set, the variety V and disparity D of each domain will continue to grow as the selected set becomes larger. However, the balance B will initially show a sharp peak when the first chosen structures are added to empty bins. Upon adding further structures, the balance will start to gradually drop. The selected subset, therefore, occupies a slightly smaller region in the material space than the full database, but the structures are better balanced, as can be observed in Fig. 5. Within the full database, many structures sample similar regions in material space and one structure hardly provides additional information that cannot be learned from the other structures. Computational high-throughput screenings that require expensive calculations can, therefore, use this subset to reduce the computational cost, while still achieving a comparable accuracy with respect to screening the whole database.

4.3 Textural features and property–property relationships

Textural properties are relatively easy to calculate and provide a first insight into the characteristics of the COF material class and the performance of individual materials. Furthermore, they can serve as criteria to filter out structures that do not meet the target design criteria. Therefore, we calculate the textural properties of all COFs in the ReDD-COFFEE database. These properties consist of the mass density and the diameters of the largest included sphere, free sphere, and included sphere along the free path. Also the gravimetric and volumetric accessible surface areas are calculated, as well as the gravimetric accessible volume and the pore fraction. In Fig. 6, property–property relationships between them are established. Additional relationships are visualized in Section S4 of the ESI.† Due to the porous nature of COFs and the fact that they are built up from organic, lightweight atoms, they possess a very low mass density, which allows for exceptionally high gravimetric properties. 2D and 3D COFs show a distinct behavior. Fig. 6c shows that the majority of the 3D COFs in our database have mass densities lower than 200 kg m^{-3} and gravimetric accessible surface areas in the range of 6000 to 10 000 m^2 g^{-1}. In contrast, the mass densities of the more densely packed 2D COFs are mostly within 200 to 600 kg m^{-3} and reach gravimetric accessible surface areas varying from 1750 to 3000 m^2 g^{-1}. On the one hand, COFs with the highest mass density are observed for structures with small pore diameters which do not accept guest molecules entering the material and therefore have no accessible surface area or volume, as is evidenced by Fig. 6a and b. The COFs with the smallest mass density, on the other hand, are obtained for the most porous materials with the largest pores. For these structures, the unit cell grows rapidly, while the accessible surface area increases to a more modest extent. Therefore, the volumetric accessible surface area of Fig. 6a and b drops to zero also for these structures. In between these two regimes of very light and very heavy materials, COFs with an ideal balance between pore volume and accessible surface area are found, which combine high volumetric and gravimetric accessible surface areas with large pore volumes.

The pore diameter, and in turn also the mass density, is not only influenced by the topology and the building blocks of the material, as one could expect, but also by the linkage that connects these building blocks (see Fig. 6d). A linkage type that has a larger spatial extent, such as the azine and (acyl)hydrazone linkages, increases the space between the linker cores and,

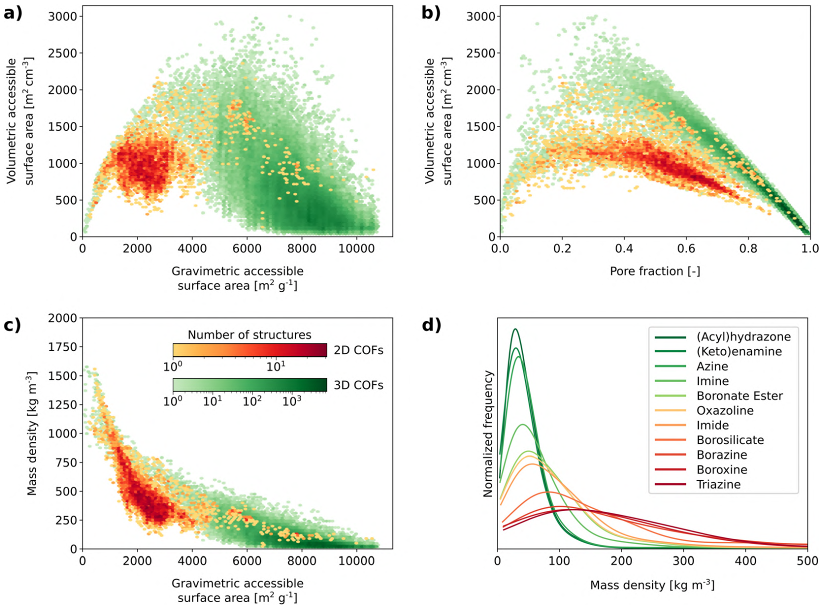


Fig. 6 Textural property–property relations of COFs in the ReDD-COFFEE database. Volumetric accessible surface area in function of (a) the gravimetric accessible surface area and (b) the pore fraction. (c) Relation between the gravimetric accessible surface area and the mass density. (d) Histograms of the mass density for each class of linkage types in the database.

therefore, also the pore diameter. This results in materials with a lower mass density as opposed to boronate ester or imide COFs, which have linkages with a smaller spatial extent. Also the linkage types that emerge during synthesis, such as the triazine and borosilicate linkages, result in higher mass densities as the linker cores of the SBUs are only separated by a single bond and are thus placed closely together. This reasoning does not take into account that interpenetrated nets can form during experimental synthesis upon increasing the pore volume, as these are described to a lesser extent in our database.[32]

These considerations are not unique to COFs. Similar trends in textural properties can be observed for other classes of nanoporous materials, such as MOFs and zeolites. In Fig. 7, a density map of the textural property–property relationships is visualized for our ReDD-COFFEE database, together with three MOF databases, *i.e.*, the experimental QMOF database[68] and the hypothetical hMOF[53] and ToBaCCo databases,[55] as well as for the database of zeolite structures of the International Zeolite Association (IZA).[97] Each relationship shows the same qualitative behavior, independent of the material class. As observed in Fig. 6a and b, the highest volumetric accessible surface areas of almost 2500 m^2 cm^{-3} are found for materials with pore fractions of 0.4 and gravimetric accessible surface areas around 4000 m^2 g^{-1}. Fig. 6c shows that these have a mass density of around 500 kg m^{-3}. As is illustrated in Fig. 6d, a higher pore fraction results in an increased gravimetric accessible surface area. COFs that approach a pore fraction of 1.0 can reach gravimetric accessible surface areas up to 9000 m^2 g^{-1}. While Fig. 6 illustrates that the COFs in our database cover the whole range of textural properties, Fig. 7 emphasizes that most of them are present in the region with the lowest mass density among the three material classes. More precisely, as is illustrated in Fig. S45 of the ESI,† 71.11% of the structures in the ReDD-COFFEE database have a pore fraction above 0.85 and a gravimetric accessible surface area larger than 7000 m^2 g^{-1}. This results in a unique combination of exceptionally high gravimetric accessible surface areas and pore volumes, together with a very large porosity as compared to MOFs and zeolites. This endows them with a large potential in adsorption applications. However, while there are COFs that have a comparable volumetric accessible surface area as the two aforementioned materials classes, most of them cover the smaller accessible surface region. Our property–property relationships show that the choice of topology, building blocks, and, importantly, linkage type, allows for a substantial freedom in the porosity reached in the synthesized COF.

4.4 High-throughput COF screening for vehicular methane storage

One of the applications for which COFs are highly promising is vehicular storage of natural gas,[142,143] due to their low mass

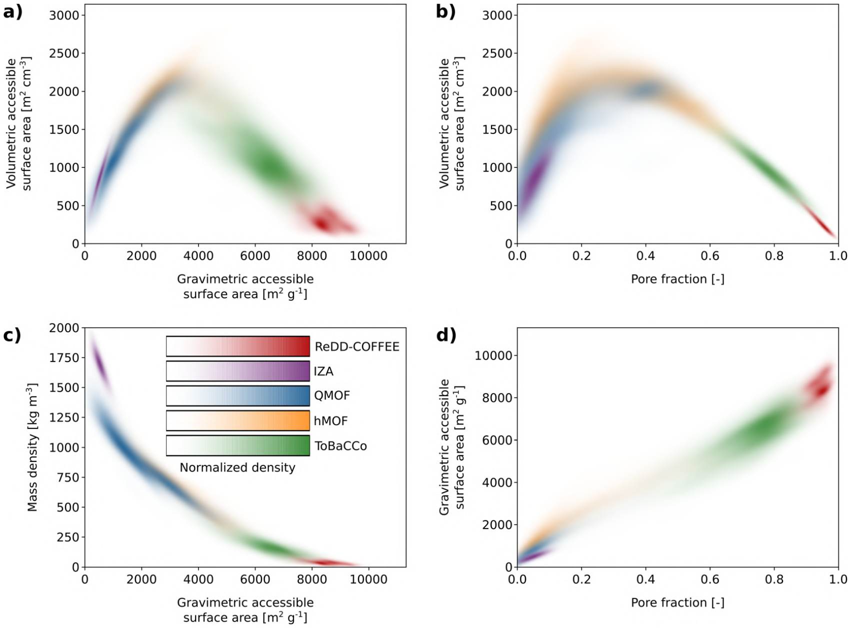


Fig. 7 Textural property–property relations for databases of different material classes. (a) Volumetric accessible surface area as a function of (a) the gravimetric accessible surface area and (b) the pore fraction. (c) Relation between the gravimetric accessible surface area and the mass density. (d) Gravimetric accessible surface area as a function of the pore fraction.

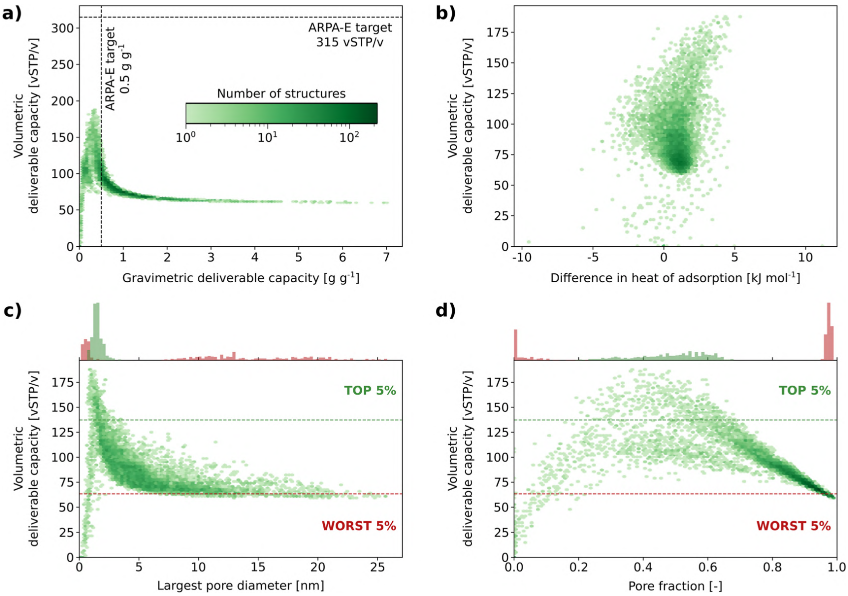


Fig. 8 Relation between vehicular methane uptake descriptors for 8394 COFs in the subset of our database. (a) Gravimetric and volumetric deliverable capacities, with the dotted lines indicating the ARPA-E targets. (b) The volumetric deliverable capacity as a function of the difference between the isosteric heat of adsorption at 65 bar and 5.8 bar. (c) and (d) The volumetric deliverable capacity in function of the largest pore diameter and pore fraction. Histograms on top of the plots indicate the regions where the top 5% (green) and worst 5% (red) performing structures in terms of the volumetric deliverable capacity are observed.

density and high internal surface (see also Fig. 7).[144,145] Natural gas, mainly consisting of methane, is an environmentally more friendly alternative than the traditional petroleum and gasoline based fuels.[142,143] However, the main challenge is that the energy density of natural gas is too low for practical applications. Several approaches to densify natural gas are proposed, such as compressed natural gas (CNG),[146] liquid natural gas (LNG),[147] and adsorbed natural gas (ANG).[148,149] In the latter approach, the gas is stored in the pores of a nanoporous material at a higher storage pressure, after which it is gradually released until the tank reaches the lower depletion pressure, when a refill is needed. The amount of gas that can be released every cycle is defined as the deliverable capacity and is a characteristic of the adopted nanoporous material. To compete with a 250 bar CNG tank, the MOVE program of the Advanced Research Projects Agency-Energy (ARPA-E) of the US Department of Energy has set targets on this deliverable capacity for nanoporous materials.[150] Candidate materials should have a volumetric deliverable capacity of at least 315 vSTP/v, whereas the gravimetric deliverable capacity should exceed 0.5 g g^{-1}.[150]

To check the performance limits of COFs for vehicular methane storage, we perform GCMC simulations on the 10 000 structures in the diverse subset of the ReDD-COFFEE database defined in Section 2.3. A storage pressure of 65 bar and a depletion pressure of 5.8 bar are applied, as imposed by the MOVE program.[51,150] For the electrostatic contribution, MBIS charges[122] are obtained from the periodic force fields and an *ab initio* calculation on a single methane molecule. The choice of van der Waals interactions is thoroughly benchmarked by reproducing the experimental adsorption isotherms of COF-1, COF-5, COF-102, and COF-103 (see Section S5.3 of the ESI†).[8] This benchmark study showed that UFF overestimates the methane uptake, while it is underestimated when the MM3 van der Waals host–guest interactions are adopted. Combining the DREIDING force field[151] to describe the host–guest interactions and the TraPPE united atom model[137] for the guest–guest interactions gives the best accuracy while maintaining a reasonable computational efficiency. Since these GCMC simulations adopt a rigid framework structure, no host–host interaction model is required.

Fig. 8a shows that the majority of the COFs in the subset (70.6%) meet the ARPA-E target of 0.5 g g^{-1} for gravimetric deliverable capacity due to their extremely low mass density. However, the lightest materials also have the largest pores and encompass the largest volumes, resulting in low volumetric deliverable capacities of about 60 vSTP/v. As discussed in Section 4.3, the heaviest materials have no accessible pores and both the volumetric and gravimetric deliverable capacities approach zero. This can also be observed in Fig. 8c and d, where the dependency of the volumetric deliverable capacity on the largest pore diameter and the pore fraction is plotted, respectively. Histograms of the structures with the 5% highest and lowest volumetric deliverable capacity support the previous claims and identify that the COFs with the highest volumetric deliverable capacity have a pore diameter between 0.7 nm and 3.4 nm and exhibit quite a broad pore fraction within the range of 0.2 to 0.7. While these materials do not have the highest gravimetric deliverable capacity, there are still structures with high volumetric deliverable capacity that meet the ARPA-E target for gravimetric deliverable capacity. The highest volumetric deliverable capacity, 187.4 vSTP/v, is observed for the boronate ester based COF ths-c3_11-01-01_06-08-01_06-08-01, which has a gravimetric deliverable capacity of 0.37 g g^{-1}. Among the structures that meet the gravimetric deliverable capacity ARPA-E target, the COF with the highest volumetric deliverable capacity is the imine COF ths-c3_11-02-04_04-03-04_04-03-04, which has a deliverable capacity of 141.1 vSTP/v and 0.50 g g^{-1}.

Whereas the methane molecules at low pressures are mainly located at the adsorption sites of the framework, they are more distributed over the pore volume at higher pressures. Therefore, the adsorption behavior is dominated by methane–framework interactions at low pressures and methane–methane interactions at high pressures, respectively. Thus, as evidenced in Fig. 8b, the deliverable capacity, *i.e.*, the difference between the methane uptake at high and low pressure, is an interplay between the two interaction types. When methane–framework interactions are less important than methane–methane interactions, the uptake at low pressures will be small and therefore, a higher deliverable capacity will be obtained. Contrarily, when the methane–framework interactions are more dominant, a lot of guests will remain bound to the framework at low pressures, decreasing the deliverable capacity. In Fig. S49 of the ESI,† the dependency of the methane uptake on the dimensionality of the framework is visualized. Despite that no COF in the database meets the ARPA-E target for the volumetric deliverable capacity, many of the here identified structures have volumetric deliverable capacities comparable with record-holding materials, such as MOF-5 (182 vSTP/v),[152] HKUST-1 (183 vSTP/v),[153] Co(bdp) (197 vSTP/v),[154] or NJU-Bai 43 (198 vSTP/v).[155] However, the hypothetical COF proposed by Mercado *et al.*[83] that consists of a carbon–carbon linked triazine framework and achieves a volumetric deliverable capacity of 216 vSTP/v, is not matched in this work, as we did not consider carbon–carbon linkages in our database. Several studies have proven that the targets imposed by the ARPA-E are too ambitious and that physical limits are preventing these objectives from being achieved by the current state-of-the-art materials.[51,156] These observations are valid also for the COFs in our ReDD-COFFEE database. However, upon functionalization, as discussed in Section 2.3, or by anchoring alkali metals to the framework,[157] an enhanced volumetric deliverable capacity could be obtained.

5 Conclusions

In this paper, the ReDD-COFFEE database of 268 687 COF structures and system-specific force fields is presented. ReDD-COFFEE is a Ready-to-use and Diverse Database of Covalent Organic Frameworks with system-specific Force field based Energy Evaluation. This database has been generated using an additive top-down approach, taking both geometric and energetic criteria into account. The structures with the lowest synthetic likelihood were filtered out using a deformation energy criterium. The combination of the structure and a system-specific force field for each COF makes the database ready-to-use in molecular simulations—for instance, for high-throughput screenings. We demonstrated the improved accuracy of the system-specific force fields over fully transferable force fields in predicting the equilibrium and single crystal geometries and experimental PXRD patterns, which provides confidence in the predictions made by our database. Furthermore, as our database is diverse in terms of geometric properties, linker cores, and linkages, it gives a representative picture of the COF material space as a whole—although we explicitly did not take a large variety of functional groups into account. Additional functionalization can be introduced to the frameworks *a posteriori*. Next to the large database, we also derived a diverse subset of 10 000 COFs with diversity metrics that are comparable or even higher than those of the full database, which allows for an efficient initial screening of the database.

We screened the textural properties of the COFs in the ReDD-COFFEE database and highlighted interesting property–property relations. Furthermore, our database is compared to databases of other material classes, *i.e.*, MOFs and zeolites, which showed that COFs possess the lowest mass densities among the studied materials. They combine high gravimetric accessible surface areas and pore fractions with low volumetric accessible surface areas. Finally, we screened the diverse subset for attractive COF storage materials for ANG in vehicular transport. Property–property relations were determined and the structural characteristics of the top candidates were selected. The highest observed volumetric deliverable capacity is comparable with the current record-holding materials. This study forms the basis for future, more elaborate screening studies focussing on gas storage and separation processes, such as carbon capture from flue gasses, as well as the mechanical stability of COFs, which will maximally benefit from the increased accuracy of the derived force fields. We hope our ReDD-COFFEE database—which is freely available on the Materials Cloud (**https://doi.org/10.24435/materialscloud:nw-3j**)—may also encourage other researchers to perform high-throughput simulations on these materials and further tap into the potential of functional COF materials.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work is supported by the Research Board of Ghent University (BOF) through a Concerted Research Action (GOA010-17). J. S. D. V. and S. M. J. R. acknowledge the Fund for Scientific Research-Flanders (FWO) for a strategic basic (SB) research fellowship (grant no. 1S94521N) and a postdoctoral fellowship (grant no. 12T3522N), respectively. V. V. S. acknowledge the Research Board of Ghent University (BOF). The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by VSC (Flemish Supercomputer Center), funded by Ghent University, FWO, and the Flemish Government—department EWI. The authors wish to thank Massimo Bocus for the design of the ReDD-COFFEE logo.

Notes and references

1 A. P. Côté, A. I. Benin, N. W. Ockwig, M. O'Keeffe, A. J. Matzger and O. M. Yaghi, *Science*, 2005, **310**, 1166–1170.
2 H. M. El-Kaderi, J. R. Hunt, J. L. Mendoza-Cortés, A. P. Côté, R. E. Taylor, M. O'Keeffe and O. M. Yaghi, *Science*, 2007, **316**, 268–272.
3 X. Feng, X. Ding and D. Jiang, *Chem. Soc. Rev.*, 2012, **41**, 6010–6022.
4 S.-Y. Ding and W. Wang, *Chem. Soc. Rev.*, 2013, **42**, 548–568.
5 C. S. Diercks and O. M. Yaghi, *Science*, 2017, **355**, eaal1585.
6 M. S. Lohse and T. Bein, *Adv. Funct. Mater.*, 2018, **28**, 1705553.
7 Y. Zeng, R. Zou and Y. Zhao, *Adv. Mater.*, 2016, **28**, 2855–2873.
8 H. Furukawa and O. M. Yaghi, *J. Am. Chem. Soc.*, 2009, **131**, 8875–8883.
9 H. Wei, S. Chai, N. Hu, Z. Yang, L. Wei and L. Wang, *Chem. Commun.*, 2015, **51**, 12178–12181.
10 N. Huang, X. Chen, R. Krishna and D. Jiang, *Angew. Chem., Int. Ed.*, 2015, **54**, 2986–2990.
11 H. Ma, H. Ren, S. Meng, Z. Yan, H. Zhao, F. Sun and G. Zhu, *Chem. Commun.*, 2013, **49**, 9773–9775.
12 Z. Li, X. Feng, Y. Zou, Y. Zhang, H. Xia, X. Liu and Y. Mu, *Chem. Commun.*, 2014, **50**, 13825–13828.
13 Q. Sun, B. Aguila, J. Perman, L. D. Earl, C. W. Abney, Y. Cheng, H. Wei, N. Nguyen, L. Wojtas and S. Ma, *J. Am. Chem. Soc.*, 2017, **139**, 2786–2793.
14 N. Huang, L. Zhai, H. Xu and D. Jiang, *J. Am. Chem. Soc.*, 2017, **139**, 2428–2434.
15 S. He, T. Zeng, S. Wang, H. Niu and Y. Cai, *ACS Appl. Mater. Interfaces*, 2017, **9**, 2959–2965.
16 S.-Y. Ding, J. Gao, Q. Wang, Y. Zhang, W.-G. Song, C.-Y. Su and W. Wang, *J. Am. Chem. Soc.*, 2011, **133**, 19816–19822.
17 Q. Fang, S. Gu, J. Zheng, Z. Zhuang, S. Qiu and Y. Yan, *Angew. Chem., Int. Ed.*, 2014, **53**, 2878–2882.
18 H.-S. Xu, S.-Y. Ding, W.-K. An, H. Wu and W. Wang, *J. Am. Chem. Soc.*, 2016, **138**, 11489–11492.
19 V. S. Vyas, F. Haase, L. Stegbauer, G. Savasci, F. Podjaski, C. Ochsenfeld and B. V. Lotsch, *Nat. Commun.*, 2015, **6**, 8508.
20 S. Chandra, T. Kundu, S. Kandambeth, R. Babarao, Y. Marathe, S. M. Kunjir and R. Banerjee, *J. Am. Chem. Soc.*, 2014, **136**, 6570–6573.
21 C. R. DeBlase, K. E. Silberstein, T.-T. Truong, H. D. Abruña and W. R. Dichtel, *J. Am. Chem. Soc.*, 2013, **135**, 16821–16824.
22 M. Dogru and T. Bein, *Chem. Commun.*, 2014, **50**, 5531–5546.
23 S. Wan, J. Guo, J. Kim, H. Ihee and D. Jiang, *Angew. Chem., Int. Ed.*, 2008, **47**, 8826–8830.
24 X. Ding, J. Guo, X. Feng, Y. Honsho, J. Guo, S. Seki, P. Maitarad, A. Saeki, S. Nagase and D. Jiang, *Angew. Chem., Int. Ed.*, 2011, **50**, 1289–1293.
25 C. Zhang, S. Zhang, Y. Yan, F. Xia, A. Huang and Y. Xian, *ACS Appl. Mater. Interfaces*, 2017, **9**, 13415–13421.
26 Q. Fang, J. Wang, S. Gu, R. B. Kaspar, Z. Zhuang, J. Zheng, H. Guo, S. Qiu and Y. Yan, *J. Am. Chem. Soc.*, 2015, **137**, 8352–8355.
27 H.-C. Zhou and S. Kitagawa, *Chem. Soc. Rev.*, 2014, **43**, 5415–5418.
28 H. Furukawa, K. E. Cordova, M. O'Keeffe and O. M. Yaghi, *Science*, 2013, **341**, 1230444.
29 R. W. Tilford, S. J. Mugavero III, P. J. Pellechia and J. J. Lavigne, *Adv. Mater.*, 2008, **20**, 2741–2746.
30 R.-R. Liang, F.-Z. Cui, R.-H. A, Q.-Y. Qi and X. Zhao, *CCS Chem.*, 2020, **2**, 139–145.
31 X. Ding, X. Feng, A. Saeki, S. Seki, A. Nagai and D. Jiang, *Chem. Commun.*, 2012, **48**, 8952–8954.
32 S. Borgmans, S. M. J. Rogge, J. S. De Vos, P. Van Der Voort and V. Van Speybroeck, *Commun. Chem.*, 2023, **6**, 5.
33 O. M. Yaghi, M. O'Keeffe, N. W. Ockwig, H. K. Chae, M. Eddaoudi and J. Kim, *Nature*, 2003, **423**, 705–714.
34 M. Eddaoudi, D. B. Moler, H. Li, B. Chen, T. M. Reineke, M. O'Keeffe and O. M. Yaghi, *Acc. Chem. Res.*, 2001, **34**, 319–330.
35 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Mater.*, 2013, **1**, 011002.
36 Y. J. Colón and R. Q. Snurr, *Chem. Soc. Rev.*, 2014, **43**, 5735–5749.
37 S. M. Moosavi, A. Nandy, K. M. Jablonka, D. Ongari, J. P. Janet, P. G. Boyd, Y. Lee, B. Smit and H. J. Kulik, *Nat. Commun.*, 2020, **11**, 4068.
38 L. Vanduyfhuys, S. Vandenbrande, T. Verstraelen, R. Schmid, M. Waroquier and V. Van Speybroeck, *J. Comput. Chem.*, 2015, **36**, 1015–1027.
39 L. Vanduyfhuys, S. Vandenbrande, J. Wieme, M. Waroquier, T. Verstraelen and V. Van Speybroeck, *J. Comput. Chem.*, 2018, **39**, 999–1011.
40 P. Z. Moghadam, S. M. J. Rogge, A. Li, C.-M. Chow, J. Wieme, N. Moharrami, M. Aragones-Anglada, G. Conduit, D. A. Gomez-Gualdron, V. Van Speybroeck and D. Fairen-Jimenez, *Matter*, 2019, **1**, 219–234.
41 L. Vanduyfhuys, T. Verstraelen, M. Vandichel, M. Waroquier and V. Van Speybroeck, *J. Chem. Theory Comput.*, 2012, **8**, 3217–3231.
42 S. Bureekaew, S. Amirjalayer, M. Tafipolsky, C. Spickermann, T. K. Roy and R. Schmid, *Phys. Status Solidi B*, 2013, **250**, 1128–1141.
43 M. Tafipolsky, S. Amirjalayer and R. Schmid, *J. Comput. Chem.*, 2007, **28**, 1169–1176.
44 M. Tafipolsky and R. Schmid, *J. Phys. Chem. B*, 2009, **113**, 1341–1352.

45 M. Tafipolsky, S. Amirjalayer and R. Schmid, *J. Phys. Chem. C*, 2010, **114**, 14402–14409.

46 J. Wieme, L. Vanduyfhuys, S. M. J. Rogge, M. Waroquier and V. Van Speybroeck, *J. Phys. Chem. C*, 2016, **120**, 14934–14947.

47 S. M. J. Rogge, A. Bavykina, J. Hajek, H. Garcia, A. I. Olivos-Suarez, A. Sepúlveda-Escribano, A. Vimont, G. Clet, P. Bazin, F. Kapteijn, M. Daturi, E. V. Ramos-Fernandez, F. X. Llabrés i Xamena, V. Van Speybroeck and J. Gascon, *Chem. Soc. Rev.*, 2017, **46**, 3134–3184.

48 R. Schmid and M. Tafipolsky, *J. Am. Chem. Soc.*, 2008, **130**, 12600–12601.

49 S. Amirjalayer, R. Q. Snurr and R. Schmid, *J. Phys. Chem. C*, 2012, **116**, 4921–4929.

50 S. Borgmans, S. M. J. Rogge, J. S. De Vos, C. V. Stevens, P. Van Der Voort and V. Van Speybroeck, *Angew. Chem., Int. Ed.*, 2021, **60**, 8913–8922.

51 C. M. Simon, J. Kim, D. A. Gomez-Gualdron, J. S. Camp, Y. G. Chung, R. L. Martin, R. Mercado, M. W. Deem, D. Gunter, M. Haranczyk, D. S. Sholl, R. Q. Snurr and B. Smit, *Energy Environ. Sci.*, 2015, **8**, 1190–1199.

52 P. G. Boyd, A. Chidambaram, E. García-Díez, C. P. Ireland, T. D. Daff, R. Bounds, A. Gładysiak, P. Schouwink, S. M. Moosavi, M. M. Maroto-Valer, J. A. Reimer, J. A. R. Navarro, T. K. Woo, S. Garcia, K. C. Stylianou and B. Smit, *Nature*, 2019, **576**, 253–256.

53 C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp and R. Q. Snurr, *Nat. Chem.*, 2012, **4**, 83–89.

54 Y. G. Chung, J. Camp, M. Haranczyk, B. J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O. K. Farha, D. S. Sholl and R. Q. Snurr, *Chem. Mater.*, 2014, **26**, 6185–6192.

55 Y. J. Colón, D. A. Gómez-Gualdrón and R. Q. Snurr, *Cryst. Growth Des.*, 2017, **17**, 5801–5810.

56 R. Anderson and D. A. Gómez-Gualdrón, *CrystEngComm*, 2019, **21**, 1653–1665.

57 J. Goldsmith, A. G. Wong-Foy, M. J. Cafarella and D. J. Siegel, *Chem. Mater.*, 2013, **25**, 3373–3382.

58 Y. J. Colón, D. Fairen-Jimenez, C. E. Wilmer and R. Q. Snurr, *J. Phys. Chem. C*, 2014, **118**, 5383–5389.

59 N. S. Bobbitt, J. Chen and R. Q. Snurr, *J. Phys. Chem. C*, 2016, **120**, 27328–27341.

60 D. A. Gómez-Gualdrón, Y. J. Colón, X. Zhang, T. C. Wang, Y.-S. Chen, J. T. Hupp, T. Yildirim, O. K. Farha, J. Zhang and R. Q. Snurr, *Energy Environ. Sci.*, 2016, **9**, 3279–3289.

61 S. Li, Y. G. Chung and R. Q. Snurr, *Langmuir*, 2016, **32**, 10368–10376.

62 G. Avci, S. Velioglu and S. Keskin, *ACS Appl. Mater. Interfaces*, 2018, **10**, 33693–33706.

63 Z. Qiao, C. Peng, J. Zhou and J. Jiang, *J. Mater. Chem. A*, 2016, **4**, 15904–15912.

64 C. Altintas, G. Avci, H. Daglar, A. N. V. Azar, I. Erucar, S. Velioglu and S. Keskin, *J. Mater. Chem. A*, 2019, **7**, 9593–9608.

65 C. Gu, J. Liu and D. S. Sholl, *J. Phys. Chem. C*, 2021, **125**, 20076–20086.

66 H. Daglar and S. Keskin, *Coord. Chem. Rev.*, 2020, **422**, 213470.

67 Y. He, E. D. Cubuk, M. D. Allendorf and E. J. Reed, *J. Phys. Chem. Lett.*, 2018, **9**, 4562–4569.

68 A. S. Rosen, S. M. Iyer, D. Ray, Z. Yao, A. Aspuru-Guzik, L. Gagliardi, J. M. Notestein and R. Q. Snurr, *Matter*, 2021, **4**, 1578–1597.

69 A. S. Rosen, V. Fung, P. Huck, C. T. O'Donnell, M. K. Horton, D. G. Truhlar, K. A. Persson, J. M. Notestein and R. Q. Snurr, *npj Comput. Mater.*, 2022, **8**, 112.

70 A. S. Rosen, J. M. Notestein and R. Q. Snurr, *J. Comput. Chem.*, 2019, **40**, 1305–1318.

71 F. H. Allen, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, **58**, 380–388.

72 Y. G. Chung, E. Haldoupis, B. J. Bucior, M. Haranczyk, S. Lee, H. Zhang, K. D. Vogiatzis, M. Milisavljevic, S. Ling, J. S. Camp, B. Slater, J. I. Siepmann, D. S. Sholl and R. Q. Snurr, *J. Chem. Eng. Data*, 2019, **64**, 5985–5998.

73 P. Z. Moghadam, A. Li, S. B. Wiggin, A. Tao, A. G. P. Maloney, P. A. Wood, S. C. Ward and D. Fairen-Jimenez, *Chem. Mater.*, 2017, **29**, 2618–2625.

74 C. Mellot Draznieks, J. M. Newsam, A. M. Gorman, C. M. Freeman and G. Férey, *Angew. Chem., Int. Ed.*, 2000, **39**, 2270–2275.

75 M. Fernandez, P. G. Boyd, T. D. Daff, M. Z. Aghaji and T. K. Woo, *J. Phys. Chem. Lett.*, 2014, **5**, 3056–3060.

76 B. J. Sikora, R. Winnegar, D. M. Proserpio and R. Q. Snurr, *Microporous Mesoporous Mater.*, 2014, **186**, 207–213.

77 R. L. Martin and M. Haranczyk, *Cryst. Growth Des.*, 2014, **14**, 2431–2440.

78 M. A. Addicoat, D. E. Coupry and T. Heine, *J. Phys. Chem. A*, 2014, **118**, 9607–9614.

79 S. Bureekaew, V. Balwani, S. Amirjalayer and R. Schmid, *CrystEngComm*, 2015, **17**, 344–352.

80 P. G. Boyd and T. K. Woo, *CrystEngComm*, 2016, **18**, 3777–3792.

81 M. Tong, Y. Lan, Z. Qin and C. Zhong, *J. Phys. Chem. C*, 2018, **122**, 13009–13016.

82 R. L. Martin, C. M. Simon, B. Medasani, D. K. Britt, B. Smit and M. Haranczyk, *J. Phys. Chem. C*, 2014, **118**, 23790–23802.

83 R. Mercado, R.-S. Fu, A. V. Yakutovich, L. Talirz, M. Haranczyk and B. Smit, *Chem. Mater.*, 2018, **30**, 5069–5086.

84 M. Tong, W. Zhu, J. Li, Z. Long, S. Zhao, G. Chen and Y. Lan, *Chem. Commun.*, 2020, **56**, 6376–6379.

85 E. Gülçay and İ. F. Eruçar, *J. Turk. Chem. Soc., Sect. A*, 2020, **7**, 65–76.

86 D. Ongari, A. V. Yakutovich, L. Talirz and B. Smit, *ACS Cent. Sci.*, 2019, **5**, 1663–1675.

87 K. S. Deeg, D. Damasceno Borges, D. Ongari, N. Rampal, L. Talirz, A. V. Yakutovich, J. M. Huck and B. Smit, *ACS Appl. Mater. Interfaces*, 2020, **12**, 21559–21568.

88 O. F. Altundal, C. Altintas and S. Keskin, *J. Mater. Chem. A*, 2020, **8**, 14609–14623.

89 M. Tong, Y. Lan, Q. Yang and C. Zhong, *Chem. Eng. Sci.*, 2017, **168**, 456–464.

90 Y. Lan, M. Tong, Q. Yang and C. Zhong, *CrystEngComm*, 2017, **19**, 4920–4926.

91 W. Li, X. Xia and S. Li, *ACS Appl. Mater. Interfaces*, 2019, **12**, 3265–3273.

92 S. Kandambeth, A. Mallick, B. Lukose, M. V. Mane, T. Heine and R. Banerjee, *J. Am. Chem. Soc.*, 2012, **134**, 19524–19527.

93 P. Puthiaraj, Y.-R. Lee, S. Zhang and W.-S. Ahn, *J. Mater. Chem. A*, 2016, **4**, 16288–16311.

94 S.-Y. Ding, M. Dong, Y.-W. Wang, Y.-T. Chen, H.-Z. Wang, C.-Y. Su and W. Wang, *J. Am. Chem. Soc.*, 2016, **138**, 3031–3037.

95 R. Anderson and D. A. Gómez-Gualdrón, *Chem. Mater.*, 2020, **32**, 8106–8119.

96 M. W. Deem, R. Pophale, P. A. Cheeseman and D. J. Earl, *J. Phys. Chem. C*, 2009, **113**, 21353–21360.

97 *Database of zeolite structures*, **http://www.iza-structure.org/databases/**, Accessed: 2022-08-01.

98 M. O'Keeffe, M. A. Peskov, S. J. Ramsden and O. M. Yaghi, *Acc. Chem. Res.*, 2008, **41**, 1782–1789.

99 M. Li, D. Li, M. O'Keeffe and O. M. Yaghi, *Chem. Rev.*, 2014, **114**, 1343–1370.

100 F. J. Uribe-Romo, J. R. Hunt, H. Furukawa, C. Klöck, M. O'Keeffe and O. M. Yaghi, *J. Am. Chem. Soc.*, 2009, **131**, 4570–4571.

101 P. Kuhn, M. Antonietti and A. Thomas, *Angew. Chem., Int. Ed.*, 2008, **47**, 3450–3453.

102 F. J. Uribe-Romo, C. J. Doonan, H. Furukawa, K. Oisaki and O. M. Yaghi, *J. Am. Chem. Soc.*, 2011, **133**, 11478–11481.

103 S. Dalapati, S. Jin, J. Gao, Y. Xu, A. Nagai and D. Jiang, *J. Am. Chem. Soc.*, 2013, **135**, 17310–17313.

104 Q. Fang, Z. Zhuang, S. Gu, R. B. Kaspar, J. Zheng, J. Wang, S. Qiu and Y. Yan, *Nat. Commun.*, 2014, **5**, 4503.

105 J. R. Hunt, C. J. Doonan, J. D. LeVangie, A. P. Côté and O. M. Yaghi, *J. Am. Chem. Soc.*, 2008, **130**, 11872–11873.

106 D. A. Pyles, J. W. Crowe, L. A. Baldwin and P. L. McGrier, *ACS Macro Lett.*, 2016, **5**, 1055–1058.

107 K. T. Jackson, T. E. Reich and H. M. El-Kaderi, *Chem. Commun.*, 2012, **48**, 8823–8825.

108 W. Kabsch, *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.*, 1976, **32**, 922–923.

109 J. P. Janet and H. J. Kulik, *J. Phys. Chem. A*, 2017, **121**, 8939–8954.

110 A. Nandy, C. Duan, J. P. Janet, S. Gugler and H. J. Kulik, *Ind. Eng. Chem. Res.*, 2018, **57**, 13973–13986.

111 J. P. Janet, F. Liu, A. Nandy, C. Duan, T. Yang, S. Lin and H. J. Kulik, *Inorg. Chem.*, 2019, **58**, 10592–10606.

112 A. Nandy, J. Zhu, J. P. Janet, C. Duan, R. B. Getman and H. J. Kulik, *ACS Catal.*, 2019, **9**, 8243–8255.

113 B. L. Bush, C. I. Bayly and T. A. Halgren, *J. Comput. Chem.*, 1999, **20**, 1495–1516.

114 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *J. Cheminf.*, 2012, **4**, 17.

115 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr, J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16 Revision C.01*, Gaussian Inc. Wallingford CT, 2016.

116 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

117 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.

118 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.

119 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

120 M. J. Frisch, J. A. Pople and J. S. Binkley, *J. Chem. Phys.*, 1984, **80**, 3265–3269.

121 N. L. Allinger, Y. H. Yuh and J. H. Lii, *J. Am. Chem. Soc.*, 1989, **111**, 8551–8566.

122 T. Verstraelen, S. Vandenbrande, F. Heidar-Zadeh, L. Vanduyfhuys, V. Van Speybroeck, M. Waroquier and P. W. Ayers, *J. Chem. Theory Comput.*, 2016, **12**, 3894–3912.

123 T. Verstraelen, P. Tecmer, F. Heidar-Zadeh, C. E. González-Espinoza, M. Chan, T. D. Kim, K. Boguslawski, S. Fias, S. Vandenbrande, D. Berrocal and P. W. Ayers, *HORTON*, 2017, 2(1), **https://github.com/theochem/horton**.

124 A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.

125 T. Verstraelen, L. Vanduyfhuys, S. Vandenbrande and S. M. J. Rogge, *Yaff, yet another force field (v1.6.0)*, **https://molmod.ugent.be/software/**.

126 A. Ghysels, T. Verstraelen, K. Hemelsoet, M. Waroquier and V. Van Speybroeck, *J. Chem. Inf. Model.*, 2010, **50**, 1736–1750.

127 T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza and M. Haranczyk, *Microporous Mesoporous Mater.*, 2012, **149**, 134–141.

128 A. F. Ismail, K. C. Khulbe and T. Matsuura, in *Fundamentals of gas permeation through membranes*, Springer International Publishing, Cham, 2015, pp. 11–35.

129 S. Nosé, *Mol. Phys.*, 1984, **52**, 255–268.

130 W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.

131 G. J. Martyna, M. L. Klein and M. Tuckerman, *J. Chem. Phys.*, 1992, **97**, 2635–2643.

132 G. J. Martyna, D. J. Tobias and M. L. Klein, *J. Chem. Phys.*, 1994, **101**, 4177–4189.

133 G. J. Martyna, M. E. Tuckerman, D. J. Tobias and M. L. Klein, *Mol. Phys.*, 1996, **87**, 1117–1157.

134 V. Favre-Nicolin and R. Černý, *J. Appl. Crystallogr.*, 2002, **35**, 734–743.

135 D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, Elsevier, 2001, vol. 1.

136 D. Dubbeldam, S. Calero, D. E. Ellis and R. Q. Snurr, *Mol. Simul.*, 2016, **42**, 81–101.

137 M. G. Martin and J. I. Siepmann, *J. Phys. Chem. B*, 1998, **102**, 2569–2577.

138 Z. Yang, H. Chen, S. Wang, W. Guo, T. Wang, X. Suo, D.-e. Jiang, X. Zhu, I. Popovs and S. Dai, *J. Am. Chem. Soc.*, 2020, **142**, 6856–6860.

139 T. Ma, E. A. Kapustin, S. X. Yin, L. Liang, Z. Zhou, J. Niu, L.-H. Li, Y. Wang, J. Su, J. Li, X. Wang, W. D. Wang, W. Wang, J. Sun and O. M. Yaghi, *Science*, 2018, **361**, 48–52.

140 Y.-B. Zhang, J. Su, H. Furukawa, Y. Yun, F. Gándara, A. Duong, X. Zou and O. M. Yaghi, *J. Am. Chem. Soc.*, 2013, **135**, 16336–16339.

141 K. Geng, T. He, R. Liu, S. Dalapati, K. T. Tan, Z. Li, S. Tao, Y. Gong, Q. Jiang and D. Jiang, *Chem. Rev.*, 2020, **120**, 8814–8933.

142 S. Yeh, *Energy Policy*, 2007, **35**, 5865–5875.

143 M. Q. Wang and H. S. Huang, *A full fuel-cycle analysis of energy and emissions impacts of transportation fuels produced from natural gas*, 2000, **https://www.osti.gov/biblio/750803**.

144 J. L. Mendoza-Cortés, S. S. Han, H. Furukawa, O. M. Yaghi and W. A. Goddard, *J. Phys. Chem. A*, 2010, **114**, 10824–10833.

145 J. L. Mendoza-Cortes, T. A. Pascal and W. A. Goddard, *J. Phys. Chem. A*, 2011, **115**, 13852–13857.

146 M. I. Khan, T. Yasmin and A. Shakoor, *Renewable Sustainable Energy Rev.*, 2015, **51**, 785–797.

147 W. Lim, K. Choi and I. Moon, *Ind. Eng. Chem. Res.*, 2013, **52**, 3065–3088.

148 V. Menon and S. Komarneni, *J. Porous Mater.*, 1998, **5**, 43–58.

149 J. Wegrzyn and M. Gurevich, *Appl. Energy*, 1996, **55**, 71–83.

150 DE-FOA-0000672, *Methane Opportunities For Vehicular Energy (Move)*, **https://arpa-e-foa.energy.gov/Default.aspx?Search=DE-FOA-0000672**, Accessed: 2022-09-21.

151 S. L. Mayo, B. D. Olafson and W. A. Goddard, *J. Phys. Chem.*, 1990, **94**, 8897–8909.

152 J. A. Mason, M. Veenstra and J. R. Long, *Chem. Sci.*, 2014, **5**, 32–51.

153 J. Möllmer, A. Möller, F. Dreisbach, R. Gläser and R. Staudt, *Microporous Mesoporous Mater.*, 2011, **138**, 140–148.

154 J. A. Mason, J. Oktawiec, M. K. Taylor, M. R. Hudson, J. Rodriguez, J. E. Bachman, M. I. Gonzalez, A. Cervellino, A. Guagliardi, C. M. Brown, P. L. Llewellyn, N. Masciocchi and J. R. Long, *Nature*, 2015, **527**, 357–361.

155 M. Zhang, W. Zhou, T. Pham, K. A. Forrest, W. Liu, Y. He, H. Wu, T. Yildirim, B. Chen, B. Space, Y. Pan, M. J. Zaworotko and J. Bai, *Angew. Chem., Int. Ed.*, 2017, **56**, 11426–11430.

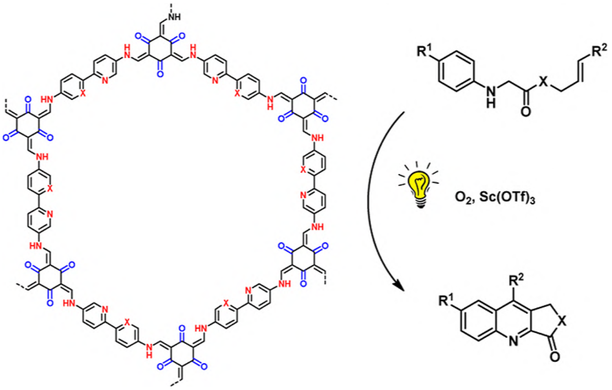
156 Y. Peng, V. Krungleviciute, I. Eryazici, J. T. Hupp, O. K. Farha and T. Yildirim, *J. Am. Chem. Soc.*, 2013, **135**, 11887–11894.

157 J. Lan, D. Cao and W. Wang, *Langmuir*, 2010, **26**, 220–226.

Paper IV

Engineering of Phenylpyridine- and Bipyridine-Based Covalent Organic Frameworks for Photocatalytic Tandem Aerobic Oxidation/Povarov Cyclization

Maarten Debruyne, Sander Borgmans, Sambhu Radhakrishnan, Eric Breynaert, Henk Vrielinck, Karen Leus, Andreas Laemont, Juul De Vos, Kuber Singh Rawat, Siebe Vanlommel, Hannes Rijckaert, Hadi Salemi, Jonas Everaert, Flore Vanden Bussche, Dirk Poelman, Rino Morent, Nathalie De Geyter, Pascal Van Der Voort, Veronique Van Speybroeck, and Christian V. Stevens*



ACS Applied Materials & Interfaces, **2023**, 15 (29), 35092–35106

The Supporting Information is available at:
https://doi.org/10.1021/acsami.3c07036

J. S. De Vos generated *ab initio* reference data for each of the three SBUs in this work, from which cluster force fields were derived. By combining them on the **hcb** topology, structural models were generated, together with a system-specific force field for each of them, and subsequently optimized.

*Corresponding author

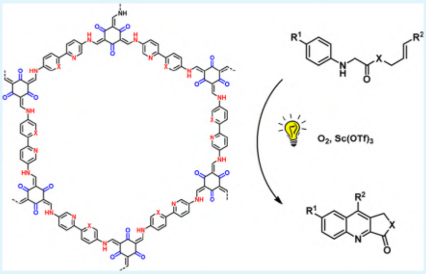
ACS APPLIED MATERIALS & INTERFACES

www.acsami.org Research Article

Engineering of Phenylpyridine- and Bipyridine-Based Covalent Organic Frameworks for Photocatalytic Tandem Aerobic Oxidation/Povarov Cyclization

Maarten Debruyne, Sander Borgmans, Sambhu Radhakrishnan, Eric Breynaert, Henk Vrielinck, Karen Leus, Andreas Laemont, Juul De Vos, Kuber Singh Rawat, Siebe Vanlommel, Hannes Rijckaert, Hadi Salemi, Jonas Everaert, Flore Vanden Bussche, Dirk Poelman, Rino Morent, Nathalie De Geyter, Pascal Van Der Voort, Veronique Van Speybroeck, and Christian V. Stevens*

Cite This: *ACS Appl. Mater. Interfaces* 2023, 15, 35092−35106 Read Online



ABSTRACT: Covalent organic frameworks (COFs) are emerging as a new class of photoactive organic semiconductors, which possess crystalline ordered structures and high surface areas. COFs can be tailor-made toward specific (photocatalytic) applications, and the size and position of their band gaps can be tuned by the choice of building blocks and linkages. However, many types of building blocks are still unexplored as photocatalytic moieties and the scope of reactions photocatalyzed by COFs remains quite limited. In this work, we report the synthesis and application of two bipyridine- or phenylpyridine-based COFs: **TpBpyCOF** and **TpPpyCOF**. Due to their good photocatalytic properties, both materials were applied as metal-free photocatalysts for the tandem aerobic oxidation/Povarov cyclization and α-oxidation of *N*-aryl glycine derivatives, with the bipyridine-based **TpBpyCOF** exhibiting the highest activity. By expanding the range of reactions that can be photocatalyzed by COFs, this work paves the way toward the more widespread application of COFs as metal-free heterogeneous photocatalysts as a convenient alternative for commonly used homogeneous (metal-based) photocatalysts.

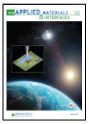
KEYWORDS: *covalent organic frameworks, photocatalysis, oxidation, Povarov reaction, heterogeneous catalysis*

■ INTRODUCTION

A booming world population and economy put ever-increasing pressure on the environment. Together with growing concerns about pollution, global warming, and the finite supply of fossil resources, this has led to the development of many different green technologies. One promising avenue is visible light photocatalysis, which utilizes the energy of visible light and converts it into chemical energy for organic reactions through a catalyst.[1−3] Based on whether the photocatalyst and the reactants exist in the same phase or not, it is classified as a homogeneous or heterogeneous photocatalyst, respectively. Typical examples of the former include metal complexes (mainly ruthenium and iridium complexes)[4] and organic dyes (fluorenone, acridinium-based photocatalysts, xanthene dyes, *etc.*),[2] whereas the latter category mainly contains inorganic semiconductors (TiO_2, CdSe, WO_3, *etc.*).[5−7] However, homogeneous catalysts, and especially the precious metal-based ones, have drawbacks such as high costs, difficult separation from the product, and low or no recyclability.

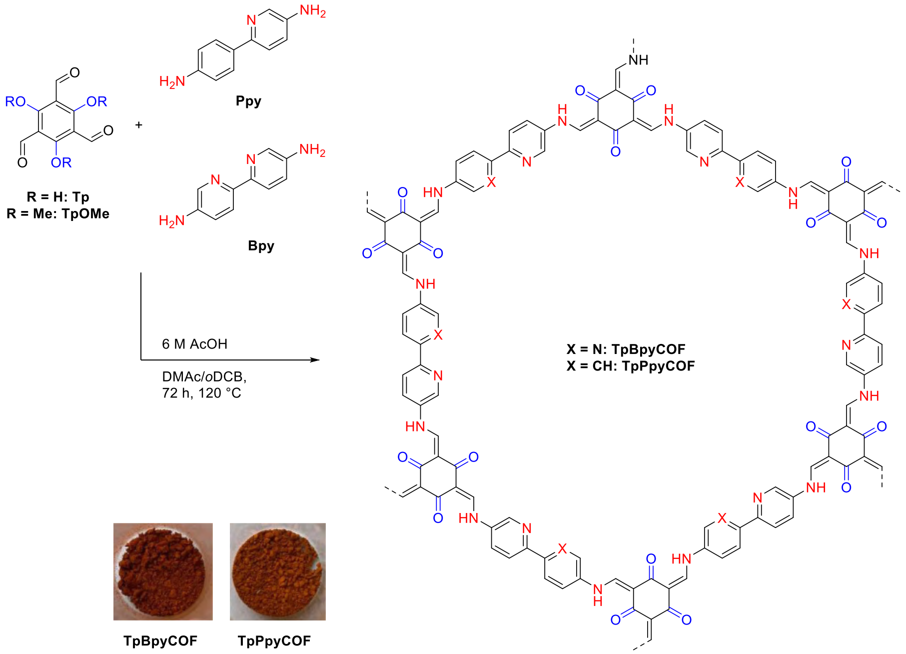
In recent years, new, heterogeneous, polymer-based photocatalysts such as graphitic carbon nitride (g-C_3N_4),[8] metal organic frameworks (MOFs),[9] and porous organic polymers (POPs)[10] have emerged to help eliminate these aforementioned deficiencies. Covalent organic frameworks (COFs) are a subclass of POPs, and like POPs, they are constructed completely from organic building blocks. Their defining characteristic is their long-range order and accompanying crystallinity. Due to their low weight, high specific surface area, and modular design, they have found application in diverse fields, such as gas sorption and separation,[11] energy storage,[12] sensing,[13] and heterogeneous catalysis.[14] COFs hold a lot of promise as heterogeneous photocatalysts, with great recyclability due to their insolubility. Their extended π-conjugated frameworks, regular pore structures, and high surface areas are

Received: May 18, 2023
Accepted: July 5, 2023
Published: July 18, 2023



ACS Publications

Scheme 1. Schematic Illustration of the Synthesized COFs[a]



[a]When **TpOMe** was condensed with **Bpy**, a material largely corresponding to **TpBpyCOF** was formed (See Supporting Information Section S2.7 for more details).

beneficial for obtaining high photocatalytic activities due to the good light-absorbing capacity and high accessibility of their active sites.[15,16] Due to their tailor-made character, the optical properties and photocatalytic efficiency of COFs can be easily adjusted by the choice of building blocks.[17−19] Nitrogen-rich moieties, when incorporated in extended frameworks, are known to confer significant photocatalytic activity. While nitrogen-rich motifs such as triazines,[16,20] porphyrins,[21,22] and benzothiadiazoles[23,24] have been applied extensively in metal-free heterogeneous photocatalysts, bipyridines and especially phenylpyridines have been less explored. Moreover, despite the advantages and the great amount of research currently being done on COFs as photocatalysts, the scope of reactions that are catalyzed by photoactive COFs is still quite limited. Mainly, three reactions are studied: the aerobic oxidation of benzylamines, the hydroxylation of aryl boronic acids, and the oxidation of sulfides. To expand this scope, we envisaged that the oxidation and subsequent transformations of *N*-aryl glycine derivatives would be feasible with COFs.

Herein, we describe the synthesis of two bipyridine- or phenylpyridine-based COFs: **TpBpyCOF** and **TpPpyCOF** and their application as photocatalysts. To evaluate their efficacy, we applied these materials in the COF- and Lewis acid-catalyzed photocatalytic aerobic oxidation/Povarov cyclization, allowing for the easy synthesis of highly substituted quinolines. Moreover, the same substrates could be converted into α-dicarbonyl compounds when the Lewis acid was omitted. The results described in this work broaden the scope of COF-catalyzed photocatalytic reactions and thus facilitate the advancement of these emerging materials as alternatives for homogeneous (metal-based) catalysts.

■ RESULTS AND DISCUSSION

Synthesis and Characterization of COFs. Two-dimensional (2D) COFs with one-dimensional open channels were synthesized by condensing 2,2′-bipyridine-5,5′-diamine (**Bpy**) or 6-(4-aminophenyl)pyridin-3-amine (**Ppy**) with 1,3,5-triformylphloroglucinol (**Tp**) by solvothermal synthesis as shown in Scheme 1. **TpBpyCOF** is already well known in the literature[25−29] and has been applied as a proton conductive solid-state electrolyte,[25] as a carrier material for palladium or cobalt to use as a heterogeneous catalyst for benzofuran synthesis[26] or for electrochemical water oxidation,[27] as a fluorescence sensing material,[28] and recently as a metal-free photocatalyst for the production of hydrogen peroxide.[29] The newly synthesized **TpPpyCOF** is the first COF synthesized using **Ppy** as a building block, and this material has potential not only as a photocatalyst but also as a platform to complex metals with its phenylpyridine moiety. Moreover, we improved the existing synthesis of **Ppy** by employing a Suzuki coupling reaction followed by reduction (see Supporting Information, Section S1.4) instead of the four-step approach used previously.[30] Next to these materials, another COF was made by condensation of **Bpy** with 2,4,6-trimethoxybenzene-1,3,5-tricarbaldehyde (**TpOMe**), which is known to give imine-linked COFs with exceptional stability.[31] While this COF has already

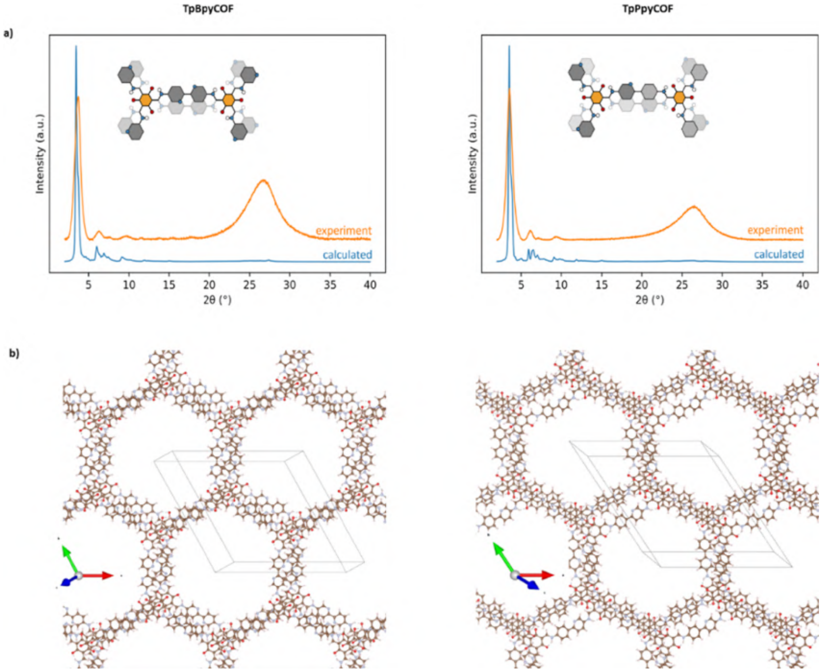


Figure 1. (a) PXRD comparison for **TpBpyCOF** and **TpPpyCOF** between experimental measurements and calculated MD averages. (b) Simulation-derived structure for **TpBpyCOF** and **TpPpyCOF**.

been reported, it was synthesized *via* a mechanochemical reaction.[31] Interestingly, during the acidic conditions of the solvothermal COF synthesis, unexpected demethylation of this material took place as signified by solid-state NMR, and a COF was formed that corresponded with **TpBpyCOF** (see Supporting Information Section S2.7).

PXRD patterns of the synthesized materials indicated relatively sharp reflections (Figure 1a), pointing to good crystallinity of the resulting COFs. The peaks around 2θ = 4° are assigned to the (100) plane and the smaller peak around 7° to the (200) plane. The broad peak around 25° corresponds to the (001) plane and originates from $\pi-\pi$ stacking between the individual layers of the COF.[32,33] SEM and TEM images of the COFs are shown in Figure 2a,b. **TpBpyCOF** and **TpPpyCOF** form large aggregates with relatively rough surfaces. The TEM images (Figure 2b) show the crystalline honeycomb-like structure. Bright-Field (BF) TEM images of **TpBpyCOF** and **TpPpyCOF** with their corresponding EDX elemental map of carbon, nitrogen, and oxygen demonstrate the uniform distributions of these atoms throughout the framework (Figures S11−S13). The permanent porosity of the synthesized COFs was assessed by N_2 sorption measurements at 77 K (Figure 3b). The calculated Brunauer−Emmett−Teller surface areas for **TpBpyCOF** and **TpPpyCOF** were 879 and 755 m^2/g, respectively. The resulting pore size distributions (PSDs) are given in Figures S1 and S2 and the pore diameters were 17.9 and 18.5 Å for **TpBpyCOF** and **TpPpyCOF**, respectively. The simulated PSDs are shown in Figures S40−S41 and possess a similar mean pore diameter but exhibit a much narrower distribution than the experimental PSDs.

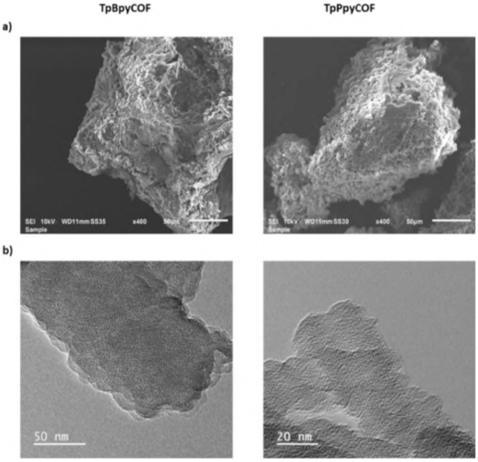


Figure 2. (a) SEM and (b) TEM images of **TpBpyCOF** and **TpPpyCOF**.

Fourier transform infrared (FTIR) spectroscopy of these materials confirmed the completion of the reaction and the

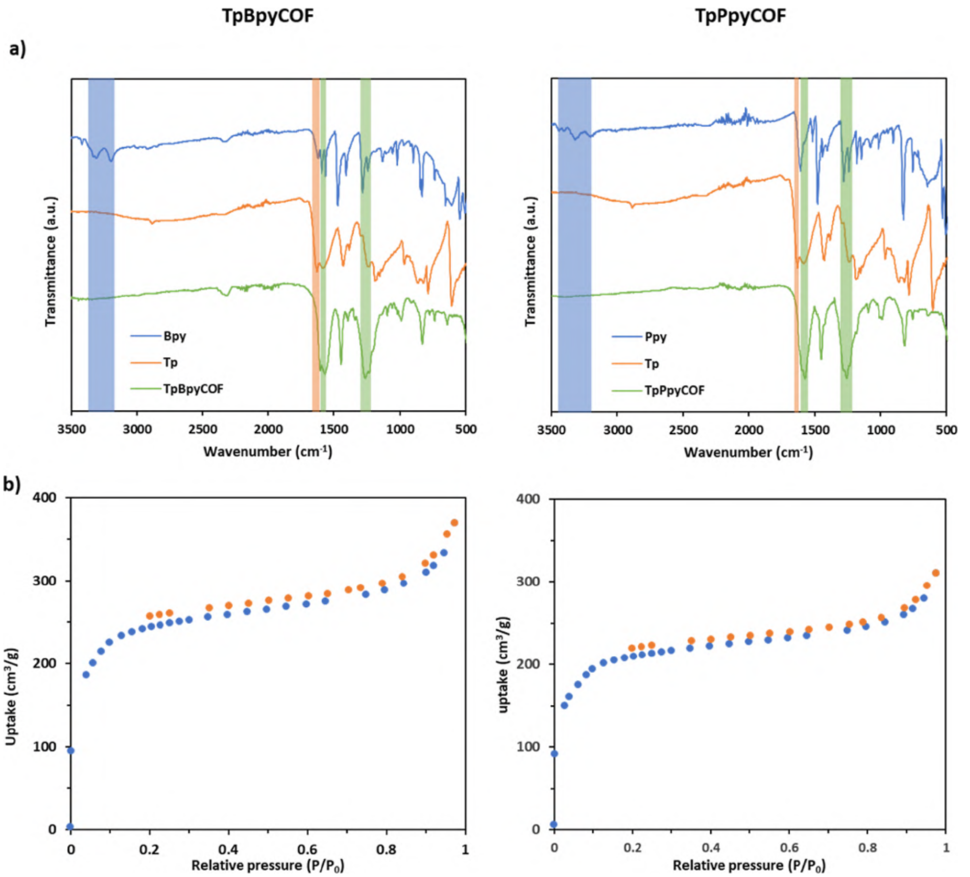


Figure 3. (a, left) Overlay of the FTIR spectra of **Bpy** (top, blue), **Tp** (middle, orange), and **TpBpyCOF** (bottom, green). (a, right) Overlay of the FTIR spectra of **Ppy** (top, blue), **Tp** (middle, orange), and **TpPpyCOF** (bottom, green). Characteristic signals originating from the amino groups, aldehydes, and β-ketoenamine linkages are marked in blue, orange, and green, respectively. (b) Nitrogen adsorption/desorption isotherms at 77 K of **TpBpyCOF** and **TpPpyCOF**.

formation of β-ketoenamine linkages (Figures 3a and S4−S10). Disappearance of the characteristic absorption bands of the **Bpy** amino group (3180−3418 cm^{-1}) and the **Tp** carbonyl (1632 cm^{-1}) in the FTIR spectrum of **TpBpyCOF** points toward a successful reaction. New peaks corresponding with the keto functionality are observed at 1604 (C═O) and 1564 cm^{-1} (C═C).[25] The peaks around 1260 cm^{-1} correspond with the C−N bond.[28] For **TpPpyCOF**, similar results were obtained; the characteristic amine absorption bands (3192−3443 cm^{-1}) disappear and new bands corresponding with the C═O and C═C absorption, which are nearly merged in this material, appear around 1595 and 1568 cm^{-1}. For additional structural characterization, the N and C 1s XPS spectra were measured. The N 1s spectra of **TpBpyCOF** and **TpPpyCOF** are shown in Figure S14. The peaks at 398.8 and 398.5 eV for **TpBpyCOF** and **TpPpyCOF**, respectively, correspond to the C═N moieties (pyridinic nitrogens). The peak at 399.9 eV for both **TpBpyCOF** and **TpPpyCOF** corresponds to the nitrogens of the enamine linkages (pyrrolic nitrogens). The N 1s spectrum of **TpBpyCOF** was deconvoluted in another smaller peak, at 401.2 eV, which was assigned to quaternary nitrogens, formed through side reactions or protonation. The C 1s spectra of **TpBpyCOF** and **TpPpyCOF** were deconvoluted into three main peaks. One was at 284.5 and 284.8 eV, for **TpBpyCOF** and **TpPpyCOF** respectively, corresponding to the C−C moieties, another at 285.6 and 286.0 eV corresponding to the C−N moieties, and a third peak at 287.9 and 288.3 eV corresponding to the carbonyl moieties.

Multinuclear NMR spectroscopic studies were performed to obtain information about the local chemical structures of these materials. The ^{1}H MAS NMR spectra (Figures 4a,b and S15) of the three COFs, **TpPPyCOF**, **TpBpyCOF** and "**TpOMeBPy-COF**", exhibited three broad resonances between 0 and 5, 5 and 10, and 10 and 15 ppm. The resonance between 5 and 10 ppm is assigned to the aromatic and olefinic protons of the linker moieties, while the downfield shifted resonance between 10 and

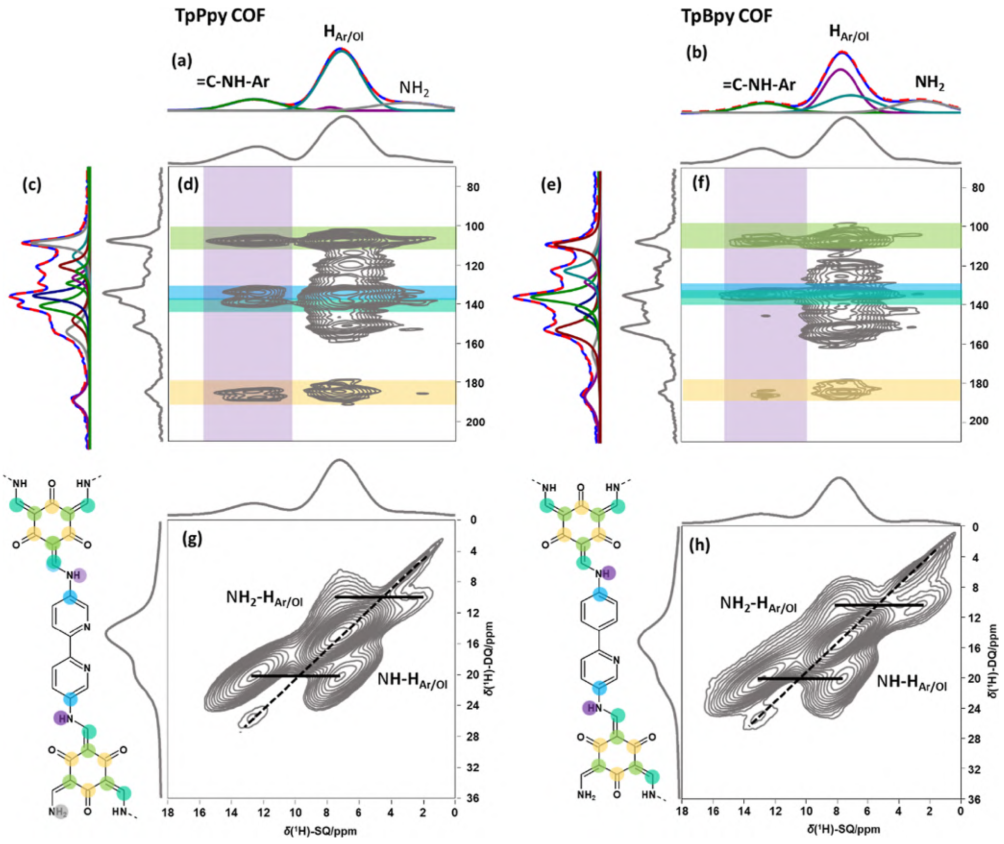


Figure 4. (a,b) ^{1}H direct excitation spectrum of **TpPpyCOF** and **TpBPyCOF**. (c) ^{1}H-^{13}C CPMAS spectrum with spectral decomposition. (d) ^{1}H-^{13}C CP-HETCOR spectrum of **TpPpyCOF**. (e) ^{1}H-^{13}C CPMAS with spectral decomposition. (f) ^{1}H-^{13}C CP-HETCOR spectrum of **TpBpyCOF**. (g,h) ^{1}H-^{1}H DQ-SQ NMR spectra of **TpPpyCOF** and **TpBpyCOF**. All the spectra were acquired at 18.8 T under 35 kHz MAS at a sample temperature of 295 K.

15 ppm is attributed to the enamine ^{1}H. The upfield resonance between 0 and 5 ppm is not an expected resonance considering the ideal chemical structure and could be attributed to the terminal $-NH_2$ groups or a composite resonance of strongly adsorbed water with the NH_2 terminals, strongly adsorbed solvents, *etc.* In the case of **"TpOMeBpyCOF"**, the presence of the enamine ^{1}H resonance and the absence of the methoxy resonance in the ^{1}H NMR spectrum (Figure S15) point toward an unexpected demethylation reaction during the COF synthesis. DQ-SQ NMR spectroscopy is a probe for the close spatial proximity between different nuclei (Figure 4g,h).[34] The existence of correlations along the diagonal (autocorrelations) indicates the appearance as pairs for the respective nuclei and the appearance of off-diagonal correlations (cross correlations) at the sum of the respective frequencies indicates their existence in close spatial proximity. The presence of the cross correlation (see Figure 4g,h, indicated as **NH-H**$_{Ar/Ol}$) of the enamine resonance with the aromatic and olefinic ^{1}H's indicates the successful linking between the **Tp** and **Bpy** or **Ppy** linkers in **TpBpyCOF** and **TpPpyCOF**, respectively. The cross correlation between the upfield ^{1}H resonance (0−5 ppm) and the aromatic/olefinic resonance (labeled as **NH**$_2$-**H**$_{Ar/Ol}$) points toward its existence in the same phase and could be attributed to the existence of terminal NH_2 groups. The ^{1}H-^{13}C CPMAS spectrum of all the COF samples exhibited the presence of carbonyl resonances around 185 ppm and multiple overlapping aromatic/olefinic ^{13}C resonances between 108 and 151 ppm (Figure S16). In line with the ^{1}H NMR spectrum, the methoxy resonances were absent in the case of **"TpOMeBpyCOF"**, confirming its demethylation, and this material exhibited resonances similar to that of **TpBpyCOF**. The ^{1}H-^{13}C CP-HETCOR spectra (Figure 4d,f) evidenced the correlation of the enamine ^{1}H resonance to the different ^{13}C nuclei in both the linkers, further confirming the local chemical structure. The correlations of the enamine resonance to the **Bpy** or **Ppy** linkers are depicted by the blue horizontal bar, while the correlation to the **Tp** linker is indicated by the yellow, green, and cyan horizontal bars. Modeling of the chemical shielding parameters

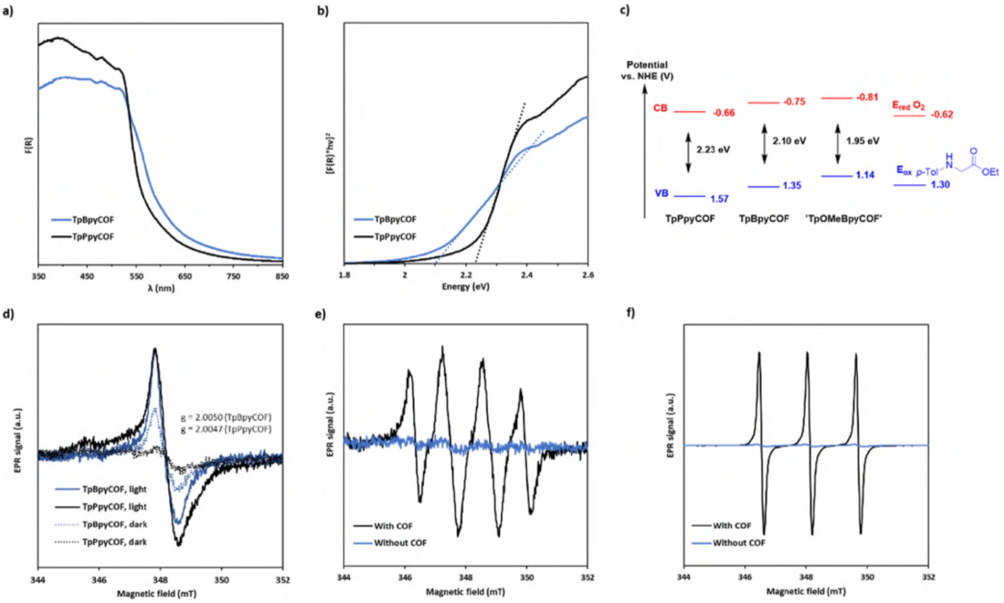


Figure 5. (a) Kubelka−Munk absorption spectra, (b) Tauc plots, (c) band structure, and (d) EPR spectra of **TpBpyCOF** and **TpPpyCOF**. (e) EPR spectrum of DMPO-OOH•. (f) TEMPO produced by **TpBpyCOF** under visible light irradiation.

(GIPAW VASP) was performed and the theoretically estimated chemical shifts were in perfect agreement with the experimental spectra for **TpBpyCOF** and **TpPpyCOF**, allowing assignment of all resonances, and confirmed the local chemical structure of these materials (Supporting Information, Section S3.6).

The atomic structure of the materials was further determined through computational modeling (see Section S3) using the measured diffraction patterns as reference data. Through judicious comparison of the experimental diffraction patterns with a collection of calculated diffraction patterns, corresponding to different possible structure models, the most representative model was identified.[35] Moreover, by performing molecular dynamics (MD) simulations at operating conditions (1 bar, 300 K), peak broadening effects due to movement of the layers were effectively included in our calculated diffraction patterns for optimal comparison. The models were generated, starting from a hexagonal layer topology (**hcb**), combinatorically varying all flexible moieties in the layer geometry that might not be able to transition at operating conditions during the MD simulation due to prohibitively large free energy barriers. Moreover, as typically only the nearest neighboring layer orientation is relevant due to screening effects and to allow for different variations between subsequent layers, two-layer unit cells were constructed. The orientations of both the pyridine (py) moieties and the enamine linkages were considered (see Figure S32). The molecular interactions in these systems were modeled using system-specific force fields, derived using the QuickFF protocol (see Section S3.1). Ultimately, this approach gave rise to a single optimal atomic structure for the materials, as visualized in Figure 1a, with excellent agreement to the experimental PXRDs. For **TpBpyCOF**, the optimal structural model has enamine linkages with an identical orientation in the same layer but inverted with respect to the layer above. The bipyridine is in a *trans* configuration and is inverted with respect to the layer above. For **TpPpyCOF**, the optimal structural model has enamine linkages with an identical orientation in the same layer but inverted with respect to the layer above. The pyridine unit is inverted with respect to the layer above and also switches positions. The orientations of asymmetric moieties in COFs, such as the linkage and the configuration of bipyridine/phenylpyridine linkers, are usually not modeled. However, by utilizing this methodology, we were able to obtain a more comprehensive understanding of these materials' configurations. Additionally, the good agreement observed between the modeled and experimental PXRDs provides evidence of the high crystallinity of these materials.

Diffuse reflectance UV−visible spectroscopy was performed to characterize the optical properties of the materials. The diffuse reflectance UV−vis spectra of the COFs were measured and transformed through the Kubelka−Munk function to obtain the absorption spectra. These spectra exhibit absorption edges around 500 nm and the absorption tails extend beyond 850 nm (Figure 5a). It can be clearly seen that the absorption edge for **TpPpyCOF** is blue-shifted compared to that of **TpBpyCOF**. The related Tauc plots (Figure 5b) were used to estimate the band gaps, which amounted to 2.10 and 2.23 eV for **TpBpyCOF** and **TpPpyCOF**, respectively, indicating that the substitution of a nitrogen atom in bipyridine to a carbon in phenylpyridine increased the band gap by 0.13 eV. The band gaps and density of states were calculated using the hybrid functional (HSE06) (Figure S35). The calculated band gaps were 2.09 and 2.14 eV for **TpBpyCOF** and **TpPpyCOF**, respectively, in good agreement with the experimental values. In both cases, the

Scheme 2. One-Pot Oxidation/Povarov Cyclization

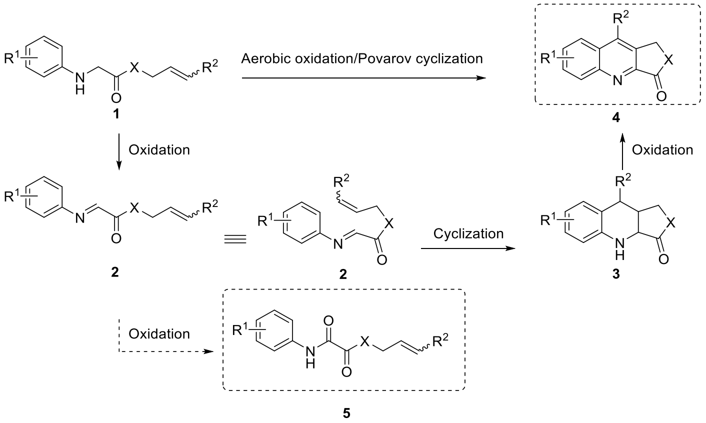
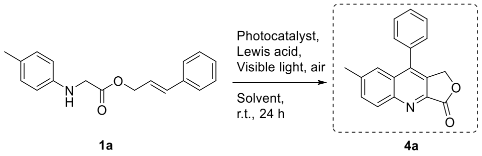


Table 1. Optimization of the Tandem Oxidation/Povarov Cyclization[a]



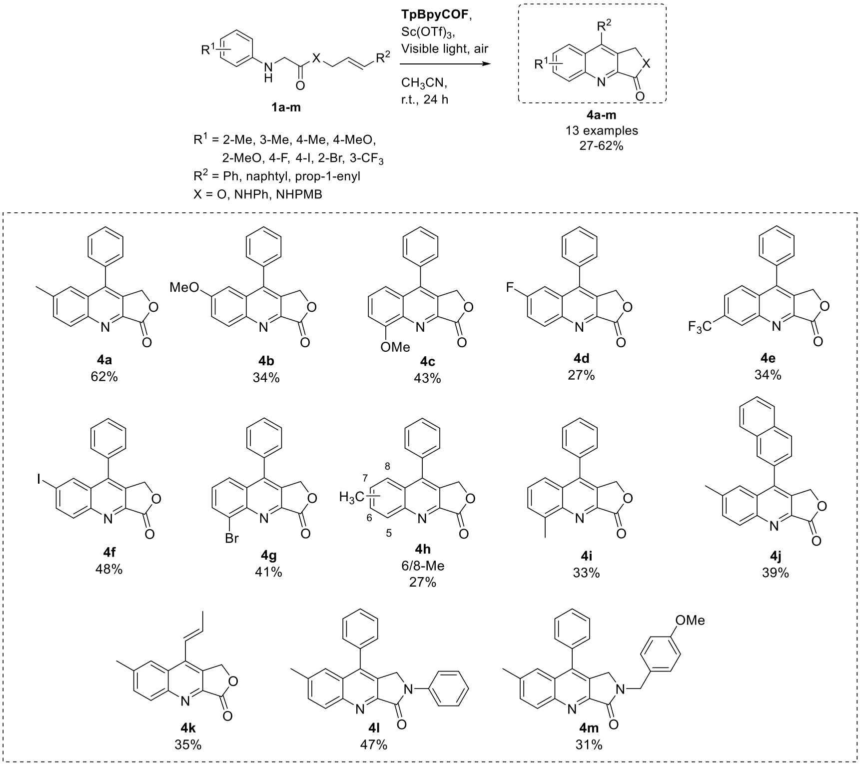
entry	photocatalyst	Lewis acid	solvent	yield (%)[b]
1	TpBpyCOF	$BF_3{\cdot}Et_2O$	CH_3CN	38
2	TpPpyCOF			32
3	$[Ru(Bpy)_3](PF_6)_2$[c]			22
4	$[Ru(Bpy)_3](PF_6)_2$[d]			43
5	TpBpyCOF	no Lewis acid		[e]
6	TpBpyCOF	$BF_3@SiO_2$		48
7		1 equiv $BF_3{\cdot}Et_2O$		26
8		$BF_3{\cdot}Et_2O$/TMSOTf		46
9		TMSOTf		55
10		$Cu(OTf)_2$		19
11		$Bi(OTf)_3$		38
12		TFA		41
13		$Dy(OTf)_3$		52
14		**$Sc(OTf)_3$**		61
15		$HN(OTf)_2$		45
16		$Zn(OTf)_2$		29
17		$AlCl_3$		25
18		$Ho(OTf)_3$		36
19		$Yb(OTf)_3$		49
20		$Er(OTf)_3$		40
21		$La(OTf)_3$		41
22		$Sm(OTf)_3$		49
23		$Eu(OTf)_3$		45
24		$HAuCl_4{\cdot}3H_2O$		33
25		$Gd(OTf)_3$		40
26		$In(OTf)_3$		40
27		$Ce(OTf)_3$		57
28		$CeCl_3{\cdot}7H_2O$		20
29		$TiCl_4{\cdot}2THF$		9
30		$B(PhF_5)_3$		14
31		5 mol % $Sc(OTf)_3$		32
32		30 mol % $Sc(OTf)_3$		39
33		10 mol % $Sc(OTf)_3$	CH_3CN[f]	52
34			$PhCF_3$	27
35			EtOH	5
36			THF	17
37			CH_2Cl_2	22
38			DMF	12
39			1,2-DCE	32
40			HFIP	n.r.[g]
41			CH_3NO_2	60
42		TMSOTf	CH_3NO_2	57
43		$B(PhF_5)_3$	CH_3NO_2	26
44	TpPpyCOF	$Sc(OTf)_3$	CH_3CN	49
45	$[Ru(Bpy)_3](PF_6)_2$[c]			12
46	$[Ru(Bpy)_3](PF_6)_2$[d]			42

[a] Reaction conditions: 0.1 mmol cinnamyl *p*-tolylglycinate **1a**, COF (10 mg), or $[Ru(Bpy)_3](PF_6)_2$ (1 or 10 mol %), [c,d] 10 mol % Lewis acid, 1 mL of CH_3CN, 26 W CFL, 24 h, r.t. [b] Determined by ^1H NMR analysis using 1,3,5-trimethoxybenzene or mesitylene as an internal standard. [c] 1 mol %. [d] 10 mol %. [e] 42% cinnamyl 2-oxo-2-(*p*-tolylamino)acetate **7a** was detected. [f] 5 mL CH_3CN. [g] No reaction.

band gaps were quite small, allowing absorption of visible light and its conversion into useful chemical energy.

Finally, the absolute positions of the valence band (VB) and conducting band (CB) of these COFs were estimated by linear sweep voltammetry under chopped light illumination.[36] To this end, the COFs were mechanically exfoliated by grinding and coated on a working electrode. This was then used in a three-electrode setup, with an Ag/AgCl reference electrode and a Pt coil counter electrode. The anodic photocurrent was measured under chopped white light illumination and an applied bias,

Table 2. TpBpyCOF and $Sc(OTf)_3$-Catalyzed Synthesis of Quinolines 4a−m[a,b]

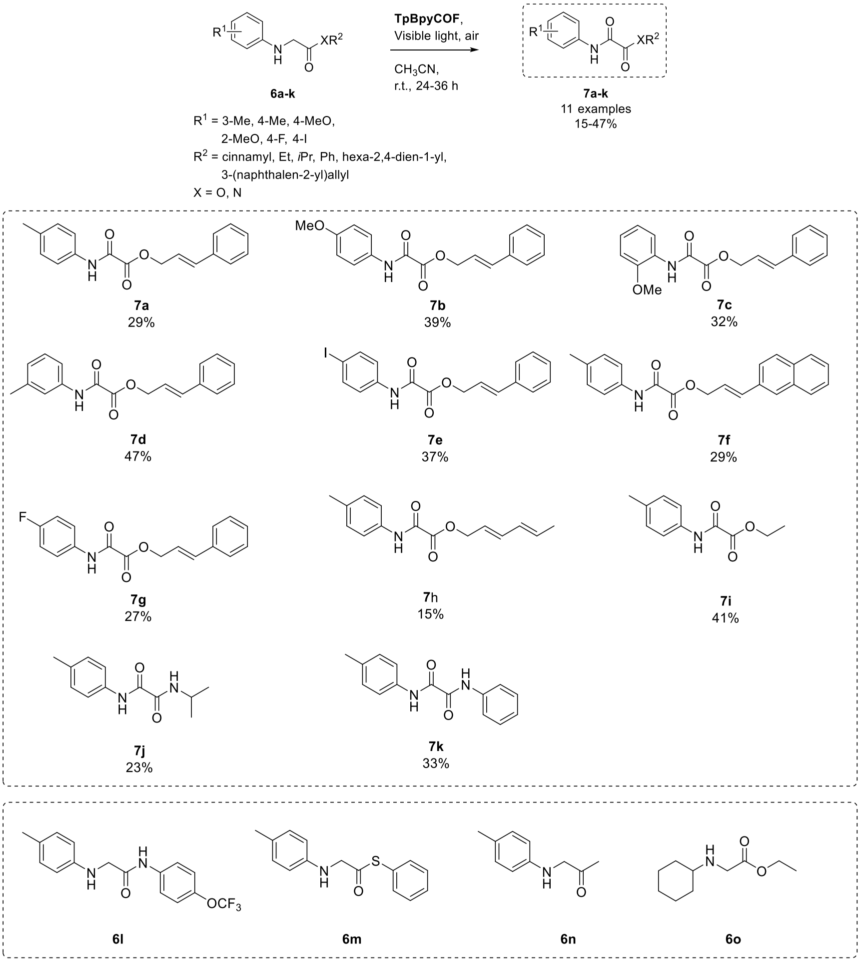


[a]Reaction conditions: 0.2 mmol substrate **1**, 20 mg of **TpBpyCOF**, 0.02 mmol $Sc(OTf)_3$, 2 mL of CH_3CN, 26 W CFL, 24 h, r.t. [b]Isolated yields are reported.

which was swept from +0.4 V to −0.4 *vs* Ag/AgCl. The resulting photocurrent decreased in magnitude and changed sign when going from +0.4 to −0.4 V (Figures S17 and S18). The energy level of the CB can be determined from the potential at which this sign change occurs. The energy levels of the VB can then be calculated using the equation: $E_{CB} = E_{VB} - E_g$ (Figure 5c).[37] To perform the targeted Povarov reaction, the amines of the glycine esters need to be oxidized and oxygen-reduced. The reduction potential of O_2 is −0.62 V (*vs* NHE)[38] and the oxidation potentials for a representative *N*-aryl glycine ester, ethyl *p*-tolylglycinate, is 1.30 V (*vs* NHE).[39] According to these potentials, both COFs can reduce oxygen to the superoxide radical, and both COFs are sufficiently oxidizing to accept electrons from a representative *N*-aryl glycine ester. Next to these direct electron transfer reactions, these COFs could also act as photosensitizers to generate singlet oxygen. Singlet oxygen can be produced by semiconductors, such as COFs, through two main mechanisms. In a Dexter-type energy transfer mechanism, one electron reduction of O_2 to $O_2^{-\bullet}$ takes place, followed by one electron oxidation. Alternatively, Förster resonance energy transfer allows the transfer of the energy of the excited state of the COF to oxygen, through dipole−dipole interactions between oxygen and the electric field produced by the separated charges.[40,41] Singlet oxygen is a powerful oxidizing agent and can oxidize the desired substrates, thereby allowing the proposed reactions to occur. By using EPR spectroscopy and **TpBpyCOF** as the model catalyst, the ability of this material to create activated forms of oxygen ($O_2^{-\bullet}$ and $^1O_2$) was confirmed. By stirring a suspension of the COF in acetonitrile with 5,5-dimethyl-1-pyrroline-*N*-oxide (DMPO), under air and visible light irradiation, the EPR spectrum of DMPO-OOH$^\bullet$, generated by reaction between DMPO and photogenerated superoxide radicals, was detected (Figure 5e).[42] Using a similar methodology, with 2,2,6,6-tetramethylpiperidine (TEMP) as the spin trap, the production of singlet oxygen was demonstrated by the detection of the characteristic spectrum of the TEMPO free radical (Figure 5f). Moreover, both **TpBpyCOF** and **TpPpy-COF** gave EPR signals in the solid state that greatly increased under visible light illumination, signifying the photogeneration of electron−hole pairs (Figure 5d).[43]

Tandem Oxidation/Povarov Cyclization. The prepared photocatalysts were applied toward the tandem aerobic oxidation/Povarov cyclization, which allows the synthesis of quinoline-fused lactones and lactams **4** from cinnamyl 2-(arylamino)-acetates and acetamides **1** (Scheme 2). Quinoline-fused lactones and lactams are an important class of

Table 3. COF-Catalyzed Synthesis of Oxalic Acid Derivatives 6[a,b]



[a]Reaction conditions: 0.2 mmol substrate **6**, 20 mg of **TpBpyCOF**, 2 mL of CH_3CN, 26 W CFL, 24–36 h, r.t. [b]Isolated yields are reported.

heterocycles and also serve as synthons for the synthesis of natural products and analogues such as Uncialamycin,[44] Luotonin A,[45] Camptothecin,[46] and quinolinecarboxamides[47] (Figure S42). The first report on this transformation used difficult-to-synthetize glyoxal imines **2** as starting materials with $BF_3·Et_2O$ as a Lewis acid to catalyze the cyclization and stoichiometric DDQ to oxidize the intermediary tetrahydroquinoline **3**.[44] However, *in situ* generation of the imines from the corresponding, easily accessible, and stable amines **1** proved to be much easier and can be achieved by a stoichiometric oxidant such as Oxone.[48] Tris(4-bromophenyl)ammoniumyl hexachloroantimonate, a radical cation salt, has also been used catalytically for this reaction with air as the terminal oxidant;[49] however, it is a moisture-sensitive irritant.[50] Moreover, tris(bipyridine)-ruthenium/$BF_3·Et_2O$[51] or rhodamine 6G/$HClO_4$[52] have been used as a photocatalyst/acid couple for this transformation. Nevertheless, a metal-free heterogeneous photocatalyst has not yet been reported for this reaction, while it can give many advantages, such as easy separation and catalyst recovery, avoiding toxic and expensive metals.

The initial reaction conditions, using 10 mol % $BF_3·Et_2O$ in acetonitrile at room temperature and exposed to air, were chosen based on literature data[51] (Table 1, entry 1−2). This gave moderate yields of 38 and 32% for **TpBpyCOF** and **TpPpyCOF**, respectively. When no Lewis acid was used, the formation of a side product, cinnamyl 2-oxo-2-(*p*-tolylamino)-acetate **7a**, was detected in 42% yield (Table 1, entry 5). To further optimize the yield and to avoid the moisture- and air-sensitive $BF_3·Et_2O$,[53] many different Lewis and Brønsted acids were screened (Table 1, entry 6−30). First, the easier to handle and heterogeneous BF_3@silica was used, which also gave acceptable yields (Table 1, entry 6). However, $Sc(OTf)_3$ gave the most promising results (Table 1, entry 14). When higher or lower amounts than the standard 10 mol % were used, lower yields were obtained (Table 1, entry 31 and 32). The screening of other solvents than acetonitrile such as $PhCF_3$, EtOH, THF, CH_2Cl_2, DMF, HFIP, or 1,2-DCE (Table 1, entry 34−41) all resulted in lower yields, except for nitromethane, which also gave promising results (Table 1, entry 41). Nitromethane was then also screened with some other Lewis acids (Table 1, entry 42−43); however, this did not improve the yields and finally CH_3CN was chosen because it is much less harmful than CH_3NO_2. Lastly, **TpPpyCOF** was again tested using $Sc(OTf)_3$ as well as $Ru(Bpy)_3(PF_6)_2$ and both gave worse results than **TpBpyCOF** (Table 1, entry 44−46). While **TpBpyCOF** and **TpPpyCOF** possess very similar surface areas and optical properties, **TpBpyCOF** does show higher yields. It is likely that **TpBpyCOF**'s energy levels are more suited for singlet oxygen activation, thereby enhancing the catalytic activity.

With the optimal conditions at hand, the scope of the reaction was examined by reacting a wide range of substrates **1a−m** with both electron-withdrawing (F, CF_3, Br, I) and electron-donating (OMe, Me) substituents in various positions (ortho-, meta-, and para-). The corresponding quinolines **4a−m** were obtained in moderate yields (27−62%, Table 2). While full conversion of the starting materials was achieved in any case, side reactions and losses during isolation resulted in moderate isolated yields. Major side products were generally not observed in ^1H NMR and LC-MS analysis of the crude (Figures S43 and S44). The over-oxidation side product **7** was present in very small quantities (<1%, ^1H NMR, Figure S43) when $Sc(OTf)_3$ was used. Therefore, it is thought that the substrates partly polymerize during the reaction to form insoluble oligomers, which lower the yield. For the synthesis of compound **4a**, the selectivity is then ~62%, as the starting material fully reacted. When using meta-substituted aryls, mixtures of the 6- and 8-substituted regioisomers were observed, of which only isomer **4e** could be isolated in the case of the trifluoromethyl substituted quinoline. Substitution of the R^2-group was also tolerated and both naphthalene- and vinylene-substituted quinolines **4j** and **4k** could be isolated with acceptable yields (39 and 35%). The photochemical synthesis of quinoline-fused lactams had not yet been reported, and to our delight, quinoline fused lactams **4l−m** could also be synthesized and isolated in 47 and 31% yields.

Next to these quinolines, a series of oxalic acid derivatives **7a−k** could be synthesized by photocatalytic α-oxidation of the glycine derivatives **6a−k** when no Lewis acid was present in the system (Table 3). For the α-oxidation of *N*-aryl glycine derivatives, there had only been incidental reports of the occurrence of the α-dicarbonyl compounds as side products during photocatalytic reactions.[51] To the best of our knowledge, **TpBpyCOF** is the first photocatalyst used (homogeneous or heterogeneous) to study this reaction in more detail and examine the substrate scope. A wide range of oxalic acid derivatives **7a−k** could be produced using **TpBpyCOF**, albeit in relatively low isolated yields (15−47%). These rather low yields were due to side reactions and losses by purification using reversed phase chromatography. For the trifluoromethoxy-substituted substrate **6l**, no product could be isolated due to the low stability of the starting material in solution. For thioester **6m**, the oxidized end product was obtained; however, it could not be fully separated from impurities. The product resulting from oxidation of ketone **6n** was obtained in only a very low yield (6%). For compound **6o**, no product was obtained, and degradation of the COF was observed as the nucleophilic amine attacks the β-ketoenamine linkage and solubilizes the COF.

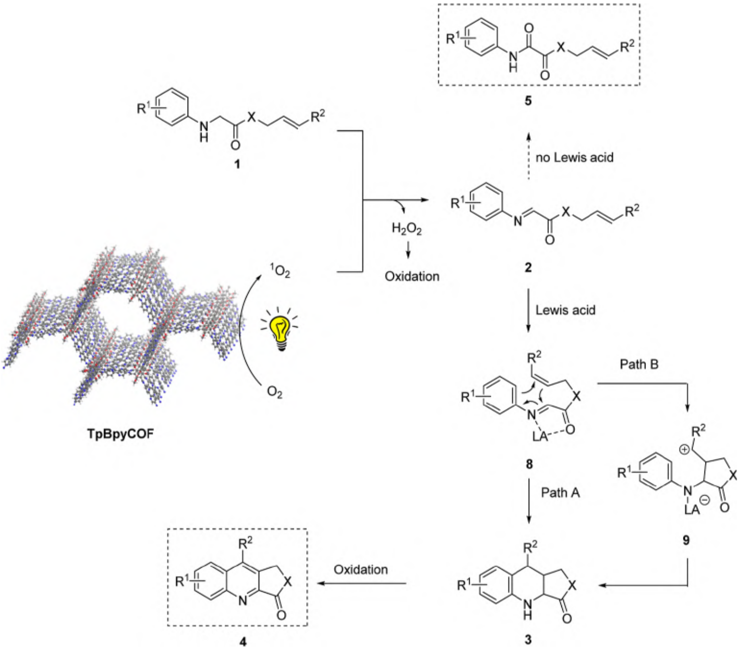
To probe the mechanism of the oxidation/Povarov cyclization, several control experiments were performed. In the dark, under argon or without COF, no product was formed (Table 4, entry 2−4). In the absence of a Lewis acid, the oxalic acid derivative was formed (Table 4, entry 5). Thus oxygen, light, the COF, and the Lewis acid were essential for the transformation. As proven by the EPR experiments (Figure 5e,f), **TpBpyCOF** is capable of photogenerating both superoxide radicals and singlet oxygen, and it was feasible for both these pathways to generate the products. To study the role of oxygen, several quenchers were added (Table 4, entry 6−14). Using quenchers for the superoxide radical, *p*-benzoquinone[54] or nitro blue tetrazolium chloride,[55] only a slight decrease in yield was observed. Similarly, *i*PrOH, a quencher for hydroxyl radicals,[56] did not lower the yield significantly. However, when using quenchers for singlet oxygen, such as L-histidine,[57] DABCO,[58] or sodium azide,[59] the reaction was inhibited effectively. This leads us to conclude that the key driver for this transformation is singlet oxygen. When quenching holes and electrons using KI[60] and $AgNO_3$,[61] respectively, the yield was also lowered. In the presence of TEMPO,[62] a known radical scavenger, the yield lowered drastically (Table 4, entry 14), suggesting the involvement of radical processes. Based on these control experiments and previous reports,[48,51] we propose the following mechanism (Scheme 3). Upon irradiation of the COF, singlet oxygen is produced by direct energy transfer[63] or by sequential oxidation/reduction of molecular oxygen.[64] The photogenerated singlet oxygen then oxidizes the amine **1** to the corresponding imine **2**, thereby generating H_2O_2.[65] The crude NMR spectra (Figure S43) clearly indicate the production of hydrogen peroxide, which is present in an approximately one-to-one ratio with the product **4**. In absence of a Lewis acid, imine **2** is further oxidized to the oxalic derivative **5**. In the presence of a Lewis acid, imine **2** undergoes an intramolecular aza-Diels−Alder reaction, which can occur either concerted (path A) or stepwise (path B).[66] The resulting tetrahydroquinoline **3** is then further oxidized to the quinoline product **4**. As bipyridine is a

Table 4. Control Experiments for the Povarov Reaction

entry	conditions[a]	result (%)[b]
1	no modification	62
2	in the dark	n.r.[c]
3	under argon	n.r.[c]
4	no COF	n.r.[c]
5	no Lewis acid	42% **7a**
6	*p*-benzoquinone	52
7	nitro blue tetrazolium chloride[d]	53
8	*i*PrOH	51
9	L-histidine	6
10	DABCO	n.r.[c]
11	NaN_3	n.r.[c]
12	$AgNO_3$	27
13	KI	n.r.[c]
14	TEMPO	10

[a] Reaction conditions: 0.1 mmol cinnamyl *p*-tolylglycinate **1a**, 10 mg of **TpBpyCOF**, 0.01 mmol $Sc(OTf)_3$, 0.1 mmol of quencher. [b] Determined by ^1H NMR analysis using 1,3,5-trimethoxybenzene as an internal standard. [c] No reaction. [d] 0.015 mmol NBT.

Scheme 3. Mechanism for the Photocatalytic Generation of Quinolines 4 and Glyoxal Derivatives 5 by TpBpyCOF



Lewis base and has the ability to coordinate Lewis acids, such as $Sc(OTf)_3$, we stirred the COF in acetonitrile for 24 h, upon which it was filtered, washed, and subjected to a catalytic experiment. This resulted in a similar result (see Table S2) to that when fresh $Sc(OTf)_3$ was used, indicating that the bipyridines of the COF heterogenize the Lewis acid and result in a catalytically active material. Lastly, reuse experiments were performed to ascertain the stability of **TpBpyCOF** in the reaction. Upon recycling the catalyst for five times, full conversion of the starting material was still achieved; however, the yield had dropped to 45%, which is still 75% of the original activity (Figure S45). Moreover, while the main chemical linkages remained intact according to FTIR, the PXRD of the recycled material seemed to indicate that it had lost its crystallinity (Figures S46 and S47). The loss of activity and crystallinity, while unfortunate, has been observed many times for photoactive COFs and can be attributed to photobleaching, partial deactivation by singlet oxygen, and also possibly side reactions with intermediates formed during the reaction.[23,67,68]

■ CONCLUSIONS

In this work, two bipyridine- or phenylpyridine-based COFs, **TpBpyCOF** and **TpPpyCOF**, were successfully synthesized and photochemically characterized. The use of bipyridine led to a smaller band gap compared with phenylpyridine, and the resulting **TpBpyCOF** was a more effective photocatalyst. This work entailed the development of the first heterogeneous photocatalyst for the aerobic oxidation/Povarov cyclization and α-oxidation of *N*-aryl glycine derivatives, which are important transformations to generate highly substituted quinolines and dicarbonyl compounds. Using **TpBpyCOF** as a recyclable photocatalyst, a wide range of quinoline-fused lactones/lactams and glyoxal derivatives could be synthesized. The findings of our study expand the potential use of COFs in photocatalysis, thereby advancing the development of these novel materials as alternatives for homogeneous (metal-based) photocatalysts.

■ EXPERIMENTAL SECTION

General Procedures. Unless stated otherwise, all reagents and solvents were purchased from commercial sources and used without further purification. Dry tetrahydrofuran and dichloromethane were obtained using the MBraun SPS-800 solvent purification system. Dry *N*,*N*-dimethylformamide was obtained by drying over 4 Å molecular sieves and storing for at least 2 days prior to use. The synthesis of the building blocks for the COFs and the substrates for the aerobic oxidation/Povarov cyclization and α-oxidation of *N*-aryl glycine derivatives is described in the Supporting Information.

IR spectra of COFs and compounds were obtained in neat form with a Shimadzu IRAffinity1S WL FTIR spectrophotometer. Nitrogen adsorption−desorption isotherms were obtained using a Micromeritics Tristar II. The samples were activated at 120 °C under vacuum overnight before measurements. Pore size distributions were calculated from N_2 sorption isotherms using quenched solid density functional theory in ASiQwin. Powder X-ray diffraction (PXRD) spectra were taken using a Bruker D8 Advance spectrometer with a copper Kα radiation source (λ = 1.54056 Å) at 40 kV and 45 mA with 1°/s scanning speed. UV−vis diffuse reflectance spectra were recorded in the solid state on a Perkin Elmer Lambda 1050 UV−vis−NIR spectrophotometer. Photoelectrochemical measurements were performed in 0.2 M KCl/NaOH buffer (pH 12) or 0.06 M $Na_2B_4O_7 \cdot 10H_2O$/NaOH buffer (pH 10) using an ALS-Japan Ag/AgCl 3 M NaCl reference electrode (+0.195 V *vs* SHE) and ALS-Japan Pt coil as the counter electrode, and the photocatalyst-coated FTO was used as the

https://doi.org/10.1021/acsami.3c07036
ACS Appl. Mater. Interfaces 2023, 15, 35092−35106

working electrode. A Bio-Logic VSP potentiostat and EC-Lab software were used to record the measurements. White light was provided by a Philips Tornado T2 CFL (23 W, 1450 lumen). XPS was used to investigate the chemical composition of the surface of the materials using the PHI 5000 VersaProbe II spectrometer. This was equipped with a monochromatic Al Kα X-ray source (hν = 1486.6 eV) operating with a beam diameter of 200 μm at 50 W. All measurements were taken using an angle of 45° between the beam and the sample and under a pressure of 10^{-6} Pa or less. The binding energies were calibrated with respect to the C−C/C−H peak of the C 1s spectrum at 285.0 eV. Survey scans and high-resolution spectra were analyzed using Multipak (v 9.6.1) software. The high-resolution spectra were deconvoluted as Gaussian−Lorentzian peaks to identify the corresponding chemical bonds. BF-TEM images were taken using a high contrast aperture in BF mode on a JEOL JEM-2200FS transmission electron microscope with a 200 kV field emission gun and the in-column energy filter (omega filter) at the UGent TEM Core Facility. The EPR spectra were recorded using a Bruker ESP300E spectrometer equipped with an HP 5350B microwave frequency counter and a Bruker NMR ER035M gaussmeter. A microwave power of 20 mW was used. The field was modulated at a frequency of 100 kHz with an amplitude of 0.1−0.2 mT. The time constant was 2.56 ms and spectra were taken by an accumulation of 20 scans of 10 s each. The magnetic fields were calibrated using the spectrum of diphenyl picryl hydrazyl (g = 2.0036). For the solid-state EPR spectra, a regular NMR tube was charged with 5 mg of the material. For spin trapping, **TpBpyCOF** (3.75 mg) was stirred in 0.75 mL of a 200 mM spin trap solution (DMPO or TEMP) in CH_3CN and illuminated for 10 min. The blanks were acquired by stirring the spin trap solution without the COF for 10 min under visible light illumination. An aliquot of these suspensions was then used to charge 2 mm outer diameter and 1.4 mm inner diameter sample tubes, overfilling the full length of the rectangular EPR cavity, which was then directly measured in the EPR spectrometer in the case of spin trapping with DMPO. As commercial TEMP, even after distillation, contained significant amounts of TEMPO, and TEMPO radicals are stable, the samples were illuminated for a longer time (2 h) and then the spectra were taken. Direct-excitation ^{1}H, ^{1}H-^{1}H double quantum-single quantum (DQ-SQ), ^{1}H-^{13}C CPMAS, and ^{1}H-^{13}C CP-HETCOR spectra were acquired on a Bruker Avance Neo 800 MHz (18.8 T) spectrometer equipped with a 1.9 mm H/X/Y triple resonance probehead. The sample was filled in a 1.9 mm ZrO_2 rotor and subjected to 35 kHz magic angle spinning (MAS). Prior to the insertion into the probe, the sample filled rotor was subjected to vacuum treatment (75 mTorr) at 353 K. Direct-excitation ^{1}H (801.25 MHz) MAS NMR experiments were performed with an excitation pulse of 140 kHz radio-frequency (RF) field strength, accumulating 32 transients with a recycle delay of 3 s. 2D ^{1}H-^{1}H DQ-SQ correlation measurements were carried out using an excitation pulse of 140 kHz RF field strength, measured using the BABA sequence.[69] The two-dimensional spectra were collected with 80 t_e increments of 28.6 μs and 16 transients in the direct dimension. ^{1}H-^{13}C CPMAS experiments were performed at a MAS rate of 35 kHz using 12,288 scans and a contact time of 1000 μs, a ^{1}H contact pulse of 120 kHz RF field strength (100−70% ramped pulse), a square pulse of 50 kHz RF field strength on ^{13}C, and 60 kHz spinal64 ^{1}H decoupling acquisition. ^{1}H-^{13}C CP-HETCOR spectra were acquired with 400 scans in the direct dimension, with 32 t_e increments of 28.57 μs. During acquisition, SPINAL64[70] ^{1}H decoupling was carried out using a pulse of 60 kHz RF field strength. All the spectra were referenced to the adamantane ^{1}H resonance at 1.81 ppm and the ^{13}C (most downfield) resonance at 38.5 ppm, respectively. ^{1}H NMR and ^{13}C NMR spectra of solutions were recorded with a Bruker Avance III HD-400 spectrophotometer at 25 °C at 400 and 100 MHz, respectively. The NMR was equipped with a 1H/BB z-gradient probe (BBO, 5 mm). All spectra were acquired through standard sequences available in the Bruker pulse program library and processed using TOPSPIN 4.1. An Agilent 1200 series HPLC system fitted with an Ascentis Express C18 column (particle size 2.7 μm, length 30 mm, internal diameter 4.6 mm) was used for HPLC(-MS) using a mixture of acetonitrile/water (5 mM NH_4OAc) as the eluent. The HPLC instrument was connected with a UV−vis detector and an Agilent 1100 series LC/MSD-type SL mass spectrometer (ESI, 4000 V) using a mass-selective single quadrupole detector. Thin-layer chromatography (TLC) for the analysis of reaction mixtures or gradient determination for chromatography was performed using glass-backed 0.25 mm Merck silica gel 60 F254 TLC plates and visualized under UV light (254 nm). Column chromatography was performed with glass columns using silica gel (particle size 35−70 μm, pore diameter 6 nm) or on a Büchi Reveleris X2 flash chromatography system (normal phase) or Grace Reveleris X1 flash chromatography system (reversed phase) using prepacked Reveleris silica or Reveleris C18 cartridges.

Synthesis of COFs. The procedure to synthesize **TpBpyCOF** was modified from a literature procedure.[25] An Agilent GC vial (size: 22.75 mm × 75 mm; 20 mm cap) was charged with 1,3,5-triformylphloroglucinol **Tp** (63.0 mg, 0.3 mmol, 1 equiv) and 2,2′-bipyridine-5,5′-diamine **Bpy** (83.7 mg, 0.45 mmol, 1.5 equiv). Dimethylacetamide (DMAc, 4.5 mL) and *o*-dichlorobenzene (*o*DCB, 1.5 mL) were added *via* the sides of the vial to flush the remaining solids from the walls. Then, 0.6 mL of 6.0 M aqueous acetic acid was added and the vial was capped. This mixture was then sonicated for 10 min, flash frozen at 77 K in liquid N_2, and degassed by three freeze−pump−thaw cycles, after which the vial was put under argon. After warming to room temperature, the vial was placed in an oven pre-heated at 120 °C. The resulting red powder was collected *via* filtration and washed sequentially with copious DMAc-DMF-H_2O-acetone-ethanol-THF. Further purification was done by Soxhlet extraction with methanol for 72 h. Finally, the material was dried under vacuum overnight yielding **TpBpyCOF** as a red powder (113 mg, 86%).

For **TpPpyCOF**, the same procedure with **Ppy** (83 mg, 0.45 mmol, 1.5 equiv) and **Tp** (63.0 mg, 0.3 mmol, 1 equiv) was used, resulting in the title COF as an orange powder (103 mg, 79%).

General Procedure for the Aerobic Oxidation/Povarov Cyclization of *N*-Aryl Glycine Derivatives. Typically, 20 mg of **TpBpyCOF**, $ScOTf_3$ (10 mg, 0.02 mmol, 0.1 equiv), 0.2 mmol of substrate **1**, and 2 mL of CH_3CN were added to a small glass test tube. This was stirred under air and irradiated with a 26 W compact fluorescent lamp (CFL; ∼10 cm distance) for 24 h. The reaction mixture was filtered over a filter paper and washed with acetone. The filtrate was then evaporated and the product was purified using column chromatography (SiO_2, hexane/acetone: 10/1 or C18, gradient CH_3CN/H_2O: 30/70−100/0).[a]

General Procedure for the α-Oxidation of *N*-Aryl Glycine Derivatives. Typically, 20 mg of **TpBpyCOF**, 0.2 mmol of substrate **6**, and 2 mL CH_3CN were added to a small glass test tube. This was stirred under air and irradiated with a 26 W CFL (∼10 cm distance) for 24−36 h. The reaction mixture was filtered through a filter paper and rinsed thoroughly with acetone. The filtrate was then evaporated and the crude product was purified using column chromatography (C18, gradient CH_3CN/H_2O: 40/60−100/0 or 50/50−100/0).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsami.3c07036.

Full experimental procedures for the linker and substrate synthesis, characterization details (pore size distributions, FTIR spectra, BF-TEM, XPS, NMR, and cyclic voltammetry), methodology for computational modeling, and ^{1}H- and ^{13}C NMR spectra of all synthesized compounds (DOCX)

The simulated PXRD patterns of all possible COF structures and CIF files for the final, optimized structures (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Christian V. Stevens − *Department of Green Chemistry and Technology, Ghent University, Ghent 9000, Belgium;*

orcid.org/0000-0003-4393-5327; Email: Chris.Stevens@Ugent.be

Authors

Maarten Debruyne − *Department of Green Chemistry and Technology, Ghent University, Ghent 9000, Belgium;* orcid.org/0000-0002-5300-8773

Sander Borgmans − *Department of Applied Physics, Ghent University, Zwijnaarde 9052, Belgium*

Sambhu Radhakrishnan − *NMR/X-ray Platform for Convergence Research (NMRCoRe) & Centre for Surface Chemistry and Catalysis: Characterisation and Application Team (COK-KAT), KU Leuven, Leuven 3001, Belgium;* orcid.org/0000-0002-0274-2759

Eric Breynaert − *NMR/X-ray Platform for Convergence Research (NMRCoRe) & Centre for Surface Chemistry and Catalysis: Characterisation and Application Team (COK-KAT), KU Leuven, Leuven 3001, Belgium;* orcid.org/0000-0003-3499-0455

Henk Vrielinck − *Department of Solid State Sciences, Ghent University, Ghent 9000, Belgium;* orcid.org/0000-0003-4861-9630

Karen Leus − *Department of Applied Physics, Faculty of Engineering and Architecture, Ghent University, Ghent 9000, Belgium*

Andreas Laemont − *Department of Chemistry, Ghent University, Ghent 9000, Belgium*

Juul De Vos − *Department of Applied Physics, Ghent University, Zwijnaarde 9052, Belgium;* orcid.org/0000-0002-2733-5762

Kuber Singh Rawat − *Department of Applied Physics, Ghent University, Zwijnaarde 9052, Belgium;* orcid.org/0000-0002-7308-4204

Siebe Vanlommel − *Department of Applied Physics, Ghent University, Zwijnaarde 9052, Belgium*

Hannes Rijckaert − *Department of Chemistry, Ghent University, Ghent 9000, Belgium;* orcid.org/0000-0002-6078-2919

Hadi Salemi − *Department of Green Chemistry and Technology, Ghent University, Ghent 9000, Belgium*

Jonas Everaert − *Department of Green Chemistry and Technology, Ghent University, Ghent 9000, Belgium;* orcid.org/0000-0002-7750-1992

Flore Vanden Bussche − *Department of Green Chemistry and Technology, Ghent University, Ghent 9000, Belgium; Department of Chemistry, Ghent University, Ghent 9000, Belgium*

Dirk Poelman − *Department of Solid State Sciences, Ghent University, Ghent 9000, Belgium;* orcid.org/0000-0002-3930-172X

Rino Morent − *Department of Applied Physics, Faculty of Engineering and Architecture, Ghent University, Ghent 9000, Belgium*

Nathalie De Geyter − *Department of Applied Physics, Faculty of Engineering and Architecture, Ghent University, Ghent 9000, Belgium*

Pascal Van Der Voort − *Department of Chemistry, Ghent University, Ghent 9000, Belgium;* orcid.org/0000-0002-1248-479X

Veronique Van Speybroeck − *Department of Applied Physics, Ghent University, Zwijnaarde 9052, Belgium;* orcid.org/0000-0003-2206-178X

Complete contact information is available at: https://pubs.acs.org/10.1021/acsami.3c07036

Author Contributions

Conceptualization was performed by M.D.B. Investigation was taken up by M.D.B. Funding acquisition was done by V.V.S., P.V.D.V., and C.S. Supervision was executed by V.V.S., P.V.D.V., and C.S. Writing—original draft was accomplished by M.D.B. Writing—review and editing were accomplished by M.D.B., S.B., S.R., H.V., D.P., and C.S. H.R. performed the TEM analyses. K.L. performed the XPS and SEM analyses under the supervision of R.M. and N.D.G. J.E., H.S., and F.V.B. helped with characterization of the COFs. A.L. performed the CV measurements. UV−vis measurements were performed under the guidance of D.P. EPR measurements were conducted with the help from H.V. All computational results were performed by S.B., J.D.V., K.S.R., and S.V. S.R. and E.B. performed the solid-state NMR analysis.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work is supported by the Research Board of Ghent University (BOF) through a Concerted Research Action (GOA20177000303) for funding. The computational resources were provided by the VSC (Flemish Super-computer Center) and Ghent University (Stevin Supercomputer Infrastructure) and funded by the Research Foundation-Flanders (FWO). NMRCoRe (E.B. and S.R.) acknowledge the European Research Council (ERC) for an Advanced Research Grant under the European Union's Horizon 2020 research and innovation program under grant agreement no. 834134 (WATUSO). NMRCoRe acknowledges the Flemish government, Department EWI for Infrastructure Investment *via* the Hermes Fund (AH.2016.134) and for financial support as International Research Infrastructure (I001321N: Nuclear Magnetic Resonance Spectroscopy Platform for Molecular Water Research).

ADDITIONAL NOTE

[a]Reversed phase for 4c,e,h,k.

REFERENCES

(1) Wang, G.-B.; Li, S.; Yan, C.-X.; Zhu, F.-C.; Lin, Q.-Q.; Xie, K.-H.; Geng, Y.; Dong, Y.-B. Covalent Organic Frameworks: Emerged High-Performance Platforms for Efficient Photocatalytic Applications. *J. Mater. Chem. A* **2020**, *8*, 6957−6983.

(2) Romero, N. A.; Nicewicz, D. A. Organic Photoredox Catalysis. *Chem. Rev.* **2016**, *116*, 10075−10166.

(3) Schultz, D. M.; Yoon, T. P. Solar synthesis: Prospects in Visible Light Photocatalysis. *Science* **2014**, *343*, 1239176.

(4) Koike, T.; Akita, M. Visible-Light Radical Reaction Designed by Ru- and Ir-Based Photoredox Catalysis. *Inorg. Chem. Front.* **2014**, *1*, 562−576.

(5) Ni, M.; Leung, M. K. H.; Leung, D. Y. C.; Sumathy, K. A Review and Recent Developments in Photocatalytic Water-Splitting Using TiO_2 for Hydrogen Production. *Renew. Sust. Energ8. Rev.* **2007**, *11*, 401−425.

(6) Andrew Frame, F.; Carroll, E. C.; Larsen, D. S.; Sarahan, M.; Browning, N. D.; Osterloh, F. E. First Demonstration of CdSe as a Photocatalyst for Hydrogen Evolution from Water under UV and Visible Light. *Chem. Commun.* **2008**, *19*, 2206−2208.

(7) Tahir, M. B.; Nabi, G.; Rafique, M.; Khalid, N. R. Nanostructured-Based WO_3 Photocatalysts: Recent Development, Activity Enhance-

ment, Perspectives and Applications for Wastewater Treatment. *Int. J. Environ. Sci. Technol.* **2017**, *14*, 2519−2542.

(8) Cao, S.; Low, J.; Yu, J.; Jaroniec, M. Polymeric Photocatalysts Based on Graphitic Carbon Nitride. *Adv. Mater.* **2015**, *27*, 2150−2176.

(9) Dhakshinamoorthy, A.; Asiri, A. M.; García, H. Metal-Organic Framework (MOF) Compounds: Photocatalysts for Redox Reactions and Solar Fuel Production. *Angew. Chem., Int. Ed.* **2016**, *55*, 5414−5445.

(10) Zhang, T.; Xing, G.; Chen, W.; Chen, L. Porous Organic Polymers: a Promising Platform for Efficient Photocatalysis. *Mater. Chem. Front.* **2020**, *4*, 332−353.

(11) Doonan, C. J.; Tranchemontagne, D. J.; Glover, T. G.; Hunt, J. R.; Yaghi, O. M. Exceptional Ammonia Uptake by a Covalent Organic Framework. *Nat. Chem.* **2010**, *2*, 235−238.

(12) Halder, A.; Ghosh, M.; Khayum M, A.; Bera, S.; Addicoat, M.; Sasmal, H. S.; Karak, S.; Kurungot, S.; Banerjee, R. Interlayer Hydrogen-Bonded Covalent Organic Frameworks as High-Performance Supercapacitors. *J. Am. Chem. Soc.* **2018**, *140*, 10941−10945.

(13) Das, G.; Biswal, B. P.; Kandambeth, S.; Venkatesh, V.; Kaur, G.; Addicoat, M.; Heine, T.; Verma, S.; Banerjee, R. Chemical Sensing in two Dimensional Porous Covalent Organic Nanosheets. *Chem. Sci.* **2015**, *6*, 3931−3939.

(14) Wu, Y.; Xu, H.; Chen, X.; Gao, J.; Jiang, D. A π-Electronic Covalent Organic Framework Catalyst: π-Walls as Catalytic Beds for Diels-Alder Reactions under Ambient Conditions. *Chem. Commun.* **2015**, *51*, 10096−10098.

(15) Liu, W.; Su, Q.; Ju, P.; Guo, B.; Zhou, H.; Li, G.; Wu, Q. A Hydrazone-Based Covalent Organic Framework as an Efficient and Reusable Photocatalyst for the Cross-Dehydrogenative Coupling Reaction of N-Aryltetrahydroisoquinolines. *ChemSusChem* **2017**, *10*, 664−669.

(16) Wei, P. F.; Qi, M. Z.; Wang, Z. P.; Ding, S. Y.; Yu, W.; Liu, Q.; Wang, L. K.; Wang, H. Z.; An, W. K.; Wang, W. Benzoxazole-Linked Ultrastable Covalent Organic Frameworks for Photocatalysis. *J. Am. Chem. Soc.* **2018**, *140*, 4623−4631.

(17) Liu, H.; Li, C.; Li, H.; Ren, Y.; Chen, J.; Tang, J.; Yang, Q. Structural Engineering of Two-Dimensional Covalent Organic Frameworks for Visible-Light-Driven Organic Transformations. *ACS Appl. Mater. Interfaces* **2020**, *12*, 20354−20365.

(18) He, S.; Yin, B.; Niu, H.; Cai, Y. Targeted Synthesis of Visible-Light-Driven Covalent Organic Framework Photocatalyst via Molecular Design and Precise Construction. *Appl. Catal. B* **2018**, *239*, 147−153.

(19) Sheng, J. L.; Dong, H.; Meng, X. b.; Tang, H. L.; Yao, Y. H.; Liu, D. Q.; Bai, L. L.; Zhang, F. M.; Wei, J. Z.; Sun, X. J. Effect of Different Functional Groups on Photocatalytic Hydrogen Evolution in Covalent-Organic Frameworks. *ChemCatChem* **2019**, *11*, 2313−2319.

(20) Wang, Y.; Liu, H.; Pan, Q.; Wu, C.; Hao, W.; Xu, J.; Chen, R.; Liu, J.; Li, Z.; Zhao, Y. Construction of Fully Conjugated Covalent Organic Frameworks via Facile Linkage Conversion for Efficient Photoenzymatic Catalysis. *J. Am. Chem. Soc.* **2020**, *142*, 5958−5963.

(21) Shi, J. L.; Chen, R.; Hao, H.; Wang, C.; Lang, X. 2D sp2 Carbon-Conjugated Porphyrin Covalent Organic Framework for Cooperative Photocatalysis with TEMPO. *Angew. Chem., Int. Ed.* **2020**, *59*, 9088−9093.

(22) Sun, N.; Wang, C.; Wang, H.; Gao, X.; Jiang, J. Photonic Switching Porous Organic Polymers toward Reversible Control of Heterogeneous Photocatalysis. *ACS Appl. Mater. Interfaces* **2020**, *12*, 56491−56498.

(23) Zhang, K.; Kopetzki, D.; Seeberger, P. H.; Antonietti, M.; Vilela, F. Surface Area Control and Photocatalytic Activity of Conjugated Microporous Poly(benzothiadiazole) Networks. *Angew. Chem., Int. Ed.* **2013**, *52*, 1432−1436.

(24) Li, R.; Ma, B. C.; Huang, W.; Wang, L.; Wang, D.; Lu, H.; Landfester, K.; Zhang, K. A. I. Photocatalytic Regioselective and Stereoselective [2 + 2] Cycloaddition of Styrene Derivatives Using a Heterogeneous Organic Photocatalyst. *ACS Catal.* **2017**, *7*, 3097−3101.

(25) Shinde, D. B.; Aiyappa, H. B.; Bhadra, M.; Biswal, B. P.; Wadge, P.; Kandambeth, S.; Garai, B.; Kundu, T.; Kurungot, S.; Banerjee, R. A Mechanochemically Synthesized Covalent Organic Framework as a Proton-Conducting Solid Electrolyte. *J. Mater. Chem. A* **2016**, *4*, 2682−2690.

(26) Bhadra, M.; Sasmal, H. S.; Basu, A.; Midya, S. P.; Kandambeth, S.; Pachfule, P.; Balaraman, E.; Banerjee, R. Predesigned Metal-Anchored Building Block for in Situ Generation of Pd Nanoparticles in Porous Covalent Organic Framework: Application in Heterogeneous Tandem Catalysis. *ACS Appl. Mater. Interfaces* **2017**, *9*, 13785−13792.

(27) Aiyappa, H. B.; Thote, J.; Shinde, D. B.; Banerjee, R.; Kurungot, S. Cobalt-Modified Covalent Organic Framework as a Robust Water Oxidation Electrocatalyst. *Chem. Mater.* **2016**, *28*, 4375−4379.

(28) Cui, W.-R.; Zhang, C.-R.; Jiang, W.; Liang, R.-P.; Qiu, J.-D. Covalent Organic Framework Nanosheets for Fluorescence Sensing via Metal Coordination. *ACS Appl. Nano Mater.* **2019**, *2*, 5342−5349.

(29) Kou, M.; Wang, Y.; Xu, Y.; Ye, L.; Huang, Y.; Jia, B.; Li, H.; Ren, J.; Deng, Y.; Chen, J.; Zhou, Y.; Lei, K.; Wang, L.; Liu, W.; Huang, H.; Ma, T. Molecularly Engineered Covalent Organic Frameworks for Hydrogen Peroxide Photosynthesis. *Angew. Chem., Int. Ed.* **2022**, *61*, No. e202200413.

(30) Xia, A.; Guo, H.; Qiu, X.; Ding, M.; Gao, L. Syntheses and Properties of Polyimides Derived from Diamines Containing 2,5-Disubstituted Pyridine Group. *J. Appl. Polym. Sci.* **2006**, *102*, 1844−1851.

(31) Halder, A.; Karak, S.; Addicoat, M.; Bera, S.; Chakraborty, A.; Kunjattu, S. H.; Pachfule, P.; Heine, T.; Banerjee, R. Ultrastable Imine-Based Covalent Organic Frameworks for Sulfuric Acid Recovery: An Effect of Interlayer Hydrogen Bonding. *Angew. Chem., Int. Ed.* **2018**, *57*, 5797−5802.

(32) Mahmood, J.; Ahmad, I.; Jung, M.; Seo, J. M.; Yu, S. Y.; Noh, H. J.; Kim, Y. H.; Shin, H. J.; Baek, J. B. Two-Dimensional Amine and Hydroxy Functionalized Fused Aromatic Covalent Organic Framework. *Commun. Chem.* **2020**, *3*, 31.

(33) Guo, J.; Xu, Y.; Jin, S.; Chen, L.; Kaji, T.; Honsho, Y.; Addicoat, M. A.; Kim, J.; Saeki, A.; Ihee, H.; Seki, S.; Irle, S.; Hiramoto, M.; Gao, J.; Jiang, D. Conjugated Organic Framework with Three-Dimensionally Ordered Stable Structure and Delocalized π Clouds. *Nat. Commun.* **2013**, *4*, 2736.

(34) Vallaey, B.; Radhakrishnan, S.; Heylen, S.; Chandran, C. V.; Taulelle, F.; Breynaert, E.; Martens, J. A. Reversible Room Temperature Ammonia Gas Absorption in Pore Water of Microporous Silica-Alumina for Sensing Applications. *Phys. Chem. Chem. Phys.* **2018**, *20*, 13528−13536.

(35) Borgmans, S.; Rogge, S. M. J.; de Vos, J. S.; Stevens, C. V.; Van Der Voort, P.; Van Speybroeck, V. Quantifying the Likelihood of Structural Models through a Dynamically Enhanced Powder X-Ray Diffraction Protocol. *Angew. Chem., Int. Ed.* **2021**, *60*, 8913−8922.

(36) Stegbauer, L.; Zech, S.; Savasci, G.; Banerjee, T.; Podjaski, F.; Schwinghammer, K.; Ochsenfeld, C.; Lotsch, B. V. Tailor-Made Photoconductive Pyrene-Based Covalent Organic Frameworks for Visible-Light Driven Hydrogen Generation. *Adv. Ener. Mater.* **2018**, *8*, 1703278.

(37) Cai, K.; Wang, W.; Zhang, J.; Chen, L.; Wang, L.; Zhu, X.; Yu, Z.; Wu, Z.; Zhou, H. Facile Construction of Olefin-Linked Covalent Organic Frameworks for Enhanced Photocatalytic Organic Transformation via Wall Surface engineering. *J. Mater. Chem. A* **2022**, *10*, 7165−7172.

(38) Liu, Z.; Su, Q.; Ju, P.; Li, X.; Li, G.; Wu, Q.; Yang, B. A hydrophilic Covalent Organic Framework for Photocatalytic Oxidation of Benzylamine in Water. *Chem. Commun.* **2020**, *56*, 766−769.

(39) Kabatc, J.; Kucybała, Z.; Pietrzak, M.; Ścigalski, F.; Pączkowski, J. Free Radical Polymerization Initiated via Photoinduced Intermolecular Electron Transfer Process: Kinetic Study 3. *Polymer (Guildf)* **1999**, *40*, 735−745.

(40) Parrino, F.; D'Arienzo, M.; Mostoni, S.; Dirè, S.; Ceccato, R.; Bellardita, M.; Palmisano, L. Electron and Energy Transfer Mechanisms: The Double Nature of TiO_2 Heterogeneous Photocatalysis. *Top. Curr. Chem.* **2022**, *380*, 2.

(41) Bellardita, M.; Ceccato, R.; Dirè, S.; Loddo, V.; Palmisano, L.; Parrino, F. Energy Transfer in Heterogeneous Photocatalysis. *Subst. Int. J. Hist. Chem* **2019**, *3*, 49−57.

(42) Ozawa, T.; Hanaki, A. Hydroxyl Radical produced by the Reaction of Superoxide Ion with Hydrogen Peroxide: Electron Spin Resonance Detection by Spin Trapping. *Chem. Pharm. bull.* **1978**, *26*, 2572−2575.

(43) Wang, Z. J.; Ghasimi, S.; Landfester, K.; Zhang, K. A. I. Molecular Structural Design of Conjugated Microporous Poly(Benzooxadiazole) Networks for Enhanced Photocatalytic Activity with Visible Light. *Adv. Mater.* **2015**, *27*, 6265−6270.

(44) Desrat, S.; van de Weghe, P. Intramolecular Imino Diels—Alder reaction: Progress Toward the Synthesis of Uncialamycin. *J. Org. Chem.* **2009**, *74*, 6728−6734.

(45) Wang, H.; Ganesan, A. Total Synthesis of the Cytotoxic Alkaloid Luotonin A. *Tetrahedron Lett.* **1998**, *39*, 9097−9098.

(46) Wall, M. E.; Wani, M. C.; Cook, C. E.; Palmer, K. H.; McPhail, A. T.; Sim, G. A. Plant Antitumor Agents. I. The Isolation and Structure of Camptothecin, a Novel Alkaloidal Leukemia and Tumor Inhibitor from Camptotheca acuminata. *J. Am. Chem. Soc.* **1966**, *88*, 3888−3890.

(47) Anzini, M.; Cappelli, A.; Vomero, S.; Seeber, M.; Menziani, M. C.; Langer, T.; Hagen, B.; Manzoni, C.; Bourguignon, J. J. Mapping and Fitting the Peripheral Benzodiazepine Receptor Binding site by Carboxamide Derivatives. Comparison of Different Approaches to Quantitative Ligand-Receptor Interaction Modeling. *J. Med. Chem.* **2001**, *44*, 1134−1150.

(48) More, D. A.; Shinde, G. H.; Shaikh, A. C.; Muthukrishnan, M. Oxone Promoted Dehydrogenative Povarov Cyclization of: N-aryl Glycine Derivatives: An Approach Towards Quinoline Fused Lactones and Lactams. *RSC Adv.* **2019**, *9*, 30277−30291.

(49) Jia, X.; Peng, F.; Qing, C.; Huo, C.; Wang, X. Catalytic Radical Cation Salt Induced $C_{sp}3$-H Functionalization of Glycine Derivatives: Synthesis of substituted Quinolines. *Org. Lett.* **2012**, *14*, 4030−4033.

(50) Earle, M. J.; Vibert, A.; Jahn, U. Tris(4-bromophenyl)aminium Hexachloroantimonate. *Encyclopedia of Reagents for Organic Synthesis*, Wiley: 2011; Vol. *1*, pp 1−6.

(51) Dong, W.; Hu, B.; Gao, X.; Li, Y.; Xie, X.; Zhang, Z. Visible-Light-Induced Photocatalytic Aerobic Oxidation/Povarov Cyclization Reaction: Synthesis of Substituted Quinoline-Fused Lactones. *J. Org. Chem.* **2016**, *81*, 8770−8776.

(52) Wang, J.; Li, L.; Guo, Y.; Li, S.; Wang, S.; Li, Y.; Zhang, Y. Visible-Light-Enabled Aerobic Oxidative $C_{sp}3$-H Functionalization of Glycine Derivatives Using an Organic Photocatalyst: Access to Substituted Quinoline-2-Carboxylates. *Org. Biomol. Chem.* **2020**, *18*, 8179−8185.

(53) Cornel, V.; Lovely, C. J. Boron Trifluoride Etherate. *Encyclopedia of Reagents for Organic Synthesis*, Wiley: 2007; Vol. *2*.

(54) Fónagy, O.; Szabó-Bárdos, E.; Horváth, O. 1,4-Benzoquinone and 1,4-hydroquinone Based Determination of Electron and Superoxide Radical Formed in Heterogeneous Photocatalytic Systems. *J. Photochem. Photobiol. A* **2021**, *407*, 113057.

(55) Choi, H. S.; Kim, J. W.; Cha, Y.-N.; Kim, C. A Quantitative Nitroblue Tetrazolium Assay for Determining Intracellular Superoxide Anion Production in Phagocytic Cells. *J. Immunoassay Immunochem.* **2006**, *27*, 31−44.

(56) Watts, R. J.; Teel, A. L. Hydroxyl Radical and Non-Hydroxyl Radical Pathways for Trichloroethylene and Perchloroethylene Degradation in Catalyzed H_2O_2 Propagation Systems. *Water Res.* **2019**, *159*, 46−54.

(57) Hartman, P. E.; Hartman, Z.; Ault, K. T. Scavenging of Singlet Molecular Oxygen by Imidazole Compounds: High and Sustained Activities of Carboxy Terminal Histidine Dipeptides and Exceptional Activity of Imidazole-4 acetic acid. *Photochem. Photobiol.* **1990**, *51*, 59−66.

(58) Ouannes, C.; Wilson, T. Quenching of singlet oxygen by tertiary aliphatic amines. Effect of DABCO (1,4-diazabicyclo[2.2.2]octane). *J. Am. Chem. Soc.* **1968**, *90*, 6527−6528.

(59) Saito, I.; Inoue, K.; Matsuura, T. Occurrence of the Singlet-Oxygen Mechanism in Photodynamic Oxidations of Guanosine. *Photochem. Photobiol.* **1975**, *21*, 27−30.

(60) Rodríguez, E. M.; Márquez, G.; Tena, M.; Álvarez, P. M.; Beltrán, F. J. Determination of Main Species Involved in the First Steps of TiO_2 Photocatalytic Degradation of Organics with the Use of Scavengers: The Case of Ofloxacin. *Appl. Catal. B* **2015**, *178*, 44−53.

(61) Wang, Y.; Li, X.; Dong, X.; Zhang, F.; Lang, X. Triazine-Based Two Dimensional Porous Materials for Visible Light-Mediated Oxidation of Sulfides to Sulfoxides With O_2. *J. Colloid Interface Sci.* **2022**, *616*, 846−857.

(62) Bhadra, M.; Kandambeth, S.; Sahoo, M. K.; Addicoat, M.; Balaraman, E.; Banerjee, R. Triazine Functionalized Porous Covalent Organic Framework for Photo-Organocatalytic E-Z Isomerization of Olefins. *J. Am. Chem. Soc.* **2019**, *141*, 6152−6156.

(63) Sun, N.; Jin, Y.; Wang, H.; Yu, B.; Wang, R.; Wu, H.; Zhou, W.; Jiang, J. Photoresponsive Covalent Organic Frameworks with Diarylethene Switch for Tunable Singlet Oxygen Generation. *Chem. Mater.* **2022**, *34*, 1956−1964.

(64) Saito, H.; Nosaka, Y. Mechanism of Singlet Oxygen Generation in Visible-Light-Induced Photocatalysis of Gold-Nanoparticle-Deposited Titanium Dioxide. *J. Phys. Chem. C* **2014**, *118*, 15656−15663.

(65) Ushakov, D. B.; Plutschack, M. B.; Gilmore, K.; Seeberger, P. H. Factors Influencing the Regioselectivity of the Oxidation of Asymmetric Secondary Amines with Singlet Oxygen. *Chem.—Eur. J.* **2015**, *21*, 6528−6534.

(66) Hermitage, S.; Jay, D. A.; Whiting, A. Evidence for the Non-Concerted [4+2]-Cycloaddition of N-Aryl Imines when Acting as Both Dienophiles and Dienes under Lewis Acid-Catalysed Conditions. *Tetrahedron Lett.* **2002**, *43*, 9633−9636.

(67) Liu, Z.; Yang, X.; Yang, Z.; Su, X.; Xie, Z.; Chen, W.; Zhang, W.; Chen, L. Quinacridone Based 2D Covalent Organic Frameworks as Efficient Photocatalysts for Aerobic Oxidative Povarov Reaction. *Appl. Catal. B* **2022**, *312*, 121406.

(68) An, W.-K.; Zheng, S.-J.; Xu, X.; Liu, L.-J.; Ren, J.-S.; Fan, L.; Yang, Z.-K.; Ren, Y.; Xu, C. Integrating Benzofuran and Heteroradialene into Donor-Acceptor Covalent Organic Frameworks for Photocatalytic Construction of Multi-substituted Olefins. *Appl. Catal. B* **2022**, *316*, 121630.

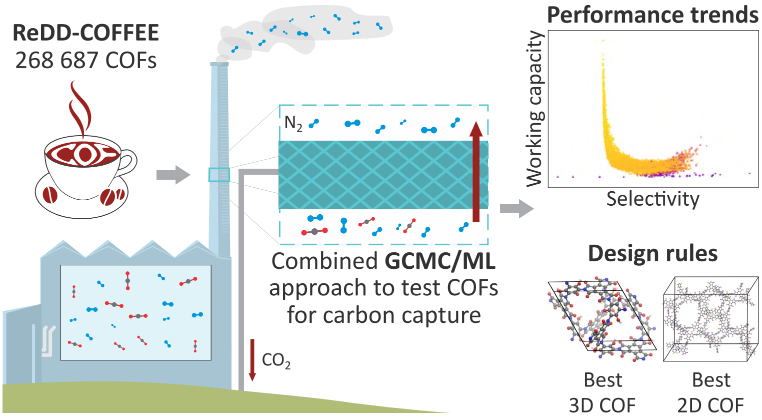
(69) Feike, M.; Demco, D. E.; Graf, R.; Gottwald, J.; Hafner, S.; Spiess, H. W. Broadband Multiple-Quantum NMR Spectroscopy. *J. Magn. Reson., Ser. A* **1996**, *122*, 214−221.

(70) Bräuniger, T.; Wormald, P.; Hodgkinson, P. Improved Proton Decoupling in NMR Spectroscopy of Crystalline Solids Using the SPINAL-64 Sequence. *Monatsh. Chem.* **2002**, *133*, 1549−1554.

Paper V

High-Throughput Screening of Covalent Organic Frameworks for Carbon Capture using Machine Learning

Juul S. De Vos, Siddharth Ravichandran, Sander Borgmans, Louis Vanduyfhuys, Pascal Van Der Voort, Sven M. J. Rogge*, and Veronique Van Speybroeck*



Chemistry of Materials, **2024**, 36 (9), 4315–4330

The Supporting Information is available at:
https://doi.org/10.1021/acs.chemmater.3c03230

J. S. De Vos proposed the screening procedure to identify promising candidates and reveal performance trends for carbon capture. He performed the force field calculations for the benchmarking and all GCMC simulations to characterize the COFs. Moreover, he trained and tested the ML algorithm that predicts the adsorption properties of all COFs. J. S. De Vos analyzed all results and prepared the initial draft of the manuscript.

*Corresponding authors

CHEMISTRY OF MATERIALS
pubs.acs.org/cm
Article

High-Throughput Screening of Covalent Organic Frameworks for Carbon Capture Using Machine Learning

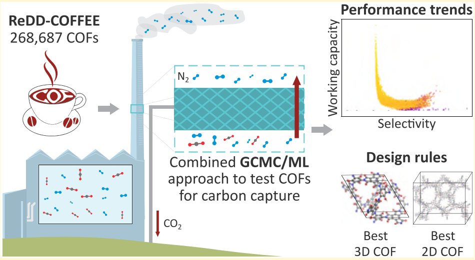
Juul S. De Vos, Siddharth Ravichandran, Sander Borgmans, Louis Vanduyfhuys, Pascal Van Der Voort, Sven M. J. Rogge,* and Veronique Van Speybroeck*

Cite This: *Chem. Mater.* 2024, 36, 4315−4330

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Postcombustion carbon capture provides a high-potential pathway to reduce anthropogenic CO_2 emissions in the short term. In this respect, nanoporous materials, such as covalent organic frameworks (COFs), offer a promising platform as adsorbent beds. However, due to the modular nature of COFs, an almost unlimited number of structures can possibly be synthesized. To efficiently identify promising materials and reveal performance trends within the COF material space, we present a computational high-throughput screening of 268,687 COFs for their ability to efficiently and selectively separate CO_2 from the flue gas of power plants using a pressure swing adsorption process. Furthermore, we demonstrate that our screening can be significantly accelerated using machine learning to identify a set of promising materials. These are subsequently characterized with high accuracy, taking into account the effects of competitive adsorption and coadsorption. Our screening reveals that imide, (keto)enamine, and (acyl)hydrazone COFs are particularly interesting for carbon capture. Additionally, the best-performing materials are 3D COFs possessing strong CO_2 adsorption sites between aromatic rings at opposite sides of pores with a diameter of 1.0 nm. In 2D COFs, a significant influence of the framework chemistry, such as imide linkages or fluoro groups, is observed. Our design rules can guide experimental researchers to construct high-performing COFs for CO_2 capture.



1. INTRODUCTION

Reducing anthropogenic greenhouse gas emissions, of which CO_2 is the most prominent, is one of the biggest challenges of today's society.[1] A promising path to significantly reduce these emissions is the capture of CO_2 from point sources, e.g., the flue gases of industrial power plants.[2,3] In this respect, covalent organic frameworks (COFs) hold great potential as adsorbents because of their large internal surface and high stability.[4,5] In addition, by virtue of their modular nature, an almost unlimited number of COFs can be envisioned.[6,7] However, this vast design space also renders the quest to find the ideal COF adsorbent cumbersome, if not impossible, when relying only on experimental synthesis and performance testing. It is precisely in this area that computational modeling may enormously facilitate the identification of the most promising materials in a fraction of the time and resources needed for their experimental synthesis. To accelerate the quest for nanoporous materials with excellent carbon capture performance, we screen the diverse ReDD-COFFEE database of 268,687 COFs for their ability to efficiently capture CO_2 molecules from the flue gas of industrial power plants.[8] Besides identifying the most promising COFs for this purpose in the database, our high-throughput screening also provides atomic insight into the driving force of favorable adsorption and selectivity behavior in these COFs. To guide experimental efforts in this topical research field, we establish several design strategies based on these insights to construct high-performing COFs for CO_2 capture.

The past decade has seen a sharp increase in natural disasters, such as extreme temperatures, heavy precipitation and pluvial floods, river floods, droughts, and storms.[9] These provide a severe warning that the rising amount of greenhouse gases in our atmosphere has unprecedented implications. Among the different greenhouse gases, CO_2 has the most significant long-term impact on the Earth's radiative balance due to its abundance and long lifespan.[1] Carbon capture and sequestration (CCS)[2,10,11] or utilization (CCU)[12,13] is one of the most viable approaches to reduce these anthropogenic CO_2 emissions in the short term. Among the different technologies

Received: December 18, 2023
Revised: March 28, 2024
Accepted: April 9, 2024
Published: April 25, 2024

being investigated to efficiently isolate the CO_2 molecules, postcombustion carbon capture of flue gases of industrial power plants has the largest potential. These flue gases mainly consist of a mixture of N_2 and CO_2 with a small amount of H_2O and other trace gases, such as O_2, SO_x, NO_x, and CO. The high interest in this technology originates from three key aspects, i.e., the relatively high concentration of CO_2 molecules in flue gases, the stationary point sources, and the easy installation on existing infrastructures.[3] The resulting outlet gas, consisting of a high-purity stream of CO_2 molecules extracted from the flue gas, can either be stored in empty oil reservoirs or other carbon sinks, or reused in industrial processes, thereby closing the carbon loop.[2,12] Key to this technology is the design of processes that selectively extract CO_2 from the other molecules present in the flue gases in an energetically efficient way, driving the need to design new separation materials. Current carbon capture technologies rely on amine scrubbing.[14] However, this process has several drawbacks, such as the high regeneration energy, fast equipment corrosion, and solvent decomposition, causing this process to be only applied for highly specialized purposes.[15]

An alternative approach is to divert the flue gas through fixed beds of nanoporous materials. These materials can separate the components of a mixture gas based on three principles, i.e., molecular sieving, kinetic separation, or equilibrium separation.[16] When the size and shape of the pores prevent one component from entering the material, molecular sieving is achieved. However, for the separation of CO_2 and N_2, the main components of flue gases, this sieving process is not very effective since both molecules have very similar sizes and shapes. Kinetic and equilibrium separations originate from a different diffusion rate and host−guest interaction strengths for the various gas components, respectively. In the context of postcombustion carbon capture, equilibrium separation is the most dominant and will, therefore, be investigated in this work.[17] Since the CO_2 molecules are only physisorbed to the pore walls rather than chemisorbed as in conventional amine scrubbing processes, their regeneration requires far less energy, which reduces the additional cost for carbon capture tremendously.[18] However, to adopt these materials efficiently in carbon capture, design rules must be established to synthesize them with targeted adsorption properties. It is exactly in this area that computational high-throughput screenings offer an efficient approach to rapidly determine such design rules, enabling researchers to significantly accelerate the quest for the ideal carbon capture adsorbent.

Examples of nanoporous materials that are proposed for carbon capture are activated carbon, zeolites, metal−organic frameworks (MOFs), and COFs.[5,19−23] Several challenges, such as reduced CO_2 selectivity or strong performance deterioration in the presence of water, limit the usage of activated carbons, MOFs, and zeolites in humid industrial environments, such as flue gases.[5,21] Therefore, the recent development of COFs, a new class of nanoporous materials, is of particular interest.[4,24,25] They are built from light organic building blocks held together by strong covalent bonds, resulting in materials with a very low mass density and high mechanical, thermal, and chemical stability. Together with their large internal surface, this makes these materials of particular interest for industrial carbon capture applications.[5,26−28] Furthermore, COFs are built up from so-called secondary building units (SBUs). Following the principle of reticular chemistry, these can be adopted to design an enormous amount of new, possibly hypothetical, materials.[6,7] While this is advantageous for their design possibilities, it makes pinpointing the most promising materials from this vast space cumbersome. To overcome this drawback, we present in this work a multistage computational screening to identify such materials efficiently and derive design rules to develop new materials.

Several attempts to collect all experimentally synthesized MOFs and COFs have resulted in a large material database,[29−34] and also the number of huge databases of hypothetical materials continues to increase rapidly.[8,35−40] With the upcoming availability of large computational resources, computational high-throughput screenings have emerged as a fast and efficient way to screen these material spaces and guide experimental researchers to highly interesting subclasses or promising candidates for further investigation.[41,42] The three main advantages of high-throughput screenings are their ability to quickly identify high-potential materials, establish trends between microscopic and macroscopic properties, and determine performance limits.[41] While this is true both for experimental and hypothetical databases, the latter can boast a broad and well-balanced representation of material space.[8,43] Furthermore, high-potential hypothetical materials identified from these high-throughput screenings also form an impetus to synthesize them experimentally, as demonstrated by Lan et al.[40] They managed to synthesize four new COFs following from an in silico design algorithm, among which two materials in a not-yet-observed 3D topology.[40]

Most high-throughput screenings have focused on exploring nanoporous materials for gas storage and separation processes.[44] Popular targets are the uptake of CH_4, which is the main component of natural gas,[8,29,35,36,38,39,45−47] or H_2 molecules[36,48−52] in MOFs and COFs due to their high potential as energy vectors to replace traditional fossil fuels. Examples of gas separation processes include natural gas purification[53−61] and pre-[57,62−64] and postcombustion[33,53,55,56,65−74] carbon capture. Recently, high-throughput screenings have also examined more complex properties, such as mechanical stability,[45,75] thermal conductivity,[76] catalytic activity,[77] and the electronic band gap.[32,78,79] Most of these high-throughput screenings adopt brute-force screening of an existing database. However, the number of available materials is quickly growing, such that brute-force screenings are not feasible anymore, even with the currently available computational resources.[40,80] Recently, several studies started adopting machine learning (ML) algorithms to speed up these high-throughput screenings.[81,82] Within the field of gas storage and separation in MOFs and COFs, ML has been applied to predict the uptake of H_2,[83−86] CH_4,[80,87−90] CO_2,[91,92] and N_2,[93] as well as the separation of CO_2/H_2,[92,94,95] CO_2/N_2,[92,96,97] CO_2/CH_4,[97,98] CH_4/H_2,[99] and noble gases.[100−102103] Other applications can be found in predicting ideal synthesis conditions,[104] mechanical stability,[75] and electronic properties.[32,78,79] Since most ML algorithms are adopted to predict material properties and reproduce performance trends, the question remains whether they can be leveraged to find the most promising materials for a given application.

In this work, we make a leap forward on this path by screening our recently developed hypothetical ReDD-COFFEE database for postcombustion carbon capture. This allows us to establish performance trends and identify subclasses with promising behavior. Furthermore, the atomic insights obtained from the screening are adopted to determine a set of design rules to guide experimental researchers toward high-performing COFs for carbon capture from the flue gases of industrial power plants. As illustrated in Figure 1, our screening procedure adopts a three-step process to discover materials within the diverse ReDD-COFFEE database with high CO_2/N_2 separation performance. Since a full characterization of all 268,687 COFs in this database would require a huge amount of computational resources, we adopt an ML approach to speed up the screening. Our ML models are trained on a diverse subset of the database and extract a set of promising COFs with excellent adsorption properties. These high-potential materials are subsequently characterized more accurately by taking into account the effects of competitive adsorption and coadsorption. Finally, the top-performing materials are analyzed to gain atomic insight into the driving forces of their improved performance for postcombustion carbon capture.

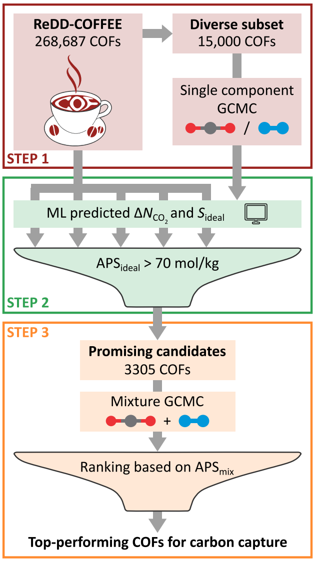


Figure 1. Visualization of the high-throughput screening procedure followed in this work. Starting from the ReDD-COFFEE database, the top-performing COFs for carbon capture in terms of the adsorbent performance score (APS) are identified in a three-step process. After an initial screening of a diverse subset using idealized single-component grand-canonical Monte Carlo (GCMC) calculations, all structures in the database are characterized using a machine learning (ML) model. The most promising COFs are tested more accurately with a mixture GCMC, from which overarching design rules are formulated.

2. METHODOLOGY

2.1. Screening Procedure. Throughout our screening procedure, COFs are investigated for their applicability in postcombustion carbon capture using a pressure swing adsorption (PSA) process between 0.1 and 10 bar. Similar to previous studies, the temperature is fixed at 298 K.[55,62,74] Two key properties of interest characterize such a process: the CO_2 working capacity, ΔN_{CO_2}, and the CO_2/N_2 selectivity, S. As defined in eq 1, the working capacity is the difference between the single-component uptake of CO_2 molecules at adsorption conditions (here, 10 bar at 298 K), $N_{CO_2}^{ads}$, and desorption conditions (here, 0.1 bar at 298 K), $N_{CO_2}^{des}$:

$$\Delta N_{CO_2} = N_{CO_2}^{ads} - N_{CO_2}^{des} \tag{1}$$

This metric indicates how many CO_2 molecules can be adsorbed and desorbed in each cycle per unit of material and thus how much COF material is required to separate a given amount of gas.

Furthermore, the CO_2/N_2 selectivity is the ratio of CO_2 molecules, $N_{CO_2}^{ads}$, to N_2 molecules, $N_{N_2}^{ads}$, adsorbed by the material at adsorption conditions, and thus determines the purity of the outlet gas. A selectivity above one indicates that the material favors CO_2 over N_2. Two methods to calculate this selectivity will be adopted in this work. To determine the ideal selectivity, S_{ideal}, the CO_2 and N_2 uptakes have to be calculated separately using single-component grand canonical Monte Carlo (GCMC) simulations, according to eq 2.

$$S_{ideal} = \frac{N_{CO_2}^{ads}}{N_{N_2}^{ads}} \text{ (single-component GCMC)} \tag{2}$$

However, these calculations do not consider the interactions between the gas components. To describe competitive adsorption and coadsorption, the mixture selectivity, S_{mix}, is calculated from a mixture GCMC calculation, mimicking the experimental conditions of the flue gas with a binary inlet gas, as defined in eq 3.

$$S_{mix} = \frac{N_{CO_2}^{ads}/y_{CO_2}}{N_{N_2}^{ads}/y_{N_2}} \text{ (mixture GCMC)} \tag{3}$$

In these simulations, the mole fractions of both components of the inlet mixture, y_{CO_2} and y_{N_2}, are provided as input parameters.

Usually, there is a trade-off between a material's working capacity and its selectivity.[56] Whereas the working capacity favors materials with large pores to maximize the number of gas molecules stored under adsorption conditions, the selectivity prefers smaller pores to maximize the interaction with the pore walls. Therefore, finding the best-performing COF is a trade-off. The adsorbent performance score (APS), defined in eq 4, combines the two key characteristics into a single parameter that allows ranking materials by their carbon capture potential.

$$APS_x = \Delta N_{CO_2} S_x \tag{4}$$

The subscript x can be "ideal" or "mix", depending on the adopted method to calculate the CO_2/N_2 selectivity. This metric results in a robust and effective ranking of the top candidates since the product of both adsorption characteristics exponentially increases for the best-performing structures. Therefore, it is less sensitive to small deviations in the predicted properties as compared to other metrics, such as the product of the working capacity with the logarithm of the selectivity.[105]

https://doi.org/10.1021/acs.chemmater.3c03230
Chem. Mater. 2024, 36, 4315−4330

Conventional adsorbents, such as activated carbon, zeolites, and MOFs, have been investigated for their CO_2/N_2 selectivity and the CO_2 working capacity. Mg-MOF-74, a widely studied MOF in the field of carbon capture, demonstrates an APS_{mix} of 600 mol/kg (ΔN_{CO_2} = 3 mol/kg, S_{mix} = 200) under vacuum swing adsorption conditions between 0.1 and 1 bar at 300 K.[106] Activated carbon has an APS_{ideal} of 22.86 mol/kg (ΔN_{CO_2} = 2.58 mol/kg, S_{ideal} = 8.86) in the same pressure range and at 293 K.[107] The adsorption properties of zeolites 13X and 5A were experimentally characterized at 298 K and between 0.1 and 1 bar. They possess an APS_{ideal} of 40.14 mol/kg (ΔN_{CO_2} = 2.3 mol/kg, S_{ideal} = 17.45) and 194.37 mol/kg (ΔN_{CO_2} = 4.18 mol/kg, S_{ideal} = 46.5), respectively.[108,109] Using pressure swing adsorption between 1 and 10 bar, zeolites NaX and NaY demonstrate a very high APS_{mix} of 3600 mol/kg (ΔN_{CO_2} = 1.2 mol/kg, S_{mix} = 3000) and 1300 mol/kg (ΔN_{CO_2} = 2.6 mol/kg, S_{ideal} = 500), respectively.[110] COFs can achieve an exceptionally high CO_2 storage capacity,[111−114] with COF-105 and COF-108 reaching 82 and 96 mol/kg at 30 bar and 300 K,[115] which exceeds the 33 mol/kg of the well-reported MOF-177.[116] At this point, however, the possible limits of COFs for CO_2 adsorption remain unknown, and it is unclear which of the COFs are best suited for this application.

To answer these questions, we follow the three-step approach in Figure 1 to establish performance trends and identify promising candidates within the COF material class. Our screening study starts from our recently developed ReDD-COFFEE database, containing a diverse set of 268,687 COFs.[8] These hypothetical materials cover a broad range of material space and sample each region equally. Whereas experimental databases, such as the CoRE and CURATED COF databases,[33,34] are biased toward linkages that are abundantly present in the literature, the hypothetical databases of Martin et al.[38] and Mercado et al.[39] only contain a limited number of linkages since they focused on a large set of linkers. All structures in this database were optimized with respect to an ab-initio-derived force field. The starting geometry of the 2D COFs contained two layers in an eclipsed AA stacking, which mostly optimized to a configuration with a small nonzero offset between the layers. This approach is justified by the fact that most COFs have a slightly inclined stacking, although there are exceptions with AB stacking.[117,118] Each material is labeled with top_SBU$_1$_SBU$_2$_···_SBU$_N$, where "top" specifies the material's topology and SBU$_i$ is the building block placed on the i^{th} set of equivalent nodes. The secondary building units (SBUs) are denoted with three numbers, the first one indicating the linker core and the last one the formed linkage with other SBUs. For example, the building block 02−03−07 corresponds to core02 and can be adopted to form an imide linkage (link07). As explained in ref 8, the middle number is introduced for technical reasons to discriminate between different parts of the linkage. The relevant linkages discussed in this work are visualized in Figure S15 of the SI.

In the first step of our screening approach, as depicted in Figure 1, we calculate the CO_2 working capacity, ΔN_{CO_2}, and the ideal CO_2/N_2 selectivity, S_{ideal}, for a representative subset of 15,000 COFs of the ReDD-COFFEE database to limit the enormous amount of computational power that would be needed to screen the whole database. Similar to our earlier work and as explained in Section S3 of the SI, the 15,000 COFs in the subset are selected iteratively in which, in each iteration, the structure that has the largest minimal distance to the already selected structures is added to the subset.[8,43] This distance is measured in the space described by the ML features defined below. As demonstrated in Figure S14 of the SI, the subset fully covers the material space described by the full database and maintains its high diversity. Therefore, the ML models trained on this subset are expected to generalize well to all COFs in the database, as is done in the next step of our high-throughput screening.

In the second step, an ML model is trained on the subset, which is split into a training set of 10,000 COFs and a test set of 5000 COFs, to predict five adsorption targets, i.e., the CO_2 uptake at adsorption conditions, $N_{CO_2}^{ads}$, the CO_2 uptake at desorption conditions, $N_{CO_2}^{des}$, the N_2 uptake at adsorption conditions, $N_{N_2}^{ads}$, the CO_2 working capacity, ΔN_{CO_2}, and the ideal CO_2/N_2 selectivity, S_{ideal}. To characterize the pore geometry of each material, 12 textural features are adopted, i.e., the largest pore diameter, the largest free pore diameter, the pore limiting diameter, the mass density, the gravimetric and volumetric accessible surface area, the gravimetric and volumetric accessible pore volume, and their nonaccessible counterparts. Furthermore, 101 revised autocorrelation functions (RACs) describe the chemical environments of the linkers, linkages, and functional groups.[119] These features, which have proven to be useful in numerous applications, determine relations of heuristic atomic properties, such as size, connectivity, and electronegativity, on the crystal graph.[8,43,119−122] A formal definition of all ML features is provided in Section S4.1 of the SI, together with an analysis of the influence of reducing the number of features. Since training the ML models with all 113 features was sufficiently fast, all features are included in the development of the final models. Together, these features are capable of providing a physical intuition for the design rules identifying good carbon capture materials, as explained in Section S4.4 of the SI. Once all ML models are trained, those that predict the working capacity and ideal selectivity most accurately are adopted to determine the adsorption characteristics of all 268,687 COFs in the database.

Neglecting the effects of competitive adsorption and coadsorption in the first two steps can drastically underestimate the absolute value of the selectivity.[60] However, as demonstrated in Section S2.3 of the SI, the performance of a set of materials can still be accurately ranked qualitatively when characterized under ideal circumstances. This motivates our choice to perform these first two steps based on computationally less expensive ideal selectivities and use their qualitative ranking to select a large set of promising materials, exhibiting APS_{ideal} > 70 mol/kg, in the third and last step of our high-throughput screening. This APS threshold is chosen sufficiently small to include all materials likely to result in a significant change in selectivity when switching from ideal to mixture conditions and to focus on COFs that can challenge conventional adsorbents. The mixture selectivities of these 3305 materials are calculated with GCMC simulations of a binary flue-gas-like mixture with a CO_2:N_2 ratio of 15:85 to incorporate the effects of competitive adsorption and coadsorption. Once these characteristics are computed, the resulting APS_{mix} values provide an effective ranking to identify the top-performing COF materials for postcombustion carbon capture.

2.2. Selection of Appropriate Force Fields. Several GCMC calculations are performed throughout our high-throughput screening. Whereas such simulations can accurately calculate adsorption isotherms, it has been shown that the resulting isotherms are very sensitive to the adopted force field parameters.[123] Therefore, the interatomic interactions must be accurately described. The partial charges that define the electrostatic interactions are directly adopted from the system-specific force fields provided in the ReDD-COFFEE database since they are fitted to ab initio reference data. For the van der Waals interactions, several candidates are proposed for the host−guest and guest−guest interactions. Whereas host−guest and guest−guest interactions can be described by the generic MM3,[124] UFF,[125] or DREIDING[126] force fields, the guest−guest interactions for CO_2 and N_2 can also be determined using the TraPPE[127] force field. Furthermore, the interactions between CO_2 molecules can be described with the EPM2 force field.[128] A thorough benchmark study is performed in Section S1 of the SI to select the most appropriate combination of van der Waals interactions.

A first set of validation data is obtained by reproducing the CO_2 and N_2 isotherms in COF-1, COF-5, COF-102, and COF-103 as measured by Furukawa and Yaghi.[129] Whereas for N_2 only low-pressure isotherms at 77 K were reported, high-pressure isotherms of CO_2 at 273 and 298 K and low-pressure isotherms at 273 K were measured. Each of these isotherms is also calculated with a total of 12 or 15 different force fields for N_2 and CO_2, respectively, which all differ in the description of the van der Waals interactions. All resulting isotherms are reported in Section S1.3 of the SI. Changing the host−guest interactions results in the most significant deviations between the different isotherms. Generally, the predicted CO_2 or N_2 uptakes are the highest when the host−guest interactions are described with the UFF parameters, whereas the lowest uptakes are obtained with MM3. Although the influence of the guest−guest interactions is more modest, it can be observed that more CO_2 molecules are adsorbed at a given pressure when a DREIDING description is adopted for this interaction type. The lowest CO_2 uptake is observed when applying the TraPPE or EPM2 force fields. Whereas the trends are less pronounced for the low-pressure N_2 isotherms, an MM3 description of the guest−guest interactions always results in the lowest uptakes, followed by UFF.

Besides comparison with the experimental isotherms, we have also calculated the Henry coefficient at an ab initio level. Typically, the Henry coefficient is calculated with Widom insertion.[130] However, since this approach requires a vast number of energy evaluations, only a computationally cheap level of theory, such as force fields, can be adopted. To obtain a highly accurate, ab initio reference of the Henry coefficient, the importance sampling scheme devised by Vandenbrande et al.[131] is adopted. In our benchmark study, the ab initio Henry coefficients of CO_2 and N_2 in COF-1 and COF-5 are calculated and compared with the force field-derived Henry coefficients computed by using the regular Widom insertion method. All Henry coefficients can be found in Section S1.2 of the SI. Again, it can be observed that adopting UFF for the host−guest interactions results in the highest uptakes for both CO_2 and N_2, followed by the DREIDING force field and the MM3 force field, respectively. Comparable values for the Henry coefficient are obtained when describing the guest−guest interactions with UFF, DREIDING, and MM3. However, since the TraPPE and EPM2 force fields specify their own partial charges, switching to these for the guest−guest interactions has a significant impact. Whereas the Henry coefficient of CO_2 drastically lowers with these force fields, adopting the TraPPE force field for the guest−guest interactions of N_2 results in a slightly higher Henry coefficient.

Overall, describing the host−guest interactions with the MM3 force field combined with representing the guest−guest interactions by the UFF force field reproduces both the CO_2 and N_2 isotherms most accurately, while providing a reasonable guess for the Henry coefficient. We will, therefore, use this combination throughout this work.

3. COMPUTATIONAL DETAILS

All GCMC calculations in this work are performed with RASPA.[132] The optimized geometries, as well as the partial charges and the host−guest van der Waals interactions, are adopted from the ReDD-COFFEE database and the ab initio derived force fields contained within. These partial charges are derived using the MBIS scheme,[133] whereas the van der Waals interactions are described with a Buckingham potential using the MM3 parameters.[124] A similar approach is followed to derive the partial charges and MM3 parameters for the CO_2 and N_2 guests. The UFF parameters that describe the guest−guest van der Waals interactions are assigned by using our in-house force field generator software. A cutoff radius of 14 Å is applied for all interactions, and tail corrections are introduced. As explained in detail in Section S2.2 of the SI, we expect that framework flexibility will have a limited influence on the APS of the best-performing materials. Therefore, to reduce the required computational resources, both the framework and guests are considered rigid. The probe-occupiable pore volume that is not accessible to the guest molecules was found to be negligible, as demonstrated in Section S2.1 of the SI, and hence was not explicitly discarded in our GCMC simulations. Each simulation cell is defined as at least twice the cutoff distance in each direction. Each simulation at elevated pressure runs for a total of 20,000 cycles, from which the first 10,000 are discarded for equilibration. As demonstrated in Section S1.3 of the SI, the simulations performed during the benchmark study illustrate that this is sufficient to obtain converged results. For the six best 2D and 3D COFs, an additional simulation is performed with 100,000 cycles to create a smooth CO_2 distribution for Figures 7 and S23−S34 of the SI. Also here, the first 10,000 steps are discarded. The Widom insertion calculations determining the Henry coefficients of CO_2 and N_2 at infinite dilution with different force fields adopt a total of 10^5 steps.

In the first step of our high-throughput screening, single-component GCMC calculations are performed for 15,000 COFs under three different conditions, i.e., for CO_2 and N_2 at adsorption conditions (10 bar, 298 K) and CO_2 at desorption conditions (0.1 bar, 298 K). A total of 20, 11, and 13 structures are discarded for the three different conditions, respectively, due to prohibitively large pores that prevented the calculation of the interaction between the guests and the framework within a reasonable time frame. Since some materials are present in more than one set of rejected structures, there are 30 COFs for which not all adsorption characteristics could be calculated. In the last step of the high-throughput screening, adsorption properties are calculated for a set of 3305 promising candidates, selected from the full ReDD-COFFEE database based on the ML predicted adsorption characteristics. 617 among them are already present in the subset examined during

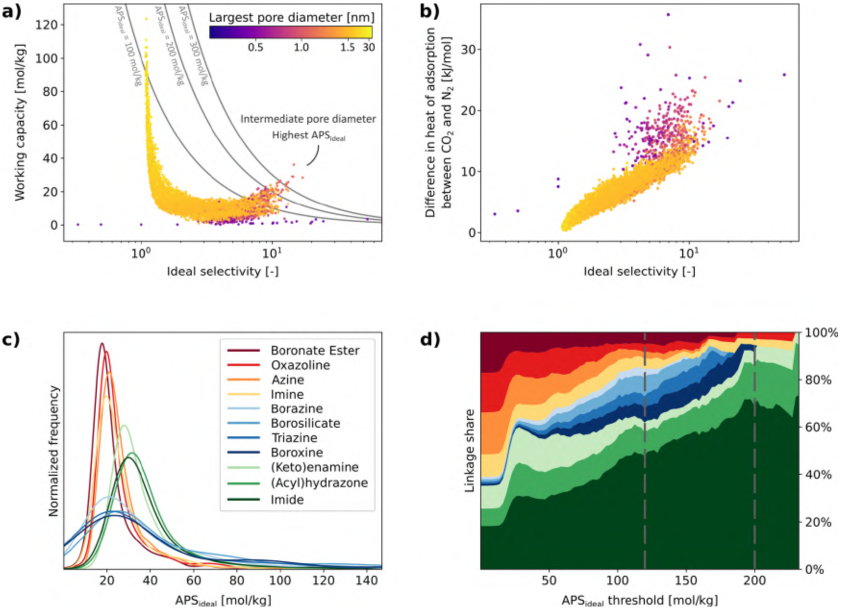


Figure 2. Analysis of screening the diverse subset of 14,970 COFs in step 1. (a) Relationship between the ideal selectivity and the working capacity. The color code, indicating the largest pore diameter, is piecewise linear between 0 and 1.5 nm and between 1.5 and 34.3 nm, i.e., the highest value of the largest pore diameter. Three thresholds for APS_{ideal} are indicated. (b) Difference of the heat of adsorption at operando conditions (see Section S3.2 of the SI) between CO_2 and N_2 uptakes as a function of the ideal selectivity and the largest pore diameter. (c) Histograms of APS_{ideal} for each class of linkage type in the database. (d) Evolution of the linkage share of different subsets with increasing threshold of APS_{ideal}. Panels a and b, as well as panels c and d, share the same color code.

the first step of the screening. Three additional single-component GCMC calculations are performed to obtain the ideal selectivity and working capacity for each of the remaining materials. Equivalent to the first step, a total of 103, 107, and 93 structures are discarded for the adsorption of CO_2 and N_2 and desorption of CO_2, respectively. For all 3305 COFs, a mixture GCMC is performed with a 15:85 ratio of CO_2 and N_2 at adsorption conditions, mimicking the composition of the flue gas of industrial power plants.[3] 12 COFs are discarded for this type of calculation since they require a prohibitively long computation time. For a total of 3180 materials, all single-component and mixture GCMC calculations were finished within a reasonable time.

All ML calculations are performed using the scikit-learn package.[134] As explained in Section 2.1, each COF is featurized with 12 textural parameters, calculated with Zeo++,[135] and 101 RACs to describe the chemical environment of the linkers, linkages, and functional groups (see Section S4.1 of the SI). To prevent bias due to the range of each feature, they are normalized to have zero mean and a standard deviation of one. The five adopted ML algorithms are kernel ridge regression, support vector regression, multilayer perceptron regression, random forest regression, and gradient boosting regression. While training all algorithms, except the multilayer perceptron, a heavier weight is given to materials with a high target value, since we are primarily interested in accurately predicting COFs with good adsorption characteristics. The hyperparameters of each algorithm are selected using shuffled 10-fold cross-validation on a diverse training set of 9973 COFs to avoid overfitting. The ones that provided the lowest mean absolute error (MAE) averaged over the ten cross-validation runs are adopted to predict the targets of the test set, which contains a total of 4997 COFs. The grid of benchmarked hyperparameters and the finally adopted ones are reported in Section S4.2 of the SI.

4. RESULTS

4.1. Step 1: Screening of the Diverse Subset.

A first indication of the carbon capture performance of COFs is obtained by calculating the CO_2 working capacity, ΔN_{CO_2}, and the ideal CO_2/N_2 selectivity, S_{ideal}, on a subset of 14,970 COFs. The results of this first step of our high-throughput screening are summarized in Figure 2. In Figure 2a, the working capacity is visualized as a function of the ideal selectivity, with the largest pore diameter of each structure indicated. The influence of switching from ideal to mixture selectivity is visualized in Figure S13 of the SI.

Figure 2a demonstrates that, in principle, the CO_2 working capacity of COFs can become very large, up to 123.3 mol/kg. However, these materials contain pores with very large diameters (>10 nm). Hence, their selectivity will be low since there is only limited interaction between the guests and the framework. In contrast, for very-small-pore COFs, with pores between 0.35 and 0.4 nm, the interactions with the pore walls are maximized, and the selectivity increases up to 53.5.

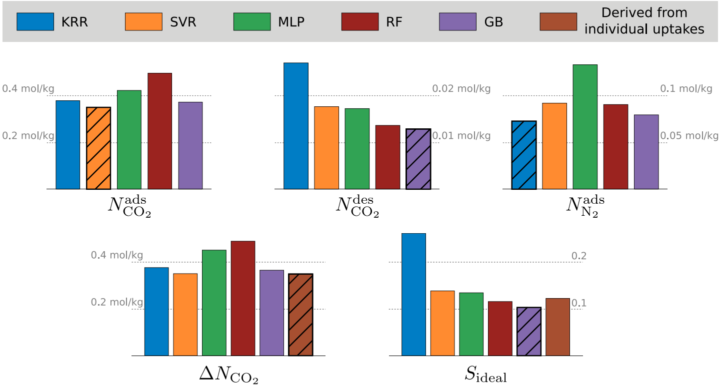


Figure 3. Mean absolute test error of all models for the five targets. Each of five ML algorithms, i.e., kernel ridge regression (KRR), support vector regression (SVR), multilayer perceptron regression (MLP), random forest regression (RF), and gradient boosting regression (GB), is trained on a set of 9973 COFs. For the working capacity and the ideal selectivity, the derived model is also reported, which is indirectly predicted from the ML predictions of the individual uptakes. The algorithm resulting in the lowest MAE is hatched.

Pores with this size are able to fit CO_2 molecules, which have a kinetic diameter of 0.330 nm, perfectly, whereas they cannot accommodate N_2 guests, which have a larger kinetic diameter of 0.364 nm.[136] Other studies have demonstrated that the materials with the highest ideal selectivity also exhibit a significant difference between the heat of adsorption of CO_2 and N_2 at infinite dilution.[55,71,72] Similar to the observations made by Altintas et al.,[56] Figure 2b illustrates that these trends remain valid at adsorption conditions (10 bar, 298 K), for which the heat of adsorption is defined in eq S3.2 of the SI. In these structures, the CO_2 molecules will bind more tightly to the adsorption sites than do the N_2 guests. These large selectivities come at the cost of having a very low working capacity, as seen in Figure 2a. Whereas the largest working capacities and selectivities are observed for COFs with a very high and very low pore diameter, respectively, both properties can also achieve relatively high values in a regime of pore diameters in between these two extremes. As displayed in Figures 2a and S16 of the SI, these structures have a pore diameter of around 1.0 nm. Since these materials are endowed with both a large working capacity and selectivity, their APS_{ideal} is maximized. With a working capacity of 35.9 mol/kg and an ideal selectivity of 14.7, the highest APS_{ideal} amounts to 527.8 mol/kg and is observed for the (acyl)hydrazone COF named twt-b_11−04−05_11−02−05_None_None. Its selectivity is comparable to the ideal selectivity of other conventional adsorbents, such as zeolites 13X and 5A, but it has a much higher working capacity. Its structure is visualized in Figure S17 of the SI, together with the materials that achieve the highest working capacity and selectivity.

To better understand the influence of the different linkage types observed in COFs on the adsorption properties, we partitioned the screened subset into different classes. When two COFs are in the same class, the same linkage type holds their building blocks together. The APS_{ideal} distribution for each of these classes is given in Figure 2c. The (acyl)-hydrazone, imide, and (keto)enamine COFs (green in Figure 2c,d) typically exhibit higher APS_{ideal} values, with medians of 33.1, 31.8, and 30.6 mol/kg, respectively. This also follows from Figure 2d, where we visualize the fraction of each class in a varying subset. Each subset is obtained by retaining only the structures with an APS_{ideal} above the threshold indicated on the *x*-axis. When no threshold is imposed, or if it equals the minimal APS_{ideal} value of 0 mol/kg, all 14,970 COFs are selected. Among the structures present in this initial screening set, the borazine, borosilicate, triazine, and boroxine linkages (blue in Figure 2c,d) possess only a collective linkage share of 3.7%. This is because the requirement of three-connected vertices in these materials strongly limits the available topologies and, therefore, also limits the number of these structures in the ReDD-COFFEE database. When increasing the APS_{ideal} threshold, the retained subset is biased toward the best-performing COFs. Although these four linkage types are observed less frequently in the initial screening set, their linkage share increases to 22.4% when the threshold achieves an intermediate value of 120 mol/kg. This can also be observed in Figure 2c, as the distributions of these linkage classes have a longer tail at these intermediate APS_{ideal} values compared to those of the other classes. The structures with a high APS_{ideal} are dominated by imide-linked COFs. Among the materials with an APS_{ideal} exceeding 200 mol/kg, 69.4% contain this linkage. Other frequently observed linkage types are (acyl)hydrazone and (keto)enamine, representing linkage shares of 16.7 and 5.6% for this APS_{ideal} threshold, respectively. From Figure 2c,d, we can conclude that structures possessing one of these linkages afford the highest APS_{ideal} values.

Since this initial screening is carried out on a diverse subset of 14,970 materials, we are confident that a screening of the entire COF database would provide the same trends in the carbon capture performance. In Section 4.2, the adsorption characteristics of all materials in our database are predicted by a machine learning model fitted to the results obtained in this screening.

4.2. Step 2: Machine Learning Prediction of the Working Capacity and Selectivity. In the next step of our high-throughput screening, it is our intention to identify the top-performing COFs for carbon capture from all 268,687 structures in the ReDD-COFFEE database. However, since the adsorption properties were calculated explicitly for only a subset of the database in Section 4.1, the CO_2 working

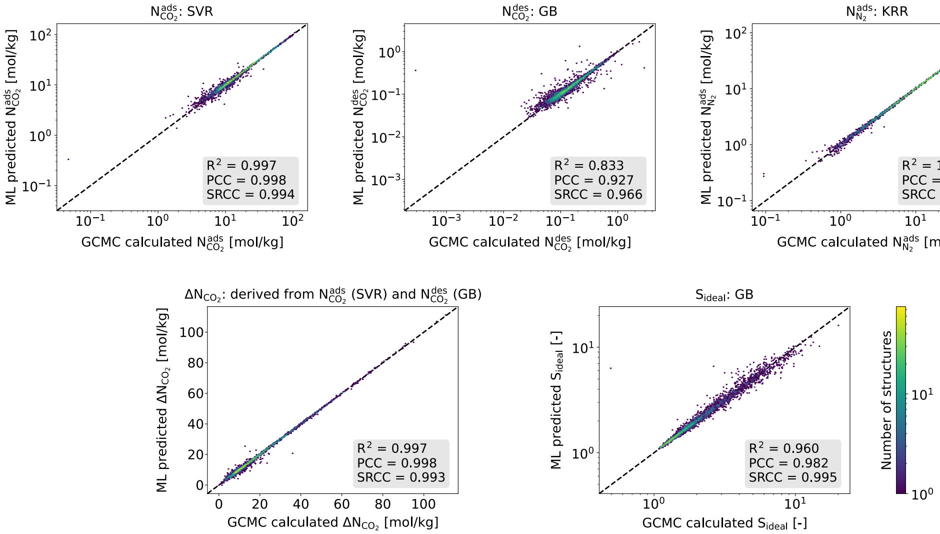


Figure 4. 2D histograms of the predictions on the test set of 4997 COFs for the best-performing model for each of five targets. These targets are the adsorption of CO_2 and N_2 at 293 K and 10 bar, the desorption of CO_2 at 293 K and 0.1 bar, the CO_2 working capacity, and the ideal CO_2/N_2 selectivity. The dashed line indicates an exact prediction. The coefficient of determination (R^2), Pearson correlation coefficient (PCC), and Spearman rank correlation coefficient (SRCC) of all predictions are reported in the insets. A zoomed-in view of these histograms is provided in Figure S19 of the SI.

capacity, ΔN_{CO_2}, and the ideal CO_2/N_2 selectivity, S_{ideal}, should still be determined for most structures. These uptake characteristics are predicted using ML models to speed up the screening procedure and avoid unnecessary GCMC calculations.

To predict the working capacity and selectivity, two approaches are adopted. On the one hand, they are directly predicted using two different ML models: one for the working capacity and one for the selectivity. On the other hand, they are also predicted indirectly by first training three ML models to predict the individual uptakes of CO_2 at adsorption and desorption conditions and N_2 at adsorption conditions, subsequently deriving the two aforementioned adsorption properties. For each of these five targets, a total of five different ML algorithms are trained, i.e., kernel ridge regression, support vector regression, multilayer perceptron regression, random forest regression, and gradient boosting regression. These models have been proven to predict adsorption properties in nanoporous materials accurately.[43,87,137] This results in 25 ML models, each trained on 9973 COFs out of the 14,970 structures characterized in Section 4.1, whereas the test set is formed by the remaining 4997 materials.

Whereas a full overview of error metrics and statistics for the training and test sets is provided in Tables S7 and S8 of the SI, respectively, the mean absolute errors (MAEs) of all models achieved on the test set are summarized in Figure 3. The ML models that predict the adsorption and desorption CO_2 uptakes and adsorption N_2 uptake most accurately are support vector regression, gradient boosting regression, and kernel ridge regression, respectively. As indicated in Figure 4, which visualizes the predictions of the ML models, these achieve coefficients of determination of 0.997, 0.833, and 1.000, respectively. Together with the Pearson and Spearman rank correlation coefficients, these statistics demonstrate that the ML models can predict the uptakes at adsorption conditions better than the one at desorption conditions since the amount of guests that can be adsorbed at the elevated pressure of 10 bar is highly correlated with the geometric properties that form part of the ML features. At 0.1 bar, the chemistry of the framework is more important, which is more difficult to describe, even with the specialized RACs.[90] This is also the case for the adsorption of the CO_2 molecules since these interact more strongly with the framework than the N_2 guests due to their increased quadrupolar moment. Therefore, the prediction of the CO_2 uptake is less accurate at adsorption conditions than that of the N_2 uptake.

As the CO_2 uptake at adsorption conditions is two orders of magnitude larger than at desorption conditions, this is the most dominant factor of the CO_2 working capacity. Therefore, the ML models trained directly on the working capacity primarily learn the CO_2 uptake at adsorption conditions and have more difficulty grasping the small corrections formed by the CO_2 uptake at desorption conditions. The resemblance between the ML models trained on the CO_2 uptake at adsorption conditions and the working capacity can be observed by the similarity in the resulting MAEs in Figure 3. A specialized ML model can learn the small deviations of CO_2 uptake at desorption conditions more accurately. This explains why the working capacity can be predicted slightly better by an indirect model derived from the individual uptakes at adsorption and desorption conditions, which attains an MAE of 0.351 mol/kg.

For the prediction of the selectivity, the interplay between the CO_2 and N_2 uptakes is more critical. This quantity is very sensitive to small fluctuations in the N_2 uptake and depends strongly on the description of the chemical environment.[102] This results in a large range of selectivity values, including some outliers with much higher selectivity compared with the majority of the materials. Whereas a quantitative prediction of these values is challenging using our single-component GCMC simulations as input, they are still ranked correctly, which is the most important for our goal to identify promising COFs for further screening. Moreover, as discussed in Section 4.1, the materials with the highest APS_{ideal} demonstrate both a relatively high working capacity and selectivity but do not have exceptionally high values for one of these properties. The results of Figure 3 show that the gradient-boosting regression model reaches the lowest MAE of 0.103 for the selectivity. The absolute values of this ideal selectivity are more challenging to predict than the ones of the working capacity, as indicated in Figure 4 by the lower coefficient of determination and Pearson correlation coefficients of 0.960 and 0.982, respectively. However, more importantly, the model still manages to predict the order of the selectivities accurately, as proven by a high Spearman rank correlation coefficient of 0.995.

To gain insight into the importance of each feature for a good prediction of the adsorption targets, a SHAP (SHapley Additive exPlanations) analysis is performed in Section S4.4 of the SI.[138] Whereas some features, such as the gravimetric accessible volume and the mass density, consistently contribute significantly to the predicted result, the trends in the relative importance of the different domains in COF chemistry depend strongly on the adopted ML model. The adsorption properties cannot be attributed to one of these domains specifically but rather emerge as a combined effect. To investigate these collective characteristics that result in outstanding COFs for carbon capture and to establish design rules interpretable by experimental scientists, a thorough investigation of the best-performing materials should be performed, as done in Section 4.4.

After this analysis of the ML models, the CO_2 working capacity and CO_2/N_2 selectivity can be predicted for all 268,687 materials in the ReDD-COFFEE database. As explained above, the working capacity is derived from the individual CO_2 uptakes at adsorption and desorption conditions, predicted by support vector regression and gradient boosting regression, respectively. In contrast, the selectivity is directly obtained by another gradient-boosting regression model. A density plot of the ML-predicted working capacities and ideal selectivities for all 268,687 structures in the database is provided in Figure 5. This strongly resembles Figure 2a on the subset of 14,970 COFs, proving the anticipated generalizability of our ML approach. To identify the most promising candidates, a threshold of 70 mol/kg for APS_{ideal} is introduced. All 3305 materials for which the APS_{ideal} exceeds this threshold are retained for further screening in step 3 below.

4.3. Step 3: Mixture Selectivity of the Most Promising COFs.

Whereas the ideal selectivity calculated in the previous steps already provides initial insight into the separation performance of COFs, it does not include the effects of competitive adsorption or coadsorption. Since the ideal selectivity is determined based on single-component GCMC calculations, the adsorption sites of the framework are available for both gas components. However, in reality, only one molecule can be adsorbed at each location. The preferred species depends on the nature of the adsorption site. Adsorbing other guest species can furthermore alter the nature of the adsorption sites or even create additional sites. These effects can result in an increase or decrease of guest uptake and, therefore, also influence selectivity. In the final step of our high-throughput screening, the mixture selectivity is calculated for the set of 3305 promising COF candidates obtained in the previous section for a 15:85 mix of CO_2:N_2. This allows us to determine the adsorption characteristics with increased accuracy and investigate the effects of competitive adsorption and coadsorption on the selectivity.

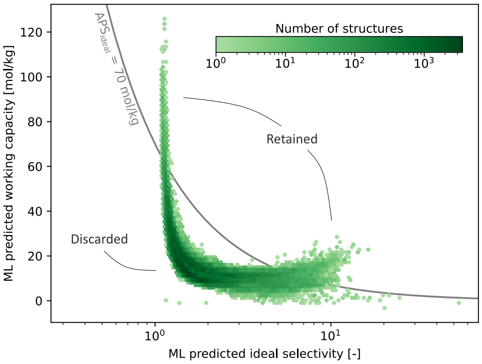


Figure 5. ML predicted CO_2 working capacity and ideal CO_2/N_2 selectivity for 268,678 COFs in the ReDD-COFFEE database. The threshold on the APS_{ideal} to retain structures for further screening is indicated in gray.

In Figure 6a, the ideal and mixture selectivities are compared with one another for the 3180 promising candidates for which all adsorption properties are successfully determined. For the large-pore structures, which combine a huge working capacity with a low selectivity, the influence of changing the inlet gas from a pure component to a mixture gas is negligible. Since there are only small interactions with the framework in these structures, both the ideal and mixture selectivities approach one. For the materials with a higher ideal selectivity, more significant corrections are observed. In case the ideal selectivity is higher than the mixture selectivity, the deviations are very small. However, the mixture selectivity can be multiple orders of magnitude higher than the ideal one. Therefore, the ideal selectivity acts as an approximate lower bound for the mixture selectivity. As such, the highest observed APS values are likely to increase even more when accounting for competitive adsorption and coadsorption. Most 3D COFs have a slightly lower mixture selectivity than ideal selectivity, and the majority of the 2D COFs have a higher mixture than ideal selectivity. These results show that the ideal selectivity is an acceptable initial guess for the separation performance of COFs, but changing the inlet gas to a mixture gas that mimics the experimental conditions significantly influences the precise adsorption characteristics.

Figure 6b visualizes the CO_2 working capacity and the CO_2/N_2 mixture selectivity. The regions with low working capacities and low selectivities are absent since these low-APS_{ideal} materials were filtered out in step 2. Similarly to Figure 6a, Figure 6b demonstrates that 2D COFs have, on average, a higher mixture selectivity but a lower spread. On the contrary, the mean mixture selectivity of 3D COFs might be lower, but due to the higher spread, the largest obtained mixture

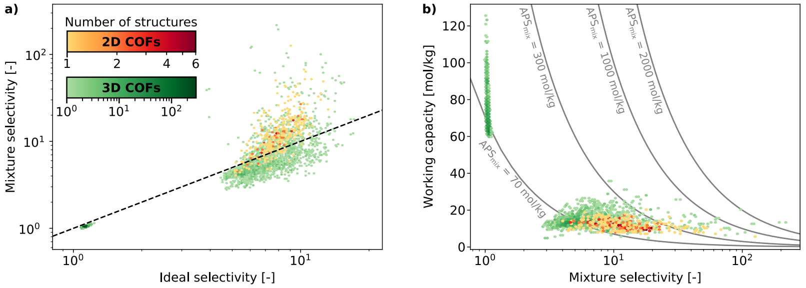


Figure 6. Analysis of screening the 3180 top-performing COFs in step 3. (a) Relationship between the ideal and the mixture selectivities. The dashed line indicates identical characteristics. (b) Relationship between the CO_2 working capacity and the mixture CO_2/N_2 selectivity. Several values of APS_{mix} are indicated in gray. In both plots, a distinction is made between the 2D (orange) and 3D (green) COFs.

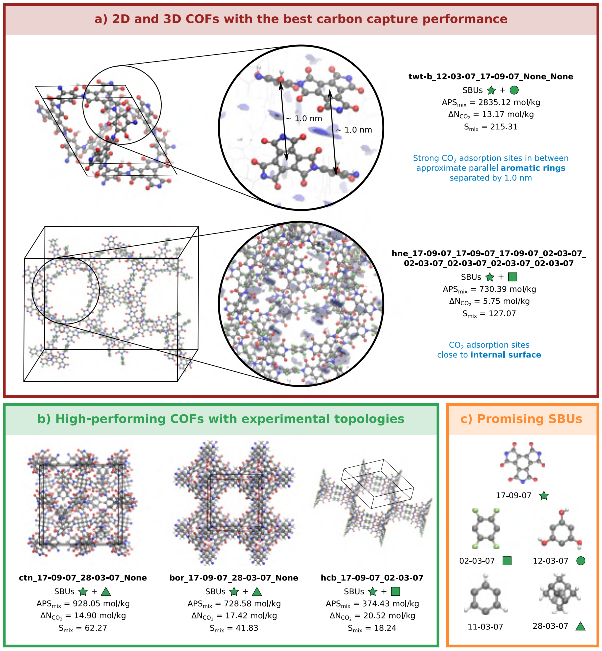


Figure 7. Overview of the best-performing COFs and candidates for experimental synthesis. (a) 2D and 3D COF with the highest APS_{mix} values, together with a detail of the distribution of CO_2 molecules at 10 bar. (b) Three materials of particular interest for experimental synthesis due to their experimentally observed topology. (c) Set of promising SBUs that constitute the materials in (a) and (b). Color code: hydrogen (white), carbon (gray), nitrogen (blue), oxygen (red), and fluorine (green). The CO_2 densities are visualized with blue isosurfaces using VMD[140] (isosurface values: 0.0108788 and 0.027197 for the best 3D COF and 0.0097033 and 0.02425825 for the best 2D COF).

selectivities are found in 3D COFs. The discrepancy in adsorption behavior between 2D and 3D COFs can be explained by the typically lower gravimetric accessible surface area in 2D COFs compared to 3D COFs, as visualized in the inset of Figure S14b of the SI. Hence, the interaction with the pore walls has a larger influence on the equilibrium separation performance of 3D COFs, resulting in a larger spread of the selectivity. The absence of the materials with a low APS_{ideal} value in Figure 6b highlights the two regimes that were discussed in Section 4.1. On the one hand are the structures with a very large pore diameter (>10 nm) that combine a high working capacity with a low selectivity around unity. Since the interactions of the guest molecules with the pore walls are very limited, these materials have only a negligible separation performance. The mixture selectivity of these materials is comparable to their ideal selectivity and therefore, also their APS does not change significantly. On the other hand, the mixture selectivity of the COFs that already had a large ideal selectivity can either decrease slightly or increase drastically. As such, the highest observed APS_{mix} values are much larger than the earlier observed APS_{ideal} values. The largest APS_{mix} amounts to 2835.1 mol/kg for the 3D imide COF twt-b_12−03−07_17−09−07_None_None, visualized in Figure 7a, which has a working capacity of 13.2 mol/kg and a mixture selectivity of 215.3. Whereas some COFs have a higher working capacity or selectivity, this material is unique because it achieves high values for both properties simultaneously. The best-performing 2D COF is the imide COF hne_17−09−07_17−09−07_17−09−07_02−03−07_02−03−07_02−03−07_02−03−07_02−03−07, again visualized in Figure 7a. This COF has an APS_{mix} of 730.4 mol/kg, resulting from a working capacity of 5.7 mol/kg and a mixture selectivity of 127.1. As can be observed in Tables S10 and S11 of the SI, both the best-performing 3D and 2D COFs suffered from much lower ideal selectivities of 7.8 and 9.0, respectively, demonstrating the need for mixture GCMC calculations to rank the best-performing COFs for carbon capture.

4.4. Design Strategies for Top-Performing COFs. The 20 structures with the highest APS_{mix} among the 2D and 3D COFs are specified in Section S5.1 of the SI. For these materials, additional evaluation metrics, such as purity and regeneration, are determined in Section S5.3 of the SI, further demonstrating their excellent performance. Among the 3D COFs, 15 out of the top-20 structures possess an imide linkage, confirming our hypothesis in Section 4.1 that materials containing this linkage type can reach high APS values. The most frequently observed building block is a phenyl ring connected to three other SBUs, as exemplified by SBUs 11−03−07 and 17−09−07 in Figure 7c, which are present in 4 and 14 of the top-20 materials, respectively. These promising building blocks are also disproportionally abundant in the top-20 2D and 3D COFs compared with the whole ReDD-COFFEE database, as demonstrated in Table S12.

When examining the distribution of CO_2 molecules within the best-performing 3D COFs, the adsorption sites with the highest CO_2 density are located between aromatic rings that are approximately parallel and have a separation of around 1.0 nm. This is illustrated in Figure 7a for the COF with the highest APS_{mix}, and in Figures S23−S34 of the SI for the other top-6 materials. The main driving force for the selectivity of this adsorption site is the favorable interaction between the delocalized aromatic system and the quadrupolar moment of CO_2.[139] The presence of these high-density pockets also explains why COFs with a pore diameter of around 1.0 nm have the highest APS, as already observed in Figure 1a. A similar conclusion was drawn by Boyd et al. in their high-throughput screening of metal−organic frameworks for carbon capture. They observed that the strongest CO_2 adsorption sites were formed by parallel aromatic rings separated by 0.7 nm.[65]

Due to their layered nature, these strong adsorption sites are usually absent in 2D COFs. The aromatic rings in these materials are mainly oriented parallel to the layers, which have an interlayer distance of around 0.3−0.4 nm, much smaller than 1.0 nm. This explains why the APS_{mix} values of 2D COFs are smaller than those of 3D COFs. The top-20 best-performing 2D COFs are all held together by an imide linkage. Furthermore, two SBUs, visualized in Figure 7c, appear frequently. The imide-terminated and fluoro-functionalized phenyl rings 17−09−07 and 02−03−07 are observed in 11 and 10 of the top-20 2D COFs, respectively. Moreover, the five best 2D COFs are built solely from these building blocks. Clearly, the chemistry of the framework environment becomes more critical in the 2D COFs. This can also be observed in the CO_2 distribution of the 2D COF with the highest APS_{mix} value in Figure 7a, which is mainly concentrated close to the pore walls. Although the effect of the fluoro group can only be compared with the limited number of other functional groups present in the ReDD-COFFEE database, such as methyl, *tert*-butyl, and hydroxyl, its increased CO_2 affinity has already been demonstrated in triazine-linked COFs.[111,141,142]

Whereas high-throughput screenings of hypothetical databases are very useful in predicting general trends and finding performance limits of material classes, the best-performing material might be very difficult or even impossible to synthesize. Even though the materials with a low synthetic likelihood are filtered out of the ReDD-COFFEE database by construction,[8] it might still be unfeasible to synthesize most of its materials. Interestingly, there are three materials within the top-20 of 2D and 3D COFs of which the topologies are experimentally observed in COFs, i.e., the **ctn**, **bor**, and **hcb** topologies,[4,24] shown in Figure 7b. Also, their building blocks have already been experimentally adopted to synthesize COFs.[143−145] Whereas these COFs are not yet synthesized, we hope that they can follow the example of the hypothetical COFs synthesized by Lan et al.[40] They form the prime candidates for experimental synthesis and validation since potential synthesis routes are readily available and can be found in Figure S36 of the SI.

5. CONCLUSIONS

In this work, the diverse ReDD-COFFEE database is screened for adopting COFs as adsorbents in postcombustion carbon capture using a pressure swing adsorption process. After an initial screening of a diverse subset of 14,970 materials, an ML algorithm is trained to accurately predict the CO_2 working capacity and the ideal CO_2/N_2 selectivity of the 268,687 materials from the ReDD-COFFEE database. This approach enables a drastic speed-up of the screening procedure, requiring only a fraction of the computational resources that would be needed to characterize the full database with GCMC. From these results, a set of 3305 promising COFs are retained for which the influence of competitive adsorption and coadsorption is investigated with mixture GCMC calculations. This reveals that the ideal selectivity provides a reliable prediction of the mixture selectivity for low selectivities. However, for high values, the mixture selectivity can be 2 orders of magnitude larger than the ideal one. The 20 best-performing 2D and 3D COFs are identified to specify design rules.

The initial screening demonstrates that both a high working capacity and a high ideal selectivity can be obtained for COFs with pore diameters of around 1.0 nm. This is explained by strong CO_2 adsorption sites in between approximately parallel aromatic rings separated by 1.0 nm, which are frequently observed in the best-performing 3D COFs. Due to the layered nature of 2D COFs, these strong adsorption sites are mostly absent in these materials. In 2D COFs, the framework chemistry, such as linkage type and functionalization, plays a dominant role. Among the studied functional groups, fluoro substituents are frequently present among the top-performing 2D COFs. For both 2D and 3D materials, the CO_2 affinity of structures possessing an imide, (acyl)hydrazone, or (keto)-enamine linkage is high. Whereas ReDD-COFFEE is a database of hypothetical COFs, three materials within the 20 best-performing 2D and 3D materials assemble in an experimentally observed topology. These are prime candidates for experimental validation due to the availability of a synthetic route.

Our best-performing ML algorithm accurately predicted the individual guest uptakes, the CO_2 working capacity, and the ideal CO_2/N_2 selectivity using only a fraction of the required computational resources compared to explicitly performing all GCMC calculations. Since the chemical environment of the framework is more difficult to describe in terms of discrete features as compared to the pore geometry, the uptake of N_2 molecules can be predicted more accurately than that of CO_2 molecules, which interact more strongly with the pore walls. Furthermore, the CO_2 uptake is characterized more precisely at adsorption than at desorption conditions due to the increased importance of the framework chemistry in this low-pressure regime. Whereas a single ML algorithm, namely, gradient boosting regression, most accurately describes the ideal selectivity, the model with the lowest MAE for the working capacity is derived from individual ML algorithms that predict the uptake of CO_2 at adsorption and desorption conditions. Especially the small desorption uptake, which forms only a minor correction on the working capacity, is described more accurately by an ML model explicitly trained for this purpose.

With this high-throughput screening, we have identified trends in the COF material space and determined design rules to guide experimental researchers in finding high-performing materials for the application of postcombustion carbon capture. We hope that the atomic insights into the adsorption behavior of COFs learned from this work may accelerate the adoption of COFs as CO_2 adsorbents in the flue gases of industrial power plants, which would accomplish a significant step forward to drastically reduce our anthropogenic CO_2 emissions.

■ ASSOCIATED CONTENT

Data Availability Statement

The relevant input files and computational data which generated the results of this work are available at 10.5281/zenodo.10701114.

SI Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.chemmater.3c03230.

> Selection of the appropriate force fields, justification of the computational methodology, additional analysis of the diverse subset and its screening, detailed workflow of the machine learning algorithm to predict the working capacity and ideal selectivity, and analysis of the top-performing COFs (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Sven M. J. Rogge − *Center for Molecular Modeling (CMM), Ghent University, Zwijnaarde 9052, Belgium;* orcid.org/0000-0003-4493-5708; Email: Sven.Rogge@UGent.be

Veronique Van Speybroeck − *Center for Molecular Modeling (CMM), Ghent University, Zwijnaarde 9052, Belgium;* orcid.org/0000-0003-2206-178X; Email: Veronique.VanSpeybroeck@UGent.be

Authors

Juul S. De Vos − *Center for Molecular Modeling (CMM), Ghent University, Zwijnaarde 9052, Belgium;* orcid.org/0000-0002-2733-5762

Siddharth Ravichandran − *Center for Molecular Modeling (CMM), Ghent University, Zwijnaarde 9052, Belgium;* orcid.org/0000-0001-8069-1099

Sander Borgmans − *Center for Molecular Modeling (CMM), Ghent University, Zwijnaarde 9052, Belgium;* orcid.org/0000-0002-9187-4293

Louis Vanduyfhuys − *Center for Molecular Modeling (CMM), Ghent University, Zwijnaarde 9052, Belgium;* orcid.org/0000-0001-6747-3388

Pascal Van Der Voort − *Centre for Ordered Materials, Organometallics and Catalysis (COMOC), Ghent University, Ghent 9000, Belgium;* orcid.org/0000-0002-1248-479X

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chemmater.3c03230

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work is supported by the Research Board of Ghent University (BOF) through a Concerted Research Action (GOA010-17). J.S.D.V. and S.M.J.R. acknowledge the Research Foundation Flanders (FWO) for a strategic basic (SB) research fellowship (grant no. 1S94521N) and a senior postdoctoral fellowship (grant no. 12T3522N), respectively. S.R. acknowledges the financial support from the Flemish Government and Flanders Innovation and Entrepreneurship (VLAIO) through the Moonshot project MOONRISE (grant no. HBC.2020.2612). L.V., S.M.J.R., and V.V.S. acknowledge the Research Board of Ghent University (BOF). The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by VSC (Flemish Supercomputer Center), funded by Ghent University, FWO, and the Flemish Government—department EWI.

■ REFERENCES

(1) Canadell, J.; Monteiro, P.; Costa, M.; Cotrim da Cunha, L.; Cox, P.; Eliseev, A.; Henson, S.; Ishii, M.; Jaccard, S.; Koven, C.; Lohila, A.; Patra, P.; Piao, S.; Rogelj, J.; Syampungani, S.; Zaehle, S.; Zickfeld, K. Climate change 2021: the physical science basis. In *Contribution of working group I to the sixth assessment report of the Intergovernmental Panel on Climate Change*; Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B., Eds.; Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 2021; pp. 673−816.

(2) Haszeldine, R. S. Carbon capture and storage: how green can black be? *Science* **2009**, *325*, 1647−1652.

(3) D'Alessandro, D. M.; Smit, B.; Long, J. R. Carbon dioxide capture: prospects for new materials. *Angew. Chem., Int. Ed.* **2010**, *49*, 6058−6082.

(4) Côté, A. P.; Benin, A. I.; Ockwig, N. W.; O'Keeffe, M.; Matzger, A. J.; Yaghi, O. M. Porous, crystalline, covalent organic frameworks. *Science* **2005**, *310*, 1166−1170.

(5) Zeng, Y.; Zou, R.; Zhao, Y. Covalent organic frameworks for CO_2 capture. *Adv. Mater.* **2016**, *28*, 2855−2873.

(6) Yaghi, O. M.; O'Keeffe, M.; Ockwig, N. W.; Chae, H. K.; Eddaoudi, M.; Kim, J. Reticular synthesis and the design of new materials. *Nature* **2003**, *423*, 705−714.

(7) Eddaoudi, M.; Moler, D. B.; Li, H.; Chen, B.; Reineke, T. M.; O'Keeffe, M.; Yaghi, O. M. Modular chemistry: secondary building units as a basis for the design of highly porous and robust metal−organic carboxylate frameworks. *Acc. Chem. Res.* **2001**, *34*, 319−330.

(8) De Vos, J. S.; Borgmans, S.; Van Der Voort, P.; Rogge, S. M. J.; Van Speybroeck, V. ReDD-COFFEE: a ready-to-use database of covalent organic framework structures and accurate force fields to enable high-throughput screenings. *J. Mater. Chem. A* **2023**, *11*, 7468−7487.

(9) Seneviratne, S.; Zhang, X.; Adnan, M.; Badi, W.; Dereczynski, C.; Di Luca, A.; Ghosh, S.; Iskandar, I.; Kossin, J.; Lewis, S.; Otto, F.; Pinto, I.; Satoh, M.; Vicente-Serrano, S.; Wehner, M.; Zhou, B. Climate change 2021: the physical science basis. In *Contribution of working group I to the sixth assessment report of the Intergovernmental Panel on Climate Change*; Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B., Eds.; Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 2021; pp. 673−816.

(10) Yang, H.; Xu, Z.; Fan, M.; Gupta, R.; Slimane, R. B.; Bland, A. E.; Wright, I. Progress in carbon dioxide separation and capture: a review. *J. Environ. Sci.* **2008**, *20*, 14−27.

(11) Smit, B.; Reimer, J. A.; Oldenburg, C. M.; Bourg, I. C. *Introduction to carbon capture and sequestration*; Imperial College Press, 2014; Vol. *1*.

(12) Markewitz, P.; Kuckshinrichs, W.; Leitner, W.; Linssen, J.; Zapp, P.; Bongartz, R.; Schreiber, A.; Müller, T. E. Worldwide innovations in the development of carbon capture technologies and the utilization of CO_2. *Energy Environ. Sci.* **2012**, *5*, 7281−7305.

(13) Joos, L.; Huck, J. M.; Van Speybroeck, V.; Smit, B. Cutting the cost of carbon capture: a case for carbon capture and utilization. *Faraday Discuss.* **2016**, *192*, 391−414.

(14) Weiland, R. H.; Dingman, J. C.; Cronin, D. B. Heat capacity of aqueous monoethanolamine, diethanolamine, N-methyldiethanolamine, and N-methyldiethanolamine-based blends with carbon dioxide. *J. Chem. Eng. Data* **1997**, *42*, 1004−1006.

(15) Rochelle, G. T. Amine scrubbing for CO_2 capture. *Science* **2009**, *325*, 1652−1654.

(16) Li, J.-R.; Kuppler, R. J.; Zhou, H.-C. Selective gas adsorption and separation in metal-organic frameworks. *Chem. Soc. Rev.* **2009**, *38*, 1477−1504.

(17) Farmahini, A. H.; Krishnamurthy, S.; Friedrich, D.; Brandani, S.; Sarkisov, L. Performance-based screening of porous materials for carbon capture. *Chem. Rev.* **2021**, *121*, 10666−10741.

(18) Ben-Mansour, R.; Habib, M.; Bamidele, O.; Basha, M.; Qasem, N.; Peedikakkal, A.; Laoui, T.; Ali, M. Carbon capture by physical adsorption: materials, experimental investigations and numerical modeling and simulations − a review. *Appl. Energy* **2016**, *161*, 225−255.

(19) Lu, X.; Jin, D.; Wei, S.; Wang, Z.; An, C.; Guo, W. Strategies to enhance CO_2 capture and separation based on engineering absorbent materials. *J. Mater. Chem. A* **2015**, *3*, 12118−12132.

(20) Férey, G.; Serre, C.; Devic, T.; Maurin, G.; Jobic, H.; Llewellyn, P. L.; De Weireld, G.; Vimont, A.; Daturi, M.; Chang, J.-S. Why hybrid porous solids capture greenhouse gases? *Chem. Soc. Rev.* **2011**, *40*, 550−562.

(21) Choi, S.; Drese, J. H.; Jones, C. W. Adsorbent materials for carbon dioxide capture from large anthropogenic point sources. *ChemSusChem* **2009**, *2*, 796−854.

(22) Bae, T.-H.; Hudson, M. R.; Mason, J. A.; Queen, W. L.; Dutton, J. J.; Sumida, K.; Micklash, K. J.; Kaye, S. S.; Brown, C. M.; Long, J. R. Evaluation of cation-exchanged zeolite adsorbents for post-combustion carbon dioxide capture. *Energy Environ. Sci.* **2013**, *6*, 128−138.

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

(24) El-Kaderi, H. M.; Hunt, J. R.; Mendoza-Cortés, J. L.; Côté, A. P.; Taylor, R. E.; O'Keeffe, M.; Yaghi, O. M. Designed synthesis of 3D covalent organic frameworks. *Science* **2007**, *316*, 268−272.

(25) Diercks, C. S.; Yaghi, O. M. The atom, the molecule, and the covalent organic framework. *Science* **2017**, *355*, No. eaal1585.

(26) Choi, Y. J.; Choi, J. H.; Choi, K. M.; Kang, J. K. Covalent organic frameworks for extremely high reversible CO_2 uptake capacity: a theoretical approach. *J. Mater. Chem.* **2011**, *21*, 1073−1078.

(27) Olajire, A. A. Recent advances in the synthesis of covalent organic frameworks for CO_2 capture. *J. CO2 Util.* **2017**, *17*, 137−161.

(28) Wang, Y.; Kang, C.; Zhang, Z.; Usadi, A. K.; Calabro, D. C.; Baugh, L. S.; Yuan, Y. D.; Zhao, D. Evaluation of schiff-base covalent organic frameworks for CO_2 capture: structure−performance relationships, stability, and performance under wet conditions. *ACS Sustainable Chem. Eng.* **2022**, *10*, 332−341.

(29) Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. Computation-ready, experimental metal−organic frameworks: a tool to enable high-throughput screening of nanoporous crystals. *Chem. Mater.* **2014**, *26*, 6185−6192.

(30) Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; Slater, B.; Siepmann, J. I.; Sholl, D. S.; Snurr, R. Q. Advances, updates, and analytics for the computation-ready, experimental metal−organic framework database: CoRE MOF 2019. *J. Chem. Eng. Data* **2019**, *64*, 5985−5998.

(31) Moghadam, P. Z.; Li, A.; Wiggin, S. B.; Tao, A.; Maloney, A. G. P.; Wood, P. A.; Ward, S. C.; Fairen-Jimenez, D. Development of a Cambridge structural database subset: a collection of metal−organic frameworks for past, present, and future. *Chem. Mater.* **2017**, *29*, 2618−2625.

(32) Rosen, A. S.; Iyer, S. M.; Ray, D.; Yao, Z.; Aspuru-Guzik, A.; Gagliardi, L.; Notestein, J. M.; Snurr, R. Q. Machine learning the quantum-chemical properties of metal-organic frameworks for accelerated materials discovery. *Matter* **2021**, *4*, 1578−1597.

(33) Ongari, D.; Yakutovich, A. V.; Talirz, L.; Smit, B. Building a consistent and reproducible database for adsorption evaluation in covalent−organic frameworks. *ACS Cent. Sci.* **2019**, *5*, 1663−1675.

(34) Tong, M.; Lan, Y.; Yang, Q.; Zhong, C. Exploring the structure-property relationships of covalent organic frameworks for noble gas separations. *Chem. Eng. Sci.* **2017**, *168*, 456−464.

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

(36) Colón, Y. J.; Gómez-Gualdrón, D. A.; Snurr, R. Q. Topologically guided, automated construction of metal−organic frameworks and their evaluation for energy-related applications. *Cryst. Growth Des.* **2017**, *17*, 5801−5810.

(37) Boyd, P. G.; Woo, T. K. A generalized method for constructing hypothetical nanoporous materials of any net topology from graph theory. *CrystEngComm* **2016**, *18*, 3777−3792.

(38) Martin, R. L.; Simon, C. M.; Medasani, B.; Britt, D. K.; Smit, B.; Haranczyk, M. In silico design of three-dimensional porous covalent organic frameworks via known synthesis routes and commercially available species. *J. Phys. Chem. C* **2014**, *118*, 23790−23802.

(39) Mercado, R.; Fu, R.-S.; Yakutovich, A. V.; Talirz, L.; Haranczyk, M.; Smit, B. In silico design of 2D and 3D covalent organic frameworks for methane storage applications. *Chem. Mater.* **2018**, *30*, 5069−5086.

(40) Lan, Y.; Han, X.; Tong, M.; Huang, H.; Yang, Q.; Liu, D.; Zhao, X.; Zhong, C. Materials genomics methods for high-throughput construction of COFs and targeted synthesis. *Nat. Commun.* **2018**, *9*, No. 5274.

(41) Colón, Y. J.; Snurr, R. Q. High-throughput computational screening of metal−organic frameworks. *Chem. Soc. Rev.* **2014**, *43*, 5735−5749.

(42) Ongari, D.; Talirz, L.; Smit, B. Too many materials and too many applications: an experimental problem waiting for a computational solution. *ACS Cent. Sci.* **2020**, *6*, 1890−1900.

(43) Moosavi, S. M.; Nandy, A.; Jablonka, K. M.; Ongari, D.; Janet, J. P.; Boyd, P. G.; Lee, Y.; Smit, B.; Kulik, H. J. Understanding the diversity of the metal-organic framework ecosystem. *Nat. Commun.* **2020**, *11*, No. 4068.

(44) Daglar, H.; Keskin, S. Recent advances, opportunities, and challenges in high-throughput computational screening of MOFs for gas separations. *Coord. Chem. Rev.* **2020**, *422*, No. 213470.

(45) Anderson, R.; Gómez-Gualdrón, D. A. Increasing topological diversity during computational "synthesis" of porous crystals: how and why. *CrystEngComm* **2019**, *21*, 1653−1665.

(46) Simon, C. M.; Kim, J.; Gomez-Gualdron, D. A.; Camp, J. S.; Chung, Y. G.; Martin, R. L.; Mercado, R.; Deem, M. W.; Gunter, D.; Haranczyk, M.; Sholl, D. S.; Snurr, R. Q.; Smit, B. The materials genome in action: identifying the performance limits for methane storage. *Energy Environ. Sci.* **2015**, *8*, 1190−1199.

(47) Tong, M.; Lan, Y.; Qin, Z.; Zhong, C. Computation-ready, experimental covalent organic framework for methane delivery: screening and material design. *J. Phys. Chem. C* **2018**, *122*, 13009−13016.

(48) Colón, Y. J.; Fairen-Jimenez, D.; Wilmer, C. E.; Snurr, R. Q. High-throughput screening of porous crystalline materials for hydrogen storage capacity near room temperature. *J. Phys. Chem. C* **2014**, *118*, 5383−5389.

(49) Goldsmith, J.; Wong-Foy, A. G.; Cafarella, M. J.; Siegel, D. J. Theoretical limits of hydrogen storage in metal−organic frameworks: opportunities and trade-offs. *Chem. Mater.* **2013**, *25*, 3373−3382.

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

(51) Bobbitt, N. S.; Chen, J.; Snurr, R. Q. High-throughput screening of metal−organic frameworks for hydrogen storage at cryogenic temperature. *J. Phys. Chem. C* **2016**, *120*, 27328−27341.

(52) Tong, M.; Zhu, W.; Li, J.; Long, Z.; Zhao, S.; Chen, G.; Lan, Y. An easy way to identify high performing covalent organic frameworks for hydrogen storage. *Chem. Commun.* **2020**, *56*, 6376−6379.

(53) Wilmer, C. E.; Farha, O. K.; Bae, Y.-S.; Hupp, J. T.; Snurr, R. Q. Structure−property relationships of porous materials for carbon dioxide separation and capture. *Energy Environ. Sci.* **2012**, *5*, 9849−9856.

(54) Qiao, Z.; Peng, C.; Zhou, J.; Jiang, J. High-throughput computational screening of 137953 metal−organic frameworks for membrane separation of a $CO_2/N_2/CH_4$ mixture. *J. Mater. Chem. A* **2016**, *4*, 15904−15912.

(55) Qiao, Z.; Zhang, K.; Jiang, J. In silico screening of 4764 computation-ready, experimental metal−organic frameworks for CO_2 separation. *J. Mater. Chem. A* **2016**, *4*, 2105−2114.

(56) Altintas, C.; Avci, G.; Daglar, H.; Nemati Vesali Azar, A.; Velioglu, S.; Erucar, I.; Keskin, S. Database for CO_2 separation performances of MOFs based on computational materials screening. *ACS Appl. Mater. Interfaces* **2018**, *10*, 17257−17268.

(57) Tong, M.; Yang, Q.; Zhong, C. Computational screening of covalent organic frameworks for CH_4/H_2, CO_2/H_2 and CO_2/CH_4 separations. *Microporous Mesoporous Mater.* **2015**, *210*, 142−148.

(58) Yan, T.; Lan, Y.; Tong, M.; Zhong, C. Screening and design of covalent organic framework membranes for CO_2/CH_4 separation. *ACS Sustainable Chem. Eng.* **2019**, *7*, 1220−1227.

(59) Altundal, O. F.; Haslak, Z. P.; Keskin, S. Combined GCMC, MD, and DFT approach for unlocking the performances of COFs for methane purification. *Ind. Eng. Chem. Res.* **2021**, *60*, 12999−13012.

(60) Altintas, C.; Erucar, I.; Keskin, S. High-throughput computational screening of the metal organic framework database for CH_4/H_2 separations. *ACS Appl. Mater. Interfaces* **2018**, *10*, 3668−3679.

(61) Basdogan, Y.; Sezginel, K. B.; Keskin, S. Identifying highly selective metal organic frameworks for CH_4/H_2 separations using computational tools. *Ind. Eng. Chem. Res.* **2015**, *54*, 8479−8491.

(62) Avci, G.; Velioglu, S.; Keskin, S. High-throughput screening of MOF adsorbents and membranes for H_2 purification and CO_2 capture. *ACS Appl. Mater. Interfaces* **2018**, *10*, 33693−33706.

(63) Aksu, G. O.; Daglar, H.; Altintas, C.; Keskin, S. Computational selection of high-performing covalent organic frameworks for adsorption and membrane-based CO_2/H_2 separation. *J. Phys. Chem. C* **2020**, *124*, 22577−22590.

(64) Aksu, G. O.; Erucar, I.; Haslak, Z. P.; Keskin, S. Accelerating discovery of COFs for CO_2 capture and H_2 purification using structurally guided computational screening. *Chem. Eng. J.* **2022**, *427*, No. 131574.

(65) Boyd, P. G.; Chidambaram, A.; García-Díez, E.; Ireland, C. P.; Daff, T. D.; Bounds, R.; Gładysiak, A.; Schouwink, P.; Moosavi, S. M.; Maroto-Valer, M. M.; Reimer, J. A.; Navarro, J. A. R.; Woo, T. K.; Garcia, S.; Stylianou, K. C.; Smit, B. Data-driven design of metal−organic frameworks for wet flue gas CO_2 capture. *Nature* **2019**, *576*, 253−256.

(66) Li, S.; Chung, Y. G.; Snurr, R. Q. High-throughput screening of metal−organic frameworks for CO_2 capture in the presence of water. *Langmuir* **2016**, *32*, 10368−10376.

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

(68) Haldoupis, E.; Nair, S.; Sholl, D. S. Finding MOFs for highly selective CO_2/N_2 adsorption using materials screening based on efficient assignment of atomic point charges. *J. Am. Chem. Soc.* **2012**, *134*, 4313−4323.

(69) Watanabe, T.; Sholl, D. S. Accelerating applications of metal−organic frameworks for gas adsorption and separation by computational screening of materials. *Langmuir* **2012**, *28*, 14114−14128.

(70) Daglar, H.; Keskin, S. Computational screening of metal−organic frameworks for membrane-based $CO_2/N_2/H_2O$ separations: best materials for flue gas separation. *J. Phys. Chem. C* **2018**, *122*, 17347−17357.

(71) Wu, D.; Yang, Q.; Zhong, C.; Liu, D.; Huang, H.; Zhang, W.; Maurin, G. Revealing the structure−property relationships of metal−organic frameworks for CO_2 capture from flue gas. *Langmuir* **2012**, *28*, 12094−12099.

(72) Tong, M.; Yang, Q.; Xiao, Y.; Zhong, C. Revealing the structure−property relationship of covalent organic frameworks for CO_2 capture from postcombustion gas: a multi-scale computational study. *Phys. Chem. Chem. Phys.* **2014**, *16*, 15189−15198.

(73) Deeg, K. S.; Damasceno Borges, D.; Ongari, D.; Rampal, N.; Talirz, L.; Yakutovich, A. V.; Huck, J. M.; Smit, B. In silico discovery of covalent organic frameworks for carbon capture. *ACS Appl. Mater. Interfaces* **2020**, *12*, 21559−21568.

(74) Altundal, O. F.; Altintas, C.; Keskin, S. Can COFs replace MOFs in flue gas separation? High-throughput computational screening of COFs for CO_2/N_2 separation. *J. Mater. Chem. A* **2020**, *8*, 14609−14623.

(75) Moghadam, P. Z.; Rogge, S. M. J.; Li, A.; Chow, C.-M.; Wieme, J.; Moharrami, N.; Aragones-Anglada, M.; Conduit, G.; Gomez-Gualdron, D. A.; Van Speybroeck, V.; Fairen-Jimenez, D. Structure-mechanical stability relations of metal-organic frameworks via machine learning. *Matter* **2019**, *1*, 219−234.

(76) Islamov, M.; Babaei, H.; Anderson, R.; Sezginel, K. B.; Long, J. R.; McGaughey, A. J. H.; Gomez-Gualdron, D. A.; Wilmer, C. E. High-throughput screening of hypothetical metal−organic frameworks for thermal conductivity. *npj Comput. Mater.* **2023**, *9*, No. 11.

(77) Rosen, A. S.; Notestein, J. M.; Snurr, R. Q. Identifying promising metal−organic frameworks for heterogeneous catalysis via high-throughput periodic density functional theory. *J. Comput. Chem.* **2019**, *40*, 1305−1318.

(78) Rosen, A. S.; Fung, V.; Huck, P.; O'Donnell, C. T.; Horton, M. K.; Truhlar, D. G.; Persson, K. A.; Notestein, J. M.; Snurr, R. Q. High-

throughput predictions of metal−organic framework electronic properties: theoretical challenges, graph neural networks, and data exploration. *npj Comput. Mater.* **2022**, *8*, No. 112.

(79) He, Y.; Cubuk, E. D.; Allendorf, M. D.; Reed, E. J. Metallic metal−organic frameworks predicted by the combination of machine learning methods and ab initio calculations. *J. Phys. Chem. Lett.* **2018**, *9*, 4562−4569.

(80) Lee, S.; Kim, B.; Cho, H.; Lee, H.; Lee, S. Y.; Cho, E. S.; Kim, J. Computational screening of trillions of metal−organic frameworks for high-performance methane storage. *ACS Appl. Mater. Interfaces* **2021**, *13*, 23647−23654.

(81) Jablonka, K. M.; Ongari, D.; Moosavi, S. M.; Smit, B. Big-data science in porous materials: materials genomics and machine learning. *Chem. Rev.* **2020**, *120*, 8066−8129.

(82) Demir, H.; Daglar, H.; Gulbalkan, H. C.; Aksu, G. O.; Keskin, S. Recent advances in computational modeling of MOFs: From molecular simulations to machine learning. *Coord. Chem. Rev.* **2023**, *484*, No. 215112.

(83) Bobbitt, N. S.; Snurr, R. Q. Molecular modelling and machine learning for high-throughput screening of metal−organic frameworks for hydrogen storage. *Mol. Simul.* **2019**, *45*, 1069−1081.

(84) Bucior, B. J.; Bobbitt, N. S.; Islamoglu, T.; Goswami, S.; Gopalan, A.; Yildirim, T.; Farha, O. K.; Bagheri, N.; Snurr, R. Q. Energy-based descriptors to rapidly predict hydrogen storage in metal−organic frameworks. *Mol. Syst. Des. Eng.* **2019**, *4*, 162−174.

(85) Anderson, G.; Schweitzer, B.; Anderson, R.; Gómez-Gualdrón, D. A. Attainable volumetric targets for adsorption-based hydrogen storage in porous crystals: molecular simulation and machine learning. *J. Phys. Chem. C* **2019**, *123*, 120−130.

(86) Thornton, A. W.; Simon, C. M.; Kim, J.; Kwon, O.; Deeg, K. S.; Konstas, K.; Pas, S. J.; Hill, M. R.; Winkler, D. A.; Haranczyk, M.; Smit, B. Materials genome in action: identifying the performance limits of physical hydrogen storage. *Chem. Mater.* **2017**, *29*, 2844−2854.

(87) Fernandez, M.; Woo, T. K.; Wilmer, C. E.; Snurr, R. Q. Large-scale quantitative structure−property relationship (QSPR) analysis of methane storage in metal−organic frameworks. *J. Phys. Chem. C* **2013**, *117*, 7681−7689.

(88) Pardakhti, M.; Moharreri, E.; Wanik, D.; Suib, S. L.; Srivastava, R. Machine learning using combined structural and chemical descriptors for prediction of methane adsorption performance of metal organic frameworks (MOFs). *ACS Comb. Sci.* **2017**, *19*, 640−645.

(89) Fanourgakis, G. S.; Gkagkas, K.; Tylianakis, E.; Froudakis, G. Fast screening of large databases for top performing nanomaterials using a self-consistent, machine learning based approach. *J. Phys. Chem. C* **2020**, *124*, 19639−19648.

(90) Pardakhti, M.; Nanda, P.; Srivastava, R. Impact of chemical features on methane adsorption by porous materials at varying pressures. *J. Phys. Chem. C* **2020**, *124*, 4534−4544.

(91) Fernandez, M.; Boyd, P. G.; Daff, T. D.; Aghaji, M. Z.; Woo, T. K. Rapid and accurate machine learning recognition of high performing metal organic frameworks for CO_2 capture. *J. Phys. Chem. Lett.* **2014**, *5*, 3056−3060.

(92) Anderson, R.; Rodgers, J.; Argueta, E.; Biong, A.; Gómez-Gualdrón, D. A. Role of pore chemistry and topology in the CO_2 capture capabilities of MOFs: from molecular simulation to machine learning. *Chem. Mater.* **2018**, *30*, 6325−6337.

(93) Fernandez, M.; Trefiak, N. R.; Woo, T. K. Atomic property weighted radial distribution functions descriptors of metal−organic frameworks for the prediction of gas uptake capacity. *J. Phys. Chem. C* **2013**, *117*, 14095−14105.

(94) Dureckova, H.; Krykunov, M.; Aghaji, M. Z.; Woo, T. K. Robust machine learning models for predicting high CO_2 working capacity and CO_2/H_2 selectivity of gas adsorption in metal organic frameworks for precombustion carbon capture. *J. Phys. Chem. C* **2019**, *123*, 4133−4139.

(95) Chung, Y. G.; Gómez-Gualdrón, D. A.; Li, P.; Leperi, K. T.; Deria, P.; Zhang, H.; Vermeulen, N. A.; Stoddart, J. F.; You, F.; Hupp, J. T.; Farha, O. K.; Snurr, R. Q. In silico discovery of metal−organic frameworks for precombustion CO_2 capture using a genetic algorithm. *Sci. Adv.* **2016**, *2*, No. e1600909.

(96) Fernandez, M.; Barnard, A. S. Geometrical properties can predict CO_2 and N_2 adsorption performance of metal−organic frameworks (MOFs) at low pressure. *ACS Comb. Sci.* **2016**, *18*, 243−252.

(97) Yao, Z.; Sánchez-Lengeling, B.; Bobbitt, N. S.; Bucior, B. J.; Kumar, S. G. H.; Collins, S. P.; Burns, T.; Woo, T. K.; Farha, O. K.; Snurr, R. Q.; Aspuru-Guzik, A. Inverse design of nanoporous crystalline reticular materials with deep generative models. *Nat. Mach. Intell.* **2021**, *3*, 76−86.

(98) Aghaji, M. Z.; Fernandez, M.; Boyd, P. G.; Daff, T. D.; Woo, T. K. Quantitative structure-property relationship models for recognizing metal organic frameworks (MOFs) with high CO_2 working capacity and CO_2/CH_4 selectivity for methane purification. *Eur. J. Inorg. Chem.* **2016**, *2016*, 4505−4511.

(99) Aksu, G. O.; Keskin, S. Advancing CH_4/H_2 separation with covalent organic frameworks by combining molecular simulations and machine learning. *J. Mater. Chem. A* **2023**, *11*, 14788−14799.

(100) Lim, Y.; Park, J.; Lee, S.; Kim, J. Finely tuned inverse design of metal−organic frameworks with user-desired Xe/Kr selectivity. *J. Mater. Chem. A* **2021**, *9*, 21175−21183.

(101) Simon, C. M.; Mercado, R.; Schnell, S. K.; Smit, B.; Haranczyk, M. What are the best materials to separate a xenon/krypton mixture? *Chem. Mater.* **2015**, *27*, 4459−4475.

(102) Ren, E.; Coudert, F.-X. Enhancing gas separation selectivity prediction through geometrical and chemical Descriptors. *Chem. Mater.* **2023**, *35*, 6771−6781.

(103) Altintas, C.; Altundal, O. F.; Keskin, S.; Yildirim, R. Machine learning meets with metal organic frameworks for gas storage and separation. *J. Chem. Inf. Model.* **2021**, *61*, 2131−2146.

(104) Moosavi, S. M.; Chidambaram, A.; Talirz, L.; Haranczyk, M.; Stylianou, K. C.; Smit, B. Capturing chemical intuition in synthesis of metal-organic frameworks. *Nat. Commun.* **2019**, *10*, No. 539.

(105) Bai, P.; Jeon, M. Y.; Ren, L.; Knight, C.; Deem, M. W.; Tsapatsis, M.; Siepmann, J. I. Discovery of optimal zeolites for challenging separations and chemical transformations using predictive materials modeling. *Nat. Commun.* **2015**, *6*, 5912.

(106) Krishna, R.; van Baten, J. M. A comparison of the CO_2 capture characteristics of zeolites and metal−organic frameworks. *Sep. Purif. Technol.* **2012**, *87*, 120−126.

(107) Xu, D.; Zhang, J.; Li, G.; Xiao, P.; Webley, P.; Zhai, Y.-C. Effect of water vapor from power station flue gas on CO_2 capture by vacuum swing adsorption with activated carbon. *J. Fuel Chem. Technol.* **2011**, *39*, 169−174.

(108) Cavenati, S.; Grande, C. A.; Rodrigues, A. E. Adsorption equilibrium of methane, carbon dioxide, and nitrogen on zeolite 13X at high pressures. *J. Chem. Eng. Data* **2004**, *49*, 1095−1101.

(109) Saha, D.; Bao, Z.; Jia, F.; Deng, S. Adsorption of CO_2, CH_4, N_2O, and N_2 on MOF-5, MOF-177, and zeolite 5A. *Environ. Sci. Technol.* **2010**, *44*, 1820−1826.

(110) Krishna, R.; van Baten, J. M. In silico screening of metal−organic frameworks in separation applications. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10593−10616.

(111) Zhao, Y.; Yao, K. X.; Teng, B.; Zhang, T.; Han, Y. A perfluorinated covalent triazine-based framework for highly selective and water−tolerant CO_2 capture. *Energy Environ. Sci.* **2013**, *6*, 3684−3692.

(112) Wang, G.; Leus, K.; Zhao, S.; Van Der Voort, P. Newly designed covalent triazine framework based on novel N-heteroaromatic building blocks for efficient CO_2 and H_2 capture and storage. *ACS Appl. Mater. Interfaces* **2018**, *10*, 1244−1249.

(113) Jena, H. S.; Krishnaraj, C.; Wang, G.; Leus, K.; Schmidt, J.; Chaoui, N.; Van Der Voort, P. Acetylacetone covalent triazine framework: an efficient carbon capture and storage material and a highly stable heterogeneous catalyst. *Chem. Mater.* **2018**, *30*, 4102−4111.

(114) Wang, G.; Tahir, N.; Onyshchenko, I.; De Geyter, N.; Morent, R.; Leus, K.; Van Der Voort, P. Novel hexaazatrinaphthalene-based covalent triazine frameworks as high-performance platforms for efficient carbon capture and storage. *Microporous Mesoporous Mater.* **2019**, *290*, No. 109650.

(115) Babarao, R.; Jiang, J. Exceptionally high CO_2 storage in covalent-organic frameworks: atomistic simulation study. *Energy Environ. Sci.* **2008**, *1*, 139−143.

(116) Millward, A. R.; Yaghi, O. M. Metal-organic frameworks with exceptionally high capacity for storage of carbon dioxide at room temperature. *J. Am. Chem. Soc.* **2005**, *127*, 17998−17999.

(117) Rawat, K. S.; Borgmans, S.; Braeckevelt, T.; Stevens, C. V.; Van Der Voort, P.; Van Speybroeck, V. How the layer alignment in two-dimensional nanoporous covalent organic frameworks impacts its electronic properties. *ACS Appl. Nano Mater.* **2022**, *5*, 14377−14387.

(118) Borgmans, S.; Rogge, S. M. J.; Vanduyfhuys, L.; Van Speybroeck, V. OGRe: optimal grid refinement protocol for accurate free energy surfaces and its application in proton hopping in zeolites and 2D COF stacking. *J. Chem. Theory Comput.* **2023**, *19*, 9032−9048.

(119) Janet, J. P.; Kulik, H. J. Resolving transition metal chemical space: feature selection for machine learning and structure−property relationships. *J. Phys. Chem. A* **2017**, *121*, 8939−8954.

(120) Nandy, A.; Duan, C.; Janet, J. P.; Gugler, S.; Kulik, H. J. Strategies and software for machine learning accelerated discovery in transition metal chemistry. *Ind. Eng. Chem. Res.* **2018**, *57*, 13973−13986.

(121) Janet, J. P.; Liu, F.; Nandy, A.; Duan, C.; Yang, T.; Lin, S.; Kulik, H. J. Designing in the face of uncertainty: exploiting electronic structure and machine learning models for discovery in inorganic chemistry. *Inorg. Chem.* **2019**, *58*, 10592−10606.

(122) Nandy, A.; Zhu, J.; Janet, J. P.; Duan, C.; Getman, R. B.; Kulik, H. J. Machine learning accelerates the discovery of design rules and exceptions in stable metal−oxo intermediate formation. *ACS Catal.* **2019**, *9*, 8243−8255.

(123) Vandenbrande, S.; Verstraelen, T.; Gutiérrez-Sevillano, J. J.; Waroquier, M.; Van Speybroeck, V. Methane adsorption in Zr-based MOFs: comparison and critical evaluation of force fields. *J. Phys. Chem. C* **2017**, *121*, 25309−25322.

(124) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.* **1989**, *111*, 8551−8566.

(125) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.

(126) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* **1990**, *94*, 8897−8909.

(127) Potoff, J. J.; Siepmann, J. I. Vapor−liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen. *AIChE J.* **2001**, *47*, 1676−1682.

(128) Harris, J. G.; Yung, K. H. Carbon dioxide's liquid-vapor coexistence curve and critical properties as predicted by a simple molecular model. *J. Phys. Chem.* **1995**, *99*, 12021−12024.

(129) Furukawa, H.; Yaghi, O. M. Storage of hydrogen, methane, and carbon dioxide in highly porous covalent organic frameworks for clean energy applications. *J. Am. Chem. Soc.* **2009**, *131*, 8875−8883.

(130) Widom, B. Some topics in the theory of fluids. *J. Chem. Phys.* **1963**, *39*, 2808−2812.

(131) Vandenbrande, S.; Waroquier, M.; Van Speybroeck, V.; Verstraelen, T. Ab initio evaluation of Henry coefficients using importance sampling. *J. Chem. Theory Comput.* **2018**, *14*, 6359−6369.

(132) Dubbeldam, D.; Calero, S.; Ellis, D. E.; Snurr, R. Q. RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Mol. Simul.* **2016**, *42*, 81−101.

(133) Verstraelen, T.; Vandenbrande, S.; Heidar-Zadeh, F.; Vanduyfhuys, L.; Van Speybroeck, V.; Waroquier, M.; Ayers, P. W. Minimal basis iterative stockholder: atoms in molecules for force-field development. *J. Chem. Theory Comput.* **2016**, *12*, 3894−3912.

(134) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.

(135) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.* **2012**, *149*, 134−141.

(136) Ismail, A. F.; Khulbe, K. C.; Matsuura, T. *Gas separation membranes: polymeric and inorganic*; Springer International Publishing: Cham, 2015; pp. 11−35.

(137) Yıldız, Z.; Uzun, H. Prediction of gas storage capacities in metal organic frameworks using artificial neural network. *Microporous Mesoporous Mater.* **2015**, *208*, 50−54.

(138) Lundberg, S.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *arXiv* **2017**.

(139) Plonka, A. M.; Banerjee, D.; Woerner, W. R.; Zhang, Z.; Nijem, N.; Chabal, Y. J.; Li, J.; Parise, J. B. Mechanism of carbon dioxide adsorption in a highly selective coordination network supported by direct structural evidence. *Angew. Chem., Int. Ed.* **2013**, *52*, 1692−1695.

(140) Humphrey, W.; Dalke, A.; Schulten, K. VMD Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33−38.

(141) Wang, G.; Leus, K.; Jena, H. S.; Krishnaraj, C.; Zhao, S.; Depauw, H.; Tahir, N.; Liu, Y.-Y.; Van Der Voort, P. A fluorine-containing hydrophobic covalent triazine framework with excellent selective CO_2 capture performance. *J. Mater. Chem. A* **2018**, *6*, 6370−6375.

(142) Wang, G.; Onyshchenko, Y.; De Geyter, N.; Morent, R.; Leus, K.; Van Der Voort, P. Straightforward preparation of fluorinated covalent triazine frameworks with significantly enhanced carbon dioxide and hydrogen adsorption capacities. *Dalton Trans.* **2019**, *48*, 17612−17619.

(143) Veldhuizen, H.; Vasileiadis, A.; Wagemaker, M.; Mahon, T.; Mainali, D. P.; Zong, L.; van der Zwaag, S.; Nagai, A. Synthesis, characterization, and CO_2 uptake of mellitic triimide-based covalent organic frameworks. *J. Polym. Sci. Part A. Polym. Chem.* **2019**, *57*, 2373−2377.

(144) Fang, Q.; Wang, J.; Gu, S.; Kaspar, R. B.; Zhuang, Z.; Zheng, J.; Guo, H.; Qiu, S.; Yan, Y. 3D porous crystalline polyimide covalent organic frameworks for drug delivery. *J. Am. Chem. Soc.* **2015**, *137*, 8352−8355.

(145) Lee, J.; Lim, H.; Park, J.; Kim, M.-S.; Jung, J.-W.; Kim, J.; Kim, I.-D. Fluorine-rich covalent organic framework to boost electrochemical kinetics and storages of K^+ ions for potassium-ion battery. *Adv. Energy Mater.* **2023**, *13*, No. 2300442.

B

List of Publications

Updated July 2024

Publications in international peer–reviewed journals

1. **Quantifying the likelihood of structural models through a dynamically enhanced powder X-ray diffraction protocol**
   S. Borgmans, S.M.J. Rogge, J.S. De Vos, C.V. Stevens, P. Van der Voort, V. Van Speybroeck
   Angewandte Chemie International Edition, **2021**, 60 (16), 8913–8922
   IF: 16.823, Number of citations: 8

2. **Exploring the phase stability in interpenetrated diamondoid covalent organic frameworks**
   S. Borgmans, S.M.J. Rogge, J.S. De Vos, P. Van Der Voort, V. Van Speybroeck
   Communications Chemistry, **2023**, 6 (1), 5
   IF: 5.9, Number of citations: 7

3. **MicMec: Developing the Micromechanical Model to Investigate the Mechanics of Correlated Node Defects in UiO-66**
   J. Vandewalle, J.S. De Vos, S.M.J. Rogge
   Journal of Physical Chemistry C, **2023**, 127 (12), 6060–6070
   IF: 3.7, Number of citations: 1

4. **ReDD-COFFEE: A ready-to-use database of covalent organic framework structures and accurate force fields to enable high-throughput screenings**
   J.S. De Vos, S. Borgmans, P. Van Der Voort, S.M.J. Rogge, V. Van Speybroeck
   Journal of Materials Chemistry A, **2023**, 11 (14), 7468–7487
   IF: 11.9, Number of citations: 11

5. **Engineering of Phenylpyridine- and Bipyridine-Based Covalent Organic Frameworks for Photocatalytic Tandem Aerobic Oxidation/Povarov Cyclization**
   M. Debruyne, S. Borgmans, S. Radhakrishnan, E. Breynaert, H. Vrielinck, K. Leus, A. Laemont, J. De Vos, K.S. Rawat, S. Vanlommel, H. Rijckaert, H. Salemi, J. Everaert, F. Vanden Bussche, D. Poelman, R. Morent, N. De Geyter, P. Van Der Voort, V. Van Speybroeck, and C.V. Stevens
   ACS Applied Materials & Interfaces, **2023**, 15 (29), 35092–35106
   IF: 9.5, Number of citations: 5

6. **High-Throughput Screening of Covalent Organic Frameworks for Carbon Capture using Machine Learning**
   J.S. De Vos, S. Ravichandran, S. Borgmans, L. Vanduyfhuys, P. Van Der Voort, S.M.J. Rogge, and V. Van Speybroeck
   Chemistry of Materials, **2024**, 36 (9), 4315–4330
   IF: 8.6, Number of citations: 1

Conference contributions

Oral presentations

1. **Quantifying the likelihood of structural models through a dynamically enhanced powder X-ray diffraction protocol**
   S. Borgmans, S.M.J. Rogge, J.S. De Vos, C.V. Stevens, P. Van Der Voort, V. Van Speybroeck
   EURMOF2021-Young Investigator Symposium, Online, September 10, 2021

2. **Exploring the phase stability in interpenetrated diamondoid covalent organic frameworks**
   S. Borgmans, S.M.J. Rogge, J.S. De Vos, P. Van Der Voort, V. Van Speybroeck
   MOF2022, Dresden, Germany, September 4–7, 2022

Poster presentations

1. **Construction of a Ready-to-Use Database to Enable Accurate High-Throughput Screenings of Covalent Organic Frameworks**
   J. De Vos, S. Borgmans, S.M.J. Rogge, R. Demuynck, P. Van Der Voort, V. Van Speybroeck
   MOFSIM2019, Ghent, Belgium, April 10–12, 2019

2. **Construction of a Ready-To-Use Database to Enable Accurate High-Throughput Screenings of Covalent Organic Frameworks**
   J. De Vos, S. Borgmans, S.M.J. Rogge, R. Demuynck, P. Van Der Voort, V. Van Speybroeck
   MOLSIM2020, Amsterdam, The Netherlands, January 6–17, 2020

3. **Quantifying the likelihood of structural models through a dynamically enhanced powder X-ray diffraction protocol**
   S. Borgmans, S.M.J. Rogge, J.S. De Vos, C.V. Stevens, P. Van Der Voort, V. Van Speybroeck
   EUROMOF2021, Online, September 13–15, 2021

4. **Exploring the phase stability in interpenetrated diamondoid covalent organic frameworks**
   S. Borgmans, S.M.J. Rogge, J.S. De Vos, P. Van Der Voort, V. Van Speybroeck
   WATOC 2020, Vancouver, Canada, July 3–8, 2022

5. **ReDD-COFFEE: a ready-to-use database of covalent organic framework structures and accurate force fields to enable high-throughput screenings**
   J.S. De Vos, S. Borgmans, P. Van Der Voort, S.M.J. Rogge, V. Van Speybroeck
   French MOFs, COFs and porous polymers conference 2023, Paris, France, May 23–24, 2023

6. **High-throughput screening of covalent organic frameworks for carbon capture using machine learning**
   J.S. De Vos, S. Ravichandran, S. Borgmans, L. Vanduyfhuys, P. Van Der Voort, S.M.J. Rogge, and V. Van Speybroeck
   MOFSIM2024, Montpellier, France, April 10–12, 2024

Master's thesis

High-throughput screening to predict the textural and adsorption properties of covalent organic frameworks using a top-down approach

Juul De Vos

Master's thesis performed at the Center for Molecular Modeling (CMM), Ghent University, 2017–2018

Supervisor: prof. dr. ir. Veronique Van Speybroeck

C

List of Software Packages

A PhD in a molecular modeling discipline obviously requires a lot of computational work. To avoid starting from scratch to implement all routines and calculations, numerous software packages are employed throughout this thesis to facilitate the simulations. An overview of the used programs is provided below, together with a short explanation of their functionality and how they contributed to this thesis.

Avogadro

Avogadro is an open-source molecular builder and visualization tool. Besides flexible high quality rendering and a powerful plugin architecture within areas such as computational chemistry, molecular modeling, bioinformatics, and materials science, it is also designed to run on Windows, Linux, and Mac operating systems. Within this thesis, Avogadro is used to build an initial geometry of the SBUs. More information can be found on https://avogadro.cc/ and in Ref. 274.

Gaussian16

Gaussian16 is the most recent release of the Gaussian series of programs.[340] Being used by scientists worldwide, it has a long history within the field of electronic structure calculation of non-periodic molecular systems, providing the most advanced modeling capabilities. Within this thesis, Gaussian16 is adopted to compute the *ab initio* optimized geometry and Hessian of cluster models, from which the cluster force fields can be derived. More information can be found on https://gaussian.com/gaussian16/.

HORTON

HORTON is an acronym for Helpful Open-source Research TOol for N-fermion systems.[341] The ambition of this program, primarily writtin in Python, is to provide a platform to rapidly test new concepts to approximately solve the quantum many-body problem. Therefore, it strongly focuses on code-readability and user-friendliness, rather than computational efficiency. Within this thesis, HORTON is employed to compute the partial charges using the MBIS partitioning scheme. These charges determine the electrostatic contribution of the used force fields. More information can be found on https://theochem.github.io/horton/.

LAMMPS

LAMMPS is an acronym for Large-scale Atomic/Molecular Massively Parallel Simulator. This is a classical molecular dynamics software that is used in the field of materials modeling. As the name suggests, the developers put much attention to the efficiency of this code and enabled parallelization over multiple processors to speed-up calculations. Within this thesis, LAMMPS is adopted to calculate the long-range forces during MD simulations via the coupling with Yaff. More information can be found on https://lammps.org/ and in Ref. 342.

Molden

Molden is a pre- and post-processing software package for electronic structure calculations. It supports the visualization of electronic orbitals from various *ab initio* codes. Within this thesis, Molden is used to automatically identify the MM3 atom types that determine the van der Waals interactions within the cluster force fields. More information can be found on https://www.theochem.ru.nl/molden/ and in Refs. 276 and 277.

MolMod

MolMod is a Python module developed at the CMM containing a variety of auxiliary tools for molecular modeling.[282] Many of the software packages developed at the CMM, such as Yaff, QuickFF, and TAMkin, are built upon this library. Furthermore, also the implementation of our automated *in silico* assembly algorithm within this thesis adopts a lot of the functionalities in MolMod. More information can be found on https://molmod.github.io/molmod/.

QuickFF

QuickFF is a Python software package developed at the CMM that can efficiently derive system-specific force fields for periodic and non-periodic molecular systems from *ab initio* data. Within this thesis, QuickFF is adopted to compute SBU cluster force fields from which periodic force fields can be determined. More information can be found on https://molmod.github.io/QuickFF/ and in Refs. 229 and 230.

RASPA

RASPA is a molecular simulation software package intended to determine adsorption and diffusion properties in flexible nanoporous materials. It contains the most recent molecular dynamics and Monte Carlo methods. Within this thesis, RASPA is employed to calculate adsorption properties using single-component and mixture GCMC calculations. More information can be found on https://iraspa.org/raspa/ and in Ref. 343.

TAMkin

TAMkin is a Python software package developed at the CMM for normal mode analysis, thermochemistry, and reaction kinetics post-processing. From these analysis methods, various thermodynamic and kinetic properties can be derived. Within this thesis, TAMkin is adopted to perform a normal mode analysis on the COF clusters with the derived force fields. This allows for a validation by comparing the resulting frequencies with the *ab initio* reference data. More information can be found on https://molmod.github.io/tamkin/ and in Ref. 344.

Yaff

Yaff is an acronym for Yet another force field.[281] Initially, this Python software package, developed at the CMM, was intended as a platform to test and validate various force fields. Due to its flexibility, it has grown to include a broad set of functionalities, including an (enhanced) molecular dynamics and Monte Carlo engine. Within this thesis, Yaff is used to perform geometry optimizations and molecular dynamics simulations. Furthermore, our implementation of the automated *in silico* assembly algorithm inherits a lot of Yaff functionalities. More information can be found on https://molmod.github.io/yaff/.

Zeo++

Zeo++ is a software package intended for the pore analysis of crystalline porous materials. Furthermore, it can also be used to *in silico* generate materials or for structure similarity calculations. Within this thesis, Zeo++ is employed to calculate the textural properties of COFs, MOFs, and zeolites. More information can be found on https://www.zeoplusplus.org/ and in Refs. 306 and 147.

D

Acknowledgements

This work was performed within the framework of a strategic basic (SB) research fellowship (grant no. 1S94521N) granted by the Research Foundation Flanders (FWO) and a teaching mandate at the Faculty of Engineering and Architecture of Ghent University.

This work was supported by the Research Board of Ghent University (BOF) through a Concerted Research Action (GOA010-17).

The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by VSC (Flemish Supercomputer Center), funded by Ghent University, FWO, and the Flemish Government – department EWI. The current infrastructure can be consulted at https://www.ugent.be/hpc/en/infrastructure.

Bibliography

[1] E. O. Pyzer-Knapp, C. Suh, R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, and A. Aspuru-Guzik, “What is high-throughput virtual screening? A perspective from organic materials discovery,” *Annu. Rev. Mater. Res.*, vol. 45, pp. 195–216, 2015.

[2] E. Photos, “The question of meteoritic versus smelted nickel-rich iron: archaeological evidence and experimental results,” *World Archaeol.*, vol. 20, pp. 403–421, 1989.

[3] R. B. Heimann, *Classic and advanced ceramics: from fundamentals to applications.* John Wiley & Sons, 2010.

[4] D. B. Harden, “Anglo-Saxon and later medieval glass in Britain: some recent developments,” *Mediev. Archaeol.*, vol. 22, pp. 1–24, 1978.

[5] G. Ceder and K. Persson, “The stuff of dreams,” *Sci. Am.*, vol. 309, pp. 36–40, 2013.

[6] J. Dalton, *A new system of chemical philosophy*, vol. 1. Cambridge University Press, 2010.

[7] J. J. Thomson, “XL. Cathode rays,” *Philos. Mag.*, vol. 44, pp. 293–316, 1897.

[8] E. Rutherford, “LXXIX. The scattering of α and β particles by matter and the structure of the atom,” *Philos. Mag.*, vol. 21, pp. 669–688, 1911.

[9] E. Schrödinger, “An undulatory theory of the mechanics of atoms and molecules,” *Phys. Rev.*, vol. 28, pp. 1049–1070, 1926.

[10] W. Kohn and L. J. Sham, “Self-consistent equations including exchange and correlation effects,” *Phys. Rev.*, vol. 140, pp. A1133–A1138, 1965.

[11] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, "Commentary: the materials project: a materials genome approach to accelerating materials innovation," *APL Mater.*, vol. 1, p. 011002, 2013.

[12] Y. Zhao, Z. Yao, C. Liu, W. Ren, G. Hautier, and L. A. Burton, "High-throughput identification of materials for silicon tandem solar cells," *Sustain. Energ. Fuels*, vol. 7, pp. 812–820, 2023.

[13] S. Adams and R. P. Rao, "High power lithium ion battery materials by computational design," *Phys. Status Solidi A*, vol. 208, pp. 1746–1753, 2011.

[14] M. de Jong, W. Chen, H. Geerlings, M. Asta, and K. A. Persson, "A database to enable discovery and design of piezoelectric materials," *Sci. Data*, vol. 2, p. 150053, 2015.

[15] E. O. Pyzer-Knapp, J. W. Pitera, P. W. J. Staar, S. Takeda, T. Laino, D. P. Sanders, J. Sexton, J. R. Smith, and A. Curioni, "Accelerating materials discovery using artificial intelligence, high performance computing and robotics," *npj Comput. Mater.*, vol. 8, p. 84, 2022.

[16] S. G. Louie, Y.-H. Chan, F. H. da Jornada, Z. Li, and D. Y. Qiu, "Discovering and understanding materials through computation," *Nat. Mater.*, vol. 20, pp. 728–735, 2021.

[17] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM J. Res. Dev.*, vol. 3, pp. 210–229, 1959.

[18] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Hung Byers, "Big data: the next frontier for innovation, competition, and productivity," *McKinsey Global Institute*, 2011.

[19] K. M. Jablonka, D. Ongari, S. M. Moosavi, and B. Smit, "Big-data science in porous materials: materials genomics and machine learning," *Chem. Rev.*, vol. 120, pp. 8066–8129, 2020.

[20] Y. Mao, Q. He, and X. Zhao, "Designing complex architectured materials with generative adversarial networks," *Sci. Adv.*, vol. 6, p. eaaz4169, 2020.

[21] J. Behler, "Perspective: machine learning potentials for atomistic simulations," *J. Chem. Phys.*, vol. 145, p. 170901, 2016.

[22] S. Vandenhaute, M. Cools-Ceuppens, S. DeKeyser, T. Verstraelen, and V. Van Speybroeck, "Machine learning potentials for metal-organic

frameworks using an incremental learning approach," *npj Comput. Mater.*, vol. 9, p. 19, 2023.

[23] V. Van Speybroeck, S. Vandenhaute, A. E. Hoffman, and S. M. J. Rogge, "Towards modeling spatiotemporal processes in metal–organic frameworks," *Trends Chem.*, vol. 3, pp. 605–619, 2021.

[24] V. Van Speybroeck, "Challenges in modelling dynamic processes in realistic nanostructured materials at operating conditions," *Phil. Trans. R. Soc. A*, vol. 381, p. 20220239, 2023.

[25] B. F. Hoskins and R. Robson, "Design and construction of a new class of scaffolding-like materials comprising infinite polymeric frameworks of 3D-linked molecular rods. A reappraisal of the zinc cyanide and cadmium cyanide structures and the synthesis and structure of the diamond-related frameworks $[N(CH_3)_4][Cu^I Zn^{II}(CN)_4]$ and Cu^I[4,4',4",4"'-tetracyanotetraphenylmethane]$BF_4.xC_6H_5NO_2$," *J. Am. Chem. Soc.*, vol. 112, pp. 1546–1554, 1990.

[26] A. P. Côté, A. I. Benin, N. W. Ockwig, M. O'Keeffe, A. J. Matzger, and O. M. Yaghi, "Porous, crystalline, covalent organic frameworks," *Science*, vol. 310, pp. 1166–1170, 2005.

[27] E. E. Ruppert, R. S. Fox, and R. D. Barnes, *Invertebrate zoology: a functional evolutionary approach.* Thomson-Brooks/Cole, 7 ed., 2004.

[28] P. Sebastian and S. B., "Nanoporous materials," *J. Nanosci. Nanotech.*, vol. 2, pp. 581–612, 2002.

[29] International Union of Pure and Applied Chemistry (IUPAC), "microporous carbon," *IUPAC Compendium of Chemical Terminology*, 2019. 3 ed., Online version 3.0.1, DOI: 10.1351/goldbook.M03909.

[30] International Union of Pure and Applied Chemistry (IUPAC), "macroporous polymer," *IUPAC Compendium of Chemical Terminology*, 2019. 3 ed., Online version 3.0.1, DOI: 10.1351/goldbook.MT07177.

[31] O. K. Farha, A. Özgür Yazaydın, I. Eryazici, C. D. Malliakas, B. G. Hauser, M. G. Kanatzidis, S. T. Nguyen, R. Q. Snurr, and J. T. Hupp, "De novo synthesis of a metal–organic framework material featuring ultrahigh surface area and gas storage capacities," *Nat. Chem.*, vol. 2, pp. 944–948, 2010.

[32] C. Wang, D. Liu, and W. Lin, "Metal–organic frameworks as a tunable platform for designing functional molecular materials," *J. Am. Chem. Soc.*, vol. 135, pp. 13222–13234, 2013.

[33] S. M. Cohen, "Postsynthetic methods for the functionalization of metal–organic frameworks," *Chem. Rev.*, vol. 112, pp. 970–1000, 2012.

[34] J. Jiang, Y. Zhao, and O. M. Yaghi, "Covalent chemistry beyond molecules," *J. Am. Chem. Soc.*, vol. 138, pp. 3255–3265, 2016.

[35] A. F. Cronstedt, "Rön och beskrifning om en obekant bärg art, som kallas zeolites," *Akad. Handl. Stockholm*, vol. 17, pp. 120–123, 1756.

[36] R. M. Barrer, "435. Syntheses and reactions of mordenite," *J. Chem. Soc.*, pp. 2158–2163, 1948.

[37] R. M. Barrer, "33. Synthesis of a zeolitic mineral with chabazite-like sorptive properties," *J. Chem. Soc.*, pp. 127–132, 1948.

[38] E. M. Flanigen, "Chapter 2. Zeolites and molecular sieves an historical perspective," in *Introduction to Zeolite Science and Practice* (H. van Bekkum, E. Flanigen, and J. Jansen, eds.), vol. 58 of *Studies in Surface Science and Catalysis*, pp. 13–34, Elsevier, 1991.

[39] D. W. Breck, W. G. Eversole, R. M. Milton, T. B. Reed, and T. L. Thomas, "Crystalline zeolites. I. The properties of a new synthetic zeolite, type A," *J. Am. Chem. Soc.*, vol. 78, pp. 5963–5972, 1956.

[40] T. B. Reed and D. W. Breck, "Crystalline zeolites. II. Crystal structure of synthetic zeolite, type A," *J. Am. Chem. Soc.*, vol. 78, pp. 5972–5977, 1956.

[41] M. Guisnet and J.-P. Gilson, *Zeolites for cleaner technologies*, vol. 3. Imperial College Press London, 2002.

[42] P. A. Wright and J. A. Connor, *Microporous Framework Solids*. The Royal Society of Chemistry, 2007.

[43] F. Guerzoni and J. Abbot, "Catalytic cracking of a gippsland reduced crude on zeolite catalysts," *J. Catal.*, vol. 147, pp. 393–403, 1994.

[44] Y. N. Nayak, S. Nayak, Y. Nadaf, N. S. Shetty, and S. L. Gaonkar, "Zeolite catalyzed Friedel-Crafts reactions: a review," *Lett. Org. Chem.*, vol. 17, pp. 491–506, 2020.

[45] S. Kulprathipanja, *Zeolites in industrial separation and catalysis.* John Wiley & Sons, 2010.

[46] A. G. Slater and A. I. Cooper, "Function-led design of new porous materials," *Science*, vol. 348, p. aaa8075, 2015.

[47] C. Baerlocher, D. Brouwer, B. Marler, and L. B. McCusker, "Database of zeolite structures." http://www.iza-structure.org/databases/. Accessed: 2022-08-01.

[48] S. Kitagawa, S. Kawata, Y. Nozaka, and M. Munakata, "Synthesis and crystal structures of novel copper(I) co-ordination polymers and a hexacopper(I) cluster of quinoline-2-thione," *J. Chem. Soc., Dalton Trans.*, pp. 1399–1404, 1993.

[49] L. R. MacGillivray, S. Subramanian, and M. J. Zaworotko, "Interwoven two- and three-dimensional coordination polymers through self-assembly of Cu cations with linear bidentate ligands," *J. Chem. Soc., Chem. Commun.*, pp. 1325–1326, 1994.

[50] O. M. Yaghi and H. Li, "Hydrothermal synthesis of a metal-organic framework containing large rectangular channels," *J. Am. Chem. Soc.*, vol. 117, pp. 10401–10402, 1995.

[51] O. M. Yaghi, G. Li, and H. Li, "Selective binding and removal of guests in a microporous metal–organic framework," *Nature*, vol. 378, pp. 703–706, 1995.

[52] D. Riou and G. Férey, "Hybrid open frameworks (MIL-n). Part 3. Crystal structures of the HT and LT forms of MIL-7: a new vanadium propylenediphosphonate with an open-framework. Influence of the synthesis temperature on the oxidation state of vanadium within the same structural type," *J. Mater. Chem.*, vol. 8, pp. 2733–2735, 1998.

[53] H. Li, M. Eddaoudi, M. O'Keeffe, and O. M. Yaghi, "Design and synthesis of an exceptionally stable and highly porous metal-organic framework," *Nature*, vol. 402, pp. 276–279, 1999.

[54] M. Eddaoudi, D. B. Moler, H. Li, B. Chen, T. M. Reineke, M. O'Keeffe, and O. M. Yaghi, "Modular chemistry: secondary building units as a basis for the design of highly porous and robust metal-organic carboxylate frameworks," *Acc. Chem. Res.*, vol. 34, pp. 319–330, 2001.

[55] M. O'Keeffe, M. A. Peskov, S. J. Ramsden, and O. M. Yaghi, "The reticular chemistry structure resource (RCSR) database of, and symbols for, crystal nets," *Acc. Chem. Res.*, vol. 41, pp. 1782–1789, 2008. Accessed: 2020-02-06.

[56] M. Rubio-Martinez, C. Avci-Camur, A. W. Thornton, I. Imaz, D. Maspoch, and M. R. Hill, "New synthetic routes towards MOF production at scale," *Chem. Soc. Rev.*, vol. 46, pp. 3453–3480, 2017.

[57] G. N. Lewis, "The atom and the molecule," *J. Am. Chem. Soc.*, vol. 38, pp. 762–785, 1916.

[58] W. Kossel, "Über Molekülbildung als Frage des Atombaus," *Ann. Phys.*, vol. 354, pp. 229–362, 1916.

[59] I. Langmuir, "The arrangement of electrons in atoms and molecules," *J. Am. Chem. Soc.*, vol. 41, pp. 868–934, 1919.

[60] R. Hoffmann, "How should chemists think?," *Sci. Am.*, vol. 268, pp. 66–73, 1993.

[61] H. M. El-Kaderi, J. R. Hunt, J. L. Mendoza-Cortés, A. P. Côté, R. E. Taylor, M. O'Keeffe, and O. M. Yaghi, "Designed synthesis of 3D covalent organic frameworks," *Science*, vol. 316, pp. 268–272, 2007.

[62] X. Feng, X. Ding, and D. Jiang, "Covalent organic frameworks," *Chem. Soc. Rev.*, vol. 41, pp. 6010–6022, 2012.

[63] S.-Y. Ding and W. Wang, "Covalent organic frameworks (COFs): from design to applications," *Chem. Soc. Rev.*, vol. 42, pp. 548–568, 2013.

[64] P. J. Waller, F. Gándara, and O. M. Yaghi, "Chemistry of covalent organic frameworks," *Acc. Chem. Res.*, vol. 48, pp. 3053–3063, 2015.

[65] C. S. Diercks and O. M. Yaghi, "The atom, the molecule, and the covalent organic framework," *Science*, vol. 355, p. eaal1585, 2017.

[66] M. S. Lohse and T. Bein, "Covalent organic frameworks: structures, synthesis, and applications," *Adv. Funct. Mater.*, vol. 28, p. 1705553, 2018.

[67] H. Furukawa and O. M. Yaghi, "Storage of hydrogen, methane, and carbon dioxide in highly porous covalent organic frameworks for clean energy applications," *J. Am. Chem. Soc.*, vol. 131, pp. 8875–8883, 2009.

[68] Y. Zeng, R. Zou, and Y. Zhao, "Covalent organic frameworks for CO_2 capture," *Adv. Mater.*, vol. 28, pp. 2855–2873, 2016.

[69] S. B. Kalidindi and R. A. Fischer, "Covalent organic frameworks and their metal nanoparticle composites: prospects for hydrogen storage," *Phys. Status Solidi B*, vol. 250, pp. 1119–1127, 2013.

[70] S.-Y. Ding, J. Gao, Q. Wang, Y. Zhang, W.-G. Song, C.-Y. Su, and W. Wang, "Construction of covalent organic framework for catalysis: Pd/COF-LZU1 in Suzuki–Miyaura coupling reaction," *J. Am. Chem. Soc.*, vol. 133, pp. 19816–19822, 2011.

[71] Q. Fang, S. Gu, J. Zheng, Z. Zhuang, S. Qiu, and Y. Yan, "3D microporous base-functionalized covalent organic frameworks for size-selective catalysis," *Angew. Chem. Int. Ed.*, vol. 53, pp. 2878–2882, 2014.

[72] X. Wang, X. Han, J. Zhang, X. Wu, Y. Liu, and Y. Cui, "Homochiral 2D porous covalent organic frameworks for heterogeneous asymmetric catalysis," *J. Am. Chem. Soc.*, vol. 138, pp. 12332–12335, 2016.

[73] L. Stegbauer, K. Schwinghammer, and B. V. Lotsch, "A hydrazone-based covalent organic framework for photocatalytic hydrogen production," *Chem. Sci.*, vol. 5, pp. 2789–2793, 2014.

[74] S. Wan, J. Guo, J. Kim, H. Ihee, and D. Jiang, "A belt-shaped, blue luminescent, and semiconducting covalent organic framework," *Angew. Chem. Int. Ed.*, vol. 47, pp. 8826–8830, 2008.

[75] J. I. Feldblyum, C. H. McCreery, S. C. Andrews, T. Kurosawa, E. J. G. Santos, V. Duong, L. Fang, A. L. Ayzner, and Z. Bao, "Few-layer, large-area, 2D covalent organic framework semiconductor thin films," *Chem. Commun.*, vol. 51, pp. 13894–13897, 2015.

[76] M. Dogru and T. Bein, "On the road towards electroactive covalent organic frameworks," *Chem. Commun.*, vol. 50, pp. 5531–5546, 2014.

[77] S. Dalapati, S. Jin, J. Gao, Y. Xu, A. Nagai, and D. Jiang, "An azine-linked covalent organic framework," *J. Am. Chem. Soc.*, vol. 135, pp. 17310–17313, 2013.

[78] Q. Fang, J. Wang, S. Gu, R. B. Kaspar, Z. Zhuang, J. Zheng, H. Guo, S. Qiu, and Y. Yan, "3D porous crystalline polyimide covalent organic frameworks for drug delivery," *J. Am. Chem. Soc.*, vol. 137, pp. 8352–8355, 2015.

[79] F. J. Uribe-Romo, J. R. Hunt, H. Furukawa, C. Klöck, M. O'Keeffe, and O. M. Yaghi, "A crystalline imine-linked 3-D porous covalent organic framework," *J. Am. Chem. Soc.*, vol. 131, pp. 4570–4571, 2009.

[80] T. Ma, E. A. Kapustin, S. X. Yin, L. Liang, Z. Zhou, J. Niu, L.-H. Li, Y. Wang, J. Su, J. Li, X. Wang, W. D. Wang, W. Wang, J. Sun, and O. M. Yaghi, "Single-crystal X-ray diffraction structures of covalent organic frameworks," *Science*, vol. 361, pp. 48–52, 2018.

[81] H. Li, H. Li, Q. Dai, H. Li, and J.-L. Brédas, "Hydrolytic stability of boronate ester-linked covalent organic frameworks," *Adv. Theory Simul.*, vol. 1, p. 1700015, 2018.

[82] L. M. Lanni, R. W. Tilford, M. Bharathy, and J. J. Lavigne, "Enhanced hydrolytic stability of self-assembling alkylated two-dimensional covalent organic frameworks," *J. Am. Chem. Soc.*, vol. 133, pp. 13975–13983, 2011.

[83] C. Qian, L. Feng, W. L. Teo, J. Liu, W. Zhou, D. Wang, and Y. Zhao, "Imine and imine-derived linkages in two-dimensional covalent organic frameworks," *Nat. Rev. Chem.*, vol. 6, pp. 881–898, 2022.

[84] A. Halder, S. Karak, M. Addicoat, S. Bera, A. Chakraborty, S. H. Kunjattu, P. Pachfule, T. Heine, and R. Banerjee, "Ultrastable imine-based covalent organic frameworks for sulfuric acid recovery: an effect of interlayer hydrogen bonding," *Angew. Chem. Int. Ed.*, vol. 57, pp. 5797–5802, 2018.

[85] S. Kandambeth, A. Mallick, B. Lukose, M. V. Mane, T. Heine, and R. Banerjee, "Construction of crystalline 2D covalent organic frameworks with remarkable chemical (acid/base) stability via a combined reversible and irreversible route," *J. Am. Chem. Soc.*, vol. 134, pp. 19524–19527, 2012.

[86] P. Puthiaraj, Y.-R. Lee, S. Zhang, and W.-S. Ahn, "Triazine-based covalent organic polymers: design, synthesis and applications in heterogeneous catalysis," *J. Mater. Chem. A*, vol. 4, pp. 16288–16311, 2016.

[87] S.-Y. Ding, M. Dong, Y.-W. Wang, Y.-T. Chen, H.-Z. Wang, C.-Y. Su, and W. Wang, "Thioether-based fluorescent covalent organic framework for selective detection and facile removal of mercury(II)," *J. Am. Chem. Soc.*, vol. 138, pp. 3031–3037, 2016.

[88] J. S. De Vos, S. Borgmans, P. Van Der Voort, S. M. J. Rogge, and V. Van Speybroeck, "ReDD-COFFEE: a ready-to-use database of covalent organic framework structures and accurate force fields to enable high-throughput screenings," *J. Mater. Chem. A*, vol. 11, pp. 7468–7487, 2023.

[89] Intergovernmental Panel on Climate Change (IPCC), "Global carbon and other biogeochemical cycles and feedbacks," in *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 673–816, Cambridge University Press, 2023.

[90] Intergovernmental Panel on Climate Change (IPCC), "Weather and climate extreme events in a changing climate," in *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth*

Assessment Report of the Intergovernmental Panel on Climate Change, pp. 1513–1766, Cambridge University Press, 2023.

[91] J. A. Martens, A. Bogaerts, N. De Kimpe, P. A. Jacobs, G. B. Marin, K. Rabaey, M. Saeys, and S. Verhelst, "The chemical route to a carbon dioxide neutral world," *ChemSusChem*, vol. 10, pp. 1039–1055, 2017.

[92] R. S. Haszeldine, "Carbon capture and storage: how green can black be?," *Science*, vol. 325, pp. 1647–1652, 2009.

[93] H. Yang, Z. Xu, M. Fan, R. Gupta, R. B. Slimane, A. E. Bland, and I. Wright, "Progress in carbon dioxide separation and capture: a review," *J. Environ. Sci.*, vol. 20, pp. 14–27, 2008.

[94] B. Smit, J. A. Reimer, C. M. Oldenburg, and I. C. Bourg, *Introduction to carbon capture and sequestration*, vol. 1. Imperial College Press, 2014.

[95] M. E. Boot-Handford, J. C. Abanades, E. J. Anthony, M. J. Blunt, S. Brandani, N. Mac Dowell, J. R. Fernández, M.-C. Ferrari, R. Gross, J. P. Hallett, R. S. Haszeldine, P. Heptonstall, A. Lyngfelt, Z. Makuch, E. Mangano, R. T. J. Porter, M. Pourkashanian, G. T. Rochelle, N. Shah, J. G. Yao, and P. S. Fennell, "Carbon capture and storage update," *Energy Environ. Sci.*, vol. 7, pp. 130–189, 2014.

[96] P. Markewitz, W. Kuckshinrichs, W. Leitner, J. Linssen, P. Zapp, R. Bongartz, A. Schreiber, and T. E. Müller, "Worldwide innovations in the development of carbon capture technologies and the utilization of CO_2," *Energy Environ. Sci.*, vol. 5, pp. 7281–7305, 2012.

[97] L. Joos, J. M. Huck, V. Van Speybroeck, and B. Smit, "Cutting the cost of carbon capture: a case for carbon capture and utilization," *Faraday Discuss.*, vol. 192, pp. 391–414, 2016.

[98] R. H. Weiland, J. C. Dingman, and D. B. Cronin, "Heat capacity of aqueous monoethanolamine, diethanolamine, N-methyldiethanolamine, and N-methyldiethanolamine-based blends with carbon dioxide," *J. Chem. Eng. Data*, vol. 42, pp. 1004–1006, 1997.

[99] G. T. Rochelle, "Amine scrubbing for CO_2 capture," *Science*, vol. 325, pp. 1652–1654, 2009.

[100] R. Ben-Mansour, M. Habib, O. Bamidele, M. Basha, N. Qasem, A. Peedikakkal, T. Laoui, and M. Ali, "Carbon capture by physical adsorption: materials, experimental investigations and numerical modeling and simulations – a review," *Appl. Energy*, vol. 161, pp. 225–255, 2016.

[101] X. Lu, D. Jin, S. Wei, Z. Wang, C. An, and W. Guo, "Strategies to enhance CO_2 capture and separation based on engineering absorbent materials," *J. Mater. Chem. A*, vol. 3, pp. 12118–12132, 2015.

[102] G. Férey, C. Serre, T. Devic, G. Maurin, H. Jobic, P. L. Llewellyn, G. De Weireld, A. Vimont, M. Daturi, and J.-S. Chang, "Why hybrid porous solids capture greenhouse gases?," *Chem. Soc. Rev.*, vol. 40, pp. 550–562, 2011.

[103] S. Choi, J. H. Drese, and C. W. Jones, "Adsorbent materials for carbon dioxide capture from large anthropogenic point sources," *ChemSusChem*, vol. 2, pp. 796–854, 2009.

[104] T.-H. Bae, M. R. Hudson, J. A. Mason, W. L. Queen, J. J. Dutton, K. Sumida, K. J. Micklash, S. S. Kaye, C. M. Brown, and J. R. Long, "Evaluation of cation-exchanged zeolite adsorbents for post-combustion carbon dioxide capture," *Energy Environ. Sci.*, vol. 6, pp. 128–138, 2013.

[105] K. Sumida, D. L. Rogow, J. A. Mason, T. M. McDonald, E. D. Bloch, Z. R. Herm, T.-H. Bae, and J. R. Long, "Carbon dioxide capture in metal–organic frameworks," *Chem. Rev.*, vol. 112, pp. 724–781, 2012.

[106] Y. J. Choi, J. H. Choi, K. M. Choi, and J. K. Kang, "Covalent organic frameworks for extremely high reversible CO_2 uptake capacity: a theoretical approach," *J. Mater. Chem.*, vol. 21, pp. 1073–1078, 2011.

[107] A. A. Olajire, "Recent advances in the synthesis of covalent organic frameworks for CO_2 capture," *J. CO2 Util.*, vol. 17, pp. 137–161, 2017.

[108] Y. Wang, C. Kang, Z. Zhang, A. K. Usadi, D. C. Calabro, L. S. Baugh, Y. D. Yuan, and D. Zhao, "Evaluation of Schiff-base covalent organic frameworks for CO_2 capture: structure–performance relationships, stability, and performance under wet conditions," *ACS Sustainable Chem. Eng.*, vol. 10, pp. 332–341, 2022.

[109] J. L. Mendoza-Cortés, S. S. Han, H. Furukawa, O. M. Yaghi, and W. A. Goddard, "Adsorption mechanism and uptake of methane in covalent organic frameworks: theory and experiment," *J. Phys. Chem. A*, vol. 114, pp. 10824–10833, 2010.

[110] J. L. Mendoza-Cortes, T. A. Pascal, and W. A. Goddard, "Design of covalent organic frameworks for methane storage," *J. Phys. Chem. A*, vol. 115, pp. 13852–13857, 2011.

[111] M. Q. Wang and H. S. Huang, “A full fuel-cycle analysis of energy and emissions impacts of transportation fuels produced from natural gas,” 2000. https://www.osti.gov/biblio/750803, DOI: 10.2172/750803.

[112] S. Yeh, “An empirical analysis on the adoption of alternative fuel vehicles: the case of natural gas vehicles,” *Energy Policy*, vol. 35, pp. 5865–5875, 2007.

[113] M. I. Khan, T. Yasmin, and A. Shakoor, “Technical overview of compressed natural gas (CNG) as a transportation fuel,” *Renew. Sustain. Energy Rev.*, vol. 51, pp. 785–797, 2015.

[114] W. Lim, K. Choi, and I. Moon, “Current status and perspectives of liquefied natural gas (LNG) plant design,” *Ind. Eng. Chem. Res.*, vol. 52, pp. 3065–3088, 2013.

[115] J. Wegrzyn and M. Gurevich, “Adsorbent storage of natural gas,” *Appl. Energy*, vol. 55, pp. 71–83, 1996.

[116] V. C. Menon and S. Komarneni, “Porous adsorbents for vehicular natural gas storage: a review,” *J. Porous Mater.*, vol. 5, pp. 43–58, 1998.

[117] “DE-FOA-0000672: methane opportunities for vehicular energy (MOVE).” https://arpa-e-foa.energy.gov/Default.aspx?Search=DE-FOA-0000672. Accessed: 2022-09-21.

[118] M. Debruyne, S. Borgmans, S. Radhakrishnan, E. Breynaert, H. Vrielinck, K. Leus, A. Laemont, J. De Vos, K. S. Rawat, S. Vanlommel, H. Rijckaert, H. Salemi, J. Everaert, F. Vanden Bussche, D. Poelman, R. Morent, N. De Geyter, P. Van Der Voort, V. Van Speybroeck, and C. V. Stevens, “Engineering of phenylpyridine- and bipyridine-based covalent organic frameworks for photocatalytic tandem aerobic oxidation/povarov cyclization,” *ACS Appl. Mater. Interfaces*, vol. 15, pp. 35092–35106, 2023.

[119] M. Ramos, A. P. S. Dias, J. F. Puna, J. Gomes, and J. C. Bordado, “Biodiesel production processes and sustainable raw materials,” *Energies*, vol. 12, p. 4408, 2019.

[120] S. Cao, J. Low, J. Yu, and M. Jaroniec, “Polymeric photocatalysts based on graphitic carbon nitride,” *Adv. Mater.*, vol. 27, pp. 2150–2176, 2015.

[121] S. M. J. Rogge, A. Bavykina, J. Hajek, H. Garcia, A. I. Olivos-Suarez, A. Sepúlveda-Escribano, A. Vimont, G. Clet, P. Bazin, F. Kapteijn,

M. Daturi, E. V. Ramos-Fernandez, F. X. Llabrés i Xamena, V. Van Speybroeck, and J. Gascon, "Metal–organic and covalent organic frameworks as single-site catalysts," *Chem. Soc. Rev.*, vol. 46, pp. 3134–3184, 2017.

[122] Y. Wu, H. Xu, X. Chen, J. Gao, and D. Jiang, "A π-electronic covalent organic framework catalyst: π-walls as catalytic beds for Diels–Alder reactions under ambient conditions," *Chem. Commun.*, vol. 51, pp. 10096–10098, 2015.

[123] G.-B. Wang, S. Li, C.-X. Yan, F.-C. Zhu, Q.-Q. Lin, K.-H. Xie, Y. Geng, and Y.-B. Dong, "Covalent organic frameworks: emerging high-performance platforms for efficient photocatalytic applications," *J. Mater. Chem. A*, vol. 8, pp. 6957–6983, 2020.

[124] N. A. Romero and D. A. Nicewicz, "Organic photoredox catalysis," *Chem. Rev.*, vol. 116, pp. 10075–10166, 2016.

[125] D. M. Schultz and T. P. Yoon, "Solar synthesis: prospects in visible light photocatalysis," *Science*, vol. 343, p. 1239176, 2014.

[126] N. W. Ockwig, O. Delgado-Friedrichs, M. O'Keeffe, and O. M. Yaghi, "Reticular chemistry: occurrence and taxonomy of nets and grammar for the design of frameworks," *Acc. Chem. Res.*, vol. 38, pp. 176–182, 2005.

[127] X. Ding, J. Guo, X. Feng, Y. Honsho, J. Guo, S. Seki, P. Maitarad, A. Saeki, S. Nagase, and D. Jiang, "Synthesis of metallophthalocyanine covalent organic frameworks that exhibit high carrier mobility and photoconductivity," *Angew. Chem. Int. Ed.*, vol. 50, pp. 1289–1293, 2011.

[128] T.-Y. Zhou, S.-Q. Xu, Q. Wen, Z.-F. Pang, and X. Zhao, "One-step construction of two different kinds of pores in a 2D covalent organic framework," *J. Am. Chem. Soc.*, vol. 136, pp. 15885–15888, 2014.

[129] G. Lin, H. Ding, D. Yuan, B. Wang, and C. Wang, "A pyrene-based, fluorescent three-dimensional covalent organic framework," *J. Am. Chem. Soc.*, vol. 138, pp. 3302–3305, 2016.

[130] S. Dalapati, M. Addicoat, S. Jin, T. Sakurai, J. Gao, H. Xu, S. Irle, S. Seki, and D. Jiang, "Rational design of crystalline supermicroporous covalent organic frameworks with triangular topologies," *Nat. Commun.*, vol. 6, p. 7786, 2015.

[131] C. Gropp, T. Ma, N. Hanikel, and O. M. Yaghi, "Design of higher valency in covalent organic frameworks," *Science*, vol. 370, p. eabd6406, 2020.

[132] B. Zhang, H. Mao, R. Matheu, J. A. Reimer, S. A. Alshmimri, S. Alshihri, and O. M. Yaghi, "Reticular synthesis of multinary covalent organic frameworks," *J. Am. Chem. Soc.*, vol. 141, pp. 11420–11424, 2019.

[133] R. W. Tilford, S. J. Mugavero III, P. J. Pellechia, and J. J. Lavigne, "Tailoring microporosity in covalent organic frameworks," *Adv. Mater.*, vol. 20, pp. 2741–2746, 2008.

[134] X. Feng, L. Liu, Y. Honsho, A. Saeki, S. Seki, S. Irle, Y. Dong, A. Nagai, and D. Jiang, "High-rate charge-carrier transport in porphyrin covalent organic frameworks: switching from hole to electron to ambipolar conduction," *Angew. Chem. Int. Ed.*, vol. 51, pp. 2618–2622, 2012.

[135] F. H. Allen, "The Cambridge Structural Database: a quarter of a million crystal structures and rising," *Acta Crystallogr. B*, vol. 58, pp. 380–388, 2002.

[136] Y. G. Chung, J. Camp, M. Haranczyk, B. J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O. K. Farha, D. S. Sholl, and R. Q. Snurr, "Computation-ready, experimental metal-organic frameworks: a tool to enable high-throughput screening of nanoporous crystals," *Chem. Mater.*, vol. 26, pp. 6185–6192, 2014.

[137] Y. G. Chung, E. Haldoupis, B. J. Bucior, M. Haranczyk, S. Lee, H. Zhang, K. D. Vogiatzis, M. Milisavljevic, S. Ling, J. S. Camp, B. Slater, J. I. Siepmann, D. S. Sholl, and R. Q. Snurr, "Advances, updates, and analytics for the computation-ready, experimental metal-organic framework database: CoRE MOF 2019," *J. Chem. Eng. Data*, vol. 64, pp. 5985–5998, 2019.

[138] P. Z. Moghadam, A. Li, S. B. Wiggin, A. Tao, A. G. P. Maloney, P. A. Wood, S. C. Ward, and D. Fairen-Jimenez, "Development of a Cambridge Structural Database subset: a collection of metal-organic frameworks for past, present, and future," *Chem. Mater.*, vol. 29, pp. 2618–2625, 2017.

[139] A. S. Rosen, S. M. Iyer, D. Ray, Z. Yao, A. Aspuru-Guzik, L. Gagliardi, J. M. Notestein, and R. Q. Snurr, "Machine learning the quantum-chemical properties of metal-organic frameworks for accelerated materials discovery," *Matter*, vol. 4, pp. 1578–1597, 2021.

[140] M. Tong, Y. Lan, Q. Yang, and C. Zhong, "Exploring the structure-property relationships of covalent organic frameworks for noble gas separations," *Chem. Eng. Sci.*, vol. 168, pp. 456–464, 2017.

[141] D. Ongari, A. V. Yakutovich, L. Talirz, and B. Smit, "Building a consistent and reproducible database for adsorption evaluation in covalent–organic frameworks," *ACS Cent. Sci.*, vol. 5, pp. 1663–1675, 2019.

[142] Y. Lan, X. Han, M. Tong, H. Huang, Q. Yang, D. Liu, X. Zhao, and C. Zhong, "Materials genomics methods for high-throughput construction of COFs and targeted synthesis," *Nat. Commun.*, vol. 9, p. 5274, 2018.

[143] C. Mellot Draznieks, J. M. Newsam, A. M. Gorman, C. M. Freeman, and G. Férey, "De novo prediction of inorganic structures developed through automated assembly of secondary building units (AASBU method)," *Angew. Chem. Int. Ed.*, vol. 39, pp. 2270–2275, 2000.

[144] C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp, and R. Q. Snurr, "Large-scale screening of hypothetical metal–organic frameworks," *Nat. Chem.*, vol. 4, pp. 83–89, 2012.

[145] M. Fernandez, P. G. Boyd, T. D. Daff, M. Z. Aghaji, and T. K. Woo, "Rapid and accurate machine learning recognition of high performing metal organic frameworks for CO_2 capture," *J. Phys. Chem. Lett.*, vol. 5, pp. 3056–3060, 2014.

[146] B. J. Sikora, R. Winnegar, D. M. Proserpio, and R. Q. Snurr, "Textural properties of a large collection of computationally constructed MOFs and zeolites," *Microporous Mesoporous Mater.*, vol. 186, pp. 207–213, 2014.

[147] R. L. Martin and M. Haranczyk, "Construction and characterization of structure models of crystalline porous polymers," *Cryst. Growth Des.*, vol. 14, pp. 2431–2440, 2014.

[148] M. A. Addicoat, D. E. Coupry, and T. Heine, "AuToGraFS: automatic topological generator for framework structures," *J. Phys. Chem. A*, vol. 118, pp. 9607–9614, 2014.

[149] S. Bureekaew, V. Balwani, S. Amirjalayer, and R. Schmid, "Isoreticular isomerism in 4,4-connected paddle-wheel metal-organic frameworks: structural prediction by the reverse topological approach," *CrystEngComm*, vol. 17, pp. 344–352, 2015.

[150] P. G. Boyd and T. K. Woo, “A generalized method for constructing hypothetical nanoporous materials of any net topology from graph theory,” *CrystEngComm*, vol. 18, pp. 3777–3792, 2016.

[151] Y. J. Colón, D. A. Gómez-Gualdrón, and R. Q. Snurr, “Topologically guided, automated construction of metal-organic frameworks and their evaluation for energy-related applications,” *Cryst. Growth Des.*, vol. 17, pp. 5801–5810, 2017.

[152] R. Anderson and D. A. Gómez-Gualdrón, “Increasing topological diversity during computational “synthesis” of porous crystals: how and why,” *CrystEngComm*, vol. 21, pp. 1653–1665, 2019.

[153] R. L. Martin, C. M. Simon, B. Medasani, D. K. Britt, B. Smit, and M. Haranczyk, “In silico design of three-dimensional porous covalent organic frameworks via known synthesis routes and commercially available species,” *J. Phys. Chem. C*, vol. 118, pp. 23790–23802, 2014.

[154] R. Mercado, R.-S. Fu, A. V. Yakutovich, L. Talirz, M. Haranczyk, and B. Smit, “In silico design of 2D and 3D covalent organic frameworks for methane storage applications,” *Chem. Mater.*, vol. 30, pp. 5069–5086, 2018.

[155] Y. J. Colón and R. Q. Snurr, “High-throughput computational screening of metal–organic frameworks,” *Chem. Soc. Rev.*, vol. 43, pp. 5735–5749, 2014.

[156] P. G. Boyd, A. Chidambaram, E. García-Díez, C. P. Ireland, T. D. Daff, R. Bounds, A. Gładysiak, P. Schouwink, S. M. Moosavi, M. M. Maroto-Valer, J. A. Reimer, J. A. R. Navarro, T. K. Woo, S. Garcia, K. C. Stylianou, and B. Smit, “Data-driven design of metal–organic frameworks for wet flue gas CO_2 capture,” *Nature*, vol. 576, pp. 253–256, 2019.

[157] Y.-S. Bae and R. Q. Snurr, “Development and evaluation of porous materials for carbon dioxide separation and capture,” *Angew. Chem. Int. Ed.*, vol. 50, pp. 11586–11596, 2011.

[158] C. E. Wilmer, O. K. Farha, Y.-S. Bae, J. T. Hupp, and R. Q. Snurr, “Structure–property relationships of porous materials for carbon dioxide separation and capture,” *Energy Environ. Sci.*, vol. 5, pp. 9849–9856, 2012.

[159] H. Daglar and S. Keskin, “Recent advances, opportunities, and challenges in high-throughput computational screening of MOFs for gas separations,” *Coord. Chem. Rev.*, vol. 422, p. 213470, 2020.

[160] C. M. Simon, J. Kim, D. A. Gomez-Gualdron, J. S. Camp, Y. G. Chung, R. L. Martin, R. Mercado, M. W. Deem, D. Gunter, M. Haranczyk, D. S. Sholl, R. Q. Snurr, and B. Smit, "The materials genome in action: identifying the performance limits for methane storage," *Energy Environ. Sci.*, vol. 8, pp. 1190–1199, 2015.

[161] M. Tong, Y. Lan, Z. Qin, and C. Zhong, "Computation-ready, experimental covalent organic framework for methane delivery: screening and material design," *J. Phys. Chem. C*, vol. 122, pp. 13009–13016, 2018.

[162] Y. J. Colón, D. Fairen-Jimenez, C. E. Wilmer, and R. Q. Snurr, "High-throughput screening of porous crystalline materials for hydrogen storage capacity near room temperature," *J. Phys. Chem. C*, vol. 118, pp. 5383–5389, 2014.

[163] J. Goldsmith, A. G. Wong-Foy, M. J. Cafarella, and D. J. Siegel, "Theoretical limits of hydrogen storage in metal–organic frameworks: opportunities and trade-offs," *Chem. Mater.*, vol. 25, pp. 3373–3382, 2013.

[164] D. A. Gómez-Gualdrón, Y. J. Colón, X. Zhang, T. C. Wang, Y.-S. Chen, J. T. Hupp, T. Yildirim, O. K. Farha, J. Zhang, and R. Q. Snurr, "Evaluating topologically diverse metal–organic frameworks for cryo-adsorbed hydrogen storage," *Energy Environ. Sci.*, vol. 9, pp. 3279–3289, 2016.

[165] N. S. Bobbitt, J. Chen, and R. Q. Snurr, "High-throughput screening of metal–organic frameworks for hydrogen storage at cryogenic temperature," *J. Phys. Chem. C*, vol. 120, pp. 27328–27341, 2016.

[166] M. Tong, W. Zhu, J. Li, Z. Long, S. Zhao, G. Chen, and Y. Lan, "An easy way to identify high performing covalent organic frameworks for hydrogen storage," *Chem. Commun.*, vol. 56, pp. 6376–6379, 2020.

[167] Z. Qiao, C. Peng, J. Zhou, and J. Jiang, "High-throughput computational screening of 137953 metal–organic frameworks for membrane separation of a $CO_2/N_2/CH_4$ mixture," *J. Mater. Chem. A*, vol. 4, pp. 15904–15912, 2016.

[168] Z. Qiao, K. Zhang, and J. Jiang, "In silico screening of 4764 computation-ready, experimental metal–organic frameworks for CO_2 separation," *J. Mater. Chem. A*, vol. 4, pp. 2105–2114, 2016.

[169] C. Altintas, G. Avci, H. Daglar, A. Nemati Vesali Azar, S. Velioglu, I. Erucar, and S. Keskin, "Database for CO_2 separation performances of MOFs based on computational materials screening," *ACS Appl. Mater. Interfaces*, vol. 10, pp. 17257–17268, 2018.

[170] M. Tong, Q. Yang, and C. Zhong, “Computational screening of covalent organic frameworks for CH_4/H_2, CO_2/H_2 and CO_2/CH_4 separations,” *Microporous Mesoporous Mater.*, vol. 210, pp. 142–148, 2015.

[171] T. Yan, Y. Lan, M. Tong, and C. Zhong, “Screening and design of covalent organic framework membranes for CO_2/CH_4 separation,” *ACS Sustainable Chem. Eng.*, vol. 7, pp. 1220–1227, 2019.

[172] O. F. Altundal, Z. P. Haslak, and S. Keskin, “Combined GCMC, MD, and DFT approach for unlocking the performances of COFs for methane purification,” *Ind. Eng. Chem. Res.*, vol. 60, pp. 12999–13012, 2021.

[173] C. Altintas, I. Erucar, and S. Keskin, “High-throughput computational screening of the metal organic framework database for CH_4/H_2 separations,” *ACS Appl. Mater. Interfaces*, vol. 10, pp. 3668–3679, 2018.

[174] Y. Basdogan, K. B. Sezginel, and S. Keskin, “Identifying highly selective metal organic frameworks for CH_4/H_2 separations using computational tools,” *Ind. Eng. Chem. Res.*, vol. 54, pp. 8479–8491, 2015.

[175] G. Avci, S. Velioglu, and S. Keskin, “High-throughput screening of MOF adsorbents and membranes for H_2 purification and CO_2 capture,” *ACS Appl. Mater. Interfaces*, vol. 10, pp. 33693–33706, 2018.

[176] G. O. Aksu, H. Daglar, C. Altintas, and S. Keskin, “Computational selection of high-performing covalent organic frameworks for adsorption and membrane-based CO_2/H_2 separation,” *J. Phys. Chem. C*, vol. 124, pp. 22577–22590, 2020.

[177] G. O. Aksu, I. Erucar, Z. P. Haslak, and S. Keskin, “Accelerating discovery of COFs for CO_2 capture and H_2 purification using structurally guided computational screening,” *Chem. Eng. J.*, vol. 427, p. 131574, 2022.

[178] S. Li, Y. G. Chung, and R. Q. Snurr, “High-throughput screening of metal–organic frameworks for CO_2 capture in the presence of water,” *Langmuir*, vol. 32, pp. 10368–10376, 2016.

[179] L.-C. Lin, A. H. Berger, R. L. Martin, J. Kim, J. A. Swisher, K. Jariwala, C. H. Rycroft, A. S. Bhown, M. W. Deem, M. Haranczyk, and B. Smit, “In silico screening of carbon-capture materials,” *Nat. Mater.*, vol. 11, pp. 633–641, 2012.

[180] E. Haldoupis, S. Nair, and D. S. Sholl, “Finding MOFs for highly selective CO_2/N_2 adsorption using materials screening based on efficient assignment of atomic point charges,” *J. Am. Chem. Soc.*, vol. 134, pp. 4313–4323, 2012.

[181] T. Watanabe and D. S. Sholl, “Accelerating applications of metal–organic frameworks for gas adsorption and separation by computational screening of materials,” *Langmuir*, vol. 28, pp. 14114–14128, 2012.

[182] H. Daglar and S. Keskin, “Computational screening of metal–organic frameworks for membrane-based $CO_2/N_2/H_2O$ separations: best materials for flue gas separation,” *J. Phys. Chem. C*, vol. 122, pp. 17347–17357, 2018.

[183] D. Wu, Q. Yang, C. Zhong, D. Liu, H. Huang, W. Zhang, and G. Maurin, “Revealing the structure–property relationships of metal–organic frameworks for CO_2 capture from flue gas,” *Langmuir*, vol. 28, pp. 12094–12099, 2012.

[184] M. Tong, Q. Yang, Y. Xiao, and C. Zhong, “Revealing the structure–property relationship of covalent organic frameworks for CO_2 capture from postcombustion gas: a multi-scale computational study,” *Phys. Chem. Chem. Phys.*, vol. 16, pp. 15189–15198, 2014.

[185] K. S. Deeg, D. Damasceno Borges, D. Ongari, N. Rampal, L. Talirz, A. V. Yakutovich, J. M. Huck, and B. Smit, “In silico discovery of covalent organic frameworks for carbon capture,” *ACS Appl. Mater. Interfaces*, vol. 12, pp. 21559–21568, 2020.

[186] O. F. Altundal, C. Altintas, and S. Keskin, “Can COFs replace MOFs in flue gas separation? High-throughput computational screening of COFs for CO_2/N_2 separation,” *J. Mater. Chem. A*, vol. 8, pp. 14609–14623, 2020.

[187] M. Islamov, H. Babaei, R. Anderson, K. B. Sezginel, J. R. Long, A. J. H. McGaughey, D. A. Gomez-Gualdron, and C. E. Wilmer, “High-throughput screening of hypothetical metal–organic frameworks for thermal conductivity,” *npj Comput. Mater.*, vol. 9, p. 11, 2023.

[188] P. Z. Moghadam, S. M. J. Rogge, A. Li, C.-M. Chow, J. Wieme, N. Moharrami, M. Aragones-Anglada, G. Conduit, D. A. Gomez-Gualdron, V. Van Speybroeck, and D. Fairen-Jimenez, “Structure-mechanical stability relations of metal-organic frameworks via machine learning,” *Matter*, vol. 1, pp. 219–234, 2019.

[189] A. S. Rosen, V. Fung, P. Huck, C. T. O'Donnell, M. K. Horton, D. G. Truhlar, K. A. Persson, J. M. Notestein, and R. Q. Snurr, "High-throughput predictions of metal–organic framework electronic properties: theoretical challenges, graph neural networks, and data exploration," *npj Comput. Mater.*, vol. 8, p. 112, 2022.

[190] Y. He, E. D. Cubuk, M. D. Allendorf, and E. J. Reed, "Metallic metal–organic frameworks predicted by the combination of machine learning methods and ab initio calculations," *J. Phys. Chem. Lett.*, vol. 9, pp. 4562–4569, 2018.

[191] A. S. Rosen, J. M. Notestein, and R. Q. Snurr, "Identifying promising metal–organic frameworks for heterogeneous catalysis via high-throughput periodic density functional theory," *J. Comput. Chem.*, vol. 40, pp. 1305–1318, 2019.

[192] S. Lee, B. Kim, H. Cho, H. Lee, S. Y. Lee, E. S. Cho, and J. Kim, "Computational screening of trillions of metal–organic frameworks for high-performance methane storage," *ACS Appl. Mater. Interfaces*, vol. 13, pp. 23647–23654, 2021.

[193] H. Demir, H. Daglar, H. C. Gulbalkan, G. O. Aksu, and S. Keskin, "Recent advances in computational modeling of MOFs: from molecular simulations to machine learning," *Coord. Chem. Rev.*, vol. 484, p. 215112, 2023.

[194] M. Pardakhti, E. Moharreri, D. Wanik, S. L. Suib, and R. Srivastava, "Machine learning using combined structural and chemical descriptors for prediction of methane adsorption performance of metal organic frameworks (MOFs)," *ACS Comb. Sci.*, vol. 19, pp. 640–645, 2017.

[195] M. Pardakhti, P. Nanda, and R. Srivastava, "Impact of chemical features on methane adsorption by porous materials at varying pressures," *J. Phys. Chem. C*, vol. 124, pp. 4534–4544, 2020.

[196] R. Anderson, J. Rodgers, E. Argueta, A. Biong, and D. A. Gómez-Gualdrón, "Role of pore chemistry and topology in the CO_2 capture capabilities of MOFs: from molecular simulation to machine learning," *Chem. Mater.*, vol. 30, pp. 6325–6337, 2018.

[197] J. P. Janet and H. J. Kulik, "Resolving transition metal chemical space: feature selection for machine learning and structure–property relationships," *J. Phys. Chem. A*, vol. 121, pp. 8939–8954, 2017.

[198] S. M. Moosavi, A. Nandy, K. M. Jablonka, D. Ongari, J. P. Janet, P. G. Boyd, Y. Lee, B. Smit, and H. J. Kulik, "Understanding the diversity of the metal-organic framework ecosystem," *Nat. Commun.*, vol. 11, p. 4068, 2020.

[199] M. Fernandez, N. R. Trefiak, and T. K. Woo, "Atomic property weighted radial distribution functions descriptors of metal–organic frameworks for the prediction of gas uptake capacity," *J. Phys. Chem. C*, vol. 117, pp. 14095–14105, 2013.

[200] H. Dureckova, M. Krykunov, M. Z. Aghaji, and T. K. Woo, "Robust machine learning models for predicting high CO_2 working capacity and CO_2/H_2 selectivity of gas adsorption in metal organic frameworks for precombustion carbon capture," *J. Phys. Chem. C*, vol. 123, pp. 4133–4139, 2019.

[201] C. Altintas, O. F. Altundal, S. Keskin, and R. Yildirim, "Machine learning meets with metal organic frameworks for gas storage and separation," *J. Chem. Inf. Model.*, vol. 61, pp. 2131–2146, 2021.

[202] N. S. Bobbitt and R. Q. Snurr, "Molecular modelling and machine learning for high-throughput screening of metal–organic frameworks for hydrogen storage," *Mol. Simul.*, vol. 45, pp. 1069–1081, 2019.

[203] B. J. Bucior, N. S. Bobbitt, T. Islamoglu, S. Goswami, A. Gopalan, T. Yildirim, O. K. Farha, N. Bagheri, and R. Q. Snurr, "Energy-based descriptors to rapidly predict hydrogen storage in metal–organic frameworks," *Mol. Syst. Des. Eng.*, vol. 4, pp. 162–174, 2019.

[204] G. Anderson, B. Schweitzer, R. Anderson, and D. A. Gómez-Gualdrón, "Attainable volumetric targets for adsorption-based hydrogen storage in porous crystals: molecular simulation and machine learning," *J. Phys. Chem. C*, vol. 123, pp. 120–130, 2019.

[205] A. W. Thornton, C. M. Simon, J. Kim, O. Kwon, K. S. Deeg, K. Konstas, S. J. Pas, M. R. Hill, D. A. Winkler, M. Haranczyk, and B. Smit, "Materials genome in action: identifying the performance limits of physical hydrogen storage," *Chem. Mater.*, vol. 29, pp. 2844–2854, 2017.

[206] G. S. Fanourgakis, K. Gkagkas, E. Tylianakis, and G. Froudakis, "Fast screening of large databases for top performing nanomaterials using a self-consistent, machine learning based approach," *J. Phys. Chem. C*, vol. 124, pp. 19639–19648, 2020.

[207] M. Fernandez, T. K. Woo, C. E. Wilmer, and R. Q. Snurr, “Large-scale quantitative structure–property relationship (QSPR) analysis of methane storage in metal–organic frameworks,” *J. Phys. Chem. C*, vol. 117, pp. 7681–7689, 2013.

[208] M. Fernandez and A. S. Barnard, “Geometrical properties can predict CO_2 and N_2 adsorption performance of metal–organic frameworks (MOFs) at low pressure,” *ACS Comb. Sci.*, vol. 18, pp. 243–252, 2016.

[209] Z. Yao, B. Sánchez-Lengeling, N. S. Bobbitt, B. J. Bucior, S. G. H. Kumar, S. P. Collins, T. Burns, T. K. Woo, O. K. Farha, R. Q. Snurr, and A. Aspuru-Guzik, “Inverse design of nanoporous crystalline reticular materials with deep generative models,” *Nat. Mach. Intell.*, vol. 3, pp. 76–86, 2021.

[210] Y. G. Chung, D. A. Gómez-Gualdrón, P. Li, K. T. Leperi, P. Deria, H. Zhang, N. A. Vermeulen, J. F. Stoddart, F. You, J. T. Hupp, O. K. Farha, and R. Q. Snurr, “In silico discovery of metal–organic frameworks for precombustion CO_2 capture using a genetic algorithm,” *Sci. Adv.*, vol. 2, p. e1600909, 2016.

[211] M. Z. Aghaji, M. Fernandez, P. G. Boyd, T. D. Daff, and T. K. Woo, “Quantitative structure-property relationship models for recognizing metal organic frameworks (MOFs) with high CO_2 working capacity and CO_2/CH_4 selectivity for methane purification,” *Eur. J. Inorg. Chem.*, vol. 2016, pp. 4505–4511, 2016.

[212] G. O. Aksu and S. Keskin, “Advancing CH_4/H_2 separation with covalent organic frameworks by combining molecular simulations and machine learning,” *J. Mater. Chem. A*, vol. 11, pp. 14788–14799, 2023.

[213] Y. Lim, J. Park, S. Lee, and J. Kim, “Finely tuned inverse design of metal–organic frameworks with user-desired Xe/Kr selectivity,” *J. Mater. Chem. A*, vol. 9, pp. 21175–21183, 2021.

[214] C. M. Simon, R. Mercado, S. K. Schnell, B. Smit, and M. Haranczyk, “What are the best materials to separate a xenon/krypton mixture?,” *Chem. Mater.*, vol. 27, pp. 4459–4475, 2015.

[215] S. M. Moosavi, A. Chidambaram, L. Talirz, M. Haranczyk, K. C. Stylianou, and B. Smit, “Capturing chemical intuition in synthesis of metal-organic frameworks,” *Nat. Commun.*, vol. 10, p. 539, 2019.

[216] M. G. Saunders and G. A. Voth, “Coarse-graining methods for computational biology,” *Annu. Rev. Biophys.*, vol. 42, pp. 73–93, 2013.

[217] M. Born and R. Oppenheimer, "Zur Quantentheorie der Molekeln," *Ann. Phys.*, vol. 389, pp. 457–484, 1927.

[218] V. Fock, "Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems," *Z. Phys.*, vol. 61, pp. 126–148, 1930.

[219] D. R. Hartree and W. Hartree, "Self-consistent field, with exchange, for beryllium," *Proc. R. Soc. A*, vol. 150, pp. 9–33, 1935.

[220] C. Møller and M. S. Plesset, "Note on an approximation treatment for many-electron systems," *Phys. Rev.*, vol. 46, pp. 618–622, 1934.

[221] J. Čížek, "On the correlation problem in atomic and molecular systems. Calculation of wavefunction components in Ursell-type expansion using quantum-field theoretical methods," *J. Chem. Phys.*, vol. 45, pp. 4256–4266, 2004.

[222] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, "A fifth-order perturbation comparison of electron correlation theories," *Chem. Phys. Lett.*, vol. 157, pp. 479–483, 1989.

[223] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Phys. Rev.*, vol. 136, pp. B864–B871, 1964.

[224] J. P. Perdew and K. Schmidt, "Jacob's ladder of density functional approximations for the exchange-correlation energy," *AIP Conf. Proc.*, vol. 577, pp. 1–20, 2001.

[225] A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange," *J. Chem. Phys.*, vol. 98, pp. 5648–5652, 1993.

[226] C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Phys. Rev. B*, vol. 37, pp. 785–789, 1988.

[227] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, "Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields," *J. Phys. Chem.*, vol. 98, pp. 11623–11627, 1994.

[228] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," *J. Chem. Phys.*, vol. 132, p. 154104, 2010.

[229] L. Vanduyfhuys, S. Vandenbrande, T. Verstraelen, R. Schmid, M. Waroquier, and V. Van Speybroeck, "QuickFF: a program for a quick and easy derivation of force fields for metal-organic frameworks from ab initio input," *J. Comput. Chem.*, vol. 36, pp. 1015–1027, 2015.

[230] L. Vanduyfhuys, S. Vandenbrande, J. Wieme, M. Waroquier, T. Verstraelen, and V. Van Speybroeck, "Extension of the QuickFF force field protocol for an improved accuracy of structural, vibrational, mechanical and thermal properties of metal-organic frameworks," *J. Comput. Chem.*, vol. 39, pp. 999–1011, 2018.

[231] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, "UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations," *J. Am. Chem. Soc.*, vol. 114, pp. 10024–10035, 1992.

[232] N. L. Allinger, Y. H. Yuh, and J. H. Lii, "Molecular mechanics. The MM3 force field for hydrocarbons. 1," *J. Am. Chem. Soc.*, vol. 111, pp. 8551–8566, 1989.

[233] S. Borgmans, S. M. J. Rogge, J. S. De Vos, P. Van Der Voort, and V. Van Speybroeck, "Exploring the phase stability in interpenetrated diamondoid covalent organic frameworks," *Commun. Chem.*, vol. 6, p. 5, 2023.

[234] T. Verstraelen, S. Vandenbrande, F. Heidar-Zadeh, L. Vanduyfhuys, V. Van Speybroeck, M. Waroquier, and P. W. Ayers, "Minimal basis iterative stockholder: atoms in molecules for force-field development," *J. Chem. Theory Comput.*, vol. 12, pp. 3894–3912, 2016.

[235] B. L. Bush, C. I. Bayly, and T. A. Halgren, "Consensus bond-charge increments fitted to electrostatic potential or field of many compounds: application to MMFF94 training set," *J. Comput. Chem.*, vol. 20, pp. 1495–1516, 1999.

[236] F. London, "Zur Theorie und Systematik der Molekularkräfte," *Z. Phys.*, vol. 63, pp. 245–279, 1930.

[237] H. A. Lorentz, "Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase," *Ann. Phys.*, vol. 248, pp. 127–136, 1881.

[238] D. Berthelot, "Sur le mélange des gaz," *Compt. Rendus*, vol. 126, p. 15, 1898.

[239] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids," *J. Am. Chem. Soc.*, vol. 118, pp. 11225–11236, 1996.

[240] S. L. Mayo, B. D. Olafson, and W. A. Goddard, "DREIDING: a generic force field for molecular simulations," *J. Phys. Chem.*, vol. 94, pp. 8897–8909, 1990.

[241] S. Bureekaew, S. Amirjalayer, M. Tafipolsky, C. Spickermann, T. K. Roy, and R. Schmid, "MOF-FF – a flexible first-principles derived force field for metal-organic frameworks," *Phys. Status Solidi B*, vol. 250, pp. 1128–1141, 2013.

[242] L. Vanduyfhuys, T. Verstraelen, M. Vandichel, M. Waroquier, and V. Van Speybroeck, "Ab initio parametrized force field for the flexible metal-organic framework MIL-53(Al)," *J. Chem. Theory Comput.*, vol. 8, pp. 3217–3231, 2012.

[243] M. Tafipolsky, S. Amirjalayer, and R. Schmid, "Ab initio parametrized MM3 force field for the metal-organic framework MOF-5," *J. Comput. Chem.*, vol. 28, pp. 1169–1176, 2007.

[244] R. Schmid and M. Tafipolsky, "An accurate force field model for the strain energy analysis of the covalent organic framework COF-102," *J. Am. Chem. Soc.*, vol. 130, pp. 12600–12601, 2008.

[245] M. Tafipolsky and R. Schmid, "Systematic first principles parameterization of force fields for metal-organic frameworks using a genetic algorithm approach," *J. Phys. Chem. B*, vol. 113, pp. 1341–1352, 2009.

[246] M. Tafipolsky, S. Amirjalayer, and R. Schmid, "First-principles-derived force field for copper paddle-wheel-based metal-organic frameworks," *J. Phys. Chem. C*, vol. 114, pp. 14402–14409, 2010.

[247] S. Amirjalayer, R. Q. Snurr, and R. Schmid, "Prediction of structure and properties of boron-based covalent organic frameworks by a first-principles derived force field," *J. Phys. Chem. C*, vol. 116, pp. 4921–4929, 2012.

[248] J. Wieme, L. Vanduyfhuys, S. M. J. Rogge, M. Waroquier, and V. Van Speybroeck, "Exploring the flexibility of MIL-47(V)-type materials using force field molecular dynamics simulations," *J. Phys. Chem. C*, vol. 120, pp. 14934–14947, 2016.

[249] A. Cauchy, "Méthode générale pour la résolution des systemes d'équations simultanées," *Comp. Rend. Sci. Paris*, vol. 25, pp. 536–538, 1847.

[250] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *J. Res. Natl. Inst. Stand.*, vol. 49, pp. 409–436, 1952.

[251] C. G. Broyden, "The convergence of a class of double-rank minimization algorithms. 1. General considerations," *IMA J. Appl. Math.*, vol. 6, pp. 76–90, 1970.

[252] R. Fletcher, "A new approach to variable metric algorithms," *Comput. J.*, vol. 13, pp. 317–322, 1970.

[253] D. Goldfarb, "A family of variable-metric methods derived by variational means," *Math. Comput.*, vol. 24, pp. 23–26, 1970.

[254] D. F. Shanno, "Conditioning of quasi-newton methods for function minimization," *Math. Comput.*, vol. 24, pp. 647–656, 1970.

[255] I. Newton, *Philosophiae naturalis principia mathematica*, vol. 1. G. Brookman, 1833.

[256] L. Verlet, "Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules," *Phys. Rev.*, vol. 159, pp. 98–103, 1967.

[257] C. Shannon, "Communication in the presence of noise," *Proc. IEEE*, vol. 37, pp. 10–21, 1949.

[258] G. D. Birkhoff, "Proof of the ergodic theorem," *Proc. Natl. Acad. Sci.*, vol. 17, pp. 656–660, 1931.

[259] J. v. Neumann, "Proof of the quasi-ergodic hypothesis," *Proc. Natl. Acad. Sci.*, vol. 18, pp. 70–82, 1932.

[260] J. W. Gibbs, *Elementary principles in statistical mechanics: developed with especial reference to the rational foundations of thermodynamics.* Charles Scribner's sons, 1902.

[261] P. Attard, "On the density of volume states in the isobaric ensemble," *J. Chem. Phys.*, vol. 103, pp. 9884–9885, 1995.

[262] G. J. M. Koper and H. Reiss, "Length scale for the constant pressure ensemble: application to small systems and relation to Einstein fluctuation theory," *J. Phys. Chem.*, vol. 100, pp. 422–432, 1996.

[263] S. Nosé, "A molecular dynamics method for simulations in the canonical ensemble," *Mol. Phys.*, vol. 52, pp. 255–268, 1984.

[264] W. G. Hoover, "Canonical dynamics: equilibrium phase-space distributions," *Phys. Rev. A*, vol. 31, pp. 1695–1697, 1985.

[265] G. J. Martyna, M. L. Klein, and M. Tuckerman, "Nosé–Hoover chains: the canonical ensemble via continuous dynamics," *J. Chem. Phys.*, vol. 97, pp. 2635–2643, 1992.

[266] G. J. Martyna, D. J. Tobias, and M. L. Klein, "Constant pressure molecular dynamics algorithms," *J. Chem. Phys.*, vol. 101, pp. 4177–4189, 1994.

[267] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein, "Explicit reversible integrators for extended systems dynamics," *Mol. Phys.*, vol. 87, pp. 1117–1157, 1996.

[268] D. Dubbeldam, A. Torres-Knoop, and K. S. Walton, "On the inner workings of Monte Carlo codes," *Mol. Simul.*, vol. 39, pp. 1253–1292, 2013.

[269] T. Kloek and H. K. van Dijk, "Bayesian estimates of equation system parameters: an application of integration by Monte Carlo," *Econometrica*, vol. 46, pp. 1–19, 1978.

[270] W. Kabsch, "Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector," *J. Appl. Crystallogr.*, vol. 21, pp. 916–924, 1988.

[271] A. Le Bail, "Whole powder pattern decomposition methods and applications: a retrospection," *Powder Diffr.*, vol. 20, p. 316–326, 2005.

[272] G. S. Pawley, "Unit-cell refinement from powder diffraction scans," *J. Appl. Crystallogr.*, vol. 14, pp. 357–361, 1981.

[273] A. Le Bail, H. Duroy, and J. Fourquet, "Ab-initio structure determination of $LiSbWO_6$ by X-ray powder diffraction," *Mater. Res. Bull.*, vol. 23, pp. 447–452, 1988.

[274] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison, "Avogadro: an advanced semantic chemical editor, visualization, and analysis platform," *J. Cheminformatics*, vol. 4, p. 17, 2012.

[275] T. Verstraelen, V. Van Speybroeck, and M. Waroquier, "ZEOBUILDER: a GUI toolkit for the construction of complex molecular structures on the

nanoscale with building blocks," *J. Chem. Inf. Model.*, vol. 48, pp. 1530–1541, 2008.

[276] G. Schaftenaar and J. H. Noordik, "Molden: a pre- and post-processing program for molecular and electronic structures*," *J. Comput.-Aided Mol. Des.*, vol. 14, pp. 123–134, 2000.

[277] G. Schaftenaar, E. Vlieg, and G. Vriend, "Molden 2.0: quantum chemistry meets proteins," *J. Comput.-Aided Mol. Des.*, vol. 31, pp. 789–800, 2017.

[278] W. Kabsch, "A solution for the best rotation to relate two sets of vectors," *Acta Crystallogr. A*, vol. 32, pp. 922–923, 1976.

[279] R. Anderson and D. A. Gómez-Gualdrón, "Large-scale free energy calculations on a computational metal–organic frameworks database: toward synthetic likelihood predictions," *Chem. Mater.*, vol. 32, pp. 8106–8119, 2020.

[280] J. S. De Vos, "Supporting information for the generation of the ReDD-COFFEE database." https://github.com/jsdvos/SupportingInformation_ReDD-COFFEE_2023/tree/master/DatabaseGeneration/StructureAssembly/scripts.

[281] T. Verstraelen, L. Vanduyfhuys, S. Vandenbrande, and S. M. J. Rogge, "Yaff, yet another force field (v1.6.0)." http://molmod.ugent.be/software/.

[282] T. Verstraelen, "Molmod software library (v1.4.5)." http://molmod.ugent.be/software/.

[283] Q.-Y. Yang, P. Lama, S. Sen, M. Lusi, K.-J. Chen, W.-Y. Gao, M. Shivanna, T. Pham, N. Hosono, S. Kusaka, J. J. Perry IV, S. Ma, B. Space, L. J. Barbour, S. Kitagawa, and M. J. Zaworotko, "Reversible switching between highly porous and nonporous phases of an interpenetrated diamondoid coordination network that exhibits gate-opening at methane storage pressures," *Angew. Chem. Int. Ed.*, vol. 57, pp. 5684–5689, 2018.

[284] Y.-X. Ma, Z.-J. Li, L. Wei, S.-Y. Ding, Y.-B. Zhang, and W. Wang, "A dynamic three-dimensional covalent organic framework," *J. Am. Chem. Soc.*, vol. 139, pp. 4995–4998, 2017.

[285] T. Sun, L. Wei, Y. Chen, Y. Ma, and Y.-B. Zhang, "Atomic-level characterization of dynamics of a 3D covalent organic framework by cryo-electron diffraction tomography," *J. Am. Chem. Soc.*, vol. 141, pp. 10962–10966, 2019.

[286] Y. Chen, Z.-L. Shi, L. Wei, B. Zhou, J. Tan, H.-L. Zhou, and Y.-B. Zhang, “Guest-dependent dynamics in a 3D covalent organic framework,” *J. Am. Chem. Soc.*, vol. 141, pp. 3298–3303, 2019.

[287] Y.-B. Zhang, J. Su, H. Furukawa, Y. Yun, F. Gándara, A. Duong, X. Zou, and O. M. Yaghi, “Single-crystal structure of a covalent organic framework,” *J. Am. Chem. Soc.*, vol. 135, pp. 16336–16339, 2013.

[288] D. Beaudoin, T. Maris, and J. D. Wuest, “Constructing monocrystalline covalent organic networks by polymerization,” *Nat. Chem.*, vol. 5, pp. 830–834, 2013.

[289] V. K. Pecharsky and P. Y. Zavalij, *Fundamentals of powder diffraction and structural characterization of materials.* Springer, 2009.

[290] L. B. McCusker, “Zeolite crystallography. Structure determination in the absence of conventional single-crystal data,” *Acta Crystallogr. A*, vol. 47, pp. 297–313, 1991.

[291] M. Falcioni and M. W. Deem, “A biased Monte Carlo scheme for zeolite structure solution,” *J. Chem. Phys.*, vol. 110, pp. 1754–1766, 1999.

[292] N. Keller, T. Sick, N. N. Bach, A. Koszalkowski, J. M. Rotter, D. D. Medina, and T. Bein, “Dibenzochrysene enables tightly controlled docking and stabilizes photoexcited states in dual-pore covalent organic frameworks,” *Nanoscale*, vol. 11, pp. 23338–23345, 2019.

[293] S. Borgmans, S. M. J. Rogge, J. S. De Vos, C. V. Stevens, P. Van Der Voort, and V. Van Speybroeck, “Quantifying the likelihood of structural models through a dynamically enhanced powder X-ray diffraction protocol,” *Angew. Chem. Int. Ed.*, vol. 60, pp. 8913–8922, 2021.

[294] J. S. De Vos, S. Borgmans, P. Van Der Voort, S. M. J. Rogge, and V. Van Speybroeck, “ReDD-COFFEE: a ready-to-use database of covalent organic framework structures and accurate force fields to enable high-throughput screenings,” *Materials Cloud Archive*, vol. 2023.13, 2023. 2023.13, DOI: 10.24435/materialscloud:nw-3j.

[295] J. R. Hunt, C. J. Doonan, J. D. LeVangie, A. P. Côté, and O. M. Yaghi, “Reticular synthesis of covalent organic borosilicate frameworks,” *J. Am. Chem. Soc.*, vol. 130, pp. 11872–11873, 2008.

[296] F. J. Uribe-Romo, C. J. Doonan, H. Furukawa, K. Oisaki, and O. M. Yaghi, “Crystalline covalent organic frameworks with hydrazone linkages,” *J. Am. Chem. Soc.*, vol. 133, pp. 11478–11481, 2011.

[297] Q. Fang, Z. Zhuang, S. Gu, R. B. Kaspar, J. Zheng, J. Wang, S. Qiu, and Y. Yan, "Designed synthesis of large-pore crystalline polyimide covalent organic frameworks," *Nat. Commun.*, vol. 5, p. 4503, 2014.

[298] D. A. Pyles, J. W. Crowe, L. A. Baldwin, and P. L. McGrier, "Synthesis of benzobisoxazole-linked two-dimensional covalent organic frameworks and their carbon dioxide capture properties," *ACS Macro Lett.*, vol. 5, pp. 1055–1058, 2016.

[299] C. R. DeBlase, K. E. Silberstein, T.-T. Truong, H. D. Abruña, and W. R. Dichtel, "β-ketoenamine-linked covalent organic frameworks capable of pseudocapacitive energy storage," *J. Am. Chem. Soc.*, vol. 135, pp. 16821–16824, 2013.

[300] P. Kuhn, M. Antonietti, and A. Thomas, "Porous, covalent triazine-based frameworks prepared by ionothermal synthesis," *Angew. Chem. Int. Ed.*, vol. 47, pp. 3450–3453, 2008.

[301] K. T. Jackson, T. E. Reich, and H. M. El-Kaderi, "Targeted synthesis of a porous borazine-linked covalent organic framework," *Chem. Commun.*, vol. 48, pp. 8823–8825, 2012.

[302] M. Li, D. Li, M. O'Keeffe, and O. M. Yaghi, "Topological analysis of metal–organic frameworks with polytopic linkers and/or multiple building units and the minimal transitivity principle," *Chem. Rev.*, vol. 114, pp. 1343–1370, 2014.

[303] Z. Yang, H. Chen, S. Wang, W. Guo, T. Wang, X. Suo, D.-e. Jiang, X. Zhu, I. Popovs, and S. Dai, "Transformation strategy for highly crystalline covalent triazine frameworks: from staggered AB to eclipsed AA stacking," *J. Am. Chem. Soc.*, vol. 142, pp. 6856–6860, 2020.

[304] S. Wan, F. Gándara, A. Asano, H. Furukawa, A. Saeki, S. K. Dey, L. Liao, M. W. Ambrogio, Y. Y. Botros, X. Duan, S. Seki, J. F. Stoddart, and O. M. Yaghi, "Covalent organic frameworks with high charge carrier mobility," *Chem. Mater.*, vol. 23, pp. 4094–4097, 2011.

[305] Z. Li, X. Feng, Y. Zou, Y. Zhang, H. Xia, X. Liu, and Y. Mu, "A 2D azine-linked covalent organic framework for gas storage applications," *Chem. Commun.*, vol. 50, pp. 13825–13828, 2014.

[306] T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza, and M. Haranczyk, "Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials," *Microporous Mesoporous Mater.*, vol. 149, pp. 134–141, 2012.

[307] A. Nandy, C. Duan, J. P. Janet, S. Gugler, and H. J. Kulik, "Strategies and software for machine learning accelerated discovery in transition metal chemistry," *Ind. Eng. Chem. Res.*, vol. 57, pp. 13973–13986, 2018.

[308] J. P. Janet, F. Liu, A. Nandy, C. Duan, T. Yang, S. Lin, and H. J. Kulik, "Designing in the face of uncertainty: exploiting electronic structure and machine learning models for discovery in inorganic chemistry," *Inorg. Chem.*, vol. 58, pp. 10592–10606, 2019.

[309] A. Nandy, J. Zhu, J. P. Janet, C. Duan, R. B. Getman, and H. J. Kulik, "Machine learning accelerates the discovery of design rules and exceptions in stable metal–oxo intermediate formation," *ACS Catal.*, vol. 9, pp. 8243–8255, 2019.

[310] A. F. Ismail, K. C. Khulbe, and T. Matsuura, *Fundamentals of gas permeation through membranes*, pp. 11–35. Cham: Springer International Publishing, 2015.

[311] M. G. Martin and J. I. Siepmann, "Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes," *J. Phys. Chem. B*, vol. 102, pp. 2569–2577, 1998.

[312] J. A. Mason, M. Veenstra, and J. R. Long, "Evaluating metal–organic frameworks for natural gas storage," *Chem. Sci.*, vol. 5, pp. 32–51, 2014.

[313] J. Möllmer, A. Möller, F. Dreisbach, R. Gläser, and R. Staudt, "High pressure adsorption of hydrogen, nitrogen, carbon dioxide and methane on the metal–organic framework HKUST-1," *Microporous Mesoporous Mater.*, vol. 138, pp. 140–148, 2011.

[314] J. A. Mason, J. Oktawiec, M. K. Taylor, M. R. Hudson, J. Rodriguez, J. E. Bachman, M. I. Gonzalez, A. Cervellino, A. Guagliardi, C. M. Brown, P. L. Llewellyn, N. Masciocchi, and J. R. Long, "Methane storage in flexible metal–organic frameworks with intrinsic thermal management," *Nature*, vol. 527, pp. 357–361, 2015.

[315] M. Zhang, W. Zhou, T. Pham, K. A. Forrest, W. Liu, Y. He, H. Wu, T. Yildirim, B. Chen, B. Space, Y. Pan, M. J. Zaworotko, and J. Bai, "Fine tuning of MOF-505 analogues to reduce low-pressure methane uptake and enhance methane working capacity," *Angew. Chem. Int. Ed.*, vol. 56, pp. 11426–11430, 2017.

[316] Y. Peng, V. Krungleviciute, I. Eryazici, J. T. Hupp, O. K. Farha, and T. Yildirim, "Methane storage in metal-organic frameworks: current records, surprise findings, and challenges," *J. Am. Chem. Soc.*, vol. 135, pp. 11887–11894, 2013.

[317] J. Lan, D. Cao, and W. Wang, "High uptakes of methane in Li-doped 3D covalent organic frameworks," *Langmuir*, vol. 26, pp. 220–226, 2010.

[318] D. M. D'Alessandro, B. Smit, and J. R. Long, "Carbon dioxide capture: prospects for new materials," *Angew. Chem. Int. Ed.*, vol. 49, pp. 6058–6082, 2010.

[319] P. Bai, M. Y. Jeon, L. Ren, C. Knight, M. W. Deem, M. Tsapatsis, and J. I. Siepmann, "Discovery of optimal zeolites for challenging separations and chemical transformations using predictive materials modeling," *Nat. Commun.*, vol. 6, p. 5912, 2015.

[320] S. Vandenbrande, M. Waroquier, V. Van Speybroeck, and T. Verstraelen, "Ab initio evaluation of Henry coefficients using importance sampling," *J. Chem. Theory Comput.*, vol. 14, pp. 6359–6369, 2018.

[321] J. S. De Vos, S. Ravichandran, S. Borgmans, L. Vanduyfhuys, P. Van Der Voort, S. M. J. Rogge, and V. Van Speybroeck, "High-throughput screening of covalent organic frameworks for carbon capture using machine learning," *Chem. Mater.*, vol. 36, no. 9, pp. 4315–4330, 2024.

[322] S. Cavenati, C. A. Grande, and A. E. Rodrigues, "Adsorption equilibrium of methane, carbon dioxide, and nitrogen on zeolite 13X at high pressures," *J. Chem. Eng. Data*, vol. 49, pp. 1095–1101, 2004.

[323] D. Saha, Z. Bao, F. Jia, and S. Deng, "Adsorption of CO_2, CH_4, N_2O, and N_2 on MOF-5, MOF-177, and zeolite 5A," *Environ. Sci. Technol.*, vol. 44, pp. 1820–1826, 2010.

[324] Z. Yıldız and H. Uzun, "Prediction of gas storage capacities in metal organic frameworks using artificial neural network," *Microporous Mesoporous Mater.*, vol. 208, pp. 50–54, 2015.

[325] E. Ren and F.-X. Coudert, "Enhancing gas separation selectivity prediction through geometrical and chemical descriptors," *Chem. Mater.*, vol. 35, pp. 6771–6781, 2023.

[326] Y. Zhao, K. X. Yao, B. Teng, T. Zhang, and Y. Han, "A perfluorinated covalent triazine-based framework for highly selective and water–tolerant CO_2 capture," *Energy Environ. Sci.*, vol. 6, pp. 3684–3692, 2013.

[327] G. Wang, K. Leus, H. S. Jena, C. Krishnaraj, S. Zhao, H. Depauw, N. Tahir, Y.-Y. Liu, and P. Van Der Voort, "A fluorine-containing hydrophobic covalent triazine framework with excellent selective CO_2 capture performance," *J. Mater. Chem. A*, vol. 6, pp. 6370–6375, 2018.

[328] G. Wang, Y. Onyshchenko, N. De Geyter, R. Morent, K. Leus, and P. Van Der Voort, "Straightforward preparation of fluorinated covalent triazine frameworks with significantly enhanced carbon dioxide and hydrogen adsorption capacities," *Dalton Trans.*, vol. 48, pp. 17612–17619, 2019.

[329] H. Veldhuizen, A. Vasileiadis, M. Wagemaker, T. Mahon, D. P. Mainali, L. Zong, S. van der Zwaag, and A. Nagai, "Synthesis, characterization, and CO_2 uptake of mellitic triimide-based covalent organic frameworks," *J. Polym. Sci. Part A Polym. Chem.*, vol. 57, pp. 2373–2377, 2019.

[330] J. Lee, H. Lim, J. Park, M.-S. Kim, J.-W. Jung, J. Kim, and I.-D. Kim, "Fluorine-rich covalent organic framework to boost electrochemical kinetics and storages of K^+ ions for potassium-ion battery," *Adv. Energy Mater.*, vol. 13, p. 2300442, 2023.

[331] L. Schaper and R. Schmid, "Simulating the structural phase transitions of metal-organic frameworks with control over the volume of nanocrystallites," *Commun. Chem.*, vol. 6, p. 233, 2023.

[332] S. Canossa, Z. Ji, C. Gropp, Z. Rong, E. Ploetz, S. Wuttke, and O. M. Yaghi, "System of sequences in multivariate reticular structures," *Nat. Rev. Mater.*, vol. 8, pp. 331–340, 2023.

[333] M. W. Deem, R. Pophale, P. A. Cheeseman, and D. J. Earl, "Computational discovery of new zeolite-like materials," *J. Phys. Chem. C*, vol. 113, pp. 21353–21360, 2009.

[334] S. Majumdar, S. M. Moosavi, K. M. Jablonka, D. Ongari, and B. Smit, "Diversifying databases of metal organic frameworks for high-throughput computational screening," *ACS Appl. Mater. Interfaces*, vol. 13, pp. 61004–61014, 2021.

[335] G. Donval, C. Hand, J. Hook, E. Dupont, M. S. Landman, M. Freitag, M. Lennox, and T. Düren, "Autonomous exploration and identification of high performing adsorbents using active learning," *ChemRxiv (Theoretical and Computational Chemistry).* May 10, 2021, DOI: 10.26434/chemrxiv.14555706.v1.

[336] A. Sriram, S. Choi, X. Yu, L. M. Brabson, A. Das, Z. Ulissi, M. Uyttendaele, A. J. Medford, and D. S. Sholl, "The Open DAC 2023 dataset and challenges for sorbent discovery in direct air capture," *arXiv (Materials Science).* November 27, 2023, DOI: 10.48550/arXiv.2311.00341.

[337] C. Chen and S. P. Ong, “A universal graph deep learning interatomic potential for the periodic table,” *Nat. Comput. Sci.*, vol. 2, pp. 718–728, 2022.

[338] I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, W. J. Baldwin, N. Bernstein, A. Bhowmik, S. M. Blau, V. Cărare, J. P. Darby, S. De, F. D. Pia, V. L. Deringer, R. Elijošius, Z. El-Machachi, E. Fako, A. C. Ferrari, A. Genreith-Schriever, J. George, R. E. A. Goodall, C. P. Grey, S. Han, W. Handley, H. H. Heenen, K. Hermansson, C. Holm, J. Jaafar, S. Hofmann, K. S. Jakob, H. Jung, V. Kapil, A. D. Kaplan, N. Karimitari, N. Kroupa, J. Kullgren, M. C. Kuner, D. Kuryla, G. Liepuoniute, J. T. Margraf, I.-B. Magdău, A. Michaelides, J. H. Moore, A. A. Naik, S. P. Niblett, S. W. Norwood, N. O’Neill, C. Ortner, K. A. Persson, K. Reuter, A. S. Rosen, L. L. Schaaf, C. Schran, E. Sivonxay, T. K. Stenczel, V. Svahn, C. Sutton, C. van der Oord, E. Varga-Umbrich, T. Vegge, M. Vondrák, Y. Wang, W. C. Witt, F. Zills, and G. Csányi, “A foundation model for atomistic materials chemistry,” *arXiv (Chemical Physics).* December 29, 2023, DOI: 10.48550/arXiv.2401.00096.

[339] M. Eckhoff and J. Behler, “From molecular fragments to the bulk: development of a neural network potential for MOF-5,” *J. Chem. Theory Comput.*, vol. 15, pp. 3793–3809, 2019.

[340] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, “Gaussian 16 Revision C.01,” 2016. Gaussian Inc. Wallingford CT.

[341] T. Verstraelen, P. Tecmer, F. Heidar-Zadeh, C. E. González-Espinoza, M. Chan, T. D. Kim, K. Boguslawski, S. Fias, S. Vandenbrande, D. Berrocal, and P. W. Ayers, “Horton 2.1.1,” 2017.

[342] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, "LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales," *Comput. Phys. Commun.*, vol. 271, p. 108171, 2022.

[343] D. Dubbeldam, S. Calero, D. E. Ellis, and R. Q. Snurr, "RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials," *Mol. Simul.*, vol. 42, pp. 81–101, 2016.

[344] A. Ghysels, T. Verstraelen, K. Hemelsoet, M. Waroquier, and V. Van Speybroeck, "TAMkin: a versatile package for vibrational analysis and chemical kinetics," *J. Chem. Inf. Model.*, vol. 50, pp. 1736–1750, 2010.