SAND2022-15643C

Sandia National Laboratories

MOFX-DB: An online database of computational adsorption data for nanoporous materials

N. Scott Bobbitt

Northwestern University

Sandia National Laboratories

Northwestern University

AIChE Annual Meeting

Molecular and Data Science Modeling of Adsorption

Phoenix, AZ

November 16, 2022, 4:00pm Convention Center N-131C

U.S. DEPARTMENT OF ENERGY NNSA National Nuclear Security Administration

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Acknowledgements

U.S. DEPARTMENT OF ENERGY
Northwestern University
Sandia National Laboratories
NIST National Institute of Standards and Technology U.S. Department of Commerce
UNIVERSITY OF MINNESOTA

Randy Snurr
Northwestern

Dan Siderius
NIST

Ilja Siepmann
U Minnesota

Ben Bucior
PhD Northwestern
Now at CSL Behring

Haoyuan Chen
U Texas-Rio Grand Valley

Kaihang Shi
Northwestern

Siderius, Daniel (Fed) <daniel.siderius@nist.gov>
to me

Sep
7
Fri

Siderius/Bucior Video Conference
View on Google Calendar

When Fri Sep 7, 2018 9am – 10am (MDT)
Where TBD: Zoom or Google Hangout
Who Daniel Siderius*

Over 4 years in the making!

Zhao Li
Northwestern

Andrew Sun
U Minnesota

Nate Tracy-Amoroso
Northwestern
Now at Microsoft

Julia Merlin
Georgia Tech

High-throughput modeling generates a lot of data...

Millions of data points about adsorption have been generated in Snurr group in the last decade.

Cost huge amount of computer time and human effort.

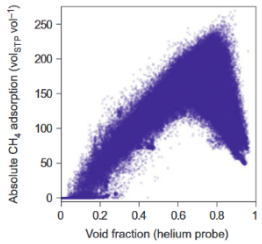
Methane storage

nature chemistry ARTICLES
PUBLISHED ONLINE: 6 NOVEMBER 2011 | DOI: 10.1038/NCHEM.1192

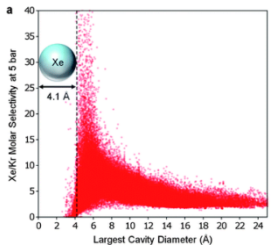
Large-scale screening of hypothetical metal–organic frameworks

Christopher E. Wilmer[1], Michael Leaf[1], Chang Yeon Lee[2], Omar K. Farha[2], Brad G. Hauser[2], Joseph T. Hupp[2] and Randall Q. Snurr[1]*

Wilmer et al, Nature Chemistry, 2012, **4**, 83–89

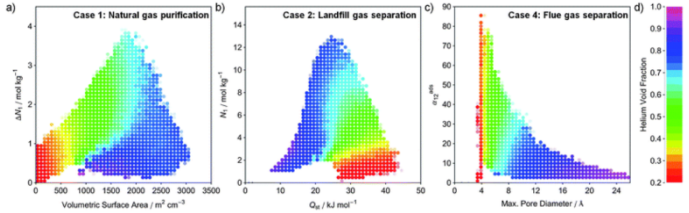


Xe/Kr separations



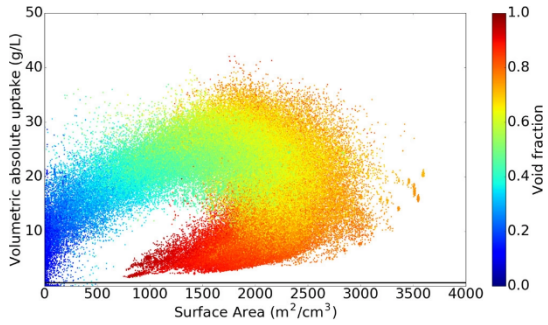
Sikora et al, Chem. Sci., 2012, **3**, 2217-2223

Carbon capture



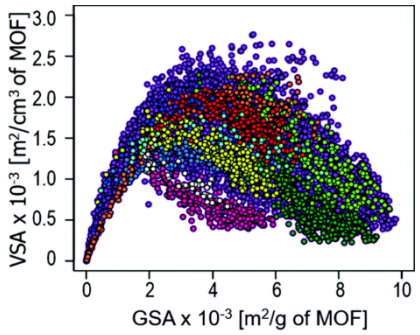
Wilmer et al, Energy Environ. Sci., 2012, **5**, 9849-9856

Hydrogen storage



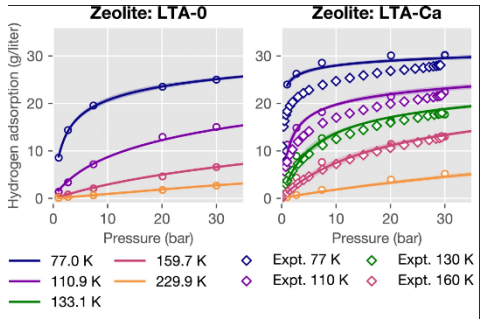
Bobbitt et al, J. Phys. Chem. C, 2016, **120**, 27328–27341

More hydrogen storage



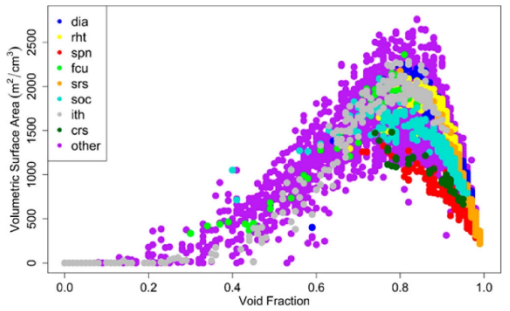
Gómez-Gualdrón & Colón et al, Energy Environ. Sci., 2016, **9**, 3279-3289

Hydrogen in zeolites



Sun et al, Science Advances, 2021, **7**, eabg3983

Methane storage, Xe/Kr separations

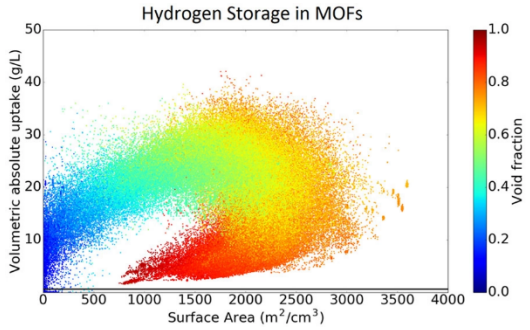


Colón & Gómez-Gualdrón, Cryst. Growth Des., 2017, **17**, 5801-5810

But what now?

This data required a lot of effort to generate. Can we get more use out of it?

Example: Ben Bucior used data from J. Phys. Chem. Paper to train machine learning model for hydrogen storage.



Bobbitt et al, J. Phys. Chem. C, 2016, **120**, 27328–27341

Training Data

Bucior et al, Mol. Syst. Des. Eng., 2019, 4, 162

Data can still be used for machine learning or finding materials for specific applications.

Problem

Much of the data published in literature is either not available or not in a standardized form.

Might be published in Supporting Information or “available upon request”

Students graduate

Postdocs take other jobs

Professors retire

Institutional knowledge is lost over time.

One study[1] reached out to almost 1800 groups to ask for data and less than 7% actually sent it.

Probability of finding valid email address for corresponding author fell by 7% per year of article age.[2]

1. Gabelica, M.; Bojčić, R.; Puljak, L., Many Researchers Were Not Compliant with Their Published Data Sharing Statement: Mixed-Methods Study. *Journal of Clinical Epidemiology* **2022**.
2. Vines, The Availability of Research Data Declines Rapidly with Article Age. *Current Biology* **2014**, *24*, 94-97

FAIR Guidelines

https://www.go-fair.org/fair-principles/

FAIR guiding principles for scientific data management and stewardship published in 2016.

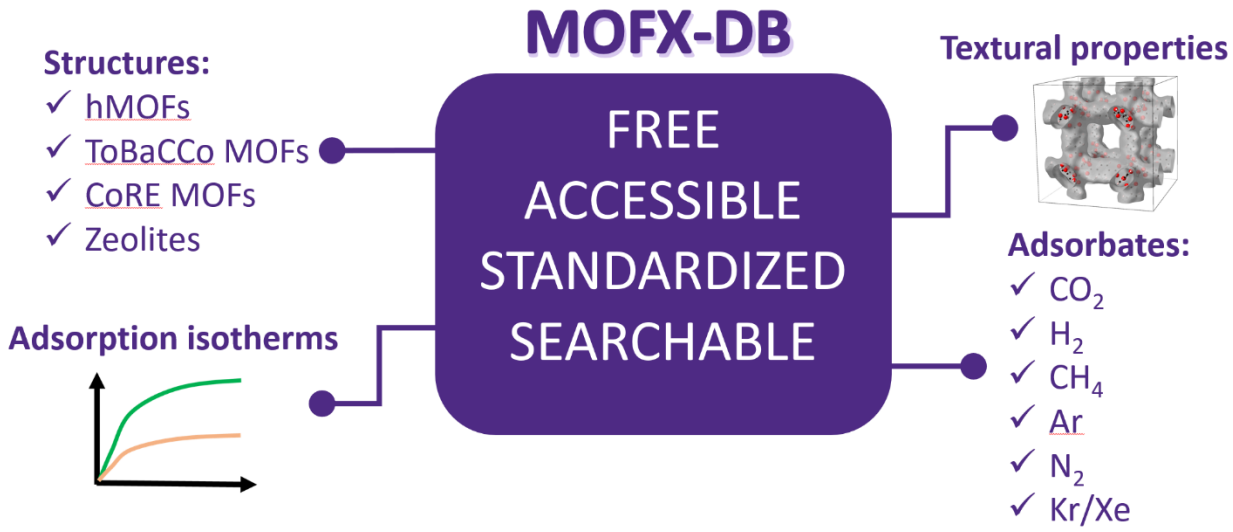
1. **F**indable: data must be easy to find for humans and computers
2. **A**ccessible: data must be easy to access, free, and open
3. **I**nteroperable: data should be able to operate with common applications and tools for analysis
4. **R**eusable**:** metadata and details should be well-described so data can be replicated and reused

Scientific Data **volume 3**, Article number: 160018 (2016)

MOFX-DB

Introducing MOFX-DB!

1. Online database of simulated adsorption data in 160,000 MOFs and 286 zeolites.
2. Over 3 million data points including H_2, CH_4, CO_2, Xe, Kr, N_2, and Ar
3. Some previously published and some new data
4. Contains all computational details and metadata (force field, GCMC parameters)
5. Textural properties for all structures
6. DOI link to original publications
7. Data stored in a JavaScript Object Notation (JSON) file that is fully compatible with NIST-ISODB database



Paper currently under review at
J. Chemical & Engineering Data

User friendly GUI

Can search by textural properties like pore sizes and surface area

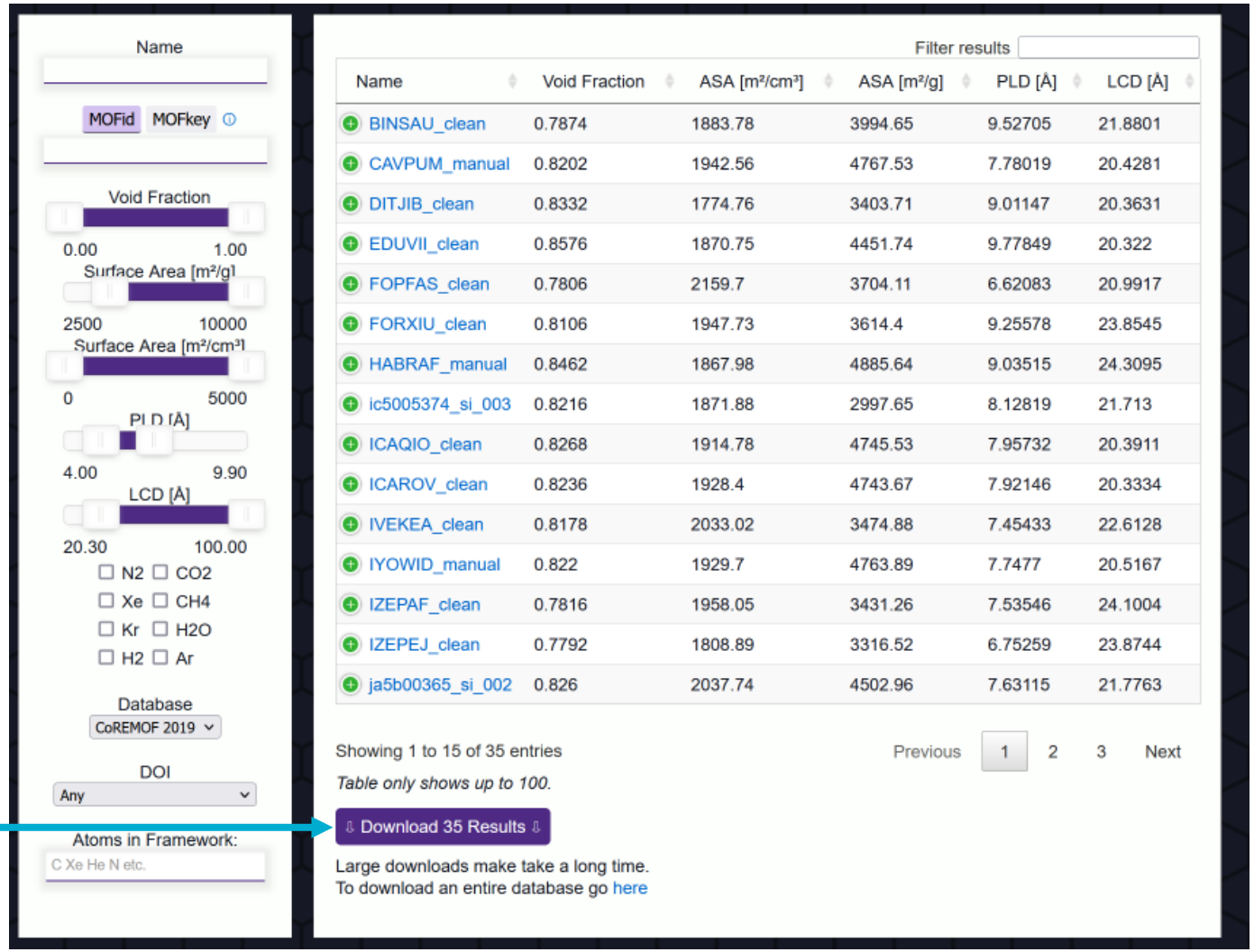
Need a MOF with

PLD between 4 – 10 Å

LCD > 20 Å

At least 2500 m^2/g surface area

Use sliders to search and find downloadable list of structures.



Click for more details

Name

MOFid MOFkey

Void Fraction
0.00 1.00

Surface Area [m²/g]
2500 10000

Surface Area [m²/cm³]
0 5000

PLD [Å]
4.00 9.90

LCD [Å]
20.30 100.00

N2 CO2
Xe CH4
Kr H2O
H2 Ar

Database
CoREMOF 2019

DOI
Any

Atoms in Framework:
C Xe He N etc.

Filter results

Name	Void Fraction	ASA [m²/cm³]	ASA [m²/g]	PLD [Å]	LCD [Å]
BINSAU_clean	0.7874	1883.78	3994.65	9.52705	21.8801
CAVPUM_manual	0.8202	1942.56	4767.53	7.78019	20.4281
DITJIB_clean	0.8332	1774.76	3403.71	9.01147	20.3631
EDUVII_clean	0.8576	1870.75	4451.74	9.77849	20.322
FOPFAS_clean	0.7806	2159.7	3704.11	6.62083	20.9917
FORXIU_clean	0.8106	1947.73	3614.4	9.25578	23.8545
HABRAF_manual	0.8462	1867.98	4885.64	9.03515	24.3095
ic5005374_si_003	0.8216	1871.88	2997.65	8.12819	21.713
ICAQIO_clean	0.8268	1914.78	4745.53	7.95732	20.3911
ICAROV_clean	0.8236	1928.4	4743.67	7.92146	20.3334
IVEKEA_clean	0.8178	2033.02	3474.88	7.45433	22.6128
IYOWID_manual	0.822	1929.7	4763.89	7.7477	20.5167
IZEPAF_clean	0.7816	1958.05	3431.26	7.53546	24.1004
IZEPEJ_clean	0.7792	1808.89	3316.52	6.75259	23.8744
ja5b00365_si_002	0.826	2037.74	4502.96	7.63115	21.7763

Showing 1 to 15 of 35 entries

Previous 1 2 3 Next

Table only shows up to 100.

⇩ Download 35 Results ⇩

Large downloads make take a long time.
To download an entire database go here

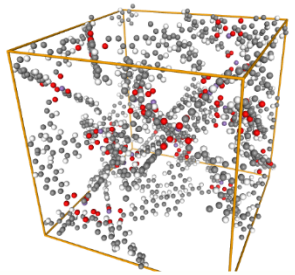
MOFXDB API Databases

Loading units: Isotherm Default

Pressure Units: Isotherm Default

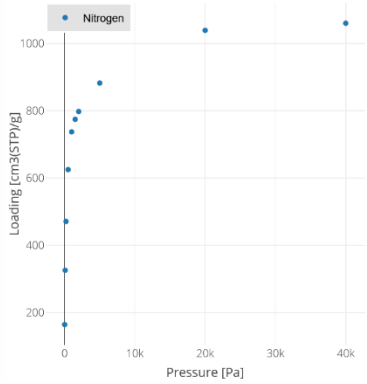
BINSAU_clean

MOFid	
MOFkey	
Database	CoREMOF 2019
Void Fraction	0.7874
ASA [m²/g]	3994.65
ASA [m²/cm³]	1883.78
PLD [Å]	9.52705
LCD [Å]	21.8801
Atoms in MOF	
Atomic Mass of Cif	10688.0
Cif Volume [Å³]	37635.1
Cif	Download

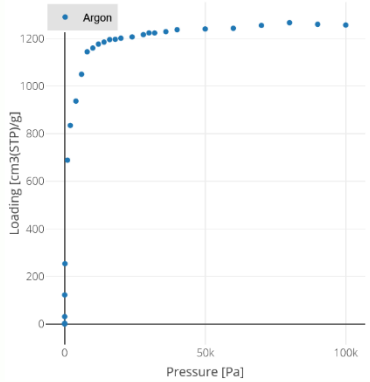


Isotherms View JSON

77K, digitized by Kaihang Shi, Zhao Li



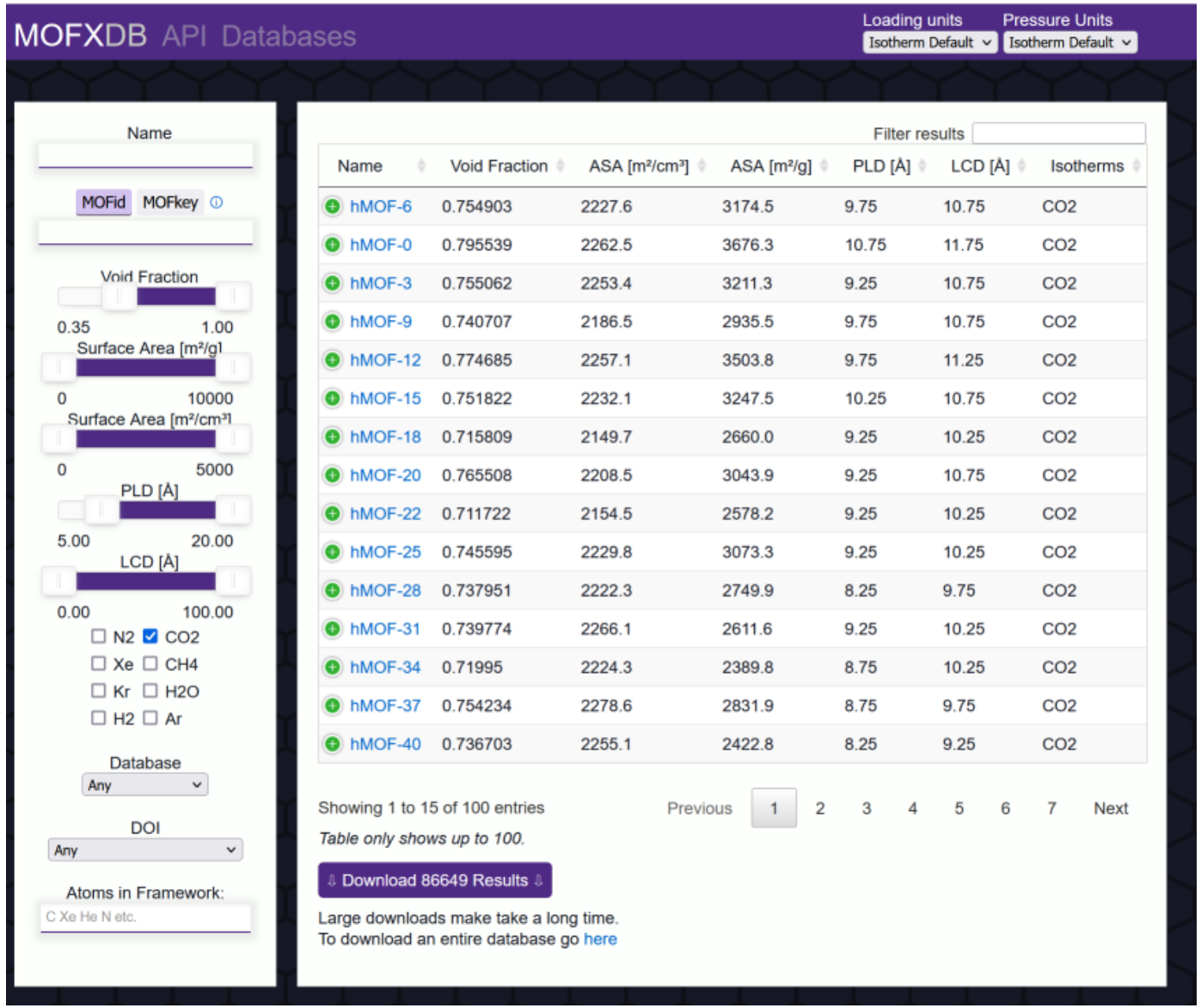
87K, digitized by Kaihang Shi, Zhao Li



Search by Adsorbate

Check boxes to search for data about particular adsorbates

Search for CO2 data in MOFs/zeolites with PLD at least 5.0 Å and void fraction greater than 0.35



Download JSON files and CIFs

Name	Type	Compressed size	Password pr...	Size	Ratio	Date modified
FORXIU_clean-(id:176567).cif	CIF File	26 KB	No	204 KB	88%	10/31/2022 11:05 PM
FORXIU_clean-(id:176567).json	JSON File	29 KB	No	216 KB	87%	10/31/2022 11:05 PM
HABRAF_manual-(id:177096).cif	CIF File	10 KB	No	48 KB	81%	10/31/2022 11:05 PM
HABRAF_manual-(id:177096).json	JSON File	12 KB	No	57 KB	81%	10/31/2022 11:05 PM
ic5005374_si_003_clean-(id:177644).cif	CIF File	10 KB	No	65 KB	85%	10/31/2022 11:05 PM
ic5005374_si_003_clean-(id:177644).json	JSON File	13 KB	No	75 KB	84%	10/31/2022 11:05 PM
ICAQIO_clean-(id:177695).cif	CIF File	4 KB	No	33 KB	89%	10/31/2022 11:05 PM
ICAQIO_clean-(id:177695).json	JSON File	6 KB	No	43 KB	87%	10/31/2022 11:05 PM
ICAROV_clean-(id:177698).cif	CIF File	4 KB	No	33 KB	89%	10/31/2022 11:05 PM
ICAROV_clean-(id:177698).json	JSON File	6 KB	No	43 KB	87%	10/31/2022 11:05 PM
IVEKEA_clean-(id:178021).cif	CIF File	17 KB	No	138 KB	88%	10/31/2022 11:05 PM
IVEKEA_clean-(id:178021).json	JSON File	20 KB	No	149 KB	88%	10/31/2022 11:05 PM
IYOWID_manual-(id:178130).cif	CIF File	4 KB	No	33 KB	89%	10/31/2022 11:05 PM
IYOWID_manual-(id:178130).json	JSON File	6 KB	No	43 KB	87%	10/31/2022 11:05 PM
IZEPAF_clean-(id:178148).cif	CIF File	5 KB	No	40 KB	88%	10/31/2022 11:05 PM
IZEPAF_clean-(id:178148).json	JSON File	7 KB	No	50 KB	87%	10/31/2022 11:05 PM
IZEPEJ_clean-(id:178149).cif	CIF File	4 KB	No	27 KB	88%	10/31/2022 11:05 PM
IZEPEJ_clean-(id:178149).json	JSON File	6 KB	No	36 KB	86%	10/31/2022 11:05 PM
ja5b00365_si_002_clean-(id:178263).cif	CIF File	7 KB	No	45 KB	87%	10/31/2022 11:05 PM
ja5b00365_si_002_clean-(id:178263).json	JSON File	9 KB	No	56 KB	85%	10/31/2022 11:05 PM
ja512973b_si_003_auto-(id:185198).cif	CIF File	10 KB	No	40 KB	78%	10/31/2022 11:05 PM
ja512973b_si_003_auto-(id:185198).json	JSON File	12 KB	No	52 KB	79%	10/31/2022 11:05 PM
ja809985t_si_002_clean-(id:178283).cif	CIF File	17 KB	No	94 KB	83%	10/31/2022 11:05 PM
ja809985t_si_002_clean-(id:178283).json	JSON File	19 KB	No	105 KB	83%	10/31/2022 11:05 PM
ja5111317_ja5111317_si_002_auto-(id:18...	CIF File	9 KB	No	40 KB	78%	10/31/2022 11:05 PM
ja5111317_ja5111317_si_002_auto-(id:18...	JSON File	11 KB	No	52 KB	79%	10/31/2022 11:05 PM
JOZWOM_clean-(id:178459).cif	CIF File	4 KB	No	30 KB	88%	10/31/2022 11:05 PM
JOZWOM_clean-(id:178459).json	JSON File	6 KB	No	39 KB	86%	10/31/2022 11:05 PM

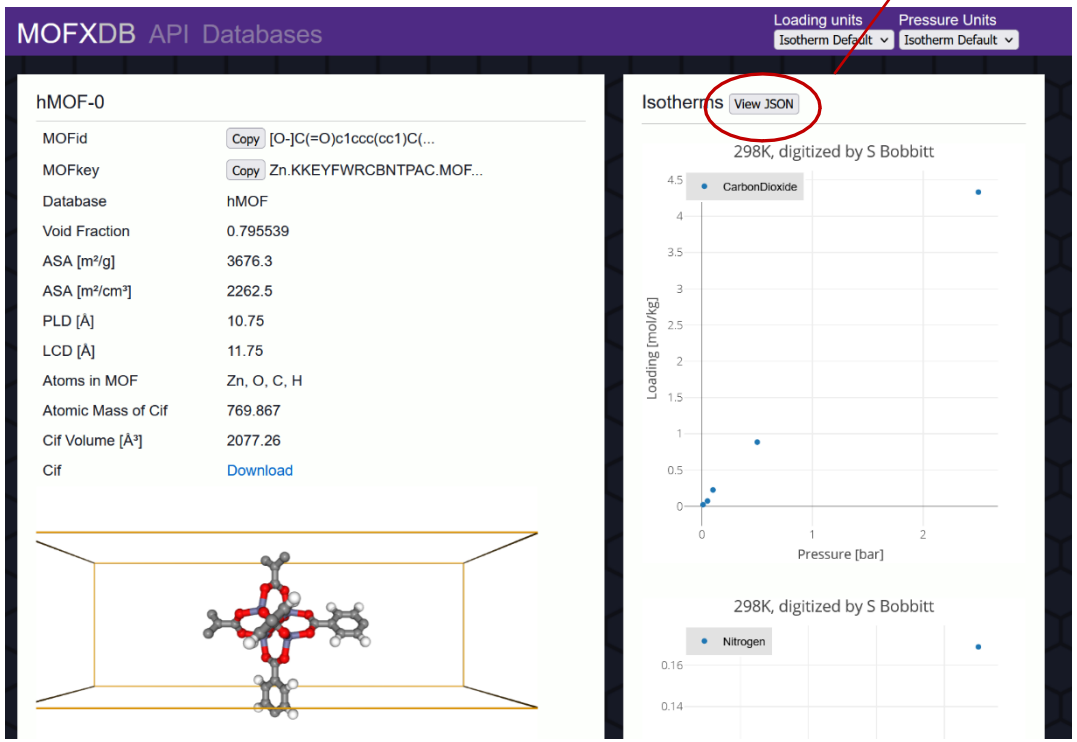
JSON file format

Single JSON file contains all adsorption data, simulation details, and meta-data (RASPA input files, cif)

Completely compatible with NIST adsorption database

NIST ISODB: **https://adsorption.nist.gov/isodb**

Dan Siderius, NIST




```
id: 15339
mofid: "[O-]C(=O)c1ccc(cc1)C(=O)[O-].[Zn][O]([Zn])([Zn])[Zn] MOFid-v1.pcu.cat0"
mofkey: "Zn.KKEYFWRCBNTPAC.MOFkey-v1.pcu\n"
hashkey: "eb66082e068e937a37608f1ebadd35947149ebd81c93a9fb14"
name: "hMOF-0"
void_fraction: 0.795539
surface_area_m2g: 3676.3
surface_area_m2cm3: 2262.5
pld: 10.75
lcd: 11.75
pxrd: null
pore_size_distribution: null
database: "hMOF"
batch_number: null
elements:
  0:
    symbol: "Zn"
    name: "Zinc"
  1:
    symbol: "O"
    name: "Oxygen"
  2:
    symbol: "C"
    name: "Carbon"
  3:
    symbol: "H"
    name: "Hydrogen"
cif: "data_functionalizedCrystal\n_audit_creation_method\t'MofGen! by Chris W
  \n_cell_angle_gamma\t90.016641\nloop_\n_atom_site_label\n_atom_site_type_
  \nO7\tO\t1.582813\t-0.570635\t-0.271317\nO8\tO\t1.453911\t-0.271191\t-0.4
  \t-0.740891\t-0.441609\nO17\tO\t1.453785\t-0.740834\t-0.570649\nC18\tC\t1
  \nC26\tC\t1.585632\t-0.950617\t-0.439056\nC27\tC\t1.585700\t-1.060341\t-0
  \t-0.115206\t-0.506237\nC36\tC\t1.451328\t-0.438653\t-0.951848\nC37\tC\t1
  \nH45\tH\t1.396672\t-0.383891\t-1.106240\nH46\tH\t1.396543\t-1.105022\t-0
url: "/mofs/15339.json"
```


We encourage the community to adopt this format as a standard for sharing adsorption data.

Compatible with Adsorption Information Format!

Can readily convert AIF to JSON which is amenable to cheminformatics, ML tools, etc.

Includes all simulation details

Includes original RASPA input files with all information (GCMC cycles, temp, cut off radius, etc.)

Lists force field info for adsorbents and adsorbates

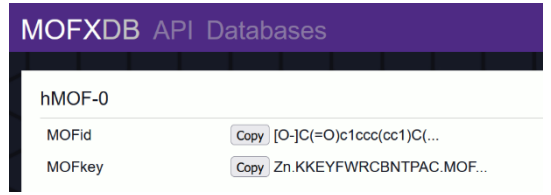
Contains DOI and link for original publications


```
▼ simin:                "SimulationType                MonteCarlo\nNumberOfCycles                2000\nNumberOfInitializationCycles  2000
                        0.8000\nExternalTemperature 77\nExternalPressure 2E5\n\n\nComponent 0 MoleculeName            hydrogen\n
                        CreateNumberOfMolecules  0\n"
  DOI:                  "10.1021/acs.jpcc.6b08729"
▼ doi_url:              "https://pubs.acs.org/doi/abs/10.1021/acs.jpcc.6b08729?casa_token=7Dx9VmjlRIkAAAAA:CbUhQ50anABy5Q_YY-QUM-aA1GCi3u
  date:                 "2021-57-03"
  temperature:          77
  adsorbent_forcefield: "UFF"
  molecule_forcefield:  "Darkrim-Levesque"
```


All of the data in MOFX-DB should be completely reproducible with the information available in the JSON file

MOFkey & MOFid

Ben Bucior



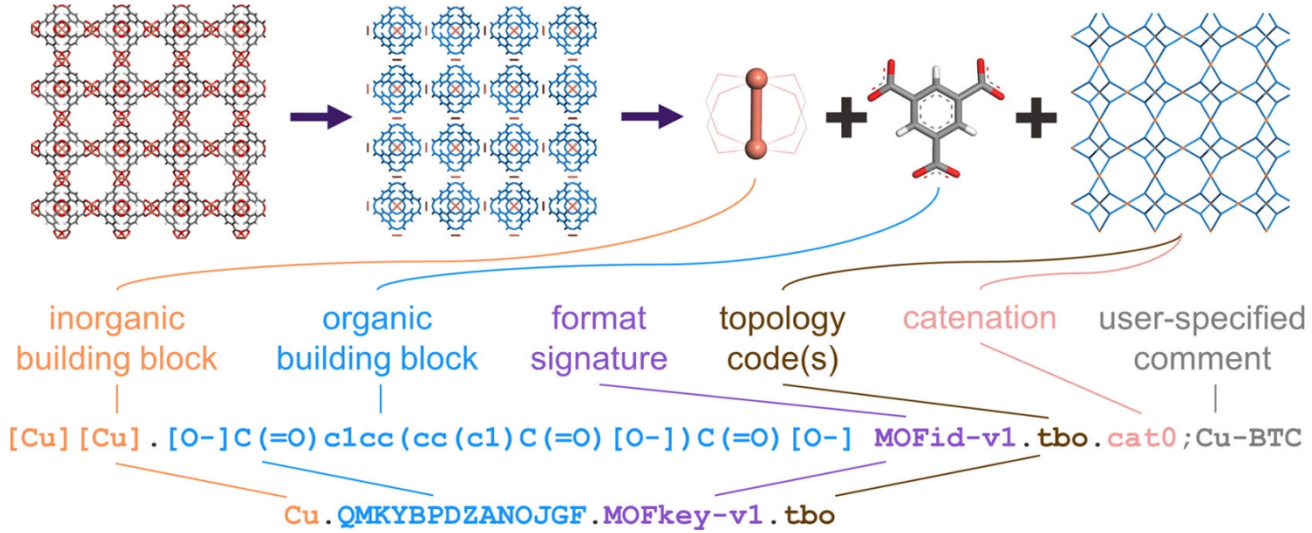
https://snurr-group.github.io/web-mofid/

Bucior et al, *Crystal Growth & Design*, **19**, 6682-6697, 2019

MOFkey and MOFid are ways to generate unique descriptors for MOF structures.

MOFid: based on SMILES

MOFkey: based on InChIKeys



MOFX-DB contains unique MOFkey and MOFid strings for every structure.

Can search by MOFkey or MOFid to find MOFs.

Can be used for cheminformatics analysis and machine learning.

Python API

https://mof.tech.northwestern.edu/api

We have an easy to use Python package: mofdb_client

Can search and fetch specific data automatically, download files, and feed data into your machine learning workflow

Use the package

Documentation on Github here

```
$ python3 -m pip install mofdb_client
```

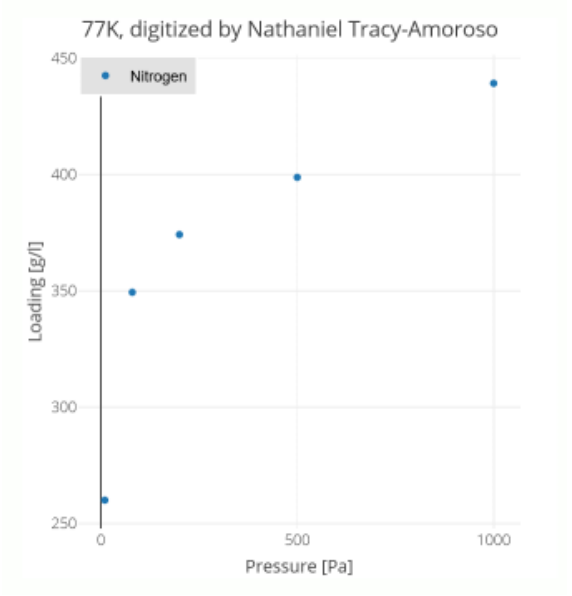
```
from mofdb_client import fetch
for mof in fetch(vf_min=0.5, vf_max=0.99):
    print(f"Mof {mof.name} has {len(mof.isotherms)} isotherms and elements {[str(e) for e in mof.elements]}")
    print(f"This mof's cif file starts with: '{mof.cif.splitlines()[1]}'")
```

```
Mof UTEWUM_clean has 2 isotherms and elements ['Cu', 'H', 'C', 'N']
This mof's cif file starts with: '_audit_creation_date              2014-07-02'
Mof ZECKID_clean has 1 isotherms and elements ['Cu', 'H', 'C', 'N', 'O']
This mof's cif file starts with: '_audit_creation_date              2014-07-02'
Mof AQOMAW_clean has 2 isotherms and elements ['N', 'C', 'H', 'Co', 'Cl', 'O']
This mof's cif file starts with: '_cell_length_a       18.8345'
Mof AQOLOJ_clean has 2 isotherms and elements ['N', 'C', 'H', 'Co', 'Cl']
This mof's cif file starts with: '_cell_length_a       18.794'
Mof SENWOZ_clean has 2 isotherms and elements ['Zn', 'H', 'C', 'O']
This mof's cif file starts with: '_audit_creation_date              2014-07-02'
Mof IYUCIQ_clean has 2 isotherms and elements ['Dy', 'O', 'N', 'C', 'H']
```

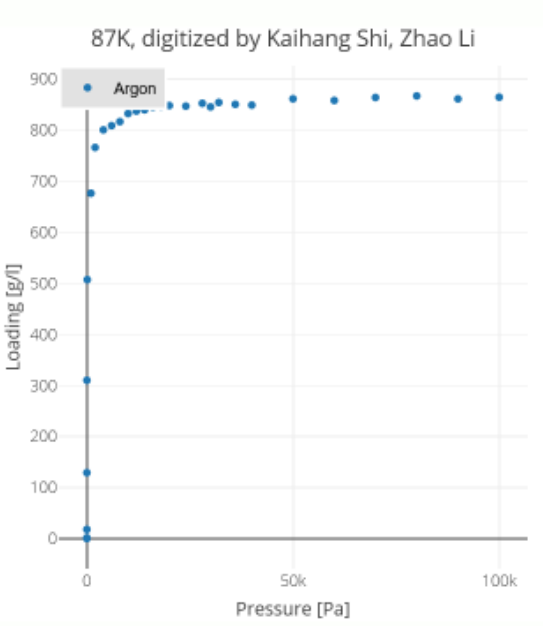
Nitrogen and Argon Isotherms

Database contains simulated nitrogen (77K) and argon (87K) isotherms for all CoREMOF structures.

Can be used as a guide for experimental activation.

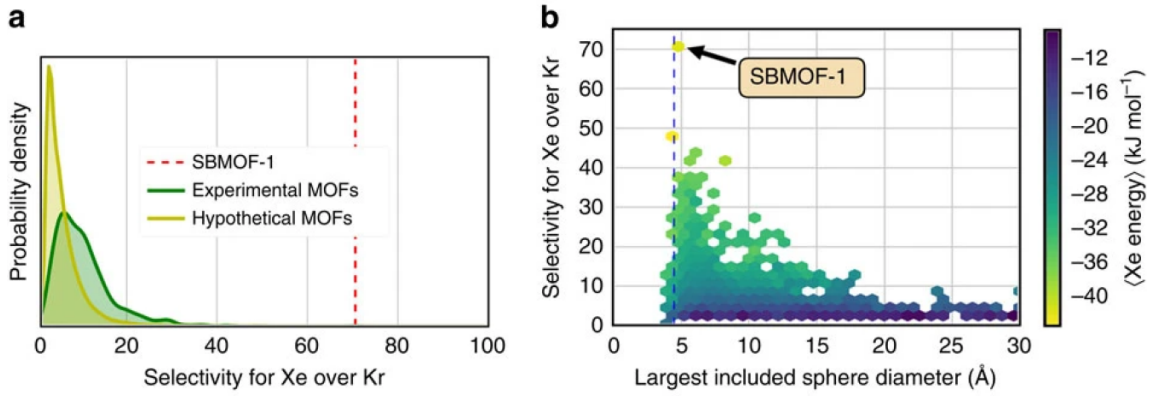


Nitrogen
77 K



Argon
87 K

Finding MOFs with good performance for Xe/Kr separation



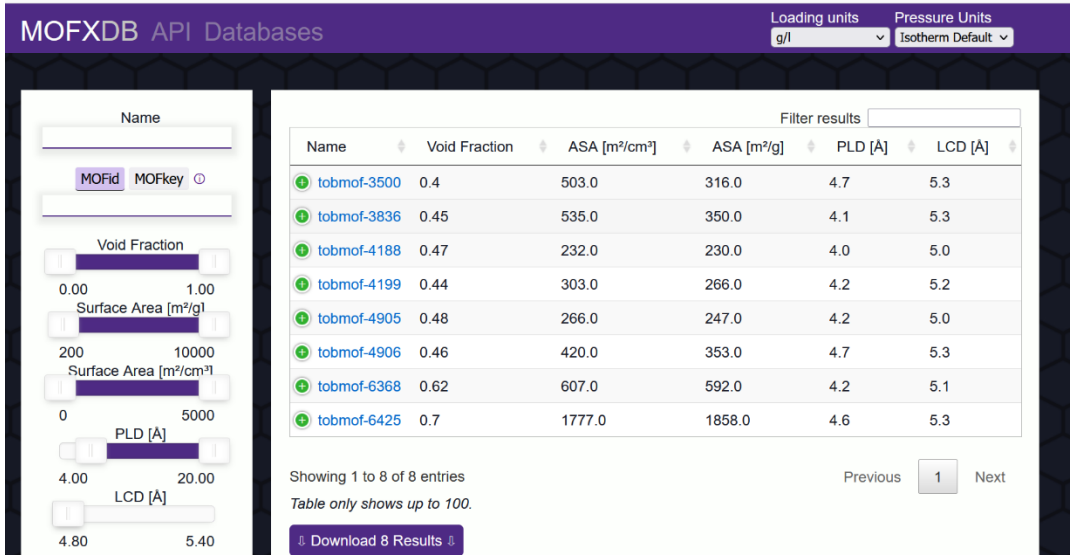
Banerjee et al, Nature Communications, 7, 11831, 2016

Banerjee found the best Xe over Kr selectivity in MOFs with LCD near 5.1 Å

SBMOF-1 has high selectivity.

Relatively low capacity (low surface area)

Search for LCD around 5.1 Å, PLD > 4.0 Å, and greater than 200 m^2/g SA



We find 8 MOFs with similar properties.

Some of them have comparable selectivity and higher capacity than SBMOF-1!

Name	Void Fraction	SA (m^2/g)	PLD (Å)	LCD (Å)	Xe (mol/kg)	Kr (mol/kg)	Selectivity
SBMOF-1*	0.25	145	4.2	5.1	1.2 (0.2 bar)	0.9 (0.8 bar)	16
tobmof-4188**	0.47	230	4	5	3.9	1	15.6
tobmof-4906**	0.46	353	4.7	5.3	2.5	0.8	12.5

*SBMOF-1: Banerjee, Nat. Comm. 2016
**tobmofs: Colon Cryst. Growth Des 2017

Multipurpose MOFs

Previous studies that generated this data focused on finding the **best** candidate for a specific application.

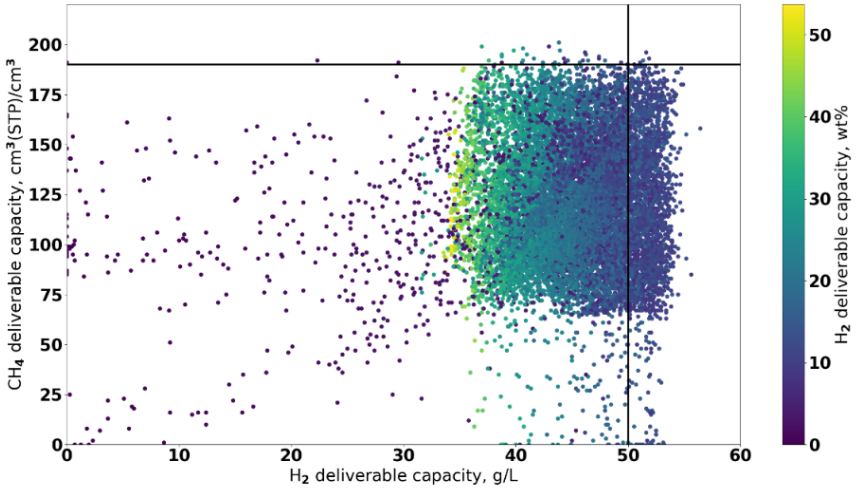
We can use data from multiple studies to find MOFs that are good for multiple applications!

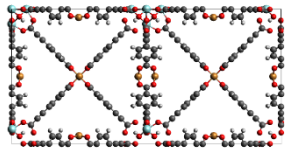
Example: Methane and hydrogen storage

Targets:

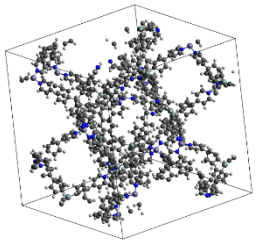
190 $cm^3(STP)/cm^3$ methane capacity

50 g/L hydrogen capacity

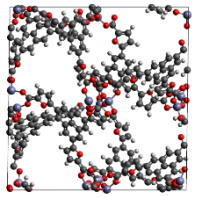




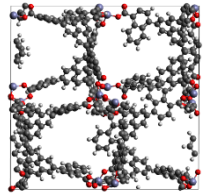
tobmof-3115



tobmof-4253



tobmof-7773



tobmof-7776

Void Frac	0.82-0.86
Pore size (LCD)	10.5-15.6 Å

Conclusions

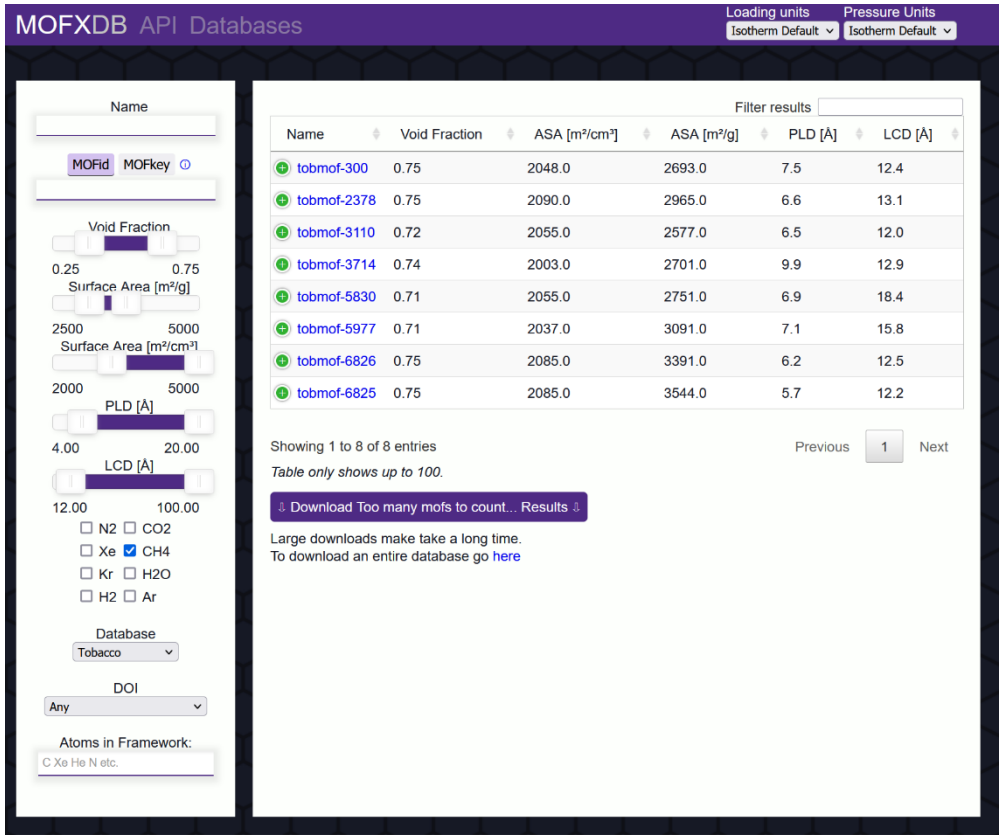
This work will (hopefully) appear soon in Journal of Chemical & Engineering Data

1. Millions of data points about adsorption in MOFs & zeolites now available
2. All simulation details and metadata available
3. Uses JSON format that is interoperable with NIST database. We hope this standard will be adopted by the community.
4. Python API to make searching and downloading data easy

Please download the data and go wild with machine learning!

URL: mof.tech.northwestern.edu

Search Commands



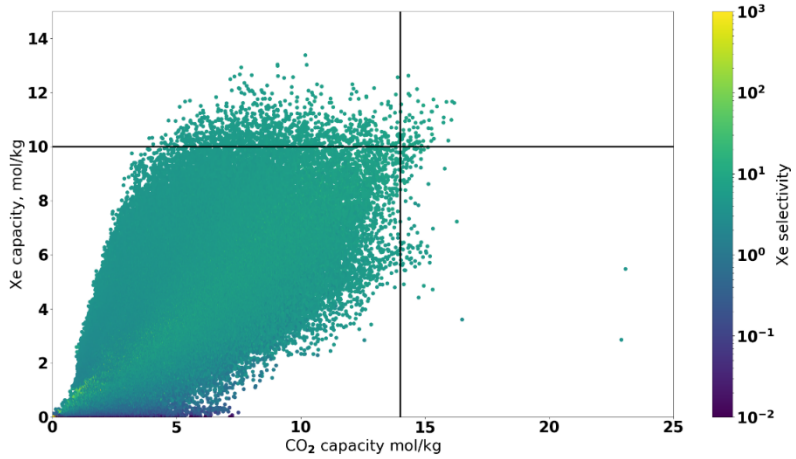
For example, to run the same search query from Figure 1, a program could issue a GET request as

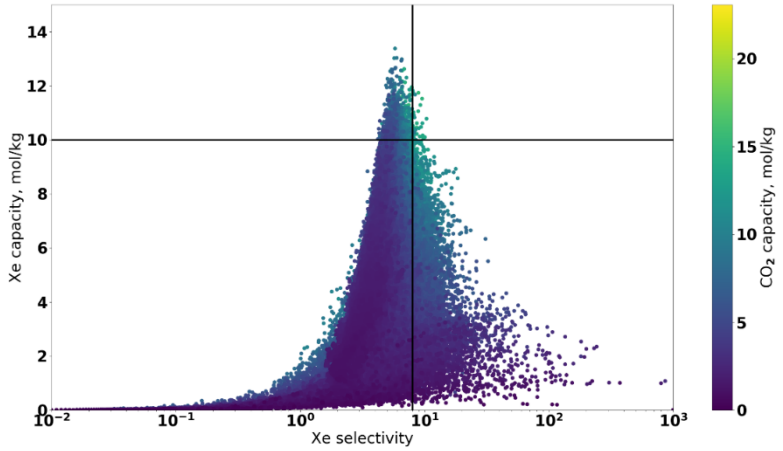
https://mof.tech.northwestern.edu/mofs.json?database=Tobacco&vf_min=0.25&vf_max=0.75&sa_m2g_min=2500&sa_m2g_max=5000&pld_min=4.0&lcd_min=12.0&gases=Methane

to encode all the relevant parameters and receive a filtered list of MOFs. Not all of the search criteria are required: the API will only filter by properties that are specified in the query. For example, to search the ToBaCCo database by surface area, a program could use a query of

https://mof.tech.northwestern.edu/mofs.json?database=Tobacco&sa_m2g_min=2500&sa_m2g_max=5000

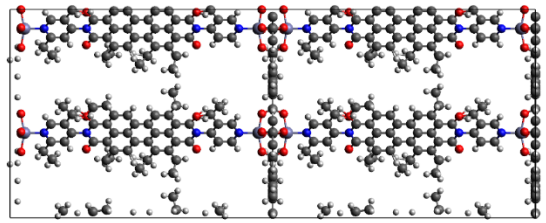
CO_2 capture and Xe/Kr separations



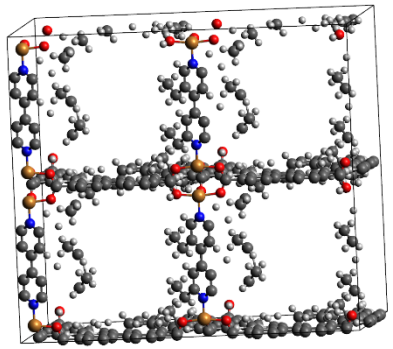


Find MOFs with good performance for both CO_2 capture (natural gas upgrading) and Xe/Kr separations

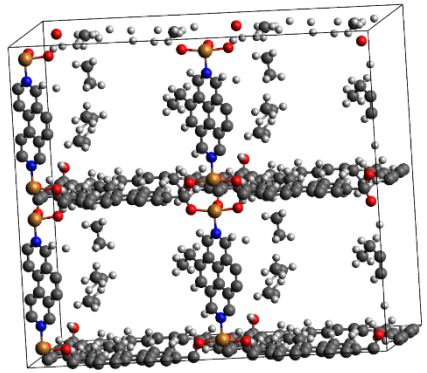
MOF name	Void Frac.	SA (m^2/g)	PLD (Å)	LCD (Å)	Xe Cap. (mol/kg)	Kr Cap. (mol/kg)	Xe select.	CO_2 Cap. (mol/kg)
hMOF-056488	0.80	3703	5.25	7.25	10.1	4.2	9.7	14.5
hMOF-36162	0.83	2754	8.25	9.75	10.4	4.9	8.4	14.7
hMOF-5067108	0.85	2842	8.25	9.75	10.1	4.9	8.2	14.6



hMOF-5056488



hMOF-36162



hMOF-5067108