



THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à Chimie ParisTech

**Origines microscopiques de la séparation xénon/krypton
dans les matériaux nanoporeux**

Microscopic origins of the xenon/krypton separation in
nanoporous materials

Présentée par

Emmanuel REN

Soutenance prévue le
XX Septembre 2023

École doctorale n°388

**Chimie Physique et
Chimie Analytique de
Paris Centre**

Spécialité

Chimie Physique

Composition du jury :

U Caroline MELLOT-DRAZNIÉKS Directrice de Recherche, Collège de France	<i>Présidente</i>
U Sofía CALERO Professeure, Universidad Pablo de Olavide	<i>Rapporteuse</i>
U Paul FLEURAT-LESSARD Professeur, Université de Bourgogne	<i>Rapporteur</i>
U Renaud DENOYEL Directeur de Recherche, Aix-Marseille Université	<i>Examineur</i>
U Alain FUCHS Professeur, PSL Université	<i>Examineur</i>
François-Xavier COUDERT Directeur de Recherche, Chimie ParisTech	<i>Directeur de thèse</i>




REMERCIEMENTS

En premier lieu, je voudrais adresser ici mes plus vifs remerciements

TABLE OF CONTENTS

General introduction	1
1 High-throughput computational screening of nanoporous materials	5
1.1 Introduction to the main screening tools	5
1.1.1 Databases	5
1.1.2 Simulation tools	7
1.1.3 Machine learning assisted screening.	7
1.2 A literature overview	8
1.2.1 Thermodynamic adsorption properties.	8
1.2.2 Transport adsorption properties	8
1.2.3 Non-adsorption properties	9
1.3 Consequences for xenon/krypton separation	9
1.3.1 Status quo	9
1.3.2 Future perspectives	9
2 Thermodynamic exploration of xenon/krypton separation	11
2.1 Preliminary analyses	11
2.1.1 Structure–selectivity relationships	11
2.1.2 Thermodynamic quantities correlations	11
2.2 Selectivity drop	11
2.2.1 Thermodynamic analyses	11
2.2.2 Detailed investigation	11
3 Adsorption molecular simulations	13
3.1 Standard simulation tools	13
3.1.1 Grand canonical monte carlo	13
3.1.2 Widom’s insertion	13
3.2 New algorithm development	13
3.2.1 Rapid Adsorption Enthalpy Surface Sampling (RAESS)	13
3.2.2 Grid Adsorption Energies Sampling (GrAES)	13
4 Adsorption properties prediction	15
4.1 Machine learning	15
4.1.1 Introduction	15
4.1.2 eXtreme Gradient Boosting	15
4.2 Ambient-pressure prediction	15
4.2.1 From infinite dilution to ambient pressure	15
4.2.2 Interpretation of the ML model	15
5 Transport properties	17
5.1 Computational simulations	17
5.1.1 Molecular dynamics	17
5.1.2 Fast kinetic Monte Carlo	17

5.2	ML modeling	17
6	Towards the next generation of screenings	19
6.1	Flexibility	19
6.1.1	Problem, literature	19
6.1.2	Snapshot.	19
6.2	Open Metal Sites.	19
6.2.1	Problem, literature	19
6.2.2	Perspectives	19
	General conclusions	21
<hr/> 		
	List of Publications	23
	Peer-reviewed papers	23
	Preprint	23
	Bibliography	25
	Résumé en français	31
	Introduction	31

GENERAL INTRODUCTION

Nanoporous materials are material

[\[Just a copy paste from last article\]](#)

Gas separation and purification are essential processes since they provide key reactants and inert gases for the chemical industry, as well as medical or food grade gases. Among them, we can find easily extractable or synthesizable molecules such as nitrogen, oxygen, carbon dioxide, noble gases, hydrogen, methane, or nitrous oxide. Moreover, gas separation is crucial in mitigating negative environmental impact at the end of industrial processes, such as facilities emitting green house gases (*e.g.* concrete or steel plants) or treating volatile radioactive wastes like ^{85}Kr . Cryogenic liquefaction or distillation is currently the mainstream technique to achieve industrial gas separation, while adsorbent beds made of nanoporous materials (activated alumina or zeolites) are mostly used as a less energy-intensive pre-purification system.[1]

A wider use of nanoporous materials could reduce the energy consumption of current separation processes since adsorption is way less energy intensive than liquefaction.[2] For instance, some prototypes involving beds of nanoporous materials have been developed for xenon/krypton separation to avoid employing cryogenic distillation.[3] For the process to be viable, materials need to perform even better and many studies focus on synthesizing ever more selective materials by leveraging all chemical intuitions around noble gas adsorption properties.[4–6] In order to speed the discovery process of novel materials with key properties, computational screening can identify factors explaining the performance and pre-select candidates for further experimental studies. As recently conceptualized by Lyu et al., a synergistic workflow combining computational discovery and experimental validation can push material discovery to the next stage.[7, 8] But to efficiently guide experimental discoveries, computational chemists are facing two major challenges: generating reliably more structures and evaluating them with fast and accurate models.

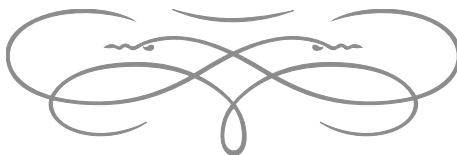
The number of nanoporous materials is potentially unlimited; for the metal–organic frameworks (MOFs) alone, over 90,000 structures have been synthesized [9] and 500,000 computationally constructed [10–12]. To deal with this ever increasing amount of structures, we need to design more efficient screening procedures as well as faster performance evaluation tools. To go beyond the time-consuming calculations over the whole dataset, computational chemists developed funnel-like screening procedures to reduce the need for expensive simulations and introduced machine learning (ML) models to replace them with faster evaluation tools.[13] To further improve the selectivity screening for Xe/Kr separation, we will need to design better performing structural and energy-based descriptors.

Simon et al. published one of the first articles on an ML-assisted screening approach for the separation of a Xe/Kr mixture extracted from the atmosphere.[14] Their model’s performance was highly relying on the Voronoi energy, which is basically an average of the interaction energies of a xenon atom at each Voronoi node.[15] To rationalize this increase in performance, we regarded this Voronoi energy as a faster proxy for the adsorption enthalpy. By comparing it to the standard Widom insertion, we found that although it is faster, it is less accurate; and we developed a more effective alternative, the

surface sampling (RAESS) using symmetry and non accessible volumes blocking.[16] Recently, Shi et al. used an energy grid to generate energy histograms as a descriptor for their ML model, which gives an exhaustive description of the infinitely diluted adsorption energies,[17] but can be computationally expensive.

All the approaches described above can have good accuracy in the prediction of low-pressure adsorption (i.e., in the limit of zero loading) but are not suitable for prediction of adsorption in the high-pressure regime, when the material is near saturation uptake. While this later task is routinely performed by Grand Canonical Monte Carlo (GCMC) simulations, there is a lack of methods at lower computational cost for high-throughput screening. To better frame our challenge, in this work we are essentially trying to predict the selectivity in the nanopores of a material at high pressure, where adsorbates are interacting with each other, while only having information on the interaction at infinite dilution. The comparison between the low and high pressure cases gives key information on the origin of the differences of selectivity. For instance, we previously showed that selectivity could drop between the low and ambient pressure cases in the Xe/Kr separation application, and it was mainly attributed to the presence of different pore sizes and potential reorganizations due to adsorbate-adsorbate interactions.[18]

This thesis presents my work on



$$1 + 1 = 2 \tag{1}$$



HIGH-THROUGHPUT COMPUTATIONAL SCREENING OF NANOPOROUS MATERIALS

1.1	Introduction to the main screening tools	5
1.1.1	Databases	5
1.1.2	Simulation tools	7
1.1.3	Machine learning assisted screening.	7
1.2	A literature overview	8
1.2.1	Thermodynamic adsorption properties.	8
1.2.2	Transport adsorption properties	8
1.2.3	Non-adsorption properties	9
1.3	Consequences for xenon/krypton separation	9
1.3.1	Status quo	9
1.3.2	Future perspectives	9



1.1 INTRODUCTION TO THE MAIN SCREENING TOOLS

1.1.1 Databases

[\[Ren2021\]](#)

In the past decade, large-scale computational screening studies have become an important part of the materials science innovation pipeline,[\[19, 20\]](#) trying to move beyond the serendipitous model of materials discovery.[\[21, 22\]](#) High-throughput computational discovery techniques are used in the generation of novel hypothetical structures for screening,[\[10, 11\]](#) as well as in trying to explore more in depth and more systematically the materials whose structure has already been published, in order to map their physical and chemical properties.[\[23–25\]](#) While the idea of large-scale exploration of materials is not new, and such databases — whether experimental or computational in the source of their data — have been around for several decades now,[\[9, 26, 27\]](#) this field has recently seen a rapid expansion enabled by several factors.

The first factor is the growth of public, open databases of materials structures (and sometimes properties).[\[28\]](#) To give only one example, projects like the Materials Genome Initiative have [\[29, 30\]](#) integrated theory, computation, synthesis, and characterization that led to the generation of vast materials datasets.[\[31, 32\]](#) Secondly, advances in the methods for construction of hypothetical structures for complex supramolecular assemblies have led to the creation of large-scale databases of hypothetical structures.[\[11, 33, 34\]](#) Thirdly, text and data mining are allowing to augment databases with content previously

thought not being machine-readable and indexable, for example by identifying unreported properties of materials in older scientific papers.[35, 36] Finally, the use of artificial intelligence techniques, such as statistical learning,[37] can enable in some cases by several orders of magnitude the scale of databases that can be screened.[38–40]

Focusing more on the larger process of digital materials discovery, and the role of large databases in materials research,[41] Boyd et al.[42] provide a broad review of the computational developments involved in the nanoporous materials genome research effort. They show that a large number of high-throughput screening studies have been targeted specifically at the question of gas phase separation, with a range of different adsorbate molecules going from noble gases, small molecules (both apolar and polar), to short alkanes and aromatics. Both Adil et al.[43] and Sturluson et al.[44] provide general reviews on the modelling in MOFs for gas separation and storage, and structure/separation relationships in MOFs for gas separation, featuring specific sections to the still open problem of Xe/Kr separation.

Chung et al. proposed a new Computation-Ready, Experimental MOF database (CoRE MOF 2019) containing over 14,000 cleaned structures.[45] The structures originated from the Cambridge Structural Database and Web of Science search. These structures went through a curation process: (i) by removing coordinates with low partial occupancies, (ii) by converting the structure to P1 symmetry, (iii) by removing free solvents (i.e., free solvent removed FSR), (iv) also removing bound solvent molecules (i.e., all solvent removed ASR), then (v) by restoring semi-automatically some disordered structures using a crystal generator. After this process, the structures are said to be “clean”.

In this study, we used the 12,020 structures of the CoRE MOF 2019-ASR (all solvent removed) database that are publicly available. Then, we extracted the non-disordered structures and the structures with a cell volume smaller than 20 nm^3 (to limit the overall calculation time). This resulted in a total of 9,668 remaining structures included in our systematic simulations to compare their selectivity and other thermodynamic quantities such as enthalpy and entropy at different pressures and compositions.

[below: Ren2022]

Before building any screening strategy or performing any computational screening, one needs to generate a set of files describing the atomic structure of the materials. Nanoporous materials can have different degrees of crystallinity from perfectly crystalline to completely amorphous. Most of the computational work is focused on crystalline structures, since the atoms are well-described within a periodic framework, which enables faster simulations. The presence of defects are also usually neglected, which could explain some of the discrepancies between simulations and experiments. And amorphous materials are described by thousands of atomic positions in order to grasp their intrinsic non-periodicity.[46] One can distinguish roughly four main classes of crystalline nanoporous materials: the inorganic zeolites (e.g. aluminosilicates, aluminophosphates), the porous polymer networks, the covalent organic frameworks (COFs) and the metal–organic frameworks (containing the zeolitic imidazolate frameworks *ie.* ZIFs and others). This diversity of nanoporous materials offer a wide range of potential candidates for any targeted applications.

The International Zeolite Association (IZA) gave a standardised set of 244 zeolites (in their idealized all-silica form) that can be used for screening purposes. To generate a dataset of structures, existing experimental database like the Cambridge Structural Database can be exploited. However, the raw structures determined experimentally by X-ray cannot be used directly as is. To obtain a computation-ready dataset, Chung *et al.* used algorithmic cleaning procedures to build the publicly available Computation-Ready Experimental MOF (CoRE MOF) database.[Chung_2014, 45] CoRE MOF 2019 contains about 14,000 MOF structures, which is the biggest experimental database. Similar approach applied to organic frameworks led to the construction of a set of 187 COFs with disorder-free and solvent-free structures.[47, 48]

These experiment-based databases can already be used in computational screenings to retrieve valuable information, but unknown structures that are yet to be discovered are not represented. To overcome the limits and biases of experimental synthesis, artificial ways of generating nanoporous material datasets can be used, which proved to be extremely efficient. The first *in silico* generated database of about 130,000 MOFs used a recursion-based assembly (or tinkertoy-like) algorithm to combine 102 building blocks.[10] Martin and Haranczyk then proposed a topology-specific structure assembly algorithm that leverage the topological information of the structures.[49] Inspired by this algorithm, topology-based databases emerged a few years later with the set of 13,000 MOF structures generated using the Topologically Based Crystal Constructor (ToBaCCo) algorithm developed by Colon, Gómez-Gualdrón and Snurr.[12] Later, Boyd and Woo proposed another topology-based algorithm using a graph theoretical approach and generated a 300,000 structures database (BW-DB) based on 46 different network topologies.[11] Similar approaches are used for other classes of materials, Deem and coworkers proposed a dataset of nearly 2.6 million hypothetical zeolite structures.[50–52] However, one could wonder if these hypothetical structures are synthesisable and can remain stable under operational conditions (e.g. thermal, mechanical, radioactive constraints). To discuss their synthetic likelihood, Anderson and Gómez-Gualdrón computed the free energies of 8,500 hypothetical structures and compared them to experimentally observed MOF structures.[53] Later, Nandy *et al.* performed a meta-analysis of thousands of articles associated to the CoRE MOF 2019 database to extract their experimental solvent-removal stability and thermal decomposition temperature.[54] These data are then leveraged in the training of multiple ML models to predict stability properties. These predictions can be very useful to gauge the relative stability of each materials and to only consider stable structures. Other types of materials have been explored, Turceni *et al.* published 60,000 organic cage structures and used machine learning to predict their stability based on the shape persistence metric.[55]

The Materials Genome Initiative, a 100 million dollar effort from the White House that aims to “discover, develop, and deploy new materials twice as fast”, led to the creation of the “Materials Project”, a centralised database containing all the above mentioned structures.[31, 56, 57] The fast development of this nanoporous materials genome motivated Boyd *et al.* to write a comprehensive review on all the initiatives on generating new data for computational analysis.[42]

Yet, the sole increase in size of the databases is not enough. One needs to add diversity to have more general knowledge on the maximum performance and the explanatory features of such performance. Moreover, the diversity of structures ensure the quality of the predicted best materials for a given application. To qualitatively or quantitatively assess the diversity of a database, inventive methodologies have been developed. For instance, Martin, Smit and Haranczyk proposed a Voronoi hologram representation as a way of measuring similarities between structures to generate geometrically diverse subsets of a database.[58] Moosavi *et al.* made a comparative study of the diversity of three well-known databases CoRE MOF 2019,[45] BW-DB[11] and ToBaCCo[Gomez_Gualdron_2016, 12] using geometrical and chemical descriptors to design a theoretical strategy for generating the most diverse set of materials.[59] Another approach consists in searching for similarities instead of differences in the materials by studying topological patterns in the data.[60] These investigations on the data structures give a solid ground to develop novel materials by objectively defining similarity, diversity and novelty. From the analysis gathered so far, one would need to radically change the approach by proposing materials with new chemistry, topology or mechanism (e.g. flexibility) in order to significantly improve the diversity of the current databases.

1.1.2 Simulation tools

1.1.3 Machine learning assisted screening

[Simon sumup]

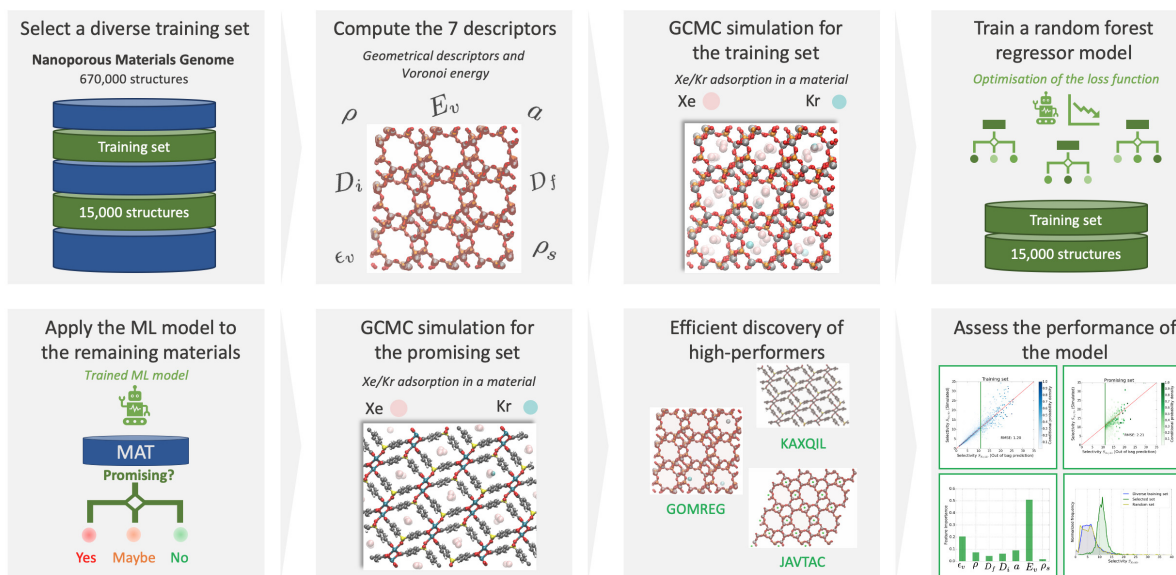


Figure 1.1 – Schematic representation of large-scale screening of nanoporous materials for Xe/Kr adsorption-based separation by Simon et al.[14], based on a combination of Grand Canonical Monte Carlo simulations and machine learning algorithm (Random Forest Regressor). The main goal of this screening is to find high-performing materials in a large dataset of both experimental and hypothetical materials. Adapted with permission from Ref. 14. Copyright 2015 American Chemical Society.

In 2015, Simon et al.[14] analyzed the Nanoporous Materials Genome,[42, 61] a database of about 670,000 experimental and hypothetical porous material structures, including MOFs, zeolites, PPNs, ZIFs, and COFs, for candidate adsorbents for xenon/krypton separations. It is possibly the largest-scale study performed in this area, both by the sheer number of frameworks involved and by the diversity of their nature. Because such a set is too big for brute-force screening with GCMC simulations, they proposed a multi-scale modelling strategy combining machine learning algorithms (trained on a diverse subset of 15,000 materials) with molecular simulations (used both to generate the ML training data, and to refine the separation properties for the top performers obtained by the ML predictor). Without going into details (see Fig. 1.1 for more details), the ML model they trained was mainly based on geometric structural descriptors, with the addition of a single energy-based descriptor: the Voronoi energy (i.e. the average energy of a xenon atom at the accessible nodes in the Voronoi partition of space). In addition to identifying and describing some top performing materials, the authors also analyzed the correlations between high Xe/Kr selectivity and the geometric properties of the frameworks, in order to “rationalize the strong link between pore size and selectivity”. In particular, by developing theoretical pore models of spherical and cylindrical geometries, they could highlight the general geometrical trends observed, but also the fact that there is a wide diversity of performance beyond the geometrical features of the frameworks.

1.2 A LITERATURE OVERVIEW

1.2.1 Thermodynamic adsorption properties

GAS STORAGE

GAS SEPARATION

1.2.2 Transport adsorption properties

KINETIC PROPERTIES

Used in breakthrough simulation

MEMBRANE MATERIALS

1.2.3 Non-adsorption properties

CATALYTIC ACTIVITY

MECHANICAL PROPERTIES

THERMAL PROPERTIES

1.3 CONSEQUENCES FOR XENON/KRYPTON SEPARATION

1.3.1 Status quo

WHAT IS DONE IN Xe/Kr SEPARATION

WHAT CAN BE LEARNED IN THE OTHER FIELDS

1.3.2 Future perspectives

Main improvement points

FASTER ENERGY SAMPLING

Integration in ml

FASTER DIFFUSION ESTIMATION

sq\$

FLEXIBILITY OMS

THERMODYNAMIC EXPLORATION OF XENON/KRYPTON SEPARATION

2.1	Preliminary analyses	11
2.1.1	Structure–selectivity relationships	11
2.1.2	Thermodynamic quantities correlations	11
2.2	Selectivity drop	11
2.2.1	Thermodynamic analyses	11
2.2.2	Detailed investigation	11



2.1 PRELIMINARY ANALYSES

2.1.1 Structure–selectivity relationships

2.1.2 Thermodynamic quantities correlations

2.2 SELECTIVITY DROP

2.2.1 Thermodynamic analyses

2.2.2 Detailed investigation

dqsdqs

ADSORPTION MOLECULAR SIMULATIONS

3.1	Standard simulation tools	13
3.1.1	Grand canonical monte carlo	13
3.1.2	Widom's insertion	13
3.2	New algorithm development	13
3.2.1	Rapid Adsorption Enthalpy Surface Sampling (RAESS)	13
3.2.2	Grid Adsorption Energies Sampling (GrAES)	13



3.1 STANDARD SIMULATION TOOLS

3.1.1 Grand canonical monte carlo

3.1.2 Widom's insertion

3.2 NEW ALGORITHM DEVELOPMENT

3.2.1 Rapid Adsorption Enthalpy Surface Sampling (RAESS)

3.2.2 Grid Adsorption Energies Sampling (GrAES)

ADSORPTION PROPERTIES PREDICTION

4.1	Machine learning	15
4.1.1	Introduction	15
4.1.2	eXtreme Gradient Boosting	15
4.2	Ambient-pressure prediction	15
4.2.1	From infinite dilution to ambient pressure	15
4.2.2	Interpretation of the ML model	15



4.1 MACHINE LEARNING

4.1.1 Introduction

4.1.2 eXtreme Gradient Boosting

4.2 AMBIENT-PRESSURE PREDICTION

4.2.1 From infinite dilution to ambient pressure

4.2.2 Interpretation of the ML model

Origins of the selectivity drop

TRANSPORT PROPERTIES

5.1	Computational simulations	17
5.1.1	Molecular dynamics	17
5.1.2	Fast kinetic Monte Carlo	17
5.2	ML modeling	17



5.1 COMPUTATIONAL SIMULATIONS

Experiment?

5.1.1 Molecular dynamics

5.1.2 Fast kinetic Monte Carlo

tutrast ctutrast ML descriptors next steps

5.2 ML MODELING

Results

TOWARDS THE NEXT GENERATION OF SCREENINGS

6.1	Flexibility	19
6.1.1	Problem, literature	19
6.1.2	Snapshot.	19
6.2	Open Metal Sites.	19
6.2.1	Problem, literature	19
6.2.2	Perspectives	19



6.1 FLEXIBILITY

Final screening step, easy integration into the workflow of current screenings

6.1.1 Problem, literature

6.1.2 Snapshot

6.2 OPEN METAL SITES

6.2.1 Problem, literature

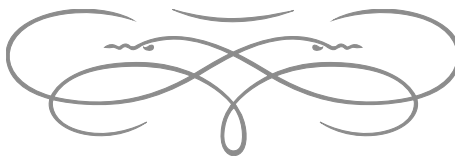
6.2.2 Perspectives

GENERAL CONCLUSIONS

The work presented in this thesis is



This work opens perspectives for



LIST OF PUBLICATIONS

PEER-REVIEWED PAPERS

1. Emmanuel Ren and François-Xavier Coudert. “Thermodynamic exploration of xenon/krypton separation based on a high-throughput screening”. In: *Faraday Discussions* 231 (2021), pp. 201–223. doi: [10.1039/D1FD00024A](https://doi.org/10.1039/D1FD00024A).
2. Emmanuel Ren, Philippe Guilbaud, and François-Xavier Coudert. “High-throughput computational screening of nanoporous materials in targeted applications”. In: *Digital Discovery* 1.4 (2022), pp. 355–374. doi: [10.1039/D2DD00018K](https://doi.org/10.1039/D2DD00018K).
3. Emmanuel Ren and François-Xavier Coudert. “Rapid adsorption enthalpy surface sampling (RAESS) to characterize nanoporous materials”. In: *Chemical Science* 14.7 (2023), pp. 1797–1807. doi: [10.1039/D2SC05810C](https://doi.org/10.1039/D2SC05810C).

PREPRINT

4. Emmanuel Ren and François-Xavier Coudert. “Gas Separation Selectivity Prediction Based on Finely Designed Descriptors”. In: *ChemRxiv* (2023).

BIBLIOGRAPHY

- [1] F. G. Kerry. *Industrial gas handbook: gas separation and purification*. CRC press, 2007.
- [2] National Academies of Sciences, Engineering, and Medicine. *A Research Agenda for Transforming Separation Science*. en. Washington, D.C.: The National Academies Press, 2019. ISBN: 978-0-309-49170-9. DOI: [10.17226/25421](https://doi.org/10.17226/25421).
- [3] D. Banerjee, C. M. Simon, S. K. Elsaïdi, M. Haranczyk, and P. K. Thallapally. “Xenon Gas Separation and Storage Using Metal–Organic Frameworks”. In: *Chem* 4 (3 2018), pp. 466–494. DOI: [10.1016/j.chempr.2017.12.025](https://doi.org/10.1016/j.chempr.2017.12.025).
- [4] L. Chen, P. S. Reiss, S. Y. Chong, D. Holden, K. E. Jelfs, T. Hasell, M. A. Little, A. Kewley, M. E. Briggs, A. Stephenson, K. M. Thomas, J. A. Armstrong, J. Bell, J. Busto, R. Noel, J. Liu, D. M. Strachan, P. K. Thallapally, and A. I. Cooper. “Separation of rare gases and chiral molecules by selective binding in porous organic cages”. In: *Nature Mater.* 13.10 (July 2014), pp. 954–960. DOI: [10.1038/nmat4035](https://doi.org/10.1038/nmat4035).
- [5] L. Li, L. Guo, Z. Zhang, Q. Yang, Y. Yang, Z. Bao, Q. Ren, and J. Li. “A Robust Squarate-Based Metal–Organic Framework Demonstrates Record-High Affinity and Selectivity for Xenon over Krypton”. In: *J. Am. Chem. Soc.* 141.23 (May 2019), pp. 9358–9364. DOI: [10.1021/jacs.9b03422](https://doi.org/10.1021/jacs.9b03422).
- [6] J. Pei, X.-W. Gu, C.-C. Liang, B. Chen, B. Li, and G. Qian. “Robust and Radiation-Resistant Hofmann-Type Metal–Organic Frameworks for Record Xenon/Krypton Separation”. In: *J. Am. Chem. Soc.* 144.7 (Feb. 2022), pp. 3200–3209. DOI: [10.1021/jacs.1c12873](https://doi.org/10.1021/jacs.1c12873).
- [7] H. Lyu, Z. Ji, S. Wuttke, and O. M. Yaghi. “Digital Reticular Chemistry”. In: *Chem* 6.9 (Sept. 2020), pp. 2219–2241. DOI: [10.1016/j.chempr.2020.08.008](https://doi.org/10.1016/j.chempr.2020.08.008).
- [8] K. M. Jablonka, A. S. Rosen, A. S. Krishnapriyan, and B. Smit. “An Ecosystem for Digital Reticular Chemistry”. In: *ACS Central Science* (Mar. 2023). DOI: [10.1021/acscentsci.2c01177](https://doi.org/10.1021/acscentsci.2c01177).
- [9] C. R. Groom, I. J. Bruno, M. P. Lightfoot, and S. C. Ward. “The Cambridge Structural Database”. In: *Acta Cryst. B* 72.2 (Apr. 2016), pp. 171–179. DOI: [10.1107/s2052520616003954](https://doi.org/10.1107/s2052520616003954).
- [10] 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”. In: *Nature Chem.* 4 (2 2012), pp. 83–89. DOI: [10.1038/nchem.1192](https://doi.org/10.1038/nchem.1192).
- [11] P. G. Boyd and T. K. Woo. “A generalized method for constructing hypothetical nanoporous materials of any net topology from graph theory”. In: *CrystEngComm* 18.21 (2016), pp. 3777–3792. DOI: [10.1039/c6ce00407e](https://doi.org/10.1039/c6ce00407e).
- [12] 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”. In: *Cryst. Growth Des.* 17 (11 2017), pp. 5801–5810. DOI: [10.1021/acs.cgd.7b00848](https://doi.org/10.1021/acs.cgd.7b00848).
- [13] E. Ren, P. Guillaud, and F.-X. Coudert. “High-throughput computational screening of nanoporous materials in targeted applications”. In: *Digital Discovery* 1.4 (2022), pp. 355–374. DOI: [10.1039/D2DD00018K](https://doi.org/10.1039/D2DD00018K).
- [14] C. M. Simon, R. Mercado, S. K. Schnell, B. Smit, and M. Haranczyk. “What Are the Best Materials To Separate a Xenon/Krypton Mixture?” In: *Chem. Mater.* 27 (12 2015), pp. 4459–4475. DOI: [10.1021/acs.chemmater.5b01475](https://doi.org/10.1021/acs.chemmater.5b01475).

- [15] C. H. Rycroft. "VORO++: A three-dimensional Voronoi cell library in C++". In: *Chaos* 19.4 (Dec. 2009), p. 041111. doi: [10.1063/1.3215722](https://doi.org/10.1063/1.3215722).
- [16] E. Ren and F.-X. Coudert. "Rapid adsorption enthalpy surface sampling (RAESS) to characterize nanoporous materials". In: *Chemical Science* 14.7 (2023), pp. 1797–1807. doi: [10.1039/D2SC05810C](https://doi.org/10.1039/D2SC05810C).
- [17] K. Shi, Z. Li, D. M. Anstine, D. Tang, C. M. Colina, D. S. Sholl, J. I. Siepmann, and R. Q. Snurr. "Two-Dimensional Energy Histograms as Features for Machine Learning to Predict Adsorption in Diverse Nanoporous Materials". In: *J. Chem. Theory Comput.* (Feb. 2023). doi: [10.1021/acs.jctc.2c00798](https://doi.org/10.1021/acs.jctc.2c00798).
- [18] E. Ren and F.-X. Coudert. "Thermodynamic exploration of xenon/krypton separation based on a high-throughput screening". In: *Faraday Discussions* 231 (2021), pp. 201–223. doi: [10.1039/D1FD00024A](https://doi.org/10.1039/D1FD00024A).
- [19] G. Hautier. "Finding the needle in the haystack: Materials discovery and design through computational ab initio high-throughput screening". In: *Comput. Mater. Sci.* 163 (2019), pp. 108–116. doi: [10.1016/j.commatsci.2019.02.040](https://doi.org/10.1016/j.commatsci.2019.02.040).
- [20] J. M. Cole. "A Design-to-Device Pipeline for Data-Driven Materials Discovery". In: *Acc. Chem. Res.* 53 (3 2020), pp. 599–610. doi: [10.1021/acs.accounts.9b00470](https://doi.org/10.1021/acs.accounts.9b00470).
- [21] A. Ludwig. "Discovery of new materials using combinatorial synthesis and high-throughput characterization of thin-film materials libraries combined with computational methods". In: *npj Comput. Mater.* 5 (1 2019), p. 0121. doi: [10.1038/s41524-019-0205-0](https://doi.org/10.1038/s41524-019-0205-0).
- [22] H. S. Stein and J. M. Gregoire. "Progress and prospects for accelerating materials science with automated and autonomous workflows". In: *Chem. Sci.* 10 (42 2019), pp. 9640–9649. doi: [10.1039/C9SC03766G](https://doi.org/10.1039/C9SC03766G).
- [23] D. A. Gómez-Gualdrón, C. E. Wilmer, O. K. Farha, J. T. Hupp, and R. Q. Snurr. "Exploring the Limits of Methane Storage and Delivery in Nanoporous Materials". In: *J. Phys. Chem. C* 118 (13 2014), pp. 6941–6951. doi: [10.1021/jp502359q](https://doi.org/10.1021/jp502359q).
- [24] M. Moliner, Y. Román-Leshkov, and A. Corma. "Machine Learning Applied to Zeolite Synthesis: The Missing Link for Realizing High-Throughput Discovery". In: *Acc. Chem. Res.* 52 (10 2019), pp. 2971–2980. doi: [10.1021/acs.accounts.9b00399](https://doi.org/10.1021/acs.accounts.9b00399).
- [25] J. L. Salcedo Perez, M. Haranczyk, and N. E. R. Zimmermann. "High-throughput assessment of hypothetical zeolite materials for their synthesizability and industrial deployability". In: *Z. Kristallogr.* 234 (7-8 2019), pp. 437–450. doi: [10.1515/zkri-2018-2155](https://doi.org/10.1515/zkri-2018-2155).
- [26] "Crystallography: Protein Data Bank". In: *Nature New Biology* 233 (42 1971), pp. 223–223. doi: [10.1038/newbio233223b0](https://doi.org/10.1038/newbio233223b0).
- [27] S. Gražulis, D. Chateigner, R. T. Downs, A. F. T. Yokochi, M. Quirós, L. Lutterotti, E. Manakova, J. Butkus, P. Moeck, and A. Le Bail. "Crystallography Open Database – an open-access collection of crystal structures". In: *J. Appl. Crystallogr.* 42 (4 2009), pp. 726–729. doi: [10.1107/S0021889809016690](https://doi.org/10.1107/S0021889809016690).
- [28] F.-X. Coudert. "Materials Databases: The Need for Open, Interoperable Databases with Standardized Data and Rich Metadata". In: *Adv. Theory Simul.* 2 (11 2019), p. 1900131. doi: [10.1002/adts.201900131](https://doi.org/10.1002/adts.201900131).
- [29] J. J. de Pablo, B. Jones, C. L. Kovacs, V. Ozolins, and A. P. Ramirez. "The Materials Genome Initiative, the interplay of experiment, theory and computation". In: *Curr. Op. Solid State Mater. Sci.* 18 (2 2014), pp. 99–117. doi: [10.1016/j.cossms.2014.02.003](https://doi.org/10.1016/j.cossms.2014.02.003).
- [30] J. J. de Pablo, N. E. Jackson, M. A. Webb, L.-Q. Chen, J. E. Moore, D. Morgan, R. Jacobs, T. Pollock, D. G. Schlom, E. S. Toberer, J. Analytis, I. Dabo, D. M. DeLongchamp, G. A. Fiete, G. M. Grason, G. Hautier, Y. Mo, K. Rajan, E. J. Reed, E. Rodriguez, V. Stevanovic, J. Suntivich, K. Thornton, and J.-C. Zhao. "New frontiers for the materials genome initiative". In: *npj Comput. Mater.* 5 (1 2019), p. 99. doi: [10.1038/s41524-019-0173-4](https://doi.org/10.1038/s41524-019-0173-4).

- [31] 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”. In: *APL Mater.* 1 (1 2013), p. 011002. doi: [10.1063/1.4812323](https://doi.org/10.1063/1.4812323).
- [32] A. Jain, K. A. Persson, and G. Ceder. “Research Update: The materials genome initiative: Data sharing and the impact of collaborative ab initio databases”. In: *APL Mater.* 4 (5 2016), p. 053102. doi: [10.1063/1.4944683](https://doi.org/10.1063/1.4944683).
- [33] M. D. Foster, O. Delgado Friedrichs, R. G. Bell, F. A. Almeida Paz, and J. Klinowski. “Chemical Evaluation of Hypothetical Uninodal Zeolites”. In: *J. Am. Chem. Soc.* 126 (31 2004), pp. 9769–9775. doi: [10.1021/ja037334j](https://doi.org/10.1021/ja037334j).
- [34] C. E. Wilmer and R. Q. Snurr. “Large-Scale Generation and Screening of Hypothetical Metal-Organic Frameworks for Applications in Gas Storage and Separations”. In: *Topics in Current Chemistry*. Springer International Publishing, 2013, pp. 257–289. doi: [10.1007/128_2013_490](https://doi.org/10.1007/128_2013_490).
- [35] V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K. A. Persson, G. Ceder, and A. Jain. “Unsupervised word embeddings capture latent knowledge from materials science literature”. In: *Nature* 571 (7763 2019), pp. 95–98. doi: [10.1038/s41586-019-1335-8](https://doi.org/10.1038/s41586-019-1335-8).
- [36] C. J. Court and J. M. Cole. “Magnetic and superconducting phase diagrams and transition temperatures predicted using text mining and machine learning”. In: *npj Comput. Mater.* 6 (1 2020), p. 38. doi: [10.1038/s41524-020-0287-8](https://doi.org/10.1038/s41524-020-0287-8).
- [37] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh. “Machine learning for molecular and materials science”. In: *Nature* 559 (7715 2018), pp. 547–555. doi: [10.1038/s41586-018-0337-2](https://doi.org/10.1038/s41586-018-0337-2).
- [38] E. Kim, K. Huang, S. Jegelka, and E. Olivetti. “Virtual screening of inorganic materials synthesis parameters with deep learning”. In: *npj Comput. Mater.* 3 (1 2017), p. 1120. doi: [10.1038/s41524-017-0055-6](https://doi.org/10.1038/s41524-017-0055-6).
- [39] G. Borboudakis, T. Stergiannakos, M. Frysali, E. Klontzas, I. Tsamardinos, and G. E. Froudakis. “Author Correction: Chemically intuited, large-scale screening of MOFs by machine learning techniques”. In: *npj Comput. Mater.* 3 (1 2017), pp. 1–7. doi: [10.1038/s41524-017-0051-x](https://doi.org/10.1038/s41524-017-0051-x).
- [40] S. Chibani and F.-X. Coudert. “Machine learning approaches for the prediction of materials properties”. In: *APL Mater.* 8 (8 2020), p. 080701. doi: [10.1063/5.0018384](https://doi.org/10.1063/5.0018384).
- [41] T. Zhou, Z. Song, and K. Sundmacher. “Big Data Creates New Opportunities for Materials Research: A Review on Methods and Applications of Machine Learning for Materials Design”. In: *Engineering* 5 (6 2019), pp. 1017–1026. doi: [10.1016/j.eng.2019.02.011](https://doi.org/10.1016/j.eng.2019.02.011).
- [42] P. G. Boyd, Y. Lee, and B. Smit. “Computational development of the nanoporous materials genome”. In: *Nature Rev. Mater.* 2 (8 2017), p. 1. doi: [10.1038/natrevmats.2017.37](https://doi.org/10.1038/natrevmats.2017.37).
- [43] K. Adil, Y. Belmabkhout, R. S. Pillai, A. Cadiau, P. M. Bhatt, A. H. Assen, G. Maurin, and M. Eddaoudi. “Gas/vapour separation using ultra-microporous metal–organic frameworks: insights into the structure/separation relationship”. In: *Chem. Soc. Rev.* 46 (11 2017), pp. 3402–3430. doi: [10.1039/C7CS00153C](https://doi.org/10.1039/C7CS00153C).
- [44] A. Sturluson, M. T. Huynh, A. R. Kaija, C. Laird, S. Yoon, F. Hou, Z. Feng, C. E. Wilmer, Y. J. Colón, Y. G. Chung, D. W. Siderius, and C. M. Simon. “The role of molecular modelling and simulation in the discovery and deployment of metal-organic frameworks for gas storage and separation”. In: *Mol. Simulat.* 45 (14-15 2019), pp. 1082–1121. doi: [10.1080/08927022.2019.1648809](https://doi.org/10.1080/08927022.2019.1648809).
- [45] 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”. In: *Journal of Chemical & Engineering Data* 64.12 (Nov. 2019), pp. 5985–5998. doi: [10.1021/acs.jced.9b00835](https://doi.org/10.1021/acs.jced.9b00835).
- [46] R. Thyagarajan and D. S. Sholl. “A Database of Porous Rigid Amorphous Materials”. In: *Chemistry of Materials* 32.18 (Aug. 2020), pp. 8020–8033. doi: [10.1021/acs.chemmater.0c03057](https://doi.org/10.1021/acs.chemmater.0c03057).

- [47] M. Tong, Y. Lan, Q. Yang, and C. Zhong. “Exploring the structure-property relationships of covalent organic frameworks for noble gas separations”. In: *Chemical Engineering Science* 168 (Aug. 2017), pp. 456–464. doi: 10.1016/j.ces.2017.05.004.
- [48] D. Ongari, A. V. Yakutovich, L. Talirz, and B. Smit. “Building a consistent and reproducible database for adsorption evaluation in covalent–organic frameworks”. In: *ACS central science* 5.10 (2019), pp. 1663–1675.
- [49] R. L. Martin and M. Haranczyk. “Construction and Characterization of Structure Models of Crystalline Porous Polymers”. In: *Crystal Growth & Design* 14.5 (Apr. 2014), pp. 2431–2440. doi: 10.1021/cg500158c.
- [50] D. J. Earl and M. W. Deem. “Toward a Database of Hypothetical Zeolite Structures”. In: *Industrial & Engineering Chemistry Research* 45.16 (Jan. 2006), pp. 5449–5454. doi: 10.1021/ie0510728.
- [51] M. W. Deem, R. Pophale, P. A. Cheeseman, and D. J. Earl. “Computational Discovery of New Zeolite-Like Materials”. In: *The Journal of Physical Chemistry C* 113.51 (Oct. 2009), pp. 21353–21360. doi: 10.1021/jp906984z.
- [52] R. Pophale, P. A. Cheeseman, and M. W. Deem. “A database of new zeolite-like materials”. In: *Physical Chemistry Chemical Physics* 13.27 (2011), p. 12407. doi: 10.1039/c0cp02255a.
- [53] R. Anderson and D. A. Gómez-Gualdrón. “Large-Scale Free Energy Calculations on a Computational Metal–Organic Frameworks Database: Toward Synthetic Likelihood Predictions”. In: *Chemistry of Materials* 32.19 (July 2020), pp. 8106–8119. doi: 10.1021/acs.chemmater.0c00744.
- [54] A. Nandy, C. Duan, and H. J. Kulik. “Using Machine Learning and Data Mining to Leverage Community Knowledge for the Engineering of Stable Metal–Organic Frameworks”. In: *Journal of the American Chemical Society* 143.42 (Oct. 2021), pp. 17535–17547. doi: 10.1021/jacs.1c07217.
- [55] L. Turcani, R. L. Greenaway, and K. E. Jelfs. “Machine Learning for Organic Cage Property Prediction”. In: *Chemistry of Materials* 31.3 (Dec. 2018), pp. 714–727. doi: 10.1021/acs.chemmater.8b03572.
- [56] T. Kalil and C. Wadia. *Materials Genome Initiative for Global Competitiveness*. Washington, 2011.
- [57] *The Materials Genome Initiative*. Available online at <https://www.mgi.gov/>. 2022.
- [58] R. L. Martin, B. Smit, and M. Haranczyk. “Addressing Challenges of Identifying Geometrically Diverse Sets of Crystalline Porous Materials”. In: *Journal of Chemical Information and Modeling* 52.2 (Dec. 2011), pp. 308–318. doi: 10.1021/ci200386x.
- [59] 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”. In: *Nature Communications* 11.1 (Aug. 2020). doi: 10.1038/s41467-020-17755-8.
- [60] Y. Lee, S. D. Barthel, P. Dłotko, S. M. Moosavi, K. Hess, and B. Smit. “Quantifying similarity of pore-geometry in nanoporous materials”. In: *Nature Communications* 8.1 (May 2017). doi: 10.1038/ncomms15396.
- [61] C. M. Simon, J. Kim, D. A. Gomez-Gualdrón, 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”. In: *Energy Environ. Sci.* 8 (4 2015), pp. 1190–1199. doi: 10.1039/C4EE03515A.

RÉSUMÉ EN FRANÇAIS

Introduction	31
------------------------	----



INTRODUCTION

[5 à 10 pages]

Les matériaux poreux sont des matériaux



RÉSUMÉ

Durant ma thèse, j'ai

MOTS CLÉS

simulation moléculaire, matériaux nanoporeux,

ABSTRACT

During my PhD, I

KEYWORDS

molecular simulation, porous materials,