THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à Chimie ParisTech
Dans le cadre d'une cotutelle avec le CEA Marcoule

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 :

Sofía CALERO Professeure, Eindhoven University of Technology	*Rapportrice*
Rochus SCHMID Professeur, Ruhr-University Bochum	*Rapporteur*
Magali BENOIT Directrice de Recherche, Université de Toulouse	*Rapportrice*
Michael BADAWI Maître de Conférences, Université de Lorraine	*Rapporteur*
NAME SURNAME TITLE, PLACE/University	*Role*
François-Xavier COUDERT Directeur de Recherche, Chimie ParisTech	*Directeur de thèse*

REMERCIEMENTS

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

TABLE OF CONTENTS

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]

[Xe/Kr applications in the industry]

This thesis presents my work on the

1

HIGH-THROUGHPUT COMPUTATIONAL SCREENING OF NANOPOROUS MATERIALS

1.1 Nanoporous Materials

1.1.1 Definition of the concept

Adsorption isotherms and geometrical descriptors

Nanoporous materials are defined as materials with a nanoscale structure constituted by pores and cavities, which some are connected by a network of channels. These pores can be empty or filled with a variety of substances called adsorbates. By adhering molecules from a liquid or gas phase into the internal surfaces of the material, we can use it in diverse applications such as gas separation and purification,[19, 20] energy storage and conversion,[21, 22] heterogeneous catalysis[23–25] drug delivery,[26, 27] or sensing.[28] By designing the chemical nature, size, shape and distribution of the pores, we can tailor the physicochemical properties to the targeted application.[29]

The process of adhering particles or molecules on surfaces is called adsorption. Adsorption occurs due to attractive forces between adsorbates and the adsorbent surface, such as Van der Waals forces, hydrogen bonding, and electrostatic interactions. The adsorption performance depends on the chemical nature of the interface, its exposed surface area and the shape of the pores. We usually characterize adsorption properties of an adsorbate compound by measuring the numbers of adsorbed molecules as a function of its pressure at a given temperature, which is called the adsorption isotherm. These isotherms can possibly be used, among other techniques, to specify the pore size distribution, accessible surface area and pore volume.[30] By using fitting models, we can also use adsorption isotherms to characterize the maximum adsorption uptake among other adsorption descriptors.[31] Using a set of experimental isotherms at close but different temperatures, we can also retrieve information on the isosteric heat of adsorption q_{st} (the negative differential of the excess enthalpy of adsorption with respect to the excess adsorption).[32] This heat of adsorption (related to the enthalpy of adsorption) can also be directly obtained using calorimetry.[33] Furthermore measurements at infinite dilution can also lead to a linear relation between the adsorbed quantity and the pressure defined by the Henry's law; another key adsorption descriptor, the Henry adsorption constant, is defined as the slope of this linear regime.[34] All of these thermodynamic quantities are most valuable in comparing experimental data to computational modeling to compare and characterize the materials suitable for a given gas adsorption process.

Most of the materials studied in this thesis will have pores with a size around the nanometer called "nanopores". The International Union of Pure Applied Chemistry (IUPAC) classifies these pores into three categories according to their size: micropores ($\leq$2 nm), mesopores (2 nm–50 nm) and macropores (>50 nm).[35] Here, we will use a single terminology (nanopore) to designate all pores of under a few nanometers. A good characterization of the nanopores of these materials is key to fine-tuning the adsorption properties.[29] The pore size distribution (PSD) can be computationally determined if we have resolved the structure of the nanoporous material (using X-ray diffraction on crystallized porous solids). This is the most accurate determination method of the PSD, but it relies on considering that the structure is perfectly rigid and crystalline so that only one structural data can characterize it. Other experimental methods rely on assumptions, model systems (e.g., cylindrical) or adsorption characteristics. For instance, stereological analyses based on plane sections cut through a porous material can evaluate the PSD.[36] The Horvath-Kawazoe (HK) method is a semi-empirical analytic model of adsorption isotherm that can extract PSD. Small angle X-ray and neutron scattering methods are non-destructive methods of pore characterization.[37] In this thesis, we will rely on computationally analyzing X-ray diffusion data to deduce pore sizes and other geometrical characteristics.

The pore volume consists in the measure of the volume of "closed" and "open" pores of nanoporous materials. Depending on the way of measuring it, different quantities are probed. Some pores could not be accessed by some adsorbate; depending on the probe size the volume calculated will not be the same. Methods that do not rely on adsorption like scattering or stereology will, however, measure the total pore volume. The porosity or void fraction would be defined as the ratio between the pore volume and the apparent framework volume. Depending on the method, we can therefore retrieve either the total porosity, the porosity opened or closed to a given probe adsorbate.

The cavities of the nanoporous material lay out an incredibly large adsorbable surface area, which is extremely useful in increasing the number of molecules in a given volume or mass of material, several thousands of square meters can be found in a gram of some nanoporous materials.[38] The higher the surface area, the more molecules can be adsorbed for storage, separation or reaction purposes; it is therefore crucial to measure this surface area using experimental and computational methods. The most extensively used method to experimentally measure surface areas from adsorption isotherms is based on the Brunauer–Emmett–Teller (BET) theory.[39] Most BET areas are calculated on N_2 isotherm at its boiling temperature (77 K); although different probe adsorbates can be considered, they are not standard.[40] However, the definition of the surface area depends highly on the condition of measurement but also on the fitting methodology; a dozen isotherms were given to 61 labs for BET area calculation, and the statistical experiment yielded to a high level of disparities in the calculation.[41]

Beyond the experimental techniques, some software like Zeo++ or PoreBlazer focus on computing pore size distributions, surface areas and void fractions using well-defined structure files.[42, 43] The definition of these values also depends on the probe size chosen to model a given adsorbate, the size of the framework atoms and the quality of the input structure. The computational values do not rely on adsorption models or on isotherm data as in the BET area, they are now relying on more comprehensive structural data. They, however, very much rely on a well-designed definition of the volume, the surface and the pore size we want to evaluate. Moreover, these values also highly depend on the radii of the framework atoms and the adsorbate we consider.[44]. In this thesis, we will rely on these computational methods to define these geometrical descriptors of nanoporous materials.

Classes of nanoporous materials

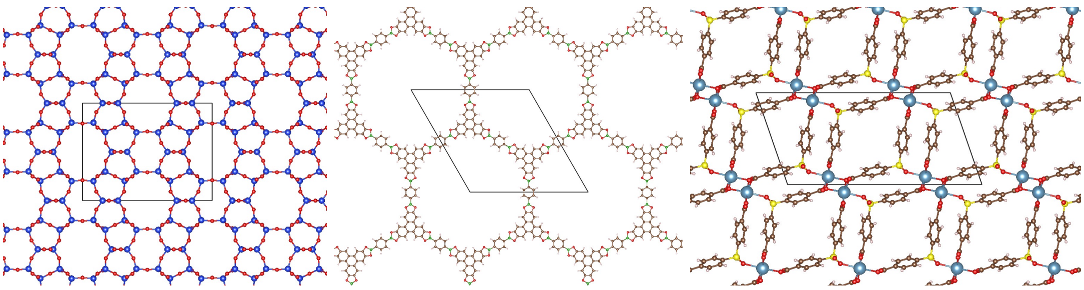


Figure 1.1: Illustration of a zeolite FER[45], a COF[46] and a MOF[47]. Color code: brown for C, white for H, red for O, blue for Si, cyan for Ca, yellow for S and green for B.

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 is also usually neglected, which could explain some discrepancies between simulations and experiments. And amorphous materials are described by thousands of atomic positions in order to grasp their intrinsic non-periodicity.[48] Activated carbons, a famous class of amorphous material, are extensively used in the industry for gas purification, but cannot be rationally studied to characterize their adsorption properties. One can distinguish roughly three main classes of crystalline nanoporous materials: the inorganic materials like zeolites (aluminosilicates or aluminophosphates), the organic materials like the porous polymer networks (PPNs) or the covalent organic frameworks (COFs) and the metal–organic frameworks (MOFs).

Zeolites are naturally occurring nanoporous aluminosilicate materials that are commonly synthesized to be used in the industry as a commercial adsorbent and heterogeneous catalyst.[49, 50] It is considered as one of the most mature nanoporous material technology at our disposal. This class of material also leaves a wide room for innovation since different Al/Si ratios of a same zeolite type pan out a wide range of structures. Furthermore, zeolite materials inspired the synthesis of zeolitic frameworks harboring different atoms such as the aluminophosphates or the zeolitic imidazolate frameworks.[51, 52]

Porous polymer networks (PPNs) are porous materials based on the already very developed polymer material technology.[31, 53, 54] However, one of the major drawback of this type of material is the formation of irreversible covalent bonds, which make the synthesis kinetically controlled leading to difficulties in crystallizing PPNs.[55] In order to create crystalline porous materials, Cote et al. figured out a way of using boron-based organic compounds to form reversible bounds, which formed thermodynamically stable materials COF-1 and COF-5.[46] This initiative was led by the group of Yaghi who was at the initiative of another very promising and well-known class of materials. A decade earlier, they pioneered a hydrothermal synthesis of a metal–organic framework presenting broad rectangular channels.[56]

Metal–organic frameworks (MOFs) are a class of nanoporous materials formed by metallic centers connected with organic linkers to form a stable crystalline solid. Even if the first synthesis of such a material was done since the early 90s,[57] and brought about a sparking interest in the scientific community a couple of decades later.[58, 59] Because plenty of combinations of linkers and metals are imaginable, an infinite amount of MOFs could theoretically be designed. Their structure can be tuned to our needs to enhance their performance in the targeted application.[60] This diversity of nanoporous materials offer a wide range of potential candidates that could be evaluated for any targeted applications.

1.1.2 Computational databases

All the previously described materials have been either synthesized and resolved using X-ray crystallography or computationally constructed. By combining almost all possible nanoporous materials, almost a million structures have been considered for separation or storage applications.[14, 61, 62] This extended database can be broken down into the synthesized materials and hypothetical ones for all the above-mentioned classes of material.

The International Zeolite Association (IZA) gave a standardized 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 databases 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.[63, 64] 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.[65, 66]

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 leverages the topological information of the structures.[67] 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-structure database (BW-DB) based on 46 different network topologies.[11] Similar approaches are used for other classes of materials, Deem and co-workers proposed a dataset of nearly 2.6 million hypothetical zeolite structures.[68–70] However, one could wonder if these hypothetical structures are synthesizable 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.[71] 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.[72] 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 material and to only consider stable structures. Other types of materials have been explored, Turcani et al. published 60,000 organic cage structures and used machine learning to predict their stability based on the shape persistence metric.[73]

The Materials Genome Initiative, 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 centralized database containing all the above-mentioned structures.[74–76] 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.[77]

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.[78] Moosavi et al. made a comparative study of the diversity of three well-known databases CoRE MOF 2019,[64] BW-DB[11] and ToBaCCo[12, 79] using geometrical and chemical descriptors to design a theoretical strategy for generating the most diverse set of materials.[80] Another approach consists in searching for similarities instead of differences in the materials by studying topological patterns in the data.[81] 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.

1.1.3 Exploring the chemical and structural space

With the development of ever-increasing nanoporous material databases, computational chemists proposed more and more inventive methods to evaluate or screen thousands of structures. Other challenges arose, such as the design of more efficient methods than the brute force screening or the analysis of big data. Two research groups in Northwestern University led by R. Snurr and J. Hupp began to address those questions, they used a “funnel-like” approach to efficiently screen about 130,000 hypothetical MOF structures.[10] To do so, they performed a first screening involving fewer steps of simulation on the whole dataset, then they extracted a subset of top-performing structures to perform a second round with more simulation steps. This procedure is repeated until a few materials are selected by a final round of simulations with reasonable accuracy. Similar “funnel-like” procedures have then been used in other fields of applications as described in the Figure 1.2. This type of screening saves precious computation time by balancing the complexity of the calculation with the amount of data to be screened. The most demanding simulations or experiments are only applied to the few most promising structures. This method can rather efficiently identify top candidates, but it can’t draw quantitative structure-property relationships (QSPR), beside facing scalability issues above a critical dataset size.

To overcome these new challenges, people are looking increasingly towards transferable models trained by a machine learning (ML) algorithm on a diverse and size-limited subsample. Ideally, such a model is transferable to potentially millions of structures and can provide valuable QSPR. For instance, Fernandez et al.[82] used multiple linear regression analysis, decision tree regression, and nonlinear support-vector machine models to extract QSPR and establish rules of designing well-performing MOFs for methane storage, while identifying promising structures. In this first work, they only used geometrical descriptors to describe methane storage,[82] but realizing the importance of chemical descriptors, they proposed the atomic property weighted radial distribution function as a powerful descriptor to predict CO_2 uptakes.[83] More importantly, they proved that ML can be used as a pre-screening tool to avoid running time-costly simulations by correctly identifying around 95 % of the top 1000 best performing materials. Recently, the same group used similar techniques to predict CO_2 working capacity as well as CO_2/H_2 selectivity in MOFs for pre-combustion carbon capture.[84]

1.2 Review of Screening Methodologies

1.2.1 Non-adsorption properties

Due to their high internal surface area, adsorption applications were a natural outlet for nanoporous materials. However, these materials can be used in many other applications. This section is dedicated to the physical and chemical properties not directly related to the adsorption process inside nanoporous materials such as catalytic activity,[87–89] mechanical properties, [90, 91] or thermal properties.[92–94] These properties require a more refined description of the atomic interactions within the material. DFT simulations are usually performed to accurately retrieve these properties. However, the computational cost required is multiplied by several orders of magnitude compared to classical simulations. The size of the datasets screened is therefore much smaller (a few hundreds maximum), and the use of ML can potentially speed

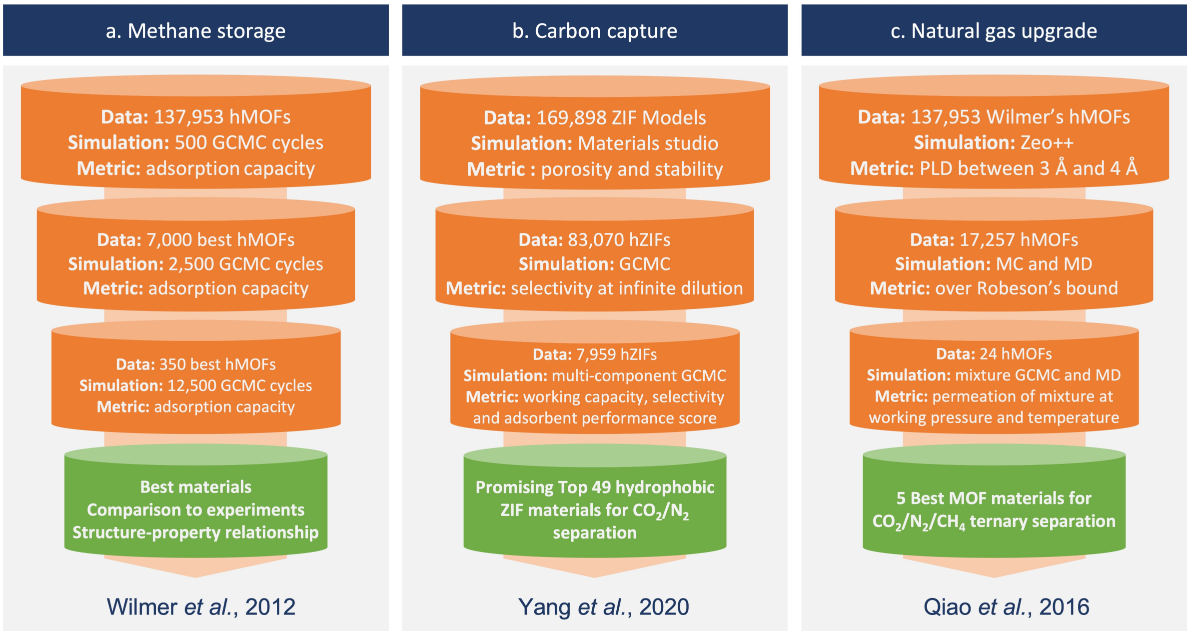


Figure 1.2: Simplified representation of typical funnel-type screening procedures, exemplified on three different applications from the published literature. (a) Wilmer et al.[10] used a series of bi-component Grand Canonical Monte Carlo (GCMC) calculations at different levels of complexity to screen a large dataset of hypothetical MOFs for methane storage application. (b) Yang et al.[85] used simulations at infinite dilution to prescreen the dataset before using computationally demanding simulations and multiple metrics to find the most promising ZIFs for carbon capture. (c) In Qiao et al.[86], transport properties were screened along standard adsorption properties to find the best materials for the targeted $CO_2/N_2/CH_4$ ternary separation; similarly, cheaper calculations at infinite dilution were carried out in a first step, before using more expensive calculations at working pressure and temperature.

up the whole process. ML is based on lower-cost descriptors,[95, 96] or it can be used in ML potentials for molecular simulations[97, 98].

Catalytic activity

Beyond adsorption properties, screening procedures have been applied to chemical properties such as catalytic activities. Heterogeneous catalysis is generally performed using metallic non-porous structures, the use of nanoporous materials can increase dramatically the active surface area and the catalytic activity. Consequently, MOFs have been demonstrated to show catalytic properties for several chemical reactions. Just to cite a few, one can think of hydrogenation, hydrolysis, oxidation, among others explicitly covered by McCarver et al. in their review.[99] Considering the sheer number of possible materials, computational studies are potentially more effective than experimental ones. Therefore, computational screenings evolved in the last decade aiming at studying more sizable datasets.

Although the vast majority of computational screenings have been done on small series, there are a few systematic screenings of bigger datasets. The scarcity of the latter can be explained by the high level of computational cost required. Here, we show some examples of such attempts by focusing on the example of C−H bond activation for the conversion of alkanes into alcohols in the presence of nitrous oxide.

Inspired by enzymatic catalysis of the reaction of small alkanes with N_2O into alcohols, Vogiatzis et al. identified seven iron-containing MOF structures out of 5,000 structures from the CoRE MOF database.[100] They found two descriptors that govern the catalytic activity: 1) the N–O dissociation energy of N_2O on the adsorption site and 2) the energy difference between two spin states of the intermediate. Using a screening on these descriptors, three structures were identified as promising for further experimental studies. The best one has been computationally demonstrated to catalytically and selectively oxidize ethane to ethanol in presence of N_2O. Moreover, the authors found that defects played a major role in the observed catalytic activity.

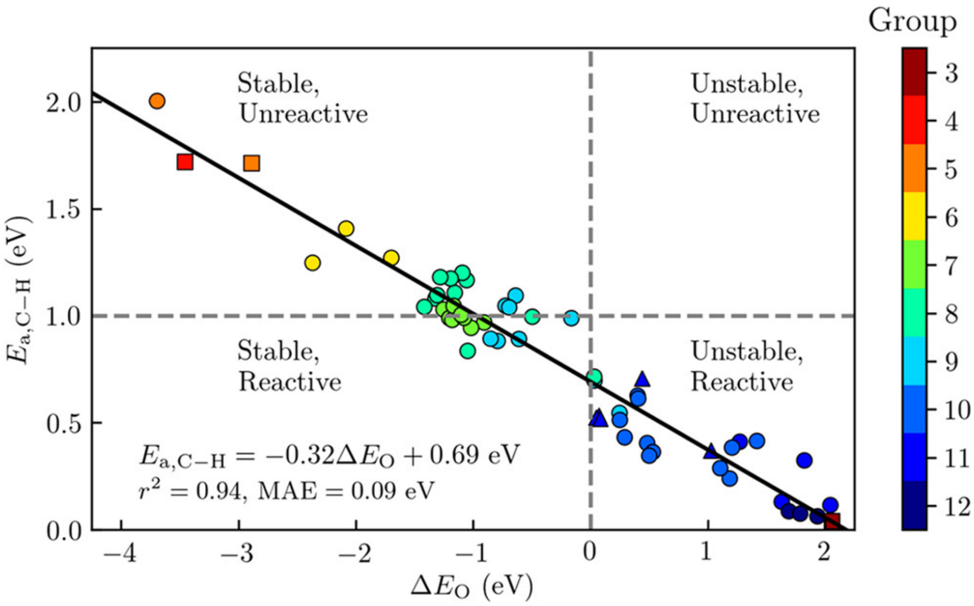


Figure 1.3: Analysis of a diverse set of experimentally derived metal–organic frameworks (MOFs) with accessible metal sites for the oxidative activation of methane. The graph shows the predicted barrier for the C–H bond activation of methane, E_a, as a function of the metal-oxo formation energy, ΔE_O. For each material, the symbol color refers to the group number of the metal in the periodic table. The best-fit line has been plotted in black, and has a mean absolute error (MAE) of 0.09 eV*. MOFs with E_a < 1 eV are classified as being reactive towards C–H bond activation and MOFs with $\Delta E_O < 0$ as having thermodynamically favored active sites when using O_2 as the reference state. Reprinted with permission from Ref. 101. Copyright 2019 American Chemical Society.*

Later, Rosen et al. enlarged the scope of materials screened to other metals.[101] From an 838 DFT-optimized MOFs subset of CoRE MOF 2014, the authors selected 168 MOFs that were likely to have open metal sites and pore-limiting diameters that allows the diffusion of the reactants. They then used a fully automated workflow to place the reactants in the adsorption site and relaxed the system using periodic DFT calculations. As shown in Figure 1.3, using the bond activation energy $E_{a,C-H}$ and the metal-oxo formation energy ΔE_O as key parameters, they classified the materials according to their relative stability and reactivity to find the best materials for the application. These energies were then analyzed using physicochemical descriptors such as the spin density on the oxygen and the metal–oxygen distance.

This type of brute force screening can be quickly cumbersome, as a result many researchers in the field are trying to find essential structure-activity relationships to accelerate future computational screenings. Several descriptors have been developed for high-throughput screenings: Butler et al. used electron removal energies to explain photocatalytic behaviors of MOFs;[102] Rosen et al. showed that the energy required to form the metal oxide intermediate was a major descriptor of the thermal catalysis of alkane oxidation by N_2O;[103] and Fumanal et al. show a screening protocol based on two energy-based descriptors to predict photocatalytic properties of MOFs.[104] Lately, Rosen et al. screened thousands of MOF structures to compare different DFT functionals and leveraged the data calculated to train machine learning models that can rapidly predict MOF band gaps.[105]

The development of ML methods are also critical in the field,[106] but the lack of centralized database with high precision descriptors is a challenge for the future of these methods. The influence of defects, the different ways of modeling MOFs as periodic structures or clusters, the diversity of structures and the stability of such structures remain open problems. Yet, it does not threaten the major role of high-throughput screenings in the early design process of any nanoporous materials for catalysis. To conclude this brief overview, we point the readers to a more exhaustive presentation of the matter.[107]

Mechanical properties

In the past decade, there has been a growing interest in the systematic study of physical properties of various classes of materials, including inorganic materials and framework materials. Among these physical properties, mechanical properties have been a topic of particular interest, as they are crucial for many applications, and at the same time can be computed by relatively standard methodologies. In particular, is it possible to calculate linear elastic constants (the second-order elastic tensor) in the zero-Kelvin limit by strain/stress or strain/energy approaches, performing a series of DFT calculations of strained structures and calculating the elastic constants. From these constants, all other mechanical properties can be evaluated by tensorial analysis,[108] including the bulk modulus, Young's modulus, shear modulus, Poisson's ratio, etc. This type of calculation can be coupled with any available quantum chemistry code,[109] and is even integrated in some packages, like CRYSTAL17.[110]

One of the first studies that investigated systematically the elastic properties of a family of materials was a 2013 study of all-silica zeolites,[111] i.e., crystalline and porous SiO_2 polymorphs. While this dealt with only 121 zeolitic frameworks out of 244 known structures, it showed that systematic studies at the DFT level were computationally tractable, and that they provided physical insight into the link between microscopic structure and macroscopic physical properties. This study demonstrated, among other things, that a few zeolites presented large negative linear compressibility (NLC), which could be linked to the wine-rack motif of their frameworks.

Outside the specific case of zeolites, other groups have applied DFT calculations of elastic constants in a high-throughput manner. de Jong et al. leveraged the structures of the Materials Project[75, 76], trying to chart the diversity of elastic properties across the whole space of inorganic crystalline compounds.[112] As shown in the Figure 1.4, they provided a database containing the full elastic information of 1,181 inorganic compounds initially, and has grown steadily since then, containing more almost 14,000 records to date.[113] This dataset has been used in two different ways by researchers in the field.

Firstly, the exploration of the database of elastic properties by tensorial analysis has allowed studying quantitatively the occurrence of certain "anomalous" or rare mechanical behavior, including negative linear compressibility, very high anisotropy, or negative Poisson's ratio (also called *auxeticity*). Indeed, such properties are considered rare and usually sought after — the materials exhibiting these anomalous behaviors are mechanical metamaterials.[114] In addition to their fundamental interest, such materials have applications in materials engineering: for example in energy dissipation (as shock absorbers and for bulletproofing), energy storage, as well as acoustics.[115] However, it was not possible until now to quantify exactly "how rare" they are. Chibani et al. showed through a systematic exploration of available mechanical properties of crystalline materials that general mechanical trends, which hold for isotropic (noncrystalline) materials at the macroscopic scale, also apply on average for crystals. Moreover, they could quantify the presence of materials with rare anomalous mechanical properties: 3% of the crystals were found to feature negative linear compressibility, and only 0.3% to exhibit complete auxeticity (negative Poisson's ratio in all directions of space).

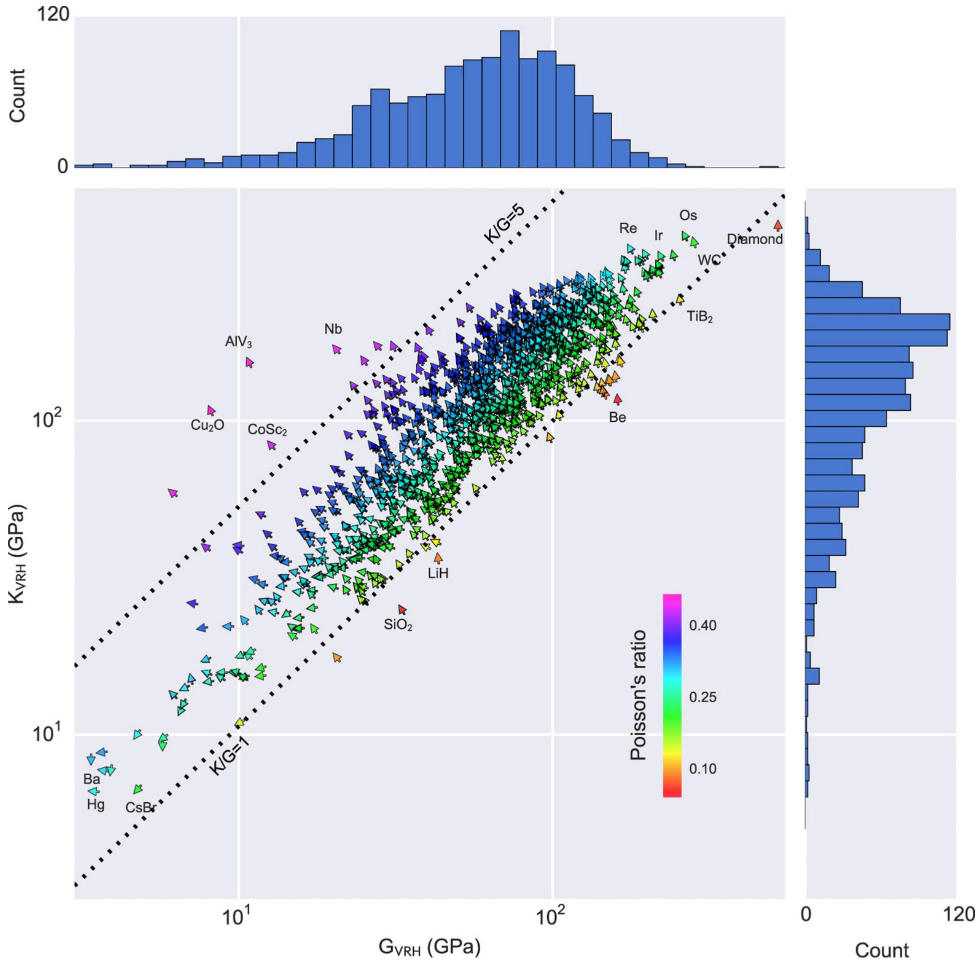


Figure 1.4: Statistical analysis of the calculated volume per atom, Poisson's ratio, bulk modulus K_{VRH} *and shear modulus* G_{VRH} *of 1,181 compounds in the Materials Project database. In the vector field-plot, arrows pointing at 12 o'clock correspond to minimum volume-per-atom and move anti-clockwise in the direction of maximum volume-per-atom, which is located at 6 o'clock. Reprinted from Ref. 112 under CC-BY license. Copyright 2015 de Jong* et al.

Secondly, the datasets of mechanical properties were used as a basis to accelerate the discovery of novel materials with targeted behavior. Dagdelen et al. used search algorithms to identify 38 candidate materials exhibiting features correlating with auxetic behavior, from more than 67,000 materials in the Materials Project database.[116] Performing DFT calculations on these 38 structures, they could identify 7 new auxetic compounds. In a more complex setup, Gaillac et al. [91] have used a multiscale modeling strategy for the fast exploration and identification of novel auxetic materials. They combined classical force fields MD simulations with DFT calculations on candidate materials, and then used this reference DFT data to train an ML algorithm. They found that the accuracy of this multiscale method exceeds the current low-computational-cost approaches for screening. In a similar work, Moghadam et al. used molecular simulation to train an artificial neural network (ANN) for the prediction of the bulk modulus of metal–organic frameworks.[117] This shows the potential of such methodologies to treat very different (chemically as well as structurally) classes of materials.

Thermal properties

While mechanical properties (in the elastic regime) have been by far the most studied physical property in nanoporous materials, others have also been occasionally screened. We can cite, in particular, the systematic study of piezoelectric tensors by de Jong et al., on almost a thousand crystalline compounds, by first-principle calculations based on density functional perturbation theory.[118] We can also cite efforts to calculate thermal properties in a high-throughput setup, using the quasi-harmonic approximation (QHA).[119] This method requires the calculation of each structure's phonon modes at various volumes, and can be coupled to any electronic structure program.[120] It is, however, quite computationally intensive, and sensitive to the parameters of the QHA methodology (range of volume, range of temperature, precision of the frequency calculation, etc.). Therefore, it has been limited so far to modest numbers of structures: a dataset of 75 inorganic structures by Toher et al.,[92] and more recently a dataset of 134 pure SiO_2 zeolites by Ducamp et al.[94] Very recent work in our group on the prediction of thermal properties through machine learning based on structural features alone indicates that thermal behavior is more difficult than mechanical behavior to predict, and might require the use of a wider set of structural descriptors or more advanced ML models.[96]

1.2.2 Transport adsorption properties

The thermodynamic properties, we will be presenting in the next section, only describes the state of equilibrium of the adsorption process. But sometimes the transient state can last long before reaching the equilibrium, which makes the process more time-consuming. Thus, the transport properties complete the thermodynamic description of the adsorption process inside a nanoporous material. For example, a low diffusion rate would mean for storage applications more time and energy needed to fill-up the tanks, or for separation applications a less selective process than expected. In more extreme cases of molecular sieves for fluid separation, the transport properties become predominant to assess the performance. One can leverage the difference of the molecules' diffusion coefficients to selectively filter gas mixtures through a nanoporous membrane.[121] Here, the main subject becomes the transient state and not the equilibrium. This section is thus dedicated to the kinetics of the adsorption process to better model the time required to reach the equilibrium or to study out-of-equilibrium processes such as molecular sieving by nanoporous membranes.

Kinetic properties

In most computational screenings, the diffusion coefficient considered is the self-diffusion coefficient that describes an infinite-dilution case. Other multi-component diffusion coefficients could be considered, but for simplicity and clarity they won't be mentioned in this review. The calculation of the self-diffusion coefficient gives a first estimation of the kinetics in a storage or a separation process in the limit of low adsorption loading.

There are two approaches to estimate the diffusion inside a porous material: the first one relies on molecular dynamics (MD) and the second one on transition state theories. In the first approach, one can analyze the mean squared displacement of the adsorbed molecule moving in the material. In the second, one identifies minimum energy path along the material to identify transition states (TS) to calculate diffusion energy barriers. The MD-based method requires fewer assumptions and is therefore more reliable than the TS-based method, but the latter is computationally more efficient in the case of low diffusion rate (diffusivity lower than 10^{-11} m^2 s^{-1}).

State-of-the-art MD simulations could calculate rather accurate diffusion coefficients, but the computational cost scales quickly with the number of structures. To use this method on a large dataset without spending too much computation time, Watanabe and Sholl prescreened the pore sizes of 1,163 MOFs to select only the structures within a certain range of PLD (pore limiting diameters).[122] A restricted list of 359 MOFs was then used to carry out MD simulations to calculate diffusion coefficients. The results of this final screening are then used to extract the most promising structures for further experimental or computational investigation. Similarly, Qiao et al. used a multistage screening to find the best membrane material within about 130,000 hypothetical MOFs for a $CO_2/N_2/CH_4$ separation.[86] They started to select materials based on pore geometry analysis; then they calculated Henry's coefficient and diffusion coefficients at infinite dilution; finally, they compared the binary permselectivities to extract 24 promising MOFs for ternary adsorption and diffusion calculation at the desired pressure and temperature conditions.

Another approach replaces MD simulations with more computationally efficient TS-based methods to determine diffusion coefficients. Haldoupis et al. developed an algorithm to identify diffusion paths by exploiting an energy grid with a clustering algorithm. The diffusion paths are then analyzed to identify the pores and the channels, and to calculate key geometric (the PLD or the largest cavity diameter) and energetic (Henry's constant, diffusion activation energy) features.[123] As illustrated in Figure 1.5, they found a clear dependence of the diffusion energy barrier to the PLD. As one of the first TS-based screenings, it is still subject to many development perspectives. For instance, the approach is limited to spherical adsorbates and rigid frameworks. Moreover, the diffusion coefficients are approximated using a simplistic hopping model for a qualitative analysis. This method is highly efficient, but the accumulation of approximations makes a quantitative systematic analysis of diffusion coefficients out of reach.

Later, Kim et al. introduced a flood fill algorithm to obtain all the points within a given energy.[124] These points are then identified as channels or blocked regions. Along the channels, local minimums of energy are defined as lattice sites and transition states are defined perpendicular to the diffusion direction. A random walk is then computed along the lattice

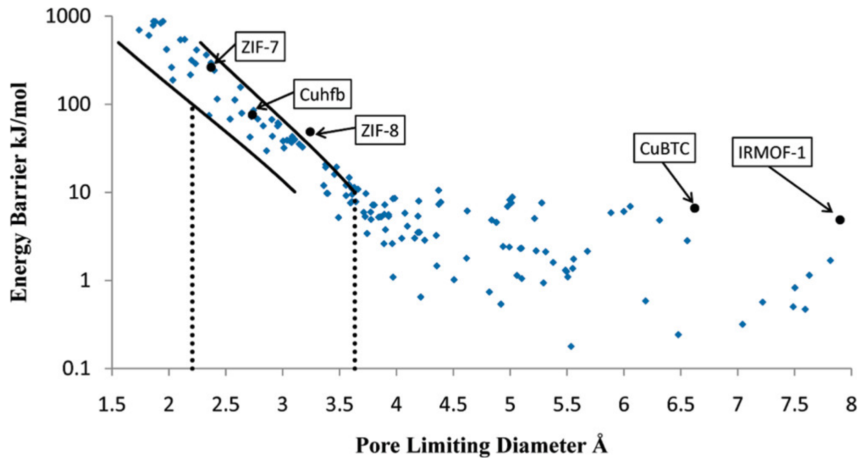


Figure 1.5: Calculated energy barrier for the diffusion of CH_4 in 216 metal–organic frameworks (MOFs), shown as a function of the pore-limiting diameter. The solid lines represent statistical upper and lower bounds on the energy barrier, in a transition state theory approach. Reprinted with permission from Ref. 123. Copyright 2010 American Chemical Society.

sites with hopping rates defined according to the activation energy. A diffusion coefficient is then calculated in each three directions of the space and an average diffusion coefficient is finally determined. A comparison with the MD method on the IZA zeolite structures shows good agreement, but there are still some discrepancies explained by correlated hops in the case of rapid diffusion or by the presence of complicated channel profiles. Inspired by this work, Mace et al. developed a similar method that progressively fill the energy grid to detect transition states, hence removing the previous restriction to orthogonal cells only.[125] The diffusion coefficient is now computed using a kinetic Monte Carlo simulation allowing the adsorbate to jump freely in all directions instead of restricting it in a single dimension. This new method, called TuTraSt, handles very complex diffusion paths (like in the AEI zeolite). This new approach seems to be promising as it is in good agreement with MD simulations, while being 2-3 orders of magnitude faster. However, the time performance could improve tremendously by translating it from Matlab to C++ and by implementing parallelization procedures.

Very recently a massively parallel GPU-accelerated string method has been implemented and shared publicly to compute very efficiently diffusion coefficients based on the transition state theory.[126] The recent developments in the prediction of diffusion coefficients in nanoporous materials point towards a promising future for the screening of transport properties applied to even larger databases. Going further, Bukowski et al. reviewed thoroughly diffusion in nanoporous solids as an attempt to connect theory to experiments.[127]

Membrane materials

In separation application, the study of the transport properties can evaluate the feasibility of the thermodynamic equilibrium, crucial for any bed separation process. If this separation is not feasible, kinetic separation or partial molecular sieving are to be considered. Some notable examples are air separation in zeolites using pressure swing adsorption,[128] N_2/O_2 separation in carbon molecular sieves,[129] or N_2 removal from natural gas.[130] In kinetic separation, the valuable metric is not the selectivity anymore, but the permselectivity, i.e. the product of the selectivity and the permeability (ratio of diffusion coefficients). Therefore, the screening of diffusion coefficients gives complementary information to the thermodynamic selectivity screenings. Here, we give some examples of such screening and the main descriptors that partially explain the computed figures of merit.

To give an overview on the potential of computational screenings to predict transport properties, we are now going to focus on the membrane separation applied to natural gas upgrading. The separation of CH_4 from N_2 and CO_2 is a crucial step of this upgrading process. In 2016, a large-scale high-throughput screening (see Figure 1.2 for the approach) of hypothetical MOF membranes for upgrading natural gas has been performed using MD simulations.[86] Qiao et al. confirmed the existence of MOF materials beyond the upper bound for N_2/CH_4 and CO_2/CH_4 separations determined by Robeson on a large set of polymeric membranes.[131] This Robeson's upper bound is systematically crossed by MOF materials in computational screenings, see as an example the Figure 1.6. This can be explained by the fact that MOFs perform better than polymeric frameworks and the simulations at this level of theory. They also identified 24 MOFs suitable for the ternary $CO_2/N_2/CH_4$ separation using a multistage screening described in the previous section.

Two years later, Qiao et al. used the same approach to study this ternary separation on a database of synthesized structures.[132] Applying machine learning techniques to their data, they performed a QSPR analysis. Using a principal component analysis, they notably found that the permeability is higher when materials have high PLD and void fraction coupled with low density and percentage of pores within a characteristic range. The opposite was found to be true for high membrane selectivity for the CO_2/CH_4 separation. Using decision tree algorithms, they gave objective procedures of selecting the best separation membranes based on some key descriptors. Finally, they studied in detail some top performing materials found by a support vector machine algorithm.

Altintas and Keskin later performed a screening on the same database for CO_2/CH_4 membrane separation to identify the best performing materials and perform more computationally demanding simulations.[133] The simulations in rigid structures at infinite dilution show numerous structures above the Robeson's upper bound as shown in the figure 1.6, this crossing of the upper-bound can be explained by either a better performance of MOF membranes compared to the polymeric membranes used by Robeson, or an overestimation due to oversimplified assumptions (infinite dilution, rigidity). But when higher pressures and flexibility are considered, the selectivity values are dropping down closer to the upper boundary, hence confirming the overestimation of the performance in screenings based on rigid approximations at infinite dilution. But the best performing materials are still above the Robeson's upper bound and can therefore be used in mixed matrix membranes with polymeric membranes. Budhathoki et al. developed a screening methodology for MOFs in mixed matrix membranes for carbon capture applications by estimating permeation values in these composite materials using a Maxwell model.[134] The authors even proposed a pricing for each material compared to their relative performance. Similar studies have been carried out on different materials, Yan et al. showed the influence of decorating COFs with different chemical compounds on the membrane selectivity.[135]

The transport properties screening is based on the calculation of diffusion coefficients at infinite dilution and in rigid molecules. There are different methods to calculate them (mainly MD and TS-based methods). Flexibility and pressure dependence are very hard to incorporate directly in

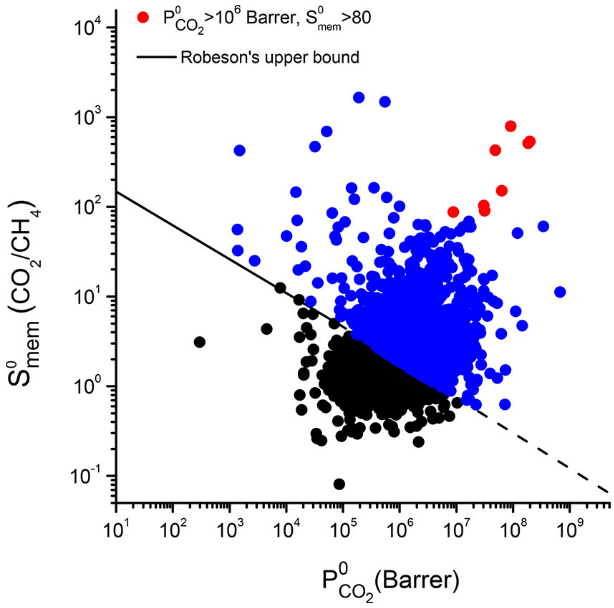


Figure 1.6: Selectivity and permeability of metal–organic framework (MOF) membranes for CO_2/CH_4 separation computed at infinite dilution by combining Grand Canonical Monte Carlo and molecular dynamics simulations.[133] The black solid line represents the Robeson's upper bound.[131, 136] MOFs that can exceed the bound are shown in blue, and the 8 top-performing MOF membranes are shown with red symbols. Reprinted with permission from Ref. 133. Copyright 2018 American Chemical Society.

the screening procedures. Researchers usually consider these factors at the end of the screening on the most promising structures because of the computational complexity of the corresponding simulations. To take account of pressure dependence, we need an MD simulation of several adsorbates that takes much more time than running single component simulations,[137, 138] which makes it harder to include in a high-throughput screening. Flexibility could be taken account by calculating snapshots and running multiple MD simulations, or by using flexible force fields, which means in both cases an increase in computational run-time. Some faster methods of quantitatively predicting the impact of flexibility on diffusion are being investigated in ZIFs and could give an interesting alternative to these expensive methodologies.[139]

1.2.3 Thermodynamic adsorption properties

In its early development, computational screening was mainly used to predict thermodynamic properties in adsorption processes. Three main applications have been identified in the associated literature: gas storage (for energy or medical applications), gas separation (noble gas, hydrocarbons, carbon dioxide, etc.) and post-combustion CO_2 capture. These applications are closely linked to urgent environmental and energy issues that are yet to be solved. Screening can guide the development of better performing materials by shedding light upon unknown structure-property relationship, probes possible theoretical limitations (unreachable targets) and identifies potential candidates that need to be experimentally tested.

Gas storage

One can leverage the high surface density of the nanoporous materials, especially the MOFs, to stock in very low-density gas. In the field of energy storage or transportation, natural gas (mainly methane) or hydrogen are considered plausible alternative fuels to replace conventional ones for transport. The US Department of Energy (US DOE) recently financed research programs and set targets for methane and hydrogen storage. Nanoporous materials could reduce energy, infrastructure and security cost due to the required compression and cooling. In this section, we are focusing on high-throughput screening for methane storage in nanoporous materials, before broadening the scope hydrogen and other perspectives.

One of the pioneering works in computational screening was published in 2012 by Wilmer et al.[10]. They performed a large-scale screening of 137,953 hypothetical MOF structures to estimate the methane storage capacity of each MOF at 35 bar and 298 K based on the US DOE standards. Back then, the US DOE set a target methane capacity value of 180 $\mathrm{vol^{STP} vol^{-1}}$ (which has since been achieved by several materials reported in the literature). In their large-scale analysis, Wilmer et al. found over 300 hypothetical MOFs that meet the targeted requirements and the best one can store up to 267 $\mathrm{vol^{STP} vol^{-1}}$, surpassing the state-of-the-art of the time. From their large dataset, a preliminary structure-property relationship analysis revealed that void fraction values of approximately 0.8 and gravimetric surface areas in a range 2500-3000 $\mathrm{m^2\,cm^3}$ resulted in the highest methane capacities. Optimal pore size is also shown to be around the size of one or two methane molecule(s). Maximization of gravimetric surface area was a common strategy in the MOF design for storage applications, but this study showed the existence of an optimal range of surface area values. Computational screenings can draw clear relationships between structural descriptors and performance. Later, a more quantitative relationship was drawn by Fernandez et al. using ML models as illustrated on Figure 1.7. Beware not to over-interpret the relation given by the response surface, since the identified maxima do not always have a physical reality, especially where there is no training data in the area pointed by the red arrows. However, it highlights promising unexplored feature space and shows potential research directions.

Since then new materials above the target have been found and the US DOE decided to set a higher target of 315 $\mathrm{vol^{STP} vol^{-1}}$. Until now, this new target is not yet reached. This is why the recent developments have focused on assessing the feasibility of such a target by accelerating the screening methods so that more data can be screened, and by interpreting the QSPR models to extract important knowledge for the design of novel materials. For instance, Gómez-Gualdrón et al. showed that even by artificially quadrupling the Lennard-Jones interaction factor ϵ and by increasing the delivery temperature by 100 K, the newly set target is only reached by a handful of MOFs.[140] This study suggests the impossibility to reach the DOE target using a preconceived (experimentally or theoretically) material to store methane. However, this theoretical limitation can be overcome by increasing the surface density of sites with high affinity with methane and by increasing the delivery temperature.

Later, a larger-scale screening on methane storage was carried out by Simon et al. on 650,000 experimental and hypothetical structures of zeolites, MOFs, and PPNs. This study confirmed that the classes of materials currently being investigated were unlikely to meet the new target. The authors suggested that it wasn't surprising since the target was based on economic arguments, while the screening is based on thermodynamic arguments.[61] This example illustrates the power of large-scale screening to settle questions of physical feasibility (if simulations are accurate) and hence avoiding experimental efforts spent on impossible tasks.

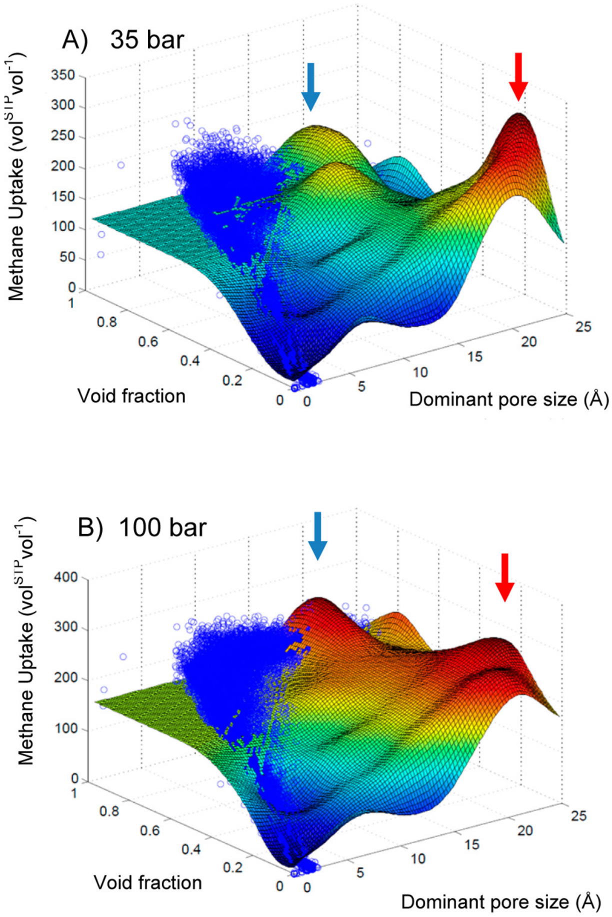


Figure 1.7: Two-dimensional response surfaces of the support vector machine (SVM) models trained by Fernandez et al. for methane storage at (A) 35 bar and (B) 100 bar using void fraction and dominant pore size. The blue dots represent the GCMC simulated uptake values. The color of the surface represents the methane storage value, from blue (the lowest values) to red (the highest values). Blue and red arrows indicate maxima on the response surface. Reprinted with permission from Ref. 82. Copyright 2013 American Chemical Society.

More recently, a dataset containing trillions of hypothetical MOFs have been screened for methane storage.[141] Lee et al. developed a methodology using machine learning combined with genetic algorithm to perform the largest screening until now. In addition to confirming most of the results (theoretical limits and QSPR) found by previous screenings, 96 MOFs were found to outperform the current world record. This study shows the scaling potential of ML-assisted screenings in handling "Big data".

Similarly, computational high-throughput screenings have been applied to other storage applications such as hydrogen storage. Computational screenings showed that cryogenic storage of hydrogen can meet the DOE target of 50 g L^{-1}.[62, 79, 142] Anderson et al. performed a large-scale screening based on neural networks to test out multiple pressure/temperature swing conditions to find that the maximal deliverable capacity cannot exceed 62 g L^{-1}.[143] Compared to the density of liquid hydrogen (72 g L^{-1}), this upper limit seems reasonable since the adsorbent material takes at least 10-20% of the tank. Here, we only showed some flagship results of the field. For a more detailed meta-analysis, Bobbitt and Snurr wrote a very complete review on computational high-throughput screening of MOFs for hydrogen storage.[144]

1.2.4 Gas separation

As a representative example of what could be done in the field of gas separation, we are going to focus on Xe/Kr separation. Nanoporous materials can be used as a safer, cheaper and less energy-intensive option for this gas separation. However, experimental design of top-performing materials can be cumbersome. Computational screenings is an ideal tool to kick-start the development of this new technology by identifying rapidly the best candidates.

Small-scale screenings

Metal–organic frameworks, and later other supramolecular porous materials like covalent organic frameworks (COFs), have been proposed for applications in separation of noble gases for a decade. With no aim of being exhaustive, we highlight some milestones in that area, both from experimental and computational point of view.

In 2012, Liu et al.[145] published an experimental study of two MOFs, HKUST-1 and Ni/DOBDC, for adsorption of Xe and Kr at ppm (part-per-million) levels in air. The target application was the removal of Xe and Kr from nuclear fuel reprocessing plants. The same group later proposed a two-column method for the separation of Kr and Xe from processed off-gases[146], based on MOF materials. At about the same time, Bae et al.[147] combined a computational Grand Canonical Monte Carlo (GCMC) study with experimental breakthrough measurements of the separation of a Xe/Kr mixture on MOF-505 and HKUST-1.

Parkes et al.[148] studied sixteen different MOF materials for Kr, Ar, and N_2 adsorption and separation, through GCMC simulations. They concluded on the potential of MOFs for separation, and a general correlation between the Henry's constant and the isosteric heat of adsorption for the three gases studied. A year later, in 2014, Chen et al.[4] demonstrated, again through a combined computational and experimental study, the potential of porous organic cages for selective binding of xenon over krypton.

Later experimental work expanded these early separation studies to different types of MOF materials. Xiong et al.[149] studied a flexible zinc tetrazolate framework for xenon selective adsorption over krypton, argon and nitrogen. Thermodynamic analysis of the adsorption isotherms at various temperatures confirmed the occurrence of a "breathing" structural transition upon Xe uptake, contributing to a high working capacity for a pressure swing adsorption (PSA) cycle. Lee et al.[150] compared the selective adsorption properties for Xe/Kr mixtures on three highly studied MOFs, namely UiO-66(Zr), MIL-100(Fe) and MIL-101(Cr), and confirmed a high potential of UiO-66(Zr) for separations under dynamic flow conditions. These authors also assessed the hydrothermal and radioactive stability of the material, a test seldom performed in the existing literature, and found it to be good. In a further study,[151] they demonstrated that Xe/Kr selectivity could be further improved by ligand substitution.

In parallel, computational studies were published to provide insight at the microscopic level into the mechanisms behind good (and bad) separation properties. Wang et al.[152] studied 6 MOFs and COFs for adsorption of Xe and Xe/N_2 separation, through GCMC simulations looking at the impact of pressure (and therefore pore filling) on selectivity. Anderson et al.[153] combined GCMC and biased MD simulations to elucidate the nature of adsorption- and diffusion-based Kr/Xe separation mechanisms in four archetypal nanoporous materials: SAPO-34, ZIF-8, UiO-66, and IRMOF-1. These authors draw a couple of general conclusions, including the fact that diffusion selectivity for krypton dominates membrane separation selectivity, and large pore cages and stiff pore windows are desirable — however the scope of these conclusions is inherently limited by the small number of materials actually studied.

In a different family of materials, Tong et al.[65] have surveyed the structure–property relationships of covalent organic frameworks (COFs) for noble gas separation, by GCMC simulations of 187 different materials for Kr/Ar, Xe/Kr and Rn/Xe separations. These authors included in their calculations some adsorption figures of merit (AFM), representative of the conditions of industrial vacuum (VSA) and pressure swing adsorption (PSA) processes.

One area that has been particularly explored is the tuning and improvement of separation properties through the presence and nature of coordinatively unsaturated sites (or open metal sites) in MOFs. In 2016, Vazhappilly et al.[154] used density functional theory (DFT) calculations of host–guest binding energies to probe the impact of the metal atoms in a specific framework (MOF-74) on Xe and Kr adsorption. Later, Zarabadi-Poor et al.[155] investigated — again through computational methods — a series of metal–BTC MOFs for recovering xenon from exhaled anesthetic gas, i.e., mixtures of CO_2, O_2, and N_2.

Large-scale computational screening

In its early stage, computational screening has been used on a small series of nanoporous materials to generate specific knowledge on some subclasses of materials. These small-scale screenings combined with experiments helped faster identification of good performing candidates, but they failed to establish general rules of design or to explore the unknown. Larger-scale screenings overcame these limitations by trying to exhaustively cover the whole spectrum of nanoporous materials.

The first large-scale computational screening on Xe/Kr adsorption-based was performed by Sikora et al. based on the same approach previously developed for methane storage by their group at the Northwestern University.[156] This study was based on the same 137,000 structures of hypothetical MOFs.[10] They calculated the Xe/Kr selectivity using Monte Carlo molecular simulations on the whole database by iteratively increasing the number of steps and selecting the best materials similar to the approach on Figure 1.2. By analyzing the relationships between pore sizes and selectivity, they confirmed a hypothesis from a smaller scale study that the pores should be between the size of 1 to 2 xenon molecules.[157] Tube-like channel was also found to favor better selectivity. Moreover, they found that top performing materials could have a selectivity around 500; but we can only conclude on the order of magnitude of the theoretical limitation of the Xe/Kr selectivity, considering the statistical uncertainty of the simulation.

Seizing the opportunity of a formidable expansion of the nanoporous materials database triggered by the Materials Genome Initiative, Simon et al. screened 670,000 experimental and hypothetical nanoporous material structures for Xe/Kr separation (see Figure 1.8).[14] It is one of the largest-scale screenings performed in this area. Inspired by the work of Fernandez and co-workers,[82] they used ML algorithms to train a model on a diverse subset of 15,000 structures. This method allowed them to run time-consuming molecular simulations only on this training set, before applying the ML model to predict the selectivity values on the larger set of structures. On top of analyzing the links between pore descriptors and selectivity, they rationalized it using theoretical pore models of spherical and cylindrical geometries to confirm the findings of Snurr and co-workers.[156, 157] By comparing the structural descriptors of good-performing

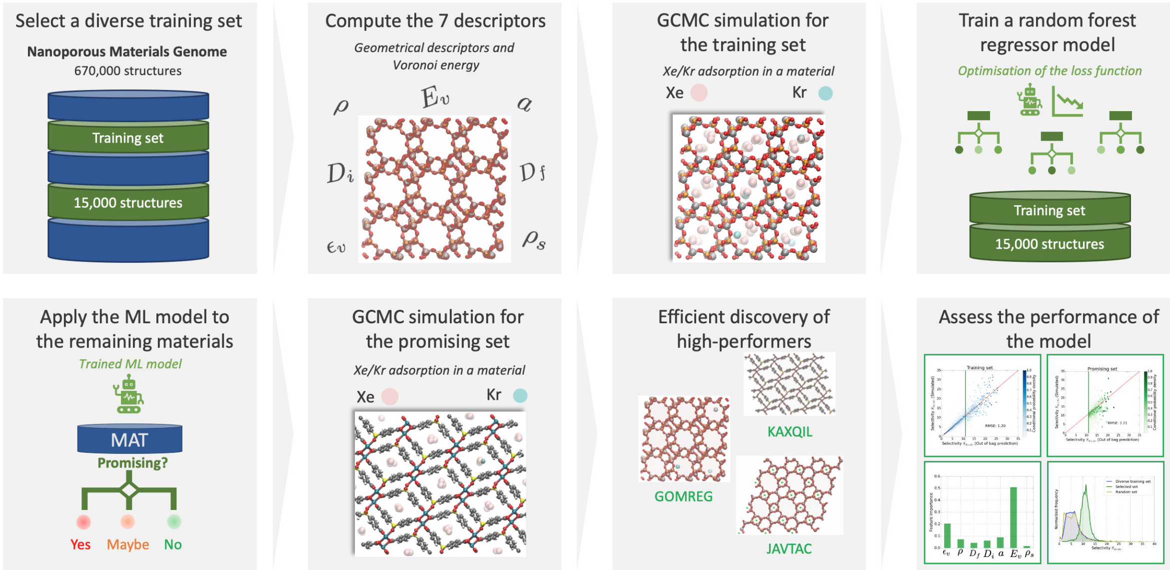


Figure 1.8: Schematic representation of large-scale screening of nanoporous materials for Xe/Kr adsorption-based separation by Simon et al.,[14] based on a combination of Grand Canonical Monte Carlo simulations and machine learning algorithm (Random Forest Regressor). The main goal of this screening 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.

and bad-performing structures, they concluded that geometrical descriptors wasn't enough to explain the performance. The analysis of a few top candidates suggests that different chemical insights could explain their good performance. For SBMOF-1 or KAXQIL,[47] an experimental MOF, its higher performance was explained by the tubelike 1D channel with a very favorable binding site formed by carbon aromatic rings. This nanoporous material was later tested using breakthrough experiments and proved to be one of the most promising candidates.[158] This close collaboration between computation and experimentation is a testimony of the potential of computational screenings to find nanoporous materials for any targeted application.

The experimental work on Xe/Kr separation on SBMOF-1 revealed discrepancies between the selectivity values obtained experimentally and computationally.[158] The assumption of rigid crystal structures in the molecular simulations could partially explain the difference observed. Witman et al. proposed that the flexibility of the materials that weren't considered in the screening of Simon et al. could explain the lower selectivity observed experimentally.[159] In this study, they screened the Henry regime separation of about 4,000 MOF structures of the CoRE MOF 2014 database[63], and found that intrinsic flexibility, i.e. the thermal vibration of the material, can make the pore size derive from the ideal value for the separation and hence lower the selectivity. This study further confirms the importance of the pore size by highlighting the effect of its evolution over time.

In 2019, Chung et al. screened the most extensive simulation-ready and experimentally synthesized MOF structures for Xe/Kr separation.[64] This study pointed out the potential of coordinated solvent molecules to fine-tune the selectivity for any separation application, since their presence can enhance selectivity in some cases. The results of their screening confirm the potential of structures such as SBMOF-1 found by Simon et al., but they also described a few structures with similar selectivity but with better xenon uptake. The authors emphasize the importance of considering other figures of merit such as the adsorption capacity. Other factors should be taken into account to find the best trade-off between all the relevant figures of merit; we could think of the kinetics of such a separation, the effect of flexibility on the performance, the stability of the materials (especially in radioactive environment), the financial aspects, and more.

After this quick overview of the different screening studies in the field of xenon/krypton separation, we are now going to detail its industrial context, the foreseen top materials that could fulfill the industrial separation and what further studies are needed to better understand the process while discovering new materials.

1.3 Separation of Xenon from Krypton

In this section, we will try to see how we can apply the above-mentioned screening methodologies to help us understand the origins of the Xe/Kr separation and identify promising materials for industrial applications.

1.3.1 Industrial applications

The industrial interest for noble gases lies first in the many applications attached to them. For instance, xenon has multiple applications in the medical (e.g. anesthesia, painkiller, imaging), [160–162] aeronautical[163, 164], lithographic[165], microelectronic[166] or lighting sectors, [167, 168] just to cite a few. To meet the demand for these noble gases, one should consider all available sources, the most obvious one being the air we breathe. Xenon and krypton have both very low atmospheric concentrations; out of a thousand liters of air, we would extract at most one tenth of a milliliter of xenon and one of krypton.[1] Nevertheless, direct extraction from the air remains the main production mean for xenon and krypton along with chemical plant off-gases that contains a higher concentration of inert gas (e.g., amonia purge gas). In these cases, the industry more commonly uses cryogenic distillation to extract xenon and krypton, which requires a compression and cooling of the gas mixture at very low temperatures. The separation process can be broken down into three steps: first the condensation of all gas with a boiling point higher than the oxygen, then the purification of oxygen resulting in a 20-80 xenon/krypton mixture, and finally the separation of xenon from krypton. In 1997, several cases of explosion of separation units were caused by the reaction of non-filtered dangerous hydrocarbons with purified liquid oxygen produced in the second step of this long process.[169, 170] The extreme chemical and physical conditions required for cryogenic distillation support the need for less energy-intensive and safer alternatives.

Industrial application	Xe/Kr composition
Extraction from ambient air[1]	20/80
Spent nuclear fuel[171]	90/10
Molten Salt Reactor[172]	?

Table 1.1: Composition of the Xe/Kr mixture in different applications.

The role of a dispatchable source of low-carbon energy can only be fulfilled by batteries charged by renewable energies (wind or solar) or by nuclear plants. However, one of the major criticisms on this source of energy concerns the management of the radioactive waste. As promising technologies in gas separation emerge, there is an increasing need for a solution for the release of very small amount of radioactive off-gases like Kr_{85} from nuclear spent fuels.[173] Furthermore, stable xenon isotopes are also produced in these spent nuclear fuels, which can be used in all the above-mentioned applications. In the context of a regained interest in nuclear energy, the fourth generation nuclear plants are projected to be built on other technologies such as the light water or the molten salt technologies.[174] Molten salt reactors would continuously produce xenon and radioactive krypton in the electricity generation process.[172] The development of gas separation units in these facilities would represent a promising source for xenon production. Yet, we can laboriously imagine deploying standard cryogenic distillation units in a nuclear facility for obvious security reasons. Consequently, nanoporous materials are considered as the alternative technology for xenon/krypton separation. Zeolites are already used as a pre-purification system,[1] and they are now projected to be used as a standalone separation system.

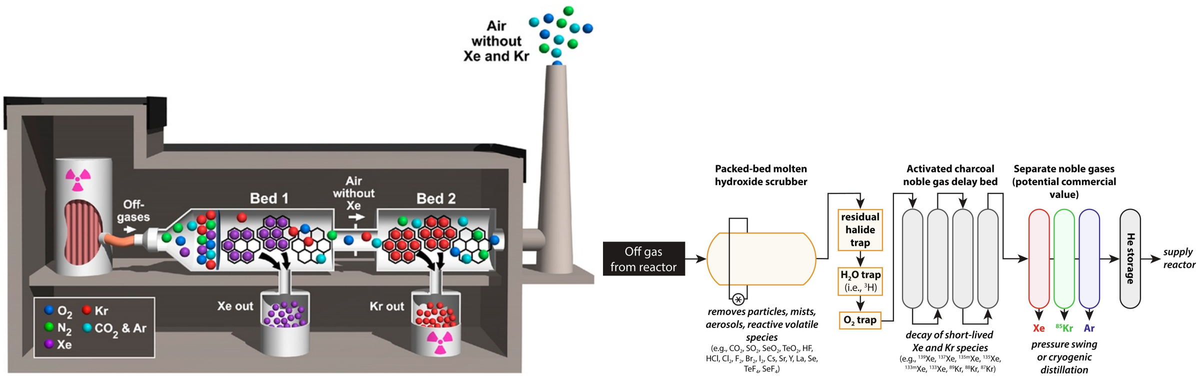


Figure 1.9: Representation of xenon/krypton separation process using porous materials in a nuclear fuel reprocessing plant and in a molten salt reactor. Reprinted with permission from reference 175 copyright ©(2014) American Chemical Society and reference 172 copyright ©(2019) Elsevier.

Banerjee et al. proposed a two-bed system with a first bed filled with MOFs designed for xenon separation and then a second one for radioactive krypton capture.[175] The authors proposed some examples of material that could be used for this separation unit; more research is needed to find out what the best materials for these separations are. In the following section, we will review the most promising materials for this separation and the structural explaining their high performance.

1.3.2 Promising materials for the separation

Several experimental reports used the strategy outlined by computational screenings to improve separation properties, as well as tuning the chemical nature of the organic linkers. The main criteria outlined by the different studies on xenon/krypton separation call for pore size tailor-made for xenon and also for maximized interactions with the framework atoms obtained either through the chemical nature or the shape of the cavities.

In the early phase of the experimental design of materials for the xenon/krypton separation, Wang et al. synthesized a cobalt MOF $Co_3(HCOO)_6$ with a selectivity of 12 that present rather narrow pores (around 5 Å) connected by zig-zag segments.[152] Later, Chen et al. synthesized a selective porous cage material by not only focusing on the pore size but more importantly on the shape of the cavity, the selectivity of around 20 was considered record high at that time. For instance, the cage windows are open for small noble gases such as krypton, whereas they close around the xenon hence maximizing the interaction.[4] Mohamed et al. also designed a material with a similar selectivity,CROFOUR-1-Ni. However the performance was now explained by the chemical nature of the chromium oxide ligands that interact more strongly with the more polarizable xenon than the krypton molecules.[176] Finally, Banerjee et al. tested a previously synthesized[47] MOF after it was identified through high-throughput screening[14] for its outstanding theoretical selectivity around 70. However, experimental measurements showed that its selectivity was not exceeding the one of the previous top materials. Similar emphases were made on the ideal pore size coupled with highly attractive framework atoms.[158]

More recently, Li et al. proposed a rigid squarate-based MOF with "perfect pore size" (comparable with the kinetic diameter of Xe), and an internal pore surface decorated with very polar hydroxyl groups. This material experimentally demonstrated record-high Xe/Kr selectivity of 60.6 at low pressure (0.2 bar) and ambient temperature.[5] Later, Pei et al. discovered even better performing materials with Xe/Kr selectivity of 74.1 and 103.4 in the same conditions 0.2 bar and 298 K. In addition to the perfectly tailored pore size, the structure features two oppositely adjacent open metal sites that strongly clamp the adsorbed xenon molecule.[6] These studies clearly show the potential of polar sites that preferentially interact with the more polarizable xenon over the krypton, hence explaining these record-breaking separation performances.

1.3.3 From the computer to the test tube

To connect back our study to computational screenings, we are now going to present one of the rare cases of direct contribution of high-throughput screenings to the lab testing of a material. In 2015, Simon et al.[14] analyzed the Nanoporous Materials Genome,[61, 77] 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. This study led to the rediscovery of SBMOF-1, a promising nanoporous material that was presented one year later.[158]

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 multiscale modeling 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. 1.8 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, which suggests the key role of the chemical nature of the cavities.

By looking at the distribution of the most selective materials ($s \geq 14$) compared to the less selective ones ($s < 14$) in the Figure 1.10, Simon et al. established a first profile of the selective materials. These materials have pore sizes of a specific diameter very close to the kinetic diameter of xenon around 4.4 Å depending on how it is defined. They have rather low surface areas and porosities (void fractions) unlike what we would normally expect since the adsorbable surface is a key reason for using nanoporous materials in adsorption applications. This behavior can be rationalized by the fact that small pores of the order of a few Å drives mechanically to smaller pore volumes and surface areas (the framework atoms have more space). The crystal density is therefore also a bit higher for these reasons. Moreover, the pore's shape is also a crucial factor since a shape closer to a sphere would interact with the xenon with more atoms, hence increasing its affinity and the selectivity. Finally, last but not least, the Voronoi energy described the physical nature of the binding between the xenon and the pore atoms, the more negative it is and the more selective the material will be. To wrap up, the ideal materials have a pseudo-spherical shape (a complete sphere would stop the diffusion of the adsorbates) with a size close to the diameter of a xenon which is rather dense and not very porous.

The chemical nature of the cavities was best described using the Voronoi energy descriptor they developed. This descriptor gives an idea of the xenon adsorption isosteric heat of the material. Given these results, more studies should focus on describing the adsorption thermodynamic quantities such as the adsorption enthalpy but also the Henry adsorption constants. This study finally leads to the synthesis and testing of one of the top performing materials in the field. However, we cannot stop but wondering why there is a discrepancy between the theoretical selectivity of around 70 of SBMOF-1 and its actual experimental selectivity of 16. In the final chapter of this thesis, we will try to give an explanation for this. In the future, such close collaboration between experimental and computational teams are crucial even if they are still too rare. A recent paper suggests that these collaborations are rare across all nanoporous fields and a lot of improvements are needed to foster cooperation between the labs.[177]

1.3.4 The future of screening

Despite the progress made, important drawbacks of the current methodologies remain. High-throughput screenings rely too much on oversimplified assumptions such as the rigidity of the framework, the absence of defects, the use of Lennard-Jones potentials and inaccurate charges. For instance, the rigidity of the framework only takes into account one conformation of the framework. Yet, thermal agitation induces a "breathing" movement of the framework with an amplitude dependent on its intrinsic flexibility. The pores of the framework can change depending on the number of adsorbates to interact more optimally with them, which can be induced by a change in pressure. The issue of flexibility is rarely tackled, and when considered, it is only on the few most selective structures given by an inaccurate screening based on the rigid crystal approximation. One can wonder about the results obtained if it is applied to larger sets of structures. Witman et al. found that flexibility applied to top performing materials can decrease the selectivity, because the pore does not have an optimal size anymore.[159] In some cases, the selectivity of a well-performing material can even increase to become a top

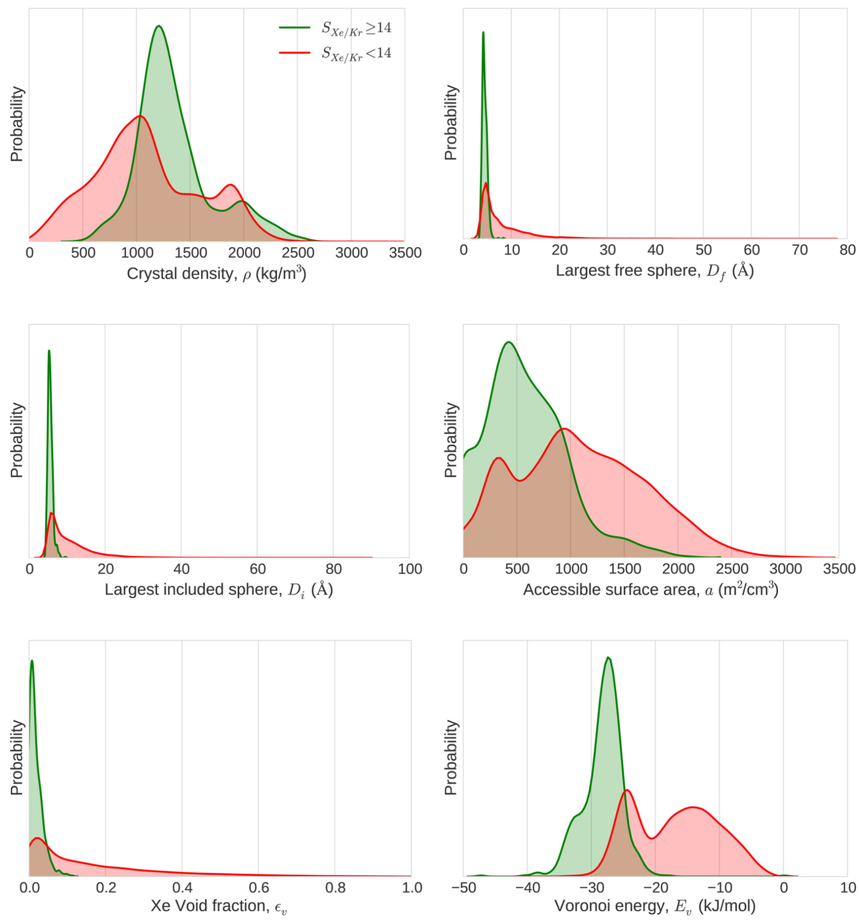


Figure 1.10: Statistical analysis of the adsorption separation of xenon/krypton mixtures by nanoporous materials. The graphs represent the distributions of structural descriptors explored by highly selective (green) and poorly selective (red) materials separately. Reprinted with permission from Ref. 14. Copyright 2015 American Chemical Society.

performing one. Computational screenings can be closer to predict experimental values of selectivity, diffusivity, and other key performance metrics.

Many open problems remain for the design of efficient high-throughput computational screenings. The connection between different properties for a given application is not systematically integrated in the screening procedures. For example, in methane storage, the working capacity of the material is the main property to optimize, but the kinetics of the adsorption/desorption or the mechanical resistance to compaction among others also need to be considered. Designing a nanoporous material is in fact a multivariate optimization problem with tacit constraints, for example the synthesizability. For instance, the diffusion coefficients of adsorbates in a xenon/krypton separation problem can help us better understand the breakthrough simulations and eventually the whole separation process in pressure or temperature swing adsorption beds. For this reason, studying transport properties along with uptake capacities and thermodynamic selectivity of the xenon/krypton separation can give a more complete picture of the industrial process we ultimately want to model.

Moreover, the transferability of the methodology to a broad range of materials is often achieved at the expense of accuracy in specific cases. And one can rightly question the universality of depending on faster but less elaborated models, which boils down to a trade-off problem between prediction accuracy and computational cost (or complexity). For instance, classical forcefields are broadly used in rigid materials for adsorption properties, but the switch to more costly *ab initio* methods or the addition of flexibility can result in a more accurate description at the expense of computational resources. The addition of polarization could be very promising since several top performing materials harbor open metal sites and highly polar sites that explain the acute affinity to xenon adsorbates.

The development of ML-assisted screenings is paired with the advances in data science techniques and algorithms. Recent advances in deep learning have enabled the development of transformer-based (the technology at the foundation of ChatGPT) machine learning models to predict adsorption properties.[178, 179] More importantly, the construction of descriptors tailored to the many possible applications is also an ongoing work. This construction work cannot be dissociated to the physical and chemical intuition of the scientists. Topological, chemical, electronic and other descriptors have been developed on top of the more common geometrical and thermodynamic descriptors, which displays the importance of strong physical chemistry knowledge. Recently Shi et al. highlighted the key role of energy histograms in predicting adsorption properties.[17] The discovery of novel relevant descriptors remains the main lever for increased performance of the ML models and is closely related to a rigorous theoretical work. For these reasons, we worked during this thesis on more accurate and faster ways of calculating these interaction energies to extract valuable energy/thermodynamic descriptors.

The development of databases is another key aspect in the promotion of data science in the field of materials science in general, and nanoporous materials chemistry in particular. The diversity of materials, the inclusion of experimental data (successful or failed), the addition of understudied classes of materials (e.g., amorphous) are all key aspects to upgrade the existing database. Even if existing attempts to create a centralized database have been initiated by the materials project,[113] this database does not contain all the existing information on each material. Furthermore, this high amount of data will need to be efficiently explored, and non-supervise deep learning algorithms have been developed to do so.[180] Coupled with synthesis robot, these methods can navigate through the unexplored databases to find the few most interesting candidates for a given targeted application.

In the future, computational high-throughput screening could be integrated more tightly into the design process of nanoporous materials, hence further improving its efficiency. The computational prescreening can be coupled with automated screenings of the most promising materials to finally identify candidates for further studies. This automated design process is described by Lyu et al. in their paper on "Digital Reticular Chemistry" and set out promising perspectives for computational screenings in the field.[7] Some studies are already pioneering this new research area by combining high-throughput characterizations, active learning algorithms and robotic synthesis.[181, 182] Another step towards faster industrialization would integrate process modeling to enrich the purely atomistic approach.

2

THERMODYNAMIC EXPLORATION OF XENON/KRYPTON SEPARATION

2.1 Characterization of Adsorption Equilibrium Properties

2.1.1 Geometrical descriptors

Before going into the details of the adsorption properties themselves, we will first introduce the different simulation techniques used to characterize the internal pore structure of a material key in interpreting the adsorption properties obtained using more complex molecular simulations. All the geometrical descriptors used in this thesis have been calculated using the Zeo++ software;[42] other tools exist,[43, 183] but the use of Voronoi decomposition of the volume speeds up the calculation making it the preferred tool for this task (efficiency gain mainly on volume calculation).[15]

Pore size

There are a multitude of pore sizes depending on the point where we measure it, all these pore sizes compose what we call a pore size distribution. Some specific values are, however, uniquely defined and used to put a single value to characterize pore sized. The diameter of the largest sphere that can diffuse freely in the structure is called D_f. The diameter D_{if} corresponds to the diameter of the largest included sphere along a free diffusion path; the diameter of the largest included sphere (not necessarily in a free diffusion path) is denoted D_i. The Figure 2.1 illustrates the difference between these pore sizes. In thermodynamic studies we will often use the term "largest cavity diameter" (LCD) instead of the largest included sphere D_i. And, the term "pore limiting diameter" will be used instead of D_i especially when studying the transport effects with the nanopores.

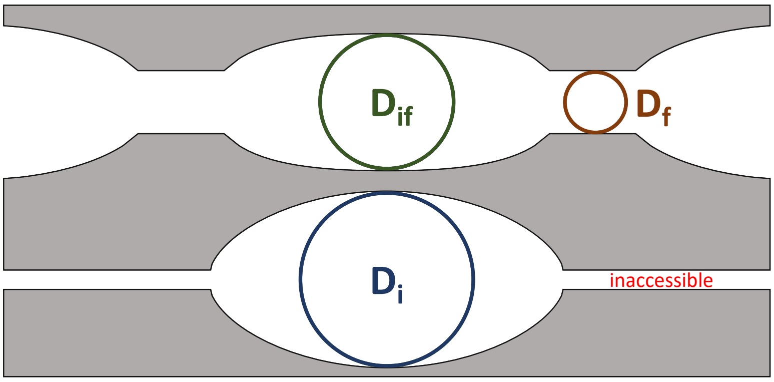


Figure 2.1: Illustration of the different pore sizes D_f, D_i and D_if. Note that in some materials D_if is equal to D_i, when the largest included sphere is also accessible through a free diffusion path.

To define these pore sizes, we need to first set the radii of the framework atoms that will shape the surrounding pores. These radii can be defined using different methods, the default mode uses the Cambridge Crystallographic Data Centre's (CCDC) radii. This method is most commonly used in the literature. We also introduced another set of radii based on the universal forcefield [184] we use for molecular simulations. The determination of these radii are inspired by an approach developed by Hung et al.[44] The atomic radii are defined as the distance where the LJ potential reaches $3k_B T/2$, for $T = 298\,K$. This type of definition can be more easily compared to the molecular simulations. For instance, we will use indexes to tell apart both methods. For example, LCD_{CCDC} corresponds to the standard definition of the LCD that uses the CCDC radii to run the Zeo++ software, while the LCD_{UFF} is associated with the definition of the atomic radii dependent on the UFF forcefield.

Surface area

The surface areas are calculated using a random sampling over the surface of the different atom surfaces. The algorithm counts only the points that do not overlap with another atom. For each atom we can therefore calculate an adsorbable surface. By summing up all the surfaces, we finally have the surface area. This "rolling ball" algorithm has been developed since 1973 by Shrake and Rupley.[185] the By defining a probe, the Voronoi tessellation defines the accessible and the non-accessible areas of the structure. Depending on where the surfaces are, they are either counted in the accessible or the non-accessible surface areas. In this chapter, we will use the accessible surface area defined by a probe of 1.2 Å; this is a computational equivalent of the experimental N_2 BET surface area.

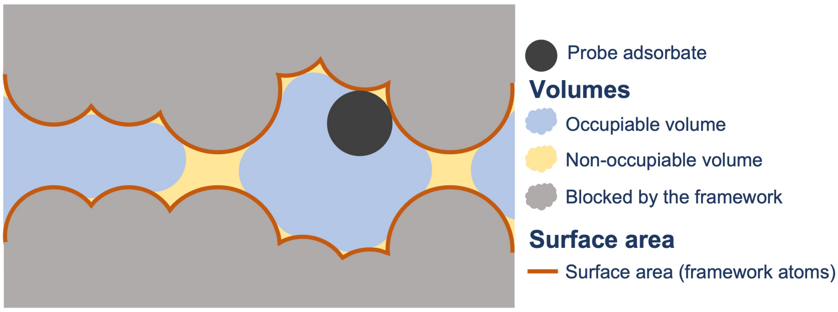


Figure 2.2: Illustration of the pore surface area and volume in a nanoporous material. As illustrated, there are different definitions of the pore volume: we can either consider the whole volume of the pores or the volume occupiable by a given probe. The surface area also changes depending on the definition. Studies have shown that occupiable volume has a better accordance with experimental data.[186]

Pore volume and porosity

The pore volume is calculated by random sampling of the accessible and inaccessible Voronoi cells. Similar algorithm, random sampling over a regular mesh. If the probe does not overlap with a framework atom, then it is counted in N. The ratio of this final N and the total number of points sampled gives the fraction of the pore volume. This ratio is called the void fraction or the porosity. Using the Voronoi decomposition, we can also define the accessible and non-accessible Voronoi cells to reduce the space the Monte Carlo simulation need to sample for the surface area and the void fraction calculations.

2.1.2 Intermolecular interaction energies

In most of the studies in this thesis, we will consider rigid structures interacting with guest adsorbates; the intramolecular interactions will not play any role in the simulations since the ionic, chemical or metallic bonds between the atoms of a same molecule are predefined at a given set of distances and remain the same. As discussed in the final chapter, this approximation can generate discrepancies between the theoretical model and the experiments. However, since the goal is to achieve screening approaches like the ones introduced in the first chapter, adding flexibility in the intramolecular interactions would reduce considerably the size of the database that could be screened. For these reasons, the term "interaction energy" will designate the guest–host and guest–guest intermolecular interactions mainly — host–host interactions would compromise the assumption on the rigidity of the framework.

In classical theory of molecular physics, the intermolecular interactions can be categorized in three different types according to their strength: (i) the ion–dipole and ion–induced dipole forces (40–600 $\mathrm{kJ\,mol^{-1}}$), (ii) the hydrogen bonding (10–50 $\mathrm{kJ\,mol^{-1}}$), and (iii) the Van der Waals forces (1–10 $\mathrm{kJ\,mol^{-1}}$). Note that these energy values are only indicative and the interaction depends on the nature of the molecules, but it allows us to rank the different foreces according to their strength; and to complete the molecular interactions picture, we can add that the ionic and covalent bonding will always be stronger than any intermolecular interactions (over 100 $\mathrm{kJ\,mol^{-1}}$). The generic term "Van der Waals interactions" actually regroups three different concepts usually called Keesom, Debye and London interactions. The Keesom interaction focuses on the electrostatic interaction between permanent multipoles (representing the electronic density around the molecules),[187] while the Debye induction force corresponds to the interaction between a multipole of a molecule and an induced multipole of another one,[188] and the London dispersion interaction occurs between instantaneous multipoles created by natural fluctuations in the electron density around polarizable atoms.[189, 190] To quantify these interactions, we can consider dipole interactions since they are the most influential in the multipole expansion of the electron density. The Keesom interaction potential V_K can therefore be reduced to the dipole–dipole interaction, which depends on the inverse third power of the distance for fixed dipoles; but in fluid phases we are interested in, the average over all the angles is rather given by the inverse sixth power as described in the equation 2.1 below:

$$V_K = -\frac{\mu_1^2\mu_2^2}{(4\pi\epsilon_0\epsilon_r)^2 r^6} \times \frac{1}{1.5 k_B T} \tag{2.1}$$

where μ_1 and μ_2 are the dipole moments of the molecules 1 and 2, ϵ_0 the vacuum dielectric permittivity and ϵ_r relative permittivity of the surrounding material, k_B the Boltzmann constant, T the temperature and r the intermolecular distance. The Debye interaction potential V_D being reduced to the permanent dipole–induced dipole interactions can now be expressed using the electric polarizability α_1 and α_2 of the molecule 2 as shown in equation 2.2.

$$V_D = -\frac{\mu_1^2\alpha_2 + \mu_2^2\alpha_1}{(4\pi\epsilon_0\epsilon_r)^2 r^6} \times \frac{1}{k_B T} \tag{2.2}$$

Finally, the London dispersion interaction potential V_L is now the fluctuating dipole–induced dipole interaction and can be expressed as follows:

$$V_L = -\frac{\alpha_1\alpha_2}{(4\pi\epsilon_0\epsilon_r)^2 r^6} \times 1.5\frac{I_1 I_2}{I_1 + I_2} \tag{2.3}$$

where I_1 and I_2 are the first ionization energies. We can note that the Van der Waals potentials are all negative (attractive interaction) and depend on the inverse sixth power of the distance — considering only the dipole moments. Before moving to the computational modelization of these long-distance intermolecular forces, we need to specify the repulsive force that occurs at very short distances; this force can be explained by the impossibility for electrons of both atoms to occupy the same quantum space as stated by Pauli in his exclusion principle.

For the system we are studying in this thesis, the adsorption of noble gases in nanoporous materials, the guest–guest and guest–host interactions can be described by the induction and dispersion interactions only. We will use a simplistic model, the Lennard-Jones (LJ) potential V_{LJ},[191] that relies on a repulsive term for the Pauli exclusion principle and an attractive term to model the attractive Van der Waals component of the interaction, as shown below:

$$V_{LJ} = 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right) \tag{2.4}$$

where ϵ is the depth of the well (minimal energy) and σ is the distance at which the potential is zero. The forcefield defines the LJ parameters ϵ and σ for either all atom pairs or only for the same type of atoms. For the commonly used universal forcefield (UFF),[184] only the parameters for atoms of the same nature are defined and the parameters of the pair atoms can be induced using combining rules. In this thesis, we will use the UFF forcefield (because it performs well compared with *ab initio* forcefield [192]) and the Lorentz-Berthelot mixing rules to combine the LJ parameters — it makes an arithmetic average of the σ values (Lorentz rule) and a geometric one of the ϵ values (Berthelot rule). Finally, to reduce the computation time, we usually set a cutoff distance at which the LJ potential can be considered negligible. At this cutoff distance, we can for example apply a shift so that the energy equals zero at the cutoffs (discontinuity of energy), just truncate (discontinuity of the force), or use a tail switching function to make the tail converge smoothly to zero near the cutoff. To make it simple, in most of the simulations in the screenings, a shifting strategy combined with a cutoff of 12 Å have been used.

Usually in adsorption simulations of other gas molecules containing partial charges, we need to calculate the Coulomb interaction between the partial charges of the host framework and those of the adsorbate — in periodic systems, the Ewald summation can be used. For noble gases, these ion–dipole and dipole–dipole interactions do not exist due to the perfect neutrality of the molecules. However, we can argue that the ion–induced dipole is not taken into account in a simplified LJ potential. To complete the whole picture of the intermolecular interactions, we would need to study the energy inducted by the surrounding framework atoms' charges on the adsorbate. Several approaches have been developed in the literature to improve the description of the intermolecular interactions by coupling the LJ potentials with an induction potential.[193, 194]

To wrap up this small section on the modelization of the intermolecular interactions in the adsorption simulations, I want to emphasize once more on the main modelization assumptions that could alter the accuracy of our method. First, the framework remains rigid during the whole simulation, which avoids the need for a molecular dynamics simulation of the framework to save time but also hides the effects of a known phenomenon.[159] Second, the polarizability of the adsorbate is not fully taken into account since the interaction with the charges of the framework are not considered; the difference in polarizability between xenon and krypton can be further exploited to enhence the selectivity even more, experimental studies suggest the key role of polar groups and open metal sites.[5, 6, 195] And finally, the complex induction and dispersion interactions are described with a two-parameter model, which cannot capture all the subtlety of the differences between a same atom in different environments for instance; it could be possible to fine-tune these parameters in very specific cases, but the overall good performance [192] of the UFF forcefield needed in a screening process can induce small errors when looking at specific cases. These assumptions have been made to find the right trade-off of the speed of computation versus a detailed description of the physical phenomena at stake.

2.1.3 Mixture adsorption: Grand Canonical Monte Carlo

As explained before, we can think of adsorption as a gas–solid or liquid–solid interfacial phenomenon, the adsorbate phase fills the accessible pore volumes depending on the physical conditions the material is put under. No simple model can predict how the adsorbates would interact with the pore surface, how many of them can fit in, what configuration is the most stable, etc. To answer these questions, we can evaluate all possible adsorption configurations that would undeniably have different numbers of adsorbate, and only keep the most thermodynamically plausible ones. To do so, these configurations will have to follow a given probability distribution, the grand canonical ensemble probability for instance, because it allows the variation of the number of molecules (adsorbate molecules) and the total energy. With the help of a Monte Carlo simulation, we can vary the energy and the loading inside the pores so that the distribution of configurations c follows this probability law:

$$\mathrm{P}_c = \frac{1}{\Xi} e^{-\beta(\mathrm{E}_c - \mu \mathrm{N}_c)} \tag{2.5}$$

where E_c and N_c are respectively the energy and the number of adsorbate particles in the configuration c. Normally the energy and the number of molecules of all particles should be considered, but for now, since the whole system is considered rigid we will only focus on the adsorbate molecules. The chemical potential μ and the temperature T inside β correspond to the ones of the gas phase in equilibrium with the adsorbent material. And the pressure and volume V are considered fixed under the rigidity assumption. The grand canonical partition function $\Xi(\mu, \mathrm{V}, \mathrm{T})$ will then be the following sum over all possible configurations:

$$\Xi(\mu, \mathrm{V}, \mathrm{T}) = \sum_c e^{-\beta(\mathrm{E}_c - \mu \mathrm{N}_c)} \tag{2.6}$$

This multiplicative constant does not need to be known in the Monte Carlo simulation we will describe now.

Beyond these theoretical considerations, the grand canonical Monte Carlo simulation, referring to a Metroplis-Hastings Monte Carlo algorithm in the context of the grand canonical thermodynamic ensemble, will need several key characteristics in order to fulfill the previous claims on the probability distribution of the configurations. Monte Carlo (MC) refers to the randomness inherent to the gambling games of the eponymous casino on the azure coast of Monaco. The MC simulations are therefore relying on randomly generating atomic configurations; however in order to do it efficiently, we need to stay as much as possible in the physically possible atomic space, while exhaustively exploring all the chemical configurations. In molecular simulations, to do so, only the initial configuration c_0 is really randomly generated, but then the algorithm has different rational moves to change the configuration with a controlled amount of randomness. The second key feature (acceptance or rejection condition) was introduced by Metropolis and co-workers that allows to reproduce any distribution with an unknown multiplicative prefactor.[196] The configuration c_1 resulting of the random move is evaluated by calculating the transition probability (like in a Markov chain) or acceptance rate $acc(c_0 \rightarrow c_1)$:

$$acc(c_0 \rightarrow c_1) = \min\left(1, e^{-\beta\left(\mathrm{E}_{c_1} - \mathrm{E}_{c_0} - \mu\left(\mathrm{N}_{c_1} - \mathrm{N}_{c_0}\right)\right)}\right) \tag{2.7}$$

Any move that has a greater probability of occurring is always accepted, if the probability is lower than the acceptance rate depends on the probability ratio. The multiplicative prefactor has no influence on the algorithm, we do not need to know the chemical space to explore beforehand, which is a valuable simplification. This sequence of a Markov-type chain can then be used to approximate the probability distribution of the grand canonical ensemble we seek to describe in the equation 2.5.

To complete the description of the grand canonical Monte Carlo (GCMC) simulation we are interested in, let us now consider the different MC moves used to generate a configuration

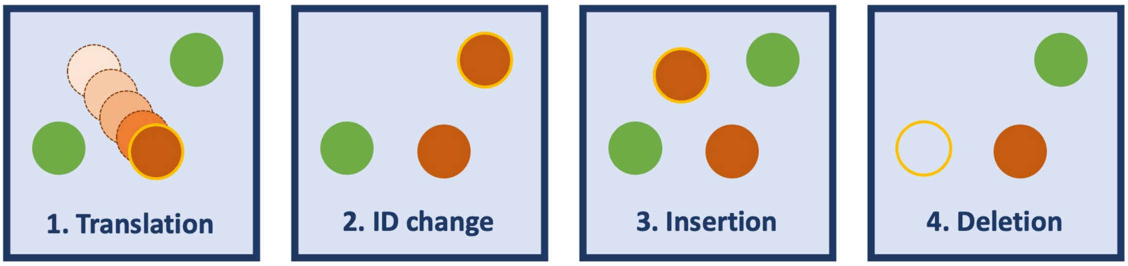


Figure 2.3: MC moves in a system of two types of monoatomic atoms (green and orange). The modification on the first box is highlighted by the yellow circle and the dragging pattern is represented by a set of dashed circles. The boxes 2 to 4 represent the moves going from the initial state represented in box 1, the corresponding move is highlighted by a yellow outer circle.

from another. Depending on the parameterization, these moves have different probabilities of occurring. For monoatomic atoms only four moves are relevant (see Figure 2.3): (i) the translation of a randomly chosen molecule with a displacement randomly chosen within a given radius, (ii) the change of identity of a randomly chosen molecule into another one, (iii) the insertion of an adsorbate molecule and (iv) the deletion of an adsorbate molecule. We deliberately omitted the rotations of the adsorbate because of the spherical symmetry of noble gases and the change of volume since the flexibility of the material framework is neglected.

By using a GCMC algorithm, we can now generate a set of configurations according to their probability of occurrence. Because the probability law is directly taken from equation 2.5, the series of configurations describe the thermodynamic equilibrium state of a nanoporous material in contact with a reservoir of a xenon-krypton mixture at a given composition, pressure and temperature. Different thermodynamic quantities can be derived from ensemble averaging: the averaging loading or uptake at a given pressure (several pressures give the isotherm) and the isosteric heat of adsorption of each adsorbate (Xe and Kr). The ratio of the uptakes q informs on the selectivity s of the thermodynamic separation process:

$$\boxed{s = \frac{q^{\mathrm{Xe}}}{q^{\mathrm{Kr}}} \times \frac{y^{\mathrm{Kr}}}{y^{\mathrm{Xe}}}} \tag{2.8}$$

where y^{Xe} and y^{Kr} designate respectively the mole fractions of Xe and Kr in the gas phase reservoir.

To characterize a separation process, we theoretically only need to perform a GCMC calculation at every pressure conditions that we are interested in. However, this type of simulation requires a lot of time to converge since we need to test out a lot of insertion/deletion moves to accurately estimate the number of adsorbed molecules and the composition of the mixture. Hence, faster methods (machine learning) are developed to estimate the selectivity at any pressure conditions.[14, 178] If we are interested in the infinite dilution case, faster methods are already available, we are now going to introduce the Widom insertion that can estimate the adsorption performances at infinite dilution by estimating the Henry adsorption constant.

2.1.4 Infinite dilution adsorption: Widom insertion

In 1963, the professor B. Widom introduced a simple method of calculation of thermodynamic properties in a material or fluid mixture.[197] Generally, this method allows accessing to the difference of internal energy before and after the insertion of a Widom test particle while fixing all other particles, therefore comparing the N-particle and (N+1)-particle states. This difference of energy $\Delta\Phi$ can then be used to deduce the excess free energy associated to it $\Delta F_{exc} = -k_B T \ln\left(\langle \exp(-\beta\Delta\Phi)\rangle\right)$, which corresponds to the excess chemical potential induced by the addition of a particle. In the domain of fluid phase equilibrium, Widom insertion is the most straightforward method to calculate a chemical potential value; however it has drawbacks in liquid-like phases because the insertable space is very narrow, and no relaxation is implemented to account for the reorganization of the surrounding particles.[198] In our case, we will be only interested in the insertion from 0 to 1 particle, where no problems of overlapping between adsorbate particles can happen. Widom insertion is in our case only a random insertion of an adsorbate into an empty nanoporous framework. By randomly sampling the void space, we obtain a distribution of interaction energies $\mathcal{E}_{int}$, the average of the Boltzmann weights associated is directly proportional to the adsorption free energy ΔG_{ads} and the Henry adsorption constant K_H. By taking the Boltzmann average of the interaction energies, we can also compute the adsorption enthalpy ΔH_{ads}. All these quantities stay only valid at infinite dilution, for higher quantities of adsorbates the previously described GCMC technique should be used.

In the infinite dilution case, this test particle insertion technique is similar to a random sampling of the adsorbable space inside a material. If the sampling is thorough enough, we can derive the following definitions of ΔG_{ads} (equation eq:gibbs), K_H (equation eq:henry) and ΔH_{ads} based on a complete sampling of the interaction energies $\mathcal{E}_{int}$.

The adsorption Gibbs free energy ΔG_{ads} is equal to the excess free energy previously calculated in a Widom insertion since the structure is rigid and PV does not fluctuate (G = F + PV).

$$\boxed{\Delta G_{ads} = -RT \ln\left(\langle \exp(-\mathcal{E}_{int}/RT)\rangle\right)} \tag{2.9}$$

To derive the Henry constant K_H, we need to consider an ideal gas. The number of adsorbed molecules n_{ads} can be expressed using the bulk density of the gas $\rho_{ads,bulk}$ and the volume of the pores V_{pore}:

$$n_{ads} = \rho_{ads,bulk} \times V_{pore} \tag{2.10}$$

The pore volume can be seen as the continuous sum of each voxel times the Boltzmann probability of presence, which is represented by the following integral of the Boltzmann factors. This integral can then be changed to the average of the Boltzmann factors:

$$V_{pore} = \int_V \exp\left(-\mathcal{E}_{int}(\mathbf{r})/RT\right) d\mathbf{r} = V\langle \exp\left(-\mathcal{E}_{int}/RT\right)\rangle \tag{2.11}$$

Let us apply the equation 2.11 and the perfect gas equation of state $P = \rho_{ads,bulk}RT$ on the bulk gas in equilibrium, we can change the equation 2.10 to:

$$\frac{n_{\mathrm{ads}}}{\mathrm{V}} = \frac{\mathrm{P}}{\mathrm{RT}} \langle \exp\left(-\mathcal{E}_{\mathrm{int}}/\mathrm{RT}\right) \rangle \tag{2.12}$$

If we now consider the gravimetric loading $\mathrm{L_{ads}}$ (in $\mathrm{mmol\,g^{-1}}$), we need to divide the equation by mass density of the framework ρ_f:

$$\mathrm{L_{ads}} = \frac{n_{\mathrm{ads}}}{\mathrm{V}\rho_f} = \frac{\langle \exp\left(-\mathcal{E}_{\mathrm{int}}/\mathrm{RT}\right) \rangle}{\rho_f \mathrm{RT}} \mathrm{P} \tag{2.13}$$

Since the Henry's law is described by $\mathrm{L_{ads}} = \mathrm{K_H} \times \mathrm{P}$, we have the final relation between the Henry adsorption constant and interaction energy distribution.

$$\boxed{\mathrm{K_H} = \frac{\langle \exp\left(-\mathcal{E}_{\mathrm{int}}/\mathrm{RT}\right) \rangle}{\rho_f \mathrm{RT}} = \frac{1}{\rho_f \mathrm{RT}} \exp\left(-\frac{\Delta \mathrm{G_{ads}}}{\mathrm{RT}}\right)} \tag{2.14}$$

Note that the ρ_f factor comes from the use of a gravimetric loading expressed in $\mathrm{mmol\,g^{-1}}$ and is not always present in the different derivations of the literature.[43] The RT factor comes from the perfect gas assumption we made, which is a good approximation in the noble gas case.

Finally, if we consider an adsorption equilibrium (e.g., $\mathrm{Xe_{(g)}} \rightleftharpoons \mathrm{Xe_{(ads)}}$), we can define an equilibrium constant $\mathrm{K_{ads}} = x_{\mathrm{ads}}/y_{\mathrm{gas}}$ where x_{ads} is the mole fraction in the adsorbed phase and y_{gas} the mole fraction in the gas phase for a given compound (e.g., Xe). For a pure gas ($y_{\mathrm{gas}} = 1$) at infinite dilution, we can apply the Henry's law to derive the following relation:

$$\mathrm{K_{ads}} = \frac{n_{\mathrm{ads}}}{n_{\mathrm{site}} y_{\mathrm{gas}}} = \frac{\mathrm{K_H P} \rho_f \mathrm{V}}{n_{\mathrm{site}}} = \mathrm{PV} \frac{\langle \exp\left(-\mathcal{E}_{\mathrm{int}}/\mathrm{RT}\right) \rangle}{n_{\mathrm{site}} \mathrm{RT}} \tag{2.15}$$

where n_{site} is the number of sites considered constant since it is much higher than n_{ads} at infinite dilution.

Now by applying the Van't Hoff equation to this infinite-dilution adsorption equilibrium constant $\mathrm{K_{ads}}$, we can derive an expression of the adsorption enthalpy at infinite dilution:

$$\Delta \mathrm{H_{ads}} = -\mathrm{R} \frac{\mathrm{d} \ln\left(\mathrm{K_{ads}(T)}\right)}{\mathrm{d}(1/\mathrm{T})} \tag{2.16}$$

Then by decomposing the logarithm on the fraction of equation 2.15,

$$\Delta \mathrm{H_{ads}} = -\frac{\mathrm{d} \ln\left(\mathrm{PV}/n_{\mathrm{site}}\right)}{\mathrm{d}(1/\mathrm{T})} - \mathrm{R} \frac{\mathrm{d} \ln\left(\langle \exp\left(-\mathcal{E}_{\mathrm{int}}/\mathrm{RT}\right) \rangle\right)}{\mathrm{d}(1/\mathrm{T})} - \mathrm{R} \frac{\mathrm{d} \ln\left(1/\mathrm{T}\right)}{\mathrm{d}(1/\mathrm{T})} \tag{2.17}$$

Then, $\mathrm{PV}/n_{\mathrm{site}}$ being a constant, we can reduce the expression to two derivatives, the first one being the logarithmic derivative of itself (1/T) and the second being the logarithmic derivative of the sum of the exponential terms.

$$\Delta \mathrm{H_{ads}} = 0 - \mathrm{R} \frac{\mathrm{d} \ln \left(\langle \exp\left(-\mathcal{E}_{\mathrm{int}}/\mathrm{RT}\right) \rangle \right)}{\mathrm{d}(1/\mathrm{T})} - \mathrm{RT} \tag{2.18}$$

Knowing that for any function f the logarithmic derivative equals the quotient of its derivative f', $\frac{\mathrm{d}\ln(f)}{\mathrm{d}x} = f'/f$, we can calculate the derivative of the average of the Boltzmann factors $\langle \exp\left(-\mathcal{E}_{\mathrm{int}}/\mathrm{RT}\right) \rangle$:

$$\Delta \mathrm{H_{ads}} = -\mathrm{R} \frac{1}{\frac{1}{\mathrm{N}} \sum e^{-\frac{\mathcal{E}_{\mathrm{int}}}{\mathrm{RT}}}} \frac{1}{\mathrm{N}} \sum \frac{\mathrm{d} \exp\left(-\mathcal{E}_{\mathrm{int}}/\mathrm{RT}\right)}{\mathrm{d}(1/\mathrm{T})} - \mathrm{RT} \tag{2.19}$$

where N corresponds to the number of points where the Widom particle has been inserted. The exponential derivative makes the energy factors come out, and we get the following expression:

$$\Delta \mathrm{H_{ads}} = -\mathrm{R} \frac{1}{\sum e^{-\frac{\mathcal{E}_{\mathrm{int}}}{\mathrm{RT}}}} \sum -\frac{\mathcal{E}_{\mathrm{int}}}{\mathrm{R}} e^{-\frac{\mathcal{E}_{\mathrm{int}}}{\mathrm{RT}}} - \mathrm{RT} \tag{2.20}$$

With some simplification, we can express the adsorption enthalpy $\Delta \mathrm{H_{ads}}$ as a Boltzmann average of the interaction energies minus a term RT that comes from the ideal gas assumption (perfect gas equation of state).

$$\boxed{\Delta \mathrm{H_{ads}} = \frac{\sum \mathcal{E}_{\mathrm{int}} e^{-\frac{\mathcal{E}_{\mathrm{int}}}{\mathrm{RT}}}}{\sum e^{-\frac{\mathcal{E}_{\mathrm{int}}}{\mathrm{RT}}}} - \mathrm{RT}} \tag{2.21}$$

From the values of the adsorption free energy and enthalpy, we can now deduce the adsorption entropy $\Delta \mathrm{S_{ads}}$ using the definition of the Gibbs free energy (G = H − TS):

$$\boxed{\Delta \mathrm{S_{ads}} = \frac{1}{\mathrm{T}} \left(\Delta \mathrm{H_{ads}} - \Delta \mathrm{G_{ads}} \right)} \tag{2.22}$$

We already defined the selectivity as the ratio of the proportion of Xe/Kr in the adsorption phase to the proportion in the gas phase in the equation 2.8. At infinite dilution, we can rewrite the selectivity using the Henry's law ($q^{\mathrm{i}} = \mathrm{V}\rho_f \mathrm{K_H^i} y^{\mathrm{i}} \mathrm{P}/n_{\mathrm{tot}}$) and simplifying the constant term $\mathrm{PV}\rho_f/n_{\mathrm{tot}}$:

$$\boxed{s = \frac{\mathrm{K_H^{Xe}} y^{\mathrm{Xe}}}{\mathrm{K_H^{Kr}} y^{\mathrm{Kr}}} \times \frac{y^{\mathrm{Kr}}}{y^{\mathrm{Xe}}} = \frac{\mathrm{K_H^{Xe}}}{\mathrm{K_H^{Kr}}}} \tag{2.23}$$

By extrapolating at the zero loading regime, the Xe/Kr selectivity can be simply expressed as the ratio of the Henry adsorption constant of xenon and krypton.

In this section, we saw that simple thermodynamic quantities such as the adsorption Gibbs free energy, enthalpy and entropy can be derived from the study of a simple adsorption equilibrium equation. In the next one, we will explore a thermodynamic characterization of the adsorption-based separation process using another equilibrium.

2.1.5 The thermodynamics behind adsorption-based separation

Now that the main simulation tools used to describe the competing adsorption of Xe/Kr binary mixtures are introduced, let us rationalize the separation process by modeling the process within a theoretical "exchange" equilibrium that corresponds to the exchange of gas phase Xe and Kr on a model adsorption site representing all the most attractive sites for a given pressure condition:

$$\mathrm{Xe}_{(g)} + \mathrm{Kr}_{(ads)} \rightleftharpoons \mathrm{Xe}_{(ads)} + \mathrm{Kr}_{(g)} \tag{2.24}$$

The equilibrium constant associated to the Equation (2.24) at any pressure for a given composition is simply the selectivity s, defined in the equation2.8, because the gas phase activities of $\mathrm{Xe}_{(g)}$ and $\mathrm{Kr}_{(g)}$ correspond the partial pressures y^{Xe} and y^{Kr}, and the adsorption phase activities of $\mathrm{Xe}_{(ads)}$ and $\mathrm{Kr}_{(ads)}$ correspond the mole fractions q^{Xe} and q^{Kr}. The Gibbs free energy at equilibrium can be directly defined using the equilibrium constant, by applying this relation to the exchange equilibrium we can define an exchange Gibbs free energy $\Delta_{\mathrm{exc}}\mathrm{G}$:

$$\boxed{\Delta_{\mathrm{exc}}\mathrm{G} = -\mathrm{RT}\ln(s)} \tag{2.25}$$

This exchange equilibrium can be seen as the substraction between the adsorption equilibria of xenon and krypton. So by applying the Hess's law of constant heat summation, we can derive an expression of the exchange enthalpy as the difference of the adsorption enthalpies between xenon and krypton within the mixture.

$$\boxed{\Delta_{\mathrm{exc}}\mathrm{H}^{\mathrm{Xe/Kr}} = \Delta_{\mathrm{ads}}\mathrm{H}^{\mathrm{Xe}} - \Delta_{\mathrm{ads}}\mathrm{H}^{\mathrm{Kr}}} \tag{2.26}$$

Moreover, the adsorption enthalpies $\Delta_{\mathrm{ads}}\mathrm{H}$ can be obtained in a GCMC calculation using a formula derived from the fluctuation theorem in statistical mechanics (see a derivation in this online article [199]):

$$\Delta_{\mathrm{ads}}\mathrm{H}^{\mathrm{Xe}} = \frac{\langle \mathrm{EN}\rangle - \langle \mathrm{E}\rangle\langle \mathrm{N}\rangle}{\langle \mathrm{N}^2\rangle - \langle \mathrm{N}\rangle^2} - \mathrm{RT} \tag{2.27}$$

where E corresponds to the energy of the adsorbates and N the total number of adsorbates at every step of the simulation. Note that this equation remains only valid for $\mathrm{N} \gg 1$, because the first step of the derivation is based on a first order Taylor expansion $\langle \mathrm{E}\rangle(\langle \mathrm{N}\rangle + 1) - \langle \mathrm{E}\rangle(\langle \mathrm{N}\rangle) = \frac{\partial\langle \mathrm{E}\rangle}{\partial\langle \mathrm{N}\rangle}$. This expression of the adsorption enthalpy echoes with the one at infinite dilution (equation 2.21), where for $\mathrm{N} \to 0$ we now have $\Delta\mathrm{H}_{\mathrm{ads}} = \langle \mathrm{E}\rangle(1) - \langle \mathrm{E}\rangle(0) - \mathrm{RT} = \langle \mathrm{E}\rangle(1) - \mathrm{RT}$.

Now that we defined the exchange Gibbs free energy and an exchange enthalpy at any pressure, we can now use the same approach as for the equation 2.22 to derive the exchange entropy:

$$\boxed{\Delta_{\mathrm{exc}}\mathrm{S} = \frac{1}{\mathrm{T}}(\Delta_{\mathrm{exc}}\mathrm{H} - \Delta_{\mathrm{exc}}\mathrm{G}) = \frac{1}{\mathrm{T}}\Delta_{\mathrm{exc}}\mathrm{H} + \mathrm{R}\ln(s)} \tag{2.28}$$

Before concluding this methodological section, we need to note that the thermodynamic quantities associated to this newly defined adsorption exchange equilibrium can be defined at different pressures and different methodologies can be used to calculate them. At infinite dilution, we would preferably use Widom insertions and the adsorption free energies and enthalpies to deduce these exchange quantities; at higher pressure, we would use the GCMC calculation to define a free energy (via the loading values) and the isosteric adsorption heat to define them. In the following study, we will focus on only two pressures: the ambient pressure (at 1 atm) and the limit of zero loading (infinite dilution). At 1 atm, the previously defined quantities will have an index 1 to differentiate them from the infinite dilution case where an index 0 will be used; for example, $\Delta_{ads}H_1^{Xe/Kr}$, $\Delta_{exc}G_1^{Xe/Kr}$ or $s_1^{Xe/Kr}$ at 1 atm, and $\Delta_{ads}H_0^{Xe/Kr}$, $\Delta_{exc}G_0^{Xe/Kr}$ or $s_0^{Xe/Kr}$ at the low-pressure limit. One final note on the simulation details, to run the GCMC calculations and the Widom insertion, we used the Raspa2 software developed by Dubbeldam et al.[200] And, the intermolecular Van der Waals interactions were described by a Lennard-Jones (LJ) potential with a cutoff distance of 12 Å. The LJ parameters of the framework atoms are given by the universal force field (UFF),[184] and the guest atoms (xenon and krypton) have their LJ parameters taken from a previous screening study.[157] All the MOFs described here are taken from the CoRE MOF 2019 database.[64]

2.2 Preliminary Analyses of the Separation Performance

As we have seen above in the existing literature, the computational screening of the nanoporous materials — both existing frameworks and hypothetical structures — for targeted adsorption properties has been the object of many studies, and several of those high-throughput screening studies have focused on noble gas separation, and Xe/Kr separation, in particular. For large-scale studies we have found that, in addition to the testing and validation of methodological developments, the screening aimed in most cases at one of three objectives: (i) to identify top performing materials for synthesis and/or characterization; (ii) to better understand the limits of possible performance, and the relationships and trade-offs between various metrics of performance (selectivity, uptake, etc); (iii) identify structure–property relationships, correlating separation performance with structural properties of the materials that can be more easily determined (i.e., at low computational cost). In this initial screening study of the thermodynamic quantities, we performed a screening of around 9,700 tridimensional structures of a preprocessed version of the CoRE MOF 2019-ASR (all solvent removed) database that are publicly available — only the non-disordered structures and the structures with a cell volume smaller than 20 nm^3 (to limit the overall calculation time) were considered. We will focus on the different relationships of the Xe/Kr selectivity has with structural descriptors based on geometrical analyses, and then with different thermodynamic descriptors (free energy, enthalpy, entropy).

2.2.1 Structure–selectivity relationships

An adsorption separation process is primarily characterized by a pivotal performance metric called the selectivity we defined in equations 2.8 and 2.23. By comparing this selectivity to geometrical descriptors calculated by the Zeo++ software,[42] we want to characterize the materials that will most likely be selective for a separation of a 20:80 Xe/Kr mixture (to compare with most literature screenings on a mixture extracted from the air). Three structural descriptors have been calculated: the accessible surface area of a N_2-sized probe of 1.2 Å, the void fraction occupiable by a 2.0 Å radius probe (roughly the size of a xenon),[186] and the diameter of the largest included sphere (D_i) using specially designed atom radii. Inspired by a recent work on the comparison of pore limiting diameters and self-diffusion coefficients,[44] we defined a list of Van der Waals radii to be read by the Zeo++ software (more details in `https://github.com/eren125/zeopp_radtable`). All Zeo++ calculations use an atomic radius that corresponds to the distance where the LJ potential reaches $3k_{\mathrm{B}}\mathrm{T}/2$, for T = 298 K.

Xenon uptake and selectivity

Before digging deeper in the structure–selectivity relationship, we will look at the relation between the xenon uptake (the number of adsorbed xenon in the GCMC simulation) and the selectivity at 1 atm. For instance, the xenon uptake could also be very important in a separation process, because it defines the working capacity of xenon produced by adsorption/desorption cycles. According to the Figures 2.5 and 2.8, first high xenon uptakes are associated with a high selectivity, but at some point very high selectivity is associated to lower uptakes. There seem to be an area where materials have selectivity over 100 and Xe uptake around 3 mmol g^{-1}, whereas an uptake over 6 mmol g^{-1} can only be obtained for a selectivity between 10 and 20. One cannot maximize both uptake and selectivity metrics at the same time, a trade-off needs to be made when designing nanoporous materials for xenon/krypton separation.[201] Different strategies have been implemented to optimize both metrics using mixed metrics such as the adsorbent performance score (APS).[202] This trade-off can be rationalized by using the different structural descriptors (pore size, surface area and volume) we presented earlier.

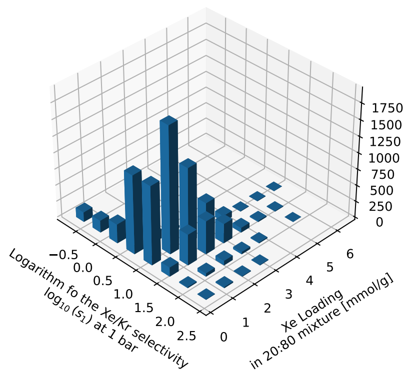


Figure 2.4: 3D histograms of in a bidimensional space formed by the Xe/Kr selectivity and the xenon uptake. The z-axis represents the number of structures with characteristics close to the one specified in x and y-axis. A base-10 logarithm has been applied to the selectivity values.

Furthermore, even if we know that it is possible to optimize either the xenon uptake or the Xe/Kr selectivity, these very successful materials are very rare inside a given diverse dataset. In the histogram presented in Figure 2.4, the number of very selective materials is very low, same for the high-capacity materials. The most frequent materials have a selectivity between 1 and 10 and an uptake below 3 mmol g^{-1}. These values can be considered the standard values of nanoporous material for Xe/Kr separation, which sets reference values to compare the various performance metrics and build a chemical intuition. A selectivity above 20 is therefore considered rather high (even if the top materials have a much higher selectivity[6]) and a xenon uptake above 4 mmol g^{-1} is also a pretty good adsorption capacity. The rarity of these top-performing materials gives the impression of searching a needle in a haystack, which has prompted some computational studies to design their algorithm to focus on finding the best materials rather than to describe equally all materials.[203, 204]

Surface area and selectivity

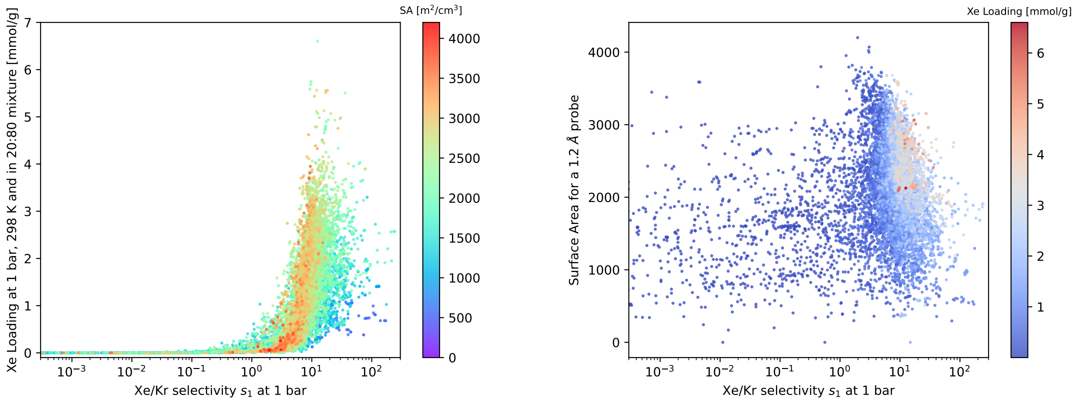


Figure 2.5: On the left: scatterplot of the xenon uptake as a function of the selectivity and labeled by the values of the surface area. On the right: scatterplot of the selectivity and the surface area labeled by the quantity of xenon adsorbed. The selectivity and uptake are calculated by a GCMC simulation of a 20:80 Xe/Kr mixture.

As demonstrated by other studies on methane storage applications by Wilmer et al.[10] and then later by Fernandez et al.[82], the methane uptake is maximal for a specific optimal range of surface area values (2500-3000 $m^2\ cm^3$). Higher values of surface area will not yield to higher values of methane uptake. This limitation also occurs for the selectivity as we can see in the right plot of the Figure 2.5. Materials with a selectivity around 5 will have any surface areas from 0 to 4000 $m^2\ cm^3$, whereas the ones with a selectivity above 40 will have a surface area below 2500 $m^2\ cm^3$. The optimal surface area for xenon uptake would on the other hand be between 2000 and 3000 $m^2\ cm^3$. The relationship between selectivity and surface area is quite complex, we cannot clearly state a precise range of surface areas that guarantees a high selectivity. This structural descriptor cannot characterize the selectivity, it needs to be coupled with other descriptors.

Looking at the 3D histogram on the Figure 2.6, we can see the breakdown of the surface area ranges for different categories of selectivity. For selectivity values higher than 92, the surface areas are most likely to be under 2000 $m^2\ cm^3$; between 92 and 35, there is a slightly wider range that goes to 2500 $m^2\ cm^3$; between 35 and 13, the interval goes even further to 3500 $m^2\ cm^3$ but is mostly centered between 1000 and 2500 $m^2\ cm^3$. This split view of the distributions gives a better grasp on what the best materials look like; however the surface area will never be a deterministic variable — we will never be available to deduce the selectivity by simply looking at the surface area, because a surface area between, for instance, 500 and 1000 $m^2\ cm^3$ have a relatively good chance of being selective but it concerns a lot of materials and it even has a higher chance of having a selectivity between 5 and 35 than a selectivity higher than that.

Void fraction and selectivity

A similar analysis of the relationship with the void fraction was also carried out by Wilmer et al. (Figure 5 of Ref. [10]) and an optimal value of the void fraction was found at around 0.8. As we can see on the plots of the Figure 2.8, the materials with the highest value of Xe uptake have void fraction values around 0.5, whereas the ones with the highest value of selectivity have much lower void fractions around 0.1. The optimal range of void fraction for uptake would

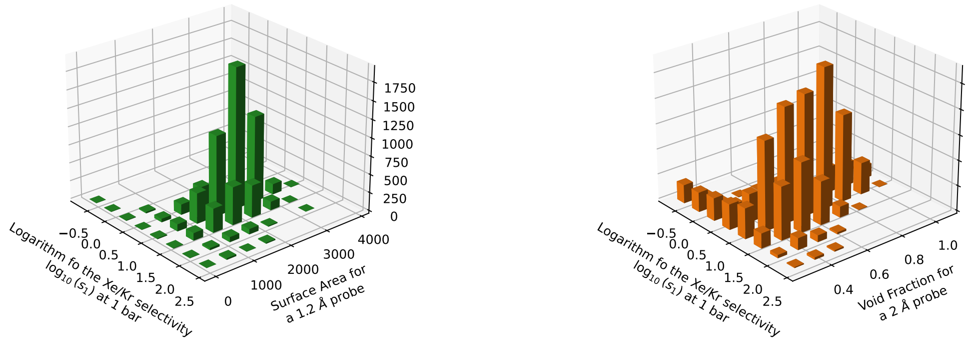


Figure 2.6: 3D histograms of in a bidimensional space formed by the Xe/Kr selectivity and the surface areas (on the left) and formed by the Xe/Kr selectivity and the pore void fractions (on the right). A base 10 logarithm has been applied to the selectivity values. Bin size increased by 2.4 (on log scale) for the selectivity, by about $500\ \mathrm{m}^2\ \mathrm{cm}^3$ *for the surface areas and by* 0.125 *for the void fraction.*

be between 0.2 and 0.6, whereas the one for selectivity is completely dissociated and is below 0.2. We can characterize a bit more finely the selectivity using the void fraction than using the surface area, even though they both give very similar results. Both descriptors describe a rather dense material with a "microporosity", in the sense of the IUPAC[35], which is characterized by medium-low pore volume and surface area.

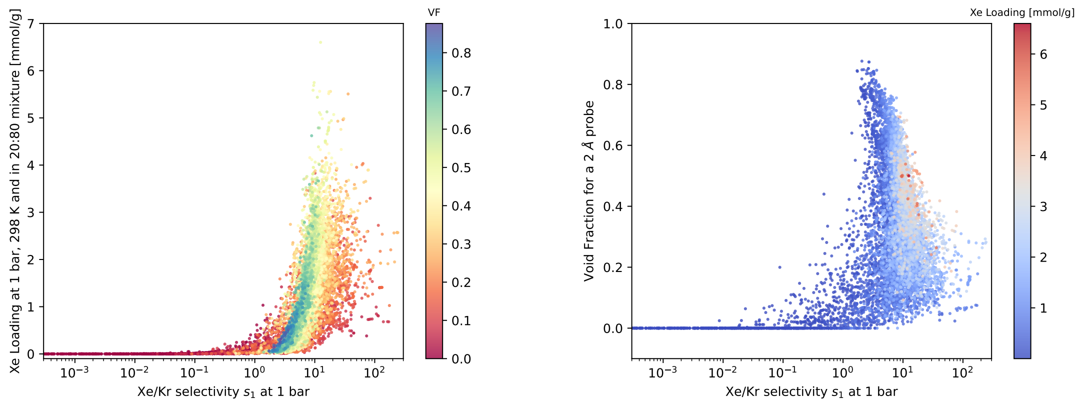


Figure 2.7: On the left: scatterplot of the xenon uptake as a function of the selectivity and labeled by the values of the void fraction. On the right: scatterplot of the selectivity and the void fraction labeled by the quantity of xenon adsorbed. The selectivity and uptake are calculated by a GCMC simulation of a 20:80 Xe/Kr mixture.

By carrying out a similar analysis than for the surface areas bu for the void fraction using the Figure 2.6 (right), we can also identify different intervals of void fractions that correspond to highly selective materials. For instance, selectivity values above 92 correspond to materials with a porosity between 0% and 37.5% (with a higher peak between 12.5% and 25%); selectivity values between 92 and 35 can be found in materials with a void fraction between 0% and 50.0% and much more frequently found for void fraction between 12.5% and 37.5%; selectivity values between 35 and 13 can be found in materials with a void fraction between 0% and 75.0% in a bell distribution centered around 31%. This center of distribution shifts toward higher values of the void fraction as lower selectivity values are considered, which suggests that a rather low porosity (below 25%) is preferable for selectivity performance. However, as we mentioned for surface areas, the void fraction is not a deterministic variable neither, we cannot predict the performance of the material solely based on this descriptor. Let us investigate whether adding another variable like the pore size as a joint variable wan better characterize the material's performance.

Pore Size and selectivity

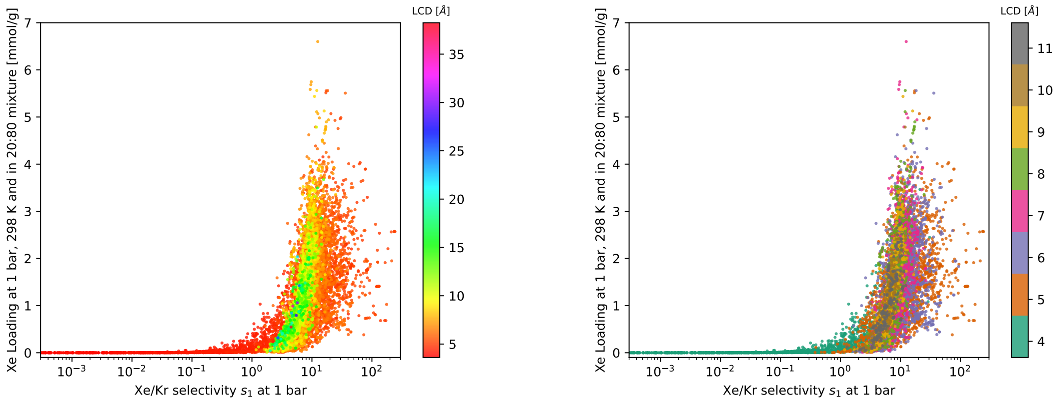


Figure 2.8: Scatterplot of the xenon uptake as a function of the selectivity (20:80) and labeled by the values of D_i *(left). The same scatterplot restricted to values of* D_i *between* (3.6 *and* 11.6 Å) *and labeled using a different color code to distinguish the most selective materials from the least selective ones. The most selective materials are colored in orange corresponding to a pore size adapted for xenon adsorption (around* 5 Å*). The least selective ones are in green, with a pore lower than the size of a xenon hence preventing its adsorption.*

If we now look at the joint effects of the void fraction and the largest cavity diameter (D_i) on the selectivity, we can note that the most selective materials are located in a very particular domain of this bidimensional descriptor space. On Figure 2.9, the structures with a selectivity over 10 are very likely to have a void fraction under 0.4 with a rather wide range of D_i. However, as we can see on the filtered version of the plot (on the right), the most selective materials (over 40) exist, on the other hand, in a very narrow range of D_i values between 4.8 and 6 Å approximately. This can be explained by the size of a xenon atom being very close to these D_i values, which allows a maximal stabilization of the xenon that we want to separate from krypton. The krypton being slightly smaller, the interaction with the pores are less favorable, hence explaining the higher selectivity we observe.

As presented in the previous chapter, Simon et al. found that the most selective materials have a pseudo-spherical shape with a size close to the diameter of a xenon which is rather dense and not very porous. By taking a slightly different approach and including the xenon uptake as a co-metric alongside the selectivity, we found very similar results by identifying specific intervals of the cavity diameter and the pore volume. However, this structure–property relationship can only be used as a description of selective materials, and no prediction could be achieved by only using structural descriptors.

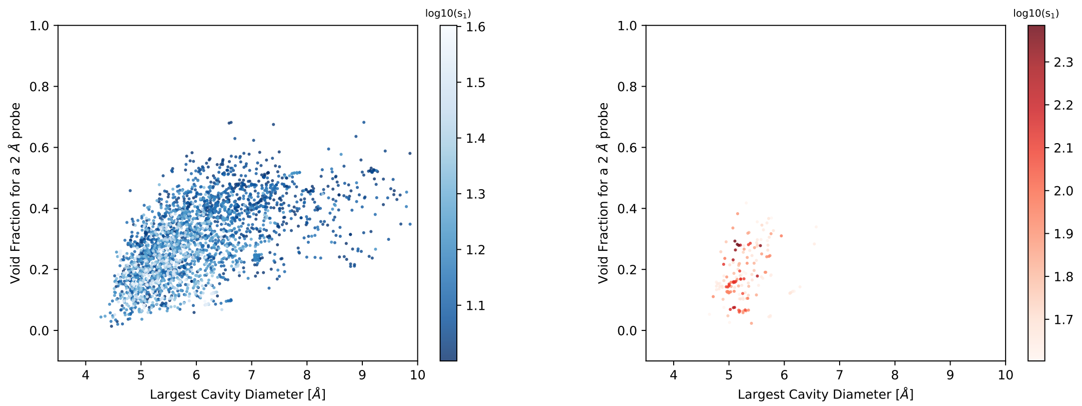


Figure 2.9: Scatterplots of the void fraction as a function of the D_i *and labeled by the log*$_1$*0 of the selectivity values. On the left, only the materials with a selectivity between* 10 *and* 40 *are considered; and on the right, selectivity values over* 40.

Effect of the composition

Finally, we previously only tackled one type of composition (20:80) associated to the extraction of xenon and krypton from a cryogenic distillation from the air (see section 1.3.1). We are now going to investigate the effects of the composition by looking at the case of xenon/krypton separation in spent nuclear fuel off-gases. In nuclear applications, the mixture has a 90:10 Xe/Kr ratio, which is much richer in xenon in comparison to the previous one. For this reason, the quantity of xenon adsorbed in the materials will mechanically be higher than previously. However, the second quotient in the formula of selectivity in equation 2.8 compensates the first one that will be logically higher. Here, we want to evaluate these two effects to see if they cancel each other out or there are different trends depending on the composition value.

As shown on the Figure 2.10, selectivity values of both compositions are quite close. However, we can note a slight decrease in performance when increasing the proportion of xenon in the mixture for some materials that are moderately selective (s between 2 and 50). This loss in performance could be explained by the fact that the materials display pores with different xenon affinities. With a lower proportion of xenon, the Xe adsorbates would access preferentially the most favorable pores and the small quantity of xenon is concentrated there. Whereas when there is a higher content of xenon, these most favorable sites begin to saturate and the xenon need to compete with krypton in much less favorable sites hence decreasing slightly the selectivity. We will see that later, we could use very similar reasoning to explain another variable that could change the selectivity.

We will now see the effect of the composition on the different analysis we carried out on the different structural descriptors. One of the major changes when considering a mixture with much more xenon is the values of the xenon uptake. The nanopores of selective materials ($1 < s_1 \leq 50$) are much more saturated in Xe and the maximum amount of xenon is therefore much higher. By comparing the Figures 2.8 and 2.11, the maximum uptake is now 11.7 mmol g^{-1} instead of 6.6 mmol g^{-1} (for the 20:80 composition). For moderately selective materials, at 20:80, the xenon competes with krypton mainly in the most selective nanopores, but when we increase the xenon content it has to compete with krypton in much less favorable sites since the other sites are already saturated. We previously stated that the most selective ($s_1 > 50$) materials could

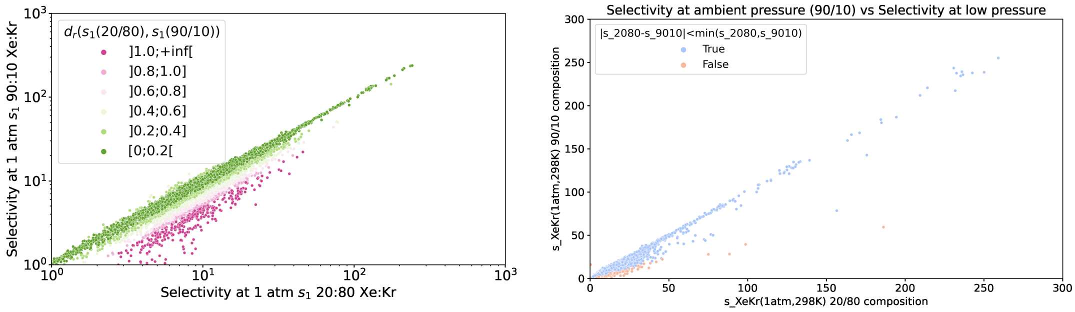


Figure 2.10: Illustration (scatterplot) of the difference of selectivity ($s_1(20:80)$ and $s_1(90:10)$) for two different Xe/Kr mixture compositions 20:80 (x-axis) and 90:10 (y-axis) at 1 atm *and* 298 K*. On the left, the axis is in log scale and the relative difference of selectivity between the two compositions is particularly high for the points labeled in purple. On the right, the axis is in linear scale and we only label the point to differentiate the materials with relative difference under and over* 1*.*

reach up to 4.0 mmol g^{-1} of xenon uptake, for a composition with a higher xenon content, and it reaches up to 4.2 mmol g^{-1} again, which is not a big change. For the most selective materials, the previous conclusion on the maximum uptake can still stand. Because of the extremely high selectivity, the change in composition does not change the nature of the adsorbed state and about the same quantities of xenon are present in the pores. Finally, we can say that a higher content in xenon does not influence a lot the most selective materials' performance but it could alter the selectivity and increase a lot the xenon uptake for some moderately selective materials.

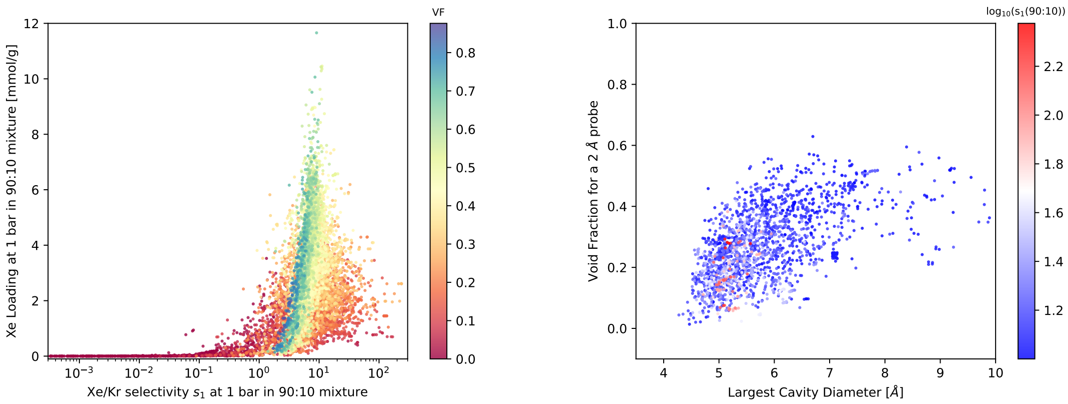


Figure 2.11: Illustration of the effect of the composition by representing the same figures as in 2.8 and 2.11 but for a 90:10 composition. On the left: scatterplot of the xenon uptake as a function of the selectivity ($s_1(90:10)$) and labeled by the values of the void fraction. On the right: scatterplots of the void fraction as a function of the D_i *and labeled by the selectivity ($s_1(90:10)$) values superior to* 10 *in log-scale.*

Finally, the composition does not affect the previously determined structural characteristics that a material needs to be very selective. As we can see on the right plot of the Figure 2.11 (right), the most selective materials still have pore size around 5 Å and porosity under 40%. This structural domain constitutes a necessary condition a selective material should have but being in this domain is not enough since less selective materials can also display these characteristics.

Now that we described the geometrical conditions needed to have a good selectivity, we will focus on the thermodynamic origins of the selectivity by focusing on energy-based quantities and the different correlations between them.

2.2.2 Thermodynamic quantities correlations at infinite dilution

In this section, our goal is not directly to address the structure–property relationships, but rather to map out the details of the thermodynamic features of Xe/Kr adsorption and separation in nanoporous materials. We used the high-throughput screening methodology as a way to map out the space of thermodynamic properties, going beyond the usual quantities of selectivity and uptake, to focus more specifically on the role of adsorption enthalpy and entropy, the differences between Xe and Kr adsorption thermodynamics, and between selectivity at low and high pressure.

To evaluate the performance of a given nanoporous material for separation in the low loading (or low pressure) limit, Henry's constants are often calculated from linear fits of low-pressure adsorption isotherm data — both experimentally and computationally. In this section, we investigate the thermodynamics of Xe and Kr adsorption at low pressure. Here, we have calculated the low-pressure adsorption properties by using the Widom insertion method [197, 205] on 9,668 structures from the dataset selected. It has higher accuracy than the fitting of isotherms, where it can be difficult to know what the extent of the linear adsorption regime is. With these simulations, we could obtain for each material the Henry's constant K and the adsorption enthalpy $\Delta_{\text{ads}}\text{H}_0$ (at the zero loading limit) for both xenon and krypton. The Xe/Kr thermodynamic selectivity s_0 in the low-pressure limit is then determined by the ratio $s_0 = \text{K}^{\text{Xe}}/\text{K}^{\text{Kr}}$ of the Henry's constants for the two gases. In the following, we look at the statistical relationships between the thermodynamic quantities at low pressure: s_0, K^{Xe}, K^{Kr}, $\Delta_{\text{ads}}\text{H}_0^{\text{Xe}}$, $\Delta_{\text{ads}}\text{H}_0^{\text{Kr}}$ and $\Delta_{\text{exc}}\text{H}_0$.

We display the distribution of thermodynamic properties of materials with favorable thermodynamic Xe/Kr selectivity ($s_0 > 1$) in Figure 2.12 — we restrict these plots to selectivity above 1, because those are the materials of interest for separation, and doing so removes several outliers with specific geometries or binding sites (but does not change the overall conclusions). We can first see that although the logarithm of the Xe Henry's constant K^{Xe} is weakly correlated to the logarithm of the selectivity s_0, this correlation is stronger for highly selective materials. Therefore, in a multistep screening study to identify the most selective materials, it could be possible to use as a "first filter" criterion based purely on Xe adsorption, discarding materials below a certain threshold (e.g., the materials with $s_0 \geq 30$ are contained in the subset with $\text{K}^{\text{Xe}} \geq 2.7\,10^{-1}$ mmol g^{-1} Pa^{-1}). The correlation between K^{Kr} and s_0, on the other hand, is weaker.

With regard to Henry's constants, we observe a broad selection of behavior, with K^{Xe} ranging from $2.6\,10^{-7}$ to $7.9\,10^{-1}$ mmol g^{-1} Pa^{-1}, and K^{Kr} ranging from $1.3\,10^{-7}$ to $5.1\,10^{-3}$ mmol g^{-1} Pa^{-1}. We also see that statistically, a high affinity for xenon usually translates into a high (relative) affinity for krypton, which is a general trend for noble gases where the adsorption sites are not strongly specific. In order to look more in detail into the thermodynamics behind this wide diversity in behavior, we plot in Figure 2.13 the enthalpies involved.

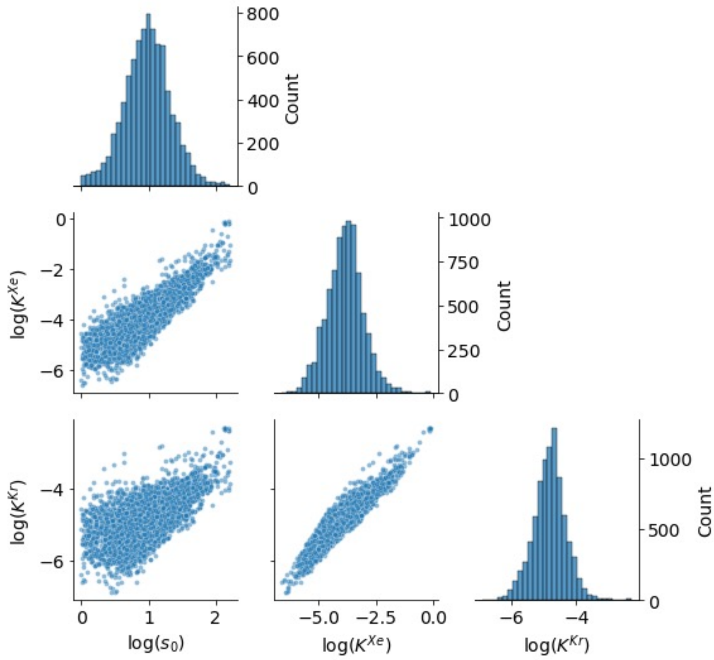


Figure 2.12: For 8,401 MOFs with favorable thermodynamic Xe/Kr selectivity ($s_0 > 1$), pair plots of $\log_{10}(s_0)$, $\log_{10}(K^{Xe})$ and $\log_{10}(K^{Kr})$ (the Henry's constants are in $\mathrm{mmol\,g^{-1}\,Pa^{-1}}$) in the off-diagonal subplots (note that the y-axis is displayed on the right side) and the distribution of each quantity are on the diagonal (note that the y-axis displayed on the right side corresponds to the count and the x-axis is correctly labeled below each subplot).

We first observe that the low-loading adsorption enthalpy of xenon ($\Delta_{ads}H_0^{Xe}$) is strongly correlated to that of krypton ($\Delta_{ads}H_0^{Kr}$). Echoing the similar correlation seen between respective Henry's constants, it suggests a rather generic physisorption mechanism is at play in the majority of materials, and that host–adsorbate affinities are mainly determined by the enthalpy. The main driver of Xe/Kr selectivity is neither the xenon nor krypton adsorption enthalpy alone (both are weakly correlated to the selectivity), but as expected their difference, $\Delta_{exc}H_0$, which is strongly correlated to $\log(s_0)$. This is further confirmed by the lack of correlation between selectivity and adsorption entropies (*cf.* Figure 2.14): the separation is mostly enthalpic in nature, and the entropy causes the dispersion in the correlation between selectivity $\log(s_0)$ and $\Delta_{exc}H_0$.

Analyzing the Figure 2.14 in more detail, the adsorption entropy of xenon and krypton being noticeably correlated, their difference (the exchange entropy) does not have a lot of variations (see Figure 2.15) compared to the enthalpy. This thermodynamic quantity plays a minor role in the selectivity performance of the materials. However it seems that the most selective materials do not have any values of exchange entropy but is centered around a value of about $-10\,\mathrm{kJ\,mol^{-1}\,K^{-1}}$. This is not a clean correlation but rather a necessary attribute of a selective material.

To emphasize one more time on the enthalpic nature of the separation by comparing the base-10 logarithm of the Henry constant (proportional to the adsorption free energy) and the adsorption enthalpy for both xenon and krypton. We can see, Figure 2.16 that the free energy can be almost totally explained by the enthalpy, which confirms the secondary role of the

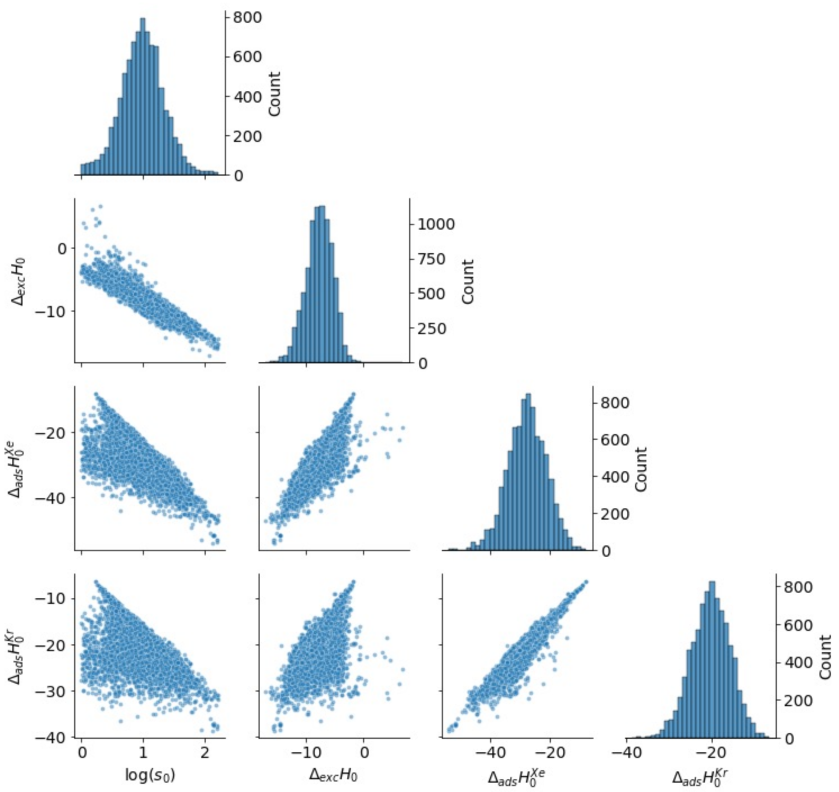


Figure 2.13: For 8,401 MOFs with favorable thermodynamic Xe/Kr selectivity (s_0 > 1), pair plots of $\log(s_0)$, $\Delta_{exc}H_0$, $\Delta_{ads}H_0^{Xe}$ *and* $\Delta_{ads}H_0^{Kr}$ *(the enthalpies are in* $\mathrm{kJ\,mol^{-1}}$*) in the off-diagonal subplots and the distribution of each quantity is on the diagonal.*

entropy that explains the variance in this linear relation. The effect of the entropy makes the correlation quite weak for the less favorable adsorption materials, but as we move to more negative values of the adsorption enthalpies, the correlation is stronger and stronger. The most selective materials have an almost negligible entropic contribution in the final free energy value (G = H − TS).

To go a little bit further in the correlation interpretation, the Figure 2.17 suggests that the entropic effect depends on the pore size. The bigger the size of the pores, the more positive the entropic term, which explains the weaker correlation for less attractive materials.

To check this, we looked at the influence of the pore size and the void fraction on the entropic term $T\Delta_{ads}S_0^{Xe}$ (see Figure 2.18). The entropy is clearly related to the pore size here represented by the LCD_{UFF} — the larger the pore the higher the entropy is likely to be. This can simply be explained by the confinement effect of the pore — a small pore gives very little possible adsorbable positions to the xenon, whereas a larger pore opens up the possible sites for the adsorption. The same trend can be observed for the pore volume represented by the void fraction here. A weak linear correlation exists between the void fraction (in log-scale) and the adsorption entropic term of xenon. We can however argue that the whole picture of the entropic behavior is not captured by these simple geometric descriptors, especially for the larger pore sizes; other effects also play a role such as the shape of the channel and cavities (e.g. tortuosity) or the whole picture of the pore size distribution that cannot be simply captured in the LCD_{UFF}.

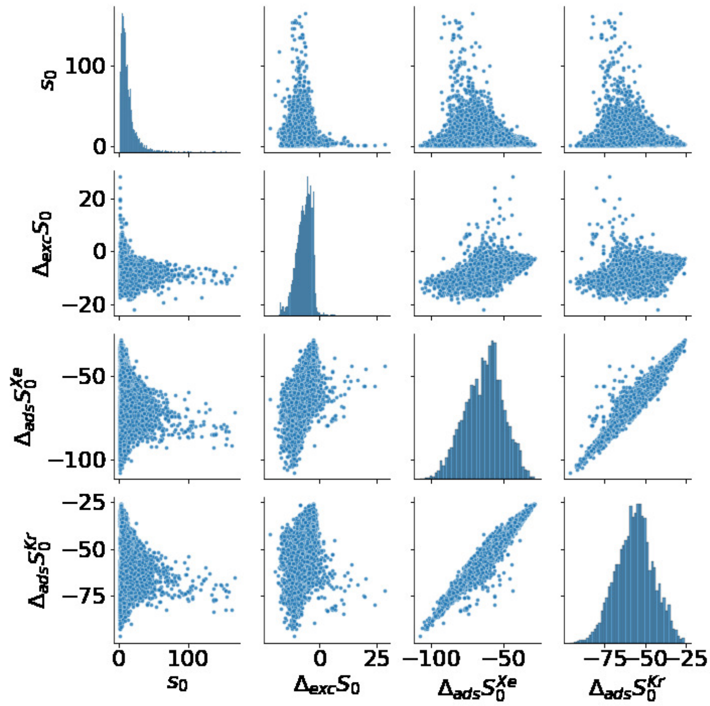


Figure 2.14: For 8,401 MOFs with favorable thermodynamic Xe/Kr selectivity ($s_0 > 1$), pair plots of s_0, $\Delta_{exc}S_0$, $\Delta_{ads}S_0^{Xe}$ and $\Delta_{ads}S_0^{Kr}$ in the off-diagonal subplots and the distribution of each quantity are on the diagonal.

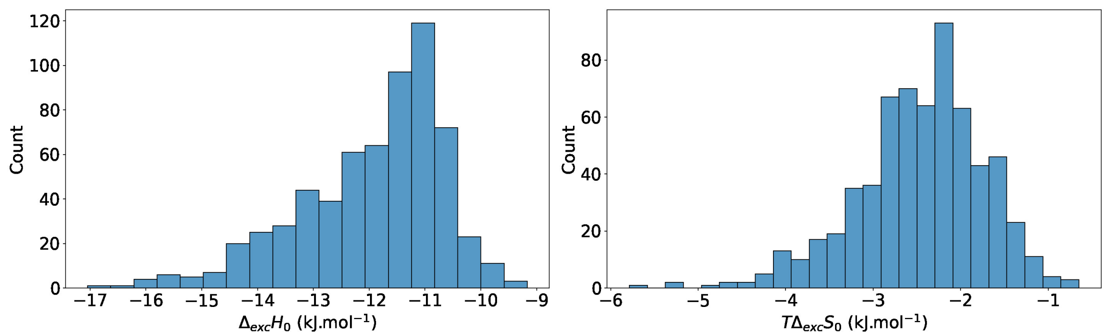


Figure 2.15: Distribution of the enthalpy $\Delta_{exc}H_0$ and entropy $T\Delta_{exc}S_0$ of exchange at low pressure on the 630 most selective structures

If we now croos these results with the previous results obtained on the influence of geometric descrioptors in the section 2.2.1, we can note that the entropic effect goes in the same direction than the enthalpic term for explaining the selecitivity, when the pore size is around the size of a xenon. The confinement of the xenon makes the entropy lower in the adsorbed phase than in the gas phase, which is even more true for pores whose size are tailored for a xenon. The second benefit of this type of pore is the optimal interaction with the surrounding framework atoms, which lowers down the enthalpic term. Both effects are going in the same way and explain the optimal selectivity for this particular pore size value (around 5 Å).

The main takeaway messages of this section are based on two relations. The first one is between the Henry constant of xenon and the selectivity: knowing the performance of xenon we can infer the Xe/Kr selectivity — the most selective materials have very high xenon affinity. The

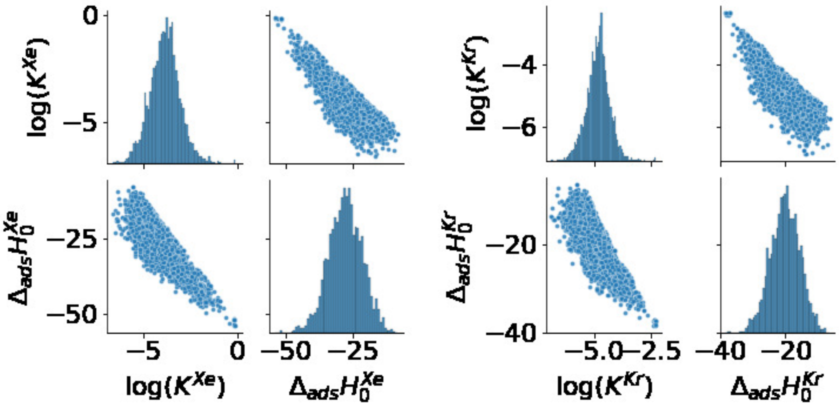


Figure 2.16: For 8,401 MOFs with favorable thermodynamic Xe/Kr selectivity ($s_0 > 1$), pair plots of $\log(\mathrm{K}_H^i)$ and $\Delta_{ads}\mathrm{H}_0^i$ in the off-diagonal subplots for both i=Xe and i=Kr and the distribution of each quantity are on the diagonal.

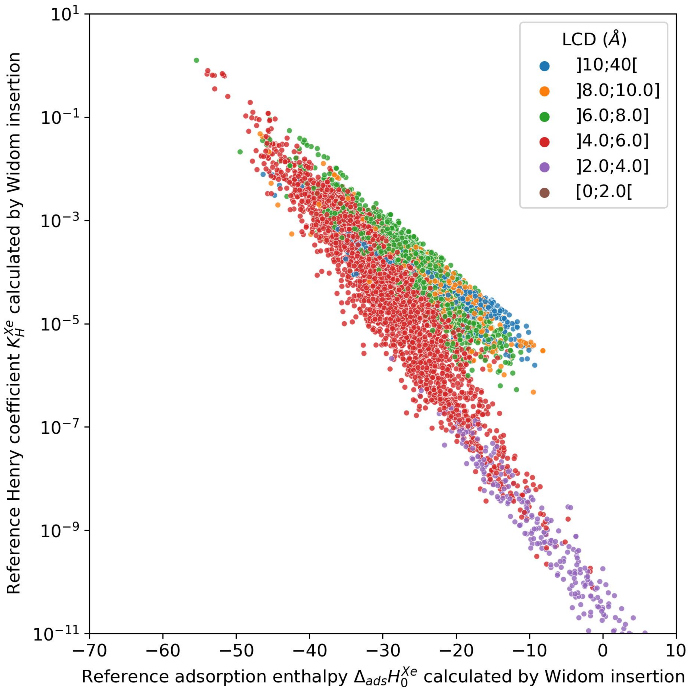


Figure 2.17: Comparison between the Xe Henry constant and Xe adsorption enthalpy labeled by categories of LCD_{UFF} values for the CoRE MOF structures.

second one concerns the relation between enthalpy and selectivity — the separation process has an enthalpic nature as a first-order approximation, which is even more true for the most selective materials. By studying the energy interactions within the material, we can understand most of the performance of it. We only looked at the thermodynamic properties at infinite dilution, in the next section we will focus on the effect of the pressure in the selectivity by focusing on a 20:80 Xe/Kr mixture at 1 atm and 298 K.

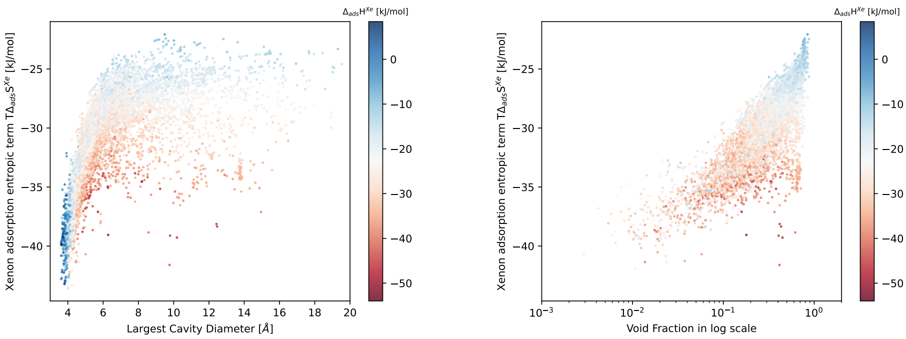


Figure 2.18: Comparison plots of the entropic term $\mathrm{T}\Delta_{ads}\mathrm{S}_0^{Xe}$ *at infinite dilution and two geometric descriptors: the* LCD_{UFF} *(left) and the void fraction (right).*

2.3 Selectivity Drop between Two Pressure Regimes

2.3.1 Thermodynamic origins

s After looking in the depth of the thermodynamics of the infinite dilution case, we will now focus on the impact of a change of working pressure on the adsorption selectivity, and analyze its thermodynamic origins. This is key to accurately assess the thermodynamics of adsorption in different working conditions for specific industrial processes, and any insight into the impact of pressure on selectivity may allow for faster screening limited at selected thermodynamic conditions.

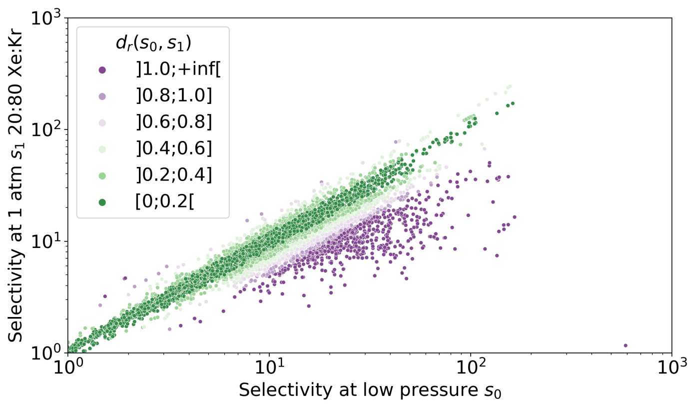


Figure 2.19: Difference of selectivity between low pressure and at a 1013 hPa *pressure for a 20:80 xenon krypton composition. The relative difference between the low-pressure selectivity and the ambient pressure is particularly high for the points labeled in purple.*

We calculated the selectivity s_1 at pressure 1 atm and ambient temperature using GCMC calculations on the entire dataset, with Xe/Kr mixture composition of 20:80 (found in a byproduct stream from air separation[1]) and 90:10 (found in the off-gas streams from nuclear waste[171]). For high-selectivity materials, we find that the impact of composition appears rather marginal (*cf.* Figure 2.10). In the following, we discuss the selectivity for the 20:80 mixture, which is the most commonly studied one in the literature. To measure the difference in selectivity between low and ambient pressures, we consider a relative difference $d_r(s_0, s_1)$ defined in equation 2.29.

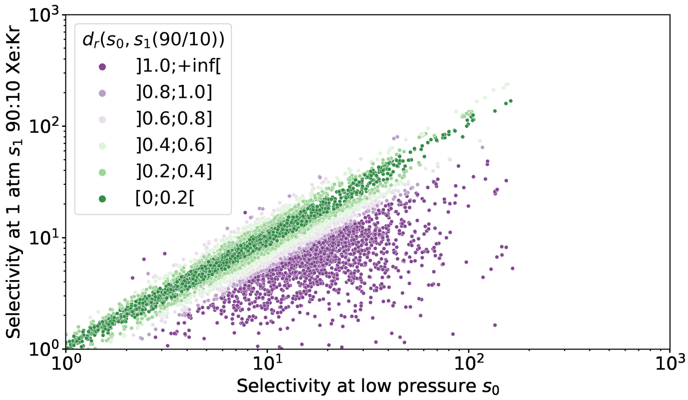


Figure 2.20: Difference of selectivity between low pressure and at a 1013 hPa *pressure for a 90:10 xenon krypton composition. The relative difference between the low-pressure selectivity and the ambient pressure is particularly high for the points labeled in purple.*

$$d_r(s_0, s_1) = \frac{|s_0 - s_1|}{\min(s_0, s_1)} \tag{2.29}$$

In Figure 2.19, the selectivity at ambient pressure s_1 is plotted against its low-pressure counterpart s_0 (for materials where $s_0 > 1$, as before). The points are color-coded according to the value of $d_r(s_0, s_1)$, in 6 discrete categories for the sake of clarity. There is some broad level of correlation, see near the diagonal with 61.5% of materials where the difference is below 20% (near the $s_0 = s_1$ line). We also see clearly that there are many more points (74.3% among the materials with $d_r(s_0, s_1) \geq 0.2$) below the first bisector ($s_1 < s_0$) than above: for these materials the selectivity s_1 at 1 atm is significantly lower than the one at low pressure s_0.

This drop in selectivity mainly concerns the materials with a relatively high selectivity $s_0 > 10$ (see Figure 2.19), and forewarns that considering solely pure-component Henry's constant (i.e., zero-pressure selectivity) for materials screening could be misleading in some cases. Although it is simpler and faster to calculate, those low-pressure results that can overestimate selectivity by more than 100% in a significant number of materials (646 out of 9,668 in our dataset). By using a thermodynamic approach, we now try to explain the reasons behind these shifts in selectivity.

If we now look at the 90:10 composition, we can note that the drop in selectivity is even more important. The selectivity with a higher proportion of xenon was already found to be higher than the selectivity for 20:80 composition (see Figure 2.11); we explained this drop by the presence of more or less favorable adsorption sites. In some materials (labeled in purple), at a low xenon content composition, the xenon and krypton mainly compete in the most favorable sites until these sites are saturated and no xenon is left to compete in the less selective sites. When we increase the Xe/Kr ratio, these less selective nanopores drive the overall selectivity down. Combined with the effect of increasing the pressure, some materials undergo both phenomena and have an exacerbated drop in selectivity compared to the selectivity at low pressure.

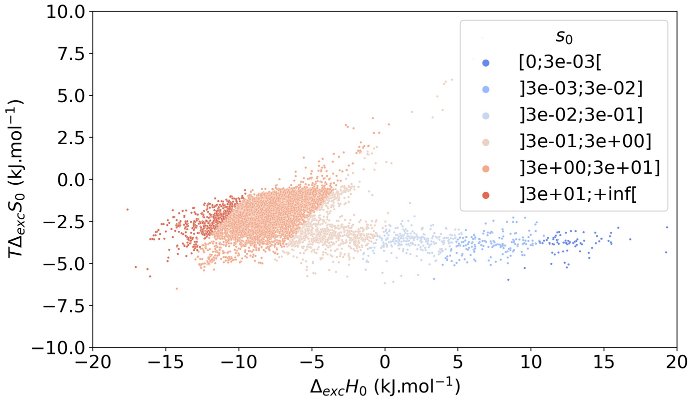


Figure 2.21: The energetic equivalent of exchange entropy $\mathrm{T}\Delta_{exc}\mathrm{S}_0$ *and enthalpy* $\Delta_{exc}\mathrm{H}_0$ *at low pressure labeled using the selectivity* s_0 *at low pressure. The limits between labels follows an affine function of slope* $1/\mathrm{T}$ *and of intercept* $-\mathrm{R}\ln(s_0^{lim})$ *where* s_0^{lim} *is the limit selectivity value (cf. Equation (2.28)). In other words, the iso-selectivity lines are all parallel lines of equation* $y = f(x)$ *where* f *is the affine function described previously.*

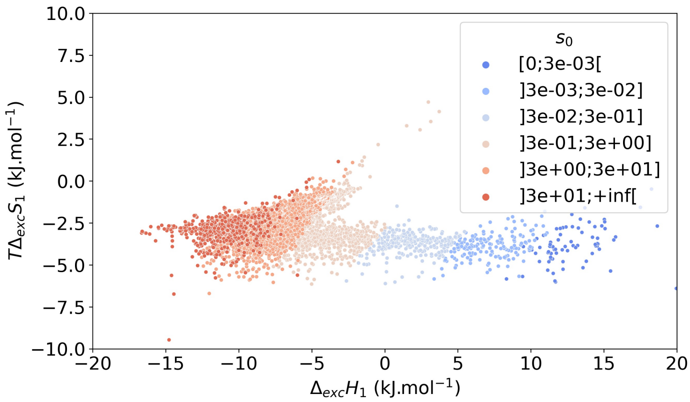


Figure 2.22: The energetic equivalent of exchange entropy $\mathrm{T}\Delta_{exc}\mathrm{S}_1$ *and enthalpy* $\Delta_{exc}\mathrm{H}_1$ *at ambient pressure labeled using the selectivity* s_0 *at low pressure. The points are layered so that the points with higher* s_0 *are always above. To see a split version of this plot, please refer to the Figure 2.24.*

To evaluate quantitatively the thermodynamic effects at play in the competitive adsorption in different regimes, we can consider thermodynamic properties of the "exchange equilibrium" predefined in equation 2.24. We plot in Figure 2.21 the exchange entropy at low pressure (plotted as $\mathrm{T}\Delta_{\mathrm{exc}}\mathrm{S}_0$) against the exchange enthalpy $\Delta_{\mathrm{exc}}\mathrm{H}_0$. In this scatterplot, the points are color-coded according to the selectivity s_0 (with discrete categories for the sake of clarity), which is related to the enthalpy and entropy through Equation 2.28 — meaning iso-selectivity lines are parallel straight lines in this scatterplot.

In the Figure 2.15, we display the distributions of the exchange enthalpy and entropy at low pressure. For the 630 most selective materials ($s_0 > 30$), the distribution of the exchange enthalpy $\Delta_{\mathrm{exc}}\mathrm{H}_0$ is centered on $-12.0\,\mathrm{kJ\,mol^{-1}}$ with a standard deviation of $1.3\,\mathrm{kJ\,mol^{-1}}$, whereas the distribution of the exchange entropy (plotted as $\mathrm{T}\Delta_{\mathrm{exc}}\mathrm{S}_0$) is centered on $-2.5\,\mathrm{kJ\,mol^{-1}}$ with a standard deviation of $0.7\,\mathrm{kJ\,mol^{-1}}$. These figures, along with the overall distribution plotted in

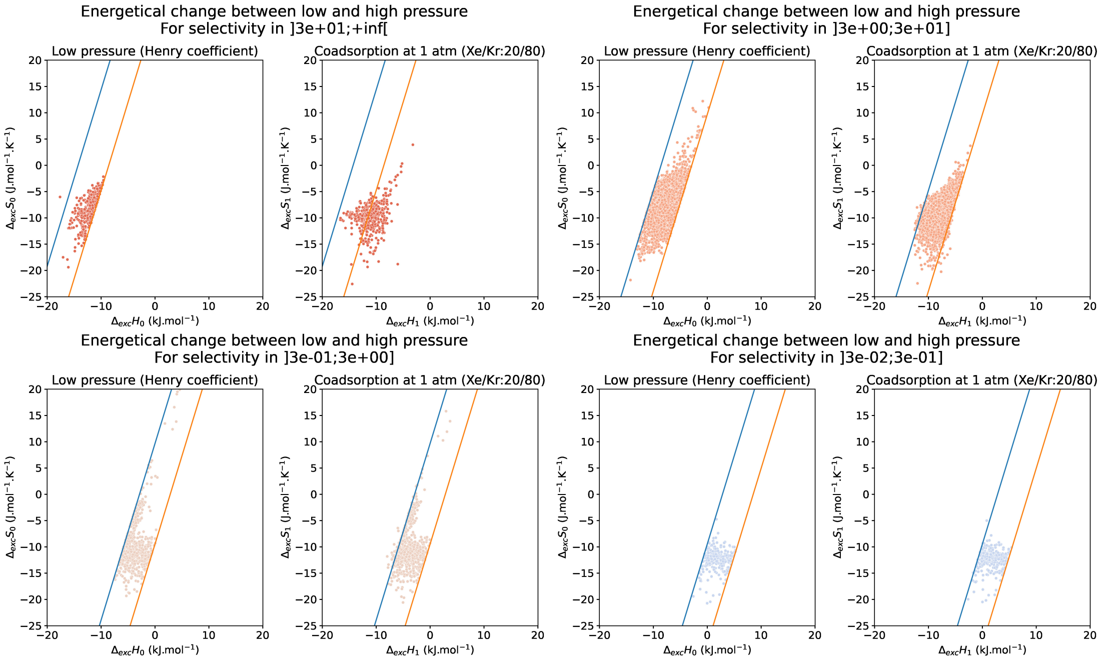


Figure 2.23: Split view of the Figure 2.21 and 2.22. The iso-selectivity lines for the limit considered are represented with blue and orange lines. We can clearly see the shift in exchange enthalpy for the structures with a selectivity higher than 30.

Figure 2.21, further confirms the moderate role of entropy in the low-pressure selectivity: it is equivalent in average to about 20% of the exchange enthalpy at low pressure.

Looking at the Figure 2.24, at ambient pressure, we can state the same conclusions on the limited influence of the entropy on the selectivity values. The distribution of the entropic term $T\Delta_{exc}S_1$ is now centered around -3 kJ mol^{-1}, which is also quite small in comparison of the values of $\Delta_{exc}H_1$. For the most selective materials, the entropic term represents about 19% of the exchange enthalpy at ambient pressure.

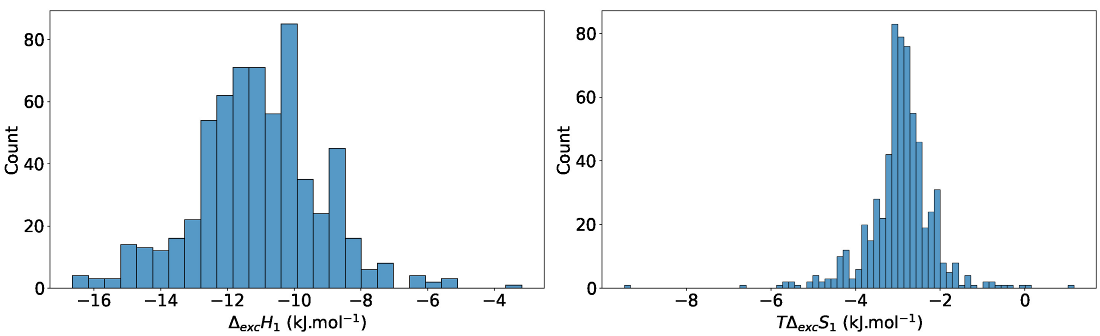


Figure 2.24: Distribution of the enthalpy $\Delta_{exc}H_1$ and entropic term $T\Delta_{exc}S_1$ of exchange at ambient pressure on the 630 most selective structures.

Figure 2.22 represents a scatterplot of the exchange entropy at P = 1 atm $\Delta_{exc}S_1$ against the exchange enthalpy at ambient pressure $\Delta_{exc}H_1$. To compare it to the Fig. 2.21, the points are color-coded according to the low-pressure selectivity s_0. Compared to the iso-selectivity s_1 straight parallel lines (*cf.* Figure 2.24), we can see that many materials with high s_0 have lower s_1 — seen as a migration of points to the right of the plot, compared to Fig. 2.21. This shift is

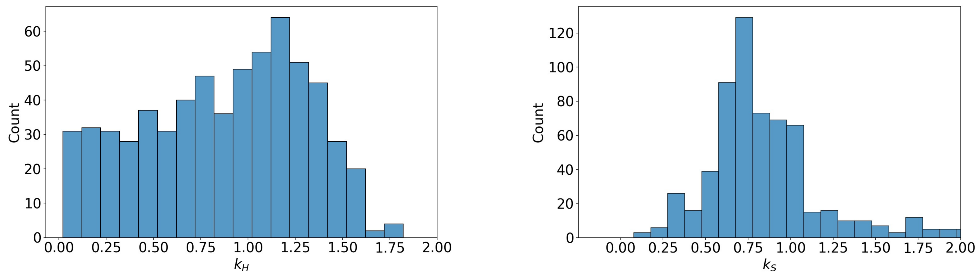


Figure 2.25: Distribution of the enthalpic k_H and entropic k_S contributions to the change of selectivity from low to ambient pressure for the 630 materials with $s_0 > 30$. k_H has a rather uniform distribution, whereas k_S has a bell-like distribution.

therefore mainly due to a higher (less favorable) exchange enthalpy, hinting at an important role of enthalpy to determine higher pressure selectivity.

To quantify this change, we consider the distributions of the exchange enthalpy $\Delta_{\text{exc}}H_1$ and the energetic equivalent of the exchange entropy $T\Delta_{\text{exc}}S_1$ at ambient pressure (Figure 2.24). The enthalpy $\Delta_{\text{exc}}H_1$ is now centered on $-11.1\,\text{kJ mol}^{-1}$ with a standard deviation of $1.9\,\text{kJ mol}^{-1}$. Compared to the zero-pressure values, the enthalpy distribution is more dispersed, showing that there are important changes in individual values, and is higher in average — majority of materials have lower ambient pressure selectivity due to enthalpic effects. This can be explained by the very general increase of adsorption enthalpy upon loading in the gas phase, which is linked to the presence of more adsorbed molecules. In fact, the correlations (Figure 2.12) suggest that highly selective materials have high affinity in xenon; therefore they feature significant uptake at 1 atm and the large Xe loading means the most favorable adsorption sites can be saturated, and further adsorption involves weaker host–guest interactions and therefore increases the average adsorption enthalpy at non-zero loading.

The entropic term $T\Delta_{\text{exc}}S_1$ is now centered on $-2.9\,\text{kJ mol}^{-1}$, with a standard deviation of $0.8\,\text{kJ mol}^{-1}$ (almost unchanged from low-pressure). The entropy is on average lower, which means an overall less favorable separation due to entropic effects: this evolution of the entropic term hints at the potential of a reorganization of the adsorbed molecules inside each material. The difference in distribution of enthalpy has, overall, more impact on the high-pressure selectivity than that of entropy. This suggests that the overall contribution of enthalpy remains more decisive than the role of entropy in the selectivity change, even at ambient pressure. This is an interesting conclusion for screening studies, because evaluation of adsorption enthalpy can be computationally faster than that of the adsorption free energy (or entropy).

To further investigate the thermodynamics of the selectivity change, we quantify in this section the contributions of enthalpy and entropy. The ratio s_1/s_0 is equal to the product $k_H \times k_S$ where k_H and k_S are the enthalpic and entropic contributions to the selectivity change defined as:

$$\begin{aligned} k_H &= \exp\left(-\frac{\Delta_{\text{exc}}H_1 - \Delta_{\text{exc}}H_0}{RT}\right) \\ k_S &= \exp\left(\frac{\Delta_{\text{exc}}S_1 - \Delta_{\text{exc}}S_0}{R}\right) \end{aligned} \tag{2.30}$$

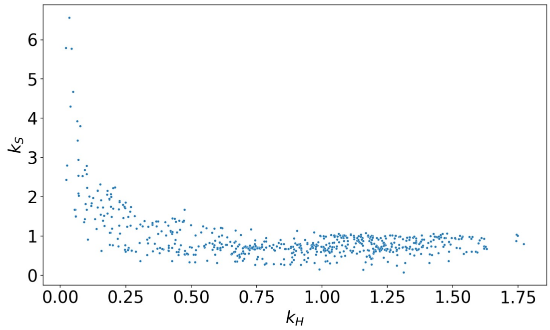


Figure 2.26: Scatterplot of the enthalpic contribution k_{H} and entropic contribution k_{S} for the 630 materials with $s_0 > 30$. The entropic compensation occurs when the enthalpic contribution is around 0.1*, else its value is around 1 and has little effect on the selectivity change.*

As we can see in Figure 2.25, the entropic contribution k_{S} has a bell-like distribution, with a mean of 0.9 and a standard deviation of 0.6. This confirms that k_{S} is close to 1, and has therefore only a marginal effect on the selectivity change. On the other hand, the enthalpic contribution k_{H} has a more uniform distribution ranging from 0.1 to 1.5, which means that enthalpy has a crucial role in the selectivity change we observe. There are a significant number of materials with a k_{H} close to zero, they correspond to the same materials highlighted in the section 2.3.1.

Furthermore, the scatterplot of k_{H} and k_{S} (shown in Figure 2.26) confirms a rather moderate effect of entropy. For most of the materials with $0.25 \leq k_{\mathrm{H}} \leq 1.75$, we see that k_{S} is close to 1. The most significant entropic contributions are found for materials where k_{H} is close to zero (typically below 0.25). If we look in more detail at the 29 materials with $k_{\mathrm{S}} > 2$, the entropic contribution k_{S} moderately compensate the enthalpic contribution as the average ratio s_1/s_0 is around 0.25. In such cases, the entropy is non-negligible and it can partially compensate the enthalpic contribution to the selectivity change, but the general trend is still given by enthalpy, since the overall selectivity is decreasing as a result.

2.3.2 Detailed investigation

In this section, we go over some of the most selective materials, as identified at low pressure and listed in Table 2.1, and we provide a detailed investigation of the thermodynamic effects behind their behavior. We can split them into three main categories: materials with a slight increase in selectivity or little change in selectivity ($s_0/s_1 > 0.8$), materials with a slight decrease in selectivity ($0.5 \leq s_0/s_1 \leq 0.8$) and materials with a significant decrease in selectivity ($s_0/s_1 < 0.5$). In this section, we investigate the origins of these different behaviors: all materials are referenced by their CSD refcode.

Before introducing the different archetypal structures that undergo different changes in selectivity, let us bring in some notions on adsorption isotherms. The isotherms are representation of the adsorbed quantity as a function of the pressure for different components at a given temperature. Here, we will only tackle the case of pure-component isotherms at 298 K. Different models have been developed to interpret these plots,[217] but here we will only use the Langmuir model as it is the most prominent equation to explain adsorption equilibria. The Langmuir model is a local model of adsorption based on the filling of a monolayer by non-interacting adsorbates. Depending on the distribution and shape of the pores, these isotherms can be either modeled by a 1-site Langmuir or a 2-site Langmuir model. At given temperature, some mono-site materials' isotherm can be described by the following equation:

$$q(\mathrm{P}) = \mathrm{N}_{\mathrm{max}} \frac{\mathrm{KP}}{1 + \mathrm{KP}} \tag{2.31}$$

where q is the adsorbed quantity of a mono-component gas, K is the adsorption equilibrium constant and P is the pressure. When the material has 2 sites, the isotherm can be described by the following equation:

$$q(\mathrm{P}) = \mathrm{N}_{\mathrm{max}} \left((1 - \alpha_2) \frac{\mathrm{K}_1\mathrm{P}}{1 + \mathrm{K}_1\mathrm{P}} + \alpha_2 \frac{\mathrm{K}_2\mathrm{P}}{1 + \mathrm{K}_2\mathrm{P}} \right) \tag{2.32}$$

where q is the loading of a given mono-component gas, K_1 and K_2 are the adsorption equilibrium constants in the respective sites, α_2 is the proportion of secondary sites, and P is the pressure.

We first study a few examples of the category of materials where ambient-pressure selectivity is close to (or even higher than) the low-pressure value. For VOKJIQ, the selectivity is multiplied by 1.5 between low and ambient pressure. We see that the adsorption enthalpy of xenon $\Delta_{\mathrm{ads}}\mathrm{H}^{\mathrm{Xe}}$ decreases from $-53.9\,\mathrm{kJ\,mol^{-1}}$ to $-61.1\,\mathrm{kJ\,mol^{-1}}$, whereas for krypton $\Delta_{\mathrm{ads}}\mathrm{H}^{\mathrm{Kr}}$ decreases from $-38.2\,\mathrm{kJ\,mol^{-1}}$ to $-44.5\,\mathrm{kJ\,mol^{-1}}$ (*cf.* Table 2.2). This increased stability of the adsorption sites upon loading is not common in nanoporous materials for rare gas adsorption, and can be linked to a cooperative effect between the adsorbed molecules. The stabilization favors the xenon molecules over the krypton molecules, due to an interatomic distance inside the pores that is a closer match to the energy well for favorable Lennard-Jones potential for xenon-xenon interactions than for krypton-krypton interactions (which is the case for a distance higher than 4.2 Å; see Figure 2.27).

Table 2.1: Enthalpic (k_H) and entropic (k_S) contributions to the selectivity change (s_1/s_0) between low and ambient pressures for some archetypal structures selected for their high s_0 selectivity at infinite dilution. Every structure is identified using a CSD Refcode and a reference the first article that mentions it. The pore size is also characterized using the diameters D_i and D_f in Å.

CSD Refcode	Ref.	s_0	s_1	s_1/s_0	k_H	k_S	D_i	D_f
VOKJIQ	[206]	157.17	242.73	1.54	1.46	1.06	5.2	3.2
KAXQIL	[47]	103.78	132.57	1.28	1.32	0.96	5.2	4.1
JUFBIX	[207]	106.11	114.83	1.08	1.08	1.00	5.3	3.0
FALQOA	[208]	162.20	171.10	1.05	1.09	0.96	5.1	3.5
GOMREG	[209]	114.14	73.83	0.65	1.01	0.64	5.8	4.0
JAVTAC	[210]	117.38	66.93	0.57	0.77	0.74	5.5	4.3
GOMRAC	[209]	124.11	47.34	0.38	0.58	0.66	5.7	3.7
MISQIQ	[211]	138.94	37.32	0.27	0.51	0.53	4.6	4.4
BAEDTA01	[212]	154.10	37.74	0.24	0.12	1.97	5.7	4.6
VIWMOF	[213]	81.13	13.24	0.16	0.04	4.30	10.2	5.3
LUDLAZ	[214]	165.68	16.42	0.10	0.16	0.63	6.7	4.2
WOJJOV	[215]	146.32	13.94	0.10	0.06	1.68	8.2	6.8
VAPBIZ	[216]	146.73	12.76	0.09	0.06	1.50	6.3	3.7

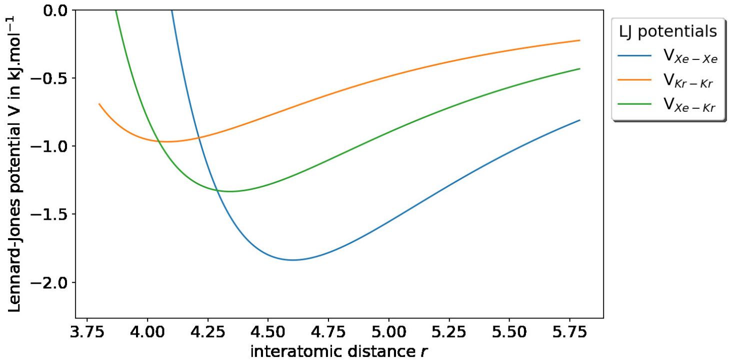


Figure 2.27: The LJ potentials for xenon and krypton interactions. The xenon-xenon interaction is more stabilizing than the krypton-krypton interaction for interatomic distance higher than 4.2 Å.

In the case of KAXQIL, the channels are one-dimensional tubes (see Figure 2.28) and the distance between two adsorption sites is approximately the unit cell parameter along the direction of the tube (5.6 Å). There the selectivity increases with pore filling, for enthalpic reasons, which we can explain by relatively simple reasoning. The Lennard-Jones potentials V_{LJ} can be estimated for all species at 5.6 Å: $V_{Xe\text{-}Xe} = -1.0\,\mathrm{kJ\,mol^{-1}}$, $V_{Kr\text{-}Kr} = -0.3\,\mathrm{kJ\,mol^{-1}}$ and $V_{Xe\text{-}Kr} = -0.5\,\mathrm{kJ\,mol^{-1}}$. In a simplistic model where all adsorbed molecules are 5.6 Å apart, the cooperative effect is higher between two xenon molecules, which explains the increased selectivity at high uptake. If we look further at the adsorption enthalpy of both xenon and krypton (*cf.* Table 2.2), they both increase: the guest molecules move from the "ideal" adsorption sites, and the guest–guest interactions do not fully compensate. The selectivity change in this material is therefore a consequence of the guest–guest interactions that rearranges the position of the adsorbates inside the nanopores.

Table 2.2: Thermodynamic quantities associated for a few archetypal structures. Henry's constant K^{Xe}, K^{Kr} *are in* $\mathrm{mmol\,g^{-1}\,Pa^{-1}}$, *loadings* q_1^{Xe} *and* q_1^{Kr} *are in* $\mathrm{mmol\,g^{-1}}$, *enthalpies* $\Delta_{ads}H_0^{Xe}$, $\Delta_{ads}H_0^{Xe}$, $\Delta_{ads}H_1^{Xe}$ *and* $\Delta_{ads}H_1^{Xe}$ *are in* $\mathrm{kJ\,mol^{-1}}$

CSD Refcode	Ref.	s_0	K^{Xe}	K^{Kr}	$\Delta_{ads}H_0^{Xe}$	$\Delta_{ads}H_0^{Kr}$	s_1	q_1^{Xe}	q_1^{Kr}	$\Delta_{ads}H_1^{Xe}$	$\Delta_{ads}H_1^{Xe}$
VOKJIQ	[206]	157	$7.92\,10^{-1}$	$5.04\,10^{-3}$	−53.9	−38.2	243	2.57	0.04	−61.1	−44.5
KAXQIL	[47]	104	$3.01\,10^{-2}$	$2.90\,10^{-4}$	−44.6	−30.5	133	1.41	0.04	−41.5	−26.8
JUFBIX	[207]	106	$1.59\,10^{-2}$	$1.50\,10^{-4}$	−45.6	−31.4	115	0.80	0.03	−45.7	−31.3
FALQOA	[208]	162	$2.23\,10^{-2}$	$1.38\,10^{-4}$	−47.3	−32.0	171	0.68	0.02	−48.6	−33.1
GOMREG	[209]	114	$9.16\,10^{-2}$	$8.03\,10^{-4}$	−44.7	−31.1	74	2.59	0.14	−47.5	−33.8
JAVTAC	[210]	117	$1.24\,10^{-1}$	$1.06\,10^{-3}$	−47.7	−33.5	67	1.50	0.09	−48.5	−34.9
GOMRAC	[209]	124	$1.17\,10^{-1}$	$9.45\,10^{-4}$	−45.6	−31.8	47	2.51	0.21	−47.3	−34.8
MISQIQ	[211]	139	$6.87\,10^{-1}$	$4.94\,10^{-3}$	−51.9	−37.4	37	2.30	0.25	−45.6	−32.8
BAEDTA01	[212]	154	$1.39\,10^{-2}$	$9.04\,10^{-5}$	−47.7	−31.7	38	1.05	11	−34.0	−23.1
VIWMOF	[213]	81	$7.87\,10^{-3}$	$9.70\,10^{-5}$	−46.3	−30.1	13	2.99	0.90	−26.0	−17.8
LUDLAZ	[214]	166	$9.04\,10^{-2}$	$5.46\,10^{-4}$	−45.4	−30.9	16	1.59	0.39	−38.3	−28.3
WOJJOV	[215]	146	$4.19\,10^{-2}$	$2.86\,10^{-4}$	−46.4	−30.7	14	2.82	0.81	−33.0	−24.4
VAPBIZ	[216]	147	$3.54\,10^{-2}$	$2.41\,10^{-4}$	−46.4	−30.5	13	2.50	0.78	−34.1	−25.3

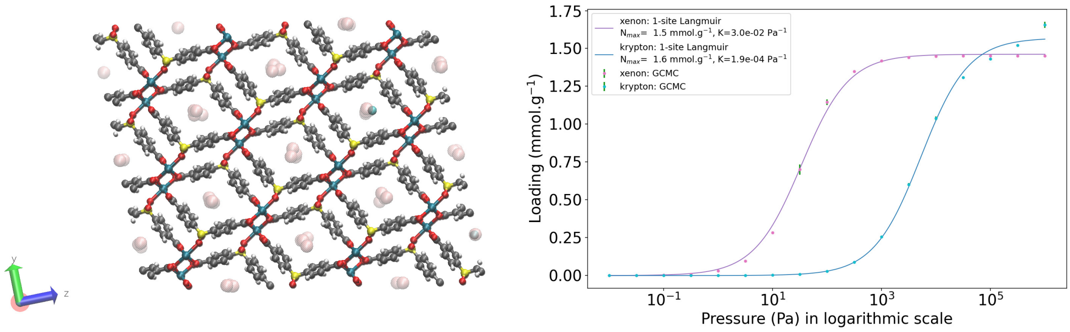


Figure 2.28: KAXQIL: On the left side, an illustration of a clean version (all solvent removed) of the calcium coordination framework [Ca(SDB)]·H_2O, where SDB = 4,4′-sulfonyldibenzoate loaded with xenon and krypton obtained by GCMC calculations. Color code: Ca in dark cyan, C in gray, O in red, H in white, S in yellow ; Xe in transparent pink and Kr in cyan for the adsorbates. The mono-component isotherms fitted with a 1-site Langmuir model (Equation 2.31) for both xenon and krypton at 298 K *is represented on the right side.*

To further corroborate the role of the guest–guest interactions, we look at another material with one-dimensional tubelike channels: JUFBIX, a cobalt(II) coordination polymer based on carboxylic acid linkers (see Figure 2.29).[207] The periodicity along the direction of the tubes is much higher at 7.2 Å. The pair interaction energies corresponding to the LJ potentials at this distance are $V_{\text{Xe-Xe}} = -0.24$ kJ mol^{-1}, $V_{\text{Kr-Kr}} = -0.06$ kJ mol^{-1} and $V_{\text{Xe-Kr}} = -0.13$ kJ mol^{-1}. By looking at the adsorption enthalpies (Table 2.1), these values are too small to affect the position of the adsorbed molecules. At high loading, the distance between adsorbed molecules is high, and every adsorption site is independent of the others. The ambient-pressure selectivity s_1 is therefore the same as the low-pressure selectivity s_0, since every guest–guest interactions are negligible. It confirms the crucial role of cooperative effects between guest molecules when considering a saturated material.

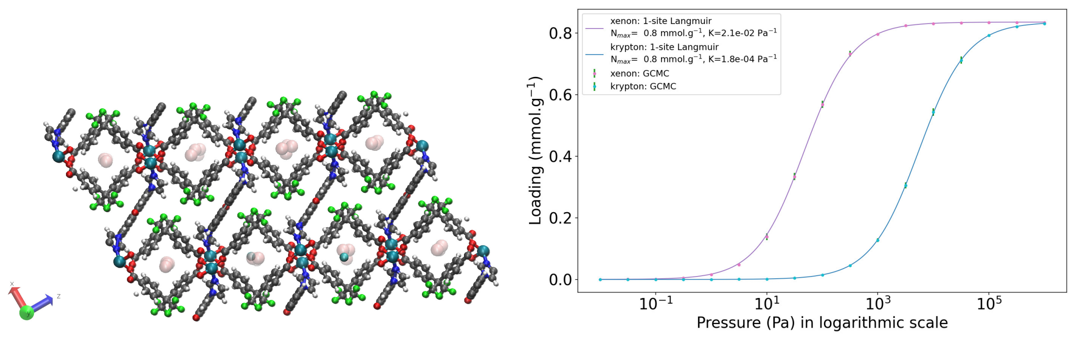


*Figure 2.29: JUFBIX: Representation of a clean version (all solvent removed) of the cobalt(II) coordination framework [Co_2(L)(ppda)$_2$]$_2$·H_2O, where the ligand L is 2,8-di(1*H*-imidazol-1-yl)dibenzofuran and the carboxylic acid ligand H_2ppda is 4,4′-(perfluoropropane-2,2-diyl)dibenzoic acid loaded with xenon and krypton obtained by GCMC calculations. Color code: Co in dark cyan, C in gray, O in red, H in white, N in blue, F in green ; Xe in transparent pink and Kr in cyan for the adsorbates. The mono-component isotherms fitted with a 1-site Langmuir model (Equation 2.31) for both xenon and krypton at* 298 K *is represented on the right side.*

GOMREG and JAVTAC are frameworks that belong to the second category of materials, with a moderate decrease in selectivity from low to ambient pressure. In GOMREG, the channels are composed of one-dimensional tubes larger than the ones found in KAXQIL or JUFBIX (see Figure 2.30 and Table 2.1). The adsorption sites are alternating from left to right inside the channel, and the adsorbed molecules organize in a "zigzag" pattern. Looking at the adsorption enthalpies, we see that both xenon and krypton have lower enthalpies by a similar margin, suggesting an equivalent stabilization for both atoms, hence the enthalpic contribution to the selectivity change is close to 1. Since krypton is smaller and less strongly tied on its adsorption site than xenon, it has more available space inside the pore space. This gives an entropic advantage to the Kr, seen in the entropic contribution k_S of 0.64 in Table 2.1. This indicates that even if enthalpic considerations mainly explain the observed changes at a statistical level, as discussed in the previous sections, for individual cases entropic considerations can be a strong factor in pressure-dependent selectivity.

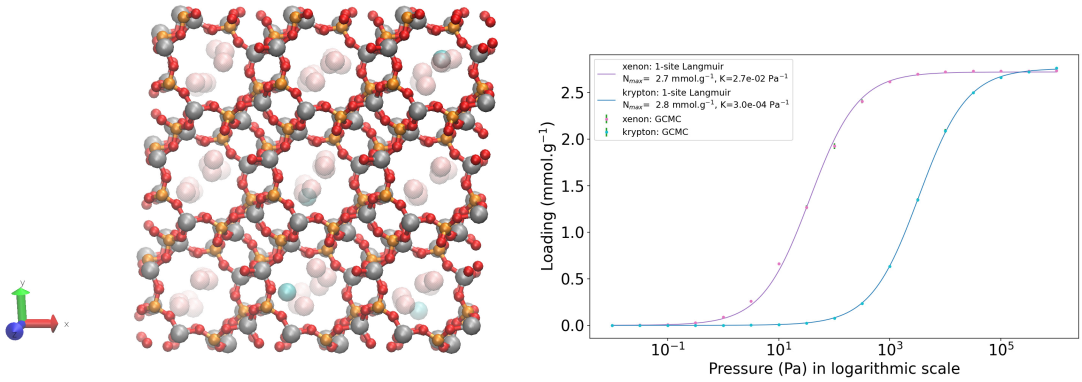


Figure 2.30: GOMREG: Representation of a clean version (all solvent removed) of this aluminophosphate $AlPO_4$-n that has a zeotype LAU topology with one-dimensional 10-ring channels loaded with xenon and krypton obtained by GCMC calculations. Color code: Al in silver, P in orange, O in red ; Xe in transparent pink and Kr in cyan for the adsorbates. The mono-component isotherms fitted with a 1-site Langmuir model (Equation 2.31) for both xenon and krypton at 298 K *is represented on the right side.*

The remaining materials discussed here form a third category, with a strong decrease in selectivity from low to ambient pressure. We look at several phenomena that can be at the root of this decrease, which is important for screening studies as it can limit the working performance of a material that appears to be a "top performer" based on zero-pressure screening.

For example, GOMRAC has a similar structure compared to GOMREG (see Figure 2.31), except for the fact that the pores and channels are smaller (see the values of the D_i, and the D_f, in Table 2.1). The distances between the adsorbed molecules — in their ideal sites — are then consequently smaller. At such distances, we can assume that the interactions between adsorbates become more stabilizing for krypton than for xenon molecules in GOMRAC (see LJ potentials at distance lower than 4.2 Å in the Figure 2.27), which translates into an enthalpic contribution k_H of 0.58. Moreover, this is compatible with the equivalent guest–guest interactions in GOMREG, as previously discussed. It explains why the difference between the adsorption enthalpies becomes smaller for GOMRAC, whereas it stays the same for GOMREG (between low and ambient pressure). This further validates the crucial role of the interactions between adsorbed molecules, and their relationship with the guest-guest distances when considering a high loading condition.

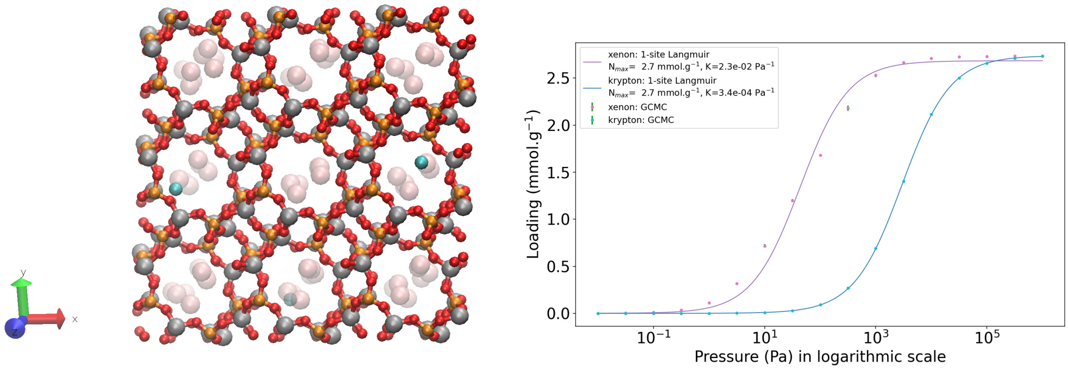


Figure 2.31: GOMRAC: Representation of a clean version (all solvent removed) of this aluminophosphate $AlPO_4$-n that has a zeotype LAU topology with one-dimensional 10-ring channels loaded with xenon and krypton obtained by GCMC calculations. Color code: Al in silver, P in orange, O in red ; Xe in transparent pink and Kr in cyan for the adsorbates. The mono-component isotherms fitted with a 1-site Langmuir model (Equation 2.31) for both xenon and krypton at 298 K *is represented on the right side. It seems that this aluminophosphate is just a smaller version of GOMREG.*

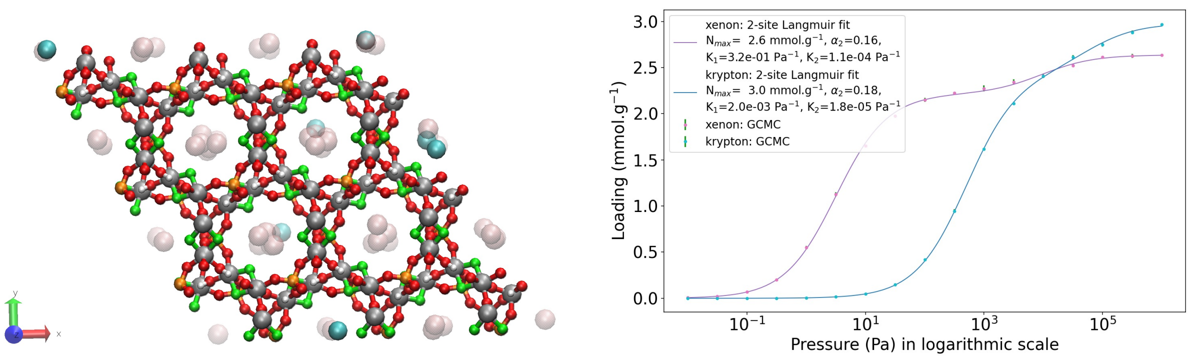


Figure 2.32: Representation of a chiral open-framework fluoroaluminophosphate $[C_4N_3H_{16}]\cdot[Al_6P_3O_{12}F_6(OH)_6]$ denoted AlPO-JU89 (referenced MISQIQ in the Cambridge structural database), which has been loaded with xenon and krypton in a GCMC simulation, on the left side.[211] Color code: Al in silver, P in orange, O in red, H in white and F in green for the framework; and Xe in transparent pink and Kr in cyan for the adsorbates. The pure-component isotherms fitted with a 2-site Langmuir model (Equation 2.32) for both xenon and krypton at 298 K *on the right side.*

If we look at the case of MISQIQ, we see that the pure-component Xe isotherm in Figure 2.32 cannot be fitted by a single-site Langmuir isotherm, but is well fitted by a two-site Langmuir model (see Figure 2.32). Visual inspection of the adsorbed density at various loadings shows that this is not a second, separate adsorption site that is populated at high loading: instead, the second step in the isotherm (representing about 20% of the uptake at full loading) is associated with a reorganization of the adsorbate molecules occurs at high loading, accompanying a contraction of the interatomic distances. In this case, the potential for a reorganization of the adsorbate in the material's nanopores leads to the change in selectivity. This reorganization can be detected on the basis of the xenon isotherm alone, and has a major role in the selectivity at ambient pressure. This repacking of the adsorbed phase is linked to a strong entropic effect, and also impacts the enthalpic contribution to selectivity.

More extreme cases of selectivity drop can occur when more than one site is available, as is the case for materials BAEDTA01, VIWMOF, LUDLAZ, WOJJOV, and VAPBIZ. The pure-component isotherms and the representation of the materials loaded in xenon and krypton molecules (presented in the supporting information of the Ref. [18] Figures S19-23) confirm the existence of at least two distinct adsorption sites in each material. The most selective sites (i.e., the most favorable for Xe) are filled in priority at low loading, and the less selective sites will then be populated when the pressure increases, leading to a net selectivity drop at ambient pressure for these materials. The different types of adsorption sites, and therefore the potential for a drop in Xe/Kr selectivity (at non-zero pressure) is a factor that could be explicitly included in screening of pure-component isotherms, without the need for explicit multi-component GCMC simulations.

2.3.3 Conclusion and introduction to the follow-up studies

In the current state of the art on Xe/Kr separation by adsorption in nanoporous materials, many studies have focused on the determination of structure/property relationships, the description of theoretical limits of performance, and the identification of top-performing materials, whether for existing experimental structures or for novel hypothetical structures yet to be synthesized. Here, we provide a study based on a high-throughput screening of the adsorption of Xe, Kr, and Xe/Kr mixtures in 12,020 experimental open-framework materials, in order to provide a better comprehension of the thermodynamics behind Xe/Kr separation in nanoporous materials and the microscopic origins of Xe/Kr selectivity at both low and ambient pressure.

The statistical correlation found between Henry's constant for Xe and Xe/Kr selectivity showed that the most selective materials are those with the highest affinity for xenon. To some degree of accuracy, we conclude that directly screening for Kr adsorption or for xenon adsorption free energy may not be necessary for a coarse-grained evaluation of a nanoporous framework selectivity. This could help build more efficient screening methodologies, for example with multistage studies with a first rough selection on Henry's constant at a low computational cost, followed by more expensive GCMC simulations on the selected materials (a gain that can be between 5 and 10-fold in our setup). Furthermore, inspection of the correlations between enthalpy and entropy contributions at low pressure showed that the adsorption-based separation process in the open frameworks studied is mainly enthalpic in nature. We intend to extend the study in the future to other classes of nanoporous materials beyond MOFs, including covalent organic frameworks, porous aromatic frameworks, purely inorganic porous frameworks such as zeolites, but also amorphous porous materials such as porous polymer membranes.

In order to use nanoporous materials to separate xenon from krypton, pressure swing adsorption (PSA) processes have been widely offered: pressure is therefore a crucial thermodynamic variable in the separation cycle. Here, we studied the difference of selectivity between a system under very low pressure (at the zero loading limit, which is calculated at relatively low computational cost) and a system at ambient pressure (closer to working conditions, but obtained at higher simulation cost). We demonstrated that the selectivity could be highly dependent on the pressure, with high low-pressure selectivity that could be maintained in some materials at ambient-pressure selectivity, while in others there would be a large drop in selectivity: a high ambient-pressure selectivity requires high low-pressure selectivity, but the reverse does not hold.

Using a thermodynamic approach to describe the separation selectivity, we showed that the differences in selectivity between the different pressures (and therefore different loading regimes of the frameworks) are mainly explained by the evolution of the adsorption enthalpies for Xe and Kr. By focusing on specific examples, we uncovered the microscopic origins of these selectivity changes, and related them to the relative roles of host–guest and guest–guest interactions. Population of different adsorption sites, or repacking of the adsorbed phase at higher loading, can lead to drastic changes in the overall selectivity. The mechanisms behind selectivity at high pressure are complex and unique to each framework, requiring a good understanding of the interactions between guest molecules constrained in the nanopores. Nevertheless, our classification of the interactions at play can help in the future to design more efficient high-throughput screening procedures.

For instance, the essentially enthalpic nature of the xenon/krypton separation process supports the need for more efficient ways of sampling the interaction energies and using them as cheap descriptors to tackle more and more numerous structures. In the next chapter, we will go over different ways of evaluating the adsorption enthalpy by comparing the computation time required and the accuracies of each of them. Finally, the influence of the partial pressure through the change in composition or in pressure questions the possible use of infinite dilution thermodynamic quantities to predict the selectivity at any pressure (GCMC). Many studies have focused on predicting the results of GCMC simulations;[14, 17, 178, 218] by using this new angle, can we achieve good results in predicting GCMC values?

Data Availability: `https://github.com/fxcoudert/citable-data/tree/master/132-Ren_FaradayDiscuss_2021`

3

ADSORPTION ENERGIES SAMPLING

3.1 Voronoi sampling

3.1.1 Theoretical considerations

In mathematics, a tessellation of a given space corresponds to a partition into non overlapping subspaces. In the Voronoi tessellation, named after Georgy Feodosevich Voronoy, a set of points (seeds) are associated to a tessellation of regions (Voronoi cells) so that each seed has a cell whose points are closer to this seed than any other seeds.[15] Applied in materials science, the Voronoi cells associated to each atom of the framework can be used to determine key geometrical descriptors (void volume, accessible surface area, pore sizes). This decomposition can also be used to sample adsorption energies as introduced by Simon et al. — an average of the interaction energies was calculates on the accessible vertices of each Voronoi cell.[14]

Equal radii

In a tridimensionnal space, let us consider the positions $(\boldsymbol{x}_k)_{k\in\{1,\dots,n\}}$ of the n points in a box B that could be periodically propagated in the whole space. For every $k \in \{1, \dots, n\}$, we can then define a subspace S_k (also called Voronoi cell) around the atom k so that any point $\boldsymbol{x}$ inside this subspace is closer to the position $\boldsymbol{x}_k$ than to any other points $\boldsymbol{x}_l$ ($l \neq k$).

$$S_k = \left\{\boldsymbol{x} \in B \mid \forall\, l \neq k,\ \|\boldsymbol{x} - \boldsymbol{x}_k\| \leq \|\boldsymbol{x} - \boldsymbol{x}_l\|\right\} \tag{3.1}$$

The set of all these 3D polyhedral subspaces S_k is then called the Voronoi partition of the space B. The edges and vertices of these polyhedra can then give valuable information of the void space between the adjacent Voronoi cells associated to them. We can quickly use them to determine the accessible and inaccessible points of the void space. For instance, a vertex $\boldsymbol{v}$ of p subspaces $\left\{V_{i_1}, \dots, V_{i_p}\right\}$ is the closest point to the atomic positions $\boldsymbol{x}_{i_1}, \dots, \boldsymbol{x}_{i_p}$ — this can simply be proven by combining the different conditions of equation 3.1. The same assessment can be done for any point on an edge adjacent to some subspaces, it will be closer to the atoms associated with these subspaces than to any other atoms.

Figure 3.1: Bidimensional illustrations of a Voronoi decomposition using three types of algorithm: (i) for equally sized circles using the equation 3.1 (`www.shadertoy.com/view/MslGD8`) (ii) for unequally sized circles using the Apollonian Voronoi decomposition condition 3.2 (`www.shadertoy.com/view/4sd3D7`) and (iii) another alforithm for unequally sized circles using the radical Voronoi condition 3.3 (`www.shadertoy.com/view/4tV3z3`). Note that the second picture shows the curved boundaries between the Voronoi cells, while the switch to the radical Voronoi decomposition gives straight line boundaries.

This regular Voronoi tessellation can only be used to separate the space for equally sized atoms because it sets the boundaries at equidistance of all the surrounding atoms, as we can see on the Figure 3.1. For unequally sized atoms, this type of definition could be undesired since the boundary can be closer to the surface of an atom than of another. The initial thought behind using a Voronoi decomposition is based on delimiting a region for each atom that is closer to this atom than any other one. The ambiguity of this definition relies on the definition of "closeness". In this regular Voronoi decompostion, we define the closeness using the distance between the center of mass of the different atoms, which is problematic for unequally distributed radii.

Unequal radii

The last definition of the Voronoi decomposition works only for equal-sized atoms because the closest region to an atom is also the closest to its center of mass, which does not apply to the complex atomic structures of nanoporous frameworks. To model the atomic radii $r_1, \ldots, r_n$ of the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ in the same box B, we can implement a so-called Apollonian Voronoi diagram.[219] For every $k \in \{1, \ldots, n\}$, the new subspaces A_k are defined as follows:

$$\mathrm{A}_k = \left\{ \boldsymbol{x} \in \mathrm{B} \mid \forall\, l \neq k,\ \|\boldsymbol{x} - \boldsymbol{x}_k\| - r_k \leq \|\boldsymbol{x} - \boldsymbol{x}_l\| - r_l \right\} \tag{3.2}$$

This new definition of the Voronoi diagram has been built on the intuitive property that stipulates that the subspace is the set of points closest to the surface of the sphere around the associated atom. It deals therefore with an unequal distribution of atomic radii, because it is now dependent on the radii. However, as we can see on the Figure 3.1, this first implementation has a very convenient definition, but the edges of the subspaces are curved, which makes it harder to use computationally.

For this reason a less intuitive implementation is more commonly used instead, which is called the radical Voronoi tesselation or power diagram or Laguerre-Voronoi diagram.[220] As we can see on the Figure 3.1, the subspaces obtained using this method are now convex polygons with straight edges instead of curved ones. The condition is way less intuitive because the condition does not rely on a simple definition. The subspaces V_k are now defined by the following condition:

$$\mathrm{V}_k = \left\{ \boldsymbol{x} \in \mathrm{B} \mid \forall\, l \neq k,\ \|\boldsymbol{x} - \boldsymbol{x}_k\|^2 - r_k^2 \leq \|\boldsymbol{x} - \boldsymbol{x}_l\|^2 - r_l^2 \right\} \tag{3.3}$$

In addition to the polyhedral form of the Voronoi cells, this new implementation presents interesting properties for porosity calculations in a framework of unequal spheres like in MOFs of zeolites.[221] First, the boundary between two overlapping spheres correspond simply to the intersection place between the spheres. Then, the boundary between non-overlapping spheres is always in the void space between the spheres. This can be simply proven by taking a point $\boldsymbol{x}$ in V_k and outside the sphere, we have $\|\boldsymbol{x} - \boldsymbol{x}_k\| \geq r_k$, which implies $\forall\, l \neq k$, $\|\boldsymbol{x} - \boldsymbol{x}_k\| \geq r_k$. The point $\boldsymbol{x}$ is therefore also not overlapping with any other atom, which means it is in the void space of the framework.

If we now consider a point $\boldsymbol{v}$ on a boundary between p Voronoi cells $\left\{\mathrm{V}_{i_1}, \ldots, \mathrm{V}_{i_p}\right\}$, this point would verify all the conditions $\|\boldsymbol{x} - \boldsymbol{x}_{i_1}\|^2 - r_{i_1}^2 = \cdots = \|\boldsymbol{x} - \boldsymbol{x}_{i_p}\|^2 - r_{i_p}^2 = \mathrm{C}$. We can know their minimum distance to the center of mass of every nearby atoms to test for possible overlapping. More specifically, in the Zeo++ software,[42] the Voronoi diagram is characterized by storing the minimum distance to the closest atoms and the index of the atoms for every vertices and edges (for edges that connect two different periodic images a periodic displacement vector is also stored). This information can be used to speed up the void fraction calculation, by skipping the volume calculations in the non-adsorbable Voronoi cells. It is also a fast way of determining the accessible and non-accessible surface areas and volumes.[42] Note that if the probe has a radius r_{probe}, then the sphere radii considered are $r_k = r_{\text{atom}} + r_{\text{probe}}$.

3.1.2 Implementation in a Screening

The use of the Voronoi decomposition of the pore space of materials for their geometric characterization has been widely employed in computational studies in the last decade,[222] in particular since it was made easily available as part of the Zeo++ software package.[223] Its use was extended recently to implement a novel sampling scheme, in a study proposing the ML-assisted screening of nanoporous materials for xenon/krypton separation. In this article, Simon et al.[14] relied on a Voronoi tessellation of the nanoporous materials and assigned the potential adsorption sites (i.e., the sampling points) at the nodes of this decomposition. The Voronoi tessellation identifies the vertices of polyhedra that correspond to the closest regions of each atom of the structure. These vertices (or *Voronoi nodes*) are the points equidistant to at least four atoms of the structure, and they can be associated with adsorption sites since they are positioned near the center of the pores.

The definition of accessibility defined by the Zeo++ software was used in a screening to find the best materials for Xe/Kr separation.[14] The interaction energies of xenon were calculated only on the accessible nodes as schemed on the Figure 3.2. The average of the energies at the accessible Voronoi nodes gives an approximation of the adsorption enthalpy. However, this sampling assumes that the nodes are close to the real, most favorable, adsorption sites. Or to put it differently, the adsorption sites need to be at the center of the pores, which is only true for structures with pore sizes close to adsorbate size. This newly defined adsorption energy descriptor was found to be one of the most influential descriptor for the ML model developed to predict the ambient-pressure selectivity.

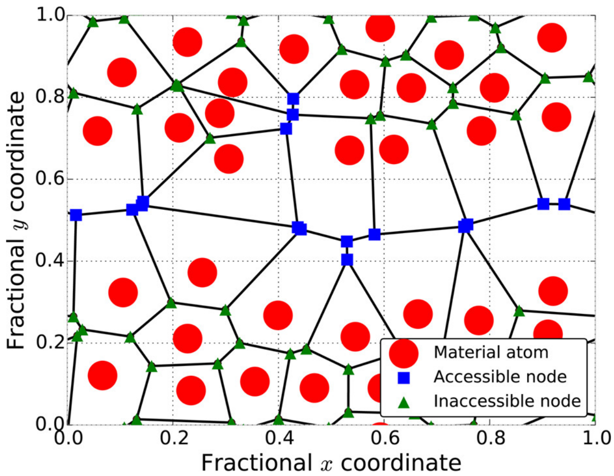


Figure 3.2: Voronoi network model of void space (2D caricature). The unit cell of a toy material is shown. Red circles represent atoms of the material; accessible and inaccessible Voronoi nodes are blue squares and green triangles, respectively. The black lines are the edges in the periodic Voronoi graph that model the void space. Taken from Ref. [14]

Starting from the initial idea of Voronoi sampling, we want to question the relevance of using a direct average of the interaction energies instead of a Boltzmann averaging in describing the adsorption enthalpy. To better understand the strengths and weaknesses of this methodology, we compared different ways of approximating adsorption enthalpies using the Voronoi sampling with more accurate infinite dilution and ambient-pressure xenon adsorption enthalpies using Widom insertions and GCMC for a 20:80 Xe/Kr mixture at 1 atm and 298 K.

3.1.3 Comparative study of the Voronoi sampling

We introduced in the previous chapter, the definition of the xenon adsorption enthalpy at infinite dilution (Widom insertion section 2.1.4) and at ambient pressure (GCMC sections 2.1.3 and 2.1.5). These methods can be considered accurate methods to calculate the adsorption enthalpy which has been proven to be strongly correlated to the logarithm of the selectivity in our previous study on the thermodynamic exploration of the xenon/krypton separation using a high-throughput screening.

Introduction of the main concepts

The Voronoi energy as initially conceptualized by Simon et al. is based on the average of the xenon interaction energies at the accessible Voronoi nodes. But because we will compare to thermodynamic simulations without blocking pockets, we will not consider the accessible Voronoi nodes but the adsorbable ones since it is closer to the simulation we want to approximate. For simplicity, we define the set of the adsorbable Voronoi nodes A as the Voronoi nodes with a negative energy value, we also apply a condition on the minimum distance (about the radius of a xenon 2 Å) to reduce the number of points to be calculated. This average on the adsorbable Voronoi nodes $E^{Xe}_{\text{voro-A}}$ can be written:

$$E^{Xe}_{\text{voro-A}} = \sum_{i \in A} E_i - RT \tag{3.4}$$

Another interesting energy descriptor could simply be the minimum of the interaction energies among the Voronoi nodes V with a minimum distance to the nearest atom is higher than 2 Å. This minimum Voronoi energy $E^{Xe}_{\text{voro-M}}$ can be written:

$$E^{Xe}_{\text{voro-M}} = \min_{i \in V} E_i \tag{3.5}$$

Finally, to get closer to the definition of the heat of adsorption defined in the previous chapter, we can build an energy descriptor using a Boltzmann averaging. This Boltzmann average of the xenon interaction energies at the Voronoi nodes V written $E^{Xe}_{\text{voro-B}}$ can be expressed as follows:

$$E^{Xe}_{\text{voro-B}} = \frac{\sum_{i \in V} E_i e^{-\beta E_i}}{\sum_{i \in V} e^{-\beta E_i}} - RT \tag{3.6}$$

Note that the −RT term is to make the expression comparable to the one of adsorption enthalpy.

We can intuitively say that the Boltzmann averaging being closer to the definition of the adsorption enthalpy it would be a better candidate as an energy descriptor, hence improving the screening methodology. To test these different methodologies, we will now compare the different energy descriptors to more accurate evaluation of the adsorption heat.

Low pressure comparison

The Widom insertion is typically used to calculate the infinite dilution adsorption properties such as the adsorption enthalpy, the Henry constant and the selectivity. The evaluation of the interaction energies of xenon at the different Voronoi nodes correspond to a low-pressure averaging and is more comparable to a Widom insertion method. It is however biased by the inhomogeneous sampling of the space, which can explain some discrepancies we could observe.

Note that in this chapter we will mainly use the standard pore size definition we find in the literature that is based on the atom radii defined by the Cambridge Crystallographic Data Centre (CCDC). The LCD that denotes to simplify the D_i

As we can see on the Figure 3.3, the average of the energies is not performing very well and is clearly less correlated to the adsorption enthalpy than the minimum interaction energy or the Boltzmann average of the interaction energies. This is because in a normal average, the high energy values have a much higher weight than in a Boltzmann average, which makes the average much more important than expected. The Voronoi average descriptor $E^{Xe}_{\text{voro-A}}$ is always higher than the infinite dilution adsorption enthalpy $\Delta_{\text{ads}}H_0^{\text{Xe}}$.

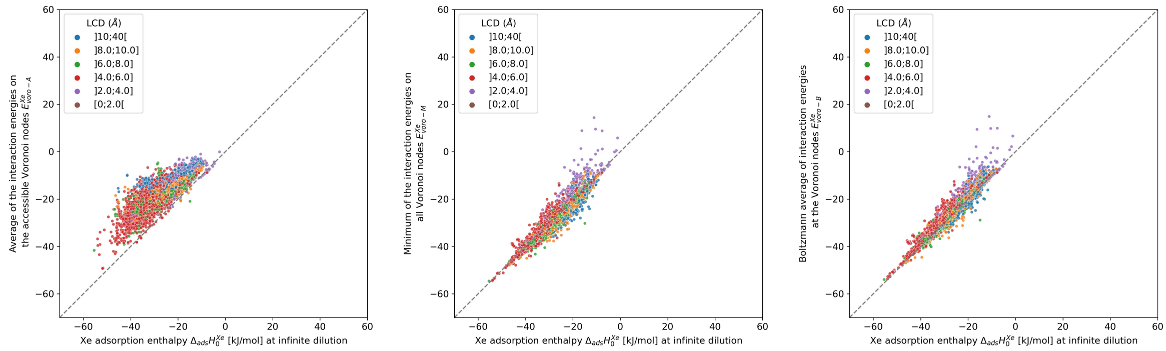


Figure 3.3: Scatterplots of the energy descriptors $E^{Xe}_{voro\text{-}A}$, $E^{Xe}_{voro\text{-}M}$ and $E^{Xe}_{voro\text{-}B}$ calculated by a Voronoi sampling compared to the enthalpies calculated by a 100k-step Widom insertion simulation of xenon in structures of CoRE MOF 2019. The points are labeled according to the largest cavity diameter (LCD_{CCDC}) belonging to one of the intervals.

The Pearson correlation coefficients corroborate our initial observation. The correlation coefficient between $E^{Xe}_{\text{voro-A}}$ and $\Delta_{\text{ads}}H_0^{\text{Xe}}$ is equal to 0.81, whereas for the minimum $E^{Xe}_{\text{voro-M}}$ it is 0.95 and for the Boltzmann average $E^{Xe}_{\text{voro-B}}$ it is 0.97. For this reason, to evaluate the relevance of a Voronoi energy sampling we should consider a Boltzmann average. As we have shown in the previous chapter, the selectivity is correlated to the difference of adsorption enthalpies of xenon and krypton. Better describing the enthalpy is a first step toward a better description of the selectivity. However, we only looked at the selectivity values at low pressure. What would happen for selectivity at higher pressure?

On the Figure 3.4, we can see that the selectivity drop between the low-pressure case and the ambient-pressure one also impact the enthalpy values of xenon. There is a reduction of the xenon affinity as the pressure goes up. Since the study of Simon et al. focused on the ambient-pressure selectivity prediction, we want to see if the energy descriptor they developed could however be used to describe the adsorption enthalpy at high pressure.

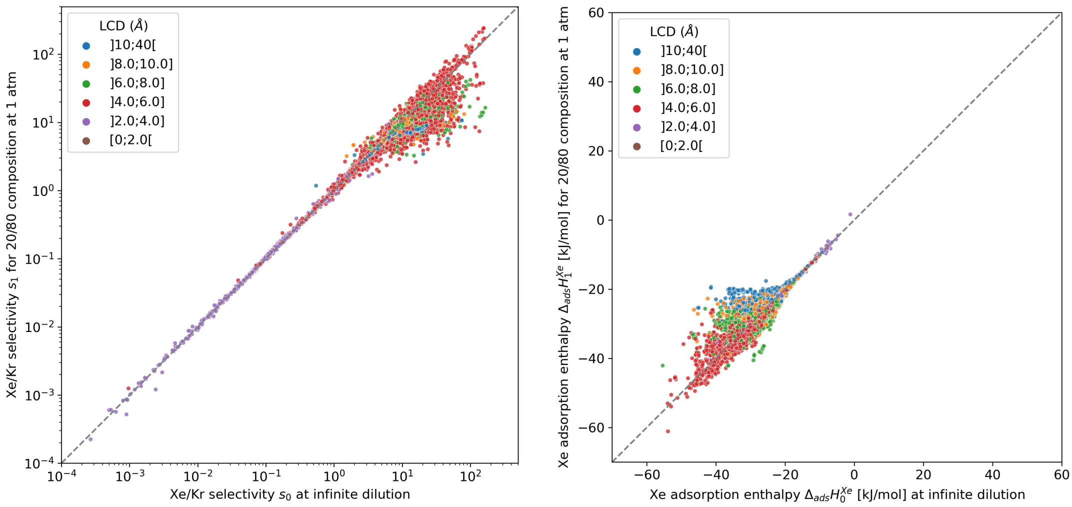


Figure 3.4: Transition

Ambient pressure

If we look at the Figure 3.5, it is not so clear which descriptor is better to describe the enthalpy at ambient-pressure. The correlation shown by the different scatterplots seem to be equally poor, and this could justify the use of a regular average rather than a Boltzmann average. The correlation coefficient for the average $\mathrm{E}^{Xe}_{\text{voro-A}}$ is now equal to 0.86, which is equivalent to the one of the minimum $\mathrm{E}^{Xe}_{\text{voro-M}}$ and slightly lower than the 0.87 for the Boltzmann average $\mathrm{E}^{Xe}_{\text{voro-B}}$.

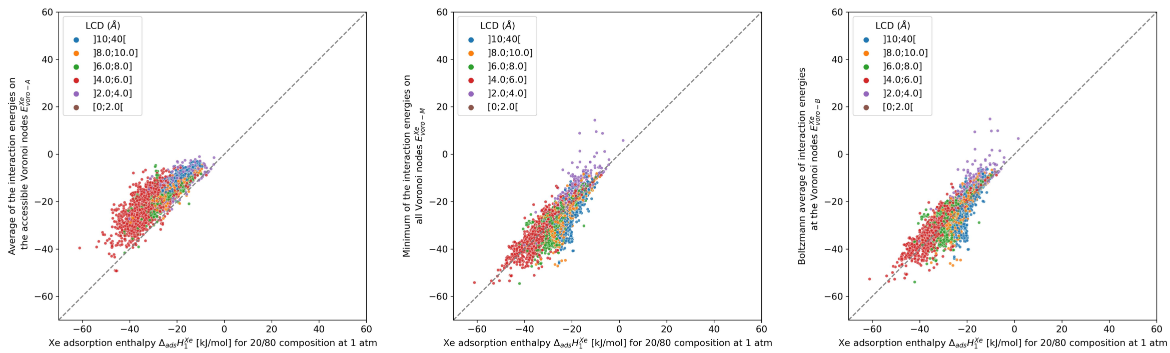


Figure 3.5: Scatterplots of the energy descriptors $\mathrm{E}^{Xe}_{voro\text{-}A}$, $\mathrm{E}^{Xe}_{voro\text{-}M}$ and $\mathrm{E}^{Xe}_{voro\text{-}B}$ calculated by a Voronoi sampling compared to the enthalpies calculated by a 100k-step GCMC simulation of xenon in structures of CoRE MOF 2019. The points are labeled according to the largest cavity diameter (LCD_{CCDC} or D_i) belonging to one of the intervals.

At higher pressure, the adsorption enthalpy has higher values, which degrades the correlation with the Boltzmann average and the minimum of the interaction energies evaluated at the Voronoi nodes. For the averaging, the over-evaluation of the energy values make it closer to the values at higher pressure. Following this idea, we later come up with another idea of averaging with bigger weights on the higher values, a Boltzmann average with a higher temperature value that will be tested in the next chapter.

3.1.4 Performance of a Voronoi Energy Sampling

We will now focus on some performance metrics associated to the Boltzmann average at the Voronoi nodes and compare it to our reference sampling, the Widom insertion with 100,000 cycles. The right plot of the Figure 3.3 compares the enthalpy computed in the Voronoi sampling with the reference adsorption enthalpy (ground truth) — showing at the same time the largest cavity diameter for each porous framework. The correlation between the values of enthalpy is very good only for a restricted number of structures with enthalpy around −50 kJ mol^{-1}. For structures with higher enthalpy, the correlation starts to degrade, and becomes very poor for small-pore structures. For the points in purple, the largest cavity diameter is lower than the kinetic diameter of a xenon, the sampling of the Voronoi nodes is clearly insufficient. In addition, the accuracy loss on the other points (larger pores) can be explained by the fact that the pores are slightly bigger and the center of the pore is not a good approximation of adsorption site position anymore: the adsorption sites are actually closer to the pore surface than to the center of the pore. This conclusion is what prompted us to propose a new sampling scheme based on the molecular surface of the pore space, which we will detail in the next sections.

The root mean square error (RMSE) and the mean absolute error (MAE) for Voronoi sampling are respectively 6.78 kJ mol^{-1} and 2.01 kJ mol^{-1}, if we consider all structures in our set, which seems too high to be useful for screening purposes. However, the non-porous materials would be screened out *a priori* in any high-throughput workflow, as they would not be of interest. We can only consider the structures with large enough cavities, higher than 3.7 Å (a bit lower than 3.96 Å Xe kinetic diameter). Thereby, the RMSE and MAE drop respectively to 2.11 kJ mol^{-1} and 1.55 kJ mol^{-1}, which can be considered acceptable for a quick estimation of the guest–host affinity, but not for accurate adsorption enthalpy calculation.

This is reinforced by the very low computational cost of the method. The Voronoi tessellation done by the Zeo++ software is extremely quick and can output the positions of the Voronoi nodes in 0.28 s (measured as an average over all the structures of the CoRE MOF 2019 database), on a typical workstation (a single Intel Xeon Platinum 8168 core at 2.7 GHz). While a simple Python for the energy calculation took around 27 s per structure, we benchmarked that a C++ optimized implementation can perform the Voronoi sampling in around 0.4 s. We only need to remember that this method takes a few hundred milliseconds per structure, while a Widom insertion needs approximately hundreds of seconds per structure. A Voronoi sampling is therefore 2 to 3 orders of magnitude quicker than a full sampling of the pore space.

This preliminary study identified a fast method for adsorption enthalpy calculation that can be widely used in screening procedures, but has limited accuracy for quantitative prediction — this sampling technique assumes that the nodes are close to the real, most favorable, adsorption sites, which is not always true. Or to put it differently, the adsorption sites need to be at the center of the pores, which is only true for structures with pore sizes close to adsorbate size. It raised important questions on the importance of selecting sampling points within the pore space of materials, and we wanted to develop an intermediate technique that is both fast and accurate for the prediction of adsorption enthalpy. For this purpose, we developed and optimized a new sampling technique that focuses the sampling on the surface of the material, which is expected to make up for the main flaws of the Voronoi sampling.

3.2 Rapid Adsorption Enthalpy Surface Sampling (RAESS)

In this section, we describe the development of our surface sampling algorithm, with the goal of being more accurate than Voronoi sampling and faster than Widom insertion. Our initial idea is based on a series of theoretical considerations: (i) the strong adsorption sites are near the surface of the material; (ii) by changing the problem from 3D to 2D sampling we can reduce the complexity; and (iii) the algorithm can scale with the number of unique atoms in the structure (and not with the size of the unit cell), which is efficient because many porous frameworks have high symmetry. The first consideration ensures that this method will be more accurate than a Voronoi sampling, and the last two made us think that a well-optimized code would be fast. To confirm these hypotheses, we will analyze both the accuracy and the speed of this new algorithm and compare them to existing methods.

3.2.1 Initial Implementation

We present here our initial implementation of the surface sampling algorithm, and its basic principles. This first implementation is a relatively basic one and already performs well compared to the other methods. In the next sections, we refine it with two additional features that will improve its accuracy and its speed.

This initial implementation speeds up the calculation of adsorption enthalpy in nanoporous materials by sampling interaction energies only near the surface. It is illustrated in Figure 3.6. For this purpose, a loop over all unique atoms (as defined by crystalline symmetry) is performed. And for each atom, a sphere around its position is sampled using a uniform distribution around it, these points will be called sampling points, and we can change the number of sampling points. The default radius chosen for the sampling spheres is the distance $r_{\text{min}} = 2^{1/6}\sigma_{ij}$ to the minimum of the LJ potential between atoms of type i (belonging to the framework) and j (the guest), corresponding to the strongest possible pair interaction (although the neighboring atoms will of course have an influence). After calculating the interaction energy at each of the sampled points, a Boltzmann average of these energies corresponds to a biased adsorption enthalpy, as described by the equation 2.21.

In order to validate the accuracy of the approximation made using this sampling, we applied this algorithm with 300,000 sampling points per unique atom. The results are illustrated by the Figure 3.12. There is a good numerical agreement with the reference calculations, the RMSE and MAE are only around 0.90 kJ mol^{-1} and 0.66 kJ mol^{-1} considering all the structures from the database. Moreover, there is no noticeable difference of RMSE when considering the structures with a pore size above 3.7 Å (as determined by the LCD_{CCDC}). Unlike Voronoi sampling, this method gives a consistent accuracy across all the structures of the database with a lower error. The fact that the RMSE error is below 1 kJ mol^{-1} is quite promising, and validates our intuition that this new sampling technique can be an intermediate between to the two previous methods (Voronoi and Widom).

After proving the good accuracy of the method, we are now exploring the computation time required. We see on Figure 3.8 that the method reaches an RMSE below 1.0 kJ mol^{-1} very quickly for an average CPU time of 1.2 s, corresponding to 2,000 sampling points per atom. This is far less than the 150 s required for a Widom insertion to be near its plateau value (converges

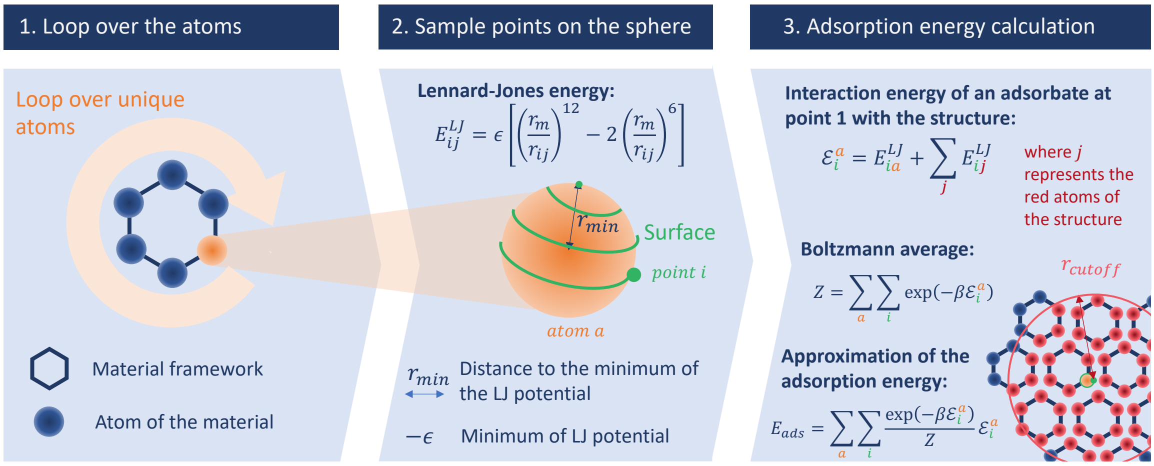


Figure 3.6: Schematic description of our surface sampling based on the three main steps of the algorithm: the loop over the unique atoms, the spiral sampling around each atom, and the energy averaging. The adsorbate is represented by the point i and is moved across all the points around the unique atoms of the structure.

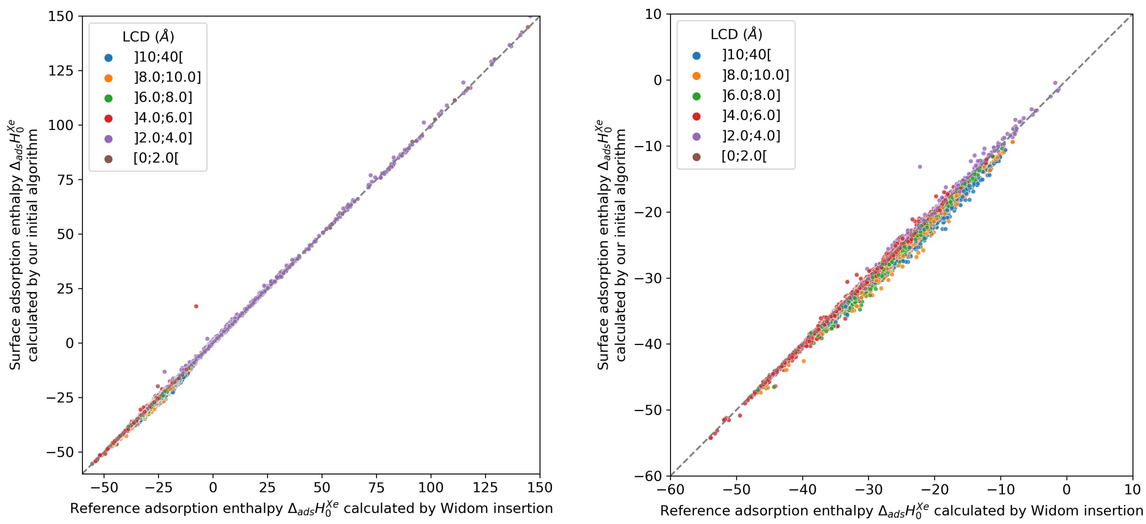


Figure 3.7: Scatterplots of the xenon surface adsorption enthalpy calculated by an initial implementation of the RAESS algorithm as a function of the xenon adsorption enthalpy calculated by a 100k-step Widom insertion simulation using two value windows. The second plot zooms on the negative values corresponding to the most selective materials.

to zero), for an RMSE of 0.10 kJ mol^{-1} with 12,000 cycles. Moreover, the Widom insertion needs around 14 s to reach a similar RMSE of 1.0 kJ mol^{-1}, which is still slower than the surface sampling. We can conclude that this initial implementation of the surface sampling is faster than a standard Widom insertion, with a good accuracy.

These results on the convergence speed and the limit values of the error can be simply rationalized by the nature of each sampling. In a surface sampling, the sampled points are biased toward the most attractive points of adsorption for the xenon, which explains the fact that the values converges very quickly because the most influential terms of the Boltzmann average are quickly gathered. However, in a Widom insertion every point of the space have equal chances of being sampled, which is extremely close to the definition of the enthalpy but requires much more time to randomly sample a very attractive adsorption site. The surface sampling by its biased nature however will inherently be less accurate since not all points are considered equally and sometimes the most optimal adsorption site could be missed, because it could in some cases be further from the sampled surface.

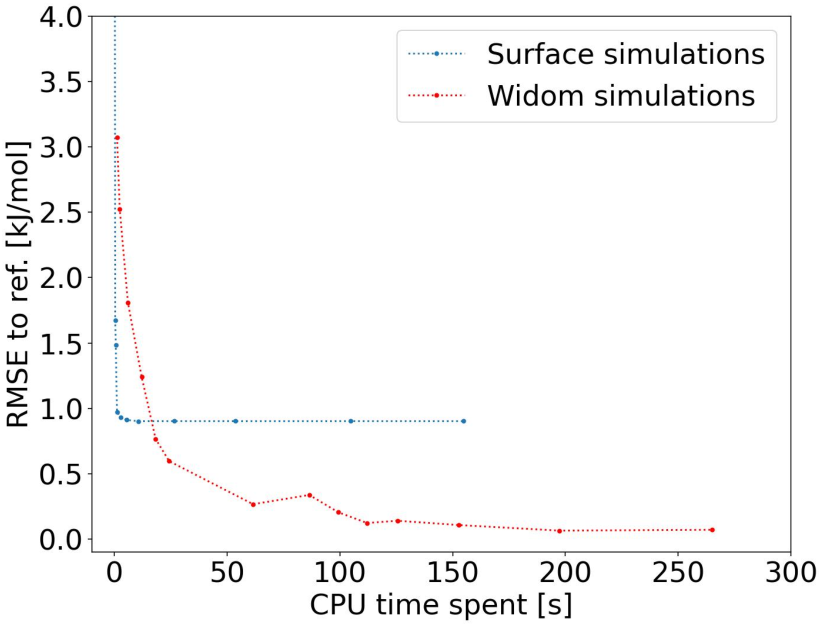


Figure 3.8: Convergence plot of the RMSE on the adsorption enthalpy for our algorithm (blue) compared to a 100k-step Widom insertion simulation (red) for xenon adsorption in all structures of the CoRE MOF 2019 database.

However, this initial implementation of the method is slower than a Voronoi sampling that only needs to sample around 1,600 points on average, instead of 13,000 sampled points on average (if we multiply by the average number of unique atoms). The sampling part would take approximately 0.15 s and the Voronoi nodes generation 0.28 s, so our surface sampling algorithm remains 2 to 3 times slower (implemented in an identical compiled language, in this case C++). In order to improve the accuracy and performance, we have further tweaked the surface sampling method, adjusting the size of the sampling sphere and adopting a fast rejection criterion. The rejection of high-energy points with little contribution to the final enthalpy value can reduce the simulation time, whereas the size of the sampling sphere can improve the accuracy. The initially chosen sphere size is only taking account of the interaction with the closest atom, we therefore chose to set it at the minimum of Lennard-Jones potential. However, the interaction with the neighboring atoms can further stabilize the adsorbate, sampling further from this minimum could in consequence increase the accuracy of our surface sampling method.

3.2.2 Performance improvement of the algorithm

Size of the sampling sphere

The validity of the initial algorithm is based on the assumption that the adsorption site is at the minimum of the Lennard-Jones potential. It will only perform well if the closest atom contributes to almost all the interaction, but in real frameworks other neighboring atoms contribute to the host/guest interaction as well. We have found that in vast majority of materials, the adsorption sites are located farther apart compared to the LJ potential minimum, in order to maximize the contribution of all atoms — and because of the dissymmetry of the interaction potential well. In order to see if this could be introduced in our algorithm, we implemented a parameter λ, and the sampling sphere radius is now defined by $R_\lambda = \lambda\sigma$, where σ is the distance at which the LJ potential is zero. If $\lambda = 2^{1/6}$, we fall back to our initial definition of the sampling sphere, and the adsorbent is at the minimum of the LJ potential of the atom. If $\lambda = 1$, the sampling sphere is at the zero of the LJ potential, and by increasing this parameter, we can check if our intuition was right.

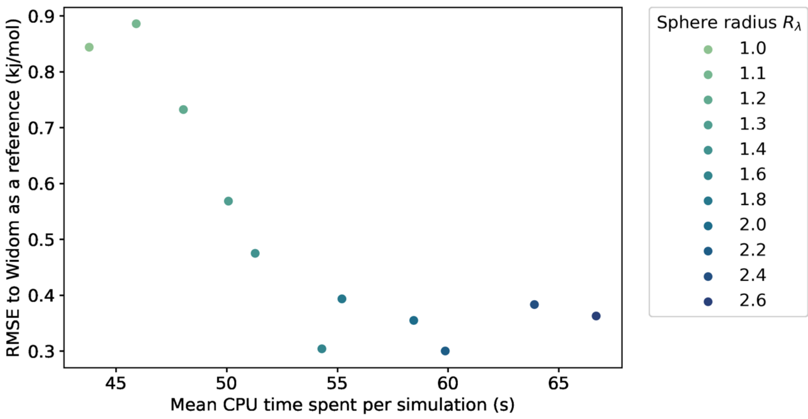


Figure 3.9: Influence of the sampling sphere radius R_λ *on the average CPU time required for a simulation of 100k sampling points and the RMSE, compared to the reference adsorption enthalpy. The averaging is done only on the structures with a largest cavity diameter (LCD_{CCDC}) higher than* 3.7 Å.

Because we have no physical model that would predict the optimal value of the sampling sphere, we followed a statistical approach. We studied the influence of the λ parameter on both the accuracy and the computation time, and the results are represented on Figure 3.9. The RMSE turns out to be relatively high around 0.90 kJ mol^{-1} for radius sphere lower than the r_{min}, it then decreases for larger values of radius to reach a plateau around 0.35 kJ mol^{-1}. We confirm that by increasing the sampling sphere radius we can improve the accuracy of our algorithm, and find that for values of λ higher than 1.6, the accuracy is stabilized. We also find that increasing the sphere radius negatively impacts the computational efficiency, since it increases the number of neighbors considered in the energy calculation.

By choosing an optimal sampling sphere, we can more than halve the error, while increasing the computation time by around 20 percent, when comparing the case $\lambda = 1.6$ with $\lambda = 1.1$ (close to r_{min}). In most cases, it will be an acceptable trade-off. However, in a case where the computation time is crucial, like in a rapid screening, the optimal choice might not be to increase the sampling sphere at $\lambda = 1.6$ but to have it lower at $\lambda = 1.4$ or $\lambda = 1.2$, and have an RMSE around 0.5 kJ mol^{-1} — still quite acceptable. The new scale parameter introduced in this section can therefore be tweaked to serve the users' purpose, whether it is to focus on the accuracy or to optimize the computation speed. If one wants to use it on a completely different database in very different conditions, then one can either choose a default value that works fine (*e.g.* $\lambda = 1.4$) or one can optimize the parameter on a small diverse sample of the unseen data.

Rejection condition

As shown above, our algorithm has better accuracy than Voronoi sampling, but its initial implementation was several times slower, which could be unsuitable for screening applications in high-throughput workflows, where the number of structures to be screened can reach one million or more. To reduce the computational expense, we thought of rejecting the points with little contribution to the final enthalpy, i.e., the largely positive interaction energies that would vanish in the exponential of the Boltzmann average.

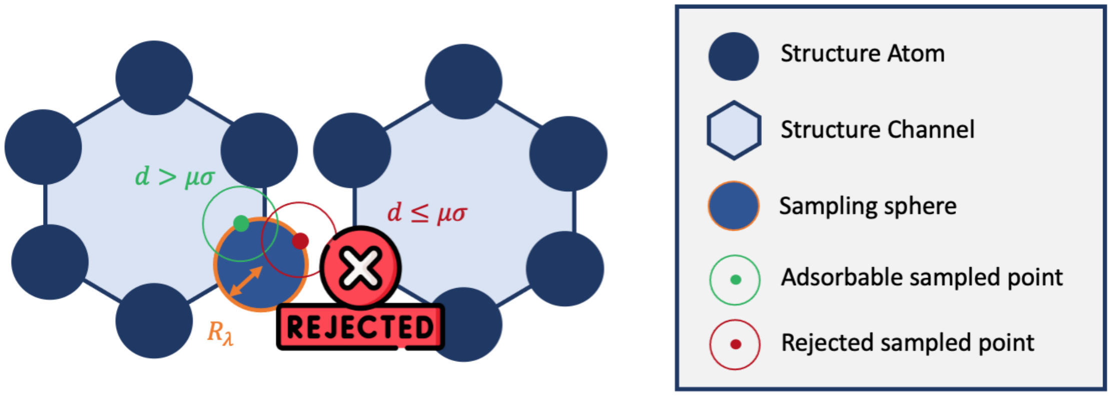


Figure 3.10: Simplified representation of the principle of rejection condition and the concept of sampling sphere inside 2D channels of a nanoporous material.

Inspired by typical methods for accessible surface calculation, we implemented a hard sphere rejection condition based on the distance to neighbors. If the adsorbate is too close to another atom of the structure, the sampling point is rejected, i.e., its energy is not calculated (or considered to be infinite). We based this distance threshold on the σ_{ij} parameter of the Lennard-Jones potential. To determine the optimal threshold, we introduced a factor μ with real values between 0 and 1, that changes the size of the hard sphere rejection condition. If the guest–host distance is lower than $d_\mu = \mu \times \sigma$, then the point is rejected. If $\mu = 0$, then there is no rejection condition. And if $\mu = 1$, we reject all points with a positive energy interaction to at least one atom of the structure. This condition could be a bit strong and points with non-negligible contribution would end up rejected. This rejection condition is schematically represented on Figure 3.10.

This rejection condition is expected to speed up the calculation, since the energy calculation is avoided for the rejected sampling points. The energy calculation accounts for the largest portion of the CPU time spent in the surface sampling. For the structure `KAXQIL`[224], the Lennard-Jones potential calculation represents up to 90% of the calculation time for 100,000 sampling points per sphere (with the initial algorithm). The higher the factor μ, the more rejections there would be. But, if too many points are rejected, the accuracy would drop. Here again, we used a statistical analysis to determine the optimal value of μ, making our sampling faster without compromising the accuracy of the enthalpy calculation. The results are displayed on Figure 3.11.

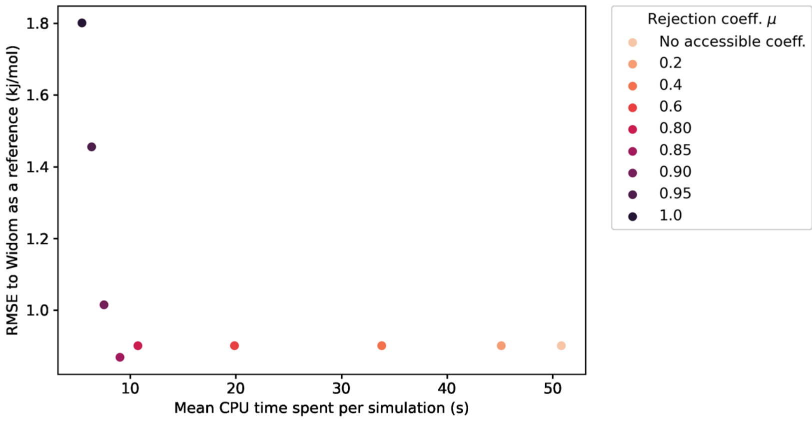


Figure 3.11: Influence of the rejection coefficient μ on the average CPU time required for a simulation of 100k sampling points and the RMSE compared to the reference adsorption enthalpy. The averaging is done only on the structures with a largest cavity diameter (LCD_{CCDC}) superior to 3.7 Å.

The values of RMSE and time on Figure 3.11 are averaged only on the most interesting structures for xenon adsorption ($LCD_{CCDC} \geq 3.7$ Å). For $\mu \leq 0.85$, increasing the value of μ improves the speed of the calculation without changing the RMSE.[1] For high values of μ, the rejection condition is too strong and we reject points with non-negligible contribution to the overall enthalpy. The RMSE increases as a consequence. If we want to keep the accuracy unchanged, the optimal value is therefore $\mu \simeq 0.85$, because it give the lowest computation time with a similar RMSE. We note that it would be possible, in specific cases, to explore higher values of μ that trade a bit more accuracy in exchange for further speed gains.

For the simulations considered in Figure 3.11, the use of a rejection condition $\mu = 0.85$ makes the simulation four times faster than the standard algorithm. As we will see in the next section, the combination of optimal values for the λ and μ parameters generates an algorithm with very interesting performance compared to Voronoi sampling or Widom insertion.

3.2.3 Final surface sampling implementation

Performance comparison

For the calculation of adsorption enthalpy, our proposed surface sampling method is a good compromise between the accuracy of Widom insertion (full sampling of the porous space) and the speed of a less accurate method such as Voronoi sampling. The performance of our algorithm, including the two new features (sampling sphere scaling and rejection criterion) is illustrated in Figure 3.13, where we can see the improvement brought by each feature and how it compares to reference simulations. All CPU times are calculated using the smallest possible number of sampling points so that the respective algorithms reach convergence. With the implementation of a rejection condition, we find that surface sampling is even quicker than Voronoi sampling. Moreover, the increase of the size of the sampling sphere makes the surface sampling much more accurate, reaching an RMSE of 0.33 kJ mol^{-1} and an MAE of 0.21 kJ mol^{-1}.

[1] In fact, what we observe is a deterioration of the accuracy for structures with small pores because the probability of rejection in a confined space is really high and all sampled points end up rejected. But these points are not considered, if we apply the condition on the cavity size ($LCD_{CCDC} \geq 3.7$ Å).

The ideal set of parameters, determined for porous materials from the CoRE MOF 2019 database, is ($\lambda = 1.6, \mu = 0.85$) in order to combine the lowest error and smallest computational cost. By combining both of these new features to the algorithm, we have a final surface sampling method with an RMSE of 0.33 kJ mol^{-1} and an average computation time of 0.34 s per structure. According to the data represented in Figure 3.13, it is about 6 times more accurate and 26% faster than Voronoi sampling, and it is also about 430 times faster than a Widom insertion with 12k cycles.

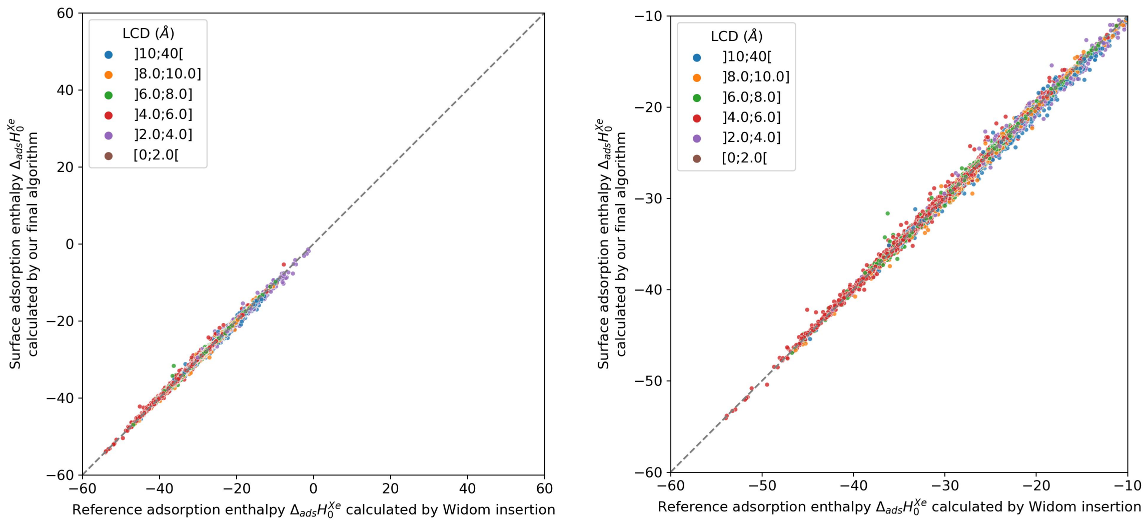


Figure 3.12: Scatterplots of the xenon surface adsorption enthalpy calculated by the final RAESS algorithm ($\lambda = 1.6$ and $\mu = 0.85$) as a function of the xenon adsorption enthalpy calculated by a 100k-step Widom insertion simulation using two value windows, in structures of CoRE MOF 2019 with $LCD_{CCDC} \geq 3.7$ Å at 298 K. The second plot zooms on the negative values corresponding to the most selective materials.

Finally, we suggest that the values of the parameters optimized in this work might need adjustment when applied to other adsorption systems. The optimal μ parameter depends on the size of the adsorbent, and it should be tweaked differently when considering another adsorbent. For instance, the set of structures used for the optimization of μ depends on the size of their cavities, and the 3.7 Å threshold chosen here would need to be changed according to the kinetic diameter of the adsorbate. Furthermore, as aforementioned in the section on the rejection condition, it is possible to trade off a bit of accuracy for faster simulations especially in high-throughput screenings where speed is extremely important. Similarly, in the case of xenon, the cost of increasing the sphere size is around 10 to 20%. On very large databases, one could consider that this increase on the required computational time is not worth the accuracy improvement, and one could decide to keep a smaller sampling sphere. If this method is transposed to different molecular systems, its parameters should be tested on the specific database and adsorbate of interest.

Calculation of Henry constant and surface area

The main goal of our sampling algorithm is to calculate adsorption enthalpy in the zero-loading limit. But the method can also calculate at the same time the Henry constant and surface area of the materials, without significant additional computational cost. The Henry constant is a

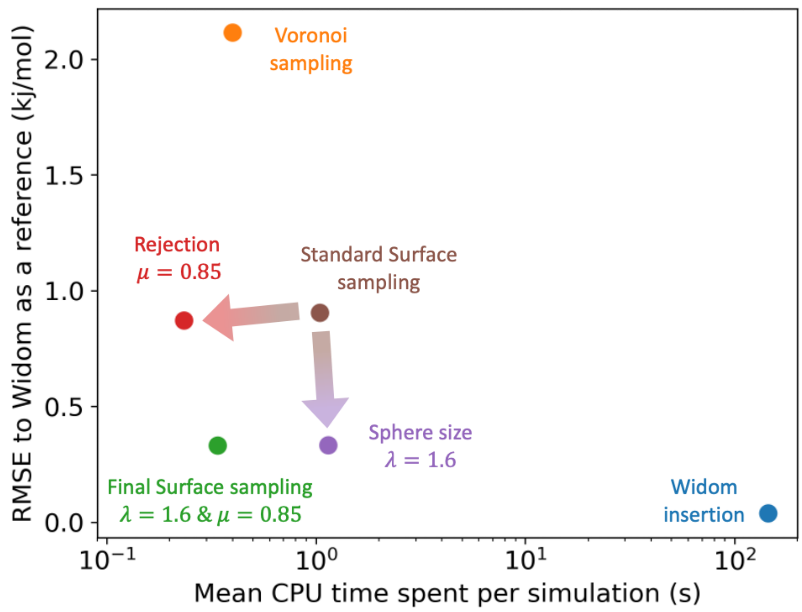


Figure 3.13: Comparison of the RMSE to the reference Widom insertion and the average computation time for different types of enthalpy calculation methods. The surface sampling calculation were all done with 2k sampling points on each sphere and the Widom simulations were done using 12k cycles. These values correspond to the value at the convergence identified using Figure 3.8.

key metric for assessing the affinity of an adsorbate to a nanoporous structure. The Xe/Kr gas selectivity at low pressure is defined as a ratio of Henry constants of Xe and Kr. This important property can be calculated using Equation 2.15 in a Widom insertion calculation. Instead of using the interaction energies at the Widom inserted points, we can now use the surface sampled points to get an approximate value for the Henry constant.

Using the optimized set of parameters for surface sampling, we assessed the performance of our algorithm on the values of Henry constant by comparing them to ground truth obtained by 100,000 cycles of Widom insertion. Since the Henry constant corresponds to the exponential of an adsorption free energy, and we are more interested in the precision on the free energy, we are using a log-scale evaluation metric. For surface sampling, the log-RMSE of K_H is equal to 0.2, which means that the order of magnitude of the values are well predicted as we can see on the Figure 3.15. If we consider the derived free energy $\Delta F_{ads} = -RT \log(\rho_f RTK_H)$, the RMSE is of the order of 1.1 kJ mol^{-1} reached in about 1 s (Figure 3.15). Whereas for Widom insertion, this level of error is also reached in a similar amount of time and 0.1 kJ mol^{-1} of RMSE is reached in about 86 s (Figure 3.15). For free energy calculation, surface sampling is still 86 times faster to converge. If consider that the main target is the adsorption enthalpy, the Henry constant is calculated with little additional computational cost and with reasonable accuracy: we get two thermodynamic properties of interest for the price of one.

The same goes for the determination of the surface area. We can adapt our algorithm to count the number of points of the sampling spheres that have a negative energy. These represent the points were a guest molecule can favorably interact, therefore when dividing it by the number of sampled points, we obtain a proportion of adsorbable area of the sphere. Summing

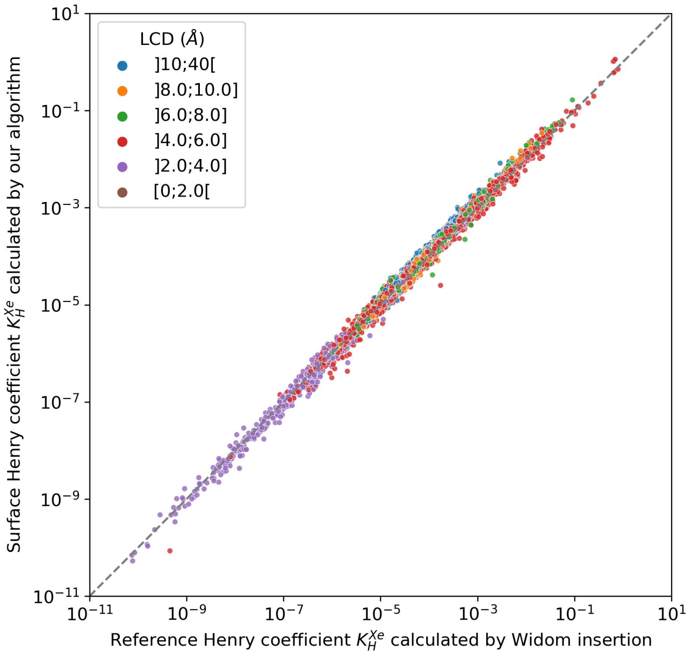


Figure 3.14: Scatterplots of the xenon Henry constants calculated by the RAESS algorithm compared to the ones calculated by a 100k-step Widom insertion simulation using two value windows.

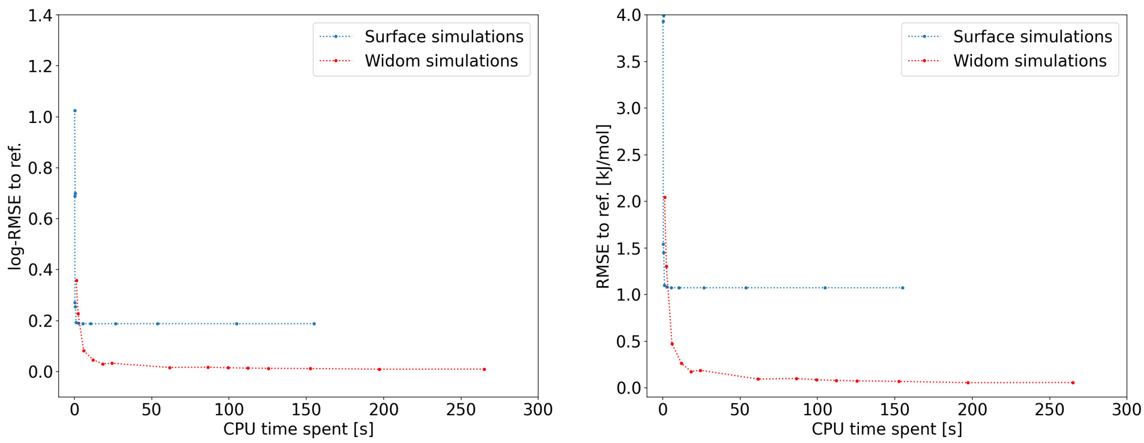


Figure 3.15: Left: convergence plot of the log-RMSE on the xenon Henry constants for both the surface sampling and the Widom insertion. Right: convergence plot of the RMSE on the xenon adsorption Gibbs free energy for the final implementation of the surface sampling and the Widom insertion.

this over all atoms, we obtain a total surface area. This implementation is summed up in Equation 3.7:

$$\mathrm{SA} = \frac{1}{\mathrm{V}} \sum_{a \in \mathrm{cell}} \frac{\mathrm{N}_{\mathrm{accessible}}(a)}{\mathrm{N}_{\mathrm{total}}} 4\pi r(a)^2 \tag{3.7}$$

where V volume of the cell a atoms of the cell; $\mathrm{N}_{accessible}(a)$ accessible points around the atom a; N_{total} sampling points; $r(a)$ radius of the sampling sphere around the atom a. When we set $\lambda = 1$, we are sampling spheres that have a radius σ, and it is equivalent as considering hard spheres all defined by σ (convention used by RASPA2 to calculate surface areas). If we compare simulation with $\lambda = 1$, we obtain surface areas that are very close to the one obtained by RASPA2 (see Figure 3.16 in SI). However, when we consider $\lambda = 1.6$, we lose the perfect accordance previously obtained and the points weakly correlated in log-scale (see Figure 3.16 in SI). The difference can be explained by the fact that the sphere size is larger, but the proportion of adsorbable points also changes. The relationship between these two adsorption surface areas is not trivial at all. Since the calculation of surface areas is quite cheap, this implementation would not be very useful, except for having a rough idea of the surface area.

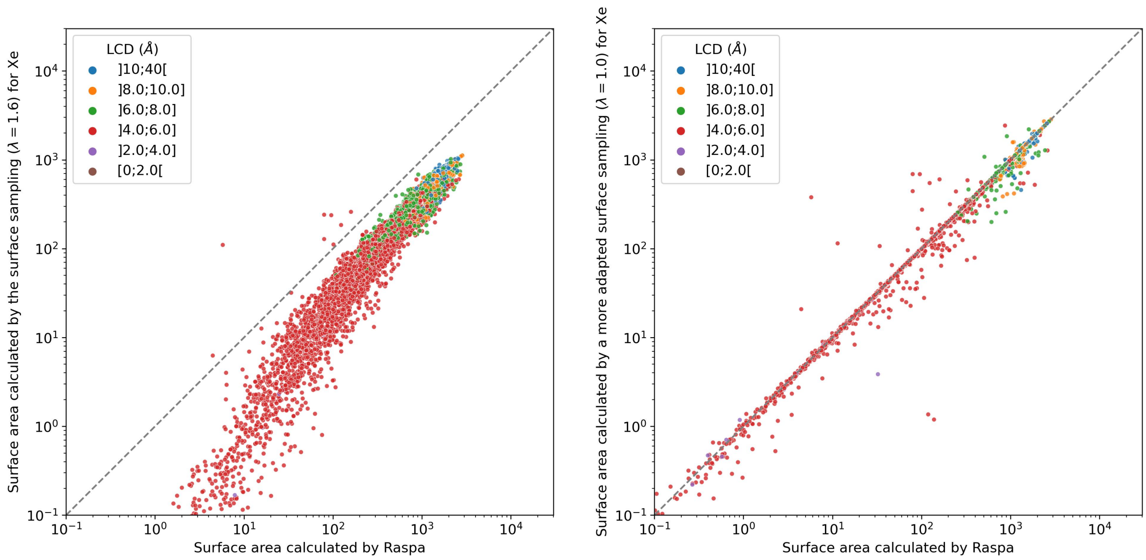


Figure 3.16: Scatterplots of the surface areas calculated by our algorithm with two different parametrizations compared to the surface area given by a Raspa surface area calculation. The left plot corresponds to the surface sampling described in the section 3.2.3 with $\lambda = 1.6$ and $\mu = 0.85$, while the right plot uses a sampling sphere near σ with $\lambda = 1.0$. The second parametrization is much closer to what a Raspa sampling based on the σ parameter of a LJ potential does, hence explaining the much better accordance.

3.2.4 Test in different configurations

After introducing the performance of our surface energy sampling algorithm for xenon and on specific materials from CoRE MOF 2019 at 298 K, we will explore other conditions to test the transferability of the methodology. First, we will use the algorithm to estimate the xenon/krypton selectivity at infinite dilution by comparing to the standard Widom insertion. Then, we want to compare the influence of the temperature on the performance since it most likely would be much less ideal since the Boltzmann weights are less concentrated on the less attractive points. Finally, we tested our algorithm on databases of different materials.

Selectivity Calculations

The selectivity value is the most important metric in evaluating the Xe/Kr separation performance of a nanoporous material. We want to see if a surface sampling technique can accurately evaluate this metric while being limited by all the approximations inherent to the technique.

A few precautions should be considered before blindly using the algorithm for selectivity prediction. When investigating the calculation of the selectivity, we noted that the rejection condition on xenon can be high since we are interested in the most favorable materials for xenon adsorption. But for krypton, we want to accurately describe very low Henry constants, because a selective material would also be a material unfavorable to krypton. For these reasons, the μ parameter needs to be chosen wisely and needs to be low enough to have accurate Kr Henry constant and then selectivity values.

As we can see on the Table 3.1, the error on the selectivity highly depends on the μ value that excludes the points at $\mu\sigma$ of a framework atom center. Logically, the lower this parameter the higher the sampled energy values can be in the Boltzmann averaging. Another reason is that when dividing by small values any small error on the values can be amplified in the quotient, and this effect can be reduced as we increase the number of points actually sampled.

rejection parameter μ	log10-RMSE to widom 100k	log10-MAE to widom 100k
0.85	0.107	0.077
0.50	0.0635	0.0402
0.20	0.0637	0.0403

Table 3.1: Influence of the rejection condition in the krypton surface simulation on the accuracy of the Xe/Kr selectivity calculation. The lower the μ parameter the more accurate the simulations are for the final selectivity calculation.

According to the quick study here, the optimal value is $\mu = 0.5$ since it gives the best accuracy for a minimal amount of time. This value will be used for krypton to make a comprehensive study of the performance on the Xe/Kr selectivity for materials from CoRE MOF 2019. To sum-up, in the following study, we will use the RAESS algorithm with $\lambda = 1.6$ and $\mu = 0.85$ for xenon and $\lambda = 1.6$ and $\mu = 0.5$ for krypton.

The selectivity can be compared directly using a log-scale plot and log-scale metrics. If we apply the log10 to the selectivities, we obtain RMSE of 0.064 and MAE of 0.04. This means that we have an error of about 0.06 when we compare orders of magnitude of the selectivity. For example, if a selectivity is predicted to be $s = 10^{-7}$, then s would be in the interval $[10^{-7.06}, 10^{-6.94}]$.

To be able to give a thermodynamic interpretation, we can use the exchange Gibbs free energy associated $\Delta_{\text{exch}}\text{G}_0^{\text{Xe/Kr}}$ to this selectivity defined in the previous chapter (equation 2.25). Using this exchange Gibbs free energy, we can assess much more easily the performance of the approach. The RMSE is about 0.36 kJ mol^{-1}. We cannot compare it to the adsorption enthalpy errors, since the ranges and interpretation are very different. Here, the selective materials have a negative $\Delta_{\text{exch}}\text{G}_0^{\text{Xe/Kr}}$, and it goes to a maximum value of about −12.7 kJ mol^{-1}. The relative error is of course higher on the Gibbs free energy. This is due to a higher uncertainty on the Henry constant and to the denominator term brought by the krypton.

To know how well the RAESS algorithm would work in real situations, we tried to compare the top 100 most selective materials given by RAESS and a Widom simulation (RASPA2). We found that 83 structures of the top 100 given by RAESS are in the top 100 given by Widom insertion. As the correlation is not perfect, it is inevitable that there is a change in the order of the top 100 given by these two methods. This number of 83% proves that the difference is quite narrow. If we enlarge to the top 150 of the Widom simulation, 94 are present in the top 100 of

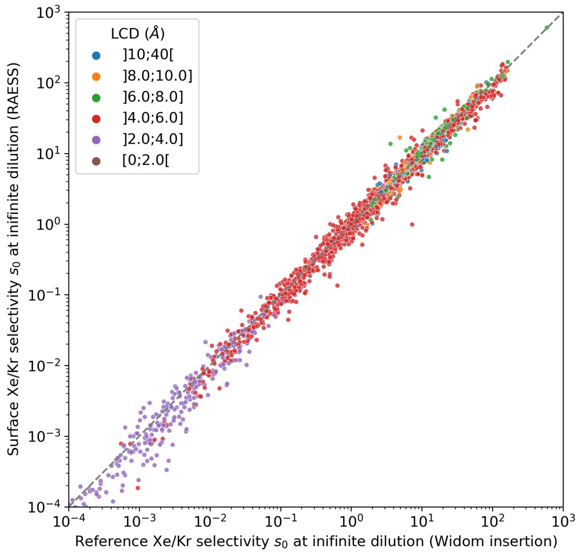


Figure 3.17: Scatterplot comparison of the Xe/Kr selectivity calculated by RAESS algorithm and the one calculated by the Widom insertion (in log scale) and labeled by the cavity size.

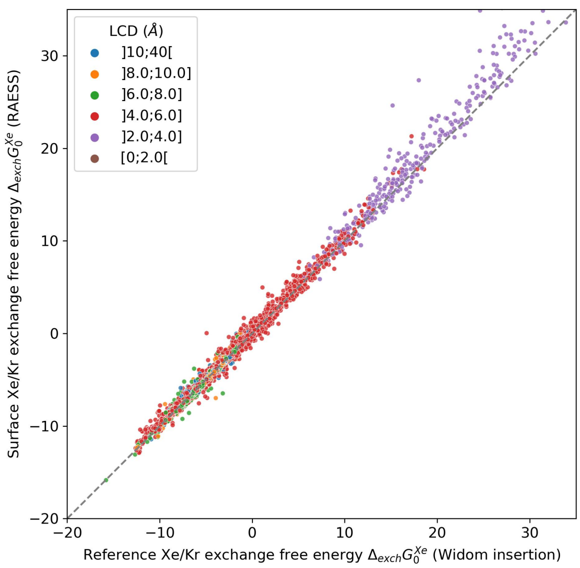


Figure 3.18: Scatterplot comparison of the exchange Gibbs free energy $\Delta_{exch}G_0^{Xe/Kr}$ calculated by the Widom insertion compared to the final implementation of RAESS (RMSE=0.36 kJ mol^{-1} and MAE=0.23 kJ mol^{-1}).

the surface simulation. We can therefore say that a vast majority of the best candidates given by the Widom insertion simulation are found by the RAESS algorithm.

A Higher Temperature

The RAESS method relies on the higher weight of the strong sites close to the surface of the pores. If we increase the temperature, the less attractive sites would play an increasing role and the accuracy of the method would drop. To grasp this limitation of the RAESS algorithm at higher temperature, we compared the results of a screening over the CoREMOF 2019 database.

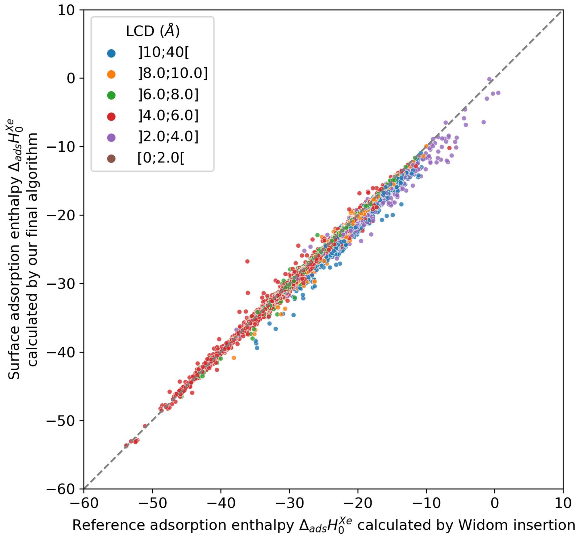


Figure 3.19: Scatterplot of the enthalpies calculated by our final algorithm (λ = 1.6 and μ = 0.85) compared to the enthalpies calculated by a 12k step Widom insertion simulation of xenon in structures of CoRE MOF 2019 with $LCD_{CCDC} \geq 3.7$ Å at 600 K.

The method is as expected less accurate, but it still gives a reasonable correlation on the performance, with an RMSE of 0.70 kJ mol^{-1} and an MAE of 0.41 kJ mol^{-1}. The errors have almost doubled when going from 298 K to 600 K. However, these limitations of the method are not crippling since adsorption processes are usually not performed at very high temperature. High temperatures are commonly used in temperature swing adsorption (TSA) to desorb the adsorbates rather than to adsorb them.

Other databases

ToBaCCo

We randomly selected 1,000 structures from the 13,511 porous frameworks of the ToBaCCo database to test the robustness of the RAESS method on a database other than CoreMOF. Since ToBaCCo contains structures with larger pores as suggested by a Moosavi et al., these materials are more unfavorable for adsorption of small molecules (such as Xe). The correlation is therefore found to be weaker than in the CoRE MOF 2019 database. This lower accuracy should be nuanced by the lack of suitability of these materials for Xe/Kr separation. Moreover, we can note that the points with weaker correlations correspond to the ones with an LCD_{CCDC} greater than 10 Å, which is not ideal to separate Xe from Kr.

The algorithm performs very well on the most adsorptive materials with xenon adsorption enthalpy values lower than −30 kJ mol^{-1}, because for these materials the adsorption sites are located near the surface. For broader pore sizes, it is interesting to note some limitations of the methodology we should be aware of. For determining the most attractive materials this shortcoming does not influence the final conclusions. Moreover, this limitation is not that damaging since the correlation although weaker still remains.

Amorphous materials

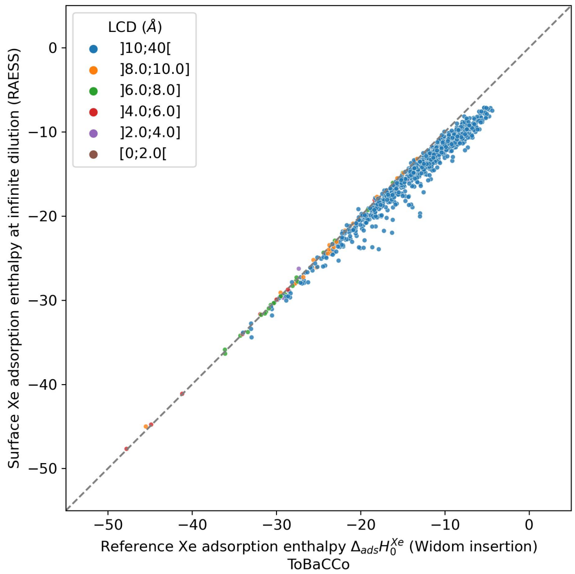


Figure 3.20: Scatterplot comparison of the xenon adsorption enthalpy calculated by the RAESS algorithm and the Widom insertion (RASPA2) on the ToBaCCo database. RMSE = 1.79 kJ mol^{-1} *and MAE =* 1.48 kJ mol^{-1}*. 915 structures have an* LCD_{CCDC} *greater than* 10 Å*.*

To further extend the possible uses cases of the RAESS algorithm, we tested our algorithm on the amorphous database [48]. The algorithm found results for 176 structures out of 196. The RASPA2 software was not able to run on these amorphous structures with our computers, running out of memory due to the large system size: therefore, there is no comparison point with a Widom simulation. However, we used another simulation that is closer to the ground truth since it samples a homogeneously distributed grid using an optimized algorithm we will present in the next section. This grid sampling managed to compute the adsorption energies of 175 structures.

The Table 3.2 gives the values of the adsorption enthalpies and the Henry constants of a few amorphous materials as well as the time they took to be computed. The sheer amount of atoms in each structure makes the CPU time required much higher than for the crystalline structures of CoRE MOF 2019. The time required are however quite manageable in a hypothetical screening procedure. If we consider all the 175 structures that could be calculated by our methods, the average time required is about 75 s per structure.

Structure Name	$\Delta_{ads}H_0^{Xe}$ (kJ mol^{-1})	K_H^{Xe} (mol kg^{-1} Pa^{-1})	CPU time (s)
aCarbon-Marks-id035	-63.55	6.98e-01	285.45
HCP-Colina-id016	-30.61	8.85e-05	3.88
Kerogen-Coasne-id010	-44.38	8.02e-03	61.2
PIM-Colina-id012	-26.39	7.00e-05	8.86

Table 3.2: Some amorphous materials' performance according to the RAESS algorithm. The results on the whole amorphous database is given in CSV format on the Github: `github.com/fxcoudert/citable-data/tree/master/154-Ren_ChemSci_2023`.

As we can see on the Figure 3.21, the accuracy of the surface sampling is rather high since by comparing it to an unbiased grid-based sampling, we obtain very similar results. The

RMSE is about 0.83 kJ mol^{-1}, which is higher than the one for CoRE MOF structures. This method could very likely be used to evaluate amorphous materials as a fast screening tool especially since the computation time required by the optimized grid sampling is about 623 s. The dimension reduction inherent to the surface sampling makes it one order of magnitude faster than conventional techniques.

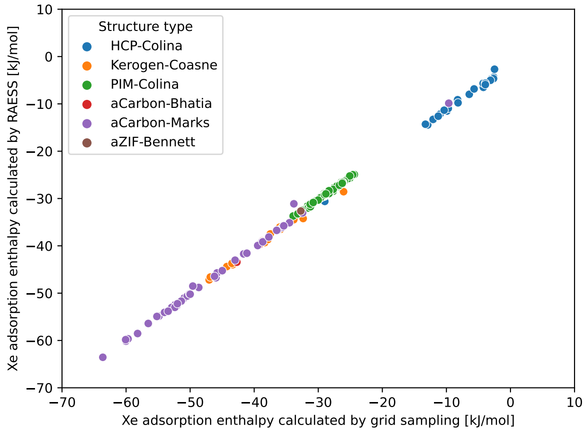


Figure 3.21: Scatterplot comparison of the xenon adsorption enthalpy calculated by the RAESS algorithm and the one calculated by a grid sampling (presented in the next section) on a database of porous rigid amorphous materials [48]. Raspa simulations could not be run on this database. Only the 175 structures computed by both methods are presented here.

3.2.5 Perspectives of surface sampling

We described a novel algorithm for the high-speed calculation of adsorption enthalpy in nanoporous materials, that takes a unique approach to reduce the sampling necessary. This new algorithm is based on the core principle of dimensional reduction, from a volume problem to a surface one. The algorithm is proven to be significantly faster than the reference Widom insertion (random sampling of porous space). Moreover, the error associated is found to be of the order of 0.4 kJ mol^{-1}, tested throughout the entire CoRE MOF 2019 database, for xenon adsorption. Even when comparing to existing very fast sampling techniques such as the Voronoi sampling, this surface sampling technique requires similar CPU time, combined with a better accuracy.

Based on these results, this algorithm has important potential for applications in the current computational analysis workflows of material databases, such as high-throughput screening studies. For instance, this algorithm can be used to get a fast approximation of the low-loading adsorption enthalpy of a molecule inside nanoporous materials. This cheap evaluation of the enthalpy can be used to screen out the structures with little affinity with the targeted adsorbate molecule. It can also be used as a thermodynamic descriptor for selectivity prediction in a machine learning model, as done by Simon et al.[14] The computational speed-up brought by this novel methodology can also enable the screening of materials databases at larger scale in the future.

We note, moreover, that the speed of our method resides in the sampling technique itself, rather than in the actual energy calculation. While we have benchmarked it in this work for a simple Lennard-Jones interaction potential, this sampling technique could equally be used to speed up samplings of space based on more expensive modeling strategies, including polarizable force fields or density functional theory (DFT) calculations. In the literature, the need for cheap *ab initio* grade thermodynamic properties is usually fulfilled by using an importance sampling method based on a classical force field.[225] In our method, the description of surface sampling is independent of any force field, and the sampling spheres can be defined according to kinetic radius, van der Waals radius or any other physically relevant distance. Consequently, given a definition of atomic radii, it is possible to define a surface on which to carry out other types of simulation such as neural network potential, DFT or any other force fields. Although the accuracy or relevance of such a sampling remains an open question, the approach will undeniably speed up the simulations. This could even be applied to calculate adsorption enthalpies while considering intrinsic structure flexibility,[159] a task whose main drawback is the high computation time required. Since surface sampling is hundreds of time faster than standard methodologies, we could use hundreds of snapshots in a flexibility-aware calculation.

Finally, although the algorithm in its present form can already be applied in a wide range of applications, additional development work could allow us to generalize it to polyatomic adsorbates. For instance, we would need to work on a definition of the molecular radius for non-spherical adsorbates as well as working all the orientation conformation of the adsorbent. We could imagine making the distance to the surface depend on the orientation of the adsorbate or sample a band volume on the surface. Although the best implementation of the surface sampling for polyatomic adsorbates remains an open question, in theory it should be possible to apply it to more complex adsorbates than the spherical noble gas. This would add more complexity to the algorithm but would not change the fundamental speedup due to surface sampling, since these orientation moves are also performed in other standard methodologies. To improve even more the accuracy, we could test hybrid samplings with multiple sampling spheres, or a combination of Voronoi nodes and sampling spheres. Another idea could be to have fractions of sphere that are oriented toward the center of pore given by the Voronoi node. In theory, having a wider variety of sampling points can only improve the sampling. There are therefore multiple possible sampling techniques that could be built around the method introduced herein. The code is made freely available on the group's GitHub (`github.com/coudertlab/RAESS`), where further development will be released.

3.3 Grid Adsorption Energies Descriptors (GrAED)

3.3.1 Implementation of the Algorithm

To build more relevant energy descriptors, we will now go back to the definitions of the adsorption enthalpy and the Henry constant (equation 2.21 and 2.14) that call for a homogeneous sampling of the adsorption space. The easiest way of sampling consists in laying a grid in the 3D space. This method is however known to be the most time-consuming on in theory. Inspired by our work on surface sampling, we designed an approach based on a symmetry-respecting grid,generated using the algorithms of the Gemmi Project,[226] where the points overlapping with framework are discarded. These new features coupled with a grid sampling makes the calculation of adsorption energies much less time-consuming while being very accurate.

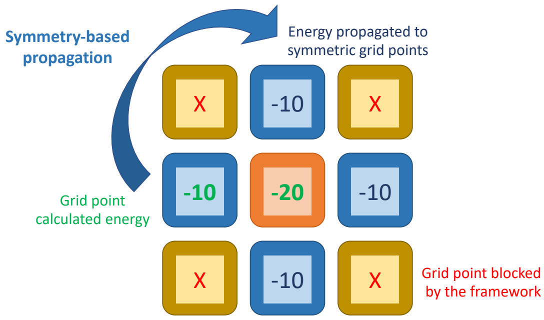


Figure 3.22: Principle of the energy sampling on a symmetry-based grid. On the 9 grid points, 4 points are blocked because they are too close to the framework atoms, 2 points are really calculated using the LJ potential and 3 points are propagated using the inner symmetry of the framework.

The corresponding algorithm has a core structure of a grid algorithm, where we need to evaluate the interaction energy at each point of the preset grid over the structure's unitcell. A naive approach would call for an expensive energy calculation at each grid point. To improve this approach, our algorithm is based on two main simplification — a quick evaluation of the framework occupied grid points and the exploitation of symmetry. The grid points that overlap with the framework's atoms have highly positive energy mainly due to the interaction with the overlapping atom. High values of energy does not contribute much to any thermodynamic quantities displayed in the section 2.1.5. For this reason by using a rejection parameter similar to the one developed for surface sampling introduced in the section 3.2.2, we can precalculate the interaction energy of the grid points within the sphere of radius $\mu \times \sigma_{g-h}$. If the value of this interaction energy is higher than a preset energy threshold E_{th}, then the associated grid point takes this values as the interaction energy and no further calculation will be performed for it. The symmetry of the grid is based on the symmetry of the structure determined using the Grid definition of the Gemmi Project. By using the symmetry operations on a grid point value, we can propagate it to the other symmetry-equivalent grid points as illustrated in the Figure 3.22, which reduces the number of time we need to calculate the interaction energy of a guest molecule at a given grid node with all the surrounding framework atoms within a given cutoff. Now that we presented the main building blocks of our optimized grid calculation, we will show how it integrates in the implementation of the algorithm.

1. We loop over the framework atoms and the grid points around a sphere of radius $\mu \times \sigma_{g-h}$, where σ_{g-h} is the distance at which the LJ potential energy between the guest atom g and the host atom is zero. The LJ potential energy between the guest molecule and the closes host atom is calculated and only the grid points with an energy lower than a predefined threshold E_{th} are considered "unvisited" and will be recalculated in the following loop, the others are considered blocked by the framework and will be considered already "visited". This first loop over the framework atoms aims at filtering out the grid points that are blocked by the framework, and we will refer to this preliminary filtering step as "blocking" in the Table 3.3.

2. A second loop over the "unvisited" grid points is performed — at each increment, if the point is "unvisited" we calculate the interaction energy between the guest and all the host atoms within the cutoff, then the symmetric images of this point are filled with the

same energy value and are considered "visited" by the algorithm. This symmetry-aware grid exploration allows the algorithm to divide the time required by the average number symmetry images — this module will be referred to as "symmetry" in the Table 3.3.

By combining both the "blocking" of the high energy grid points and the "symmetry" based calculation of the interaction energies, we built a "fast" version of the grid calculation algorithm that can compete with our previously developed rapid surface sampling method (RAESS). To control the trade-off between accuracy and computation time, we can vary the spacing between the grid points — the computation time is theoretically inversely proportional to the cube of the spacing.And, for some values of this spacing, this algorithm can even be faster than the surface sampling on the CoRE MOF database where the symmetry plays an important role (see Table 3.3).

3.3.2 Performance on the Adsorption Equilibrium

The good performance of the grid sampling on the structures of CoRE MOF 2019 database can be explained by their rather small pores and their high order of symmetry. For instance, the average void fraction for a 1.2 Å probe radius is equal to 0.16 and the average number of symmetric images is equal to 5.8 (most MOFs present symmetry operations). On average, the "blocking" procedure means that only ∼16% of the grid points really need to be calculated. And, the "symmetry" procedure implies that only ∼17% of points need to be considered, and the combination of both theoretical reduces the number of useful points to only 2.7% of the grid. This leads to a significant reduction in the CPU time of the calculation while keeping the accrurary level compared to the naive grid approach, as shown in Table 3.3.

Energy sampling method	Average CPU time (s)	RMSE on adsorption enthalpy ($\mathrm{kJ\,mol^{-1}}$)
Grid – naive – 0.1 Å	71.3	0.03
Grid – blocking – 0.1 Å	18.8	0.03
Grid – symmetry – 0.1 Å	16.8	0.02
Grid – fast – 0.1 Å	4.8	0.02
Grid – fast – 0.3 Å	0.16	0.23
RAESS[16]	0.34	0.34
Widom[197] (12k cycles)	150	0.01

Table 3.3: Performance comparison of the new grid method to other standard techniques used to calculate the xenon adsorption enthalpies. The RMSE is calculated by comparing to the values given by a 100k-steps Widom insertion considered as the ground truth. The associated calculations are performed on the structures with an LCD_{CCDC} over 3.7 Å of CoRE MOF 2019 database with a single Intel Xeon Platinum 8168 core at 2.7 GHz. [incohérence sur les rmse refaire avec LCD_{CCDC}]

As we can see of the Figure 3.23, the approach does not damage the accuracy of the adsorption enthalpy and of the Henry constant. There is an almost perfect accordance between the Widom insertion method and the grid-based approach for a very finely meshed grid (0.12 Å spacing). This was expected since both methods are unbiased sampling of the adsorption energies. If we really want to nit-pick, we can note that very small differences between both methods can be detected for a few structures. [give rmse for enthalpy and henry]

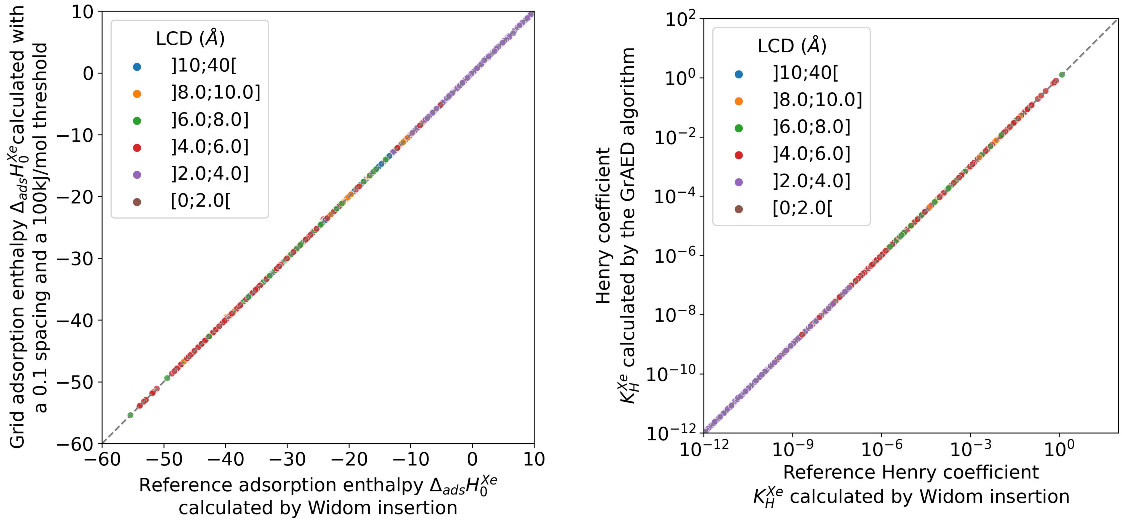


Figure 3.23: Comparison of the adsorption enthalpies (left) and the Henry constants (right) calculated by the optimized grid energy sampling (for a 0.1 Å *spacing, a rejection parameter* $\mu = 0.8$ *and an energy threshold* E_{th} *of* $100\,\mathrm{kJ\,mol^{-1}}$*) and by the Widom insertion of RASPA with 100,000 cycles.*

From the energy values of this grid, we can now calculate many useful descriptors of the adsorption process. We have seen the performance on the Xe adsorption enthalpy and the Xe Henry constant. But as mentioned in the section 2.1.5, we can also derive the Xe adsorption Gibbs free energy and the Xe adsorption entropy. If we now consider the krypton in addition to the xenon, we can naturally evaluate the Kr adsorption thermodynamic quantites but also the exchange thermodynamic quantities and especially the Xe/Kr selectivity (the key metric in evaluating the separation process we are interested in).

3.3.3 Performance on the Exchange Equilibrium

To characterize the concurrent adsorption of the binary mixture of xenon and krypton we usually use the selectivity. But the uncertainty will mechanically increase the uncertainty since the selectivity is a quotient of the Henry constants of the concurrent adsorbates. In this section, we want to measure this error and see if it is usable to characterize the separation.

[Figure compare widom]

Selectivity well predicted >

[Performance on exchange metics]

[Figure compare GCMC with averaging at 298K]

3.3.4 Characterization of the Ambient-pressure Selectivity

Thermodynamic quantities

comparison with ambient-pressure quantities

High temperature quantities

In this section we show figures that support the use of higher temperature averaging to better understand the ambient-pressure case.

[] We chose to use 900 K to improve the correlation of key metrics like the adsorption enthalpy of xenon. This improvement impacts the exchange free energy metrics as well as the adsorption enthalpy associated with the separation of xenon from krypton. The exchange Gibbs free energy and the adsorption enthalpy of xenon at ambient pressure are better correlated to the equivalent values at lower pressure and higher temperature (900 K) than to the 298 K case. These observation supports the use of higher temperature averaging in the final ML model we present in the main paper.

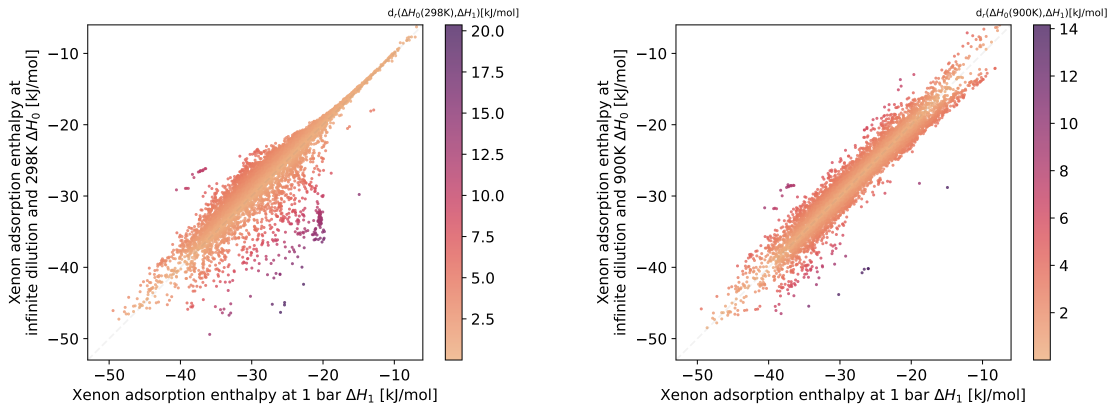


Figure 3.24: Scatterplots of the low-pressure xenon adsorption enthalpy at 298 K *(left) and at* 900 K *(right) calculated by the GrAED algorithm against the ambient-pressure xenon adsorption enthalpy at* 298 K*. Using a higher temperature Boltzmann averaging, the correlation with the ambient-pressure case of interest is much higher, the R2 coefficient improves from* 0.80 *to* 0.92 *for instance. The RMSE also decreased from* 2.87 kJ mol^{-1} *to* 1.76 kJ mol^{-1}.

Statistical characterization

Average / Std / skew / kurtosis

Perspectives of these energy descriptors

github.com/coudertlab/GrAED

[transition into ML models]

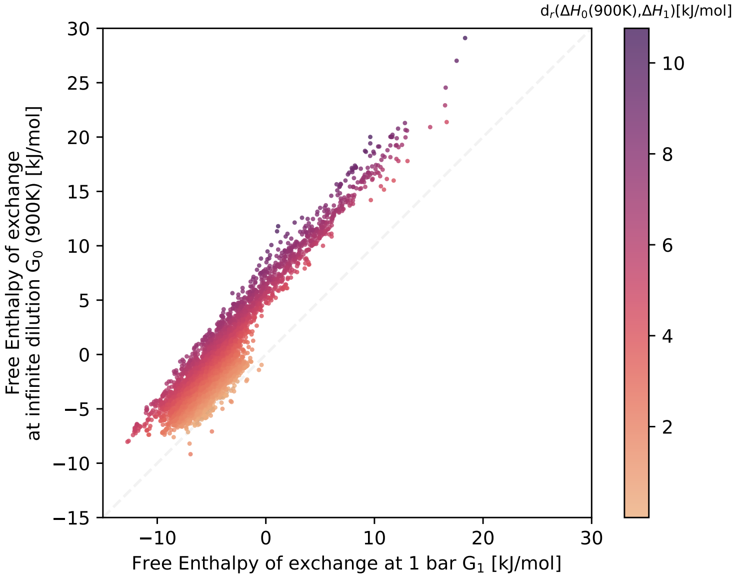


Figure 3.25: Comparison plot between the low-pressure exchange free energy at 900 K *calculated by the GrAED algorithm and the ambient-pressure exchange free energy at* 298 K *calculated by Widom insertion.*

4

STATISTICAL LEARNING OF ADSORPTION PROPERTIES

[t-SNE somewhere ?]

4.1 Machine Learning

Machine learning (ML) models have been widely used to characterize adsorption, transport, catalytic or mechanical properties, just to cite a few. It can in some cases replace very time consuming simulations with simpler calculation of key descriptors that can help the model predict the desired properties. In other cases, it is used to describe the structure–property relationships learned by the ML model. However,machine learning is not a silver bullet, we cannot blindly apply it on any applications; it requires a thorough work on understanding the key variables that will improve the prediction. By using our work on the thermodynamic descriptors and our knowledge on the effect of pressure on the selectivity, we will build a machine learning model to characterize the separation of xenon from krypton at ambient pressure.

4.1.1 What is machine learning

To understand how machine learning, first, we need to understand how a computer accomplishes a task. The human operator plays a key role in the process — after designing the solution through theoretical considerations, he needs to write down a list of instructions, called an algorithm, that specifies every needed actions given the circumstances so that the computer achieves the desired outcome. In physical or chemical sciences, these algorithms usually articulates the different components of a theoretical model, which can be an equation without analytical solutions, an analytical expression or a probabilistic problem, just to cite a few. The previous chapters typically presented such algorithms for the simulation of adsorption processes; for instance, the GCMC simulations are based on the statistical physics of the phase equilibrium between a gas phase and adsorption phase inside a nanoporous material, and a Monte Carlo model is used to reproduce the statistics associated to the grand canonical ensemble. The energy sampling algorithms along with the Widom insertion are also good examples of how the computer can help the theoretician model the systems under specific chemical and physical conditions.

A machine learning model is also based on an algorithm, but the goal is very different from the above-mentioned examples — it doesn't aim at giving all the details of the computation according to proven theoretical principles. As implied by the name, the ambition would be to learn underlying relations within the input data so that it performs the task itself. The machine learning (ML) algorithm is then the list of instructions that sepecifies how the machine is going to learn from the data. For example, clustering algorithms can distinguish different classes of elements within a disordered dataset so that new concepts emerge; this type of machine learning algorithm is called unsupervised learning because we do not predfine or prelabel the data and the machine helps us understand the underlying structure in the data. The class of algorithm we want to study is rather the supervised learning model, which learns from labeled data the relation between the label and the characteristics (called features or descriptors) from a given set of data points, and can predict the label from unlabeld data using the characteristics. For example, if we want to predict the weather of tomorrow, the model could use the past weather of similar dates to infer if it will rain tomorrow; the history of the weather are the features of the ML model and the future weather is the target variable or the label of the data.

To articulate the differences between a standard algorithm and an ML algorithm, let me introduce a fascinating board game called Go. This game is traditionnally played with 2 players on a 19 by 19 board, where each player places black/white pieces to controle the maximum of boxes. Based on these simple rules, different algorithms have been developped to make the computer play the game. The first Go program was written in the late 60s to mimick the pattern recognition of Go players when estimating the "score" through an influence function,[227] and from the 80s to the beginning of the 21st century the first Go programs capable of playing were releases. These programs were based on simple alpha-beta search algorithms that seeks at testing every possible moves (while pruning the less promising ones); while they were working very well in other games like chess (IBM's Deep(er) Blue beat the world champion of chess in 1995), in Go these type of programs were only at the level of a novice player. The difference of performance lies in the combinatorics behind both games, the game of chess has a number of legal positions lower than 10^{47},[228] while for the game of Go there are approximately 10^{171}

legal positions.[229, 230] The state space to explore is incomparably greater and a boost in the computing power that improved the performance for computer chess is not going to make a difference for Go. A drastic reduction of the space to be explored is needed for a computer program to work. The biggest improvement came when in 2007, Coulom introduced a Monte Carlo tree search.[231] This algorithm uses heuristics to distinguish between bad an good moves according to human perception of the game, a probability of selection is assigned to the moves according to their potential (policy), potential moves are randomly picked according to this probability; the average outcomes associated with a parent move gives the value of the move. The computer Go is now more efficient in the evaluation of the moves using a Monte Carlo sampling, and it can now play with average amateur players, but it is nowhere near surpassing them. Up until now, the algorithms are based on human knowledge that the programmer implements directly in the computer using machine instructions. Statistics and randomness are used to orient the machine towards the best moves and reduce their predictibility, but the statistics that identifies the moves are based on human heuristics that are usually not generalizable. The big revolution brought about by machine learning in the field aims at better evaluating these statistics using the data from already played games. By using a dataset of 30 million moves, the Alpha Go is based on the same Monte Carlo tree search framework but it replaces the formulas behind the probability of searching a move by a machine learning model called the “policy network” and the one behind evaluating the confidence in winning of the position by a value “network”.[232] Alpha Go was the first computer program to beat a world champion in 2016. One year later, to further emancipate from human knowledge, an improved version, Alpha Go zero generates its own data by playing games against itself to train a similar machine learning structure than presented before. This new version beats 100 times out of 100 the former version,[233] which marks a new era of domination of computer go over the best player in the world, and the defeat of another top player just confirms the advent of this new era.

In this example, we can see how the machine learned the value of each moves by compiling the knowledge of huge datasets in a deep neural network. The main difference between conventional approaches to algorithmic and machine learning is very well illustrated in the previous example; the goal is not tell the computer how to play using player knowledge implemented in formulas and explicit instructions, but it is to give an explicit framework with flexible parameters that the model needs to learn using a database. In other words, the parameters of a model are fitted to match the values of a database, while being capable of generalizing in situations outside of the database (this notion of generalizability will be further discussed in the following sections). In this section, the goal is not to give a complete overview of all existing models but rather to introduce the main concepts of ML through the example of the model we will use for our problem of selectivity performance prediction.

4.1.2 How to accurately learn

The Elements of Statistical Learning[234]

[From algorithms to machine learning. An algortihmic way of finding a relation in data. Give example of clustering and Go, selectivity example simulation / ml.]

[Supervised / unsupervised learning: foucs on supervised]

[cross validation, overfitting, hyperparameter optimization]

Hyperparameter fine-tuning

We used the training data to perform a random search of hyperparameters, with 5-fold cross-validation to evaluate the root mean squared errors (RMSE) of the model. The range of search explored for each hyperparameter is made available in the SI. After this search, a set of optimal hyperparameters were identified, that give an average RMSE of 0.36 $\mathrm{kJ\,mol^{-1}}$; we used it to build our final model. A convergence plot of the model performed using 5-fold cross-validations is given in Figure S6. Given this configuration, the model is tested on the prior defined test-set and interpretation tools are used to better understand the structure-property relationships in play.

Explainable AI

The final model is trained on the predefined training set using XGBoost with the fine-tuned hyperparameters. By testing it on the test set, we measure the accuracy of our approach, however, it is interesting to extract chemical insight into the hidden relationship between the predicted value and the descriptors, to better understand the thermodynamic origins of the performance. In this work, we used the Shapley values,[235] a game theory concept developed by Shapley in 1953, to measure the contribution of each descriptor in the predicted value. This tool is used locally to understand for a given structure how their characteristics had contributed to the prediction. To draw structure-property relationships, we would need to use a global interpretation methods such as the SHapley Additive exPlanations (SHAP) method thoroughly detailed in the online book *Interpretable Machine Learning* of Christoph Molnar.[236] The SHAP tool developed by Lundberg and Lee [237] is based on a faster algorithm adapted to tree-based ML models like gradient boosting, TreeSHAP, and integrates useful global interpretation modules like SHAP feature importance and dependence plot.

4.1.3 Tree-based models

Regression tree

Random Forest

Gradient Boosting

random forest boosting eXtreme Gradient Boosting

4.2 Prediction of the selectivity

4.2.1 Intro

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]

In this article, we combined a grid-based approach with core components of our previously developed RAESS algorithm [16] to design a new adsorption energy sampling technique. Moreover, a statistical characterization of the pore size and energy distributions has been performed to inform the model on a potential selectivity drop. By combining these two approaches, we propose a set of useful ML descriptors for fast and accurate ambient-pressure selectivity prediction, and we highlight its performance on the case of xenon/krypton separation in the CoRE MOF 2019 database[64].

4.2.2 The machine learning model

We chose to use eXtreme Gradient Boosting (XGBoost) as the machine learning framework for our predictive model because of its accuracy, efficiency and simplicity of use. Its performance has long been proven since 17 out 29 Kaggle challenge winning solutions were based on this algorithm in 2015. The XGBoost system is highly scalable and parallelized, which gives very fast model training.[238] Compared to more standard tree-based algorithms such as random forest (commonly used in the field [14]), the boosting component of the algorithm means that it learns from previous mistakes and puts higher weights on the problematic data points, hence improving the accuracy of the final ML model.

In the next sections, we introduce new descriptors for nanoporous materials, as well as new concepts of feature engineering based on energy and pore size histograms. The ML features presented have been selected by progressively filtering out the less influential ones on the accuracy of the final model, see the complete list in Table S1-3 of Supporting Information (SI). The influence or importance are defined later in a section dedicated to the interpretation of the model. The hyperparameters of the model were fine-tuned using random search to design the best performing final model. Finally, the influence of the pre-selected descriptors on the final model is interpreted using a unified approach.

4.2.3 Target variable

We want to predict the Xe/Kr ambient-pressure selectivity faster than standard techniques. To obtain reference values (ground truth), we used the Raspa2 software[200] to run grand canonical Monte Carlo (GCMC) calculations of 20-80 Xe/Kr mixtures at 298 K and 1 atm on our cleaned database. The van der Waals interactions are described by a Lennard-Jones (LJ) potential with a cutoff distance of 12 Å. The LJ parameters of the framework atoms are given by the universal force field (UFF),[184] and the guest atoms (xenon and krypton) have their LJ parameters taken from a previous screening study.[157] The study only focuses on a given Xe/Kr composition usually obtained by cryogenic distillation of ambient air [1] as a first step towards predicting other mixtures at different physical conditions (*e.g.* Xe/Kr mixtures out of nuclear off-gases). In the broader scope, this methodology could be adapted to the desired application with some tweaks on the descriptors calculation (*e.g.* CO_2/CH_4 separation).

We decided to use a logarithmic transform of the selectivity instead of the raw value because we are more interested in the order of magnitude of the selectivities than to predict the higher values of selectivity — an ML model that predicts selectivity values can lower down the errors by focusing the prediction more on the higher values than the lower ones. By focusing on the logarithmic transform of the selectivity, we can better separate the different orders of magnitude of the selectivities. This approach distributes more evenly the efforts on all the different values of selectivities. Moreover, this logarithmic transform is related to a thermodynamic quantity that we elaborate later in the section 2.1.5; it can therefore be easily compared with the energy descriptors we introduced in this article.

4.2.4 Database and data preparation

To test our methodology on a set of realistic MOFs, we chose to screen the 12,020 all-solvent removed (ASR) structures of the CoRE MOF 2019 database[64]. After removing the disordered and the non-MOF structures as well as the ones with a large unitcell volume of 20 nm^3, we obtained a set of 9,748 structures. Then we analyze the string information given by the Zeo++ software[239] to reduce the number to 9,177 by removing the structures that are not tridimensional, where solvents are still detected (wrongly classified in all-solvent removed), or where the metal is radioactive or fissile (e.g., Pu-MOF TAGCIP[240], Np-MOF KASHUK[241], U-MOF ABETAE[242] or Th-MOF ASAMUE[243]) — this can be a source of risks in a nuclear waste processing plant. Furthermore, we added a condition on the largest cavity diameter (LCD) to keep only the structures that can accept a xenon molecule: 8,529 structures have a LCD higher than 4 Å (approximately the size of a xenon molecule). This is equivalent to removing the structures with very unfavorable adsorption enthalpies, that are not promising for our adsorption-based separation (see previous work [16]).

Then, the descriptors summarized below (and fully detailed in Supporting Information) were calculated on this restrained dataset. At this stage, 370 structures failed to be calculated in GCMC and 82 have no standard deviation for the pore distribution (skewness and kurtosis cannot be retrieved). A final dataset of 8,077 structures was therefore used to perform our ML-assisted method of screening the Xe/Kr adsorption selectivity. Based on this final set, 20% were randomly used for the test set and 80% were used to train our model. The goal is to learn from the training set a relationship between the descriptors and the target ambient-pressure selectivity in order to evaluate the performance on the test set. A CSV file of training and test sets can be found in the data availability section.

4.2.5 Geometrical and chemical ML descriptors

Looking at a number of different research papers on supervised ML for the prediction of adsorption properties,[14, 71, 82, 244, 245] we see that some descriptors are recurrent: 1) geometrical descriptors obtained by software like Zeo++ [239] such as the surface area (SA), the void fraction (VF), the largest cavity diameter (LCD) and the pore limiting diameter (PLD); and 2) physical and chemical descriptors such as the framework's density, the framework's molar mass, the percentage of carbon (C%), nitrogen (N%), oxygen (O%), hydrogen but also halogen, nonmetals, metalloids and metals, and the degree of unsaturation. Although these descriptors are very versatile and used in many ML models, they however fail to provide specific information for our ML task. As shown by Simon et al., energy descriptors are greatly influential in ML models for selectivity prediction.

The geometric analysis of the crystalline porous materials is typically based on the van der Waals (vdW) radii predefined by the Cambridge Crystallographic Data Centre (CCDC). This force field-independent choice can create a gap between the geometrical descriptors and the thermodynamic values obtained through molecular simulations. Inspired by a recent work on the comparison of PLDs and self-diffusion coefficients,[44] we defined a list of vdW radii to be read by the Zeo++ software (more details in https://github.com/eren125/zeopp_radtable). In this study, all Zeo++ calculations use an atomic radius that corresponds to the distance where the LJ potential reaches $3k_\mathrm{B}T/2$, for T = 298 K.

The SA exposed to different probe sizes (1.2 Å, 1.8 Å and 2.0 Å) were tested. The probe occupiable volume was chosen to measure the void fraction (VF) for different adsorbent by using probe sizes of 1.8 Å (close to the radius of krypton) and 2.0 Å (close to that of xenon). This definition of the pore volume was found to be in better agreement with experimental nitrogen isotherms.[186]

Because our goal is to predict the difference between the low-pressure selectivity and the ambient-pressure (for a given gas mixture composition), some of these descriptors have very little importance, and the key factor is the difference of accessible volume and the affinity of the remaining pore volume with xenon, compared to krypton. The intuition developed in the previous study sketched the role of a diverse distribution of pores with different xenon affinities.[18] For all these reasons, from all the "standard" descriptors taken from the literature, we kept only the following 7 descriptors: C%, N%, O%, LCD ("D_i_vdw_uff298"), PLD ("D_f_vdw_uff298"), SA for a 1.2 Å probe ("ASA_m2/cm3_1.2") and VF for a 2.0 Å probe ("PO_VF_2.0"). We also built a new descriptor ΔVF void fraction values, the difference of volumes occupiable by xenon (2.0 Å) and by krypton (1.8 Å). All these descriptors along with other pore size distribution based geometrical descriptors are presented in detail in the Table S1 of the Supplementary Information (SI).

4.2.6 Pore size distribution

To generate a histogram of pore sizes (or pore size distribution, PSD), Monte Carlo steps are used to measure the frequency of every accessible pore sizes binned by 0.1 Å.[246] This histogram can then be used to generate descriptors based on statistical parameters that describes the overall location, the dispersion, the shape and the modality of the distribution. In addition to the mean and standard deviation of the distribution, we introduced two additional moments: the skewness (γ) corresponds to the third standardized moment and measures the asymmetry of a distribution; and the kurtosis (k), being the fourth standardized moment, measures the relative weight of the tails of the distribution. Knowing the importance of characterizing the number of different pore sizes suspected to be at the origin of the selectivity drop observed, we tried to find a simple descriptor to measure the number of modes in the distribution. The Sarle's bimodality coefficient, $\mathrm{BC} = (\gamma^2 + 1)/k$, represents a simple quantification of how far we are from the unimodality based only on skewness and kurtosis.[247] Finally to further measure the diversity of pores, we introduced an effective number $n_{\mathrm{eff}} = \mathrm{N}^2 / \sum n_i^2$ of pore sizes, where N is the total number of points in the histogram and n_i the number of points associated with the i^{th} bin. This number is very similar to a statistical number widely used in other scientific fields: in political science it is used to measure the effective number of political parties, [248], in ecology the inverse Simpson's index evaluates the species diversity in an ecosystem,[249] or in quantum physics the inverse participation number measures the degree of localization of a wave-function.[250] This effective number of pore sizes gives an idea of the diversity of pore sizes (considering a binning of 0.1 Å). A high effective number would mean that multiple pore sizes are highly represented in the structure; this descriptor gives an idea of how scattered the pore sizes are. All these descriptors carries information on the form of the PSD needed to figure out the loading and selectivity situation in the framework near saturation uptake, which is crucial to predict the ambient-pressure selectivity.

4.2.7 From infinite dilution to ambient pressure

The low-pressure selectivity provides a first intuition of the selectivity at higher pressure, as demonstrated in our previous work showing a correlation between the selectivity at both pressures.[18] If we adopt the Gibbs free energy formalism (Equation 2.25), which correspond to a logarithmic transform of the selectivities, this correlation is confirmed and highlighted on Figure 4.1. We can also note that although a majority of structures have similar selectivities in both pressure conditions, a handful of structures experience a selectivity drop at higher pressure. The zero-loading selectivity is always higher or similar to the ambient-pressure one, it gives therefore a solid ground on which to build an efficient prediction model. The second ingredient for a good prediction model is to build explanatory descriptors related to this selectivity drop phenomenon. One of the main causes to the selectivity drop being the presence of bigger pores that are less attractive xenon, therefore additional information on the pore size distributions or the energy landscape would be helpful for this task.

To incorporate information on the pore size diversity of the materials, we carried out statistical measurements on the PSD. By analyzing them, we detected explanatory factors at the origin of the observed selectivity drop. A high degree of multi-modality in the distribution would mean a diverse set of pores, which can lead to a selectivity drop if the pores are significantly different one from another. The more distant is the average pore size from the largest cavity diameter the higher the chance of observing a selectivity drop, because a big difference between the pore sizes bring about lower selectivities. All these statistics are designed to give as much knowledge as possible on a hypothetical selectivity drop and on the quantitative estimation of its magnitude.

To better quantify the change of selectivity, it could be interesting to give statistics on the distribution of interaction energies for xenon and krypton calculated by our grid algorithm. These statistics include moments of different orders (up to 4) of the energy distribution, which

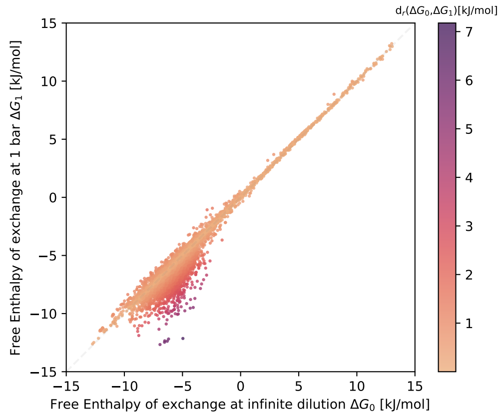


Figure 4.1: Comparison between the Gibbs free energy of exchange at low pressure ΔG_0 and ambient pressure ΔG_1 labeled by the relative distance between them. This plot is equivalent to a logarithmic plot of the selectivities at these two pressure conditions.

informs on the adsorbate–adsorbent interaction energies in the nanopores at higher loading. The shape of the energy distribution can help assess quantitatively the change in selectivity. We can consider this as a way of compressing the whole energy distribution into a few statistical values, which is a standard method used in the field of data science to tackle distribution data. The same approach has also been applied to the Boltzmann weighted distributions to generate temperature specific descriptors for the energy distributions.

By using different temperatures, we noted that the infinite dilution adsorption enthalpies at higher temperatures can be better correlated to the adsorption enthalpy at ambient pressure. The minimum error was found for the adsorption enthalpy at 900 K, which gives an RMSE of 1.76 kJ mol^{-1} instead of 2.87 kJ mol^{-1} for the 298 K case. This new type of descriptor is very interesting since it better performs around the high selectivity region, where the standard Boltzmann average at 298 K loses its accuracy (see Figure S1). As we can see in the Figure S7, the exchange free energy at 900 K and the excess of free energy compared to the 298 K case are the second and third most influential descriptors of our ML model. They are complementary to the exchange free energy at 298 K to predict selectivities at higher pressures.

By combining the above-mentioned features with more standard geometrical descriptors, we trained an ML model for the ambient pressure selectivity that identifies the origins of the selectivity drop and gives promising prediction results.

4.2.8 ML model performance

In this section, we present the performance of the ML model that learned the joint effects of all the newly introduced descriptors to detect and evaluate the observed drop between the easily accessible low-pressure selectivity and the more computationally demanding ambient-pressure selectivity. A GCMC simulation of a 20-80 xenon/krypton gas mixture takes in average 2.400 s per structure on the CoRE MOF 2019 database, while our grid-based adsorption calculation only takes about 5 s per structure (on a single Intel Xeon Platinum 8168 core at 2.7 GHz). To compute all features needed for our prediction, we would need less than a minute per structure, which is way faster than the 40 minutes required for a GCMC calculation. The ML-based approach has a very clear speed advantage over standard molecular simulations. But to be a good substitute, it needs to keep a good level of accuracy on an unseen set of structures.

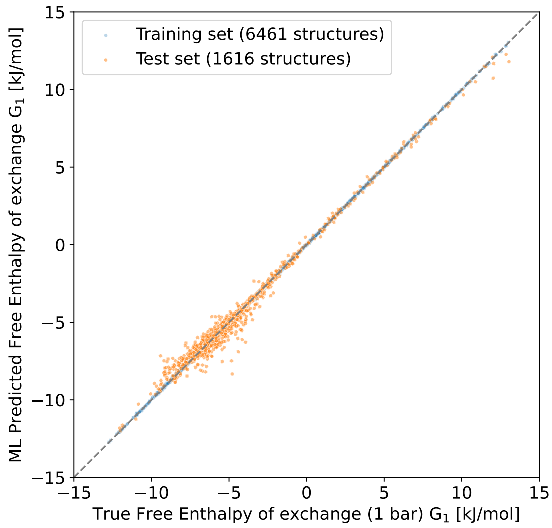


Figure 4.2: Scatter plot of the exchange free energy predicted by the model. There is a good agreement between the predicted and true values. On the test set, there is an RMSE of 0.37 kJ mol^{-1} *and an MAE of* 0.21 kJ mol^{-1}. *This plot is equivalent to the scatter plot between the logarithm of the ambient-pressure selectivities (see Figure S5 of the SI). The corresponding errors for the ambient-selectivity are 2.5 and 1.1 for respectively the RMSE and MAE of the selectivity, and 0.065 and 0.038 for the RMSE and MAE of its base-10 logarithm.*

We defined a set of 80% randomly chosen structures out of the final dataset to train and fine-tune the parameters of our model. A randomized search over a range of maximum depths, learning rates, sizes of feature samples used by tree and by level, sizes of data sample and alpha regularization parameters has been performed and a set of hyperparameters have been chosen to minimize the average RMSE computed using a 5-fold cross-validation. The ranges used in the randomized search as well as the final hyperparameters set are given in SI. By using this parameterization, our XGBoost model has an RMSE of 0.37 kJ mol^{-1} and an MAE of 0.21 kJ mol^{-1} on the exchange Gibbs free energies of the test set of 1,616 structures. If we convert back these results to the selectivity values, the RMSE on the selectivities would be 2.5 and 0.07 on the logarithm base 10 of the selectivity, which means that the order of magnitude of the selectivity is known with a very good accuracy. To prove that this good performance is not fortuitous, we used a 5-fold cross-validation procedure on the whole dataset and found an average RMSE of 0.36 kJ mol^{-1} with a standard deviation of 0.01 kJ mol^{-1}, which is consistent with the performance given by a standard train/test split.

This method can later be used in a screening procedure that relies on cheap descriptors to skim off obviously undesirable structures to only keep the promising structures for the final ML model evaluation. This is the reason why, as previously explained in the methods, only the 3D MOF structures with an LCD above 4 Å are kept because they have a positive xenon affinity, which is a necessary condition for a good Xe/Kr selectivity. Our model being very good at predicting the ambient pressure selectivity of structures with good xenon affinity, the proposed screening procedure, illustrated Figure 4.3, would include (i) a check of the nature of the structure to insure it is a 3D MOF structure, (ii) then a filter on the LCD value (above 4 Å), (iii) a pre-evaluation of the Xe/Kr selectivity at infinite dilution using the grid-based method, and (iv) finally the ML evaluation to keep only structures above a certain threshold of ambient-pressure selectivity (*e.g.* 30). We could eventually evaluate more thoroughly the top structures using GCMC simulations, *ab initio* calculations or adsorption experiments.

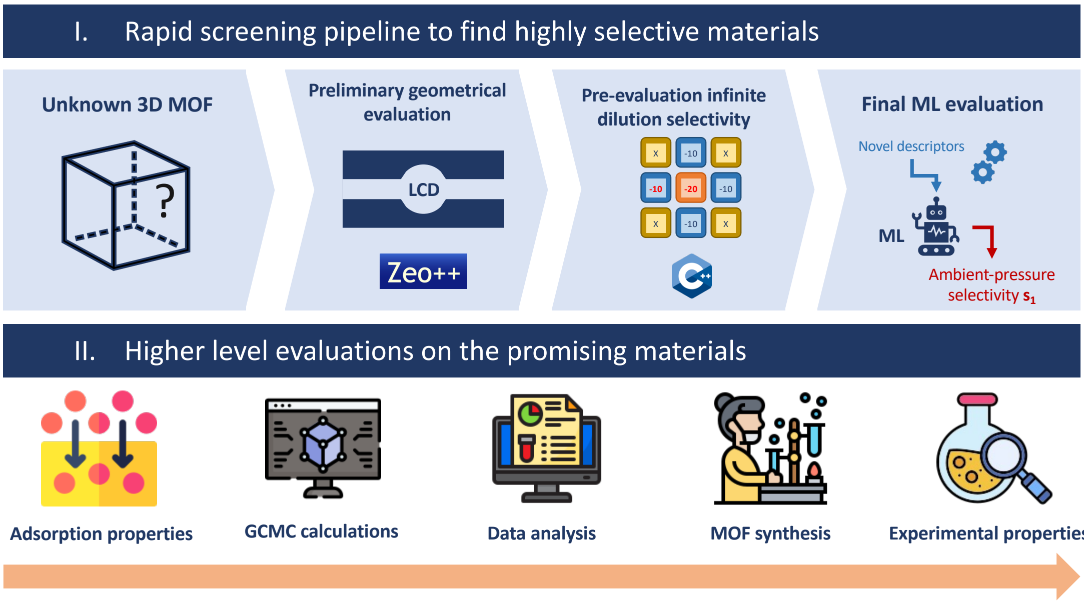


Figure 4.3: An illustration of the screening procedure that could be used to find highly selective materials.

4.2.9 Opening the black box

To better understand the intuition behind this selectivity drop, we used the SHAP[236, 237] library of interpretation models to draw relationships between the descriptors and the predicted ambient-pressure selectivity. This code library is based on the calculation of Shapley values[235] that measure the contribution of each descriptor to the prediction to locally interpret our ML model. This interpretation model untangles the interdependence between the descriptors to extract an individual contribution. To go beyond the local interpretation, we can rapidly compute the Shapley values for the whole dataset using faster algorithms;[237] scatter plots of the contribution as a function of the descriptor values called SHAP dependence plots can then be drawn to make a more global interpretation of our ML model. Knowing a descriptor value we could then infer, with a certain level of uncertainty, how it changes the final predicted value, which highlights unknown structure–property relationships. Finally, we can use the mean absolute Shapley values of each feature on the training set to measure the feature importance (see Figure S7 and S8).

Global interpretability

To rank the descriptors according to their average impact on the magnitude of the model output, we can use the mean absolute Shapley values of each descriptor. The importance plot associated with these values are presented in Figure S8. Even if the low-selectivity exchange Gibbs free energy has a SHAP importance value way above the others, it only serves as a baseline where only a correlation close to the one presented Figure 4.1 can be reached; the other descriptors play a major role in moving the outliers of the figure closer to the diagonal line. Energy descriptors play a major role in the model's prediction, and geometry-based new descriptors while playing a more secondary role are key in evaluating the gaps between the low-pressure case with the ambient-pressure one that we are interested in. To dig deeper into the mechanisms that allows the model to predict the selectivity with a very good accuracy — the RMSE and MAE on the test set's selectivity being respectively 2.5 and 1.1 — we are now going to look into the SHAP dependence plots of each interesting descriptor that plots the contribution to the predicted value as a function of the actual descriptor value.

To make a global interpretation, we applied the partial dependence module provided by the SHAP library on our model. Although other methods to compute dependence plots exist (*e.g.* partial dependence plots),[236] we can keep a good level of consistency between our global and local interpretations by using the same underlying theory. The SHAP dependence plots of all the descriptors of the Figures S9 and S10, these plots have a rather distinct form, directions and shape, which is encouraging for the interpretability of our model. By looking at the profile of the dependence plots, we can extract valuable information on how the ML model predicts the ambient-pressure selectivity.

The most important descriptor is obviously the exchange free energy "G_0" associated to the low-pressure selectivity, its contribution has a very strong positive linear correlation (see Figure 4.4), which gives a base value on top of which the other contributions will either reduce the free energy (more selective) or increase it (less selective). The model can be interpreted as the combination of a baseline combined with smaller tweaks that estimate the magnitude of the deviation from the ideal low dilution case. For instance, the next two descriptors "G_900K" (900 K low-pressure exchange free energy) and "G_Xe_900K" (900 K low-pressure xenon adsorption free energy) continue to build up the baseline by providing information on the low-pressure selectivity, but they start giving a glimpse of deviations needed to differentiate between the structures experiencing a drop with the ones that keep their selectivity. As we can see in the SI (Figure S1 and S2), the thermodynamic quantities at high pressure is closer to the 900 K case than to the ambient temperature one, these two descriptors informs naturally on the selectivity at higher pressure. For "G_900K" (see Figure 4.4), blue points (corresponding to a "G_0" of around $-8\,\mathrm{kJ\,mol^{-1}}$) can have either negative or negligible contributions depending on the value; values below $-4\,\mathrm{kJ\,mol^{-1}}$ give a negative contribution with a linear relation, whereas values between -4 and $5\,\mathrm{kJ\,mol^{-1}}$ give constantly almost zero contribution. This type of domain differentiation illustrates how the model can identify structures with a selectivity drop based on the values of a descriptor. We will see more telling examples of how the contribution to the selectivities are determined using the values of the remaining descriptors.

The U-shape of some SHAP dependence plots can highlight optimal values for the associated descriptors. For instance, the optimal value of "D_i_vdw_uff298" is around 5.1 (see Figure 4.4) and the optimal average of pore sizes is around 5.6. These optimal values match with the physical need of having pores of the size of a xenon to be more attractive to it, which was identified in several papers in the literature. We can note that these values are a bit higher than the ones mentioned in the literature due to the different definition of the atom radii.[44] Moreover, values of "delta_G0_298_900" between 4 and $6\,\mathrm{kJ\,mol^{-1}}$ (see Figure 4.4) have a higher chance of giving a negative contribution, which means a lower ambient-pressure selectivity.

These sweet spots constitute valuable hints to tell the truly selective materials from the others. Some SHAP dependence plots have a rather linear domain for the most selective structures (in blue) — the difference of pore volumes between Xe and Kr sized probes "delta_VF_18_20" have a good linear contribution (see Figure 4.4), which means that the lower the more selective the structure will be. The same can be said for the standard deviations of the PSD "pore_dist_std" and of the Boltzmann weighted krypton interaction energies distribution "enthalpy_std_krypton". The optimal values for these descriptors are zero, the closest to zero it is the more negative the contribution will be and the more selective the structure at ambient pressure.

Sometimes the optimal values are not around well identified values but are contained within larger domains with threshold values separating them. For instance, the difference between the LCD and the average pore size "delta_pore" has a threshold value around 0.3 Å below which the contribution for the most selective structures (blue) is negative (see Figure 4.4); even though no clear correlations can be found, we can at least find a threshold value (about 0.23) below which there is higher probability of having a high ambient-pressure selectivity. The same type of domain splits can be found for the average of krypton interaction energies distribution "mean_grid_krypton" (at around 15), the Boltzmann weighted xenon interaction energies distribution "enthalpy_std_xenon" (at around 2.5), the difference of exchange entropic term between the ambient temperature "delta_TS0_298_900" (at around 3) and high temperature and the effective number associated to the PSD "pore_dist_neff" (at around 2.3). These domains separate structures that are selective at low pressure, which is key to telling apart the structures with a selectivity drop at ambient pressure from the ones without.

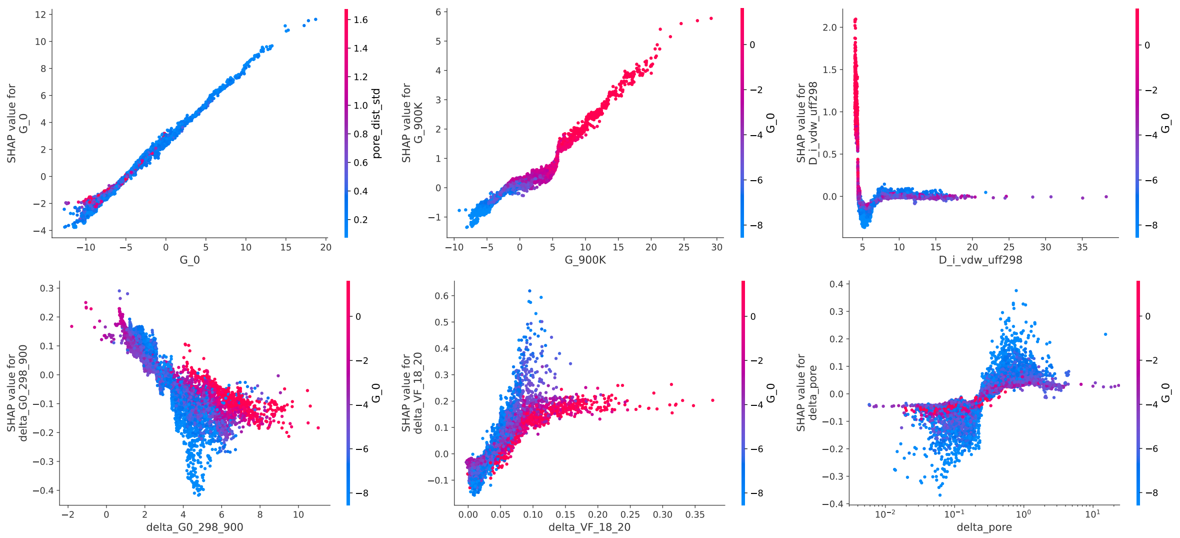


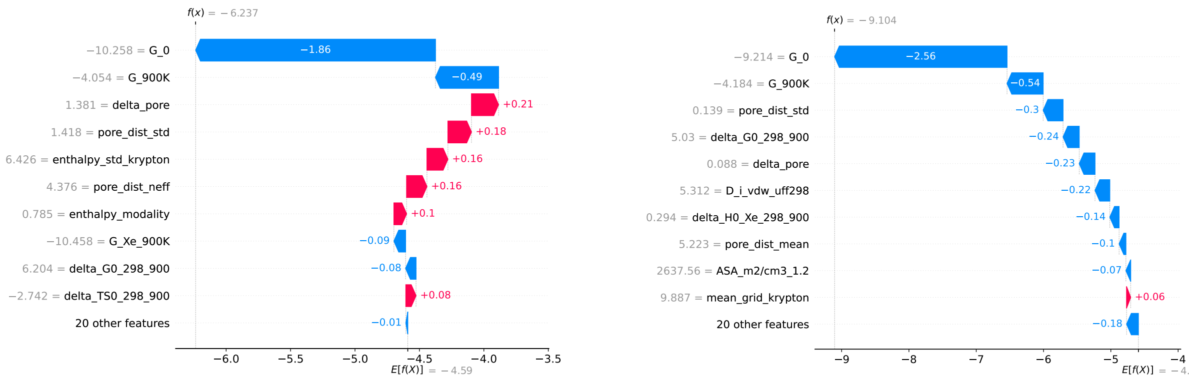
Figure 4.4: Some SHAP dependence plots that are analyzed in the main article. The 18 top descriptors' SDPs can be found in the SI.

Local interpretability

To put into practice our previous analysis, let's look at some archetypal structures and how the model predicted the selectivity based on the descriptor values. We chose two MOF structures from the test set, their CSD code being respectively VIWMIZ and BIMDIL. Both structures are selective at low pressure but the first one decreases in selectivity while the other maintains it at ambient pressure. It will be interesting to see what the model does to tell apart these two completely different behaviors.

VIWMIZ is part of the highly selective structures that experience a selectivity drop at ambient pressure. If we convert back the free energy values to selectivity values, its selectivity is 62.8 at infinite dilution and 14.5 at ambient pressure. The ML model manages to give a close prediction of 12.0 for the ambient-pressure selectivity based on the given values of the descriptors. If we only look at "G_0", it has one of the most negative values, which explains the rather high negative contribution of −1.81. However, the −0.57 contribution of "G_900K" is rather low compared to other materials (see Figure 4.4), since a value of −4.05 is not the most negative considering all structures. On the other hand, the remaining descriptors have values in the domain of positive contributions, which lead to the drop of the selectivity. For example, the difference of pore sizes "delta_pore" has a value of 1.38 Å (above the threshold of 0.23 Å), which contributes +0.25 to the predicted selectivity and is consistent with the value ranges of the associated dependence plot. By reporting the values to the dependence plots, the same analyses can be made on the other positive contributions of the Figure 4.5: "pore_dist_std" is above the threshold of 0.4, "enthalpy_std_krypton" is above 2.5 kJ mol^{-1}, "pore_dist_neff" is above 2.3, "delta_TS0_298_900" is below 3 kJ mol^{-1} and "enthalpy_modality" is around 0.75 where positive contributions are more commonly observed. However, the "delta_G0_298_900" value is a bit too close to its optimal value, which explains its negative contribution in this particular prediction. The rest of the features have almost negligible contributions and are detailed in the Figure S11. By analyzing the contributions of each descriptor to the prediction given by our model, we can understand the underlying features of the VIWMIZ structure that explains the selectivity drop at higher pressure. The shape of the xenon and krypton energy distributions ("enthalpy_std_krypton" and "enthalpy_modality") and of the PSD ("pore_dist_std" and "pore_dist_neff") as well as the void fraction difference "delta_pore" are key descriptors at the origin of the lower selectivity at ambient pressure compared to the ideal infinite dilution case. Intuitively, one can easily understand that effective number of pores exceeding 2 can mean the presence of different pore sizes, which is consistent with the presence of pores that are less attractive to the xenon and leads necessarily to less selectivity. The previous statement is also very much consistent with a high standard deviation of the PSD or the Boltzmann weighted krypton interaction energy distribution. One can also conceive that a much larger difference between the average pore size and the LCD could mean a high disparity in pore sizes that leads to the presence of larger pores more and more loaded as the pressure rises. The entropic term is however way more complex to interpret and opens unexplored ways of tackling the problem of selectivity drop at higher pressure unraveled by our previous study[18].

The second structure BIMDIL is also among the most selective with a selectivity at low pressure of 41.0, while maintaining it to 41.2 at ambient pressure. The model manages to predict this stability of the selectivity by giving a value of 40.0. Consequently, the first contribution of "G_0" is among the most negative ones and set a baseline of −2.4 for the upcoming contributions. The contributions of "G_900K" and "G_900K" are not the highest possible but they continue to lower down the value of the predicted selectivity. It is the joint contributions of the other descriptors that will really discriminate between the two structures and decide why this one will keep its selectivity. Unlike the previously analyzed structure, this one has a "delta_pore" value below 0.3 Å, which explains the negative Shapley value it has for our prediction. The contribution of "delta_G0_298_900" that was only a little negative for the other one, is now playing a major role since it is right within the range of between 4 and 6 kJ mol^{-1} (see Figure 4.5). We can also verify that "pore_dist_std" is now below the threshold instead of being above for the other



(a) VIWMIZ: true $\Delta_{exc}G_1 = -6.63$ *(b) BIMDIL: true* $\Delta_{exc}G_1 = -9.20$

Figure 4.5: Main contributions of the descriptors on the selectivity prediction of two archetypal examples. The descriptor labels used are detailed in the Table S1 and S2 of the SI.

structure. We can confirm that the other contributions are also following the rules implied by the SHAP dependence plots, no apparent anomalies are detected, and the joint efforts of all the descriptors tend to give a lower free energy value, which lead to the conservation of the selectivity value at higher pressure. The set of descriptor values is clearly very different from the previous structure, many values are in opposite contribution domains, which explains how the model manages to disentangle the highly selective structures to find out the ones that would keep their selectivity at higher pressure.

These two examples allow us to understand a bit more how the model tells apart the structures that will lose selectivity at higher pressure from the ones that will not. Most of the dependence plots can give very strong association between the descriptors and their effects; the outliers are rare enough that the inner logic of our model can be understood. As developed previously, the first three descriptors set a baseline on few information on the eventual drop of selectivity; then the other descriptors contribution are either positive, negligible or negative depending on the domain of values the descriptor is in. For instance, the average pore size and the largest cavity diameter need to be around very specific values to maximize the chance of keeping the selectivity at higher pressure, which was highlighted by previous works that emphasize on the importance of pore sizes close to the size of xenon for Xe/Kr separation.[14, 18] The difference of entropy between the ambient temperature and 900 K is surprising descriptor that separates selective structures depending on whether its value is within a given range. The difference of void fraction occupied by xenon and krypton is also very interesting since it affects the selectivity differently depending on whether it is highly selective or not, and the contribution is more or less proportional to its value. Different ways of measuring the disparity of the PSD and interaction energy distribution are key in sorting highly selective structures (in blue on the dependence plot Figure 4.4) between the ones maintaining their performance and the ones decreasing in selectivity. Among others, we can find the difference between the average pore size and the LCD, as well as the standard deviation of the PSD or of the Boltzmann weighted energy distribution that would behave very differently according to the domain in which the value lies. The SHAP dependence plots, partially plotted in the main text and entirely available in the SI, are very valuable reading grid to understand the mechanisms behind our ML model and more broadly to what it understood from the origins of Xe/Kr separation.

4.3 Conclusions and perspectives

In order to better understand separation processes inside nanoporous materials, we performed a machine learning prediction of Xe/Kr ambient-pressure selectivity that is faster than standard GCMC calculations. For MOF structures of the CoRE MOF 2019 database, a xenon/krypton selectivity evaluation would take less than a minute, while an equivalent GCMC calculation takes around 40 min. Unlike most of the selectivity predictions of the literature, we chose to predict a selectivity in the logarithmic scale, because it focuses more on the order magnitude than the exact value of the selectivity of highly selective materials. Moreover, the conversion to an exchange Gibbs free energy allows a more thermodynamic approach based on enthalpy, entropy and free energy values. The challenge was then to predict a free energy equivalent of the ambient-pressure selectivity by using the low-pressure selectivity along with key energy, geometrical and chemical descriptors. The final, fully-optimized ML model performs very well with an RMSE of 0.36 $\mathrm{kJ\,mol^{-1}}$, which corresponds to a 0.06 RMSE on the base-10 log of the selectivity.

One of our more specific goals was to uncover underlying reasons of a selectivity drop at high pressure observed on some highly selective materials at low pressure. Previous studies found that a high diversity of pore sizes and channel sizes that favor adsorbate reorganizations could be at the origin of this phenomenon.[18] By applying interpretability tools, we found quantitative factors that explain the conservation or the drop of the selectivity for highly selective materials. Depending on energy averaging at 900 K, on statistical characterizations of the energy or pore size distributions, and on the difference of volumes occupiable we have a structure either with a selectivity similar to the low-pressure case or that is less selective at higher pressure. All the quantitative rules are contained in a complex ensemble of decision trees constructed by our XGBoost model, and they can be extracted to build rule of thumbs in order to back our intuition on the Xe/Kr selectivity in MOF structures.

The final ML model can be used in a well-designed workflow to find the best performing materials. For instance, we could filter out the structures with pores that cannot fit a xenon in, then we could use a first calculation of the low-pressure selectivity to filter out the selectivities below a given threshold, finally we can use the model to remove the structures that would experience a selectivity drop. We tested our methodology on the Xe/Kr separation as proof of concept since it is one of the simplest adsorption system (monoatomic species with no electrostatic interactions). A similar approach can be generalized to other separation applications by calculating the infinite dilution energies with a more standard method (*e.g.* Widom's insertion) and by adapting the descriptor definitions to fit the adsorbates of interest.

This study ambitions to add new descriptor ideas to help the development of ever more efficient screening methodologies to find the best materials for target applications. However, like many other studies on the topic, this one also relies on a few strong assumptions — the simulations are performed in rigid frameworks with non-polarized classical force fields. As suggested in the literature, the most selective materials ever synthesized for Xe/Kr separation are all based on the effect of open-metal sites that uses the difference of polarizability between the two molecules to efficiently separate them.[5, 6] Moreover, the structures can be made flexible using flexible force fields with adapted simulation methodologies[251] or by using multiple rigid simulations of snapshots from NPT simulations[159]. It would be possible to improve the simulations at the cost of CPU times, if we coupled it with a reduction of simulation time like the one presented in this article. The quest of ever faster evaluation tools will allow us to investigate more complex properties and uncover structures with ever more interesting characteristics.

5

XENON AND KRYPTON TRANSPORT PROPERTIES

5.1 Current state of the art

Experiment? [reprendre la review daglar]

5.1.1 Molecular dynamics

5.1.2 Transition state theory

[what is a transition state for diffusion?]

[how to detect TS]

Fast kinetic Monte Carlo tutrast[125]

5.2 ML modeling

Results

5.3 Transport properties of xenon/krypton separation

5.3.1 Why studying diffusion for xenon krypton

[kinetic KAXQIL]

5.3.2 Correlations

5.4 Fast diffusion calculation algorithm

5.4.1 Implementation in C++

5.4.2 Preliminary results

5.4.3 Visualization tool

[Take some examples for the vizualisation with comparison to the pore size and diffusion coefficient]

5.4.4 ML model training

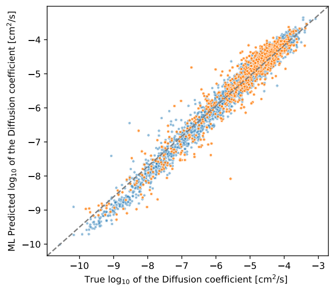


Figure 5.1

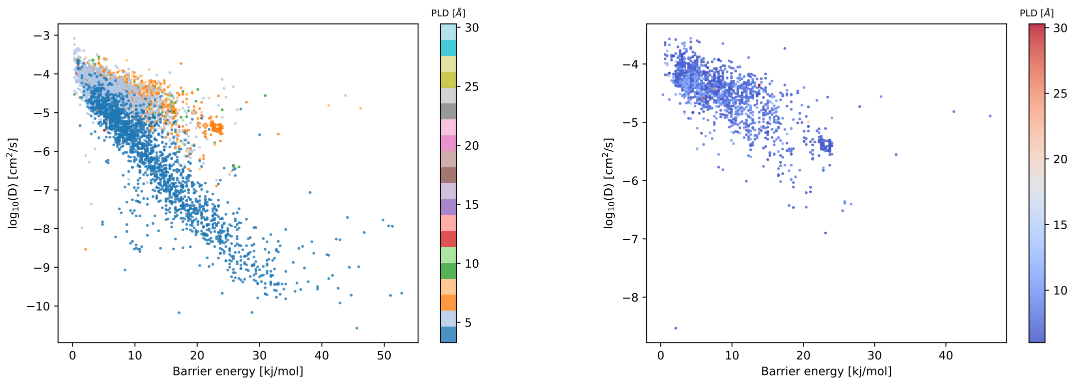


Figure 5.2

ML descriptors next steps

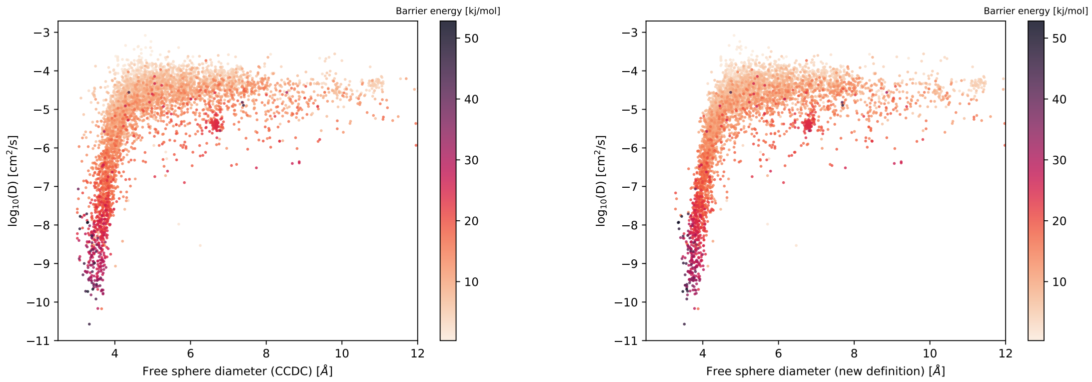

Barrier energy [kj/mol]
log10(D) [cm2/s]
Free sphere diameter (CCDC) [Å]
Barrier energy [kj/mol]
log10(D) [cm2/s]
Free sphere diameter (new definition) [Å]

Figure 5.3

6

TOWARDS THE NEXT GENERATION OF SCREENINGS

6.1 Flexibility

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

6.1.1 Problem, literature

6.1.2 Database approach

6.1.3 Perspectives: Snapshot method

6.2 Noble Gas Polarizability

6.2.1 Problem definition

Best materials use polarization effects [5, 6]

Talk about the order of magnitude of the different interactions > charge–(induced dipole high magnitude)

Standard methods failing to describe oms[195]

[faire référence à 2-thermo partie sur les interactions2.1.2]

6.2.2 Studying the polarization

Inspired by works on the subject[193, 194]

[essayer d'ajouter la polarisabilité pour PEI et al. et Li et al.]

Xe/Kr difference of polarisability Open Metal Sites/polar groups [20220421_pres]

[Not the best material, but interesting discussion on open metal site effect] Tao et al.[252] looked at tuning (and improving) the selective adsorption of Xe over Kr by MOF open metal sites in the UTSA-74 framework structure.

6.2.3 Perpectives

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.
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.
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.

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.

[3] D. Banerjee, C. M. Simon, S. K. Elsaidi, 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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[13] E. Ren, P. Guilbaud, 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.

[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.

[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.

[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.

[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.

[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.

[19] J.-R. Li, R. J. Kuppler, and H.-C. Zhou. "Selective gas adsorption and separation in metal–organic frameworks". In: *Chem. Soc. Rev.* 38.5 (2009), p. 1477. DOI: 10.1039/b802426j.

[20] S. Lagorsse, F. Magalhães, and A. Mendes. "Xenon recycling in an anaesthetic closed-system using carbon molecular sieve membranes". In: *Journal of Membrane Science* 301.1-2 (Sept. 2007), pp. 29–38. DOI: 10.1016/j.memsci.2007.05.032.

[21] R. Morris and P. Wheatley. "Gas Storage in Nanoporous Materials". In: *Angew. Chem. Int. Ed.* 47.27 (June 2008), pp. 4966–4981. DOI: 10.1002/anie.200703934.

[22] T. Qiu, Z. Liang, W. Guo, H. Tabassum, S. Gao, and R. Zou. "Metal–Organic Framework-Based Materials for Energy Conversion and Storage". In: *ACS Energy Letters* 5.2 (Jan. 2020), pp. 520–532. DOI: 10.1021/acsenergylett.9b02625.

[23] A. T. Bell. "The Impact of Nanoscience on Heterogeneous Catalysis". In: *Science* 299.5613 (Mar. 2003), pp. 1688–1691. DOI: 10.1126/science.1083671.

[24] B. K. Singh, S. Lee, and K. Na. "An overview on metal-related catalysts: metal oxides, nanoporous metals and supported metal nanoparticles on metal organic frameworks and zeolites". In: *Rare Met.* 39.7 (Feb. 2019), pp. 751–766. DOI: 10.1007/s12598-019-01205-6.

[25] V. Pascanu, G. G. Miera, A. K. Inge, and B. Martín-Matute. "Metal–Organic Frameworks as Catalysts for Organic Synthesis: A Critical Perspective". In: *J. Am. Chem. Soc.* 141.18 (Apr. 2019), pp. 7223–7234. DOI: 10.1021/jacs.9b00733.

[26] J. D. Rocca, D. Liu, and W. Lin. "Nanoscale Metal–Organic Frameworks for Biomedical Imaging and Drug Delivery". In: *Acc. Chem. Res.* 44.10 (June 2011), pp. 957–968. DOI: 10.1021/ar200028a.

[27] M. C. Bernini, D. Fairen-Jimenez, M. Pasinetti, A. J. Ramirez-Pastor, and R. Q. Snurr. "Screening of bio-compatible metal–organic frameworks as potential drug carriers

using Monte Carlo simulations". In: *J. Mater. Chem. B* 2.7 (2014), pp. 766–774. DOI: 10.1039/c3tb21328e.

[28] A. Breslin. *Guidance for air sampling at nuclear facilities. [Radiation monitoring].* Tech. rep. Nov. 1976. DOI: 10.2172/7326039.

[29] Z. Yan, Y. Gong, C.-T. Yang, X. Wu, B. Liu, Q. Liu, S. Xiong, and S. Peng. "Pore Size Reduction by Methyl Function in Aluminum-Based Metal–Organic Frameworks for Xenon/Krypton Separation". In: *Crystal Growth & Design* 20.12 (Nov. 2020), pp. 8039–8046. DOI: 10.1021/acs.cgd.0c01283.

[30] J. Rouquerol, D. Avnir, C. W. Fairbridge, D. H. Everett, J. M. Haynes, N. Pernicone, J. D. F. Ramsay, K. S. W. Sing, and K. K. Unger. "Recommendations for the characterization of porous solids (Technical Report)". In: *Pure and Applied Chemistry* 66.8 (Jan. 1994), pp. 1739–1758. DOI: 10.1351/pac199466081739.

[31] C. Wang, C. Li, E. R. C. Rutledge, S. Che, J. Lee, A. J. Kalin, C. Zhang, H.-C. Zhou, Z.-H. Guo, and L. Fang. "Aromatic porous polymer network membranes for organic solvent nanofiltration under extreme conditions". In: *J. Mater. Chem. A* 8.31 (2020), pp. 15891–15899. DOI: 10.1039/c9ta10190j.

[32] D. Nicholson and T. Stubos. "Simulation of Adsorption in Micropores". In: *Membrane Science and Technology*. Elsevier, 2000, pp. 231–256. DOI: 10.1016/s0927-5193(00)80011-x.

[33] J. A. Dunne, R. Mariwala, M. Rao, S. Sircar, R. J. Gorte, and A. L. Myers. "Calorimetric Heats of Adsorption and Adsorption Isotherms. 1. O_2, N_2, Ar, CO_2, CH_4, C_2H_6, and SF_6 on Silicalite". In: *Langmuir* 12.24 (Jan. 1996), pp. 5888–5895. DOI: 10.1021/la960495z.

[34] V. Finsy, S. D. Bruyne, L. Alaerts, D. D. Vos, P. A. Jacobs, G. V. Baron, and J. F. Denayer. "Shape selective adsorption of linear and branched alkanes in the Cu3(BTC)2 metal-organic framework". In: *From Zeolites to Porous MOF Materials - The 40th Anniversary of International Zeolite Conference, Proceedings of the 15th International Zeolite Conference.* Elsevier, 2007, pp. 2048–2053. DOI: 10.1016/s0167-2991(07)81098-6.

[35] K. S. W. Sing. "Reporting physisorption data for gas/solid systems with special reference to the determination of surface area and porosity (Recommendations 1984)". In: *Pure and Applied Chemistry* 57.4 (Jan. 1985), pp. 603–619. DOI: 10.1351/pac198557040603.

[36] J. M. Haynes. "Stereological analysis of pore structure". In: *Matériaux et Constructions* 6.3 (May 1973), pp. 175–179. DOI: 10.1007/bf02479030.

[37] A. Radlinski, M. Mastalerz, A. Hinde, M. Hainbuchner, H. Rauch, M. Baron, J. Lin, L. Fan, and P. Thiyagarajan. "Application of SAXS and SANS in evaluation of porosity, pore size distribution and surface area of coal". In: *International Journal of Coal Geology* 59.3-4 (Aug. 2004), pp. 245–271. DOI: 10.1016/j.coal.2004.03.002.

[38] O. K. Farha, I. Eryazici, N. C. Jeong, B. G. Hauser, C. E. Wilmer, A. A. Sarjeant, R. Q. Snurr, S. T. Nguyen, A. Ö. Yazaydın, and J. T. Hupp. "Metal–Organic Framework Materials with Ultrahigh Surface Areas: Is the Sky the Limit?" In: *J. Am. Chem. Soc.* 134.36 (Aug. 2012), pp. 15016–15021. DOI: 10.1021/ja3055639.

[39] E. Detsi, E. D. Jong, A. Zinchenko, Z. Vuković, I. Vuković, S. Punzhin, K. Loos, G. ten Brinke, H. D. Raedt, P. Onck, and J. D. Hosson. "On the specific surface area of nanoporous materials". In: *Acta Materialia* 59.20 (Dec. 2011), pp. 7488–7497. DOI: 10.1016/j.actamat.2011.08.025.

[40] Y. Tian and J. Wu. "A comprehensive analysis of the BET area for nanoporous materials". In: *AIChE Journal* 64.1 (Aug. 2017), pp. 286–293. DOI: 10.1002/aic.15880.

[41] J. W. M. Osterrieth et al. "How Reproducible are Surface Areas Calculated from the BET Equation?" In: *Advanced Materials* 34.27 (May 2022), p. 2201502. DOI: 10.1002/adma.202201502.

[42] 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". In: *Microporous and Mesoporous Materials* 149.1 (Feb. 2012), pp. 134–141. DOI: 10.1016/j.micromeso.2011.08.020.

[43] L. Sarkisov, R. Bueno-Perez, M. Sutharson, and D. Fairen-Jimenez. "Materials Informatics with PoreBlazer v4.0 and the CSD MOF Database". In: *Chem. Mater.* 32.23 (Nov. 2020), pp. 9849–9867. DOI: 10.1021/acs.chemmater.0c03575.

[44] T.-H. Hung, Q. Lyu, L.-C. Lin, and D.-Y. Kang. "Transport-Relevant Pore Limiting Diameter for Molecular Separations in Metal–Organic Framework Membranes". In: *J. Phys. Chem. C* 125.37 (Sept. 2021), pp. 20416–20425. DOI: 10.1021/acs.jpcc.1c05959.

[45] P. A. Vaughan. "The crystal structure of the zeolite ferrierite". In: *Acta Crystallogr* 21.6 (Dec. 1966), pp. 983–990. DOI: 10.1107/s0365110x66004298.

[46] A. P. Côté, A. I. Benin, N. W. Ockwig, M. O'Keeffe, A. J. Matzger, and O. M. Yaghi. "Porous, Crystalline, Covalent Organic Frameworks". In: *Science* 310.5751 (Nov. 2005), pp. 1166–1170. DOI: 10.1126/science.1120411.

[47] D. Banerjee, Z. Zhang, A. M. Plonka, J. Li, and J. B. Parise. "A Calcium Coordination Framework Having Permanent Porosity and High CO_2/N_2 Selectivity". In: *Cryst. Growth Des.* 12 (5 2012), pp. 2162–2165. DOI: 10.1021/cg300274n.

[48] 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.

[49] G. A. Ozin, A. Kuperman, and A. Stein. "Advanced Zeolite, Materials Science". In: *Angew. Chem. Int. Ed. Engl.* 28.3 (Mar. 1989), pp. 359–376. DOI: 10.1002/anie.198903591.

[50] Y. Ma, W. Tong, H. Zhou, and S. L. Suib. "A review of zeolite-like porous materials". In: *Microporous and Mesoporous Materials* 37.1-2 (May 2000), pp. 243–252. DOI: 10.1016/s1387-1811(99)00199-7.

[51] Z. Wang, J. Yu, and R. Xu. "Needs and trends in rational synthesis of zeolitic materials". In: *Chem. Soc. Rev.* 41.5 (2012), pp. 1729–1741. DOI: 10.1039/c1cs15150a.

[52] B. Chen, Z. Yang, Y. Zhu, and Y. Xia. "Zeolitic imidazolate framework materials: recent progress in synthesis and applications". In: *J. Mater. Chem. A* 2.40 (2014), pp. 16811–16831. DOI: 10.1039/c4ta02984d.

[53] W. Lu, D. Yuan, D. Zhao, C. I. Schilling, O. Plietzsch, T. Muller, S. Bräse, J. Guenther, J. Blümel, R. Krishna, Z. Li, and H.-C. Zhou. "Porous Polymer Networks: Synthesis, Porosity, and Applications in Gas Storage/Separation". In: *Chem. Mater.* 22.21 (Oct. 2010), pp. 5964–5972. DOI: 10.1021/cm1021068.

[54] S. Che and L. Fang. "Porous Ladder Polymer Networks". In: *Chem* 6.10 (Oct. 2020), pp. 2558–2590. DOI: 10.1016/j.chempr.2020.08.002.

[55] X. Feng, X. Ding, and D. Jiang. "Covalent organic frameworks". In: *Chem. Soc. Rev.* 41.18 (2012), p. 6010. DOI: 10.1039/c2cs35157a.

[56] O. M. Yaghi and H. Li. “Hydrothermal Synthesis of a Metal-Organic Framework Containing Large Rectangular Channels”. In: *J. Am. Chem. Soc.* 117.41 (Oct. 1995), pp. 10401–10402. DOI: 10.1021/ja00146a033.

[57] B. F. Abrahams, B. F. Hoskins, and R. Robson. “A new type of infinite 3D polymeric network containing 4-connected, peripherally-linked metalloporphyrin building blocks”. In: *J. Am. Chem. Soc.* 113.9 (Apr. 1991), pp. 3606–3607. DOI: 10.1021/ja00009a065.

[58] R. J. Kuppler, D. J. Timmons, Q.-R. Fang, J.-R. Li, T. A. Makal, M. D. Young, D. Yuan, D. Zhao, W. Zhuang, and H.-C. Zhou. “Potential applications of metal-organic frameworks”. In: *Coordination Chemistry Reviews* 253.23-24 (Dec. 2009), pp. 3042–3066. DOI: 10.1016/j.ccr.2009.05.019.

[59] H. Furukawa, K. E. Cordova, M. O’Keeffe, and O. M. Yaghi. “The Chemistry and Applications of Metal-Organic Frameworks”. In: *Science* 341.6149 (Aug. 2013). DOI: 10.1126/science.1230444.

[60] A. Ejsmont, J. Andreo, A. Lanza, A. Galarda, L. Macreadie, S. Wuttke, S. Canossa, E. Ploetz, and J. Goscianska. “Applications of reticular diversity in metal–organic frameworks: An ever-evolving state of the art”. In: *Coordination Chemistry Reviews* 430 (Mar. 2021), p. 213655. DOI: 10.1016/j.ccr.2020.213655.

[61] 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”. In: *Energy Environ. Sci.* 8 (4 2015), pp. 1190–1199. DOI: 10.1039/C4EE03515A.

[62] 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”. In: *Chemistry of Materials* 29.7 (Mar. 2017), pp. 2844–2854. DOI: 10.1021/acs.chemmater.6b04933.

[63] 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”. In: *Chemistry of Materials* 26.21 (Oct. 2014), pp. 6185–6192. DOI: 10.1021/cm502594j.

[64] 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.

[65] 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.

[66] 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.

[67] 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.

[68] 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.

[69] 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.

[70] 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.

[71] 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.

[72] 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.

[73] 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.

[74] T. Kalil and C. Wadia. *Materials Genome Initiative for Global Competitiveness*. Washington, 2011.

[75] *The Materials Genome Initiative*. Available online at https://www.mgi.gov/. 2022.

[76] 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 Materials* 1.1 (July 2013), p. 011002. DOI: 10.1063/1.4812323.

[77] 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.

[78] 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.

[79] 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”. In: *Energy & Environmental Science* 9.10 (2016), pp. 3279–3289. DOI: 10.1039/c6ee02104b.

[80] 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.

[81] 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.

[82] 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". In: *The Journal of Physical Chemistry C* 117.15 (Apr. 2013), pp. 7681–7689. DOI: 10.1021/jp4006422.

[83] 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". In: *The Journal of Physical Chemistry C* 117.27 (July 2013), pp. 14095–14105. DOI: 10.1021/jp404287t.

[84] H. Dureckova, M. Krykunov, M. Z. Aghaji, and T. K. Woo. "Robust Machine Learning Models for Predicting High CO2 Working Capacity and CO2/H2 Selectivity of Gas Adsorption in Metal Organic Frameworks for Precombustion Carbon Capture". In: *The Journal of Physical Chemistry C* 123.7 (Jan. 2019), pp. 4133–4139. DOI: 10.1021/acs.jpcc.8b10644.

[85] L. Yang, C. Shi, L. Li, and Y. Li. "High-throughput model-building and screening of zeolitic imidazolate frameworks for CO2 capture from flue gas". In: *Chinese Chemical Letters* 31.1 (Jan. 2020), pp. 227–230. DOI: 10.1016/j.cclet.2019.04.025.

[86] Z. Qiao, C. Peng, J. Zhou, and J. Jiang. "High-throughput computational screening of 137953 metal–organic frameworks for membrane separation of a CO2/N2/CH4 mixture". In: *Journal of Materials Chemistry A* 4.41 (2016), pp. 15904–15912. DOI: 10.1039/c6ta06262h.

[87] A. K. Singh, K. Mathew, H. L. Zhuang, and R. G. Hennig. "Computational Screening of 2D Materials for Photocatalysis". In: *J. Phys. Chem. Lett.* 6 (6 2015), pp. 1087–1098. DOI: 10.1021/jz502646d.

[88] J. Greeley, T. F. Jaramillo, J. Bonde, I. Chorkendorff, and J. K. Nørskov. "Computational high-throughput screening of electrocatalytic materials for hydrogen evolution". In: *Nature Mater.* 5 (11 2006), pp. 909–913. DOI: 10.1038/nmat1752.

[89] S. Back, K. Tran, and Z. W. Ulissi. "Discovery of Acid-Stable Oxygen Evolution Catalysts: High-Throughput Computational Screening of Equimolar Bimetallic Oxides". In: *ACS Appl. Mater. Interfaces* 12 (34 2020), pp. 38256–38265. DOI: 10.1021/acsami.0c11821.

[90] S. Chibani and F.-X. Coudert. "Systematic exploration of the mechanical properties of 13 621 inorganic compounds". In: *Chem. Sci.* 10 (37 2019), pp. 8589–8599. DOI: 10.1039/C9SC01682A.

[91] R. Gaillac, S. Chibani, and F.-X. Coudert. "Speeding Up Discovery of Auxetic Zeolite Frameworks by Machine Learning". In: *Chem. Mater.* 32 (6 2020), pp. 2653–2663. DOI: 10.1021/acs.chemmater.0c00434.

[92] C. Toher, J. J. Plata, O. Levy, M. de Jong, M. Asta, M. B. Nardelli, and S. Curtarolo. "High-throughput computational screening of thermal conductivity, Debye temperature, and Grüneisen parameter using a quasiharmonic Debye model". In: *Phys. Rev. B* 90 (17 2014), p. 174107. DOI: 10.1103/PhysRevB.90.174107.

[93] S. Sarikurt, T. Kocabaş, and C. Sevik. "High-throughput computational screening of 2D materials for thermoelectrics". In: *J. Mater. Chem. A* 8 (37 2020), pp. 19674–19683. DOI: 10.1039/D0TA04945J.

[94] M. Ducamp and F.-X. Coudert. "Systematic Study of the Thermal Properties of Zeolitic Frameworks". In: *The Journal of Physical Chemistry C* 125.28 (July 2021), pp. 15647–15658. DOI: 10.1021/acs.jpcc.1c03975.

[95] J. D. Evans and F.-X. Coudert. "Predicting the Mechanical Properties of Zeolite Frameworks by Machine Learning". In: *Chemistry of Materials* 29.18 (Aug. 2017), pp. 7833–7839. DOI: 10.1021/acs.chemmater.7b02532.

[96] M. Ducamp and F.-X. Coudert. "Prediction of Thermal Properties of Zeolites through Machine Learning". In: *J. Phys. Chem. C* 126 (3 2022), pp. 1651–1660. DOI: 10.1021/acs.jpcc.1c09737.

[97] M. Eckhoff and J. Behler. "From Molecular Fragments to the Bulk: Development of a Neural Network Potential for MOF-5". In: *Journal of Chemical Theory and Computation* 15.6 (May 2019), pp. 3793–3809. DOI: 10.1021/acs.jctc.8b01288.

[98] P. Friederich, F. Häse, J. Proppe, and A. Aspuru-Guzik. "Machine-learned potentials for next-generation matter simulations". In: *Nature Materials* 20.6 (May 2021), pp. 750–761. DOI: 10.1038/s41563-020-0777-6.

[99] G. A. McCarver, T. Rajeshkumar, and K. D. Vogiatzis. "Computational catalysis for metal-organic frameworks: An overview". In: *Coordination Chemistry Reviews* 436 (June 2021), p. 213777. DOI: 10.1016/j.ccr.2021.213777.

[100] K. D. Vogiatzis, E. Haldoupis, D. J. Xiao, J. R. Long, J. I. Siepmann, and L. Gagliardi. "Accelerated Computational Analysis of Metal–Organic Frameworks for Oxidation Catalysis". In: *The Journal of Physical Chemistry C* 120.33 (Aug. 2016), pp. 18707–18712. DOI: 10.1021/acs.jpcc.6b07115.

[101] A. S. Rosen, J. M. Notestein, and R. Q. Snurr. "Structure–Activity Relationships That Identify Metal–Organic Framework Catalysts for Methane Activation". In: *ACS Catalysis* 9.4 (Mar. 2019), pp. 3576–3587. DOI: 10.1021/acscatal.8b05178.

[102] K. T. Butler, C. H. Hendon, and A. Walsh. "Electronic Chemical Potentials of Porous Metal–Organic Frameworks". In: *Journal of the American Chemical Society* 136.7 (Feb. 2014), pp. 2703–2706. DOI: 10.1021/ja4110073.

[103] 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". In: *Journal of Computational Chemistry* 40.12 (Feb. 2019), pp. 1305–1318. DOI: 10.1002/jcc.25787.

[104] M. Fumanal, G. Capano, S. Barthel, B. Smit, and I. Tavernelli. "Energy-based descriptors for photo-catalytically active metal–organic framework discovery". In: *Journal of Materials Chemistry A* 8.8 (2020), pp. 4473–4482. DOI: 10.1039/c9ta13506e.

[105] 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". In: *npj Computational Materials* 8.1 (May 2022). DOI: 10.1038/s41524-022-00796-6.

[106] 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". In: *Matter* 4.5 (May 2021), pp. 1578–1597. DOI: 10.1016/j.matt.2021.02.015.

[107] A. S. Rosen, J. M. Notestein, and R. Q. Snurr. "Realizing the data-driven, computational discovery of metal-organic framework catalysts". In: *Current Opinion in Chemical Engineering* 35 (Mar. 2022), p. 100760. DOI: 10.1016/j.coche.2021.100760.

[108] A. Marmier, Z. A. Lethbridge, R. I. Walton, C. W. Smith, S. C. Parker, and K. E. Evans. "ElAM: A computer program for the analysis and representation of anisotropic elastic properties". In: *Computer Physics Communications* 181 (12 2010), pp. 2102–2115. DOI: 10.1016/j.cpc.2010.08.033.

[109] R. Golesorkhtabar, P. Pavone, J. Spitaler, P. Puschnig, and C. Draxl. "ElaStic: A tool for calculating second-order elastic constants from first principles". In: *Computer Physics Communications* 184 (8 2013), pp. 1861–1873. DOI: 10.1016/j.cpc.2013.03.010.

[110] R. Dovesi, A. Erba, R. Orlando, C. M. Zicovich-Wilson, B. Civalleri, L. Maschio, M. Rérat, S. Casassa, J. Baima, S. Salustro, and B. Kirtman. "Quantum-mechanical condensed matter simulations with CRYSTAL". In: *WIREs Comput Mol Sci* 8 (4 2018), p. 171. DOI: 10.1002/wcms.1360.

[111] F.-X. Coudert. "Systematic investigation of the mechanical properties of pure silica zeolites: stiffness, anisotropy, and negative linear compressibility". In: *Phys. Chem. Chem. Phys.* 15 (38 2013), p. 16012. DOI: 10.1039/c3cp51817e.

[112] M. de Jong, W. Chen, T. Angsten, A. Jain, R. Notestine, A. Gamst, M. Sluiter, C. Krishna Ande, S. van der Zwaag, J. J. Plata, C. Toher, S. Curtarolo, G. Ceder, K. A. Persson, and M. Asta. "Charting the complete elastic properties of inorganic crystalline compounds". In: *Sci Data* 2 (1 2015), p. 345. DOI: 10.1038/sdata.2015.9.

[113] *The Materials Project.* Available online at https://materialsproject.org/. 2022.

[114] F.-X. Coudert and J. D. Evans. "Nanoscale metamaterials: Meta-MOFs and framework materials with anomalous behavior". In: *Coordination Chemistry Reviews* 388 (2019), pp. 48–62. DOI: 10.1016/j.ccr.2019.02.023.

[115] J. U. Surjadi, L. Gao, H. Du, X. Li, X. Xiong, N. X. Fang, and Y. Lu. "Mechanical Metamaterials and Their Engineering Applications". In: *Adv. Eng. Mater.* 21 (3 2018), p. 1800864. DOI: 10.1002/adem.201800864.

[116] J. Dagdelen, J. Montoya, M. de Jong, and K. Persson. "Computational prediction of new auxetic materials". In: *Nat Commun* 8 (1 2017), p. 124. DOI: 10.1038/s41467-017-00399-6.

[117] P. Z. Moghadam, S. M. Rogge, A. Li, C.-M. Chow, J. Wieme, N. Moharrami, M. Aragones-Anglada, G. Conduit, D. A. Gomez-Gualdron, V. V. Speybroeck, and D. Fairen-Jimenez. "Structure-Mechanical Stability Relations of Metal-Organic Frameworks via Machine Learning". In: *Matter* 1.1 (July 2019), pp. 219–234. DOI: 10.1016/j.matt.2019.03.002.

[118] M. de Jong, W. Chen, H. Geerlings, M. Asta, and K. A. Persson. "A database to enable discovery and design of piezoelectric materials". In: *Sci Data* 2 (1 2015), p. 746. DOI: 10.1038/sdata.2015.53.

[119] A. Togo, L. Chaput, I. Tanaka, and G. Hug. "First-principles phonon calculations of thermal expansion in Ti_3SiC_2, Ti_3AlC_2, and Ti_3GeC_2". In: *Phys. Rev. B* 81 (2010), p. 174301. DOI: 10.1103/PhysRevB.81.174301.

[120] A. Togo and I. Tanaka. "First principles phonon calculations in materials science". In: *Scripta Materialia* 108 (2015), pp. 1–5. DOI: 10.1016/j.scriptamat.2015.07.021.

[121] E. S. Miandoab, S. H. Mousavi, S. E. Kentish, and C. A. Scholes. "Xenon and Krypton separation by membranes at sub-ambient temperatures and its comparison with cryogenic distillation". In: *Separation and Purification Technology* 262 (May 2021), p. 118349. DOI: 10.1016/j.seppur.2021.118349.

[122] T. Watanabe and D. S. Sholl. "Accelerating Applications of Metal–Organic Frameworks for Gas Adsorption and Separation by Computational Screening of Materials". In: *Langmuir* 28.40 (July 2012), pp. 14114–14128. DOI: 10.1021/la301915s.

[123] E. Haldoupis, S. Nair, and D. S. Sholl. "Efficient Calculation of Diffusion Limitations in Metal Organic Framework Materials: A Tool for Identifying Materials for Kinetic Separations". In: *Journal of the American Chemical Society* 132.21 (May 2010), pp. 7528–7539. DOI: 10.1021/ja1023699.

[124] J. Kim, M. Abouelnasr, L.-C. Lin, and B. Smit. "Large-Scale Screening of Zeolite Structures for CO2 Membrane Separations". In: *Journal of the American Chemical Society* 135.20 (May 2013), pp. 7545–7552. DOI: 10.1021/ja400267g.

[125] A. Mace, S. Barthel, and B. Smit. "Automated Multiscale Approach To Predict Self-Diffusion from a Potential Energy Field". In: *Journal of Chemical Theory and Computation* 15.4 (Feb. 2019), pp. 2127–2141. DOI: 10.1021/acs.jctc.8b01255.

[126] M. Zhou and J. Wu. "Massively Parallel GPU-Accelerated String Method for Fast and Accurate Prediction of Molecular Diffusivity in Nanoporous Materials". In: *ACS Applied Nano Materials* 4.5 (May 2021), pp. 5394–5403. DOI: 10.1021/acsanm.1c00727.

[127] B. C. Bukowski, F. J. Keil, P. I. Ravikovitch, G. Sastre, R. Q. Snurr, and M.-O. Coppens. "Connecting theory and simulation with experiment for the study of diffusion in nanoporous solids". In: *Adsorption* 27.5 (Apr. 2021), pp. 683–760. DOI: 10.1007/s10450-021-00314-y.

[128] D. M. Ruthven and S. Farooq. "Air separation by pressure swing adsorption". In: *Gas Separation & Purification* 4.3 (1990), pp. 141–148. DOI: 10.1016/0950-4214(90)80016-E.

[129] C. R. Reid and K. M. Thomas. "Adsorption of Gases on a Carbon Molecular Sieve Used for Air Separation: Linear Adsorptives as Probes for Kinetic Selectivity". In: *Langmuir* 15.9 (Mar. 1999), pp. 3206–3218. DOI: 10.1021/la981289p.

[130] Y. Wang and R. T. Yang. "Chemical Liquid Deposition Modified 4A Zeolite as a Size-Selective Adsorbent for Methane Upgrading, CO2 Capture and Air Separation". In: *ACS Sustainable Chemistry & Engineering* 7.3 (Jan. 2019), pp. 3301–3308. DOI: 10.1021/acssuschemeng.8b05339.

[131] L. M. Robeson. "Correlation of separation factor versus permeability for polymeric membranes". In: *Journal of membrane science* 62.2 (1991), pp. 165–185. DOI: 10.1016/0376-7388(91)80060-J.

[132] Z. Qiao, Q. Xu, and J. Jiang. "High-throughput computational screening of metal-organic framework membranes for upgrading of natural gas". In: *Journal of Membrane Science* 551 (Apr. 2018), pp. 47–54. DOI: 10.1016/j.memsci.2018.01.020.

[133] C. Altintas and S. Keskin. "Molecular Simulations of MOF Membranes and Performance Predictions of MOF/Polymer Mixed Matrix Membranes for CO_2/CH_4 Separations". In: *ACS Sustainable Chemistry & Engineering* 7.2 (Dec. 2018), pp. 2739–2750. DOI: 10.1021/acssuschemeng.8b05832.

[134] S. Budhathoki, O. Ajayi, J. A. Steckel, and C. E. Wilmer. "High-throughput computational prediction of the cost of carbon capture using mixed matrix membranes". In: *Energy & Environmental Science* 12.4 (2019), pp. 1255–1264. DOI: 10.1039/c8ee02582g.

[135] T. Yan, Y. Lan, M. Tong, and C. Zhong. "Screening and Design of Covalent Organic Framework Membranes for CO2/CH4 Separation". In: *ACS Sustainable Chemistry & Engineering* 7.1 (Nov. 2018), pp. 1220–1227. DOI: 10.1021/acssuschemeng.8b04858.

[136] L. M. Robeson. "The upper bound revisited". In: *Journal of Membrane Science* 320 (1-2 2008), pp. 390–400. DOI: 10.1016/j.memsci.2008.04.030.

[137] S. Keskin and D. S. Sholl. "Screening Metal-Organic Framework Materials for Membrane-based Methane/Carbon Dioxide Separations". In: *The Journal of Physical Chemistry C* 111.38 (Aug. 2007), pp. 14055–14059. DOI: 10.1021/jp075290l.

[138] S. Keskin and D. S. Sholl. "Efficient Methods for Screening of Metal Organic Framework Membranes for Gas Separations Using Atomically Detailed Models". In: *Langmuir* 25.19 (July 2009), pp. 11786–11795. DOI: 10.1021/la901438x.

[139] C. Han, Y. Yang, and D. S. Sholl. "Quantitatively Predicting Impact of Structural Flexibility on Molecular Diffusion in Small Pore Metal–Organic Frameworks—A Molecular Dynamics Study of Hypothetical ZIF-8 Polymorphs". In: *The Journal of Physical Chemistry C* 124.37 (Aug. 2020), pp. 20203–20212. DOI: 10.1021/acs.jpcc.0c05942.

[140] 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: *The Journal of Physical Chemistry C* 118.13 (Mar. 2014), pp. 6941–6951. DOI: 10.1021/jp502359q.

[141] 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". In: *ACS Applied Materials & Interfaces* 13.20 (May 2021), pp. 23647–23654. DOI: 10.1021/acsami.1c02471.

[142] N. S. Bobbitt, J. Chen, and R. Q. Snurr. "High-Throughput Screening of Metal–Organic Frameworks for Hydrogen Storage at Cryogenic Temperature". In: *The Journal of Physical Chemistry C* 120.48 (Nov. 2016), pp. 27328–27341. DOI: 10.1021/acs.jpcc.6b08729.

[143] 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". In: *The Journal of Physical Chemistry C* 123.1 (Dec. 2018), pp. 120–130. DOI: 10.1021/acs.jpcc.8b09420.

[144] N. S. Bobbitt and R. Q. Snurr. "Molecular modelling and machine learning for high-throughput screening of metal-organic frameworks for hydrogen storage". In: *Molecular Simulation* 45.14-15 (Apr. 2019), pp. 1069–1081. DOI: 10.1080/08927022.2019.1597271.

[145] J. Liu, P. K. Thallapally, and D. Strachan. "Metal–Organic Frameworks for Removal of Xe and Kr from Nuclear Fuel Reprocessing Plants". In: *Langmuir* 28 (31 2012), pp. 11584–11589. DOI: 10.1021/la301870n.

[146] J. Liu, C. A. Fernandez, P. F. Martin, P. K. Thallapally, and D. M. Strachan. "A Two-Column Method for the Separation of Kr and Xe from Process Off-Gases". In: *Ind. Eng. Chem. Res.* 53 (32 2014), pp. 12893–12899. DOI: 10.1021/ie502156h.

[147] Y.-S. Bae, B. G. Hauser, Y. J. Colón, J. T. Hupp, O. K. Farha, and R. Q. Snurr. "High xenon/krypton selectivity in a metal-organic framework with small pores and strong adsorption sites". In: *Micropor. Mesopor. Mater.* 169 (2013), pp. 176–179. DOI: 10.1016/j.micromeso.2012.11.013.

[148] M. V. Parkes, C. L. Staiger, J. J. Perry IV, M. D. Allendorf, and J. A. Greathouse. "Screening metal–organic frameworks for selective noble gas adsorption in air: effect of pore size and framework topology". In: *Phys. Chem. Chem. Phys.* 15 (23 2013), p. 9093. DOI: 10.1039/c3cp50774b.

[149] S. Xiong, Q. Liu, Q. Wang, W. Li, Y. Tang, X. Wang, S. Hu, and B. Chen. "A flexible zinc tetrazolate framework exhibiting breathing behaviour on xenon adsorption and selective adsorption of xenon over other noble gases". In: *J. Mater. Chem. A* 3 (20 2015), pp. 10747–10752. DOI: 10.1039/C5TA00460H.

[150] S.-J. Lee, T.-U. Yoon, A.-R. Kim, S.-Y. Kim, K.-H. Cho, Y. K. Hwang, J.-W. Yeon, and Y.-S. Bae. "Adsorptive separation of xenon/krypton mixtures using a zirconium-based metal-organic framework with high hydrothermal and radioactive stabilities". In: *J. Hazard. Mater.* 320 (2016), pp. 513–520. DOI: 10.1016/j.jhazmat.2016.08.057.

[151] S.-J. Lee, S. Kim, E.-J. Kim, M. Kim, and Y.-S. Bae. "Adsorptive separation of xenon/krypton mixtures using ligand controls in a zirconium-based metal-organic framework". In: *Chem. Eng. J.* 335 (2018), pp. 345–351. DOI: 10.1016/j.cej.2017.10.155.

[152] H. Wang, K. Yao, Z. Zhang, J. Jagiello, Q. Gong, Y. Han, and J. Li. "The first example of commensurate adsorption of atomic gas in a MOF and effective separation of xenon from other noble gases". In: *Chem. Sci.* 5.2 (2014), pp. 620–624. DOI: 10.1039/c3sc52348a.

[153] R. Anderson, B. Schweitzer, T. Wu, M. A. Carreon, and D. A. Gómez-Gualdrón. "Molecular Simulation Insights on Xe/Kr Separation in a Set of Nanoporous Crystalline Membranes". In: *ACS Appl. Mater. Interfaces* 10 (1 2017), pp. 582–592. DOI: 10.1021/acsami.7b14791.

[154] T. Vazhappilly, T. K. Ghanty, and B. N. Jagatap. "Computational Modeling of Adsorption of Xe and Kr in M-MOF-74 Metal Organic Frame Works with Different Metal Atoms". In: *J. Phys. Chem. C* 120 (20 2016), pp. 10968–10974. DOI: 10.1021/acs.jpcc.6b02782.

[155] P. Zarabadi-Poor and R. Marek. "In Silico Study of (Mn, Fe, Co, Ni, Zn)-BTC Metal–Organic Frameworks for Recovering Xenon from Exhaled Anesthetic Gas". In: *ACS Sustainable Chem. Eng.* 6 (11 2018), pp. 15001–15006. DOI: 10.1021/acssuschemeng.8b03475.

[156] B. J. Sikora, C. E. Wilmer, M. L. Greenfield, and R. Q. Snurr. "Thermodynamic analysis of Xe/Kr selectivity in over 137 000 hypothetical metal–organic frameworks". In: *Chem. Sci.* 3 (7 2012), p. 2217. DOI: 10.1039/c2sc01097f.

[157] P. Ryan, O. K. Farha, L. J. Broadbelt, and R. Q. Snurr. "Computational screening of metal-organic frameworks for xenon/krypton separation". In: *AIChE Journal* 57.7 (Sept. 2010), pp. 1759–1766. DOI: 10.1002/aic.12397.

[158] D. Banerjee, C. M. Simon, A. M. Plonka, R. K. Motkuri, J. Liu, X. Chen, B. Smit, J. B. Parise, M. Haranczyk, and P. K. Thallapally. "Metal–organic framework with optimally selective xenon adsorption and separation". In: *Nature Communications* 7.1 (June 2016). DOI: 10.1038/ncomms11831.

[159] M. Witman, S. Ling, S. Jawahery, P. G. Boyd, M. Haranczyk, B. Slater, and B. Smit. "The Influence of Intrinsic Framework Flexibility on Adsorption in Nanoporous Materials". In: *Journal of the American Chemical Society* 139.15 (Apr. 2017), pp. 5547–5557. DOI: 10.1021/jacs.7b01688.

[160] S. C. Cullen and E. G. Gross. "The anesthetic properties of xenon in animals and human beings, with additional observations on krypton". In: *Science* 113.2942 (1951), pp. 580–582. DOI: 10.1126/science.113.2942.580.

[161] T. F. Holsträter, M. Georgieff, K. J. Föhr, W. Klingler, M. E. Uhl, T. Walker, S. Köster, G. Grön, and O. Adolph. "Intranasal application of xenon reduces opioid requirement and postoperative pain in patients undergoing major abdominal surgery: a randomized

controlled trial". In: *The Journal of the American Society of Anesthesiologists* 115.2 (2011), pp. 398–407. DOI: 10.1097/aln.0b013e318225cee5.

[162] J. G. Mammarappallil, L. Rankine, J. M. Wild, and B. Driehuys. "New Developments in Imaging Idiopathic Pulmonary Fibrosis With Hyperpolarized Xenon Magnetic Resonance Imaging". In: *Journal of Thoracic Imaging* 34.2 (Mar. 2019), pp. 136–150. DOI: 10.1097/rti.0000000000000392.

[163] M. Patterson, J. Foster, T. Haag, V. Rawlin, G. Soulas, and R. Roman. "NEXT: NASA's Evolutionary Xenon Thruster". In: *38th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit*. American Institute of Aeronautics and Astronautics, July 2002. DOI: 10.2514/6.2002-3832.

[164] I. Coxhill and D. Gibbon. "A Xenon Resistojet Propulsion System for Microsatellites". In: *41st AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit*. American Institute of Aeronautics and Astronautics, July 2005. DOI: 10.2514/6.2005-4260.

[165] I. S. Abramov, E. D. Gospodchikov, and A. G. Shalashov. "Extreme-Ultraviolet Light Source for Lithography Based on an Expanding Jet of Dense Xenon Plasma Supported by Microwaves". In: *Phys. Rev. Applied* 10.3 (Sept. 2018). DOI: 10.1103/physrevapplied.10.034065.

[166] F. I. Chang, R. Yeh, G. Lin, P. B. Chu, E. G. Hoffman, E. J. Kruglick, K. S. J. Pister, and M. H. Hecht. "Gas-phase silicon micromachining with xenon difluoride". In: *SPIE Proceedings*. Ed. by W. Bailey, M. E. Motamedi, and F.-C. Luo. SPIE, Sept. 1995. DOI: 10.1117/12.220933.

[167] G. D. Jarman, E. W. R. Barlow, and L. Boersma. "Application of Long-Arc Xenon Lighting for Plant Growth Experiments 1". In: *Agronomy Journal* 66.5 (Sept. 1974), pp. 703–706. DOI: 10.2134/agronj1974.00021962006600050029x.

[168] T. Tanaka, T. Hayashi, T. Nagayama, T. Yanagidaira, and Y. Inui. "Proposal of novel degradation diagnosis method for photovoltaic module employing xenon flash lighting system and detector capacitor". In: *Energy Conversion and Management* 186 (Apr. 2019), pp. 450–461. DOI: 10.1016/j.enconman.2019.02.059.

[169] F. S. D.
bibinitperiod E. Ministery. *Explosion in an Air Distillation Unit.* Accessed: 2023-04-27.

[170] F. S. D.
bibinitperiod E. Ministery. *Explosion in a petrochemical plant.* Accessed: 2023-04-27.

[171] S. M. Auerbach, K. A. Carrado, and P. K. Dutta. *Handbook of zeolite science and technology.* CRC press, 2003.

[172] B. J. Riley, J. McFarlane, G. D. DelCul, J. D. Vienna, C. I. Contescu, and C. W. Forsberg. "Molten salt reactor waste and effluent management strategies: A review". In: *Nuclear Engineering and Design* 345 (Apr. 2019), pp. 94–109. DOI: 10.1016/j.nucengdes.2019.02.002.

[173] J. O. Blomeke and J. J. Perona. *MANAGEMENT OF NOBLE-GAS FISSION-PRODUCT WASTES FROM REPROCESSING SPENT FUELS.* Tech. rep. Jan. 1969. DOI: 10.2172/4731193.

[174] D. LeBlanc. "Molten salt reactors: A new beginning for an old idea". In: *Nuclear Engineering and Design* 240.6 (June 2010), pp. 1644–1656. DOI: 10.1016/j.nucengdes.2009.12.033.

[175] D. Banerjee, A. J. Cairns, J. Liu, R. K. Motkuri, S. K. Nune, C. A. Fernandez, R. Krishna, D. M. Strachan, and P. K. Thallapally. "Potential of Metal–Organic Frameworks for

Separation of Xenon and Krypton". In: *Acc. Chem. Res.* 48.2 (Dec. 2014), pp. 211–219. DOI: 10.1021/ar5003126.

[176] M. H. Mohamed, S. K. Elsaidi, T. Pham, K. A. Forrest, H. T. Schaef, A. Hogan, L. Wojtas, W. Xu, B. Space, M. J. Zaworotko, and P. K. Thallapally. "Hybrid Ultra-Microporous Materials for Selective Xenon Adsorption and Separation". In: *Angew. Chem. Int. Ed.* 55.29 (May 2016), pp. 8285–8289. DOI: 10.1002/anie.201602287.

[177] A. Li, R. Bueno-Perez, D. Madden, and D. Fairen-Jimenez. "From computational high-throughput screenings to the lab: taking metal–organic frameworks out of the computer". In: *Chem. Sci.* 13.27 (2022), pp. 7990–8002. DOI: 10.1039/d2sc01254e.

[178] Y. Kang, H. Park, B. Smit, and J. Kim. "A multi-modal pre-training transformer for universal transfer learning in metal–organic frameworks". In: *Nature Machine Intelligence* 5.3 (Mar. 2023), pp. 309–318. DOI: 10.1038/s42256-023-00628-2.

[179] Z. Cao, R. Magar, Y. Wang, and A. B. Farimani. "MOFormer: Self-Supervised Transformer Model for Metal–Organic Framework Property Prediction". In: *J. Am. Chem. Soc.* 145.5 (Jan. 2023), pp. 2958–2967. DOI: 10.1021/jacs.2c11420.

[180] H. Park, S. Majumdar, X. Zhang, J. Kim, and B. Smit. "Inverse design of metal-organic frameworks for direct air capture of CO2 via deep reinforcement learning". In: (Apr. 2023). DOI: 10.26434/chemrxiv-2023-71mjq-v2.

[181] R. L. Greenaway, V. Santolini, M. J. Bennison, B. M. Alston, C. J. Pugh, M. A. Little, M. Miklitz, E. G. B. Eden-Rump, R. Clowes, A. Shakil, H. J. Cuthbertson, H. Armstrong, M. E. Briggs, K. E. Jelfs, and A. I. Cooper. "High-throughput discovery of organic cages and catenanes using computational screening fused with robotic synthesis". In: *Nature Communications* 9.1 (July 2018). DOI: 10.1038/s41467-018-05271-9.

[182] S. M. Moosavi, A. Chidambaram, L. Talirz, M. Haranczyk, K. C. Stylianou, and B. Smit. "Capturing chemical intuition in synthesis of metal-organic frameworks". In: *Nature Communications* 10.1 (Feb. 2019). DOI: 10.1038/s41467-019-08483-9.

[183] E. L. First and C. A. Floudas. "MOFomics: Computational pore characterization of metal–organic frameworks". In: *Microporous and Mesoporous Materials* 165 (Jan. 2013), pp. 32–39. DOI: 10.1016/j.micromeso.2012.07.049.

[184] A. K. Rappé, C. J. Casewit, K. Colwell, W. A. Goddard III, and W. M. Skiff. "UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations". In: *J. Am. Chem. Soc.* 114.25 (1992), pp. 10024–10035.

[185] A. Shrake and J. Rupley. "Environment and exposure to solvent of protein atoms. Lysozyme and insulin". In: *J. Mol. Bio.* 79.2 (Sept. 1973), pp. 351–371. DOI: 10.1016/0022-2836(73)90011-9.

[186] D. Ongari, P. G. Boyd, S. Barthel, M. Witman, M. Haranczyk, and B. Smit. "Accurate Characterization of the Pore Volume in Microporous Crystalline Materials". In: *Langmuir* 33.51 (July 2017), pp. 14529–14538. DOI: 10.1021/acs.langmuir.7b01682.

[187] W. Keesom. "The second viral coefficient for rigid spherical molecules, whose mutual attraction is equivalent to that of a quadruplet placed at their centre". In: *Proc. R. Acad. Sci* 18 (1915), pp. 636–646.

[188] J. K. Roberts and W. J. C. Orr. "Induced dipoles and the heat of adsorption of argon on ionic crystals". In: *Trans. Faraday Soc.* 34 (1938), p. 1346. DOI: 10.1039/tf9383401346.

[189] F. London. "Zur theorie und systematik der molekularkräfte". In: *Zeitschrift für Physik* 63.3-4 (1930), pp. 245–279.

[190] M. Polanyi. "Section III.—Theories of the adsorption of gases. A general survey and some additional remarks. Introductory paper to section III". In: *Transactions of the Faraday Society* 28 (1932), pp. 316–333.

[191] J. E. Lennard-Jones. "On the determination of molecular fields.—I. From the variation of the viscosity of a gas with temperature". In: *Proc. R. Soc. Lond. A* 106.738 (Oct. 1924), pp. 441–462. DOI: 10.1098/rspa.1924.0081.

[192] J. G. McDaniel, S. Li, E. Tylianakis, R. Q. Snurr, and J. R. Schmidt. "Evaluation of Force Field Performance for High-Throughput Screening of Gas Uptake in Metal–Organic Frameworks". In: *J. Phys. Chem. C* 119.6 (Feb. 2015), pp. 3143–3152. DOI: 10.1021/jp511674w.

[193] V. Lachet, A. Boutin, B. Tavitian, and A. H. Fuchs. "Computational Study of *p*-Xylene/*m*-Xylene Mixtures Adsorbed in NaY Zeolite". In: *J. Phys. Chem. B* 102.46 (Oct. 1998), pp. 9224–9233. DOI: 10.1021/jp980946j.

[194] T. M. Becker, J. Heinen, D. Dubbeldam, L.-C. Lin, and T. J. H. Vlugt. "Polarizable Force Fields for CO_2 and CH_4 Adsorption in M-MOF-74". In: *J. Phys. Chem. C* 121.8 (Feb. 2017), pp. 4659–4673. DOI: 10.1021/acs.jpcc.6b12052.

[195] J. J. Perry, S. L. Teich-McGoldrick, S. T. Meek, J. A. Greathouse, M. Haranczyk, and M. D. Allendorf. "Noble Gas Adsorption in Metal–Organic Frameworks Containing Open Metal Sites". In: *J. Phys. Chem. C* 118.22 (May 2014), pp. 11685–11698. DOI: 10.1021/jp501495f.

[196] N. Metropolis and S. Ulam. "The Monte Carlo Method". In: *Journal of the American Statistical Association* 44.247 (Sept. 1949), pp. 335–341. DOI: 10.1080/01621459.1949.10483310.

[197] B. Widom. "Some Topics in the Theory of Fluids". In: *J. Chem. Phys.* 39 (11 1963), pp. 2808–2812. DOI: 10.1063/1.1734110.

[198] I. Nezbeda and J. Kolafa. "A New Version of the Insertion Particle Method for Determining the Chemical Potential by Monte Carlo Simulation". In: *Molecular Simulation* 5.6 (Feb. 1991), pp. 391–403. DOI: 10.1080/08927029108022424.

[199] C. Simon. *Computing the heat of adsorption in a grand canonical Monte Carlo simulation.* Ed. by github. Available online at http://corysimon.github.io/articles/hoa/. 2016.

[200] D. Dubbeldam, S. Calero, D. E. Ellis, and R. Q. Snurr. "RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials". In: *Mol. Simulat.* 42.2 (2016), pp. 81–101.

[201] C. Zhang, X. Dong, Y. Chen, H. Wu, L. Yu, K. Zhou, Y. Wu, Q. Xia, H. Wang, Y. Han, and J. Li. "Balancing uptake and selectivity in a copper-based metal–organic framework for xenon and krypton separation". In: *Separation and Purification Technology* 291 (June 2022), p. 120932. DOI: 10.1016/j.seppur.2022.120932.

[202] V. A. Solanki and B. Borah. "High-Throughput Computational Screening of 12,351 Real Metal–Organic Framework Structures for Separation of Hexane Isomers: A Quest for a Yet Better Adsorbent". In: *J. Phys. Chem. C* 124.8 (Feb. 2020), pp. 4582–4594. DOI: 10.1021/acs.jpcc.9b11196.

[203] A. Deshwal, C. M. Simon, and J. R. Doppa. "Bayesian optimization of nanoporous materials". In: *Mol. Syst. Des. Eng.* 6.12 (2021), pp. 1066–1086. DOI: 10.1039/d1me00093d.

[204] L. T. Glasby and P. Z. Moghadam. “Hydrogen storage in MOFs: Machine learning for finding a needle in a haystack”. In: *Patterns* 2.7 (July 2021), p. 100305. DOI: 10.1016/j.patter.2021.100305.

[205] D. Frenkel and B. Smit. *Understanding molecular simulation: from algorithms to applications.* Vol. 1. Elsevier, 2001. Chap. 11, pp. 269–287.

[206] H. Lu, Y. Yan, X. Tong, W. Yan, J. Yu, and R. Xu. “The structure-directing effect of *n*-propylamine in the crystallization of open-framework aluminophosphates”. In: *Sci. China Chem.* 57 (1 2014), pp. 127–134. DOI: 10.1007/s11426-013-4980-z.

[207] T. Wang, C. Zhang, Z. Ju, and H. Zheng. “Solvent-induced synthesis of cobalt(II) coordination polymers based on a rigid ligand and flexible carboxylic acid ligands: syntheses, structures and magnetic properties”. In: *Dalton Trans.* 44 (15 2015), pp. 6926–6935. DOI: 10.1039/C5DT00578G.

[208] G.-L. Zhuang, W.-W. Chen, G.-N. Zeng, J.-G. Wang, and W.-L. Chen. “Position of substituent dependent dimensionality in Ln–Cu heterometallic coordination polymers”. In: *CrystEngComm* 14 (2 2012), pp. 679–683. DOI: 10.1039/C1CE05864A.

[209] X. Song, J. Li, Y. Guo, Q. Pan, L. Gan, J. Yu, and R. Xu. “Syntheses and Characterizations of Transition-Metal-Substituted Aluminophosphate Molecular Sieves $|(C_3N_2H_5)_8|[M_8Al_{16}P_{24}O_{96}]$ (M = Co, Mn, Zn) with Zeotype LAU Topology”. In: *Inorg. Chem.* 48 (1 2009), pp. 198–203. DOI: 10.1021/ic801405e.

[210] E. R. Cooper, C. D. Andrews, P. S. Wheatley, P. B. Webb, P. Wormald, and R. E. Morris. “Ionic liquids and eutectic mixtures as solvent and template in synthesis of zeolite analogues”. In: *Nature* 430 (7003 2004), pp. 1012–1016. DOI: 10.1038/nature02860.

[211] X. Tong, W. Yan, J. Yu, and R. Xu. “A chiral open-framework fluoroaluminophosphate with enantiomeric excess in the bulk product”. In: *Chem. Commun.* 49 (96 2013), p. 11287. DOI: 10.1039/c3cc47241h.

[212] S. Chen, S. Hoffmann, Y. Prots, J.-T. Zhao, and R. Kniep. “Preparation, Crystal Structures and Thermal Decomposition of Ba_2(EDTA) and Ba_2(EDTA)·2.5H_2O”. In: *Z. anorg. allg. Chem.* 636 (9-10 2010), pp. 1710–1715. DOI: 10.1002/zaac.201000044.

[213] S. Yuan, Y.-K. Deng, W.-M. Xuan, X.-P. Wang, S.-N. Wang, J.-M. Dou, and D. Sun. “Spontaneous chiral resolution of a 3D (3,12)-connected MOF with an unprecedented ttt topology consisting of cubic $[Cd_4(\mu^3\text{-}OH)_4]$ clusters and propeller-like ligands”. In: *CrystEngComm* 16 (19 2014), p. 3829. DOI: 10.1039/c4ce00028e.

[214] S. C. McKellar, A. J. Graham, D. R. Allan, M. I. H. Mohideen, R. E. Morris, and S. A. Moggach. “The effect of pressure on the post-synthetic modification of a nanoporous metal–organic framework”. In: *Nanoscale* 6 (8 2014), pp. 4163–4173. DOI: 10.1039/C3NR04161A.

[215] A. Comotti, S. Bracco, P. Sozzani, S. Horike, R. Matsuda, J. Chen, M. Takata, Y. Kubota, and S. Kitagawa. “Nanochannels of Two Distinct Cross-Sections in a Porous Al-Based Coordination Polymer”. In: *J. Am. Chem. Soc.* 130 (41 2008), pp. 13664–13672. DOI: 10.1021/ja802589u.

[216] X.-D. Zheng, M. Zhang, L. Jiang, and T.-B. Lu. “A pair of 3D homochiral metal–organic frameworks: spontaneous resolution, single-crystal-to-single-crystal transformation and selective adsorption properties”. In: *Dalton Trans.* 41 (6 2012), pp. 1786–1791. DOI: 10.1039/C1DT11825K.

[217] M. A. Al-Ghouti and D. A. Da'ana. "Guidelines for the use and interpretation of adsorption isotherm models: A review". In: *Journal of Hazardous Materials* 393 (July 2020), p. 122383. DOI: 10.1016/j.jhazmat.2020.122383.

[218] L. Li, Y. Zhao, H. Yu, Z. Wang, Y. Zhao, and M. Jiang. "An XGBoost Algorithm Based on Molecular Structure and Molecular Specificity Parameters for Predicting Gas Adsorption". In: *Langmuir* (May 2023). DOI: 10.1021/acs.langmuir.3c00255.

[219] S. V. Anishchik and N. N. Medvedev. "Three-Dimensional Apollonian Packing as a Model for Dense Granular Systems". In: *Phys. Rev. Lett.* 75 (23 Dec. 1995), pp. 4314–4317. DOI: 10.1103/PhysRevLett.75.4314.

[220] F. Aurenhammer. "Power Diagrams: Properties, Algorithms and Applications". In: *SIAM Journal on Computing* 16.1 (1987), pp. 78–96. DOI: 10.1137/0216006.

[221] S. C. van der Marck. "Network Approach to Void Percolation in a Pack of Unequal Spheres". In: *Phys. Rev. Lett.* 77 (9 Aug. 1996), pp. 1785–1788. DOI: 10.1103/PhysRevLett.77.1785.

[222] 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". In: *Microporous and Mesoporous Materials* 149.1 (Feb. 2012), pp. 134–141. DOI: 10.1016/j.micromeso.2011.08.020.

[223] M. Pinheiro, R. L. Martin, C. H. Rycroft, A. Jones, E. Iglesia, and M. Haranczyk. "Characterization and comparison of pore landscapes in crystalline porous materials". In: *Journal of Molecular Graphics and Modelling* 44 (2013), pp. 208–219. DOI: 10.1016/j.jmgm.2013.05.007.

[224] D. Banerjee, Z. Zhang, A. M. Plonka, J. Li, and J. B. Parise. "A Calcium Coordination Framework Having Permanent Porosity and High CO_2/N_2 Selectivity". In: *Cryst. Growth Des.* 12.5 (Mar. 2012), pp. 2162–2165. DOI: 10.1021/cg300274n.

[225] S. Vandenbrande, M. Waroquier, V. V. Speybroeck, and T. Verstraelen. "Ab Initio Evaluation of Henry Coefficients Using Importance Sampling". In: *J. Chem. Theo. Comput.* 14.12 (Oct. 2018), pp. 6359–6369. DOI: 10.1021/acs.jctc.8b00892.

[226] M. Wojdyr. "GEMMI: A library for structural biology". In: *JOSS* 7.73 (May 2022), p. 4200. DOI: 10.21105/joss.04200.

[227] A. L. Zobrist. *Feature extraction and representation for pattern recognition and the game of Go*. The University of Wisconsin-Madison, 1970.

[228] F. Labelle. *Number of legal Go positions*. Available online at http://wismuth.com/chess/chess.html. 2016.

[229] J. Tromp and G. Farnebäck. "Combinatorics of Go". In: *Computers and Games*. Springer Berlin Heidelberg, 2007, pp. 84–99. DOI: 10.1007/978-3-540-75538-8_8.

[230] J. Tromp. *Number of legal Go positions*. Ed. by github. Available online at https://tromp.github.io/go/legal.html.

[231] R. Coulom. "Efficient Selectivity and Backup Operators in Monte-Carlo Tree Search". In: *Computers and Games*. Springer Berlin Heidelberg, 2007, pp. 72–83. DOI: 10.1007/978-3-540-75538-8_7.

[232] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and

D. Hassabis. “Mastering the game of Go with deep neural networks and tree search”. In: *Nature* 529.7587 (Jan. 2016), pp. 484–489. DOI: 10.1038/nature16961.

[233] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis. “Mastering the game of Go without human knowledge”. In: *Nature* 550.7676 (Oct. 2017), pp. 354–359. DOI: 10.1038/nature24270.

[234] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer New York, 2009. DOI: 10.1007/978-0-387-84858-7.

[235] L. S. Shapley et al. “A value for *n*-person games”. In: (1953).

[236] C. Molnar. *Interpretable machine learning*. Available online at https://christophm.github.io/interpretable-ml-book/. 2023.

[237] S. M. Lundberg and S.-I. Lee. “A Unified Approach to Interpreting Model Predictions”. In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Vol. 30. Curran Associates, Inc., 2017.

[238] T. Chen and C. Guestrin. “XGBoost: A Scalable Tree Boosting System”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’16. San Francisco, California, USA: ACM, 2016, pp. 785–794. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785.

[239] 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”. In: *Microporous Mesoporous Mater.* 149.1 (Feb. 2012), pp. 134–141. DOI: 10.1016/j.micromeso.2011.08.020.

[240] J. Diwu, A.-G. D. Nelson, S. Wang, C. F. Campana, and T. E. Albrecht-Schmitt. “Comparisons of Pu(IV) and Ce(IV) Diphosphonates”. In: *Inorg. Chem.* 49.7 (Apr. 2010), pp. 3337–3342. DOI: 10.1021/ic100184q.

[241] N. P. Martin, J. März, C. Volkringer, N. Henry, C. Hennig, A. Ikeda-Ohno, and T. Loiseau. “Synthesis of Coordination Polymers of Tetravalent Actinides (Uranium and Neptunium) with a Phthalate or Mellitate Ligand in an Aqueous Medium”. In: *Inorg. Chem.* 56.5 (Feb. 2017), pp. 2902–2913. DOI: 10.1021/acs.inorgchem.6b02962.

[242] L. Jouffret, M. Rivenet, and F. Abraham. “Linear Alkyl Diamine-Uranium-Phosphate Systems: U(VI) to U(IV) Reduction with Ethylenediamine”. In: *Inorg. Chem.* 50.10 (Apr. 2011), pp. 4619–4626. DOI: 10.1021/ic200345j.

[243] L. Liang, R. Zhang, J. Zhao, C. Liu, and N. S. Weng. “Two actinide-organic frameworks constructed by a tripodal flexible ligand: Occurrence of infinite $\{(UO_2O_2(OH)_3\}_{4n}$ and hexanuclear $\{Th_6O_4(OH)_4\}$ motifs”. In: *J. Solid State Chem.* 243 (Nov. 2016), pp. 50–56. DOI: 10.1016/j.jssc.2016.07.026.

[244] G. S. Fanourgakis, K. Gkagkas, E. Tylianakis, and G. E. Froudakis. “A Universal Machine Learning Algorithm for Large-Scale Screening of Materials”. In: *J. Am. Chem. Soc.* 142.8 (Feb. 2020), pp. 3814–3822. DOI: 10.1021/jacs.9b11084.

[245] M. Pardakhti, P. Nanda, and R. Srivastava. “Impact of Chemical Features on Methane Adsorption by Porous Materials at Varying Pressures”. In: *J. Phys. Chem. C* 124.8 (Jan. 2020), pp. 4534–4544. DOI: 10.1021/acs.jpcc.9b09319.

[246] M. Pinheiro, R. L. Martin, C. H. Rycroft, A. Jones, E. Iglesia, and M. Haranczyk. "Characterization and comparison of pore landscapes in crystalline porous materials". In: *J. Mol. Graph. Model.* 44 (July 2013), pp. 208–219. DOI: 10.1016/j.jmgm.2013.05.007.

[247] N. Tarbă, M.-L. Voncilă, and C.-A. Boiangiu. "On Generalizing Sarle's Bimodality Coefficient as a Path towards a Newly Composite Bimodality Coefficient". In: *Mathematics* 10.7 (Mar. 2022), p. 1042. DOI: 10.3390/math10071042.

[248] M. Laakso and R. Taagepera. ""Effective" Number of Parties". In: *Comparative Political Studies* 12.1 (Apr. 1979), pp. 3–27. DOI: 10.1177/001041407901200101.

[249] E. H. Simpson. "Measurement of Diversity". In: *Nature* 163.4148 (Apr. 1949), pp. 688–688. DOI: 10.1038/163688a0.

[250] B. Kramer and A. MacKinnon. "Localization: theory and experiment". In: *Rep. Prog. Phys.* 56.12 (Dec. 1993), pp. 1469–1564. DOI: 10.1088/0034-4885/56/12/001.

[251] D. Bousquet, F.-X. Coudert, and A. Boutin. "Free energy landscapes for the thermodynamic understanding of adsorption-induced deformations and structural transitions in porous materials". In: *J. Chem. Phys.* 137 (4 2012), p. 044118. DOI: 10.1063/1.4738776.

[252] Y. Tao, Y. Fan, Z. Xu, X. Feng, R. Krishna, and F. Luo. "Boosting Selective Adsorption of Xe over Kr by Double-Accessible Open-Metal Site in Metal–Organic Framework: Experimental and Theoretical Research". In: *Inorg. Chem.* 59.16 (July 2020), pp. 11793–11800. DOI: 10.1021/acs.inorgchem.0c01766.

RÉSUMÉ EN FRANÇAIS

Introduction

[5 à 10 pages]

Les matériaux poreux sont des matériaux

RÉSUMÉ

Durant ma thèse, j'ai

MOTS CLÉS

simulation moléculaire, matériaux nanoporeux, criblage haut-débit, adsorption, apprentissage statistique

ABSTRACT

During my PhD, I

KEYWORDS

molecular simulation, nanoporous materials, high-throughput screening, adsorption, machine learning