UNIVERSITY OF CRETE

DEPARTMENT OF CHEMISTRY

GENERAL POSTGRADUATE PROGRAM

IN PHYSICAL CHEMISTRY

LABORATORY OF THEORETICAL AND COMPUTATIONAL CHEMISTRY

MASTER THESIS

«MULTI SCALE INVESTIGATION OF HYDROGEN ADSORPTION IN METAL-ORGANIC FRAMEWORKS»

Rafaela Maria Giappa

Supervisor: George E. Froudakis

HERAKLION 2020

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΧΗΜΕΙΑΣ

ΓΕΝΙΚΟ ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ

ΤΟΜΕΑΣ ΦΥΣΙΚΟΧΗΜΕΙΑΣ

ΕΡΓΑΣΤΗΡΙΟ ΘΕΩΡΗΤΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΗΣ ΧΗΜΕΙΑΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΔΙΠΛΩΜΑ ΕΙΔΙΚΕΥΣΗΣ

«ΘΕΩΡΗΤΙΚΗ ΜΕΛΕΤΗ ΠΡΟΣΡΟΦΗΣΗΣ ΥΔΡΟΓΟΝΟΥ ΣΕ ΜΕΤΑΛΛΟ-ΟΡΓΑΝΙΚΑ ΣΚΕΛΕΤΙΚΑ ΥΛΙΚΑ ΜΕ ΜΕΘΟΔΟΥΣ ΠΟΛΛΑΠΛΗΣ ΚΛΙΜΑΚΟΣ»

Ραφαέλα Μαρία Γιάππα

Επιβλέπων Καθηγητής: Γεώργιος Ε. Φρουδάκης

ΗΡΑΚΛΕΙΟ 2020

στον Αττίλα

Thesis Committee

George E. Froudakis (Supervisor)

Professor

Department of Chemistry, University of Crete

Pantelis N. Trikalitis

Professor

Department of Chemistry, University of Crete

Athanasios Stubos

Research Director

National Research Center Demokritos

Study hard what interests you the most, in the most undisciplined, irreverent, and original manner possible.

Richard P. Feynman

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my supervisor and mentor Prof. George Froudakis for his patient guidance, enthusiastic encouragement and useful critique. Throughout this scientific journey I was offered life lessons that will always be remembered and appreciated. Subsequently, I would like to thank my thesis committee; Prof. Pantelis Trikalitis and Dr. Athanasios Stubos for their insightful comments and remarks. In my quest of building an interdisciplinary scientific toolbox, I would like to highlight special thanks to Prof. Pantelis Trikalitis for our first year's discussions that helped me smoothen the transition from the physical to the chemical scientific language. To this point, I wish to express my appreciation and thanks to all the members of the group; I am particularly grateful to Dr Emmanuel Tyllianakis for his patient temper, guidance and scientific lessons resulting in a productive collaboration. My sincere gratitude to Dr Emmanuel Klontzas for being -along with Prof. George Froudakis- the first person to introduce me to the world of MOFs and guide me through my first quantum chemistry calculations. I would also like to express my gratitude to Dr Taxiarchis Stergiannakos for our fruitful discussions and Dr George Fanourgakis for his kind personality and keen scientific eye that make me look forward to a potential common project cooperation.

This thesis was based on research conducted during my internship period in Toyota Motor Europe in Brussels; this project has been a collaboration of Toyota and the University of Crete. The internship opportunity has been a great chance for gaining valuable professional experience and developing core personal aspects. For this reason, I thank Toyota for teaching me the virtue of continuous improvement through challenge, Kaizen (改善) and Genchi Genbutsu (現地現物). I would like to express my deep gratitude to all Toyota members, one by one, for their teamwork and genuine respect for people. During my internship journey I collaborated closely with Dr Marco di Gennaro, Dr Konstantinos Gkagkas, Dr Emmanuel Tyllianakis and Prof. George Froudakis. I would like to thank Dr Konstantinos Gkagkas for being the senior coordinator of the project and personally, for offering valuable advice for living abroad and guiding important steps. Special thanks to Dr Marco di Gennaro for our close collaboration in the same wavelength, for his patient guidance and his kind try of passing me valuable knowledge and expertise.

To this point, I would like to thank the Department of Chemistry and the University of Crete, for offering me an amazing education journey that molded my enthusiastic scientific point of view to see and interpret the vast world of knowledge with awe and respect. It would be a serious oversight not to express my gratitude to many colleagues and friends for sharing the same interdisciplinary enthusiasm and being active members of fruitful scientific discussions. Special thanks belong to my dear friend George Vailakis for standing by my side and sharing the majority of our scientific evolution. Last but not least, I would like to thank my family; both two- and four-legged. To my parents I express my endless love and eternal gratitude for always believing in me, being supportive and teaching me milestone lessons. My finest feelings belong to Attilas, for "just" being in my life, reminding me to keep my head high and teaching me that sky is the limit.

CURRICULUM VITAE

PERSONAL INFORMATION

Full name	Rafaela Maria Giappa
Date of Birth	04/09/1994
Place of Birth	Athens, Greece
Nationality	Greek
Address	Department of Chemistry, University of Crete, P. O. BOX 2208, Heraklion, 71003, Greece
Telephone	+30 6951508371 +30 6951688868
E-mail	rafaelagiappa@gmail.com

EDUCATION

10/2017 – 09/2020	**Master of Science (MSc) Degree in Physical Chemistry**, Department of Chemistry, University of Crete, Greece
09/2012 – 09/2017	**Bachelor of Science (BSc) Degree in Physics**, Department of Physics, University of Crete, Greece (Grade: 7.92/10)
06/2012	High school graduate, 4th General High School of Ilioupolis, Athens, Greece (Grade: 18.7/20)

WORKING EXPERIENCE

10/2019 - 04/2020	**Intern, R&D, Toyota Motor Europe**, Brussels
Spring semester 2017-2018	Physics II (XHM-017) **Teaching Assistant**, Department of Chemistry, University of Crete
Fall semester 2017-2018	Physical Chemistry II (XHM-444) **Laboratory Assistant**, Department of Chemistry, University of Crete

COMPUTATIONAL SKILLS

Computational Chemistry programs and packages	**Gaussian, Turbomole, ORCA, RASPA, Avogadro, Molden, gOpenMol, Materials Studio**.
Operating Systems	Microsoft Windows (Vista, 7, 8, 10) , Linux (Debian)
Other Software	Word, Excel, PowerPoint, OriginPro

FOREIGN LANGUAGES

- Certificate of Proficiency in **English**-University of Michigan
  Proficient user (CEFR Level: **C2**)
- Diploma in **French** language
  DELF B2 Independent user (CEFR Level: **B2**)

PUBLICATIONS

2019 **"Water-Stable 2-D Zr MOFs with Exceptional UO_2^{2+} Sorption Capability"**
Nikos Panagiotou, Ioanna Liatsou, Anastasia Pournara, Giasemi K. Angeli, Rafaela Maria Giappa, Manolis J. Manos, Emmanuel Tylianakis, Pantelis N. Trikalitis, George E. Froudakis, Ioannis Pashalides, Anastasios J. Tasiopoulos,
J. Mater. Chem. A, 2020,8, 1849-1857.

2019 **"Design of novel nano-porous materials for hydrogen storage"**
Giappa R.M, Froudakis G.
Proceedings of the Greek Academy of Science for the energy research in Greece

In preparation **"Multi-scale investigation of hydrogen adsorption in metal organic frameworks"**
Giappa R.M., Tylianakis E., Froudakis G.

In preparation **"Theoretical study of water interaction with functionalized benzene molecules"**
Giappa R.M., Klontzas E., Tylianakis E., Froudakis G.

CONFERENCES

27-31/05/2019 **E-MRS Spring Meeting 2019, Nice, France**
"Strategically functionalized MOFs for enhanced water adsorption"
Giappa R.M., Klontzas E., Tylianakis E., Froudakis G.
(**Presentation)**

14-19/10/2018 **7th International Symposium on Transparent Conductive Materials (IS-TCMs), Platanias, Crete, Greece**
"Theoretical study of water interaction with functionalized benzene molecules"
Giappa R.M., Klontzas E., Froudakis G.
(**Poster)**

24-28/09/2018	**1st International Conference on Nanotechnologies and Bionanoscience (NanoBio), Heraklion, Crete, Greece** "Theoretical study of water interaction with functionalized benzene molecules." Giappa R.M., Klontzas E., Froudakis G. (**Poster)**
22/11/2018	**VI Workshop in the memory of Dr. Maria Hatzimarinaki, Department of Chemistry, University of Crete, Heraklion, Greece** "Theoretical study of water interaction with functionalized benzene molecules" Giappa R.M., Klontzas E., Froudakis G. (**Presentation)**
17-19/09/2018	**XXXIII Panhellenic Conference on Solid State Physics and Materials Science, University of Cyprus, Nicosia, Cyprus** "Theoretical study of water interaction with functionalized benzene molecules" Giappa R.M., Klontzas E., Froudakis G. (**Presentation)**
25/06/2018	**20th Postgraduate Chemistry Students Conference, University of Crete, Heraklion, Greece** "Theoretical study of water interaction with functionalized benzene molecules" Giappa R.M., Klontzas E., Froudakis G. (**Presentation)**

ABSTRACT

Nowadays humanity is facing two urgent, interconnected problems; the exhaust of conventional energy resources and the need to reduce CO_2 emissions and move towards a sustainable carbon emission free economy. Within the scope of sustainability, hydrogen economy presents a possible solution to both problems. Hydrogen is a completely clean, non-toxic burning fuel which when consumed in a fuel cell, the only by-product is pure water. Fuel-cell technology is currently under intensive research and development in view of the expected benefits in facing the environmental problems and the gradual depletion of conventional fuel reserves. The main obstacle for the commercial development of fuel-cell powered vehicles is the efficient hydrogen storage. Among the existing technologies, an alternative way of storing hydrogen is the adsorption in porous materials, such as Metal Organic Frameworks (MOFs). Many different strategies have been up-to-date employed to enhance the hydrogen storage capacity of MOFs. The objective in this project is to design new materials that have improved sorption capacities through enhancing their interaction with H_2.

Within the scope of the multiscale approach, we investigate *in silico* the H_2 capacity enhancement in several MOF-type nanomaterials, when these are properly functionalized with a variety of chemical species. Driven by the fact that the organic linkers of most MOFs have aromatic backbones such as benzene, we screen simple aromatic systems of the form C_6H_5X (where X stands for different functional groups) for their binding strength towards hydrogen. Ab initio calculations are employed to calculate the binding energy and favorable positions of hydrogen with a series of strategically selected functionalized benzenes. Subsequently, GCMC calculations with model potentials derived from the ab initio calculations, are employed to study the trend obtained from the meticulous ab initio search. The results from our bottom-up approach lead us to conclude that this functionalization strategy can be applied to various porous materials (MOFs, COFs, etc.), in order to enhance their hydrogen storage performance, especially at cryogenic temperatures.

Key words: sustainability, fuel-cells, hydrogen storage, MOFs, functionalization, multi-scale, ab initio, GCMC.

ΠΕΡΙΛΗΨΗ

Σήμερα, η ανθρώπινη δραστηριότητα εξαρτάται σχεδόν εξολοκλήρου από μη ανανεώσιμες πηγές ενέργειας όπως το φυσικό αέριο, ο άνθρακας και το πετρέλαιο προκειμένου να ικανοποιηθούν οι ολοένα αυξανόμενες ενεργειακές ανάγκες του πλανήτη. Η αναπόφευκτη μελλοντική εξάντληση των ορυκτών αυτών καυσίμων σε συνδυασμό με την έκλυση ρύπων κατά τη χρήση τους, καθιστούν επιτακτική την ανάγκη ανεύρεσης ενός νέου ενεργειακού μοντέλου.

Το υδρογόνο αποτελεί έναν φορέα ενέργειας με πολλές προοπτικές που αναμένεται να έχει σημαντική συμβολή στην κατεύθυνση της επίλυσης του ενεργειακού προβλήματος. Συγκεκριμένα, η τεχνολογία των κυψελίδων καυσίμου (fuel cells) υπόκειται σε εντατική έρευνα και ανάπτυξη εξαιτίας τόσο των πλεονεκτημάτων της στην μείωση των περιβαλλοντικών επιπτώσεων όσο και της σταδιακής εξάντλησης των συμβατικών καυσίμων. Σήμερα, το βασικότερο εμπόδιο στην εμπορική αξιοποίηση της τεχνολογίας των κυψελίδων καυσίμου σε οχήματα, είναι το πρόβλημα της αποθήκευσης του υδρογόνου στο χώρο του οχήματος. Ανάμεσα στις υπάρχουσες τεχνολογίες, μια εναλλακτική μέθοδος αποθήκευσης του υδρογόνου η οποία ερευνάται εντατικά, είναι η προσρόφηση του σε πορώδη υλικά, όπως τα μέταλλο-οργανικά σκελετικά υλικά (Metal Organic Frameworks-MOFs). Η προσροφητική ικανότητα των υλικών αυτών δύναται να ενισχυθεί μέσω συγκεκριμένων στρατηγικών βελτίωσης. Κύριο στόχο της παρούσας μελέτης αποτελεί η κατάλληλη τροποποίηση τους ώστε να ενισχυθεί η αλληλεπίδραση τους με το H_2, με αποτέλεσμα να προκύψουν νέα υλικά με βελτιωμένες ενέργειες δέσμευσης. Ο σχεδιασμός, η τροποποίηση και η μελέτη των υλικών αυτών πραγματοποιείται με υπολογιστικές μεθόδους πολλαπλής κλίμακας (multi-scale treatment).

Στην παρούσα διατριβή μελετήθηκε με τη χρήση μεθόδων πολλαπλής κλίμακας η ενίσχυση της αλληλεπίδρασης μέσω του κατάλληλου εμπλουτισμού του υλικού με σειρά στρατηγικά επιλεγμένων λειτουργικών ομάδων. Καθώς ο οργανικός υποκαταστάτης πολλών MOF αποτελείται από βενζολικούς δακτυλίους, υπολογίστηκε με κβαντικές μεθόδους υψηλής ακρίβειας (ab initio) η αλληλεπίδραση του H_2 με σειρά τροποποιημένων δακτυλίων την μορφής C_6H_5X (όπου X η εκάστοτε λειτουργική μονάδα) ως προς την ενέργεια δέσμευσης. Στη συνέχεια πραγματοποιήθηκαν κλασικές προσομοιώσεις Grand Canonical Monte Carlo (GCMC) σε συγκεκριμένες MOF δομές (IRMOF-8, IRMOF-16, IRMOF-14), προκειμένου να

διερευνηθεί εάν οι υψηλότερες ενέργειες δέσμευσης του H_2 από τους τροποποιημένους οργανικούς υποκαταστάτες οδηγούν σε ενίσχυση της προσροφητικής ικανότητας του τροποποιημένου MOF υλικού. Η στρατηγική τροποποίησης του οργανικού υποκαταστάτη με τις ισχυρότερες ως προς την αλληλεπίδρασή του με το υδρογόνο λειτουργικές ομάδες μπορεί να επεκταθεί τόσο σε άλλες MOF δομές όσο και σε άλλο είδος πορωδών δομών (όπως π.χ. τα Covalent Organic Frameworks (COFs)), προκειμένου να ενισχυθεί η ικανότητα προσρόφησης τους σε υδρογόνο.

Λέξεις Κλειδιά: ανανεώσιμες πηγές ενέργειας, υδρογόνο, κυψελίδες καυσίμου, μέταλλο-οργανικά σκελετικά υλικά, μέθοδοι πολλαπλής κλίμακας, λειτουργικές ομάδες, GCMC.

CONTENTS

CHAPTER 3 **Results and Discussion**

LIST OF FIGURES

Figure 34: Equilibration and production phases in an MC simulation.

Figure 35: 2-D representation of periodic boundary conditions (left). Example of the minimum image convention (right).

Figure 36: Some of the most important electrophilic substitutions.

Figure 37: Examples of electron donating and electron withdrawing functional groups (left). Comparison of the electrostatic potential surface of benzene, an electron donating FG (-OCH_3) and an electron withdrawing FG (-NO_2) (right).

Figure 38: A 1-D representation of the energy of all possible conformers of a simple molecule.

Figure 39: The optimized monomer geometries for all the functionalized benzenes that will be studied thoroughly for their binding strength towards hydrogen.

Figure 40: Example of various initial configurations of the dimer C_6H_5-NH_2…H_2 that were created and optimized. The energetically most stable configuration (global minimum) is shown on the right.

Figure 41: Global minima of the optimized geometries for the least interacting FGs. Binding energies up to -4.6 kJ/mol (the binding energy of benzene (reference)).

Figure 42: Global minima of the optimized geometries for FGs with moderate binding energies. Values from -4.7 to -5.0 kJ/mol.

Figure 43: Global minima of optimized geometries for FGs with binding energies from -5.1 to -5.4 kJ/mol.

Figure 44: Global minima of optimized geometries for the best performing FGs, with energy enhancement over 20% when compared to benzene.

Figure 45: The electrostatic potential maps of the structures under study.

Figure 46: Electron-density redistribution plots of the optimized geometries for the least interacting FGs, with binding energies up to -4.6 kJ/mol.

Figure 47: Electron-density redistribution plots of the optimized geometries for FGs with moderate binding energies from -4.7 to -5.0 kJ/mol.

Figure 48: Electron-density redistribution plots of the optimized geometries for FGs with binding energies from -5.1 to -5.4 kJ/mol.

Figure 49: Electron-density redistribution plots of the optimized geometries for the best performing FGs.

Figure 50: Fitting of the (ε, σ) parameters of the Dreiding potential on the QM data obtained from the ab initio scan of H_2 over benzene.

Figure 51: Fitting of the (ε, σ) parameters for the H_2…C_6H_5-$OCONH_2$ interaction.

Figure 52: Fitting of the (ε, σ) parameters for the H_2…C_6H_5-SO_2NH_2 interaction.

Figure 53: Fitting of the (ε, σ) parameters for the H_2…C_6H_5-OPO_3H_2 interaction.

Figure 54: Fitting of the (ε, σ) parameters for the H_2…C_6H_5-OSO_3H interaction.

Figure 55: Structures of the IRMOF-n (n=1-8, 10, 12, 14, 16) series.

Figure 56: The IRMOFs selected and functionalized for this study along with the cluster models chosen to represent their organic linker part.

Figure 57: The cluster model selected to represent the metal corner of the IRMOF series.

Figure 58: Volumetric (g/L) and gravimetric (%wt) hydrogen uptake at T=300K of IRMOF-08 and IRMOF-08-n (n: -OSO_3H, -$OCONH_2$, -SO2NH_2).

Figure 59: Volumetric (g/L) and gravimetric (%wt) hydrogen uptake at T=77K of IRMOF-08 and IRMOF-08-n (n: -OSO_3H, -$OCONH_2$, -SO2NH_2).

Figure 60: Isosteric heat of adsorption (kJ/mol) of IRMOF-08 and IRMOF-08-n (n: -OSO_3H, -$OCONH_2$, -SO2NH_2).

Figure 61: Snapshots of IRMOF-08 and IRMOF-08-$OCONH_2$ at T=77K.

Figure 62: Snapshots of IRMOF-08 and IRMOF-08-SO_2NH_2 at T=77K.

Figure 63: Snapshots of IRMOF-08 and IRMOF-08-OSO_3H at T=77K.

Figure 64: Volumetric (g/L) and gravimetric (%wt) hydrogen uptake at T=300K of IRMOF-16 and IRMOF-16-n (n: -OSO_3H, -OPO_3H_2).

Figure 65: Isosteric heat of adsorption (kJ/mol) of IRMOF-16 and IRMOF-16-n (n: -OSO_3H, -OPO_3H_2).

Figure 66: The IRMOF-16-OSO_3H linker (left) and the robust IRMOF-14-OSO_3H linker (right).

LIST OF TABLES

1.1 Hydrogen Storage

Our planet is facing the urgent need to reduce CO_2 emissions and move towards a sustainable carbon emission free economy. Global carbon emission from fossil fuel combustion accounts for 90% of all emissions from human activity, making fossil fuels the principal CO_2 emission source. Meanwhile, fossil fuels reserves face depletion as global fossil fuel consumption is on the rise. Whilst depleting reserves could become a pressing issue 50-100 years from now, an important limit to fossil fuel production is climate change as CO_2 emissions remain trapped in the atmosphere for long periods of time causing global temperature to increase. Today, there is a general shift towards environmental awareness, which has led to the rise of a number of alternative energy sources. Decreased emissions and reduction of environmental pollution are major benefits associated with the use of sustainable, green energy solutions. Within the scope of sustainability and with a great potential, hydrogen economy is in its early days.

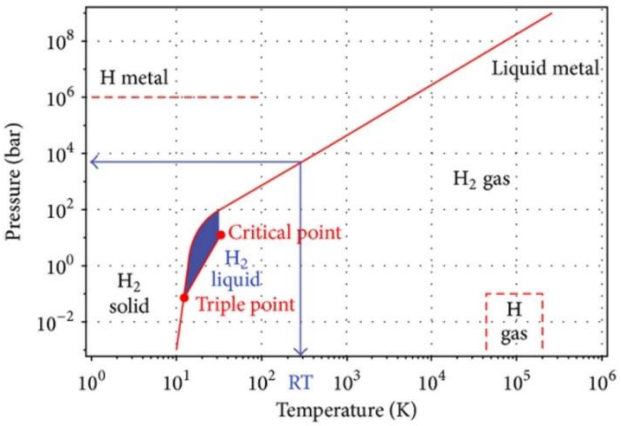


Figure 1: Primitive phase diagram of hydrogen showing the phase behavior with changes in temperature and pressure [1].

Hydrogen is a completely clean, non-toxic burning fuel with pure water as the only by-product when consumed in a fuel cell.

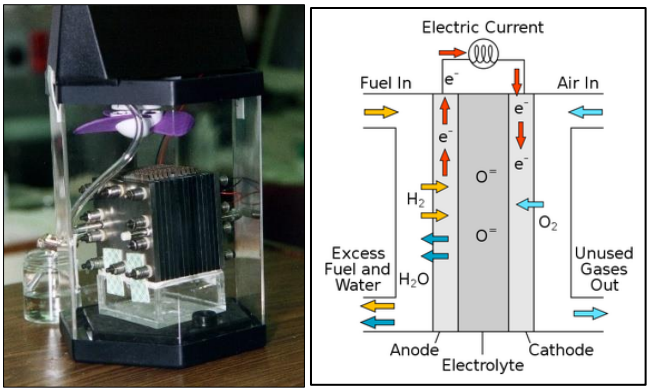


Figure 2: Demonstration model of a fuel cell (black layered cube) in its enclosure (left). Scheme of a proton-conducting fuel cell (right) [2].

Classified as an alternative energy fuel under the Energy Policy Act of 1992 [3], hydrogen is an energy carrier that can deliver or store a significant amount of energy. On a mass basis, hydrogen has nearly three times the energy content of gasoline, with 120 MJ/kg versus 44 MJ/kg respectively. On the volume basis, however, the situation is reversed; the same amount of hydrogen requires more storage space than gasoline with liquid hydrogen having a density of 8MJ/L whereas gasoline 32 MJ/L [4].

Today, hydrogen is most commonly used in petroleum refining and fertilizer production, while transportation and utilities are emerging markets. The extensive use of hydrogen as a fuel is being hindered, amongst others, by the lack of effective on-board hydrogen storage solutions. In other words, any practical application for vehicles must include a storage solution competitive with current gas technology. The major challenge is to effectively store enough hydrogen (~6 kg) to enable a practical driving range (~500 km) in a reasonably sized onboard tank. Current fuel cell vehicles, manufactured and sold by automotive companies such as Honda, Hyundai and Toyota, store hydrogen in high pressure (~700 bar) carbon-fiber-reinforced fuel tanks [5]. Still, even at such high pressures, 6 kg of liquid hydrogen require a tank of 150L volume. Such high storage pressure is also an area of concern for consumer safety and compression expenditures.

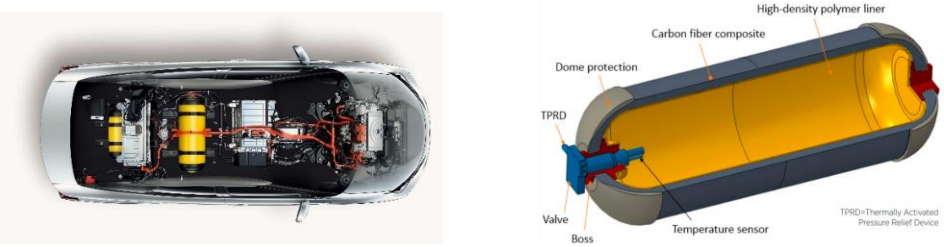


Figure 3: A look on the inside of Toyota Mirai (left) [5] and the main parts of its carbon-fiber reinforced fuel tank (right) [6]. The Mirai tank has an internal volume of 122.4 L, with volumetric energy density up to 4.9 MJ/L [7].

The goal is to design low-cost, light-weight materials that can reversibly and rapidly store hydrogen near ambient conditions. The U.S. Department of Energy (DOE) has set the ultimate technical targets of onboard hydrogen storage for light-duty fuel cell vehicles to be 2.2 kWh/kg for the usable gravimetric and 1.7 kWh/L for the usable volumetric capacities; amounts equivalent to 6.5 wt% H_2 and 50 gr H_2/L respectively, for the complete storage system (i.e. all stored hydrogen, media, reactants and system components). For 2025, these targets are 1.8 kWh/kg (5.5 wt% H_2) and 1.3 kWh/L (40 gr H_2/ L system) [8]. To overcome the challenges in the existing technologies as well as meet the DOE targets, several strategic pathways have been proposed and are currently an active area of multidisciplinary research. Long-term solutions focus on both cryo-compressed and material-based hydrogen storage technologies with the latter including adsorbents, metal hydrides and chemical storage materials as their major representatives [9].

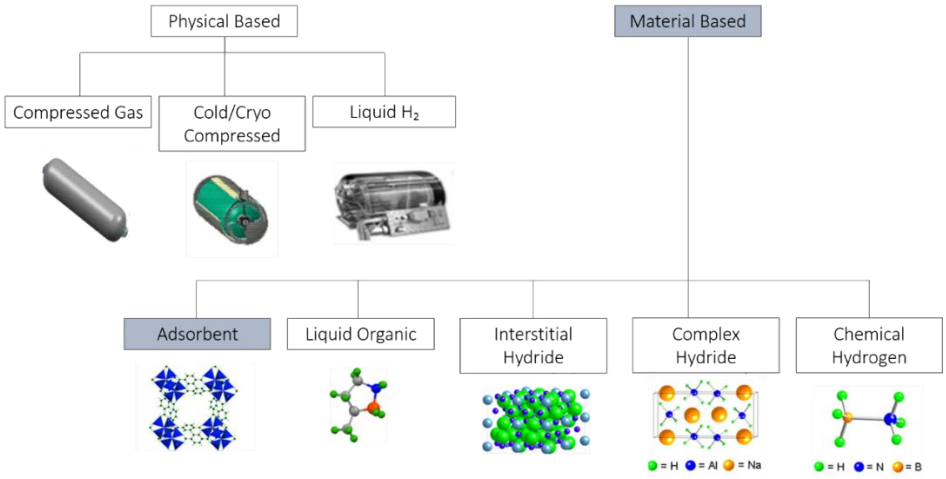


Figure 4: Classification of the existing technologies for hydrogen storage.

Depending on the type of materials and therefore the type of the interaction with molecular hydrogen, the storage system adsorbs hydrogen either physically (binding energies of ~ 4-10 kJ/mol) or chemically (~50-100 kJ/mol). Between the two, physisorption shows considerable promise since hydrogen adsorption is fast and fully reversible. Physisorption strongly correlates with surface area, with higher surface areas leading to greater gas uptake.

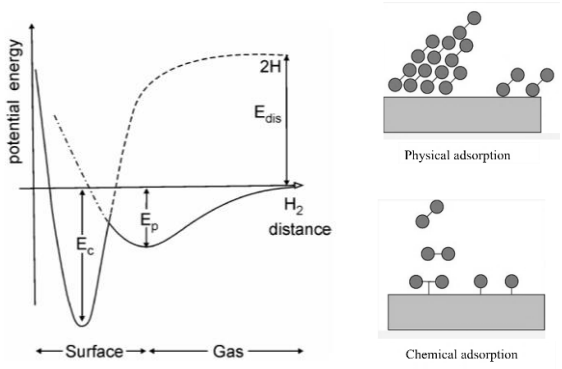


Figure 5: Potential energy curve for chemisorbed and physisorbed hydrogen as a function of the distance from the adsorbent's surface. The minima of the two curves correspond to the equilibrium distance for physisorbed (E_p) and chemisorbed (E_c) hydrogen.

Nanoporous[1] materials such as metal-organic frameworks (MOFs) are attractive candidates possessing high surface areas so that hydrogen could be stored in smaller, more manageable volumes and at relatively low pressures.

[1] The size of the pores in nanoporous materials is generally 100 nanometers or smaller. They can be subdivided into three categories according to IUPAC [95]; microporous (0.2-2 nm), mesoporous (2-50 nm), microporous (>50 nm).

1.2 Metal Organic Frameworks

We have a class of materials in which we can change the components nearly at will. There is no other class of materials where one can do that.
-Omar Yaghi

Introduced by O. Yaghi in 1995 [10], Metal Organic Frameworks (MOFs) are a class of crystalline materials that belong to the broader category of coordination[2] polymers and have attracted substantial attention in the last decade due to their unique properties. They are porous materials that stand out for their ultrahigh porosity (up to 90% free volume), extended internal surface areas (beyond 6000 m^2/gr), tunable pore size and chemical composition. Their porosity allows the diffusion of guest molecules into the bulk structure, with the shape and size of pores governing the shape and size selectivity of the guests to be incorporated.

Figure 6: MOFs in powder and pellet form [Source: Monash University].

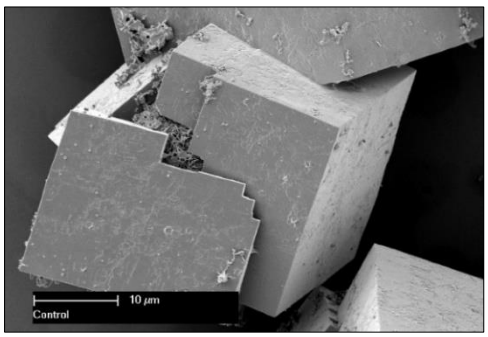


Figure 7: Scanning electron microscope image of a MOF crystal [11].

[2] A coordination polymer is an inorganic or organometallic polymer structure containing metal cation centers linked by ligands. More formally, a coordination polymer is a coordination compound with repeating coordination entities extending in 1, 2, or 3 dimensions.

Their framework is composed of inorganic metal ions (primary building unit (PBU)) or metal clusters (secondary building units (SBUs)) bridged by coordinated multidentate organic ligands, such as carboxylates. Their geometry is determined by the coordination number, coordination geometry of the metal ions, and the nature of the functional groups. SBU geometries may vary, with different number of points of extension such as octahedron (six points), trigonal prism (six points), square paddle-wheel (four points), and triangle (three points). In principle, a bridging ligand (ditopic, tritopic, tetratopic, or multitopic linkers) reacts with a metal ion with more than one vacant or labile site. The final framework topology of MOF is determined by both SBU connectors and organic ligand linkers [12].

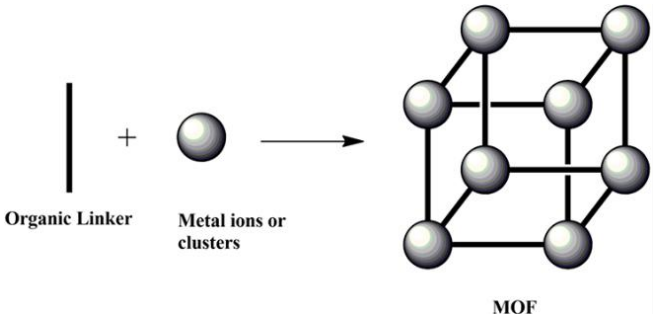


Figure 8: Simplified scheme for the node-and-connector approach to prepare MOFs [12]. Different metal ions or clusters are mixed together with organic linkers using a suitable solvent. Coordination polymerization takes place between the precursors, resulting in a cross-linked network showing potential voids.

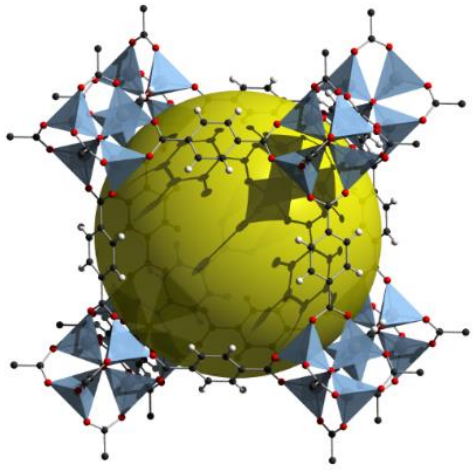


Figure 9: The structure of MOF-5 or IRMOF-1, the most well studied and representative example of MOFs. Composed of zinc cations (the inorganic part) and 1,4-benzene dicarboxylate (BDC, the organic part). The yellow sphere shows the empty pore space [Source: Wikipedia].

Building blocks can be combined in various ways to create novel materials. Complex networks containing different topologies and specific functions morphologies are constructed and employed for a wide variety of applications. The large porosity allows their applications in adsorption and separation of gaseous molecules, catalysis, microelectronics, optics, sensing applications, bioreactors, drug delivery, and others [13].

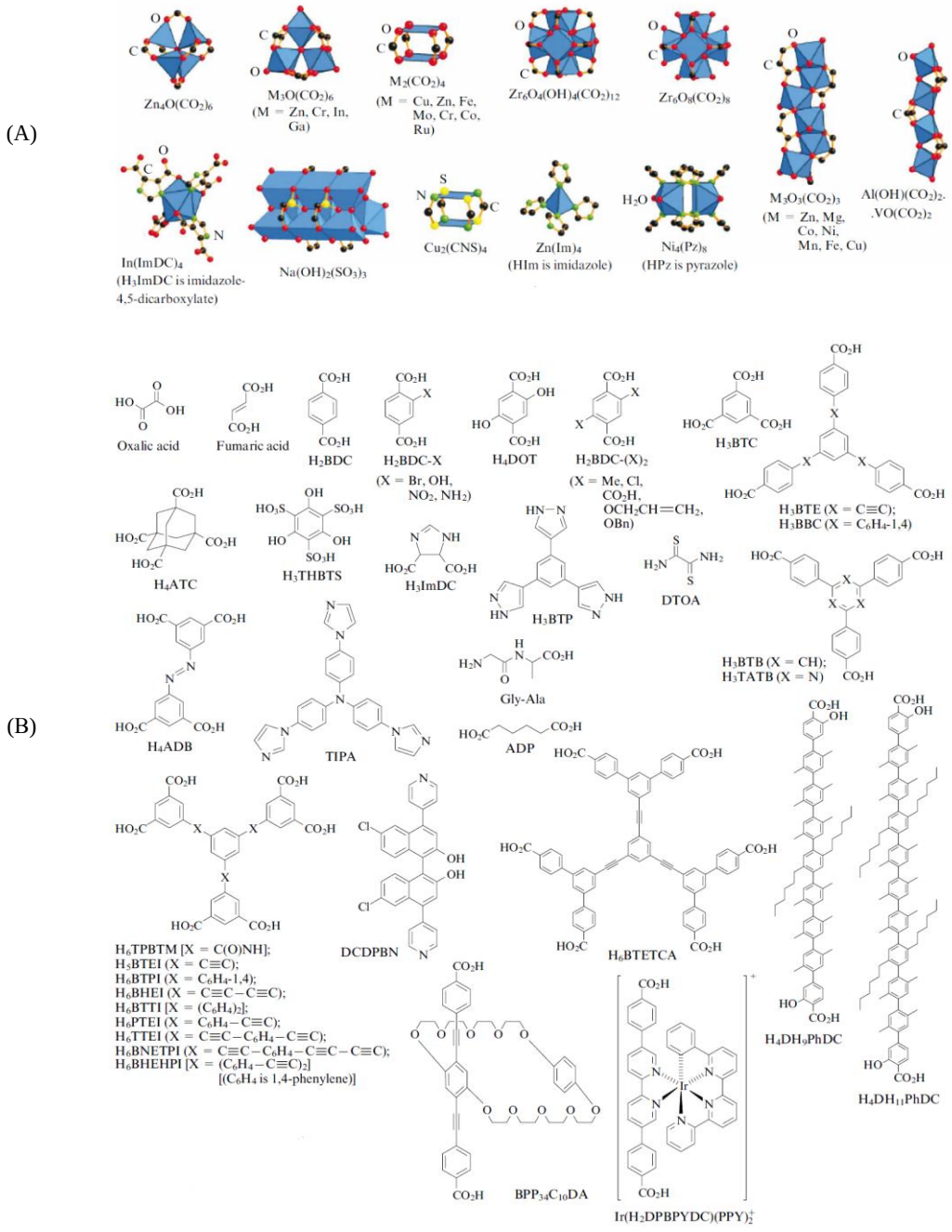


Figure 10: Building blocks of MOFs. Representative list of (A) Metal clusters (the polyhedral demonstrate the metal coordination) and (B) Organic linkers used to construct various types of MOFs [14].

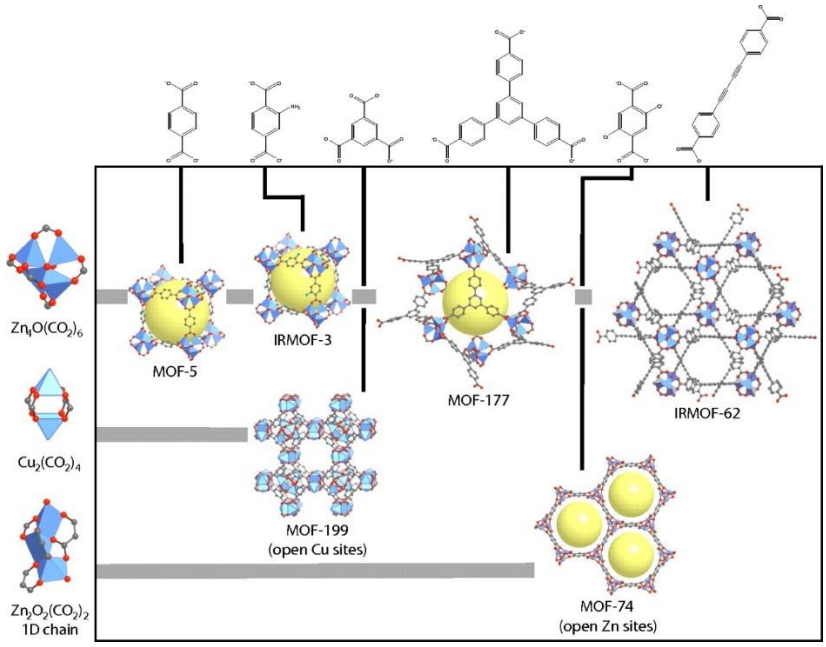


Figure 11: Indicative examples of various combinations of organic linkers and metal clusters resulting in different MOF[3] structures.

MOFs can be characterized as "tailored materials", as they can be designed to have the desired properties for specific applications. Stretching the limits of their physical properties, MOFs' composition, structure and pore metrics can be varied, granting them versatile performances in various research fields. The pore size is allowed to be tuned and spatial cavity arrangement be controlled, by judicious selection of metal centers and organic ligands and also by adjusting their conditions of synthesis.

[3] Abbreviation MOF is usually used as a general name of the class of compounds. When followed by an ordinal number, it denotes an individual metal-organic framework (e.g. MOF-5, MOF-74, etc.). Other abbreviations include naming by initials of institution or place of discovery (e.g. UiO-n: Universitetet i Oslo, MIL-n: Materials of Institut Lavoisier, etc.)

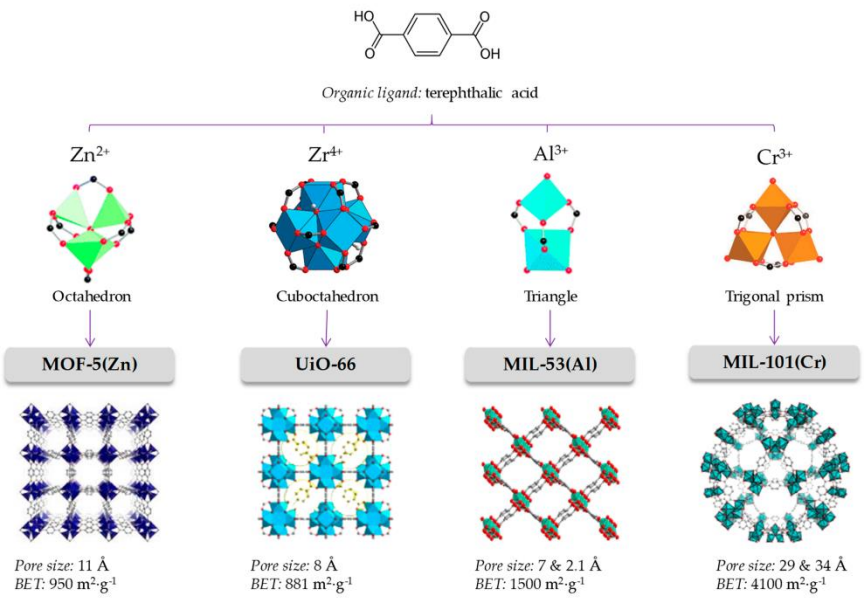


Figure 12: The selection of the organic linker and connection geometry of the metal cluster lead to the desired topology. Each framework topology has its characteristic pore size and available surface [15].

The synthesis of MOFs involves reaction conditions and methods such as solvothermal, ionothermal, diffusion, microwave methods, ultrasound-assisted, template-directed syntheses, and others [16]. At their discovery, the method for synthesis of MOFs is solvothermal. Typically, metal precursors and organic linkers are dissolved in solvent and placed in a closed reaction vessel for the formation and self-assembly of MOF crystals. The common solvents used include N,N-dimethyl formamide (DMF), N,N-diethyl formamide (DEF), methanol, ethanol, and acetonitrile. The synthesis temperature is generally below 220°C, and the crystallization time varies from several hours to several tens of days.

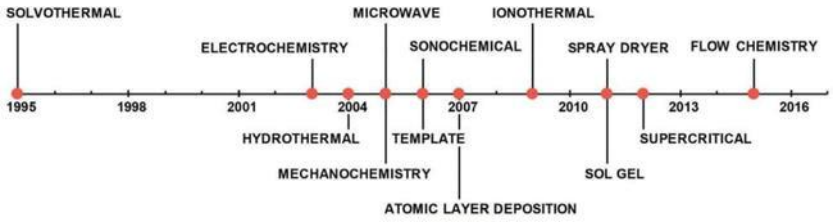


Figure 13: Timeline of the most common synthesis approaches for MOFs. [17]

Metal-Organic Frameworks are currently an extremely active area of research that holds great promises for advanced technological applications especially in gas-sorption/storage and separation. Through the great variety of organic molecules and metals that can be used, over the past 20 years the field has proliferated a huge number of publications with over 20,000 MOF structures published.

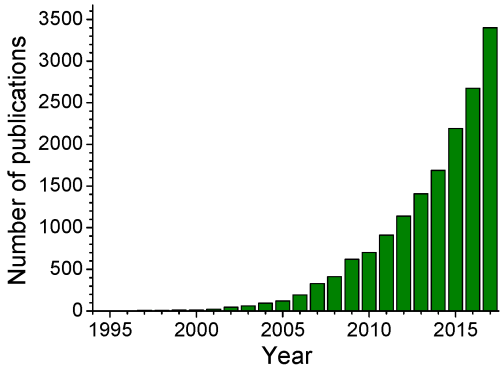


Figure 14: Diagram illustrating the dynamics of interest in MOFs with the number of publications for the period 1994-2017 [Source: Web of Science].

Nowadays MOF databases of theoretically predicted ("hypothetical") and experimentally synthesized ("real") MOFs are being developed for managing the enormous amounts of data. They also give access to search more easily for desired properties, and MOFs with their tailorable pore size and functionality seem to be ideal for the computational design of novel functional materials.

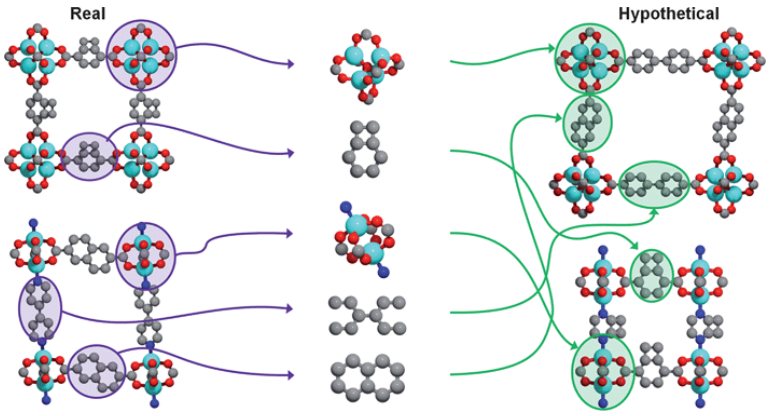


Figure 15: Visual summary of the hypothetical MOF generation strategy. Crystal structures of existing metal–organic frameworks are obtained from X-ray diffraction data (left) and are subsequently divided into building blocks (middle) that can then be recombined to form new, hypothetical metal–organic frameworks (right) [18].

After O. Yaghi first reported the high H_2 uptake capacity of MOF-5 in 2003 [19], MOFs have been intensively researched as hydrogen adsorbents. In order to be promising for H_2 storage applications, a porous material should show an uptake[4] that is ideally large and the adsorption should strong and reversible. Many experimental studies show that MOFs have very high H_2 adsorption capacities, up to 10 wt% at 77 K and 100 bar. For practical hydrogen storage materials, volumetric and gravimetric hydrogen capacities are the key factors that determine the size and weight of the MOF-filled tank required to store a certain amount of hydrogen for reasonable driving range. Therefore, they must be optimized so the tank is neither too large nor too heavy.

Recently, the metal–organic frameworks M2(m-dobdc) (M = Co, Ni; m-dobdc4– = 4,6-dioxido-1,3-benzenedicarboxylate) and the isomeric frameworks M2(dobdc) (M = Co, Ni; dobdc4– = 1,4-dioxido-1,3-benzenedicarboxylate), which are known to have open metal cation sites that strongly interact with H2, were evaluated by Jeffrey R. Long and co-workers for their usable volumetric H2 storage capacities over a range of near-ambient temperatures relevant to on-board storage [20]. The testing of the four different compounds showed that the MOF called Ni_2(m-dobdc) presented the highest hydrogen-storage capacity over a range of pressures and temperatures. At ambient temperature and a much lower tank pressure than used in current hydrogen vehicles, Ni_2(m-dobdc) set a new record for hydrogen storage capacity of 11.9 g of fuel per liter of MOF crystal. The MOF had a significantly greater storage capacity than compressed hydrogen gas under the same conditions.

[4] The uptake curves are also called adsorption isotherms and the uptake can be presented as gravimetric (gr of H_2 per gr of system) or volumetric (gr of H_2 per L of system).

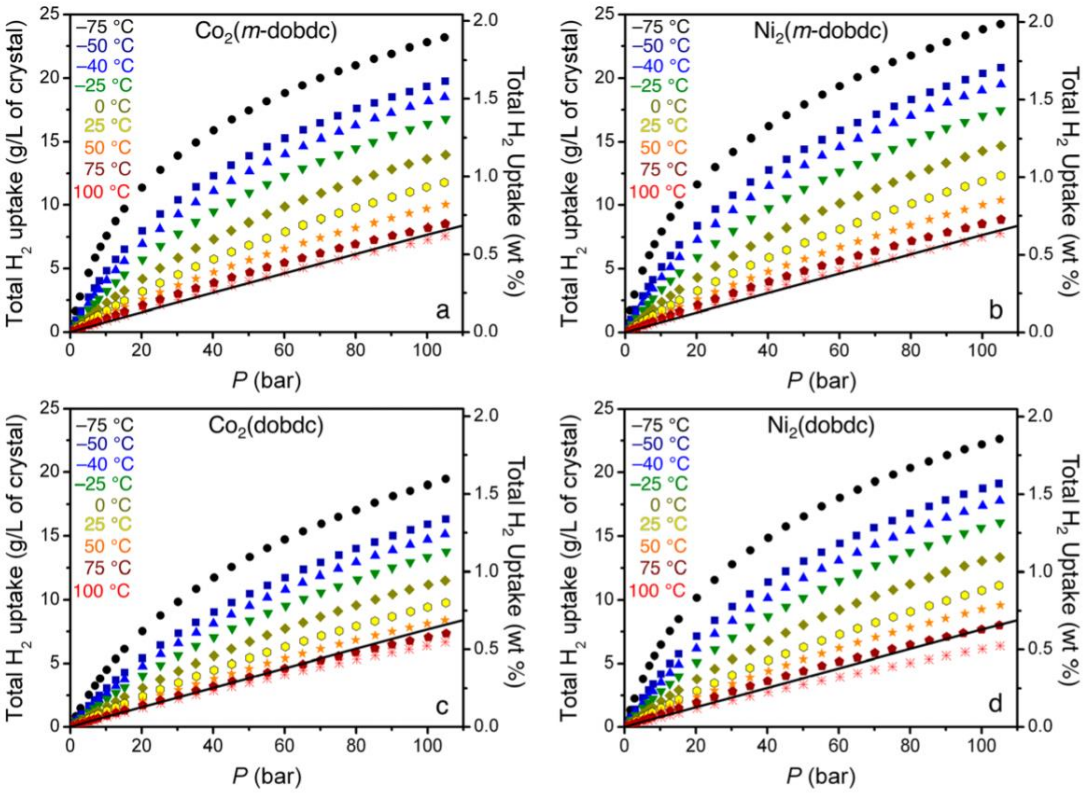


Figure 16: Hydrogen adsorption isotherms for (a) Co2(m-dobdc), (b) Ni2(m-dobdc), (c) Co2(dobdc), and (d) Ni2(dobdc) at −75 (black circles), −50 (navy squares), −40 (blue triangles), −25 (green upside-down triangles), 0 (gold diamonds), 25 (yellow hexagons), 50 (orange stars), 75 (dark red pentagons), and 100 °C (bright red crosses) measured between 0 and 100 bar and plotted in terms of total volumetric and gravimetric capacity. The black line in each plot represents the volumetric density of pure compressed H2 at 25 °C [20].

Lin et al. synthesized and studied three MOFs and found that the maximum amount of H_2 adsorbed increased with increasing pore size, while the maximum adsorbate density decreased with increasing pore size, indicating that the contrasting adsorbed H_2 density with increasing pore size suggests that an optimum pore size exists [21]. Therefore, it was concluded that a strategy of only increasing pore volume may not give the optimum hydrogen storage material. Volumetric hydrogen uptake generally decreases with increasing pore size, because H_2 molecules in the void central space of large pores do not experience the attractive forces from pore surfaces. Molecular simulation results of Farha et al. [22] show hydrogen "monolayer" adsorption to be proportional to the surface area, whereas hydrogen "pore filling" adsorption is proportional to the pore volume. Other calculations suggest that pores with a width of 9 Å are optimal to maximize the hydrogen capacities at 77 K and 100 bar. [23] With this pore size, the overlap of the interaction from multiple walls leads to the highest H_2 affinity.

Many MOFs with suitable pores for H_2 are constructed; their high H_2 storage capacities are summarized in Table 1.

Compound	H_2 uptake (wt%)	Temperature (K)	Pressure (bar)
MOF-5	10	77	100
MOF-177	7.5	77	70
IRMOF-6	4.63	77	45
IRMOF-8	3.4	77	33.7
SNU-4	3.7	77	50
Mn-BTT	6.9	77	90
Cu-BTT	3.7	77	90
Fe-BTT	4.1	77	95
Ni-MOF-74	2.95	77	10
Co-MOF-74	3.15	77	10
HKUST-1	3.6	77	50
SNU-5	5.22	77	50
SNU-6	10.0	77	70
PCN-6	1.9	77	1
PCN-6′	1.1	77	1
PCN-10	6.84	30	3.5
PCN-12	3.05	77	1
PCN-12′	2.4	77	1
PCN-20	6	77	35
PCN-46	7.2	77	60
PCN-66	6.65	77	45
NOTT-140	6	77	20

Table 1: Hydrogen uptakes of selected MOFs [24].

However, the strategy of only varying the pore size may not give the optimum hydrogen storage material and many different strategies have been up-to-date employed towards enhancing hydrogen storage capacity of MOFs. Since interaction of hydrogen molecules with the framework is weak, to increase hydrogen capacities, it is essential that both the surface area and the framework–hydrogen interaction strength are increased. Other than control of the pore size (e.g. increase of BET surface area, impregnation, catenation), some strategies include introduction of open metal sites in the metal oxide part as well as in the organic linker, doping of alkali elements onto the organic linker, functionalization of the organic linker and hydrogen spill-over [25].

As a way to increase the H_2 physisorption energy of MOFs, the functionalization of organic linkers shows a good effect and is the focus of this study. Currently, organic linkers of most MOFs have aromatic backbones such as benzene and naphthalene. According to the ab initio calculations performed so far, the H2 physisorption energy to aromatic organic linkers increases with the addition of $–NH_2$, $–CH_3$, and –OH groups due to their ability to enrich the aromatic system electronically [26].

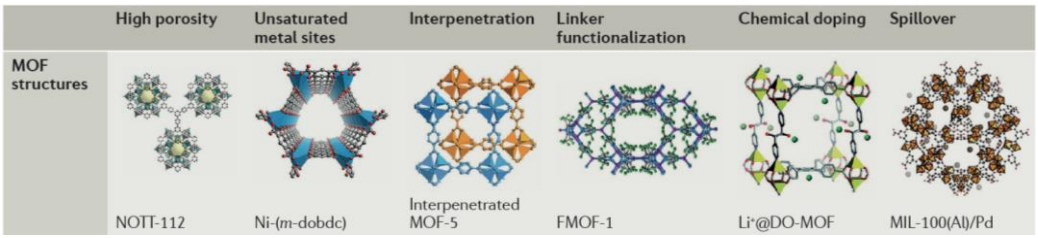


Figure 17: Strategies for improving hydrogen storage capacities in MOFs along with examples of the corresponding MOF candidates.

The design flexibility and tuning of properties of MOFs have complicated the search and identification of optimal compositions as the parameter space is very extended. To this direction, computational methods are able to accelerate the demanding quest; recently, many high-throughput studies have searched the real and hypothetical MOF databases to identify MOFs that simultaneously exhibit high volumetric and gravimetric hydrogen uptakes.

Database	Number of MOFs screened	H_2 storage condition
CoRE+UM (real)	5,309	Usable: (77 K, 100 bar) → (77 K, 5 bar) Usable: (77 & 298 K, 100 bar) → (77 & 298 K, 1 bar)
Northwestern (hypothetical)	137,953	Usable: (77 & 298 K, 100 bar) → (77 & 298 K, 1 bar) Usable: (77 K, 100 bar) → (77 K, 2 bar) Total: 1, 50, and 100 atm at 77 K
ToBaCCo (hypothetical)	13, 512	Total: 100 bar at 130, 200, and 243 K Usable: (77 K, 100 bar) → (77 K, 5 bar)
Mg-MOFs (hypothetical)	18,383	Usable: (243 K, 100 bar) → (243 K, 2 bar) Total: 243 K & 100 bar
UM (real)	~4,000	Total: 77 K & 35 bar

Table 2: Summary of recent high-throughput calculations of hydrogen storage in MOFs [27].

Notably, in the work of Siegel and coworkers [27], nearly 500,000 MOFs were screened computationally and the most promising candidates were assessed experimentally. Three MOF candidates, SNU-70, UMCM-9, and PCN-610/NU-100 were found to surpass the capacities of IRMOF-20, the record holder for balanced hydrogen capacity.

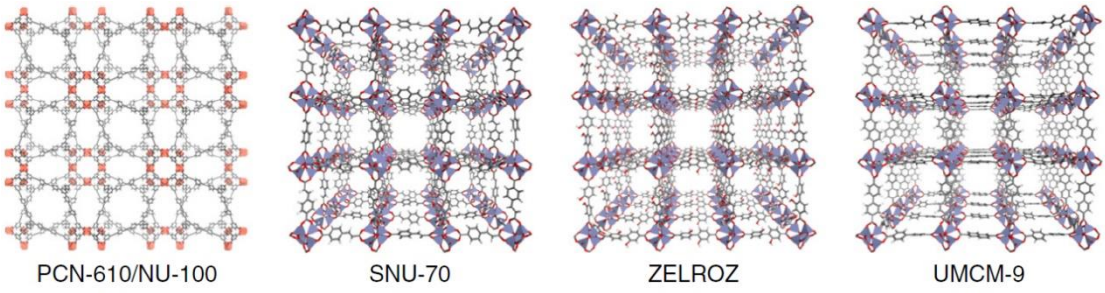


Figure 18: Crystal structures of MOFs identified by computational screening and were assessed experimentally afterwards [27]. Atom colors: C dark gray, H white, O red, Cu orange, Zn blue.

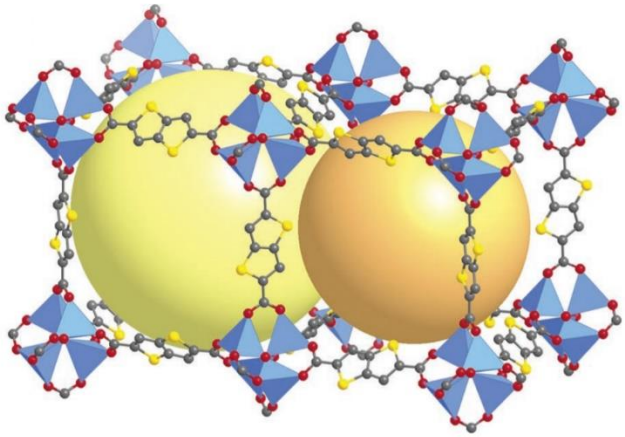


Figure 19: The structure of IRMOF-20, the record-holder for balanced gravimetric and volumetric hydrogen capacities [28]. Atom colors: C dark gray, H white, O red, S orange, Zn blue.

1.3 Purpose of this study

The main scope of this study is to identify promising functional groups that enhance the hydrogen storage capacities of MOFs and provide guidance for future material development. To this purpose, a combination of quantum chemical and classical simulation methodology is vital, demonstrating the importance of multiscale modeling on the studies of adsorptive properties of MOFs.

Aiming to identify a promising functional group that binds strongly the hydrogen molecule and also build a valuable database for further theoretical and experimental studies, the interaction of hydrogen with a series of strategically functionalized aromatic molecules is studied in the first place, by using quantum mechanical methods. Taking one step further, electrostatic potential maps of the functionalized moieties and electron density redistribution plots were created in order to provide insights on the nature of the interaction of hydrogen with the functionalized molecules. Last but not least, by performing GCMC studies we would be able to evaluate if stronger interaction energies provided by the introduction of a functional group in the structure of the framework would lead to an enhanced hydrogen uptake at different thermodynamic conditions.

CHAPTER 2 Computational Methodology

2.1 Multiscale modeling of materials

The arts celebrate multiple perspectives.
One of their major lessons is that there are
many ways to see and interpret the world.
-Elliot W. Eisner

Some of the most fascinating and intriguing problems in all scientific fields involve multiple spatial or temporal scales. Important problems include multiscale modeling of fluids, solids, polymers, proteins, as well as various physical and chemical phenomena like adsorption, chemical reactions, and diffusion.

Multi-scale materials modeling in general can be defined as an approach that combines existing as well as emerging methods from diverse scientific disciplines to bridge the wide range of time and length scales that are inherent in a number of essential phenomena and processes in materials science and engineering.

Materials are made up of atoms and the interactions among them at the microscopic level determine the behavior of the material at the macroscopic scale. The growth of multiscale modeling is strongly coupled with the unique advantages of this approach. Multiscale modeling can help in the direction of developing new materials, reduce product costs through innovations in material design, decrease the number of costly large-systems scale experiments and increase product quality and performance.

In the context of materials simulations, one can distinguish four characteristic length levels;

- the atomic scale ($\sim 10^{-9}$ m or a few nanometers),
- the microscopic scale ($\sim 10^{-6}$ m or a few micrometers),
- the mesoscopic scale ($\sim 10^{-4}$ m or hundreds of micrometers) and
- the macroscopic scale ($\sim 10^{-2}$ m or centimeters and beyond).

Phenomena at each length scale typically have a corresponding timescale. According to the four length scales mentioned above, the timescale ranges roughly from femtoseconds to picoseconds, to nanoseconds, to milliseconds and beyond [29].

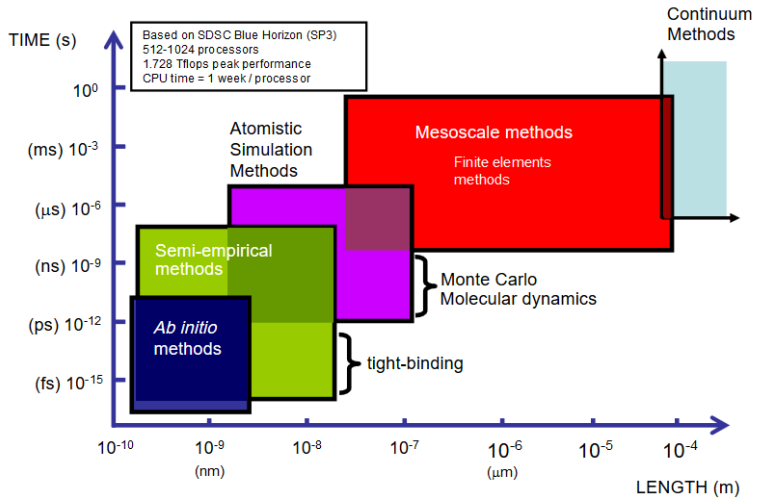


Figure 20: Molecular simulation methods at multiple length- and time scales. Hierarchical multiscale simulations utilize information obtained from simulations at a given (lower) characteristic length and time scales as an input for the next (upper) scale simulations.

At each length and timescale, efficient computational methods have been developed over the years to handle the relevant phenomena. Multi-scale *molecular* modeling consists of two basic categories of computational methods that can be used in the study of materials; quantum mechanical calculations and simulations based on classical physics.

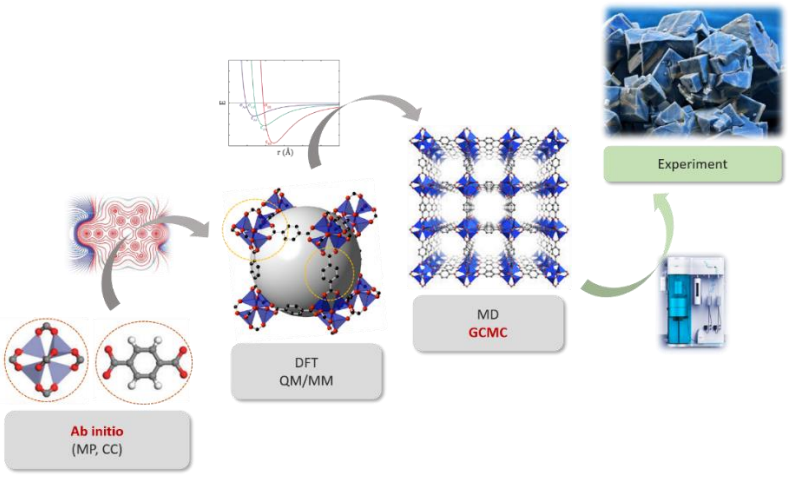


Figure 21: Multi-scale methodology scheme, showing the different levels of theory and the corresponding size of systems under study.

Quantum chemistry (QM) methods can be employed to treat electrons explicitly at the atomic scale and can provide structural, electronic and dynamic properties of the system with high accuracy. Using no parameters to the system other than solving the Schrodinger equation from first principles, they cannot handle more than a few tens of atoms as their computational cost increases dramatically with the number of electrons of the system. On the other hand, methods based on Density Functional Theory, can be readily applied to systems containing several hundred of atoms.

Classical simulations include several types of methods, such as Monte Carlo (MC), Molecular Dynamics (MD) and Molecular Mechanics (MM). Having the advantage of being able to treat larger systems, they can provide insight into atomic processes involving considerably larger systems, reaching up to millions of atoms and timescales up to a microsecond. They take into consideration the thermodynamic conditions of the system under study. However, their major disadvantage is that their results and accuracy depend on the parameters. For material properties at the microscopic scale, MD/MC simulations are usually performed employing classic interatomic potentials, which can often be derived from highly accurate QM or DFT calculations.

Overall, within the concept of multiscale approach in general, we try to predict the performance and behavior of materials across all relevant length and timescales and try to achieve a balance among accuracy, efficiency, and realistic description. In this project, we followed a bottom-up approach, beginning with ab initio calculations to study the strength and the nature of the interaction between hydrogen guest molecules individual parts of MOF structures. Going one step up, we introduce the topology of our systems by studying the whole organic linker and metal cluster with DFT methods. The success of our computational model depends heavily on the step of the accurate description of the intermolecular interactions between the guest hydrogen molecules and the MOF atoms. In our approach, we use our ab initio derived results to fit the parameters of an already existing interatomic potential onto these results. Having an accurate description of the intermolecular interaction, we can perform Monte Carlo simulations in the Grand Canonical ensemble. From these simulations we measure the adsorption profile of our system, we extract the adsorption isotherms and we can directly compare to experiment as GCMC calculations give us directly the number of hydrogen molecules inside the MOF for different thermodynamic conditions.

2.2 Quantum Chemistry Methods

I consider that I understand an equation
when I can predict the properties of each solution
without actually solving it.
- P. Dirac

Quantum Chemistry is the application of quantum mechanical postulates in the study of molecular systems. Many of the properties of atoms, molecules and solids can be calculated by determining the eigenfunctions of the many-body Hamiltonian. While Schrodinger equation of the hydrogen atom can be solved exactly, for systems having more than a few electrons the exact solution is an impossible task. As P. Dirac commented in 1929, "The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble."

Since molecular binding energies vary from hundredths to hundreds of kcal/mol depending on the governing nature of the interaction, the daunting challenge is to develop theoretical and computational approaches that are capable of accurately describing molecular binding over this wide range of values. Several methods have been formulated to find accurate approximations. There are two broad classes of methods, wave function-based and density-based, each subdivided into different approaches. In wave function-based methods, the wave function is written explicitly whereas in density-based, energy is written in terms of electron density.

In this chapter, the basic concepts and approximations of ab initio and density functional methods are presented, in an attempt to provide a conceptual framework for the quantum chemistry calculations conducted in this study.

2.2.1 Ab initio methods

In order to present the postulates of ab initio approaches we begin by formulating the electronic problem. With the time-independent Schrodinger equation as the starting point, we introduce the Born-Oppenheimer approximation as well as the basic ideas of the Hartree-Fock method. Subsequently, considering the limitations of the Hartree-Fock theory, we focus on more sophisticated methods that incorporate the effect of electron correlation and are able to give quantitatively accurate predictions.

A. Born-Oppenheimer Approximation and the concept of the Potential Energy Surface

Named after Max Born and J. Robert Oppenheimer in 1927 [30], the Born-Oppenheimer approximation (BO) lies at the core of quantum chemistry and molecular physics. The approach rests on the fact that since the masses of the nuclei are much greater than the masses of electrons (the resting mass of the lightest nucleus, the proton, is 1836 times heavier than the resting mass of one electron), the electrons can adjust almost instantaneously to any changes in the positions of the nuclei. Hence, to a good approximation, one can consider the electrons in a molecule to be moving in the field of fixed nuclei and the motion of the electrons can therefore be decoupled from the motion of the nuclei.

The BO or "clamped nuclei" approximation is intrinsic to the concept of the potential energy surface (PES). Considering a specific nuclear configuration for the molecular system under study, the Schrodinger equation is solved for the electrons in the static electric field generated by the fixed nuclei. Then, selecting a different nuclear geometry and repeating the calculations, a different solution to the electronic problem is derived. These solutions constitute the PES, whose global minimum corresponds to the most stable nuclear configuration.

Many sorts of macroscopic concepts in Chemistry, such as energetics, reactivity, equilibria, etc., are traced back to the notion of the PES. The dimensionality of a PES equals the internal degrees of freedom of the system, i.e. 3N-6 (or 3N-5 for linear molecules).

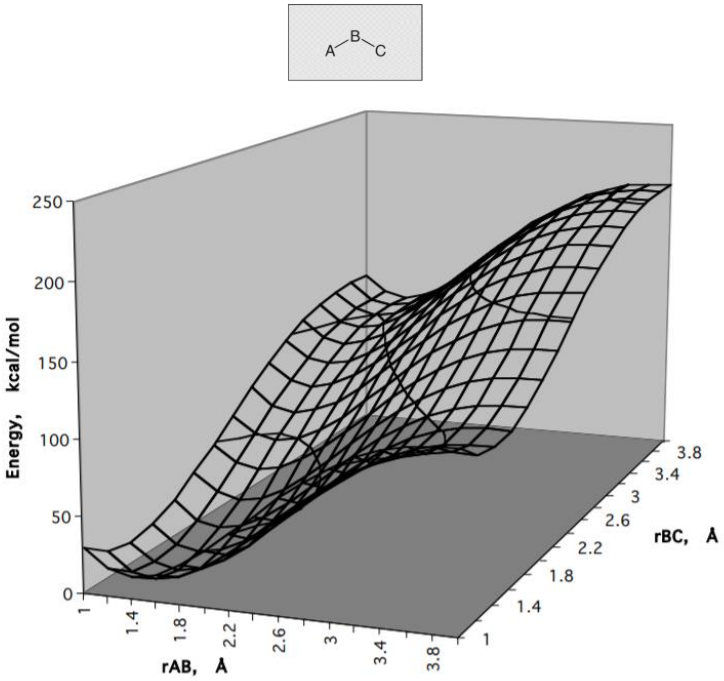


Figure 22: A 2-dimensional slice of a molecule with three degrees of freedom; the bond angle is held fixed and defines the slice [31].

Due to the importance of the BO approximation, we consider a more detailed description.

Our main interest is to find approximate solutions of the non-relativistic time-independent Schrodinger equation $\hat{H}|\Psi\rangle = E|\Psi\rangle$ (2.1). In atomic units, the Hamiltonian operator for a system of *N* electrons and *M* nuclei described by position vectors $\boldsymbol{r_i}$ and $\boldsymbol{R_A}$ respectively, is

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{\boldsymbol{r_{iA}}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{\boldsymbol{r_{ij}}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{\boldsymbol{R_{AB}}} \quad (2.2)$$

where $\boldsymbol{r_{iA}}$ is the distance between the ith electron and the Ath nucleus, $\boldsymbol{r_{ij}}$ the distance between the ith and jth electron and $\boldsymbol{R_{AB}}$ the distance between the Ath and Bth nucleus. M_A and Z_A are the mass and the atomic number of nucleus *A*, respectively.

Written in a more compact form,

$$\hat{H} = \hat{T}_e + \hat{T}_N + \hat{V}_{eN} + \hat{V}_{ee} + \hat{V}_{NN} \quad (2.3)$$

where

$$\hat{T}_e = -\frac{1}{2}\sum_{i=1}^{N} \nabla_i^2 \quad (2.3\alpha),$$

$$\hat{T}_N = -\sum_{A=1}^{M} \frac{1}{2M_A} \nabla_A^2 \quad (2.3\beta),$$

$$\hat{V}_{eN} = -\sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{\boldsymbol{r_{iA}}} \quad (2.3\gamma),$$

$$\hat{V}_{ee} = \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{\boldsymbol{r_{ij}}} \quad (2.3\delta),$$

$$\hat{V}_{NN} = \sum_{A=1}^{M}\sum_{B>A}^{M} \frac{Z_A Z_B}{\boldsymbol{R_{AB}}} \quad (2.3\varepsilon).$$

The two first terms represent the kinetic energy of the electrons and the nuclei, respectively; the third term represents the Coulomb attraction between electrons and nuclei; the fourth and fifth terms represent the repulsion between electrons and between nuclei, respectively.

In the light of the BO approximation, the kinetic energy of the nuclei, $\hat{T}_N$, can be neglected and the repulsion between the nuclei, $\hat{V}_{NN}$, can be considered to be constant. The remaining terms of (2.3) constitute the electronic Hamiltonian, $\hat{H}_{elec} = \hat{T}_e + \hat{V}_{eN} + \hat{V}_{ee}$ (2.4) , i.e. the Hamiltonian describing the motion of N electrons in the field of M point charges. The term $\hat{V}_{eN}$ is large and cannot be neglected; however, we can make the **R** dependence parametric so that total wave function is given as $|\Psi_{tot}\rangle = |\Psi_{elec}(\{\boldsymbol{r_i}\};\{\boldsymbol{R_A}\})\rangle\, |\Psi_{nucl}(\{\boldsymbol{R_A}\})\rangle$ (2.5) where Ψ_{elec} depends *explicitly* on the electronic coordinates $\{\mathbf{r_i}\}$ and *parametrically* on the nuclear coordinates $\{\mathbf{R_A}\}$. Having found the electronic energy eigenvalues by solving the electronic Schrodinger equation $\hat{H}_{elec}|\Psi_{elec}\rangle = E_{elec}|\Psi_{elec}\rangle$ (2.6), the total energy E_{tot} is derived by simply adding the term of the constant nuclear repulsion. $E_{tot} = E_{elec} + \sum_{A=1}^{M}\sum_{B>A}^{M} \frac{Z_A Z_B}{\boldsymbol{R_{AB}}}$ (2.7). Equations (2.4) to (2.7) constitute the electronic problem.

Once the electronic problem is solved, the problem for the motion of the nuclei can be tackled if we consider nuclei to move in the average field generated by the electrons. As the electrons move much faster than the nuclei, it is a reasonable approximation in (2.3) to replace the electronic coordinates by their average values. The nuclear Hamiltonian is thus:

$$\hat{H} = -\sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 + \left\langle -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{\boldsymbol{r_{iA}}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{\boldsymbol{r_{ij}}}\right\rangle + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{\boldsymbol{R_{AB}}}$$

$$= -\sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 + E_{elec}(\{\boldsymbol{R}_A\}) + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{\boldsymbol{R_{AB}}}$$

$$= -\sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 + E_{tot}(\{\boldsymbol{R}_A\}) \quad (2.8)$$

The total energy $E_{tot}(\{\boldsymbol{R}_A\})$ constitutes the potential energy surface. Thus, the nuclei move in a potential energy surface obtained by solving the electronic problem.

The solutions of the nuclear Schrodinger equation, $\hat{H}_{nucl}|\Psi_{nucl}\rangle = E|\Psi_{nucl}\rangle$, describe the vibration, rotation, and translation of a molecule, and E is the BO approximation to the total energy of (2.1) and includes the electronic, vibrational, rotational and translational energy.

B. Hartree-Fock Approximation

The electronic Schrodinger equation that results from the time-independent Schrodinger equation after invoking the Born-Oppenheimer approximation is still insoluble due to the interaction of the electrons. The essence of the Hartree-Fock approximation is to replace the complicated many-electron problem by a one-electron problem in which electron-electron repulsion is treated in an average way.

The electronic problem for the simplest atom, hydrogen, which has only one electron can be solved exactly. Therefore, it is reasonable to start by assuming that electrons do not interact with each other (i.e. that $\hat{V}_{ee} = 0$). Taking this assumption, the Hamiltonian would be separable and the total electronic wavefunction describing the motions of N non-interacting electrons would be of the form

$$\Psi_{HP}(\mathbf{r_1}, \mathbf{r_2}, \ldots, \mathbf{r_N}) = \varphi_1(\mathbf{r_1})\varphi_2(\mathbf{r_2}) \ldots \varphi_N(\mathbf{r_N}) \quad (2.9)$$

which is known as a *Hartree Product*. However, this functional form fails to satisfy the *antisymmetry principle*, which states that a wavefunction describing fermions should be antisymmetric with respect to the interchange of any set of space-spin coordinates. Equation (2.9) includes only spatial coordinates. In order to have a complete description of the electron,

we have to take into account not only the three spatial coordinates **r**, but also a coordinate for spin. We call a generic spin coordinate ω, and the complete set of space-spin coordinates $\mathbf{x} = \{\mathbf{r}, \omega\}$. We change the notation for orbitals from $\varphi(\mathbf{r})$, a *spatial orbital*, to $\chi(\mathbf{x})$, a *spin orbital*, where the spin orbital is just the product of a spatial orbital and either the $\alpha(\omega)$ or $\beta(\omega)$ spin function, for spin up and spin down respectively. With the full set of coordinates, the Hartree Product becomes

$$\Psi_{HP}(\mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_N}) = \chi_1(\mathbf{x_1})\chi_2(\mathbf{x_2}) \dots \chi_N(\mathbf{x_N}) \quad (2.10)$$

For a two-electron problem, the antisymmetry principle can be satisfied by a wavefunction of the form

$$\Psi(\mathbf{x_1}, \mathbf{x_2}) = \frac{1}{\sqrt{2}}[\chi_1(\mathbf{x_1})\chi_2(\mathbf{x_2}) - \chi_1(\mathbf{x_2})\chi_2(\mathbf{x_1})] \quad (2.11)$$

Rewriting the above functional form as

$$\Psi(\mathbf{x_1}, \mathbf{x_2}) = \frac{1}{\sqrt{2}}\begin{vmatrix} \chi_1(\mathbf{x_1}) & \chi_2(\mathbf{x_1}) \\ \chi_1(\mathbf{x_2}) & \chi_2(\mathbf{x_2}) \end{vmatrix} \quad (2.12)$$

we can easily see that if one tries to put two electrons in the same orbital at the same time, i.e. if $\chi_1 = \chi_2$, then $\Psi(\mathbf{x_1}, \mathbf{x_2}) = 0$ which is nothing more than the Pauli exclusion principle, which is a consequence of the antisymmetry principle. Generalizing to *N* electrons, the antisymmetric wavefunction is written as a determinant of spin orbitals which is called *Slater determinant*

$$\Psi = \frac{1}{\sqrt{N!}}\begin{vmatrix} \chi_1(\mathbf{x_1}) & \chi_2(\mathbf{x_1}) & \dots & \chi_N(\mathbf{x_1}) \\ \chi_1(\mathbf{x_2}) & \chi_2(\mathbf{x_2}) & \dots & \chi_N(\mathbf{x_2}) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{x_N}) & \chi_2(\mathbf{x_N}) & \dots & \chi_N(\mathbf{x_N}) \end{vmatrix} \quad (2.13)$$

where one can see that all electrons are indistinguishable and that each electron is associated with every orbital. Interchanging the coordinates of two electrons corresponds to interchanging two rows of the Slater determinant, which changes the sign of the determinant. Thus, Slater determinants meet the requirement of the antisymmetry principle. When two electrons occupy the same spin-orbital (i.e. two columns are equal) the determinant is zero; no more than one

electron can occupy a spin-orbital (Pauli exclusion principle). Writing the determinant in a more compact form:

$$\Psi = |\chi_i(\mathbf{x_1})\chi_j(\mathbf{x_2}) \dots \chi_k(\mathbf{x_N})\rangle = |\chi_i\chi_j \dots \chi_k\rangle = |ij \dots k\rangle \quad (2.14)\,,$$

where the normalization factor is implied.

Knowing the functional form of the wavefunction, we proceed to the notation of the Hamiltonian. We define a one-electron operator $h(i)$ as

$$h(i) = -\frac{1}{2}\nabla_i^2 - \sum_A \frac{Z_A}{r_{iA}} \quad (2.15)$$

and a two-electron operator $v(i,j)$ as

$$v(i,j) = \frac{1}{r_{ij}} \quad (2.16)$$

and we can write the electronic Hamiltonian in the simpler form of

$$\widehat{H_{el}} = \sum_i h(i) + \sum_{i<j} v(i,j) + \hat{V}_{NN} \quad (2.17)$$

Since $\hat{V}_{NN}$ is just a constant for a fixed set of nuclear coordinates {**R**} we will ignore it as it does not change the eigenfunctions but only shifts them. The electronic energy, assuming that the wavefunction is normalized will be given by

$$E_{el} = \langle\Psi|\widehat{H_{el}}|\Psi\rangle \quad (2.18)$$

For symmetric energy expressions, we can employ the *variational theorem*, which states that the energy is always an upper bound to the true energy.

Hence, one can obtain better approximate wavefunctions by varying their parameters until the energy within the given functional space is minimized; the correct molecular orbitals being those who minimize the electronic energy E_{el}. The molecular orbitals can be expressed as a linear combination of a set of atomic orbital basis functions (LCAO).

It can be shown that the Hartree-Fock energy E_{el} can be rewritten in terms of integrals of the one- and two-electron operators as

$$E_{HF} = \sum_i \langle i|h|i\rangle + \frac{1}{2}\sum_{ij}\{\langle ii|jj\rangle - \langle ij|ji\rangle\} \quad (2.19)$$

where the one-electron integral is

$$\langle i|h|j\rangle = \int d\mathrm{x}_1 \chi_i^*(\mathrm{x}_1)\, h(\mathbf{r}_1)\, \chi_j(\mathrm{x}_1) \quad (2.20)$$

and the two-electron integral is

$$\langle ij|kl\rangle = \int d\mathrm{x}_1 d\mathrm{x}_2 \chi_i^*(\mathrm{x}_1)\,\chi_j(\mathrm{x}_1)\frac{1}{r_{12}}\chi_k^*(\mathrm{x}_2)\chi_l(\mathrm{x}_2) \quad (2.21)$$

There exist efficient algorithms for computing such one- and two-electron integrals, with two-electron integrals to be computationally demanding. In order to minimize the Hartree-Fock energy expression with respect to changes in the orbitals ($\chi_i \rightarrow \ \chi_i + \delta\chi_i$) we use Lagrange's method of undetermined multipliers and eventually arrive at the Hartree-Fock equations defining the orbitals

$$h(\mathrm{x}_1)\chi_i(\mathrm{x}_1) + \sum_{i\neq j}\left[\int d\mathrm{x}_2 \left|\chi_j(\mathrm{x}_2)\right|^2 r_{12}^{-1}\right]\chi_i(\mathrm{x}_1) - \sum_{j\neq i}\left[\int d\mathrm{x}_2 \chi_j^*(\mathrm{x}_2)\,\chi_i(\mathrm{x}_2) r_{12}^{-1}\right]\chi_j(\mathrm{x}_1) = \epsilon_i\chi_i(\mathrm{x}_1) \quad (2.22)$$

where ϵ_i is the energy eigenvalue associated with the orbital χ_i.

- The term $\sum_{i\neq j}\left[\int d\mathrm{x}_2 \left|\chi_j(\mathrm{x}_2)\right|^2 r_{12}^{-1}\right]\chi_i(\mathrm{x}_1)$ is called the *Coulomb term* and expresses the Coulomb interaction of an electron in spin orbital χ_i with the average charge distribution of the other electrons. We define a Coulomb operator as

  $$J_j(\mathrm{x}_1) = \int d\mathrm{x}_2 \left|\chi_j(\mathrm{x}_2)\right|^2 r_{12}^{-1} \quad (2.23)$$

  which gives the average local potential at point x_1 due to charge distribution from the electron in orbital χ_j.

- The term $\sum_{j\neq i}\left[\int d\mathrm{x}_2 \chi_j^*(\mathrm{x}_2)\,\chi_i(\mathrm{x}_2) r_{12}^{-1}\right]\chi_j(\mathrm{x}_1)$ looks like the Coulomb term except that it switches or exchanges spin orbitals χ_i and χ_j. It arises from the antisymmetry requirement of the wavefunction and does not have a simple classical analog. We can define an exchange operator in terms of its action on an arbitrary spin orbital χ_i as

  $$K_j(\mathrm{x}_1)\,\chi_i(\mathrm{x}_1) = \left[\int d\mathrm{x}_2 \chi_j^*(\mathrm{x}_2)\,\chi_i(\mathrm{x}_2) r_{12}^{-1}\right]\chi_j(\mathrm{x}_1) \quad (2.24)$$

In terms of these Coulomb and exchange operators, the Hartree-Fock equations become:

$$\left[h(\mathrm{x}_1) + \sum_{j\neq i} J_j(\mathrm{x}_1) - \sum_{j\neq i} K_j(\mathrm{x}_1)\right]\chi_i(\mathrm{x}_1) = \epsilon_i\chi_i(\mathrm{x}_1) \quad (2.25)$$

where it is clear that Hartree-Fock equations are eigenvalue equations.

If we realize that $[J_j(\mathrm{x}_1) - K_j(\mathrm{x}_1)]\,\chi_i(\mathrm{x}_1) = 0$ then we can introduce the *Fock operator* as:

$$f(\mathrm{x}_1) = h(\mathrm{x}_1) + \sum_j [J_j(\mathrm{x}_1) - K_j(\mathrm{x}_1)] \quad (2.26)$$

where we can define $u^{HF}(\mathrm{x}_1) = \sum_j [J_j(\mathrm{x}_1) - K_j(\mathrm{x}_1)]$ as the Hartree-Fock potential which is

actually the *average* potential experienced by the ith electron due to the presence of the other electrons. The Hartree-Fock equations thus become:

$$f(\mathrm{x}_1)\chi_i(\mathrm{x}_1) = \epsilon_i\chi_i(\mathrm{x}_1) \quad (2.27)$$

The Hartree-Fock potential $u^{HF}(\mathrm{x}_1)$, or equivalently the "field" seen by the ith electron, depends on the spin orbitals of the other electrons (i.e. the fock operator depends on its eigenfunctions). Hence, the HF equation is nonlinear and must be solved iteratively; the procedure is called the self-consistent-field (SCF) method. The basic idea of the SCF method is the following; by making an initial guess at the spin orbitals, one can calculate the average field u^{HF} seen by each electron and then solve the eigenvalue equation for a new set of spin orbitals. Using these new spin orbitals, one can obtain new fields and repeat the procedure until self-consistency is reached (i.e. until the fields no longer change and the spin orbitals used to construct the Fock operator are the same as its eigenfunctions).

The solution of the HF eigenvalue problem yields a set $\{\chi_k\}$of orthonormal HF spin orbitals with orbital energies $\{\epsilon_k\}$. The *N* spin orbitals with the lowest energies are the *occupied* spin orbitals. The Slater determinant formed from these orbitals is the HF ground state wavefunction and is the best variational approximation to the ground state of the system.

The integro-differential Hartree-Fock equations can be solved numerically only for atoms due to the spherical symmetry of their spin orbitals. Roothaan and Hall showed that by expanding the molecular orbitals into an appropriate spatial basis (i.e. a set of atomic orbital basis functions), the integro-differential equation can be converted to a set of equations that can be solved with linear algebra tools. Denoting the atomic orbital basis functions as $\tilde{\chi}$, we have the expansion

$$\chi_i = \sum_{\mu=1}^{K} C_{\mu i}\tilde{\chi}_\mu \quad (2.28)$$

for each spin orbital *i*.

This leads to

$$f(\mathrm{x}_1)\sum_v C_{vi}\tilde{\chi}_v(\mathrm{x}_1) = \epsilon_i\sum_v C_{vi}\tilde{\chi}_v(\mathrm{x}_1) \quad (2.29)$$

Left multiplying by $\tilde{\chi}_{\mu}^{*}(\mathrm{x}_1)$ and integrating yields a matrix equation

$$\sum_{v} C_{vi} \int d\mathrm{x}_1 \tilde{\chi}_{\mu}^{*}(\mathrm{x}_1)\, f(\mathrm{x}_1)\, \tilde{\chi}_{v}(\mathrm{x}_1) = \epsilon_i \sum_{v} C_{vi} \int d\mathrm{x}_1 \tilde{\chi}_{\mu}^{*}(\mathrm{x}_1)\, \tilde{\chi}_{v}(\mathrm{x}_1) \quad (2.30)$$

that can be simplified by introducing the matrix element notation

$$S_{\mu v} = \int d\mathrm{x}_1 \tilde{\chi}_{\mu}^{*}(\mathrm{x}_1)\, \tilde{\chi}_{v}(\mathrm{x}_1) \quad (2.31)$$

$$F_{\mu v} = \int d\mathrm{x}_1 \tilde{\chi}_{\mu}^{*}(\mathrm{x}_1)\, f(\mathrm{x}_1)\, \tilde{\chi}_{v}(\mathrm{x}_1) \quad (2.32)$$

to the

$$\sum_{v} F_{\mu v} C_{vi} = \epsilon_i \sum_{v} S_{\mu v} C_{vi} \quad (2.33)$$

or as matrices to the

$$FC = SC\epsilon \quad (2.34)$$

where ϵ is a diagonal matrix of the orbital energies ϵ_i. Considering the actual calculation of HF wavefunctions, we have to be more specific about the form of the spin orbitals. There are two types of spin orbitals; restricted spin orbitals, which are constrained to have the same spatial function for α and β spin functions, and unrestricted spin orbitals, which have different spatial functions for α and β spins. The restricted closed-shell set of spin orbitals leads to restricted closed-shell wavefunctions via the Roothaan-Hall equations. An unrestricted open-shell set of spin orbitals leads to unrestricted open-shell wavefunctions via the Pople-Nesbet equations. The Roothaan-Hall equation is a generalized matrix eigenvalue equation of the form

$$FC = SC\epsilon \quad (2.34)$$

These equations have a non-trivial solution only if the secular equation is satisfied;

$$\det(F - \epsilon_{\alpha} S) = 0 \quad (2.35)$$

and can be solved with the use of the self-consistent method where in each cycle, a new set of C coefficients is produced until the convergence criteria are met. By procedure analogous of those used to derive the Roothaan equations, one can derive the corresponding unrestricted open-shell Pople-Nesbet equations

$$F^{\alpha} C^{\alpha} = S C^{\alpha} \epsilon^{\alpha} \quad (2.36)$$

$$F^{\beta} C^{\beta} = S C^{\beta} \epsilon^{\beta} \quad (2.37)$$

These equations can be solved in a manner similar to the way Roothaan's equations are solved, except that F^{α} and F^{β} depend on both C^{α} and C^{β}, thus the two matrix eigenvalue problems must be solved simultaneously.

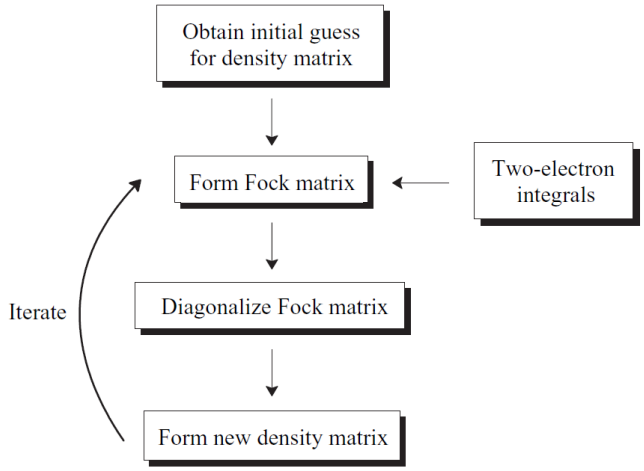


Figure 23: Illustration of the SCF procedure.

The quality of the initial guess orbitals influences the number of iterations necessary for achieving convergence. The simplest way to generate a set of initial orbitals is to diagonalize the Fock matrix consisting only of one-electron contributions, i.e. totally neglecting the electron-electron repulsion. This is generally a poor guess; more sophisticated procedures involve taking the initial set of MO coefficients from a semi empirical calculation such as Extended Hückel Theory (EHT) [32]. The EHT method has the advantage that it is readily parametrized for all elements.

C. Electron Correlation and Post-Hartree-Fock Methods

Within the Hartree–Fock method, the antisymmetric wave function is approximated by a single Slater determinant. The single-determinant approximation does not take into account Coulomb correlation, leading to a total electronic energy different from the exact solution. Although with a sufficiently large basis set the HF wavefunction is able to account for ~99% of the total energy, the remaining ~1% is more than significant for describing chemically important effects such as dispersion. In order to calculate total energies with a chemical accuracy of ~1 kcal/mol, one has to use more sophisticated methods including electron correlation and large basis sets.

The correlation energy is defined as the difference between the exact total energy of the system and the Hartree-Fock energy. The Hartree-Fock ground state is of the single determinant form: $|\Psi_0\rangle = |\chi_1\chi_2 \cdots \chi_a\chi_b \cdots \chi_N\rangle$, comprising of *N* occupied spin-orbitals. A determinant that differs from $|\Psi_0\rangle$ by *N* spin-orbitals is called a *N-tuply excited determinant.* In order to take into account multiple determinants, one can consider the HF ground state to be a reference state and classify other possible determinants by how they differ from the reference state. The additional determinants can be used to represent approximate excited states of the system, or if used in a linear combination with $|\Psi_0\rangle$ they can provide a more accurate description of the ground state or any excited state of the system. A singly excited determinant is one in which an electron occupying a spin-orbital χ_a has been excited to a virtual spin-orbital χ_r:

$$|\Psi_a^r\rangle = |\chi_1\chi_2 \cdots \chi_r\chi_b \cdots \chi_N\rangle \quad (2.38)$$

A doubly excited determinant is one in which two electrons have been promoted from χ_a and χ_b to χ_r and χ_s respectively.

$$|\Psi_{ab}^{rs}\rangle = |\chi_1\chi_2 \cdots \chi_r\chi_s \cdots \chi_N\rangle \quad (2.39)$$

We can write the exact wavefunction for any state of the system as

$$|\Psi_0\rangle = c_0|\Psi_0\rangle + \sum_{ar} c_a^r|\Psi_a^r\rangle + \sum_{\substack{a<b \\ r<s}} c_{ab}^{rs}|\Psi_{ab}^{rs}\rangle + \sum_{\substack{a<b<c \\ r<s<t}} c_{abc}^{rst}|\Psi_{abc}^{rst}\rangle + \cdots \quad (2.40)$$

The infinite set of *N*-electron determinants $\{|\Psi_i\rangle\} = \{|\Psi_0\rangle, |\Psi_a^r\rangle, |\Psi_{ab}^{rs}\rangle, \ldots\}$ is a complete set for the expansion of any *N*-electron wavefunction. The exact energies of the ground and excited states of the system are the eigenvalues of the Hamiltonian matrix formed from the complete set $\{|\Psi_i\rangle\}$. The lowest eigenvalue of the Hamiltonian matrix, *E*, is the exact nonrelativistic ground state energy of the system within the BO approximation. The difference between this exact energy and the Hartree-Fock-limit energy, E_0, is the correlation energy $E_{corr} = E - E_0$.

There are three main methods for accounting the effect of electron correlation: Configuration Interaction (CI) [33], Many-body Perturbation theory (MBPT), and Coupled Cluster theory (CC) [34,35].

Møller–Plesset perturbation theory (MP)

The Møller–Plesset perturbation theory is a special case of Rayleigh-Schrödinger perturbation theory (PT) which is a systematic procedure for finding the correlation energy. In PT, the total Hamiltonian of the system is divided in two parts; a zero-order part, H_0 , which has known eigenfunctions and eigenvalues, and a small perturbation part, V.

$$H = H_0 + \lambda V \quad (2.41)$$

The exact energy is expressed as an infinite sum of contributions of increasing complexity

$$E = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \cdots \quad (2.42)$$

and the exact wavefunction in terms of the approximate wavefunctions as

$$|\Psi_i\rangle = \left|\Psi_i^{(0)}\right\rangle + \lambda \left|\Psi_i^{(1)}\right\rangle + \lambda^2 \left|\Psi_i^{(2)}\right\rangle + \cdots \quad (2.43)$$

Substituting these expansions into the Schrödinger equation, collecting terms according to powers of λ, multiplying the resulting equations by $\Psi_i^{(0)}$ and integrating over all space yields the following expressions for the n^{th}-order (MPn) energies

$$E_i^{(0)} = \left\langle \Psi_i^{(0)} \middle| H \middle| \Psi_i^{(0)} \right\rangle \quad (2.44)$$

$$E_i^{(1)} = \left\langle \Psi_i^{(0)} \middle| V \middle| \Psi_i^{(0)} \right\rangle \quad (2.45)$$

$$E_i^{(2)} = \left\langle \Psi_i^{(0)} \middle| V \middle| \Psi_i^{(1)} \right\rangle \quad (2.46)$$

$$E_i^{(3)} = \left\langle \Psi_i^{(0)} \middle| V \middle| \Psi_i^{(2)} \right\rangle \quad (2.47)$$

Since we are interested in obtaining a perturbation expansion for the correlation energy, in MP theory we choose the Hartree-Fock Hamiltonian as our zeroth-order Hamiltonian

$$H_0 = \sum_i f(i) = \sum_i [h(i) + u^{HF}(i)] \quad (2.48)$$

and

$$V = \sum_i \sum\nolimits_{j>i} {r_{ij}}^{-1} + V^{HF} = \sum_i \sum\nolimits_{j>i} {r_{ij}}^{-1} - \sum_i u^{HF}(i) \quad (2.49)$$

while the Hartree-Fock wavefunction $|\Psi_0\rangle$ is an eigenfunction of H_0

$$H_0 \left|\Psi_0\right\rangle = E_0^{(0)} \left|\Psi_0\right\rangle \quad (2.50)$$

where

$$E_0^{(0)} = \sum_a \varepsilon_a \quad (2.51)$$

is the zero[th] order energy and ε_a the energies of the spin-orbitals.

For the first order energy we have:

$$E_0^{(1)} = \langle \Psi_0 | V | \Psi_0 \rangle = -\frac{1}{2} \sum_{ab} \langle ab \| \, ab \rangle \quad (2.52)$$

and the Hartree-Fock energy is the sum of zeroth and first order energies

$$E_0 = E_0^{(0)} + E_0^{(1)} = \sum_a \varepsilon_a - \frac{1}{2} \sum_{ab} \langle ab || ab \rangle \quad (2.53)$$

Thus, the first correction of the Hartree-Fock energy corresponds to the second order of perturbation theory, i.e. the MP2.

$$E_0^{(2)} = \sum_{\substack{a<b \\ r<s}} \frac{\left| \langle \Psi_0 | \sum_i \sum_{j>i} {r_{ij}}^{-1} | \Psi_{ab}^{rs} \rangle \right|^2}{\varepsilon_a + \varepsilon_b - \varepsilon_r - \varepsilon_s} = \sum_{\substack{a<b \\ r<s}} \frac{|\langle ab || rs \rangle|^2}{\varepsilon_a + \varepsilon_b - \varepsilon_r - \varepsilon_s} \quad (2.54)$$

Expressions for higher order terms follow similarly, although with much greater algebraic and computational complexity. Because of its relatively modest computational cost, second-order Møller-Plesset perturbation (MP2) theory is one of the most widely used methods in quantum chemistry for studying noncovalent interactions. Depending on the size of the molecular system under study, more accurate calculations, such as coupled-cluster calculations with single, double, and perturbative triple excitations [CCSD(T)], may be computationally prohibited; the cost of benchmark-quality CCSD(T) formally scales as $O(N^7)$, where N is proportional to the system size, whereas MP2 scales as $O(N^5)$.

D. Basis Set Approximation

One of the approximations inherent in essentially all ab initio methods is the introduction of a basis set. A basis set is a set of mathematical functions from which the wavefunction is constructed. Expanding an unknown function, such as a molecular orbital, in a set of known functions is not an approximation if the basis set is complete, i.e. an infinite number of functions is used. However, in actual calculations, the expansion to an infinite set of functions is an impossible task. Thus, one has to identify mathematical functions that allow wavefunctions to approach the complete basis set limit in as efficient a manner as possible. In principle, the smaller the basis set, the poorer the representation. The type of basis functions used also influences the accuracy; the better a single basis function is able to reproduce the unknown function, the fewer basis functions are necessary for achieving a given level of accuracy.

There are two types of basis functions commonly used in electronic structure calculations; Slater Type Orbitals (STO) and Gaussian Type Orbitals (GTO). Slater type orbitals have the functional form

$$\varphi_{n,l,m}^{STO}(\zeta, \mathbf{r}) = \varphi_{n,l,m}^{STO}(\zeta, r, \theta, \varphi) = Nr^{n-1}e^{-\zeta r}Y_l^m(\theta, \varphi) \quad (2.55)$$

where N is a normalization constant and Y_l^m are spherical harmonic functions. The exponential dependence on the distance between the nucleus and the electron ($e^{-\zeta r}$), reflects the exact orbitals for the hydrogen atom. However, STOs do not have any radial nodes; nodes are introduced by constructing linear combinations of STOs. The calculation of the two-electron integrals cannot be performed analytically with the use of STOs.

Gaussian type orbitals can be written in terms of polar coordinates as:

$$\varphi_{n,l,m}^{GTO}(\alpha, \mathbf{r}) = \varphi_{n,l,m}^{GTO}(\alpha, r, \theta, \varphi) = Nr^{n-1}e^{-\alpha r^2}Y_l^m(\theta, \varphi) \quad (2.56)$$

or Cartesian coordinates as:

$$\varphi_{i,j,k}^{GTO}(\alpha, x, y, z) = Nx^i y^j z^k e^{-\alpha(x^2+y^2+z^2)} \quad (2.57)$$

where α is an exponent controlling the width of the GTO, and i, j, k, are non-negative integers that dictate the nature of the orbital in a Cartesian sense.

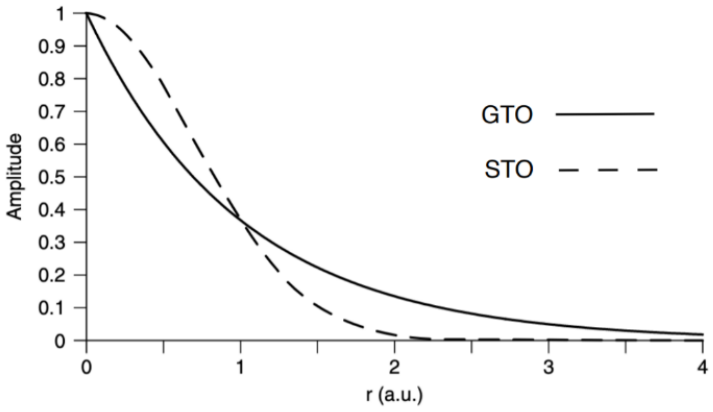


Figure 24: Slater vs Gaussian-type orbital [36].

GTOs with $i = j = k = 0$ have spherical symmetry and are called s-type GTOs, with $i = j = k = 1$ are p-type GTOs, with $i = j = k = 2$ are d-type GTOs, etc.

In terms of computational efficiency, GTOs are superior to STOs due to the ease of which the required integrals can be calculated. However, the r^2 exponential dependence makes GTOs inferior to STOs in two ways; GTOs fail to represent the proper behavior near the nucleus and also decay too fast with the distance. In order to combine the best feature of GTOs (i.e. computational efficiency) with that of STOs (proper radial shape), a linear combination of

GTOs is used to approximate STOs. The linear combination of Gaussians is referred to as a "contracted" basis function, and the individual Gaussians from which it is formed are the "primitive" Gaussians.

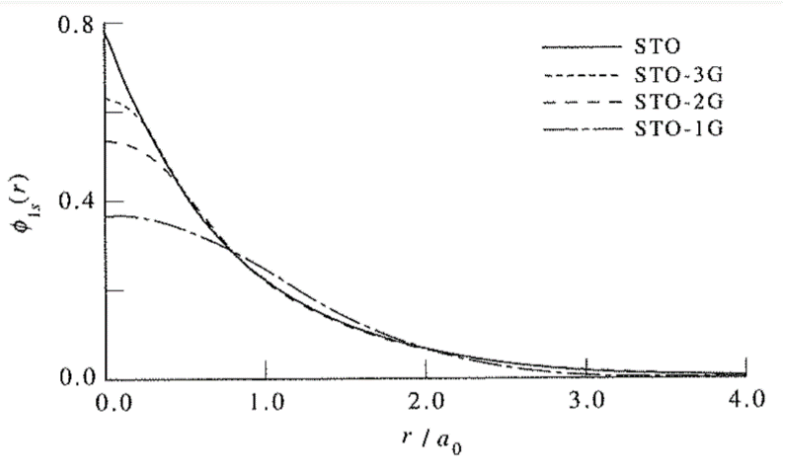


Figure 25: Approximating an STO with a linear combination of GTOs [37].

The STO-nG basis sets, where n is the number of primitive Gaussians are known as "single-ζ" basis sets and constitute the most common "minimal" basis set.
Given the number of the functions used per orbital , basis sets can be classified to Single Zeta, Double Zeta (DZ), Triple Zeta (TZ), Quadruple Zeta (QZ) and Quintuple Zeta (5Z) which correspond to basis sets with one, two, three, four, five functions for each orbital respectively.
Another classification is based on the fact that the chemical bonding occurs between valence orbitals; valence orbitals can vary widely as a function of chemical bonding whereas core orbitals are only weakly affected. "Split-valence" basis sets are basis sets in which there are multiple basis functions corresponding to each valence atomic orbital and are called valence double, triple, quadruple-zeta, and so on, basis sets. Amongst the most widely used split-valence basis sets are those of Pople *et al*. These basis sets include 3-21G, 6-21G, 4-31G, 6-31G, and 6-311G. In the same category belong also the basis sets developed by Dunning and coworkers, with the difference that these basis sets were optimized not only for HF calculations, but also for calculations including electron correlation. Examples include cc-pVDZ, cc-pVTZ, etc. where the acronym stands for "correlation-consistent polarized Valence (Double/Triple/etc.) Zeta".
Molecular orbitals require more mathematical flexibility than atoms. This flexibility is added via polarization functions which correspond to quantum numbers of higher angular momentum.

For independent-particle wavefunctions, where electron correlation is not considered, the first set of polarization functions describes adequately most of the important charge polarization effects. Adding a single set of polarization functions (i.e. *p*-functions on hydrogens and *d*-functions on heavy atoms) to the DZ basis, forms a *Double Zeta plus Polarization* (DZP) type basis. If methods including electron correlation are used, higher angular momentum functionals are essential. If two sets of polarization functions are added to a TZ basis, a *Triple Zeta plus Double Polarization* (TZ2P or TZPP) type basis is obtained.

The highest energy MOs of anions, highly excited electronic states, and loose supramolecular complexes, tend to be much more spatially diffuse than other MOs. In order to provide the necessary flexibility to allow a weakly bound electron to localize far from the remaining density, basis sets are "augmented" with diffuse basis functions. In the Pople family of basis sets [38], the presence of diffuse functions is indicated by a '+' in the basis set name (e.g. 6-31+G(d)). In the Dunning family of cc-pV*n*Z basis sets [39], diffuse functions on all atoms are indicated by the 'aug' prefix (e.g. aug-cc-pVTZ).

A well performing for general-purpose computations family of basis sets are the Karlsruhe basis sets [40] by Ahlrichs and coworkers that were designed as the second-generation default (or "def2") basis sets in the Turbomole [41] program (e.g. def2-TZVPP is a Triple Zeta basis with two sets of Polarization functions).

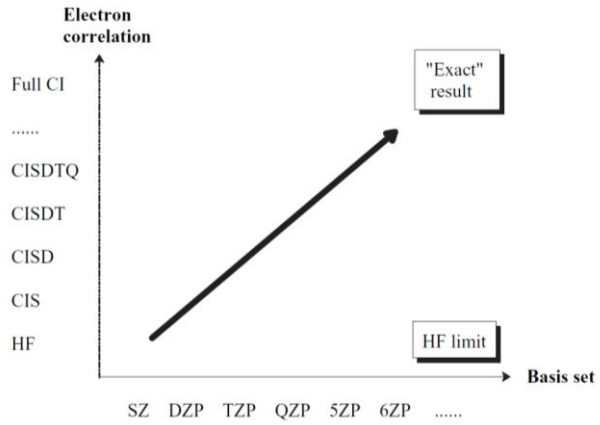


Figure 26: Convergence to the exact solution for methods that include electron correlation. The larger the one-electron expansion (basis set size) and the larger the many-electron expansion (number of determinants), the better the results.

Basis Set Superposition Error (BSSE)

In calculating molecular interactions, the use of *finite* basis sets introduces the so-called “basis set superposition error (BSSE)”. It consists of the fact that when the atoms of interacting molecules approach one another, their basis functions overlap. Each monomer "borrows" functions from other nearby components, effectively increasing its basis set. This effect increases as the atoms orientate themselves in closer distances, creating an effectively varying basis set with the interatomic distance.

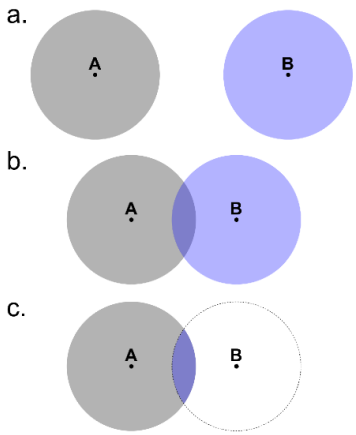


Figure 27: Origin of the basis set superposition error. (a) A and B far apart from each other; (b) basis orbitals overlap when A couples to B; (c) the region in blue indicates the part of the basis orbitals centered at B that are available to describe A.

In the case of a dimer, interaction energies (ΔE) between two atoms or molecules A and B are calculated as the difference between the product complex AB and its components A and B:

$$\Delta E = E_{AB}^{AB}(AB) - E_{A}^{A}(A) - E_{B}^{B}(B) \quad (2.58)$$

where the subscript denotes the geometry of the system and the superscript the used basis sets. The interaction energies calculated according to (2.58) lead to artificial stabilization energies with severe complications for systems with dominant dispersion forces or hydrogen bonds.

One obvious solution to the BSSE is the use of extremely large basis sets. This is, however, hardly feasible for most of the systems. In order to eliminate the BSSE, Boys and Bernardi introduced the counterpoise correction (CP) [42]. In the counterpoise correction, the artificial stabilization is countered by letting the separate atoms improve their basis sets by borrowing functions of an empty basis set. To realize such an empty basis set, a ghost atom is used. Ghost atoms have the basis orbitals of the according atom, but no charge or mass; they are basis set

functions which have no electrons or protons. The original Boys and Bernardi formula for the interaction energy between fragments A and B is:

$$\Delta E^{CP} = E_{AB}^{AB}(AB) - E_{A}^{A}(A) - E_{B}^{B}(B) - [E_{A}^{AB}(AB) - E_{A}^{AB}(A) + E_{B}^{AB}(AB) - E_{B}^{AB}(B)] \quad (2.59)$$

Here, $E_X^Y(Z)$ is the energy of fragment X calculated at the optimized geometry of fragment Y with the basis set of fragment Z.

Thus, one needs to do a total the following series of calculations; First, optimize the geometry of the dimer and the monomers with some basis set Z in order to get $E_{AB}^{AB}(AB)$, $E_{A}^{A}(A)$ and $E_{B}^{B}(B)$, i.e. the uncorrected interaction energy ΔE. Subsequently, delete fragment A (B) from the optimized structure of the dimer and re-run the single point calculation with basis set Z to get $E_{B}^{AB}(AB)$ and $E_{A}^{AB}(A)$.

Finally, calculate the energies of A and B at the dimer geometry but with the dimer basis set to compute $E_{A}^{AB}(AB)$ and $E_{B}^{AB}(B)$.

2.2.2 Density Functional Theory

From the previous analysis it has become clear that solving the many-particle equation for the ground state wavefunction and energy is a daunting task even for the smallest system. The problem with these methods is the huge computational effort, which makes it virtually impossible to apply them efficiently to larger, more complex systems. An alternative route is to shift the focus from the wavefunction Ψ, to a fundamental observable, namely the electronic density n.

In 1964, Hohenberg and Kohn laid the foundations of Density Functional Theory (DFT) [43]. In this theory, the key variable is the electron density $n(\mathbf{r})$

$$n(\mathbf{r}) = N \int \Psi^*(\mathbf{r}, \mathbf{r_2}, \dots, \mathbf{r_N}) \Psi(\mathbf{r}, \mathbf{r_2}, \dots, \mathbf{r_N}) d\mathbf{r}_2 d\mathbf{r}_N \quad (2.60)$$

which is a function of the three spatial coordinates. The electron density and the total energy depend parametrically from the nuclear positions as $n(\mathbf{r}) = n(\mathbf{r}; \mathbf{R_1}, \mathbf{R_2}, \dots, \mathbf{R}_M)$ and $E = E(\mathbf{R_1}, \mathbf{R_2}, \dots, \mathbf{R}_M)$ respectively.

DFT was made possible by the existence of two ingeniously simple theorems put forward and proven by Hohenberg and Kohn. The first H–K theorem demonstrates that the ground state properties of a many-electron system are *uniquely* determined by an electron density.

Since the density depends on only three spatial coordinates, the many-body problem of N electrons with $3N$ spatial coordinates can be reduced to three spatial coordinates, through the use of functionals of the electron density. The second H–K theorem states that a universal functional for the energy can be defined in terms of the electron density. The exact ground state is the global minimum value of this functional.

Within the BO approximation, the electronic energy, E_{elec}, can be written as a functional of the electron density as

$$E_{elec}[n(\mathbf{r})] = T[n(\mathbf{r})] + V_{eN}[n(\mathbf{r})] + V_{ee}[n(\mathbf{r})] \quad (2.61)$$

where $T[n(\mathbf{r})]$ is the kinetic energy of the electrons, $V_{eN}[n(\mathbf{r})]$ is the nuclear-electron attraction energy, and $V_{ee}[n(\mathbf{r})]$ is the electron-electron interaction energy.

- V_{eN} can be computed as

$$V_{eN}[n(\mathbf{r})] = \sum_{A=1}^{M} \int \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} n(\mathbf{r})\, d\mathbf{r} \quad (2.62)$$

- V_{ee} can be separated in two parts; $J[n(\mathbf{r})]$ and $Q[n(\mathbf{r})]$ where J is the classical electron-electron repulsion energy that can be computed according to

$$J[n(\mathbf{r})] = \frac{1}{2} \iint \frac{n(\mathbf{r_1}) n(\mathbf{r_2})}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 \quad (2.63)$$

  and Q is the quantum electron-electron interaction energy that has no classical analog and cannot be computed analytically.

- T is the kinetic energy of the electrons which cannot be directly written in terms of the electron density since it contains a derivative term

$$T[n(\mathbf{r})] = -\frac{1}{2} \int \Psi^*(\mathbf{r_1}, \dots, \mathbf{r_N}) \nabla^2 \Psi(\mathbf{r_1}, \dots, \mathbf{r_N}) d\mathbf{r} \quad (2.64)$$

  In order to solve this problem, instead of the full many-particle system, an auxilliary system of single-particle orbitals is considered, namely the Kohn-Sham orbitals, that have the same ground state density as the real system.

$$n(\mathbf{r}) = \sum_{i=1}^{N} |\varphi_i(\mathbf{r})|^2 \quad (2.65)$$

  The kinetic energy can then be written as the sum of the kinetic energies of the Kohn-Sham orbitals. The kinetic energy T can be therefore expressed as the single-particle kinetic energy (T_s) plus a correction ΔT, i.e., $T = T_s + \Delta T$, where

$$T_s[\{\varphi_i(\mathbf{r})\}] = -\frac{1}{2} \int \varphi_i{}^*(\mathbf{r}) \nabla^2 \varphi_i(\mathbf{r}) d\mathbf{r} \quad (2.66)$$

Since the single-particle kinetic energy T_s is not equal to the kinetic energy of the real many-particle kinetic energy T, the difference ΔT between these two terms is combined with Q, to define the exchange-correlation energy $E_{xc}[n(\mathbf{r})]$; namely,

$$E_{xc}[n(\mathbf{r})] = T[n(\mathbf{r})] - T_s[\{\varphi_i(\mathbf{r})\}] + Q[n(\mathbf{r})] \quad (2.67)$$

The only unknown term is now the exchange-correlation functional with the objective of DFT being the development of accurate approximate functionals for $T[n(\mathbf{r})]$ and $Q[n(\mathbf{r})]$. The exchange & correlation energy is often represented as a sum of an exchange functional $E_x[n(\mathbf{r})]$ and a correlation functional $E_c[n(\mathbf{r})]$ and is defined as the difference between the exact total energy of a system and the classical Hartree energy. The exchange energy is the Pauli repulsion (omitted in the Hartree term). Correlations are a result of the collective behavior of the electrons to screen and decrease the Coulombic interaction. Unlike the exchange term, the correlations become more pronounced for opposite spins, since the electrons are more likely to occupy nearby locations. Both exchange and correlation tend to keep electrons apart.

In the past 30 years, many non-empirical and semi-empirical density functionals have been developed. A major difficulty in their development is that density functionals are not systematically improvable, i.e. there is no guarantee that finer functional forms yield more accurate results across all types of interactions. One possible DFT hierarchy is represented by the "Jacob's Ladder" of John Perdew.

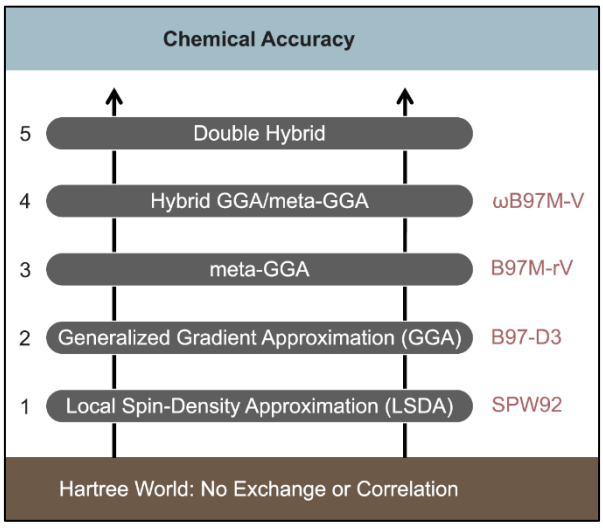


Figure 28: Perdew's metaphorical Jacob's Ladder [44].

The ladder is composed of five rungs corresponding to increasingly sophisticated models for the unknown exchange-correlation functional and has its foundations in the "Hartree World"

where the exchange-correlation energy is zero and the electron-electron interaction is provided solely by $J[n(\mathbf{r})]$. Moving up the ladder introduces additional ingredients into the functional form, culminating in the "Heaven" of chemical accuracy.

The simplest exchange-correlation approximation is the **Local Spin-Density Approximation (LSDA)** with functionals depending only on the electron density. The electron density can be locally considered as a homogeneous electron gas; this is a serious approximation for atoms and molecules since their electron density changes drastically with distance. The LSDA exchange functional has an exact analytic form

$$E_x{}^{LDA}[n(\mathbf{r})] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\int n(\mathbf{r})^{4/3}\,d\boldsymbol{r} \quad (2.68)$$

while there is no exact analytic form for the LSDA correlation functional; it can be given by parametrization on accurate Quantum Monte Carlo data. Some popular LSDA exchange functionals are the VWN5 [45], PZ81 [46], and PW92 [47].

In order to improve upon the errors that LSDA approximation introduces, one has to account inhomogeneities in the density. The **Generalized Gradient approximation (GGA)** functionals account for this effect by depending also on the density gradient, ∇n. Popular GGA exchange functionals include B88 [48], PW91 [49], and PBE [50]. Popular GGA correlation functionals include P86 [51], LYP [52], PW91 [49], and PBE [50]. These components can be combined to define GGA exchange-correlation functionals, with probably the most popular one to be the PBE exchange-correlation functional. In addition to combining separable exchange and correlation functionals, it is possible to semi-empirically parametrize GGA exchange-correlation functionals.

Going one step up the Jacob's Ladder, one can further improve the accuracy by introducing on the functional a dependence on the Laplacian of the density, $\nabla^2 n$, or the kinetic energy density; both these ingredients capture second derivative information and are actually related. Functionals that depend on either of these, are known as meta-Generalized Gradient Approximations (meta-GGA or mGGA). Popular non-empirical meta-GGA exchange-correlation functionals include PKZB [53], TPSS [54], MS0 [55]. Semi-empirical meta-GGA include functionals such as M06-L [56].

However, there are three additional limitations to the aforementioned exchange-correlation functionals; self-interaction, long-range dynamic correlation (i.e. dispersion) and strong correlation. In order to account for self-interaction, the local exchange functional can be

replaced by the exact exchange functional (Hartree-Fock), while employing a local correlation functional that gives exactly zero correlation energy for any one-electron system (such as LYP). The idea of mixing only a global fraction of exact exchange with the exchange-correlation functional is a successful solution that defines the next rung of Jacob's Ladder; the **Hybrid GGA functionals**. Most global hybrid GGA functionals have an exact exchange mixing parameter between 20%-25%. The most popular non-empirical global hybrid GGA functional is PBE0 [57] (with 25% exact exchange), while the most popular semi-empirical global hybrid GGA is B3LYP [58] (with 20% exact exchange). B3LYP functional has the form

$$E_{xc}^{B3LYP} = E_{xc}^{LDA} + \alpha_0(E_x^{HF} - E_x^{LDA}) + \alpha_x(E_x^{GGA} - E_x^{LDA}) + \alpha_c(E_c^{GGA} - E_c^{LDA}) \quad (2.69)$$

where $\alpha_0 = 0.20$, $\alpha_x = 0.72$, and $\alpha_c = 0.81$ are the three empirical parameters.

A weakness of local and hybrid exchange-correlation functionals is their inability to properly account for dispersion interactions. On improving the description of non-covalent interactions, Grimme and co-workers developed the DFT-D method [59–62] which is a damped, atom-atom empirical potential that can be used as a correction on top of the aforementioned functionals. So far, three generations of DFT-D tails have been developed; D1 [59], D2 [60], and D3 [61]. The D3 tail correction can be used either with the original damping function, D3(0), or the Becke-Johnson's, D3(BJ) [62]. In general, dispersion corrections should be used with functionals that tend to underbind both strong and weak non-covalent interactions.

Finally, all of the aforementioned density functionals may fail at describing multi-reference systems that are strongly correlated as HF and KS-DFT are both single-determinant methods. Thus, it may be necessary to include information from multiple determinants.

2.3 Classical Approaches

Classical simulations can be used to study structural, diffusion, and adsorption properties of porous materials. Molecular Mechanics or Force-Field (FF) methods use *classical* models to predict the energy of a molecule as a function of its conformation. While in Quantum Chemistry methods one of the major challenges is to calculate the electronic energy for a given nuclear configuration to give a potential energy surface, in Force Field (FF) methods, this step is bypassed by writing the electronic energy as a parametric function of the nuclear coordinates and fitting the parameters to experimental or higher-level computational data. The "building blocks" in FF methods are atoms; electrons are not considered as individual particles. Bonding information are provided explicitly, rather than being the result of solving the electronic Schrodinger equation.

In this chapter we start by introducing briefly what is a FF, which are the main interactions and terms that come into play in a standard FF and what are the challenges of an accurate description of the interatomic forces. Force field-based approaches include Monte Carlo (MC) and Molecular Dynamics (MD) simulations; Herein, Monte Carlo in the Grand Canonical ensemble (GCMC) is presented since it is applied for predicting macroscopic properties (such as adsorption isotherms, isosteric heats, etc.) of guest hydrogen molecules in porous MOF structures.

2.3.1 Force-Field Energy

The potential energy of any system can be written as a sum of different potentials with simple interpretation (bond deformations, electrostatics, dispersion forces, etc.). For an arbitrary geometry of a molecule is expressed as a superposition of valence (or bonded) interactions (E_{val}) that depend on the specific connections (bonds) of the structure and non-bonded interactions (E_{nb}) that depend only on the distance between atoms [63];

$$E = E_{val} + E_{nb} \quad (2.70)$$

Bonded interactions consist of bond stretching, angle bending, torsional components, etc. relating atoms that are linked by covalent bonds, while non-bonded interactions include van der Waals and electrostatic components. Additional non-bonded terms can be included, such as polarization and hydrogen bonding interactions.

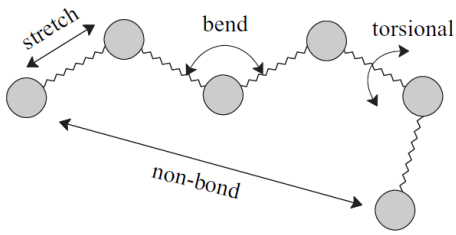


Figure 29: Illustration of the fundamental force field energy terms. The molecular mechanics model considers atoms as spheres and bonds as springs.

Therefore, the classical molecular energy (U) can be described as a Taylor expansion in bonds, bends, torsions, etc.

$$U = \sum_{bonds} u_b(r) + \sum_{bends} u_\theta(\theta) + \sum_{torsions} u_\varphi(\varphi) + \sum_{non-bonding} u_{nb}(r) + \sum_{out-of-plane\ bends} u_\chi(\chi) + \sum_{bond-bond} u_{bb}(r, r') + \sum_{bond-bend} u_{b\theta}(r, \theta) + \sum_{bend-bend} u_{\theta\theta'}(\theta, \theta') + \sum_{bond-torsion} u_{b,\varphi}(r, \varphi, r') + \sum_{bend-torsion} u_{b,\varphi}(r, \varphi, \theta') + \cdots \quad (2.71)$$

The above equation is historically referred to as a *force field*[5].

Non-bonding interactions are usually divided into van der Waals (vdW) and electrostatic interactions, with the former being dominant in physical adsorption processes and the latter significantly affecting the adsorbate uptake.

Van der Waals interactions between two atoms arise from the balance between repulsive and attractive forces with repulsion due to the overlap of the electron clouds of both atoms and the attractive component resulting from the interactions between induced dipoles. The Lennard-Jones (12-6) potential [64] is commonly used to describe vdW interactions due to its computational simplicity.

$$u_{ij}(r) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad (2.72)$$

[5] The name force field arose from the lowest order approximation using only springs with *force constants*.

where ε and σ being the effective interaction and size parameters of a particle. The r^{-12} term is the Pauli repulsion that dominates at short ranges whereas the r^{-6} term describes the dispersion attraction at long ranges.

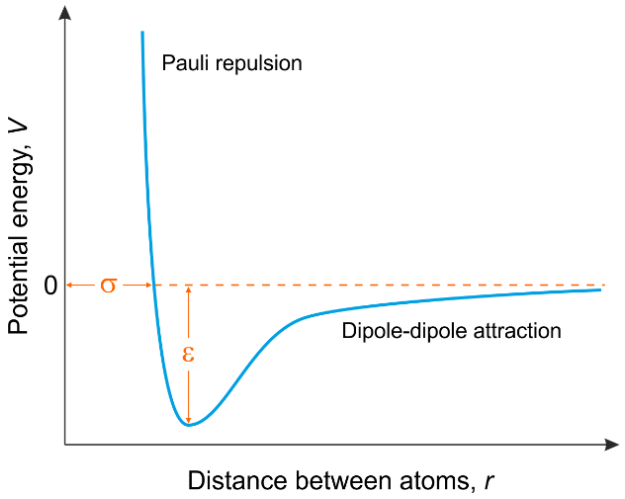


Figure 30: The Lennard-Jones potential.

Van der Waals forces act between any pair of atoms belonging to different molecules, but they also intervene between sufficiently separated atoms that belong to the same molecule. In principle, a set of parameters (σ_{ij} and ε_{ij}) should be defined for each pair of atoms, but most FFs specify the σ and ε per atom-type and use a set of mixing rules (e.g. Lorentz-Berthelot [65]) to calculate the interaction between two particles. For example, the Lorentz-Berthelot mixing rules [65] for UFF [66] and DREIDING [63] give $\sigma_{ij} = \frac{\sigma_{ii}+\sigma_{jj}}{2}$ (2.73) and $\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}$ (2.74).

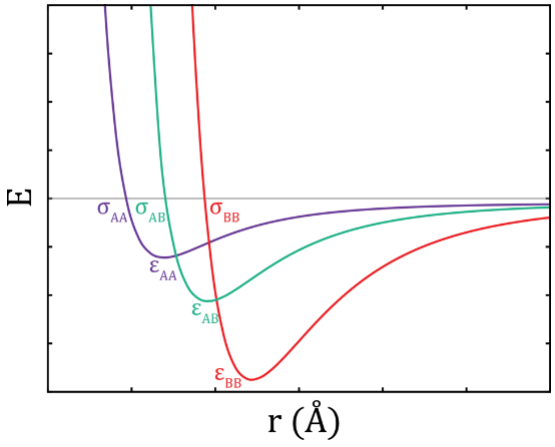


Figure 31: Lorentz-Berthelot combination rule for the Lennard-Jones potential.

Although the molecular electron density can be obtained with high accuracy from high-level QM calculations, the problem of reducing such density to a manageable description to be used in a molecular mechanics simulation is not trivial. Most partition schemes assign partial charges to each atom; the electrostatic energy is then computed according to[6]

$$u_{ij}(r) = \frac{1}{8\pi\varepsilon_0}\sum_{i\neq j}\frac{q_i q_j}{|r_i - r_j|} \quad (2.75)$$

where ε_0 is the electric permittivity of vacuum and q_i the partial charge of atom i. In order to treat the non-converging behavior of the above equation, the Ewald summation [67] is used. The appropriate way to obtain reliable partial charges consists of performing an ab initio calculation and then derive them from the QM potential. However, they cannot be derived unambiguously as atomic charges are not experimental observables; many different methods have been developed to determine them (e.g. CHELPG) and they do not always produce the same distribution of point charges.

Name	Use	Energy Term
Harmonic	Bond stretch	$k(l-l_0)^2$
Harmonic	Angle bend	$k(\theta-\theta_0)^2$
Cosine	Torsion	$k[1+\cos(n\theta)]$
Leonard–Jones 6–12	van der Waals	$4k\left(\frac{A}{r}\right)^{12}-\left(\frac{B}{r}\right)^{6}$
Leonard–Jones 10–12	van der Waals	$4k\left(\frac{A}{r}\right)^{12}-\left(\frac{B}{r}\right)^{10}$
Coulomb	Electrostatic	$\frac{q_1 q_2}{4\pi\varepsilon_0 r}$
Taylor	Stretch-bend	$k(\theta-\theta_0)[(l_1-l_{1_0})(l_2-l_{2_0})]$
Morse	Bond stretch	$D_e[1-e^{-\alpha(l-l_0)}]^2$

l-bond length.
θ-bond angle.
k, *α*, *A*, *B*-constants particular to the elements in a certain hybridization state.
n-an integer.
r-nonbond distance.
q-charge.
D_e-dissociation energy.

Table 3: Common Force Field terms.

Some examples of popular generic force fields include AMBER [68], COMPASS [69], DREIDING [63], UFF [66], CHARMM [70], GROMOS [71], etc. Each FF has its particular strengths and weaknesses related to the data and procedure employed in its parametrization, so the final choice depends on the particular problem being considered.

A complete description of molecular adsorption in porous materials should account for adsorbate-adsorbent and adsorbate-adsorbate intermolecular interactions, as well as

[6] The following expression counts all interactions and divides by a factor of 2 to compensate for double counting.

intramolecular interactions for the adsorbate and the adsorbent. For adsorbate-adsorbate interactions, well-determined FFs are available for various adsorbate molecules (e.g. the Darkrim-Levesque model for H_2 [72]). The intramolecular interactions for the adsorbent (i.e. the framework) can be omitted if the adsorbent framework can reasonably be approximated as rigid. If the size of the adsorbates (i.e. guest molecules) are not comparable to the window sizes of the host framework, then lattice flexibility of the framework does not have a strong influence on the adsorption of the guest molecules and intramolecular interactions of the adsorbent are not needed. The most challenging part of defining an appropriate FF for a system of molecules adsorbed in nanoporous materials is the accurate description of the adsorbate-adsorbent interactions; in our case, the interaction between H_2 guest molecules and the MOF atoms.

2.3.2 Ab initio derived Force-Fields

While generic and experimentally-derived FFs work well for many simple systems, they often fail to describe interactions in more complex porous materials. Since first-principles quantum mechanical approaches are capable of accurately predicting intermolecular interactions, deriving FFs from QM data without experimental input is a promising solution. A first-principles FF consists of a suitable interatomic potential form with a set of parameters that can reproduce an underlying set of data from QM calculations. In general, deriving a reliable and accurate first-principles FF for a new system requires the following:

1. A suitable adsorbent model that can capture all essential characteristics of the porous framework. For MOF materials that consist of metal corners and organic linkers, cluster models are constructed in order to represent different local fragments of the framework. For example, in order to derive FFs for adsorption of H_2 in MOFs, Han et al. [73] used a benzene molecule and a metal-based cluster to stand for the organic linker and the metal node part of their framework under study.
2. A reliable QM method that can accurately characterize the adsorbate-adsorbent interactions. The accuracy of an ab initio derived FF depends first and foremost on the accuracy of the QM method used; the QM method should correctly describe dispersion energies, since they are critical in the description of molecular adsorption in porous materials. MP2 that includes electron correlation can be reliably used -if not computationally prohibited- to calculate the adsorbate-adsorbent interaction energies.

3. A sufficient number of adsorbate-adsorbent configurations to represent all important sites of adsorption. Ideally, one should choose and perform QM calculations of configurations that represent all of the important degrees of freedom on the overall PES. One or more potential energy surfaces of interactions are acquired by letting adsorbate molecules follow one or more certain paths to the adsorption sites of the adsorbent models. Typically, the intermolecular distance between adsorbate and adsorbent site is adjusted to be closer or longer than the energy minimized state.
4. A physically meaningful FF potential form that fits well to the QM data. The non-bonded intermolecular interactions consist of van der Waals and electrostatic components. Especially for the case of guest molecules with permanent dipole or quadrupole moments, the accuracy of electrostatic components is crucial. The most common method to treat electrostatic interactions between adsorbate and adsorbent molecules is by point charge models; first assigning partial point charges to atoms and then summing up the pairwise interaction among these point charges. For adsorbate molecules, well-determined models are available where the atomic charges are assigned to reproduce experimental quadrupole moments (e.g. the DL model for H_2). For adsorbent molecules, atomic charges are determined from QM calculations; a common scheme being CHELPG [74]. For the description of vdW interactions the Lennard-Jones (12-6) potential [64] is commonly used due to its computational simplicity.
5. A reliable parametrization algorithm. Once the form of the FF potential function is known and used to calculate the corresponding PES, one has to determine the parameters that best describe the acquired QM data. This is generally achieved by optimizing an objecting function that measures the difference between the FF and QM energies.

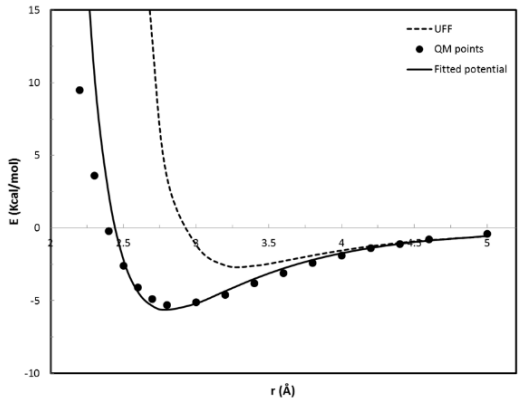


Figure 32: Example of the FF fitting from first principles. Beginning with the UFF potential form, parameters are varied to produce the QM data, resulting in a good agreement (solid line).

2.3.3 Monte Carlo simulations

To study thermodynamic properties at the molecular level, one needs to collect information about the positions of the atoms averaged over a long time. Monte Carlo (MC) method [75] is a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results and is useful for simulating systems with many coupled degrees of freedom. Each microscopic state of the system is a point in the 6*N* phase space. The measurements of macroscopic variables such as temperature *T*, volume *V* and pressure *P* involve taking averages over the phase-space curve of the system.

MC method focuses on static properties; the system does not evolve over time. Intrinsic principle to the MC method is ergodicity, which states that time averaging is equivalent to the ensemble[7] given an infinitely long simulation "time". The average behavior of the macroscopic system in equilibrium is given by the average taken over a suitable ensemble consisting of an infinite number of randomized mental copies of the system of interest. Each state[8] of the system does not depend on the previous state. In principle, each molecular state can be created from scratch; however, for computational efficiency, most MC algorithms base a new snapshot on the modification of the current (old) snapshot by performing trial moves. Specifically, at each iteration, the algorithm picks a candidate for the next sample value based on the current sample value. Then, with some probability, the candidate is either accepted (in which case the candidate value is used in the next iteration) or rejected (in which case the candidate value is discarded, and current value is reused in the next iteration). Common moves include the translation and/or rotation of a molecule, insertion, deletion, etc.

The expectation value of a variable A is given by $\langle A \rangle = \sum_i A_i P_i$ that sums over all microstates in the phase space with A the value of the variable A_i in the i_{th} microstate and P_i its probability. Instead of calculating the whole phase space, the MC algorithm scans the most probable microstates and dumps the less probable as the contribution is minimal to the expectation value (importance sampling). If $P_{eq}(o)$ and $P_{eq}(n)$ denote the probability of finding the system in states (o) and (n) respectively, and $a(o \to n)$ and $a(n \to o)$ denote the conditional probability to perform a trial move from $o \to n$ and $n \to o$, respectively, then an arbitrary initial distribution eventually relaxes to the equilibrium distribution when:

[7] a group of microstates with the same macroscopic state (e.g. N, V and T)

[8] The states of the system are also called *snapshots.*

$$P_{eq}(o)a(o \to n)P_{acc}(o \to n) = P_{eq}(n)a(n \to o)P_{acc}(n \to o) \quad (2.76)$$

In equilibrium, the flow from the old state o to any other state n is exactly equal to the reverse flow. In the Metropolis algorithm [76], $a(o \to n) = a(n \to o)$ and the acceptance rule is chosen as $P_{acc}(o \to n) = min\left(1, \frac{P_{eq}(n)}{P_{eq}(o)}\right)$ (2.77).

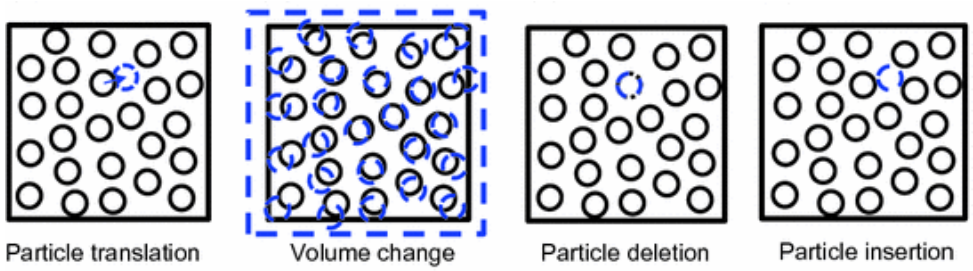


Figure 33: Classical MC moves.

2.3.4 Grand Canonical Monte Carlo of adsorption processes

In adsorption studies, we are interested in the amount of the molecules adsorbed as a function of pressure and temperature of the material with which the adsorbent is in contact. To this direction, the most common ensemble for adsorption is the grand canonical or μVT [77]. In this ensemble, the chemical potential μ, the volume V and the temperature T are held fixed, while the number of particles N is allowed to fluctuate.

In the Grand Canonical ensemble, the partition function is given by

$$\Xi(\mu, V, T) = \sum_{N=0}^{\infty} \frac{V^N e^{\beta\mu N}}{\Lambda^{3N} N!} \int e^{-\beta U(\boldsymbol{s}^N;\boldsymbol{h})}\, d^N\boldsymbol{s} \quad (2.78)$$

where $\boldsymbol{s}$ are the positions in fractional coordinates[9] and $\Lambda = \left(\frac{h^2\beta}{2\pi m}\right)^{1/2}$ (2.79) is the thermal de Broglie wavelength which represents the critical length scale at which interactions are neglected. The probability of a particular configuration is

$$P(\boldsymbol{s}^N, V) \propto \frac{V^N e^{\beta\mu N}}{\Lambda^{3N} N!} e^{-\beta U(\boldsymbol{s}^N;\boldsymbol{h})} \quad (2.80);$$

this is the distribution that will be sampled. The acceptance rules for particle move, insertion and deletion in the grand canonical ensemble are thus:

$acc(s \to s') = min\left(1, e^{-\beta[U_n(\boldsymbol{s}^N;\boldsymbol{h}) - U_o(\boldsymbol{s}^N;\boldsymbol{h})]}\right)$ (2.81) for particle move;

$acc(N \to N+1) = min\left(1, \frac{Ve^{\beta\mu}}{\Lambda^3(N+1)} e^{-\beta[U_n(\boldsymbol{s}^{(N+1)};\boldsymbol{h}) - U_o(\boldsymbol{s}^N;\boldsymbol{h})]}\right)$ (2.82) for particle insertion;

$acc(N \to N-1) = min\left(1, \frac{\Lambda^3 N}{Ve^{\beta\mu}} e^{-\beta[U_n(\boldsymbol{s}^{(N-1)};\boldsymbol{h}) - U_o(\boldsymbol{s}^N;\boldsymbol{h})]}\right)$ (2.83) for particle deletion.

In order to measure the chemical potential μ, we consider the Widom method [78], often referred to as the "particle-insertion" method. For sufficiently large N, the chemical potential is given by $\mu = -\frac{1}{\beta}\ln(Q_{N+1}/Q_N)$ (2.84), where Q is the partition function in the NVT ensemble.

Taking into account that $Q(N,V,T) = \frac{1}{\Lambda^{3N}N!}\int V^N e^{-\beta U(\boldsymbol{s}^N;\boldsymbol{h})} d\boldsymbol{s}^N$ (2.85), the chemical potential becomes $\mu = \mu_{ideal} + \mu_{excess}$ (2.86), an expression that decouples the ideal-gas

[9] For ensembles where the simulation cell is allowed to change, it is more convenient to redefine the positions in fractional coordinates using $\boldsymbol{s} = \boldsymbol{h}^{-1}\boldsymbol{r}$ (the factor h^3 is the phase-space volume). In fractional coordinates, it is easier to describe a volume change while leaving the particle positions the same.

contribution from that of an excess part. From the aforementioned expression, μ_{ideal} can be calculated analytically (e.g. via the Peng-Robinson EOS [79]), while the μ_{excess} is measured via Widom's method. We can write

$$\mu_{exc} = -\frac{1}{\beta}\ln\langle \int d\boldsymbol{s}^{(N+1)}\, e^{-\beta\left(U_n(\boldsymbol{s}^{(N+1)};\boldsymbol{h})-U_o(\boldsymbol{s}^{N};\boldsymbol{h})\right)}\rangle_N \quad (2.87)$$

where $\langle \ldots \rangle_N$ denotes a canonical ensemble averaging over the configurational space of the *N*-particle system. This ensemble average can be sampled by the conventional Metropolis scheme. In order to compute the average of the integral over the position *N+1,* we begin by an *NVT* MC simulation on the system of *N* particles. At frequent intervals during the simulation (e.g. after every MC trial move) we insert a ghost atom in our system, i.e. we randomly generate a coordinate $\boldsymbol{s}^{(N+1)}$, and compute the $e^{-\beta\left(U_n(\boldsymbol{s}^{(N+1)};\boldsymbol{h})-U_o(\boldsymbol{s}^{N};\boldsymbol{h})\right)}$ value. By averaging this value over all generated trial positions, we obtain the desired average. The significance of the Widom method lies in the simplicity of its scheme and to the insight into the meaning of the chemical potential that it provides – an insight often hard to be extracted from most statistical thermodynamics' definitions.

In order to set-up and run a GCMC simulation, an energy model should be defined for all the possible interatomic interactions. The simulation starts from an initial configuration of the system; the crystal structure of the parent (i.e. unmodified) MOF can be taken from crystallographic information files available from structural databases, whereas the positions of the atoms of the introduced FGs can be specified from ab initio of DFT optimizations of the corresponding functionalized part. Each simulation is divided in two phases; an equilibration phase (where the system evolves from the initial configuration to achieve stable structural and thermodynamic properties) and a production phase (where all the desired properties are computed).

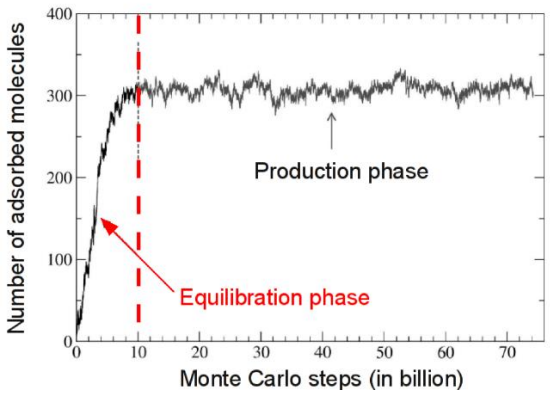


Figure 34: Equilibration and production phases in an MC simulation [80].

The duration of the simulation is measured in 'MC steps' or 'MC cycles. An MC step is actually an MC move, either accepted or rejected. Since one needs to sample longer if there are more molecules in the system, an MC cycle takes the number of particles into account; in each cycle, one MC move per particle is attempted on average. To avoid poor sampling at low densities, the number of steps per cycle is set to have a lower limit of 20. Thus, a cycle is defined as $N_{cycles} = \max(20, N)\, N_{steps}$ (2.88).

To be able to represent and calculate properties of an infinite periodic system, one has to employ periodic boundary conditions (PBC). The simulation box is surrounded by an infinite number of replicas of itself and only the *N* atoms inside the main cell are considered explicitly; when a molecule in the main box moves, its periodic image in each of the surrounding boxes moves in exactly the same way. If a molecule leaves the central box, one of its images will enter the box through the opposite face. The minimum image convention is a common form of PBC where each individual particle in the simulation interacts with the closest image of the remaining particles in the system.

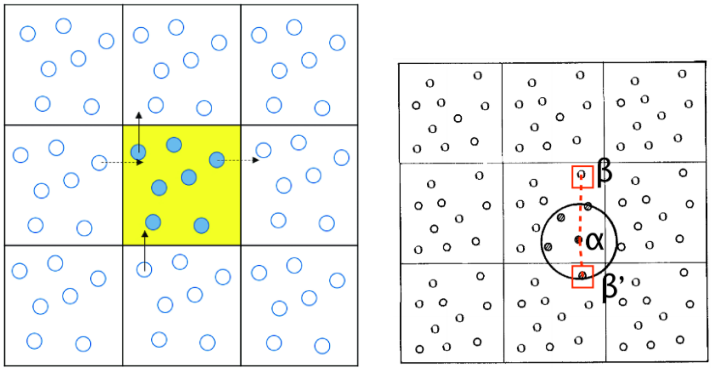


Figure 35: 2-D representation of periodic boundary conditions (left). The central cell (filled with yellow) represents the simulation box. Filled circles represent particles in the simulation box and open circles represent their periodic image in other cells. Bold and dashed lines show the movement of two particles near the boundary; as a particle leaves the simulation box, its image enters the box from the opposite end [81]. Example of the minimum image convention; the interaction will be computed between α and the image β' (right) [82].

In order for the particles not "see" their own image, the smallest perpendicular width of the chosen unit cell should be larger than twice the spherical cut-off; this requirement determines the minimum amount of crystallographic unit cells to be used in the simulation.

3.1 Functional group screening – ab initio results

A promising strategy for increasing the H_2 adsorption capabilities of MOFs is the functionalization of their organic linker components to incorporate stronger adsorption sites. To this purpose, functional groups were strategically selected to be screened for their binding strength towards hydrogen. The selection was based on electrophilic aromatic substitution[10] pathways.

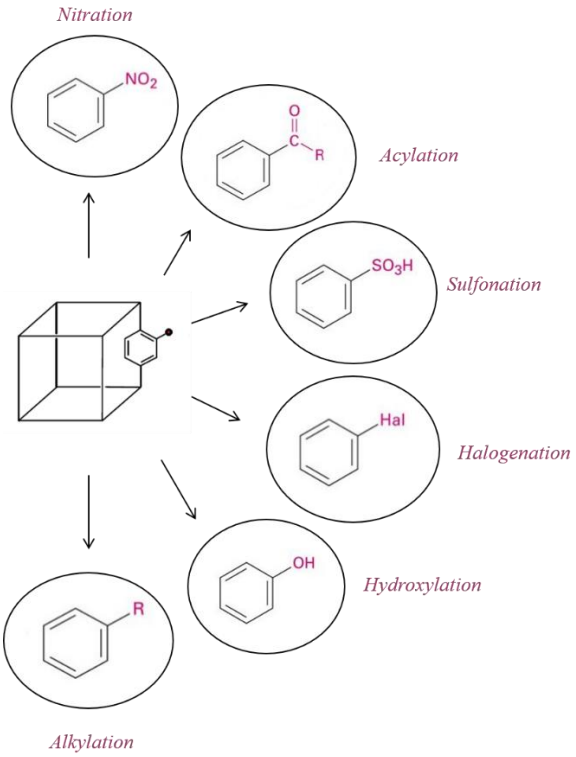


Figure 36: Some of the most important electrophilic substitutions.

Benzene has six π electrons delocalized in six *p* orbitals that overlap above and below the plane of the ring. Electrophilic substituents can exert resonance and inductive effects; resonance effects occur through the pi system and can be either electron donating (e.g. $-OCH_3$) where pi electrons are pushed towards the arene, or electron withdrawing (e.g. -CH=O) where pi

[10] The electrophilic aromatic substitution is an organic reaction in which an atom that is attached to an aromatic system (usually the hydrogen atom) is replaced by an electrophile.

electrons are drawn away from the ring; inductive effects occur through the sigma system due to electronegativity effects. They, too, can be either electron donating (e.g. -Me) where sigma electrons are pushed towards the arene, or electron withdrawing (e.g. -CF_3).

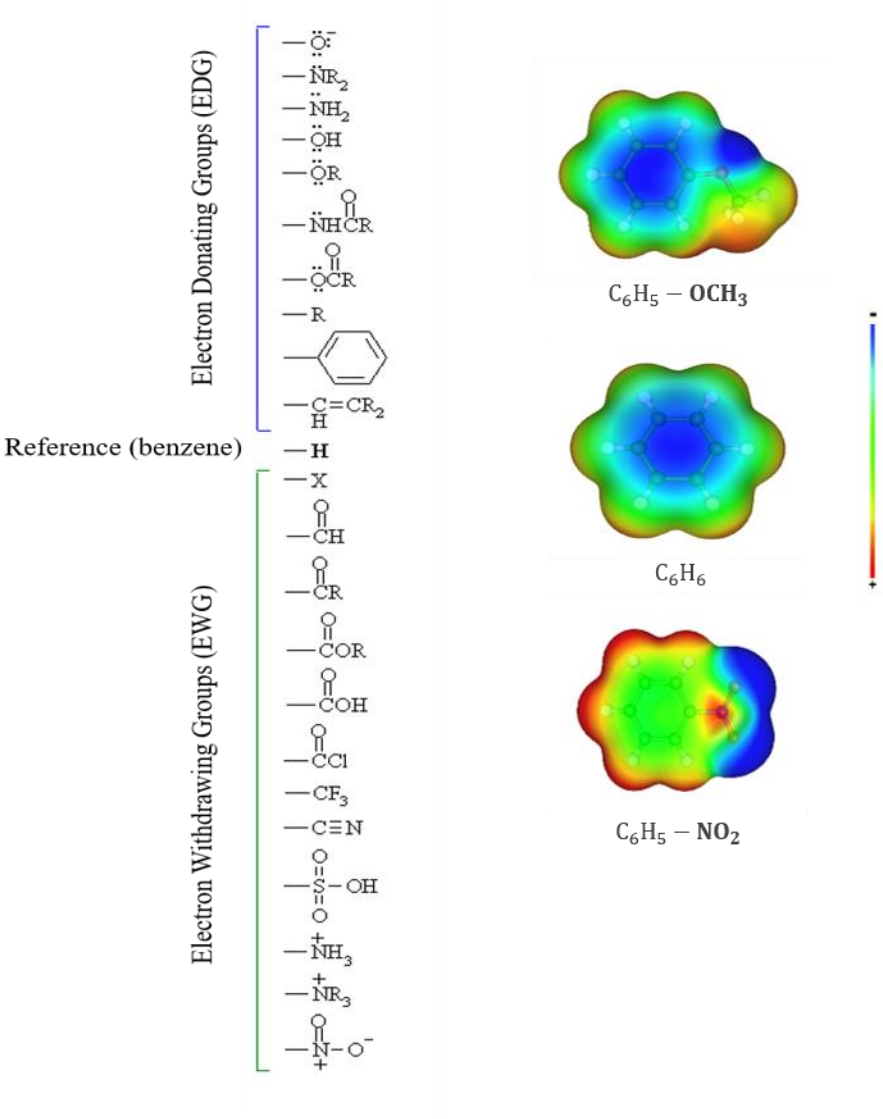


Figure 37: Examples of electron donating and electron withdrawing functional groups (left). Comparison of the electrostatic potential surface of benzene, an electron donating FG (-OCH_3) and an electron withdrawing FG (-NO_2) (right). Color spectrum with red as the highest electrostatic potential energy value regions and blue as the lowest.

In order to study the strength of the interaction of the chosen monomers towards hydrogen, in the first place we have to find their most stable conformation. One cannot easily guess the shape of all possible conformations, especially for larger and more complex molecules. Conformational search algorithms serve as an automated means for generating many different

conformers and them comparing them based on their relative energies. In this study, we performed a thorough conformational search for each monomer, using the conformational search tools of Avogadro [83] (i.e. systematic rotor, random rotor, weighted rotor and genetic algorithm search) along with extensive literature search for experimental data. The energetically most favorable configurations for each monomer are shown in Fig. 39.

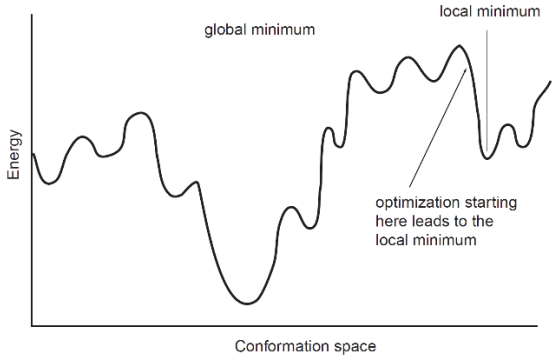


Figure 38: A 1-D representation of the energy of all possible conformers of a simple molecule. Geometry optimization methods start with an initial geometry and then change that geometry to find an energetically more stable conformation.

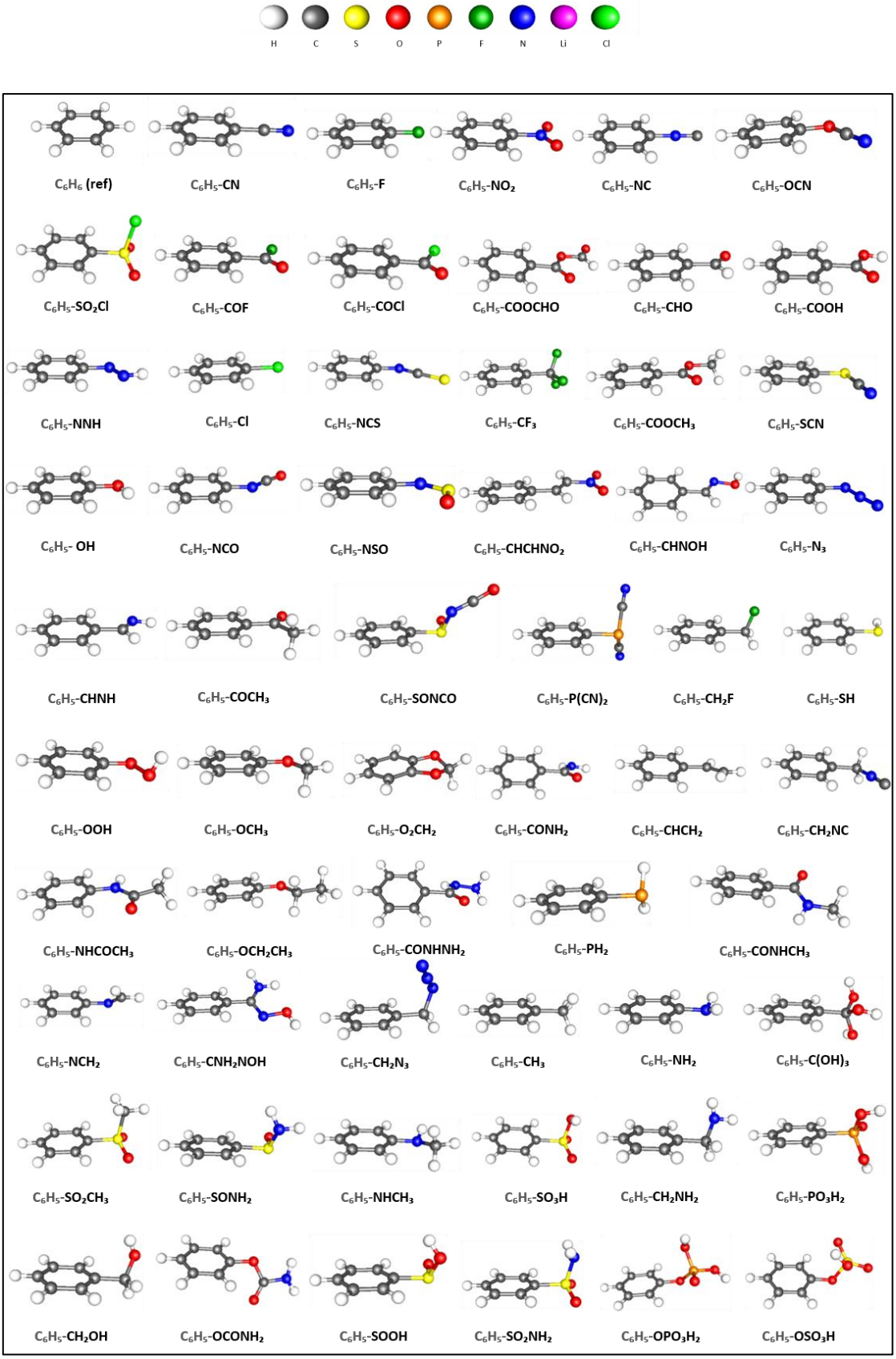


Figure 39: The optimized monomer geometries for all the functionalized benzenes that were studied thoroughly for their binding strength towards hydrogen.

Subsequently, in an attempt to identify both local minima and the energetically most favorable configuration of each system, various initial configurations for the dimer C_6H_5–X…H_2 were considered and optimized.

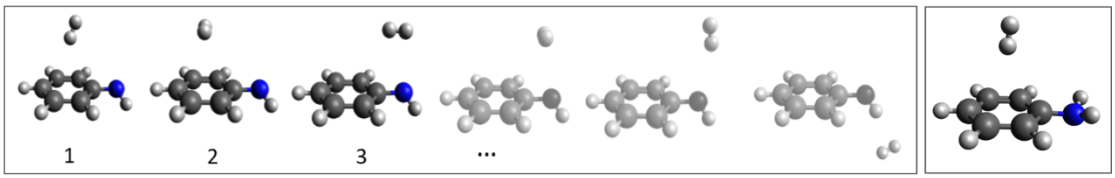


Figure 40: Example of various initial configurations of the dimer C_6H_5-NH_2…H_2 that were created and optimized. The energetically most stable configuration (global minimum) is shown on the right.

Binding sites along with their energy values are obtained from ab initio calculations performed with the ORCA 4.2 software package [84]. By virtue of consistency, all geometry optimizations were held at the MP2 level of theory using the approximate resolution of the identity (RI-MP2) [85] and the frozen-core approximation. For all atoms, calculations were performed in the def2-TZVPP basis and the appropriate auxiliary basis set for the RI approximation was assigned with the automatic auxiliary basis set algorithm implemented in ORCA. All structures were optimized without any symmetry constraints. Numerical frequency calculations were performed to all the different optimized structures arisen, to verify them as stationary points on the potential energy surface. Binding energy values were corrected for the Basis Set Superposition Error (BSSE) by the full counterpoise method proposed by Boys and Bernardi [42].

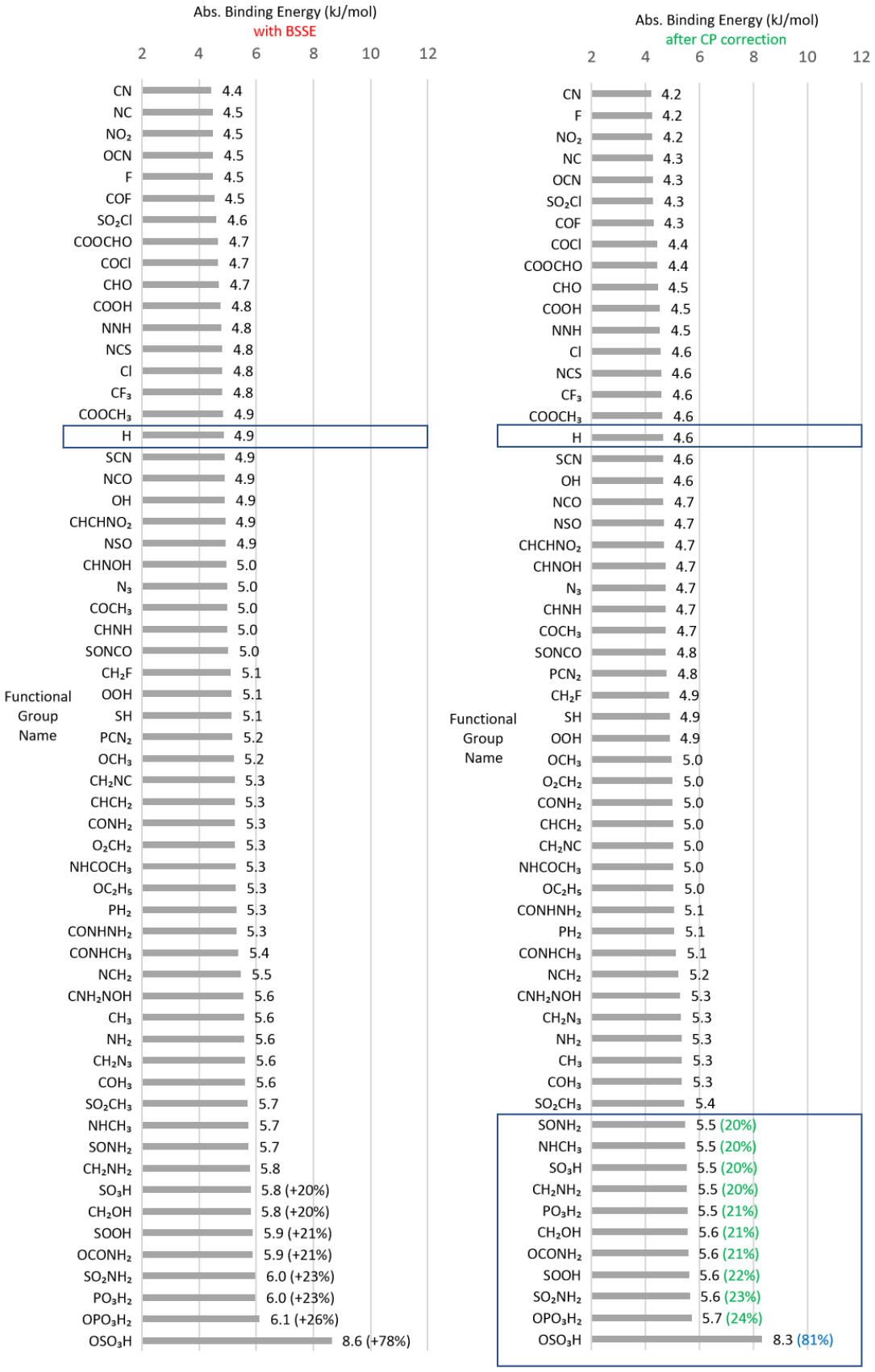


Table 4: Sorted binding energy values for all the FGs in this study. 11 FGs stand out for their binding strength towards hydrogen, with 20 up to 80% enhancement when compared with benzene.

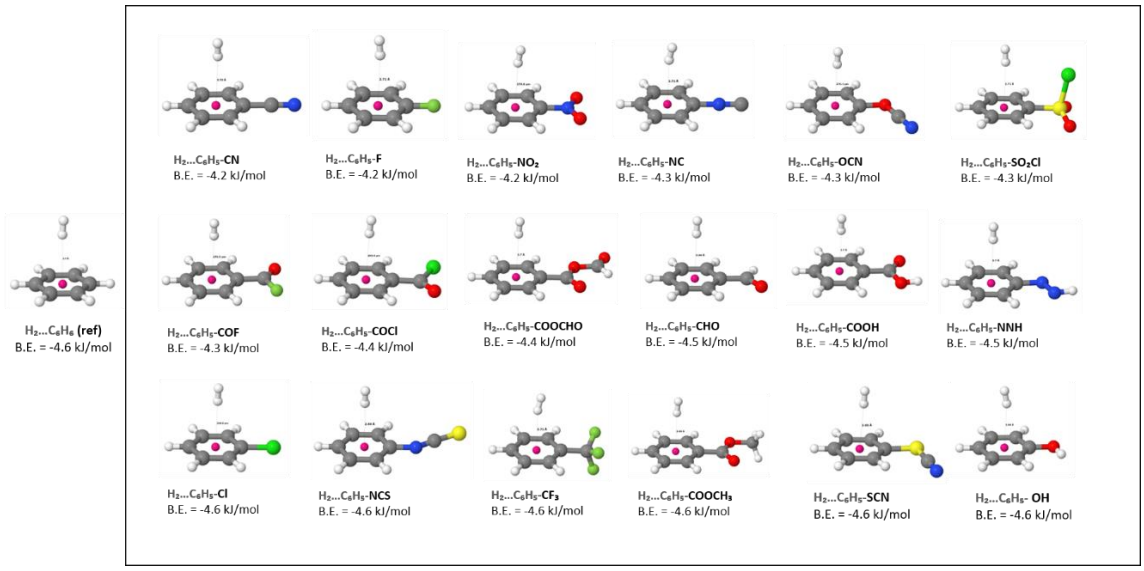


Figure 41: Global minima of the optimized geometries for the least interacting FGs. Binding energies up to -4.6 kJ/mol (the binding energy of benzene). The center of the ring is illustrated as a pink sphere to guide the eye.

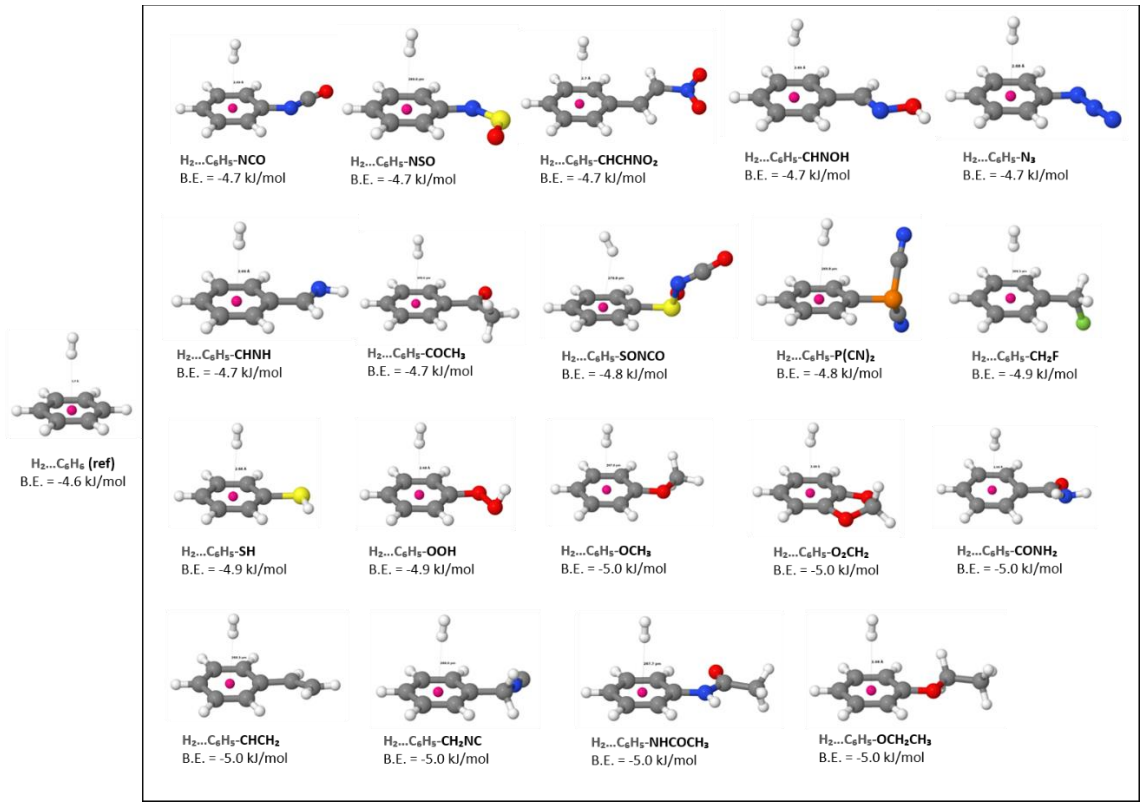


Figure 42: Global minima of the optimized geometries for FGs with moderate binding energies. Values from -4.7 to -5.0 kJ/mol.

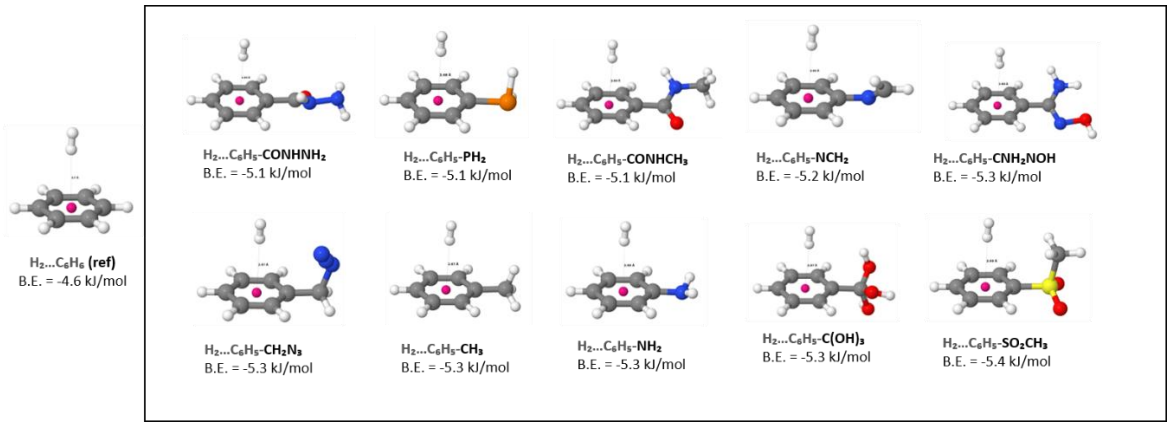


Figure 43: Global minima of optimized geometries for FGs with binding energies from -5.1 to -5.4 kJ/mol.

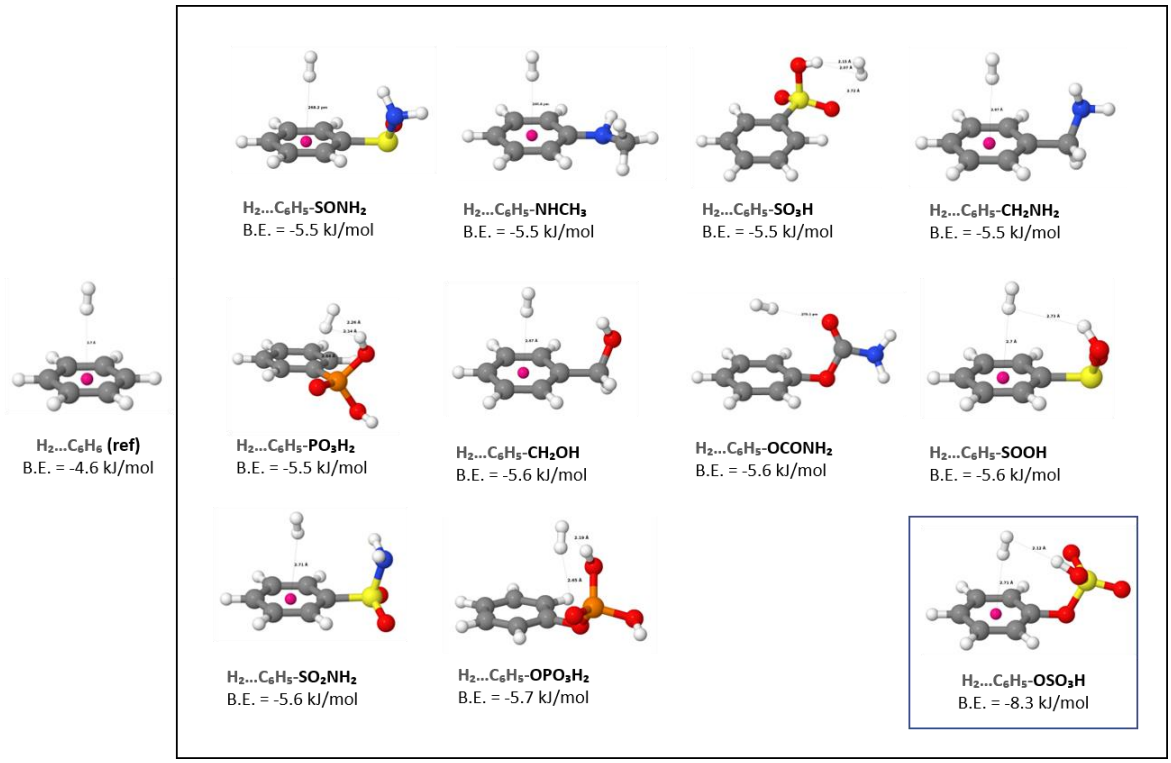


Figure 44: Global minima of optimized geometries for the best performing FGs, with energy enhancement over 20% when compared to benzene. The best performing FG is **-OSO_3H**, which shows **80%** enhancement of binding strength.

For the majority of the systems studied, the most energetically favorable position for H_2 is above the ring; a position where dihydrogen's axis is almost perpendicular to the aromatic plane and points to about the middle of the ring. As binding energy towards hydrogen increases, the inclination of the H_2 axis with respect to the normal of the aromatic plane deviates and

differences in the distance between the aromatic plane and the closer H of H_2 are observed. It is thus clear that hydrogen interacts with the pi system of the functionalized rings and the strength of the interaction correlates with the ability of the substituents to enrich the aromatic system electronically. Few FGs, namely the C_6H_5-SO_3H, C_6H_5-PO_3H_2, C_6H_5-OPO_3H_2 and OSO_3H, were able to outperform the delocalized aromatic electron distribution and thus exhibit an increased interaction, with binding energies >20% enhanced when compared to benzene. From these, C_6H_5-OSO_3H stands out with an enhancement of ~80%. This is a special case as it combines the interaction with the delocalized cloud and the interaction of the polarized hydrogen of the FG with the quadrupole moment of the H_2 molecule. An interesting case of a well performing FG is C_6H_5-$OCONH_2$ in which hydrogen seems to take advantage of both the delocalized pi system of the ring and the FG's influence.

In an attempt to gain insights on the nature of the interaction of hydrogen with the functionalized benzenes, electrostatic potential maps of the monomers and electron density redistribution plots of the complexes were generated.

The electrostatic potential that is created by a system of nuclei and electrons is formulated directly from Coulomb's law and is a physical observable that can be determined both experimentally and computationally [86]. When evaluated in the surface of a molecule, it shows how the molecule is "seen" by an approaching reactant and thus is a useful guide to the molecule's reactive behavior. The surface of the molecule can be a considered as contour of its electron density ($\rho(\mathbf{r})$), with $\rho(\mathbf{r}) = 0.001$ au ($\frac{\text{electrons}}{\text{bohr}^3}$) being a good choice as it lies beyond the Van der Waals radii of the atoms comprising most molecules. Colored electrostatic potential maps were generated for all the monomers C_6H_5-X, by mapping the calculated electrostatic potentials onto the 0.001 au electron density isosurface, using gOpenMol [87,88]. The color scale indicates regions of high and low electrostatic potential, depending on the electron density of these regions. The electrostatic potential in all illustrated structures ranges from +0.03 to -0.03 $\text{Hartree} \cdot e^{-1}$.

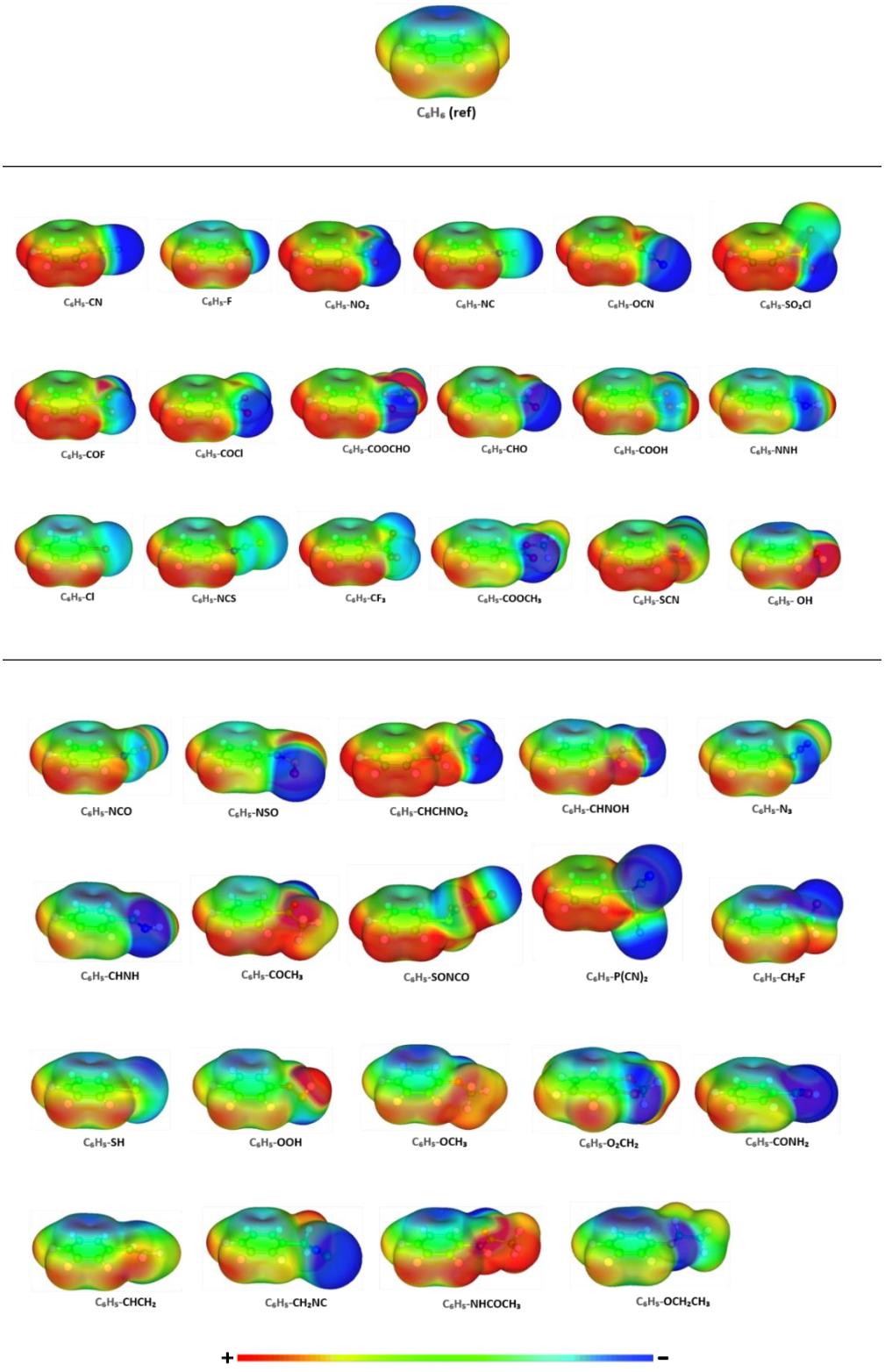


Figure 45: The electrostatic potential maps of the structures under study. Calculated at the MP2/def2-TZVPP level with ORCA 4.2 [84] and visualized with gOpenMol [87,88]. With blue and red the regions of low and high electrostatic potential respectively, ranging from -0.03 to +0.03 Hartree $\cdot\, e^{-1}$.

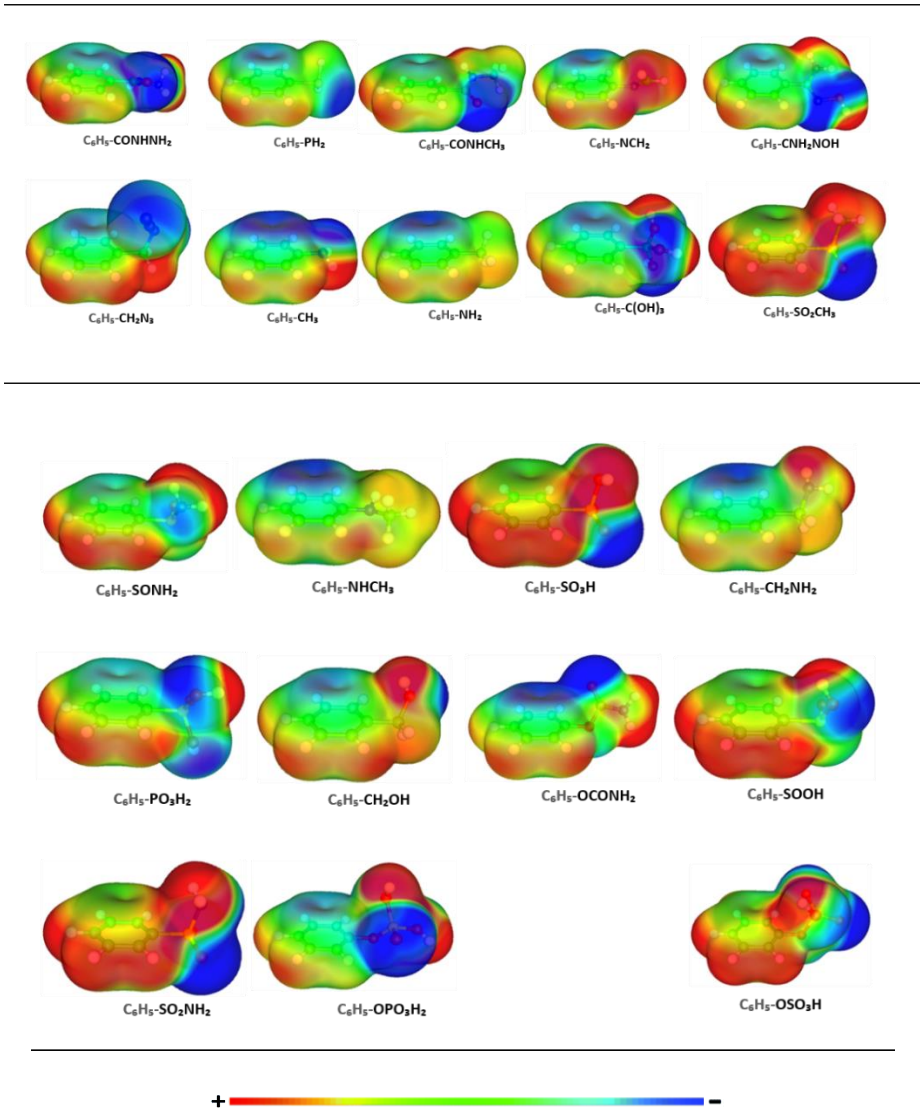


Figure 45: The electrostatic potential maps of the structures under study. Calculated at the MP2/def2-TZVPP level with ORCA 4.2 [84] and visualized with gOpenMol [87,88]. With blue and red the regions of low and high electrostatic potential respectively, ranging from -0.03 to +0.03 Hartree $\cdot\, e^{-1}$.

It can be seen that for the weakest interacting candidates (top set of maps) the FG withdraws electrons from the pi electron cloud of the ring, leaving a more neutral (green) electrostatic potential. For the strongest interacting candidates (bottom set of maps), however, some FGs seem to withdraw electron density from the ring (e.g. -OSO_3H) whereas others donate electrons reinforcing the electron density above the ring (e.g. -$OCONH_2$). Electrostatic potential maps are useful in predicting the geometry of the molecular system and the nature of the interactions between two molecules especially when electrostatic interactions are dominant; in our case that

the electrostatic interactions are not dominant to the binding of H_2, the electrostatic potential maps alone are not sufficient to explain the trend obtained from the FG screening.

In order to provide a visualization of the changes in the electron density upon interaction with hydrogen, electron density difference plots were calculated at the MP2/def2-TZVPP level of theory, as the difference of the electron density of the dimer minus the sum of the isolated monomers within the conformation of the dimer. The density of each monomer at the complex geometry was calculated in the presence of ghost basis functions of the other monomer. Densities were plotted with a contour value of 0.001 au by using gOpenMol [87,88]. With blue and green the regions that gain and lose electron density upon the formation of the complex, respectively.

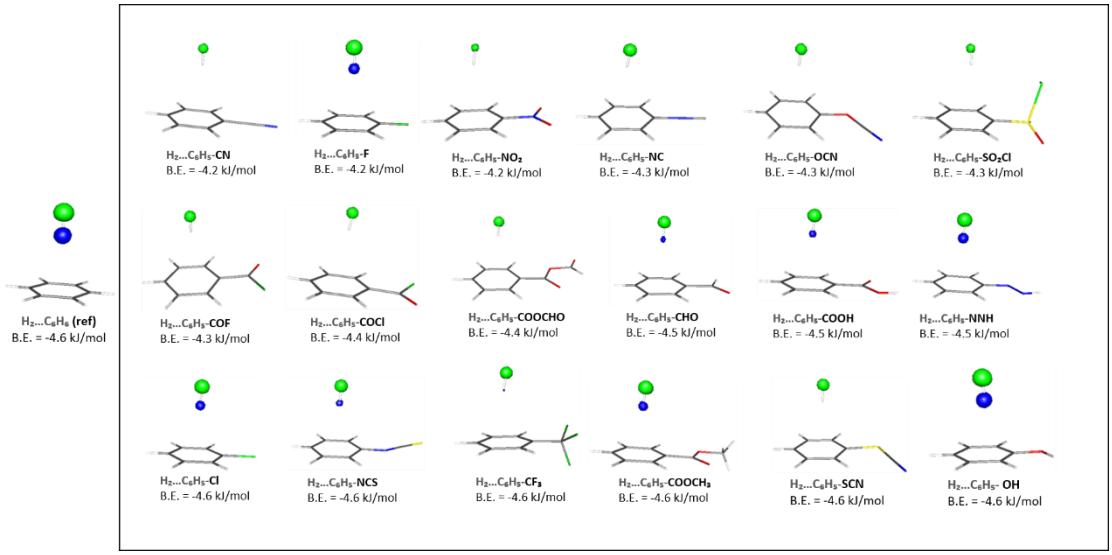


Figure 46: Electron-density redistribution plots of the optimized geometries for the least interacting FGs, with binding energies up to -4.6 kJ/mol.

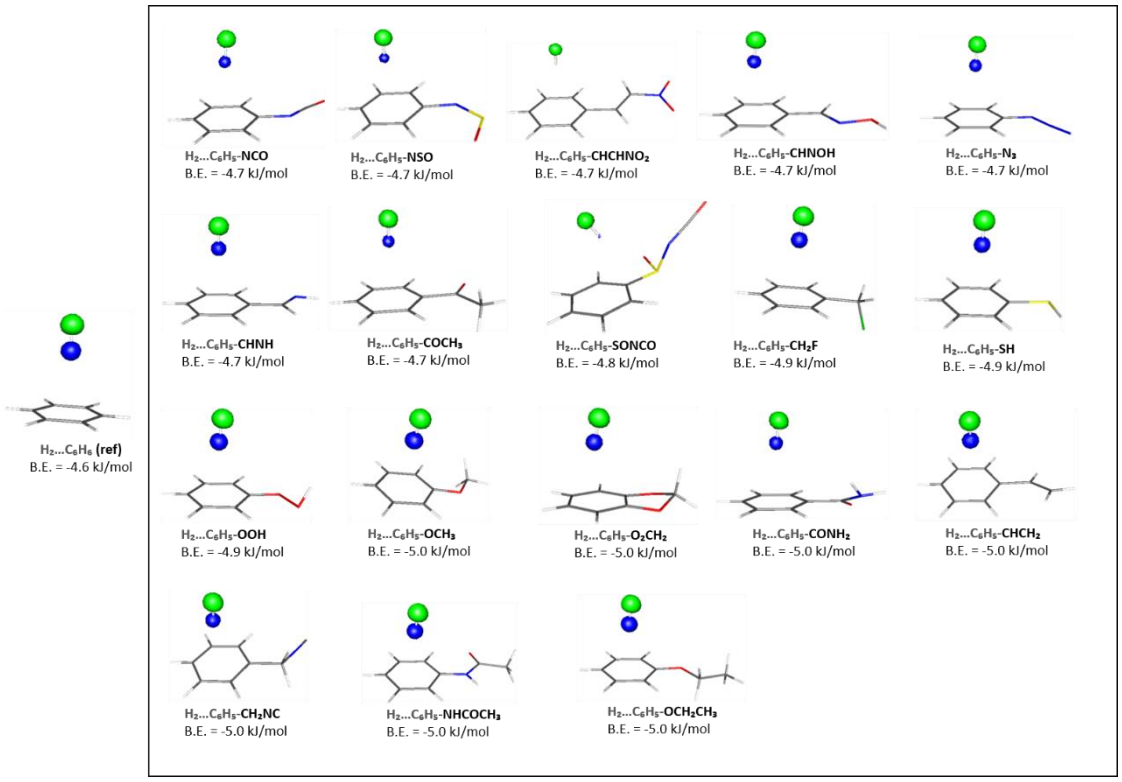


Figure 47: Electron-density redistribution plots of the optimized geometries for FGs with moderate binding energies from -4.7 to -5.0 kJ/mol.

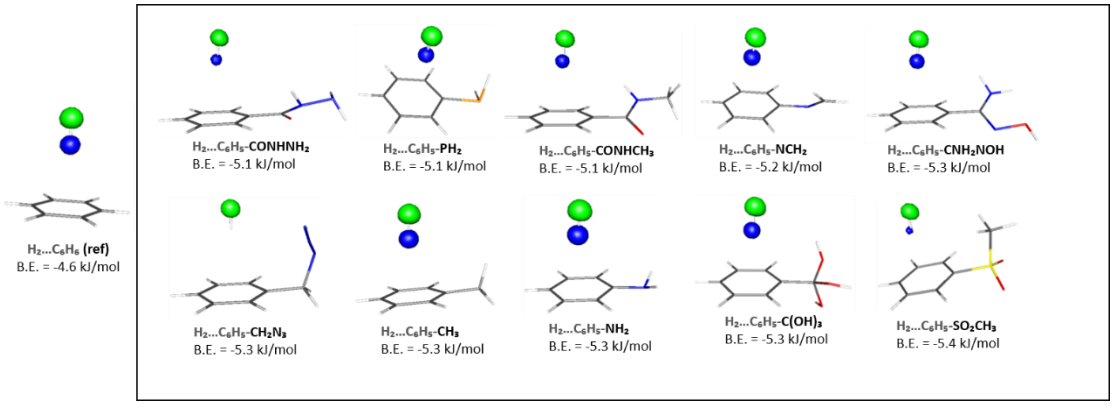


Figure 48: Electron-density redistribution plots of the optimized geometries for FGs with binding energies from -5.1 to -5.4 kJ/mol.

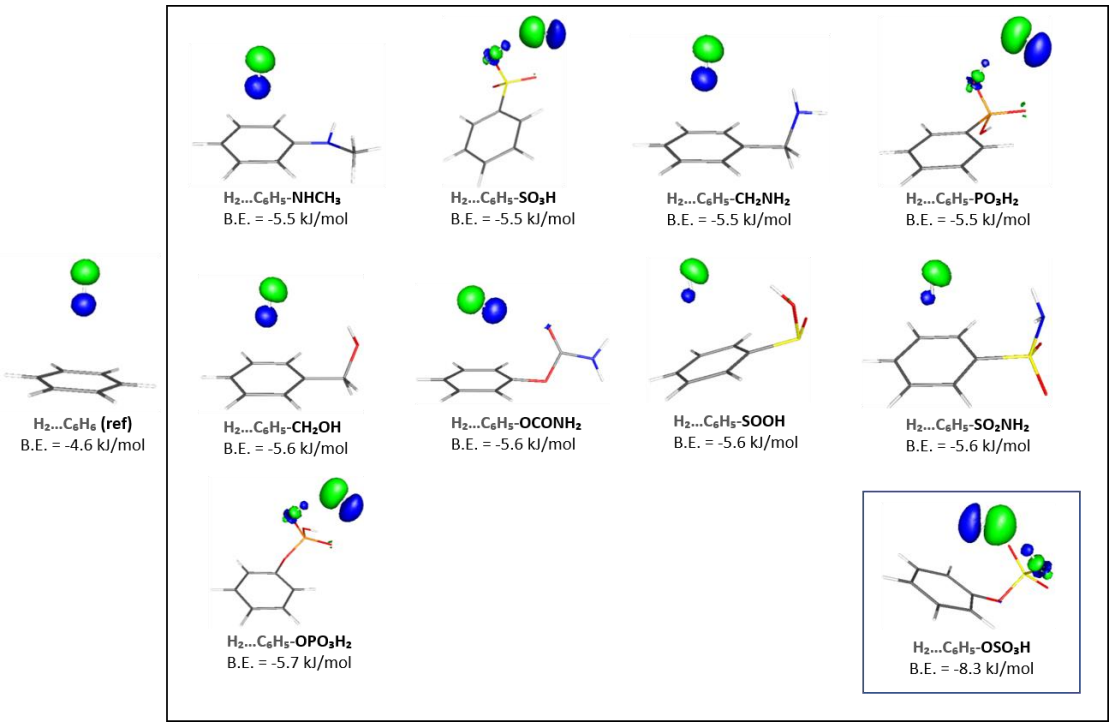


Figure 49: Electron-density redistribution plots of the optimized geometries for the best performing FGs.

Hydrogen is a "hard" molecule with a HOMO-LUMO gap of about 11.19 eV that is not likely to be significantly polarized by electric fields unless they are strong. Due to its low-lying σ_g orbital and high-lying σ_u^* orbital, it does not readily donate or receive electrons. However, its permanent quadrupole moment can interact with fields within the MOF environment, such as those created from the introduction of FGs in the organic linker.

As discussed before, for most of the systems, the most energetically favorable site for the H_2 molecule is a position above the ring, where the axis is almost perpendicular to the aromatic plane and point to about the middle of the ring. Although challenged by the large HOMO-LUMO gap, some charge transfer interaction can occur between hydrogen and the functionalized ring and is worth examined on future studies.

The largest electron density redistributions are observed for the strongest interacting functional groups, whereas for the weakest interacting FGs hydrogen's polarization is smaller or equal to the case of benzene. Thus, it is evident that the binding strength is analogous to the extent of the electron redistribution on the and $H_2…C_6H_5$-X complex; similar binding strengths of interaction correlate with similar ways of interaction.

Among the strongest interacting candidates, $H_2…C_6H_5$-SO_3H, $H_2…C_6H_5$-PO_3H_2, $H_2…C_6H_5$-OPO_3H_2 and $H_2…C_6H_5$-OSO_3H exhibit polarization of the electron cloud of hydrogen and prominent polarization of the electron density of the atoms of the FGs. The strong polarization effect comes with the fact that these FGs draw the hydrogen molecule from the pi system of the ring towards their side. For the cases of $H_2…C_6H_5$-$OCONH_2$, $H_2…C_6H_5$-SOOH, and $H_2…C_6H_5$-SO_2NH_2, hydrogen has changed its orientation in such a way that exploits both the pi system of the ring and the FG's field. Thus, although the electron energy redistribution is not that big, the binding strength is high.

3.2 Towards GCMC simulations

From the ab initio screening of the functional groups, 4 of the best performing candidates were selected for further GCMC studies, namely -$OCONH_2$, -SO_2NH_2, -OPO_3H_2 and -OSO_3H. In order to proceed to the GCMC simulations, we have to calculate the interatomic potential parameters for all different species of interacting particles and select a suitable MOF structure to introduce the chosen functionalities. The introduction of the different FGs requires the optimization of the whole linker structure in order to obtain accurate positions of the FG atoms. Last but not least, atomic charges should be assigned in all MOF atoms.

Interatomic Potential fitting

In order to describe accurately the intermolecular interactions between the guest hydrogen molecules and the MOF atoms, we used ab initio derived results and fitted the parameters of the DREIDING [63] potential to these results. At the beginning we optimized the benzene ring and applied the CHELPG population analysis [74] in order to get the point charges for the ring atoms. Then, we made use of an in-home code that takes as input:

1. The geometries for each ab-initio point.
2. The point charges for the ring as calculated using CHELPG analysis [74].
3. The ε and σ for the guest hydrogen molecule, together with the point charges. These data are taken from literature according to the model we use for GCMC simulations and are not meant to be fitted; these values are kept constant.
4. The ε and σ values for the atoms of the ring. These are the parameters that need to be fitted.

For each one of the different geometries of the ab-initio points, the code calculates the distances between each atom of the guest molecule and each atom of the ring. For every pair of each distance the code calculates the total energy (in our case this is the sum of Lennard Jones potential plus the coulomb interactions) as the sum of energies for all the pairs for the specific point. This is the classical energy calculated for the specific point. The procedure is repeated for all different points and at the end we get the curve for the classical potential and compare with the one taken from QM calculations.

In order to fit the parameters, we change the ε and σ values for the various atoms of the ring (i.e. we change the values corresponding to step (4) from the aforementioned list) and the code repeats the procedure. This is repeated until we get a fair agreement between classical and quantum points.

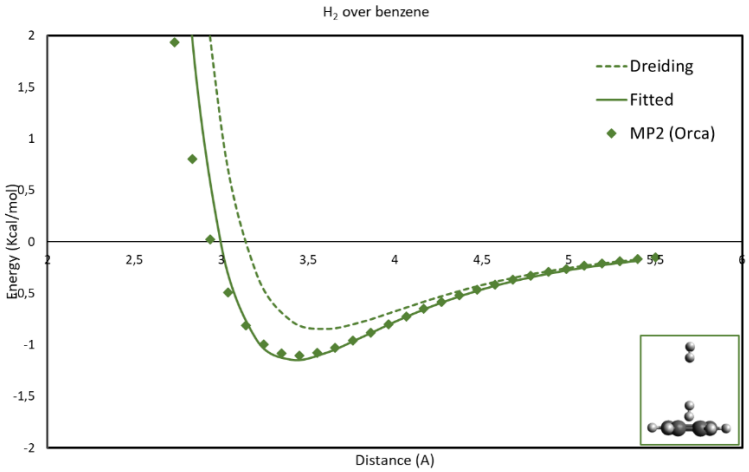


Figure 50: Fitting of the (ε, σ) parameters of the Dreiding potential on the QM data obtained from the ab initio scan of H_2 over benzene.

Subsequently, in order to accurately model the interaction of H_2 with the functional group atoms, we fitted the parameters of the DREIDING [63] potential form to the interaction energy values calculated with ab initio methods. To that purpose, functionalized linkers were represented with functionalized benzenes. The demanding/non trivial step is that we have to fit parameters for many different atom types at the same time, since the functional group contains many different atoms whose parameters ε and σ need to be fitted. Changing the values for many different atom types at the same time is not preferred since it is very hard to get a reliable set of parameters. For this reason, before starting the whole procedure we examine the local minima of the dimers H_2…C_6H_5-X as taken from the ab initio calculations of the screening for the stronger binding energies. Our aim is to find minima around the FG where for each different minimum, a different functional group atom plays the dominant role for the interaction energy for the specific site. In this way, when we fit the specific curve, we mainly deal with the parameters ε and σ for the specific atom type. By doing so for all different atoms, we get the ε and σ values for all atoms of the functional group.

In order to obtain the QM points, i.e. the interaction energies between the hydrogen molecule and the C_6H_5-X moiety, a rigid scan of 40 single point calculations was performed in the

MP2/def2-TZVPP level for each functional group selected. During these calculations, we held fixed the position of the functionalized benzene at the global (or local) minimum energy configuration of the dimer and sampled a selected distance, moving hydrogen from 5.5 to 1.5 Å towards the functional group position.

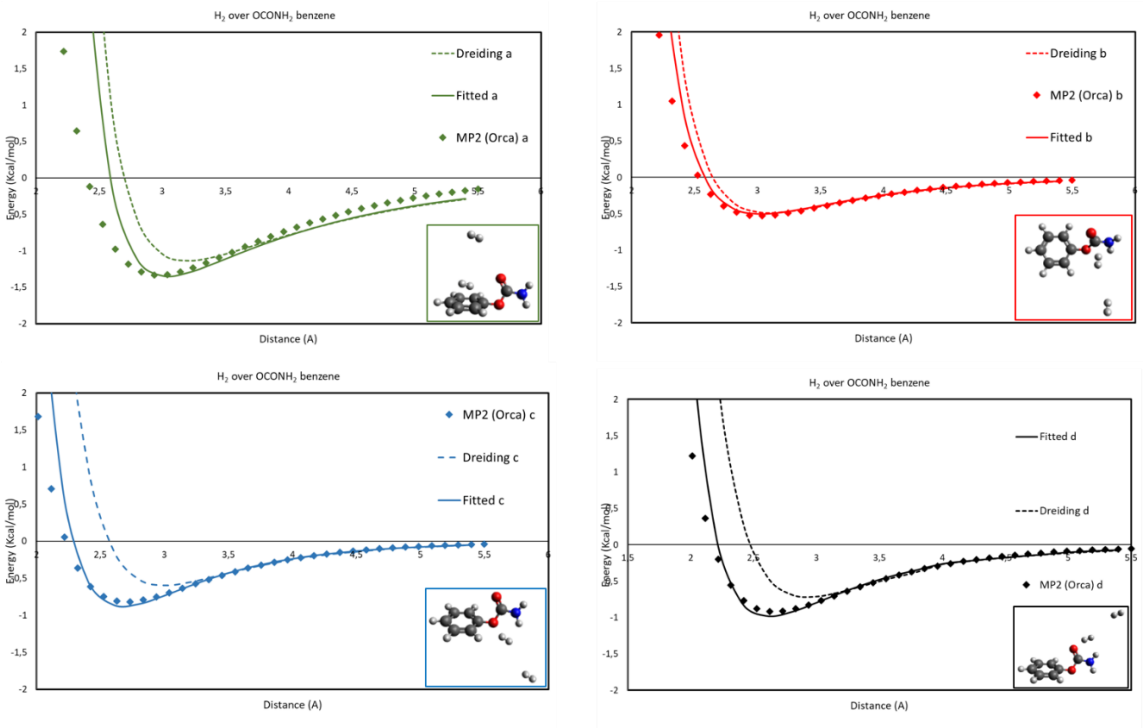
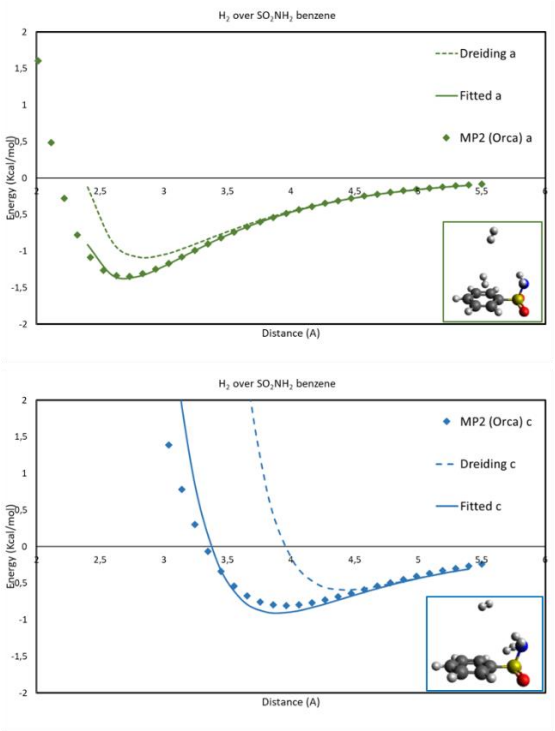


Figure 51: Fitting of the (ε, σ) parameters for the H_2…C_6H_5-$OCONH_2$ interaction. With green, the ab initio scanning around the global minimum conformation of the dimer; red, blue and black correspond to scanning directions around local minima.



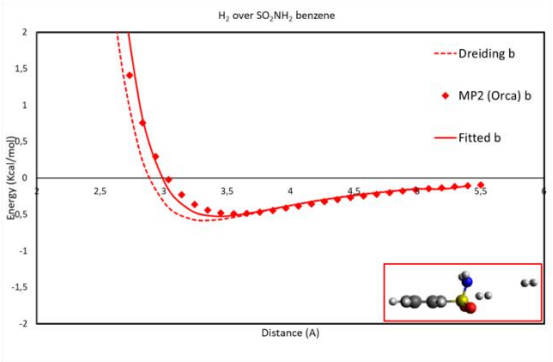
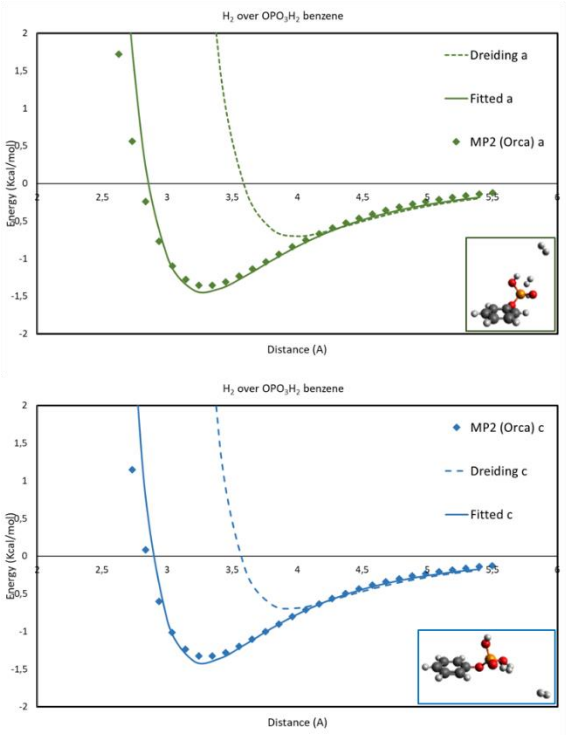


Figure 52: Fitting of the (ε, σ) parameters for the $H_2\ldots C_6H_5\text{-}SO_2NH_2$ interaction. With green, the ab initio scanning around the global minimum conformation of the dimer; red and blue correspond to scanning directions around local minima.



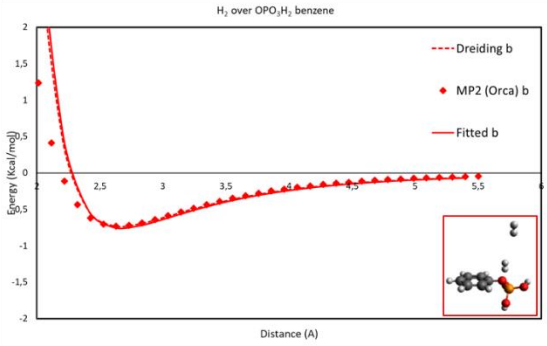


Figure 53: Fitting of the (ε, σ) parameters for the $H_2\ldots C_6H_5\text{-}OPO_3H_2$ interaction. With green, the ab initio scanning around the global minimum conformation of the dimer; red and blue correspond to scanning directions around local minima.

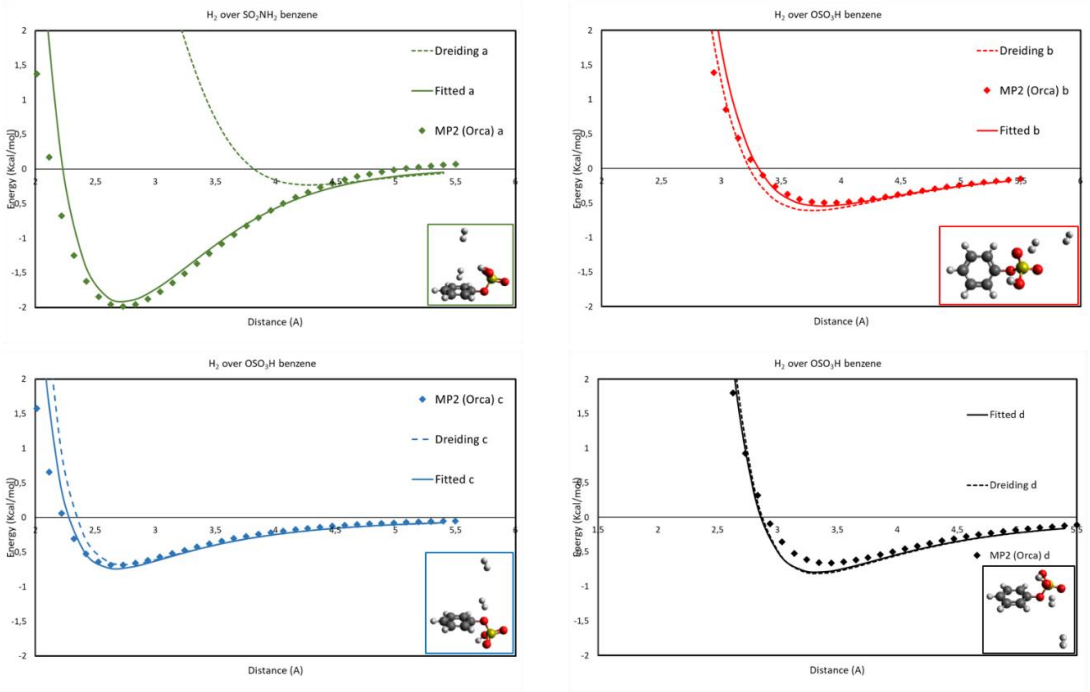


Figure 54: Fitting of the (ε, σ) parameters for the H_2…C_6H_5-OSO_3H interaction. With green, the ab initio scanning around the global minimum conformation of the dimer; red, blue and black correspond to scanning directions around local minima.

The pronounced difference between DREIDING and the ab initio curve in some of the cases, verifies the importance of the described fitting procedure and demonstrates the imminent danger of employing classical FFs without first checking their validity on our system under study.

MOF selection and preparation

For applying our linker functionalization strategy, numerous MOF structures as well as other porous materials' structures (e.g. COFs, ZIFs, etc.) are suitable. Herein, we choose IRMOF-08 and IRMOF-16 as simple cases with their organic linker part being appropriate for chemical modification. IRMOF-08 was functionalized with -$OCONH_2$, -SO_2NH_2, and -OSO_3H, while the bulkier -OPO_3H_2 was introduced to the IRMOF-16 structure along with -OSO_3H. As it will be shown from our results, in order to shed light on the interesting results that come up for the IRMOF-16 series, IRMOF-14 was also selected for functionalization with the -OPO_3H_2 and -OSO_3H FGs.

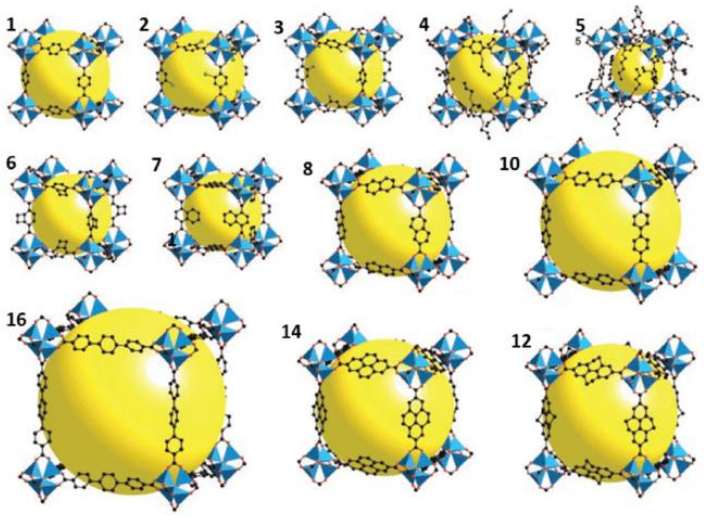


Figure 55: Structures of the IRMOF-n (n=1-8, 10, 12, 14, 16) series [89]. IR stands for IsoReticular, implying that these structures share the same metal corner and differentiate on their organic linker part.

In order to study the FG-modified and parent selected IRMOFs for their uptake, we first have to run ab initio calculations to determine the positions of the FGs inside the MOF structures as well as the partial charges of all MOF atoms. To this purpose, we used the cluster approximation. We separated the ligand from its environment and terminated the hydroxyl groups with a Li atom to represent the effect of the charge density of the missing metal cluster. The different organic linkers were doubly functionalized, as shown in the following Fig. 56. The metal corner was calculated separately, as shown in Fig. 57. Since MP2 is computationally expensive for a system of this many atoms, all geometry optimizations were performed at the B3LYP [58] level of theory, employing the RI approximation [90]. Def2-TZVP basis set together with the corresponding auxiliary basis set for the RI approximation were used. All structures were optimized with the ORCA 4.2 [84] package and without any symmetry constraints. The optimized minimum-energy structures were verified as stationary points on the PES by performing numerical frequency calculations. Partial charges for atoms in each model cluster were calculated with the CHELPG population analysis [74] and balanced afterwards to yield a neutral MOF structure.

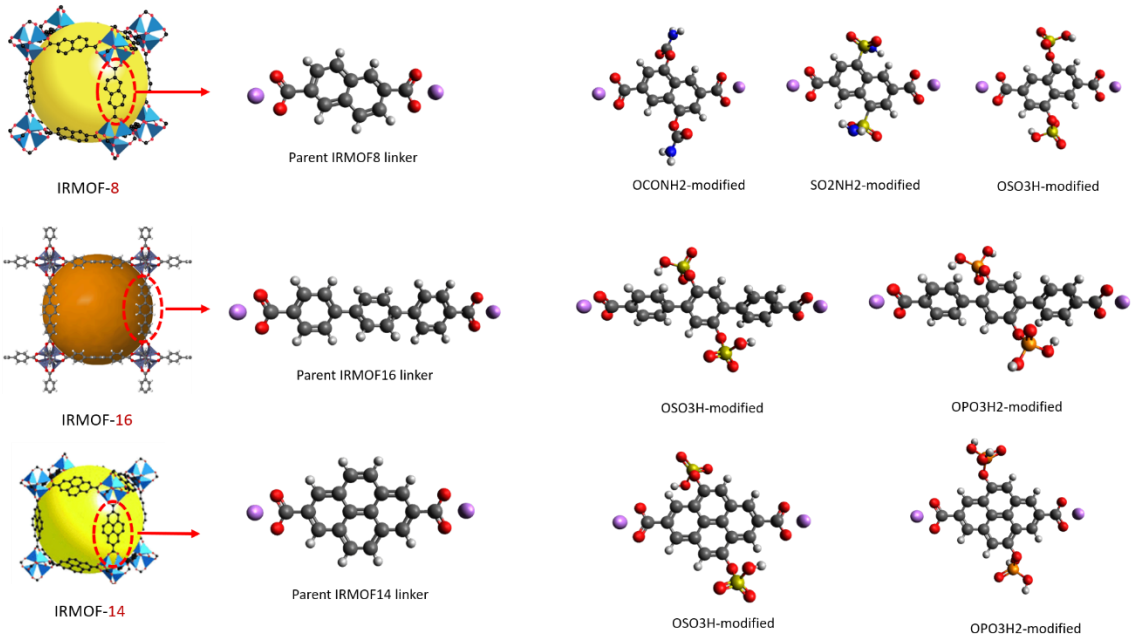


Figure 56: The IRMOFs selected and functionalized for this study along with the cluster models chosen to represent their organic linker part. All structures shown are optimized with the B3LYP/def2-TZVP method.

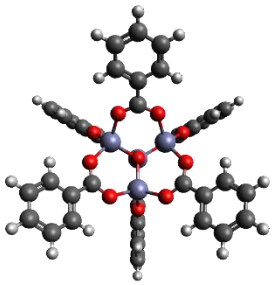


Figure 57: The metal cluster model selected to represent the metal corner of the IRMOF series.

3.3 GCMC results

In order to evaluate the trend obtained from the ab initio screening, we studied the hydrogen uptake profile of both the parent structure of IRMOF-8, IRMOF-16, IRMOF-14 and the functionalized with the best X candidates. Linkers were doubly functionalized. Positions of the IRMOF framework atoms were obtained from crystallographic information files available at the Cambridge Structural Database (CSD) [91]. Functionalized linker geometries were taken from geometry optimizations performed at the B3LYP/def2-TZVP level of theory.

H_2 interactions with all framework other than the functional group atoms were calculated employing DREIDING parameters [63] for framework atoms and empirically derived parameters for H_2. Interactions of hydrogen with the functional group atoms were calculated employing the parameters obtained from the ab initio force-field fitting. All parameters were mixed using the Lorentz-Berthelot mixing rules [65,92].

Coulomb interactions were taken into account by calculating the partial atomic charges. These calculations were performed at the B3LYP/def2-TZVP level of theory for both the metal cluster and the modified organic linker, employing the CHELPG [74] method. The Darkrim-Levesque (D-L) model [72] was used to represent for H2 where charges of +0.468 are placed on the H nuclei and a -0.936 charge at the center of mass in order to represent the H2 quadrupole. The H-H bond length is held fixed at 0.741 Å. All Coulomb interactions were handled using the Ewald summation technique [67]. Quantum effects were taken into account by the Feynman-Hibbs effective potential [93].

GCMC calculations [89] were performed with the RASPA software package [94]. Periodic boundary conditions were applied in all three dimensions. All Lennard-Jones interactions were calculated up to a cut-off distance of 12.8 Å. Simulations were performed in supercells incorporating enough repeat units such that all edge lengths were greater than 25.6 Å, i.e. twice the LJ cut-off radius. Framework was considered rigid with all framework atoms held fixed throughout the simulations. For each simulation point, 50000 cycles were performed for system equilibration followed by an additional 50000 cycles for sampling over the ensemble averages. Each cycle comprised of N steps, where N was the number of molecules in the system at the current loading. Monte Carlo moves attempted are translation, reinsertion, creation and deletion. The total gravimetric and volumetric uptake isotherms at 77 and 300K were obtained.

IRMOF-08

IRMOF-08 was functionalized with -$OCONH_2$, -SO_2NH_2, and -OSO_3H. The results from the GCMC simulations are summarized in Fig. 58 and Fig. 59 where the volumetric (g/L) and gravimetric (%wt) uptake are shown for T=300K and T=77K respectively.

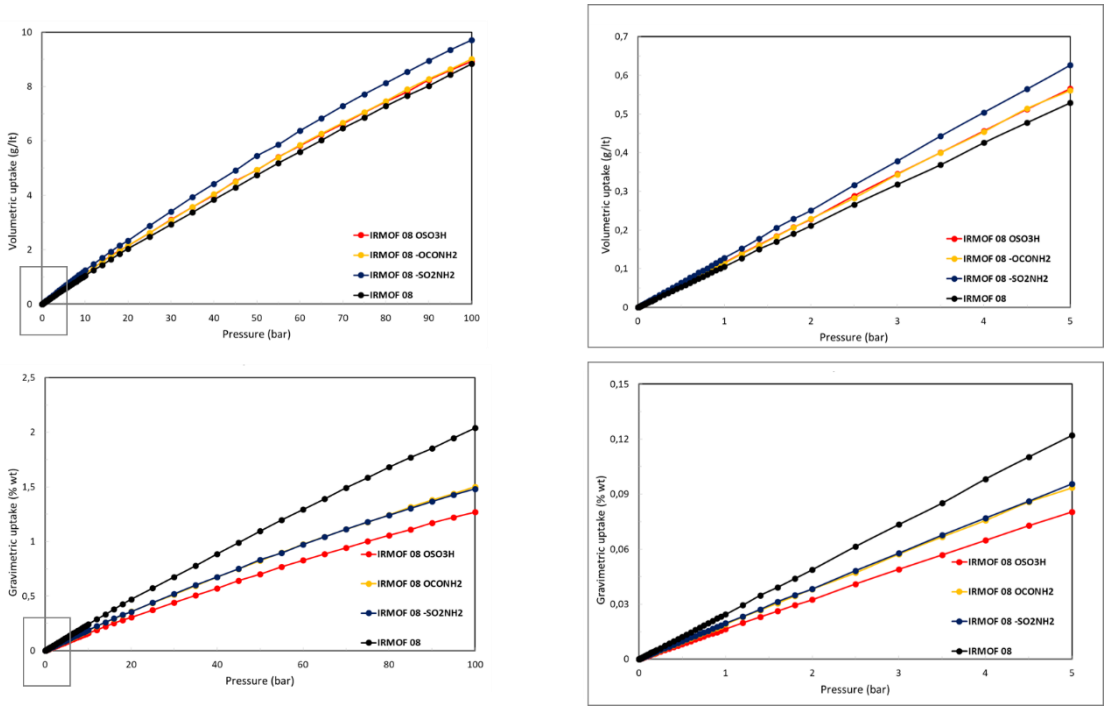


Figure 58: Volumetric (g/L) and gravimetric (%wt) hydrogen uptake at **T=300K** of IRMOF-08 and IRMOF-08-n (n: -OSO_3H, -$OCONH_2$, -$SO2NH_2$).

For T=300 K, it can be seen that volumetric uptake is enhanced from the introduction of FGs whereas for the gravimetric uptake the unmodified IRMOF-08 outperforms all FGs. This discrepancy is due to the fact that the uptake is influenced, among others, by three competing factors; the strength of the binding sites, the bulk and the weight of the FGs. In the case of the volumetric uptake, although FGs are bulky, the stronger binding sites that they introduce in the framework offer an enhancement up to 10-20% for high loadings. For the gravimetric uptake on the other hand, the strength of the introduced binding sites is not able to compensate the weight of the FGs. In general, at ambient conditions (T=300K) the binding energy enhancement from the functionalization is weak to compete with the thermal energy of H_2.

At cryogenic conditions (T=77K) the resulting isotherms are shown in Fig. 59.

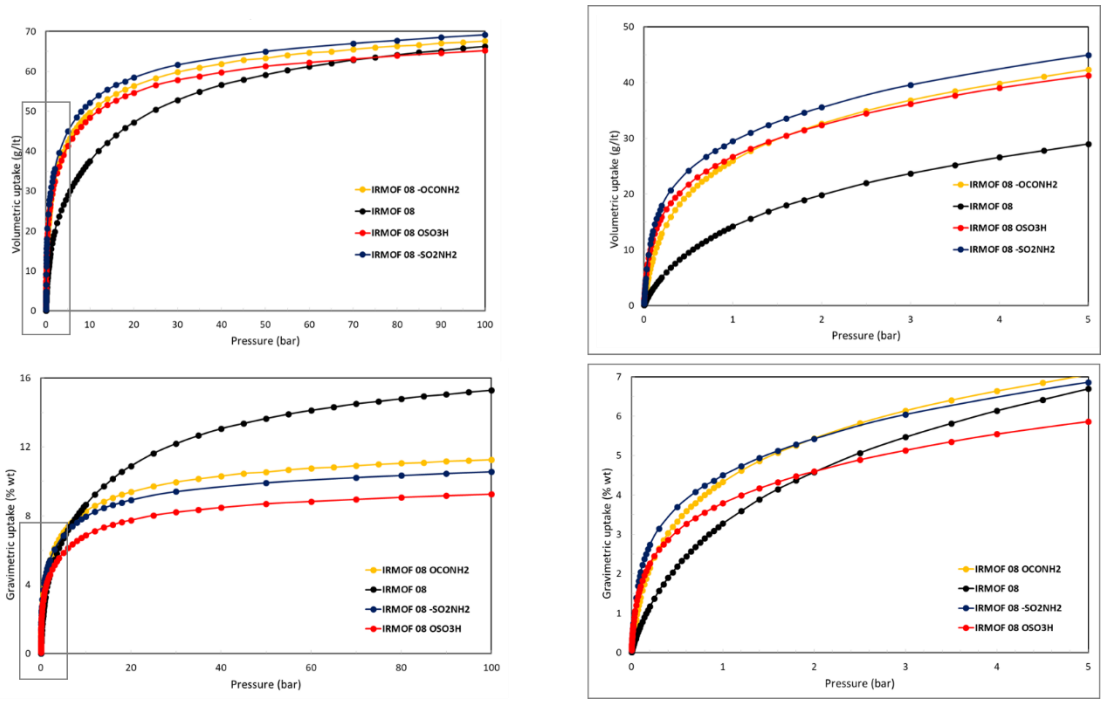


Figure 59: Volumetric (g/L) and gravimetric (%wt) hydrogen uptake at **T=77K** of IRMOF-08 and IRMOF-08-n (n: -OSO_3H, -$OCONH_2$, -$SO2NH_2$).

The volumetric uptake is enhanced in the case of the functionalized IRMOF-08 compared to the parent structure, with an enhancement up to ~60% for the case of IRMOF-08-SO_2NH_2 at P=5bar. At low loadings (i.e. P < 5bar) the first hydrogen molecules fill the structure and "feel" the effect of the stronger binding sites that FGs introduce. At higher loadings, where more and more hydrogens enter the pore, their interaction with the FGs is screened by the H_2-H_2 interactions. Thus, the percentage of enhancement drops but is still present; in the case of IRMOF-08-SO_2NH_2 is 10-20%. Overall, the strength of the introduced binding sites of all three different FGs beats the effect of their bulk for the whole pressure range, with -SO_2NH_2 being the most promising FG.

The gravimetric uptake shows an interesting behavior where a crossing of the different isotherms is observed. At high loadings (i.e. P > 10bar), parent IRMOF-08 outperforms the functionalized candidates; the strength of the introduced binding sites cannot compensate the weight of the FGs, resulting in a prominent difference. At low loadings (i.e. P < bar), on the other hand, the increased strength of the introduced binding sites competes the weight of the FGs; at very low loadings (i.e. P < 1 bar) where the effect of the increased binding strength is not yet significantly screened, all functionalized structures outperform their unmodified

counterpart. IRMOF-08-SO_2NH_2 and IRMOF-08-$OCONH_2$ outperform the parent IRMOF-08 for pressures up to 5bar, whereas the crossing for the case of IRMOF-08-OSO_3H and parent IRMOF-08 occurs at P=2bar. For P > 2 bar, parent IRMOF-08 outperforms its OSO_3H-functionalized counterpart. This effect can be attributed also to the fact that the OSO_3H-functionalized linker is heavier that the SO_2NH_2- and $OCONH_2$-functionalized.

Isosteric heats of adsorption for H_2 on parent and modified IRMOF-08 as functions of the uptake are shown in Fig. 60 and can be compared to the binding energy values obtained from the ab initio calculations. Isosteric heat of adsorption serves as an indicator of the strength of the interaction between an adsorbate and a solid adsorbent.

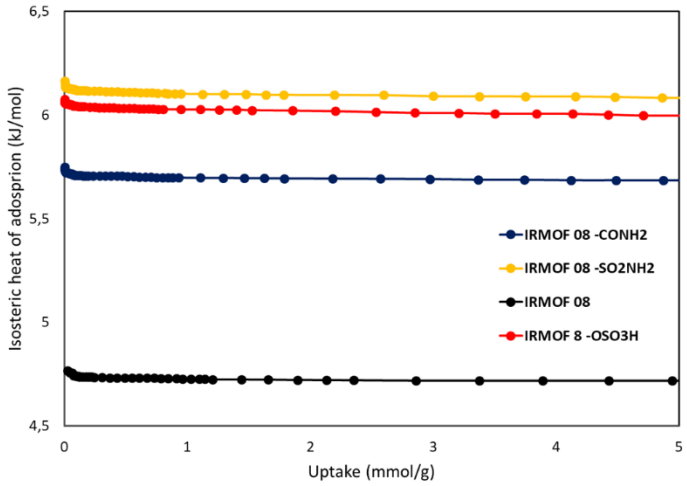


Figure 60: Isosteric heat of adsorption (kJ/mol) of IRMOF-08 and IRMOF-08-n (n: -OSO_3H, -$OCONH_2$, -$SO2NH_2$).

The qualitative trends show all structures to bind stronger at lower loadings and then leveling off. At zero loadings, i.e. when the first hydrogen molecules interact with the framework, isosteric heats of adsorption are comparable to the QM results and can be an indicator of the quality of the GCMC calculations.

Isosteric heats presented at Fig. 60 are in good agreement with the QM results; at zero loading, IRMOF-SO_2NH_2 and IRMOF-OSO_3H have similar isosteric heats at ~6.1 kJ/mol with IRMOF-$OCONH_2$ having a lower value at ~5.8 kJ/mol, whereas the unmodified IRMOF-08 structure shows a significantly lower value, below 5 kJ/mol. From the ab initio calculations, the absolute binding energies of H_2…C_6H_6, H_2…C_6H_5-$OCONH_2$, H_2…C_6H_5-SO_2NH_2 and H_2…C_6H_5-OSO_3H are 4.6, 5.6, 5.6 and 8.3 kJ/mol respectively. On an initial estimate based on the QM results, one would expect that -OSO_3H FG would give the bigger enhancement when introduced in the framework. However, this is not the case since the ab initio screening was carried out on

the simplified model of C_6H_5-X, whereas the linker of the IRMOF-08 is naphthalene dicarboxylate; although the pi system of benzene and naphthalene is comparable, differences are reasonable and expected. In addition, the ab initio calculations were employed in an isolate molecular environment and not within the pores of a robust material.

In Fig. 61-63 we present snapshots for the parent and the FG-modified IRMOF-08 taken from the GCMC simulations at T=77K and low to moderate pressures (i.e. 0.1, 1, and 5 bar). The enhancement of the hydrogen uptake can be clearly seen as one can directly compare the number of hydrogen molecules inside the cell.

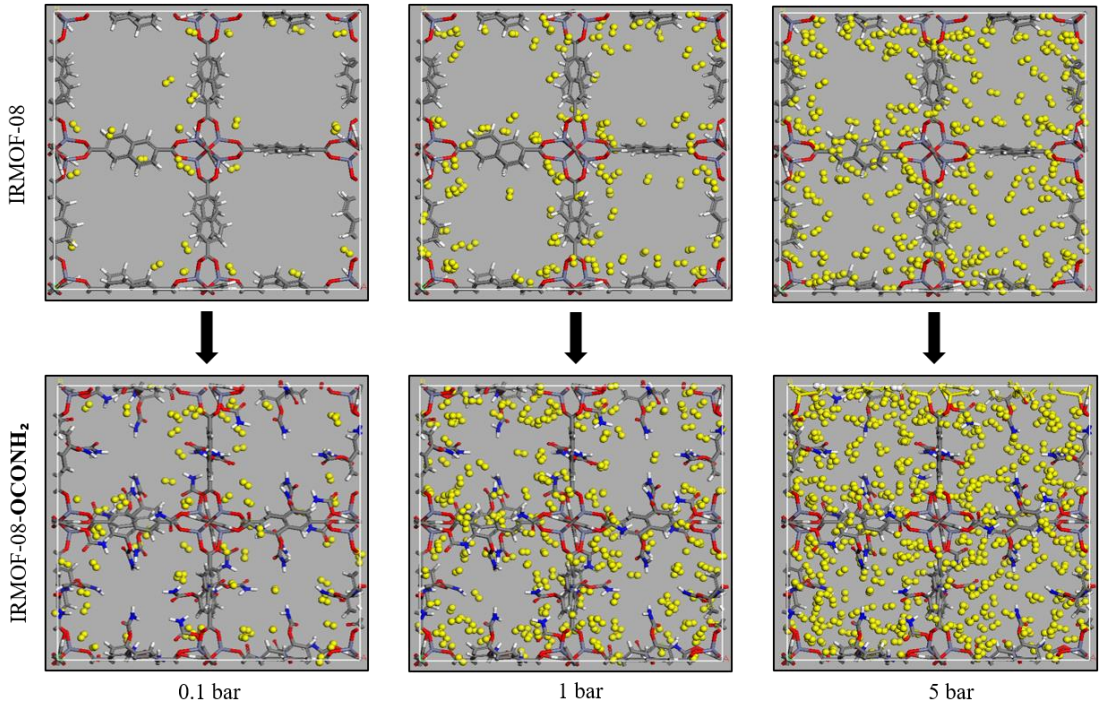


Figure 61: Snapshots of IRMOF-08 and IRMOF-08-$OCONH_2$ at T=77K.

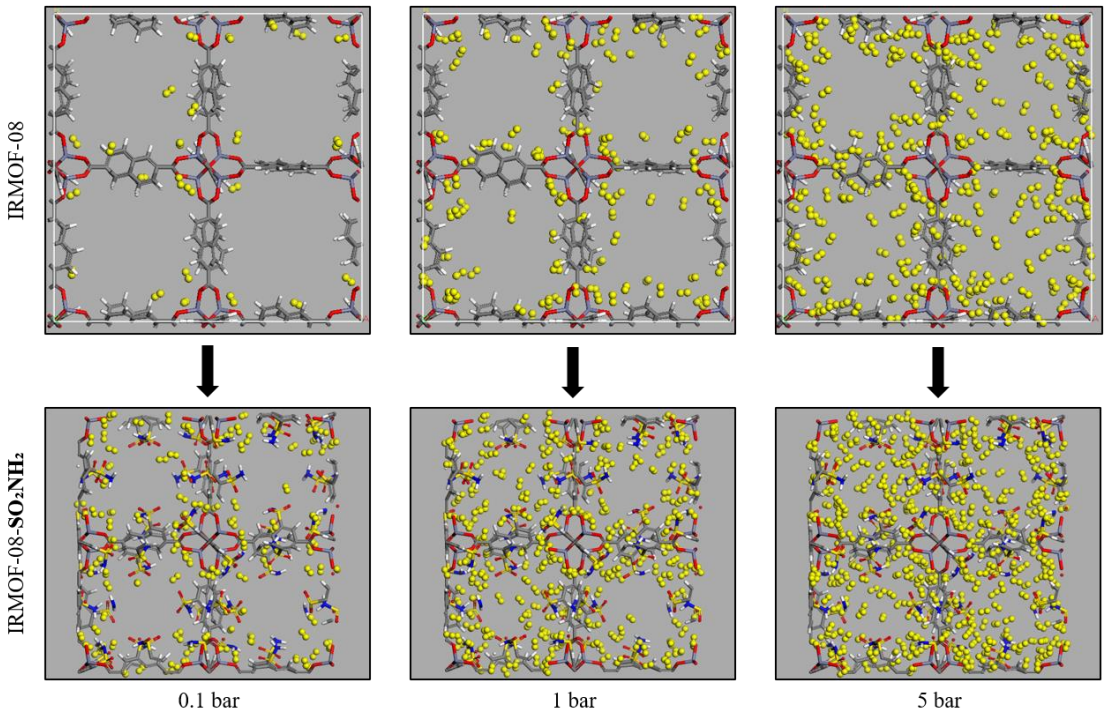


Figure 62: Snapshots of IRMOF-08 and IRMOF-08-SO_2NH_2 at T=77K.

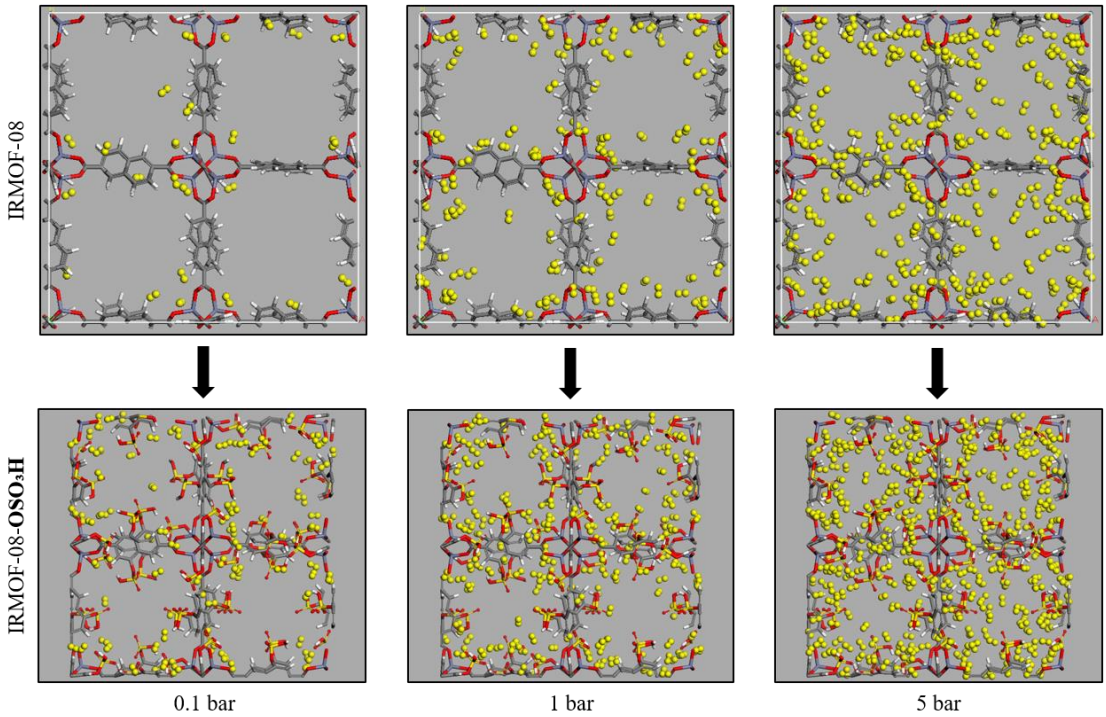


Figure 63: Snapshots of IRMOF-08 and IRMOF-08-OSO_3H at T=77K.

IRMOF-16

IRMOF-16 was selected to be functionalized with the bulkier FGs -OSO_3H and -OPO_3H_2. The results from the GCMC simulations are summarized in Fig. 64 and Fig. 65 where the volumetric (g/L) and gravimetric (%wt) uptake are shown for T=300K and T=77K respectively.

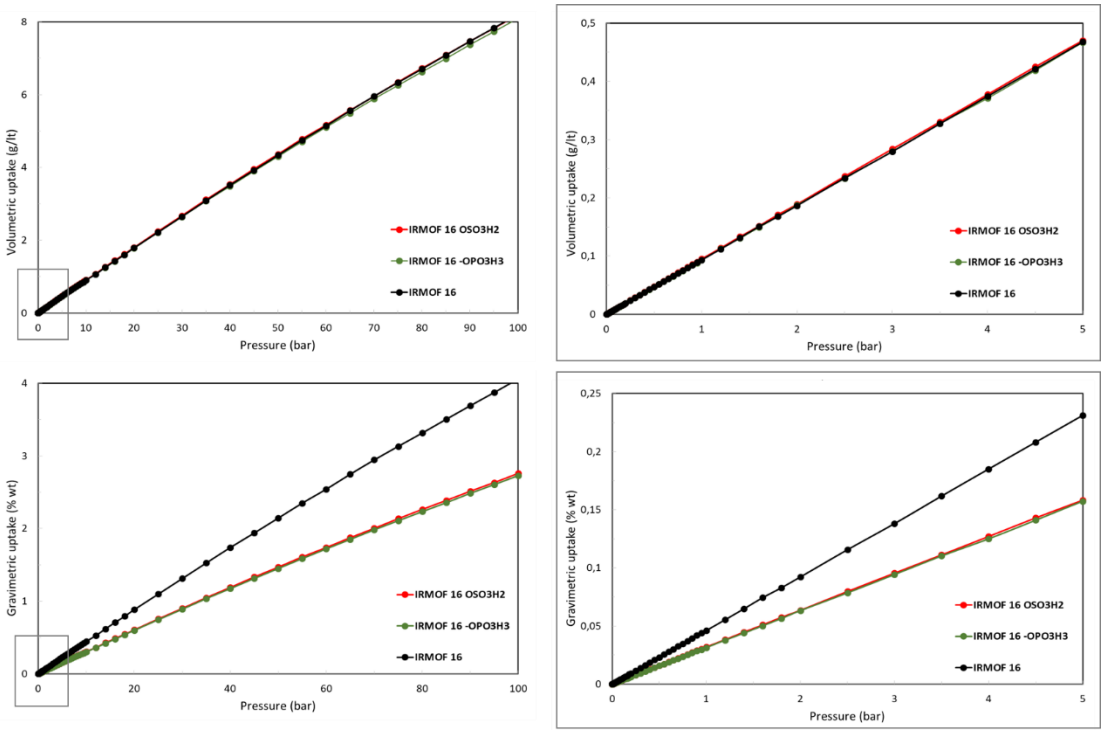


Figure 64: Volumetric (g/L) and gravimetric (%wt) hydrogen uptake at **T=300K** of IRMOF-16 and IRMOF-16-n (n: -OSO_3H, -OPO_3H_2).

For T=300K, it can be seen that the volumetric uptake shows no enhancement by the introduction of FGs in the structure; in fact, the uptake profile is the exactly the same for IRMOF-16 and IRMOF-16-n, reveling a perfect counterbalance of the gain from the increased interaction and the loss of volume.

The gravimetric uptake shows that the strength of the introduced binding sites is not able to compensate the weight of the FGs, with the unmodified structure clearly outperforming the functionalized counterparts for all pressures.

The fact that there is no difference at all in the volumetric uptake at all pressures, is in need of further research. In the first place, we survey the isosteric heats of adsorption in order to establish the validity of the performed GCMC calculations.

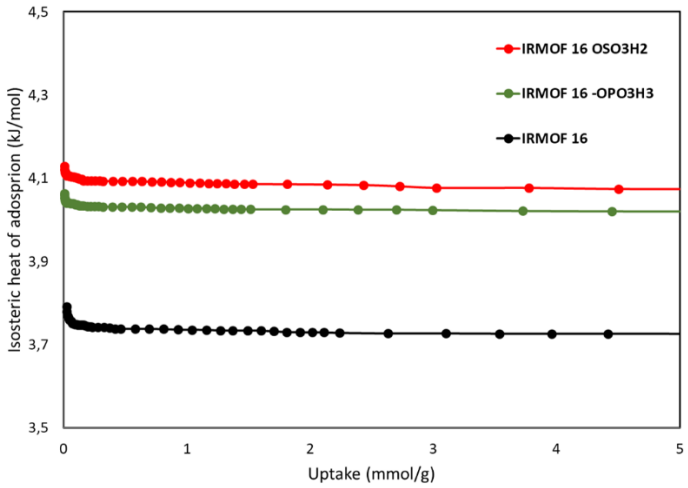


Figure 65: Isosteric heat of adsorption (kJ/mol) of IRMOF-16 and IRMOF-16-n (n: -OSO_3H, -OPO_3H_2).

Isosteric heats of Fig. 65 show that at zero loading, IRMOF-16-OSO_3H and IRMOF-16-OPO_3H_2 have similar isosteric heats around ~4.1 kJ/mol whereas the unmodified IRMOF-16 structure shows a lower value of ~3.8 kJ/mol. From the ab initio calculations, the absolute binding energies of H_2…C_6H_6, H_2…C_6H_5-OSO_3H and H_2…C_6H_5-OPO_3H_2 are 4.6, 8.3 and 5.7 kJ/mol respectively. Focusing on the case of -OSO_3H, the isosteric heat shows half binding strength compared to the ab initio calculation. Again, the ab initio screening was carried out on the simplified model of C_6H_5-X, whereas the linker of the IRMOF-16 is terphenyl dicarboxylate thus differences are reasonable and expected; however, these differences cannot be attributed to this reason.

Taking a closer look at the linker, we observe that the double functionalization of the terphenyl dicarboxylate results in a stable geometry with a slight rotation of the middle phenyl ring in order for the bulky FGs to be incorporated. This effect is shown in Fig. 66. The slight rotation of the middle phenyl ring affects the pi electron distribution of the linker, therefore affecting the binding performance of the FGs towards hydrogen. This effect leads us to the reasonable hypothesis that the functionalization of the IRMOF-16 linker results in the loss of the strong binding sites due to rotation of the linker. In order to check the validity of our hypothesis we chose to functionalize a structure of the IRMOF family with a robust linker, i.e. with no rotational degrees of freedom; IRMOF-14 served as the appropriate candidate with its pyrene dicarboxylate linker shown in Fig. 66.

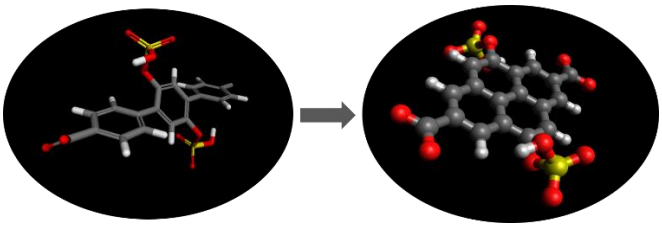


Figure 66: The IRMOF-16-OSO_3H linker (left) and the robust IRMOF-14-OSO_3H linker (right).

For T=77K, the following isotherms were obtained:

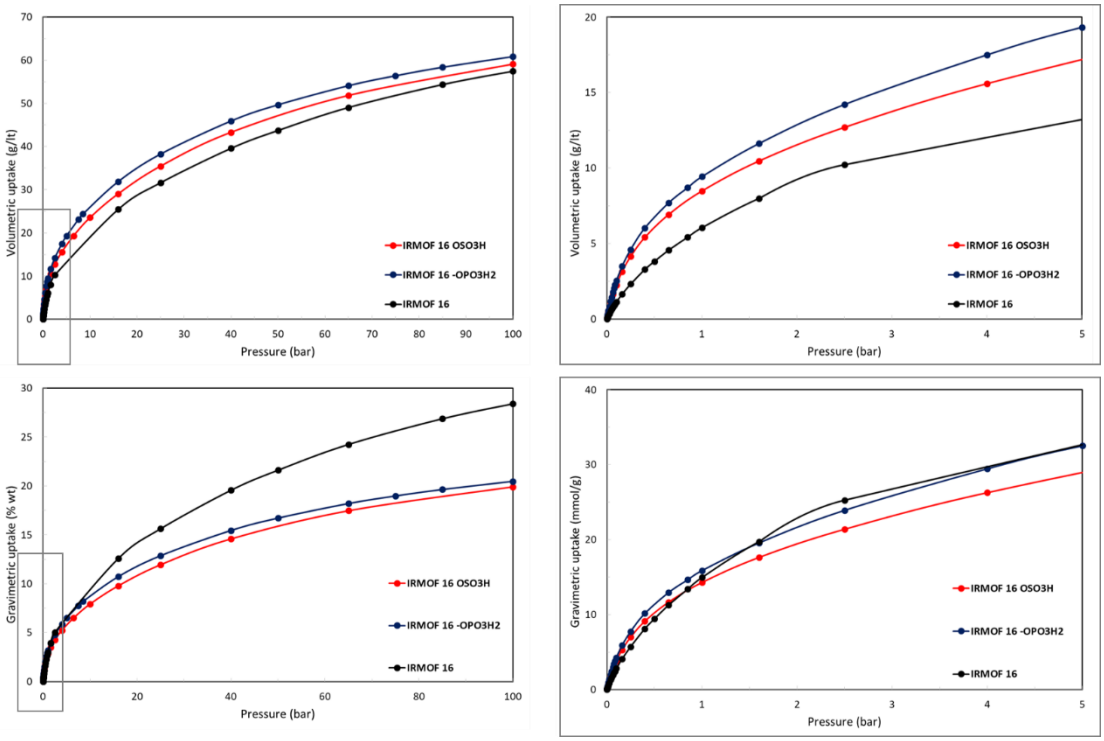


Figure 67: Volumetric (g/L) and gravimetric (%wt) hydrogen uptake at **T=77K** of IRMOF-16 and IRMOF-16-n (n: -OSO_3H, -OPO_3H_2).

The volumetric uptake is enhanced in the case of the functionalized IRMOF-16 compared to the parent structure, with an enhancement more than 50% for both FGs at P=5bar. From the two FGs, -OPO_3H_2 has a greater uptake for all the range of pressures. For the gravimetric uptake on the other hand, the situation is reversed, with the unfunctionalized stucture presenting a significantly greater uptake at high loadings. At low pressures, parent IRMOF-16 and IRMOF-16-OPO_3H_2 have comparable gravimetric uptakes, outperforming that of IRMOF-16-OSO_3H.

In Fig. 68-69 we present snapshots for the parent and the FG-modified IRMOF-16 taken from the GCMC simulations at T=77K and low to moderate pressures (i.e. 0.1, 1, and 5 bar).

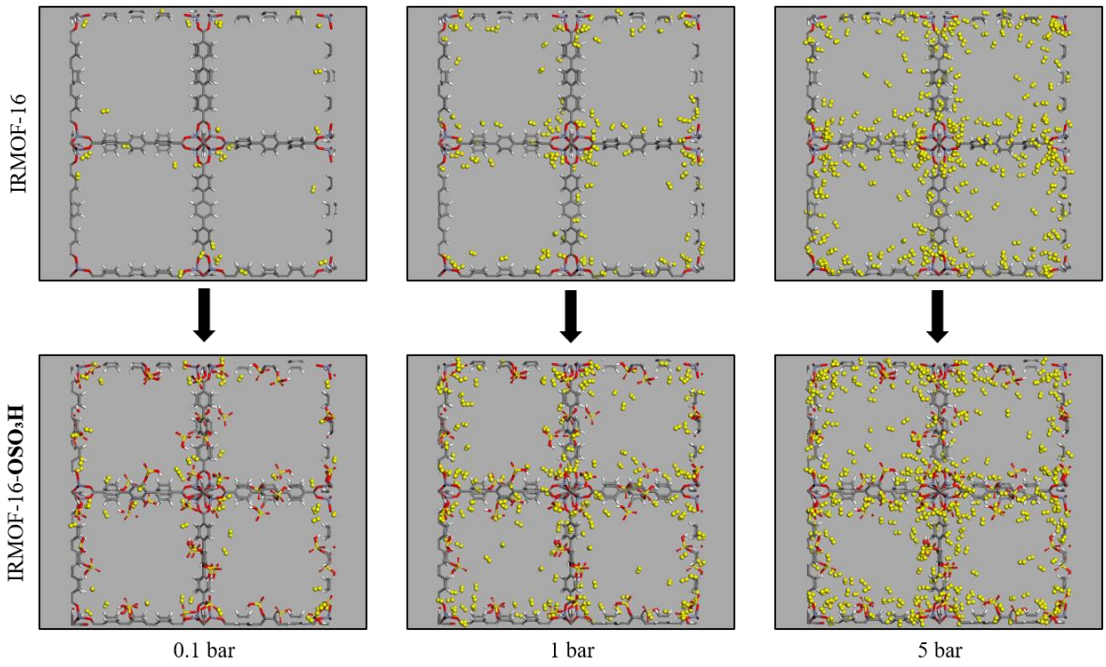


Figure 68: Snapshots of IRMOF-16 and IRMOF-16-OSO_3H at T=77K.

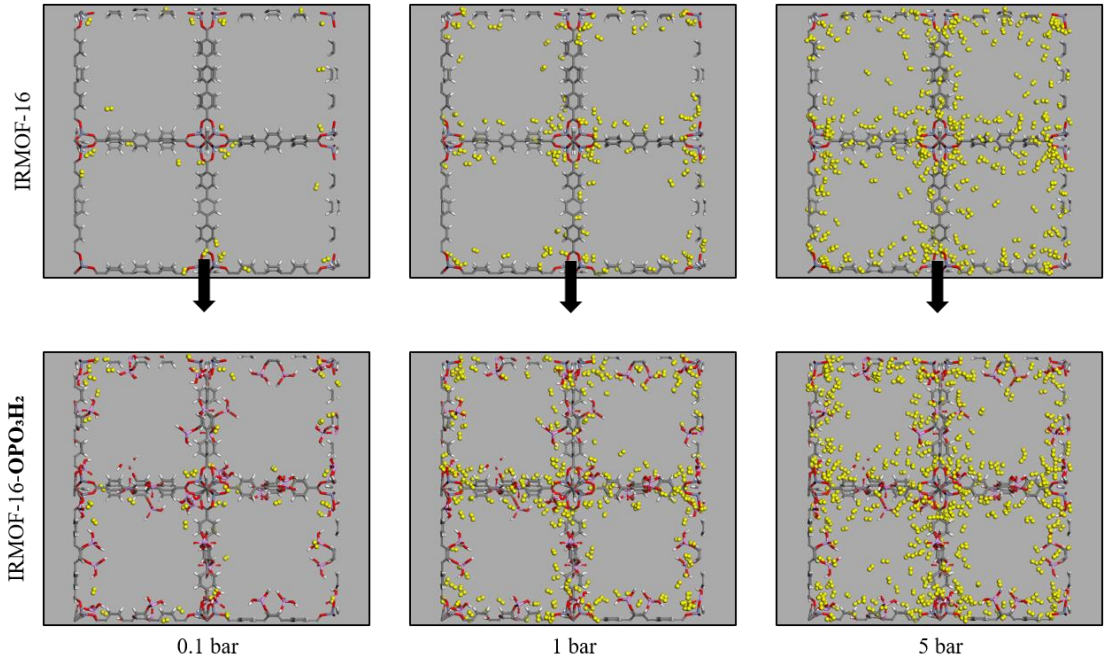


Figure 69: Snapshots of IRMOF-16 and IRMOF-16-OPO_3H_2 at T=77K.

IRMOF-14

IRMOF-14 was functionalized with the -OSO_3H and -OPO_3H_2 FGs. The results from the GCMC simulations are summarized in Fig. 70 and Fig. 71 where the volumetric (g/L) and gravimetric (%wt) uptake are shown for T=300K and T=77K respectively.

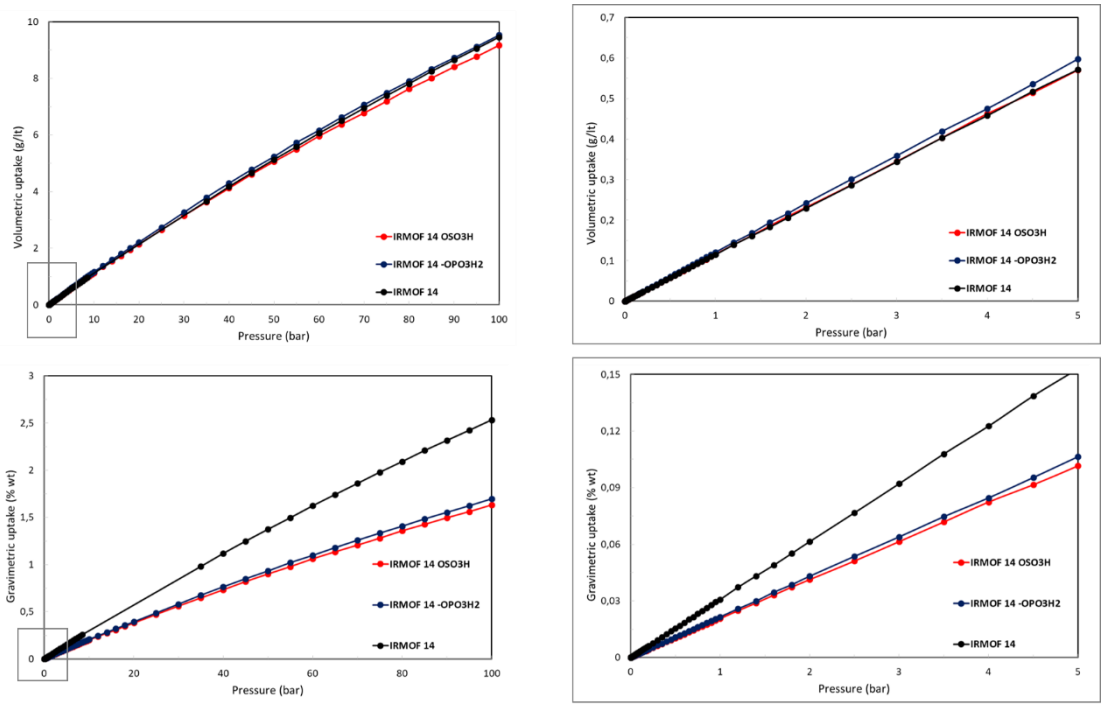


Figure 70: Volumetric (g/L) and gravimetric (%wt) hydrogen uptake at **T=300K** of IRMOF-14 and IRMOF-14-n (n: -OSO_3H, -OPO_3H_2).

For T=300K, the volumetric uptake shows no notable enhancement by the introduction of FGs in the structure; the uptake profile is almost the same for IRMOF-14 and IRMOF-14-n. The gravimetric uptake shows that the strength of the introduced binding sites is not able to compensate the weight of the FGs, with the unmodified structure clearly outperforming the functionalized counterparts for all pressures.

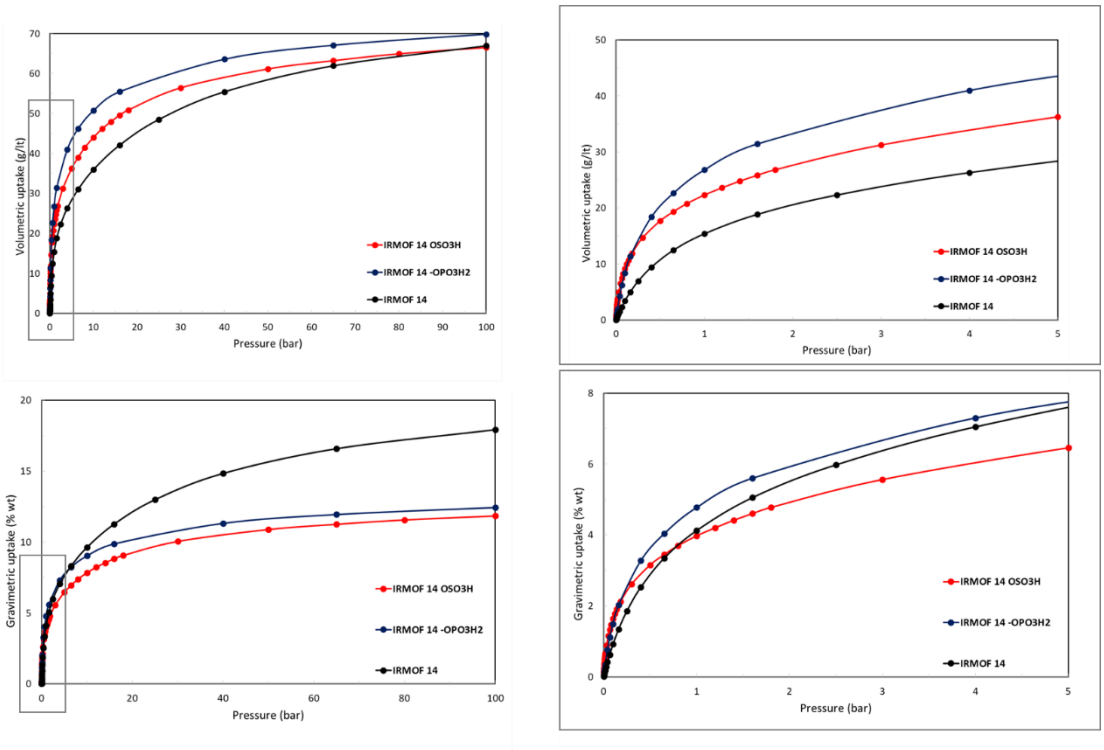


Figure 71: Volumetric (g/L) and gravimetric (%wt) hydrogen uptake at **T=77K** of IRMOF-14 and IRMOF-14-n (n: -OSO_3H, -OPO_3H_2).

For T=77K, the volumetric uptake is significantly enhanced in the case of the functionalized IRMOF-14 compared to the parent structure, with a ~20% enhancement for OSO_3H- and more than ~40% for OPO_3H_2-functionalized at P=5bar. From the two FGs, -OPO_3H_2 shows greater uptake for all the range of pressures. For the gravimetric uptake on the other hand, the situation is again reversed for high loadings, with the unfunctionalized stucture presenting a significantly greater uptake. At P >1bar, parent IRMOF-14 and IRMOF-14-OPO_3H_2 outperform IRMOF-14-OSO_3H, while at very low pressures (P~0.1bar), both functionalized candidates show a much greater uptake than that of the parent structure.

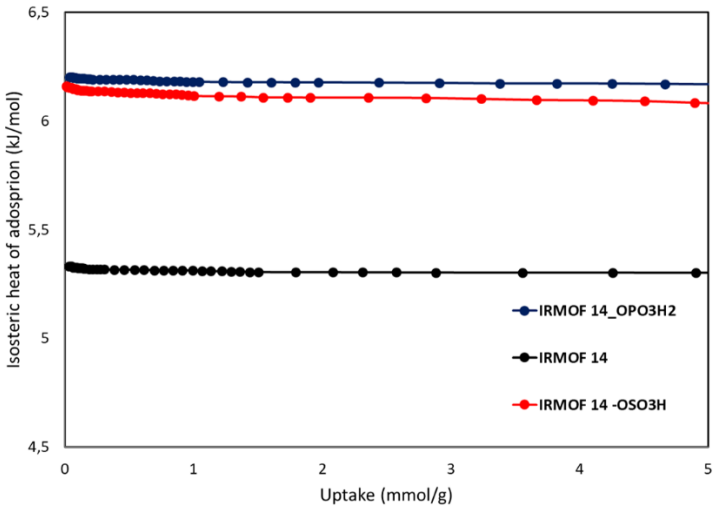


Figure 72: Isosteric heat of adsorption (kJ/mol) of IRMOF-14 and IRMOF-14-n (n: -OSO_3H, -OPO_3H_2).

Isosteric heats of Fig. 72 show that at zero loading, IRMOF-14-OSO_3H and IRMOF-14-OPO_3H_2 have similar isosteric heats ~6.2 kJ/mol whereas the unmodified IRMOF-14 structure shows a lower value of ~5.3 kJ/mol. From the ab initio calculations, the absolute binding energies of H_2…C_6H_6, H_2…C_6H_5-OSO_3H and H_2…C_6H_5-OPO_3H_2 are 4.6, 8.3 and 5.7 kJ/mol respectively. Since the ab initio screening was carried out on the simplified model of C_6H_5-X, whereas the linker of the IRMOF-14 is the pyrene dicarboxylate, differences are reasonable and expected. In addition, and as mentioned before, in the case of IRMOF-08, the ab initio calculations were employed in an isolate molecular environment and not within the pores of a robust material.

In Fig. 73-74 we present snapshots for the parent and the FG-modified IRMOF-14 taken from the GCMC simulations at T=77K and low to moderate pressures (i.e. 0.1, 1, and 5 bar).

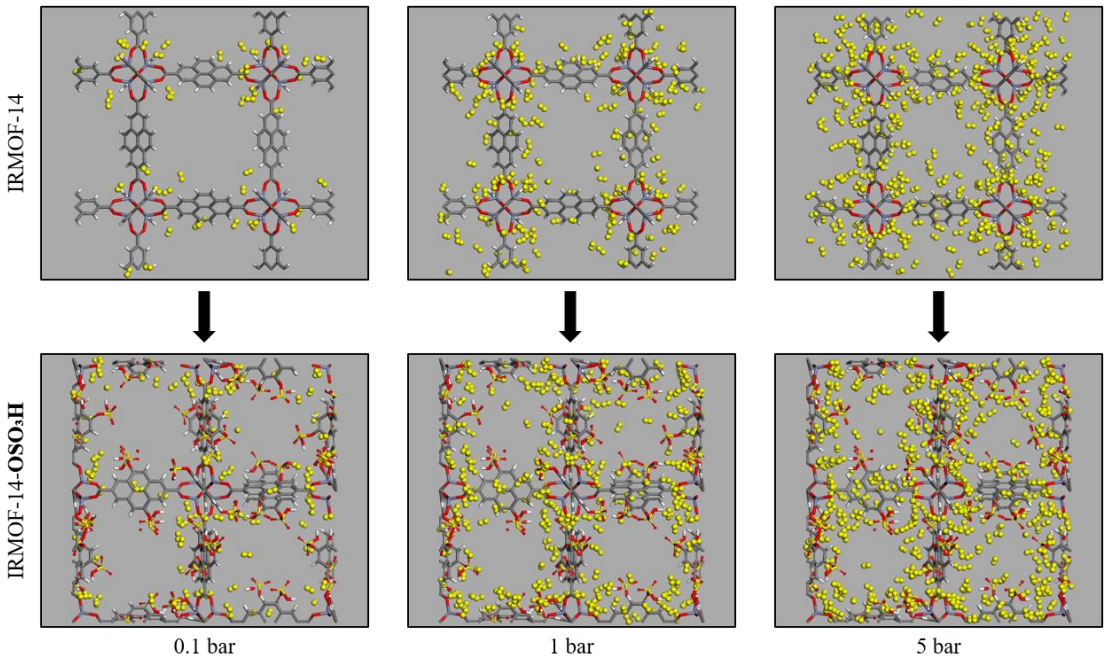


Figure 73: Snapshots of IRMOF-14 and IRMOF-14-OSO_3H at T=77K.

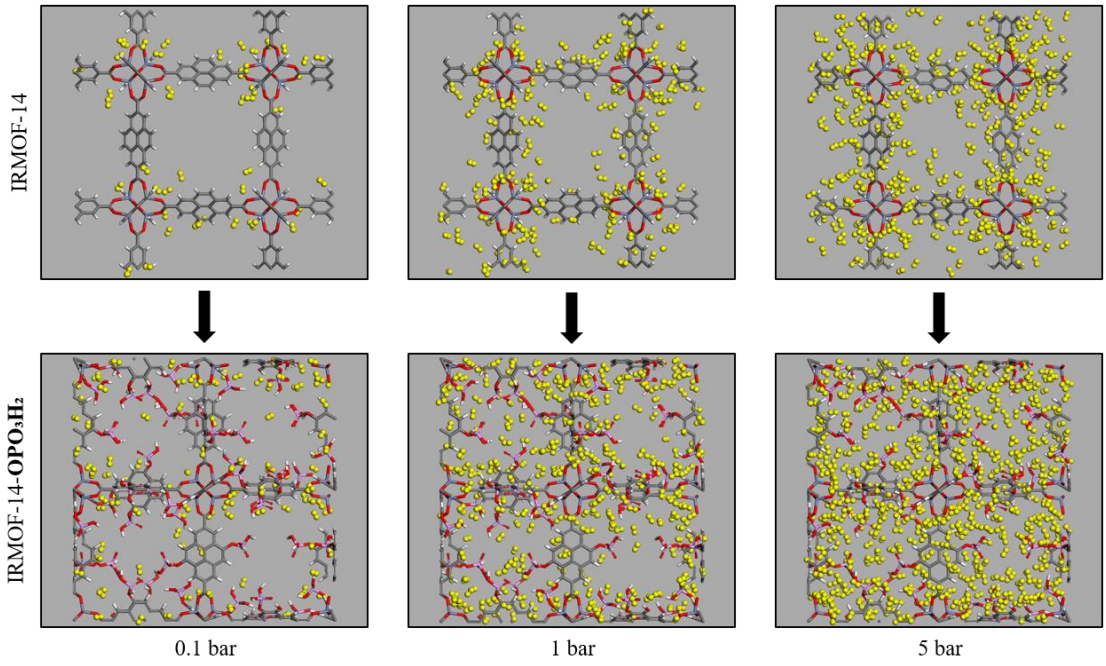


Figure 74: Snapshots of IRMOF-14 and IRMOF-14-OPO_3H_2 at T=77K.

CONCLUDING REMARKS

In this work, we studied the effect of linker functionalization of MOF structures on the enhancement of their hydrogen uptake capacities, by means of multi-scale molecular simulations. In the first place, we studied with ab initio methods the binding strength towards hydrogen of 58 functionalized benzenes of the form C_6H_5-X (X: the different FGs). From this set, 37 FGs (Fig. 42-44) showed higher binding energies compared to benzene, with 11 of them presenting a binding enhancement over 20% (Fig. 44). Notably, -OSO_3H FG stands out with an enhancement of ~80%. In the context of surface modification, the FGs that exhibit an enhancement over 20% are recommended as promising for potential linker functionalization of MOF candidates. In order to describe accurately the intermolecular interactions between the guest hydrogen molecules and the MOF atoms, we used ab initio derived results and fitted the parameters of the DREIDING potential onto these results. All of our studied cases indicate that available Force-Fields fail to describe accurately our systems, highlighting the importance of the fitting procedure. The effect of linker functionalization on the H_2 adsorption uptake of MOFs was studied employing Grand Canonical Monte Carlo simulations at cryogenic and ambient conditions. We functionalized three different MOF candidates from the IRMOF family (IRMOF-08, IRMOF-16 and IRMOF-14) with the FGs that showed the strongest binding towards hydrogen according to our ab initio calculations (-$OCONH_2$, -SO_2NH_2, -OSO_3H and -OPO_3H_2). A significant enhancement of the volumetric uptake at 77K was observed for all cases. While structural features clearly influence the packing of H_2 molecules within an adsorbent, of equal importance are the specific binding locations within the pores and the relative binding strength at these sites. Taking into consideration the weight of the FGs as well as other important factors (e.g. pore environment), the adsorption is in agreement with the strength of interaction between the adsorbate (H_2) and the organic parts of the adsorbent (MOF). However, within the scope of practical applications, the absolute uptake percentages illustrate clearly that considerable work is still required in the field.

Overall, our work has set up a systematic and transferable database that can be used for experimental and further computational studies on the enhancement of hydrogen uptake capacities of porous materials in general (i.e. MOFs, COFs, ZIFs, etc.).

REFERENCES

[1] A. Züttel, *Hydrogen Storage Methods*, Naturwissenschaften **91**, 157 (2004).

[2] *Wikipedia: Fuel Cell*, https://en.wikipedia.org/wiki/Fuel_cell.

[3] *Energy Policy Act 1992*, https://afdc.energy.gov/files/pdfs/2527.pdf.

[4] Engineering-ToolBox, *Fossil and Alternative Fuels - Energy Content*, https://www.engineeringtoolbox.com/fossil-fuels-energy-content-d_1298.html.

[5] *Toyota Mirai Fuel Cell*, www.toyota.com/content/ebrochure/2018/mirai_FuelCellTech.pdf.

[6] *Fuel Cell Technologies Office, Hydrogen Storagee*, https://www.energy.gov/sites/prod/files/2017/03/f34/fcto-h2-storage-fact-sheet.pdf.

[7] E. Rivard, M. Trudeau, and K. Zaghib, *Hydrogen Storage for Mobility: A Review*, Materials (Basel). **12**, (2019).

[8] *DOE Technical Targets for Onboard Hydrogen Storage for Light-Duty Vehicles*, https://www.energy.gov/eere/fuelcells/doe-technical-targets-onboard-hydrogen-storage-light-duty-vehicles.

[9] U. Eberle, M. Felderhoff, and F. Schüth, *Chemical and Physical Solutions for Hydrogen Storage*, Angew. Chemie - Int. Ed. **48**, 6608 (2009).

[10] O. M. Yaghi, G. Li, and H. Li, *Selective Binding and Removal of Guests in a Microporous Metal–Organic Framework*, Nature **378**, 703 (1995).

[11] *ScienceImage*, https://www.scienceimage.csiro.au/pages/about/.

[12] E. Sharmin and F. Zafar, *Introductory Chapter: Metal Organic Frameworks (MOFs)*, in *Metal-Organic Frameworks* (2016).

[13] H. Furukawa, K. E. Cordova, M. O'Keeffe, and O. M. Yaghi, *The Chemistry and Applications of Metal-Organic Frameworks*, Science (80-.). **341**, (2013).

[14] V. V Butova, M. A. Soldatov, A. A. Guda, K. A. Lomachenko, and C. Lamberti, *Metal-Organic Frameworks: Structure, Properties, Methods of Synthesis and Characterization*, Russ. Chem. Rev. **85**, 280 (2016).

[15] P. Rocío-Bautista, I. Taima-Mancera, J. Pasán, and V. Pino, *Metal-Organic Frameworks in Green Analytical Chemistry*, Separations **6**, 1 (2019).

[16] N. Stock and S. Biswas, *Synthesis of Metal-Organic Frameworks (MOFs): Routes to Various MOF Topologies, Morphologies, and Composites*, Chem. Rev. **112**, 933 (2012).

[17] M. Rubio-Martinez, C. Avci-Camur, A. W. Thornton, I. Imaz, D. Maspoch, and M. R. Hill, *New Synthetic Routes towards MOF Production at Scale*, Chem. Soc. Rev. **46**, 3453 (2017).

[18] C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp, and R. Q. Snurr, *Large-Scale Screening of Hypothetical Metal–Organic Frameworks*, Nat. Chem. **4**, 83 (2012).

[19] N. L. Rosi, J. Eckert, M. Eddaoudi, D. T. Vodak, J. Kim, M. O'Keeffe, and O. M. Yaghi, *Hydrogen Storage in Microporous Metal-Organic Frameworks*, Science (80-.). **300**, 1127 LP (2003).

[20] M. T. Kapelewski, T. Runčevski, J. D. Tarver, H. Z. H. Jiang, K. E. Hurst, P. A. Parilla, A. Ayala, T. Gennett, S. A. Fitzgerald, C. M. Brown, and J. R. Long, *Record High Hydrogen Storage Capacity in the Metal-Organic Framework Ni2(m-Dobdc) at Near-Ambient Temperatures*, Chem. Mater. **30**, 8179 (2018).

[21] X. Lin, J. Jia, X. Zhao, K. M. Thomas, A. J. Blake, G. S. Walker, N. R. Champness, P. Hubberstey, and M. Schröder, *High H2 Adsorption by Coordination-Framework Materials*, Angew. Chemie Int. Ed. **45**, 7358 (2006).

[22] D. A. Gómez-Gualdrón, T. C. Wang, P. García-Holley, R. M. Sawelewa, E. Argueta, R. Q. Snurr, J. T. Hupp, T. Yildirim, and O. K. Farha, *Understanding Volumetric and Gravimetric Hydrogen Adsorption Trade-off in Metal–Organic Frameworks*, ACS Appl. Mater. Interfaces **9**, 33419 (2017).

[23] V. V Simonyan, P. Diep, and J. K. Johnson, *Molecular Simulation of Hydrogen Adsorption in Charged Single-Walled Carbon Nanotubes*, J. Chem. Phys. **111**, 9778 (1999).

[24] H. Li, K. Wang, Y. Sun, C. T. Lollar, J. Li, and H. C. Zhou, *Recent Advances in Gas Storage and Separation Using Metal–Organic Frameworks*, Mater. Today **21**, 108

(2018).

[25] J. L. C. Rowsell and O. M. Yaghi, *Strategies for Hydrogen Storage in Metal–Organic Frameworks*, Angew. Chemie Int. Ed. **44**, 4670 (2005).

[26] O. Hübner, A. Glöss, M. Fichtner, and W. Klopper, *On the Interaction of Dihydrogen with Aromatic Systems*, J. Phys. Chem. A **108**, 3019 (2004).

[27] A. Ahmed, S. Seth, J. Purewal, A. G. Wong-Foy, M. Veenstra, A. J. Matzger, and D. J. Siegel, *Exceptional Hydrogen Storage Achieved by Screening Nearly Half a Million Metal-Organic Frameworks*, Nat. Commun. **10**, (2019).

[28] J. L. C. Rowsell and O. M. Yaghi, *Effects of Functionalization, Catenation, and Variation of the Metal Oxide and Organic Linking Units on the Low-Pressure Hydrogen Adsorption Properties of Metal-Organic Frameworks*, J. Am. Chem. Soc. **128**, 1304 (2006).

[29] G. Lu and E. Kaxiras, *An Overview of Multiscale Simulations of Materials*.

[30] M. Born and R. Oppenheimer, *Zur Quantentheorie Der Molekeln*, Ann. Phys. **389**, 457 (1927).

[31] *Lecture Notes*, http://pollux.chem.umn.edu/8021/Lectures/Preamble_1.pdf.

[32] R. Hoffmann, *An Extended Hückel Theory. I. Hydrocarbons*, J. Chem. Phys. **39**, 1397 (1963).

[33] C. David Sherrill and H. F. Schaefer, *The Configuration Interaction Method: Advances in Highly Correlated Approaches*, in edited by P.-O. Löwdin, J. R. Sabin, M. C. Zerner, and E. B. T.-A. in Q. C. Brändas, Vol. 34 (Academic Press, 1999), pp. 143–269.

[34] H. G. Kümmel, *A Biography of the Coupled Cluster Method*, in *Recent Progress in Many-Body Theories*, Vol. Volume 6 (WORLD SCIENTIFIC, 2002), pp. 334–348.

[35] I. Shavitt and R. J. Bartlett, *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory* (Cambridge University Press, Cambridge, 2009).

[36] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models* (2004).

[37] D. A. McQuarrie and J. D. Simon, *Physical Chemistry: A Molecular Approach* (1997).

[38] R. Ditchfield, W. J. Hehre, and J. A. Pople, *Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules*, J. Chem. Phys. **54**, 724 (1971).

[39] T. H. Dunning, *Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen*, J. Chem. Phys. **90**, 1007 (1989).

[40] F. Weigend and R. Ahlrichs, *Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy*, Phys. Chem. Chem. Phys. **7**, 3297 (2005).

[41] S. G. Balasubramani, G. P. Chen, S. Coriani, M. Diedenhofen, M. S. Frank, Y. J. Franzke, F. Furche, R. Grotjahn, M. E. Harding, C. Hättig, A. Hellweg, B. Helmich-Paris, C. Holzer, U. Huniar, M. Kaupp, A. Marefat Khah, S. Karbalaei Khani, T. Müller, F. Mack, B. D. Nguyen, S. M. Parker, E. Perlt, D. Rappoport, K. Reiter, S. Roy, M. Rückert, G. Schmitz, M. Sierka, E. Tapavicza, D. P. Tew, C. van Wüllen, V. K. Voora, F. Weigend, A. Wodyński, and J. M. Yu, *TURBOMOLE: Modular Program Suite for Ab Initio Quantum-Chemical and Condensed-Matter Simulations*, J. Chem. Phys. **152**, 184107 (2020).

[42] S. F. Boys and F. Bernardi, *The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors*, Mol. Phys. **19**, 553 (1970).

[43] P. Hohenberg and W. Kohn, *Inhomogeneous Electron Gas*, Phys. Rev. **136**, B864 (1964).

[44] N. Mardirossian and M. Head-Gordon, *Thirty Years of Density Functional Theory in Computational Chemistry: An Overview and Extensive Assessment of 200 Density Functionals*, Mol. Phys. **115**, 2315 (2017).

[45] S. H. Vosko and L. Wilk, *Influence of an Improved Local-Spin-Density Correlation-Energy Functional on the Cohesive Energy of Alkali Metals*, Phys. Rev. B **22**, 3812 (1980).

[46] J. P. Perdew and A. Zunger, *Self-Interaction Correction to Density-Functional Approximations for Many-Electron Systems*, Phys. Rev. B **23**, 5048 (1981).

[47] J. P. Perdew and Y. Wang, *Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy*, Phys. Rev. B **45**, 13244 (1992).

[48] A. D. Becke, *Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior*, Phys. Rev. A **38**, 3098 (1988).

[49] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, *Atoms, Molecules, Solids, and Surfaces: Applications of the Generalized Gradient Approximation for Exchange and Correlation*, Phys. Rev. B **46**, 6671 (1992).

[50] J. P. Perdew, K. Burke, and M. Ernzerhof, *Generalized Gradient Approximation Made Simple*, Phys. Rev. Lett. **77**, 3865 (1996).

[51] J. P. Perdew, *Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas*, Phys. Rev. B **33**, 8822 (1986).

[52] C. Lee, W. Yang, and R. G. Parr, *Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density*, Phys. Rev. B **37**, 785 (1988).

[53] J. P. Perdew, S. Kurth, A. Zupan, and P. Blaha, *Accurate Density Functional with Correct Formal Properties: A Step Beyond the Generalized Gradient Approximation*, Phys. Rev. Lett. **82**, 2544 (1999).

[54] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, *Climbing the Density Functional Ladder: Nonempirical Meta--Generalized Gradient Approximation Designed for Molecules and Solids*, Phys. Rev. Lett. **91**, 146401 (2003).

[55] J. Sun, R. Haunschild, B. Xiao, I. W. Bulik, G. E. Scuseria, and J. P. Perdew, *Semilocal and Hybrid Meta-Generalized Gradient Approximations Based on the Understanding of the Kinetic-Energy-Density Dependence*, J. Chem. Phys. **138**, 44113 (2013).

[56] Y. Zhao and D. G. Truhlar, *A New Local Density Functional for Main-Group Thermochemistry, Transition Metal Bonding, Thermochemical Kinetics, and Noncovalent Interactions*, J. Chem. Phys. **125**, 194101 (2006).

[57] C. Adamo and V. Barone, *Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model*, J. Chem. Phys. **110**, 6158 (1999).

[58] A. D. Becke, *A New Mixing of Hartree–Fock and Local Density-functional Theories*, J.

Chem. Phys. **98**, 1372 (1993).

[59] S. Grimme, *Accurate Description of van Der Waals Complexes by Density Functional Theory Including Empirical Corrections*, J. Comput. Chem. **25**, 1463 (2004).

[60] S. Grimme, *Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction*, J. Comput. Chem. **27**, 1787 (2006).

[61] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, *A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu*, J. Chem. Phys. **132**, 154104 (2010).

[62] S. Grimme, S. Ehrlich, and L. Goerigk, *Effect of the Damping Function in Dispersion Corrected Density Functional Theory*, J. Comput. Chem. **32**, 1456 (2011).

[63] S. L. Mayo, B. D. Olafson, and W. A. Goddard, *DREIDING: A Generic Force Field for Molecular Simulations*, J. Phys. Chem. **94**, 8897 (1990).

[64] J. E. Jones and S. Chapman, *On the Determination of Molecular Fields. —II. From the Equation of State of a Gas*, Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character **106**, 463 (1924).

[65] H. A. Lorentz, *Ueber Die Anwendung Des Satzes Vom Virial in Der Kinetischen Theorie Der Gase*, Ann. Phys. **248**, 127 (1881).

[66] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, *UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations*, J. Am. Chem. Soc. **114**, 10024 (1992).

[67] D. J. Adams, *On the Use of the Ewald Summation in Computer Simulation*, J. Chem. Phys. **78**, 2585 (1983).

[68] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules*, J. Am. Chem. Soc. **117**, 5179 (1995).

[69] H. Sun, *COMPASS: An Ab Initio Force-Field Optimized for Condensed-Phase ApplicationsOverview with Details on Alkane and Benzene Compounds*, J. Phys. Chem.

B **102**, 7338 (1998).

[70] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, *CHARMM: The Biomolecular Simulation Program*, J. Comput. Chem. **30**, 1545 (2009).

[71] W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger, and W. F. van Gunsteren, *The GROMOS Biomolecular Simulation Program Package*, J. Phys. Chem. A **103**, 3596 (1999).

[72] F. Darkrim and D. Levesque, *Monte Carlo Simulations of Hydrogen Adsorption in Single-Walled Carbon Nanotubes*, J. Chem. Phys. **109**, 4981 (1998).

[73] S. S. Han, W.-Q. Deng, and W. A. Goddard III, *Improved Designs of Metal–Organic Frameworks for Hydrogen Storage*, Angew. Chemie Int. Ed. **46**, 6289 (2007).

[74] C. M. Breneman and K. B. Wiberg, *Determining Atom-centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis*, J. Comput. Chem. **11**, 361 (1990).

[75] D. Frenkel, *Introduction to Monte Carlo Methods*, in *Computational Soft Matter: From Synthetic Polymers to Proteins* (2004).

[76] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *Equation of State Calculations by Fast Computing Machines*, J. Chem. Phys. **21**, 1087 (1953).

[77] D. Nicholson and N. G. Parsonage, *Computer Simulation and the Statistical Mechanics of Adsorption* (London: Academic Press., 1982).

[78] B. Widom, *Some Topics in the Theory of Fluids*, J. Chem. Phys. **39**, 2808 (1963).

[79] D.-Y. Peng and D. B. Robinson, *A New Two-Constant Equation of State*, Ind. Eng. Chem. Fundam. **15**, 59 (1976).

[80] *Notes: Basics of Molecular Modelling*, http://hebergement.u-psud.fr/fabien-

cailliez/Downloads/basics_molecularmodeling_1314.pdf.

[81] R. S. Katiyar and P. K. Jha, *Molecular Simulations in Drug Delivery: Opportunities and Challenges*, WIREs Comput. Mol. Sci. **8**, e1358 (2018).

[82] *Lecture Notes*, http://www.grs-sim.de/cms/upload/Carloni/Tutorials/CMB/ComputerSimulationMethods_2.pdf.

[83] *Avogadro: An Open-Source Molecular Builder and Visualization Tool. Version 1.XX*.

[84] F. Neese, *Software Update: The ORCA Program System, Version 4.0*, WIREs Comput. Mol. Sci. **8**, e1327 (2018).

[85] F. Weigend, M. Häser, H. Patzelt, and R. Ahlrichs, *RI-MP2: Optimized Auxiliary Basis Sets and Demonstration of Efficiency*, Chem. Phys. Lett. **294**, 143 (1998).

[86] J. S. Murray and P. Politzer, *The Electrostatic Potential: An Overview*, Wiley Interdiscip. Rev. Comput. Mol. Sci. **1**, 153 (2011).

[87] L. Laaksonen, *A Graphics Program for the Analysis and Display of Molecular Dynamics Trajectories*, J. Mol. Graph. **10**, 33 (1992).

[88] D. L. Bergman, L. Laaksonen, and A. Laaksonen, *Visualization of Solvation Structures in Liquid Mixtures*, J. Mol. Graph. Model. **15**, 301 (1997).

[89] M. Eddaoudi, J. Kim, N. Rosi, D. Vodak, J. Wachter, M. Keeffe, and O. M. Yaghi, *Systematic Design of Pore Size and Functionality in Isoreticular MOFs and Their Application in Methane Storage*, Science (80-.). **295**, 469 LP (2002).

[90] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, *Auxiliary Basis Sets to Approximate Coulomb Potentials (Chem. Phys. Letters 240 (1995) 283-290)*, Chem. Phys. Lett. **242**, 652 (1995).

[91] C. R. Groom, I. J. Bruno, M. P. Lightfoot, and S. C. Ward, *The Cambridge Structural Database*, Acta Crystallogr. Sect. B **72**, 171 (2016).

[92] D. Berthelot, *Sur Le Mélange Des Gaz*, Comptes Rendus Hebd. Des Séances l'Académie Des Sci. **126**, 1703 (1898).

[93] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill: New York, 1965).

[94] D. Dubbeldam, S. Calero, D. E. Ellis, and R. Q. Snurr, *RASPA: Molecular Simulation Software for Adsorption and Diffusion in Flexible Nanoporous Materials*, Mol. Simul. **42**, 81 (2016).

[95] M. Vert, Y. Doi, K.-H. Hellwich, M. Hess, P. Hodge, P. Kubisa, M. Rinaudo, and F. Schué, *Terminology for Biorelated Polymers and Applications*, Pure Appl. Chem. **84**, 377 (2012).