UNIVERSITY OF CALIFORNIA
RIVERSIDE

Density Functional Theory for High-Throughput Screening of Chemicals and Materials

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemical and Environmental Engineering

by

Jia Fu

June 2015

Dissertation Committee:
Dr. Jianzhong Wu, Chairperson
Dr. David Kisailus
Dr. Chia-En Chang

The Dissertation of Jia Fu is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

I feel grateful to many people for their great support over the years. First and foremost, I would like to thank my supervisor Prof. Jianzhong Wu for his superb and insightful guidance. Learned from him, I start to know how to think as a theoretician, to solve problems in an elegant way and to enjoy the beauty of statistical mechanics. His strong enthusiasm motivate me and shows an excellent example on how to be a successful chemical engineer and professor.

I sincerely appreciate my PhD committee, Prof. David Kisailus, Prof. Chia-en A. Chang for their great advice and help in my qualifying exam and final defense. I also wish to express my special thanks to Prof. De-en Jiang, Prof. Shuangliang Zhao, Dr. Zhehui Jin, Dr. Ke Wang, Dr. Yu Liu and Dr. Jehoon Kim for their valuable support during my PhD studying. Also all the good friends that I have made in Riverside for their encouragement throughout this special period.

What's more, I greatly appreciate the endless support and patience of my parents during this long journey, who support and encourage me more than anyone else to pursue an academic career.

In this dissertation, some parts are the reprint of the materials appeared in Journal of Physical Chemistry Letter (Volume 4, Issue 21, Page 3687-3691, October 2013), Journal Computer-Aided Molecular Design (Volume 28, Issue 3, Page 299-304, March 2014), Chemical Engineering Science (Volume 126, Page 370-382, April 2015), Langmuir (Volume 29, Issue 42, Page 12997-13002, 2013), Journal of Physical Chemistry C (Volume 119, Issue 10, Page 5374-5385, 2015), Journal of Colloid and Interface Science (Volume 438, Page 191-195, January 2015). This work is financially sponsored by the US Department of Energy (DOE) (DE-FG02-06ER46296 and DE-AC02-05CH11231), the National Science Foundation (NSF-CBET-0852353, NSF-CBET-1404046). This work utilizes supercomputer from the National Energy Research Scientific Computing Center (NERSC).

ABSTRACT OF THE DISSERTATION

Density Functional Theory for High-Throughput Screening of Chemicals and Materials

by

Jia Fu

Doctor of Philosophy
Graduate Program in Chemical and Environmental Engineering
University of California, Riverside, June 2015
Professor Jianzhong Wu, Chairperson

Multiscale modelling is an essential part in chemical engineering field, recently with the developments in physical and computer sciences, computer modelling plays a more and more important role in quantitative predictions of chemical reactions and thermodynamic data, covering broad applications to problems of practical concern. Whereas classical simulations and quantum density functional theory methods have become routine tools for engineering applications, the power of classical density functional theory has not been widely recognized, which has already shown great success in many model systems. Considering its efficiency and accuracy, it's necessary to establish a systematic way to utilize this method applied in realistic applications. During this PhD research, I mainly focus on developing classical density functional theory for large-scale screening of chemicals and materials, especially for organic compounds solvation free energy and gas storage and transport properties in nanoporous materials.

For the hydration free energy calculation part, we try to build a reliable multiscale procedure to predict the water solvation free energies of organic compounds at ambient conditions, by combination of quantum mechanics, statistical mechanics and molecular mechanics force fields. Using the experimental data of hydration free energies as the benchmark, we show that the new method, taking the explicit solvent molecules into consideration, shows a comparable accuracy as full atomistic molecular simulations, but over one magnitude faster calculation speed. We also find that the theoretical results are sensitive to the input generation methods, i.e. selection of quantum-mechanical methods for determining atomic charges and solute configurations, the assignment of the van der Waals parameters and the solvent models. In generally speaking, HF/6-31G/Vac/OPLS/ChelpG gives the best prediction, in particular those less hydrophilic ones, the overall average unsiged error is 1.35 kcal/mol compared to 700 common organic molecular hydration free energy database, close to the experimental measurement chemical accuracy limit 1 kcal/mol.

On the other hand, we have investigated the four representative versions of non-local density functionals as efficient methods to predict gas adsorption in both simple slit pore model system and a large library of real metal-organic frameworks (MOFs) under a broad range of temperatures and pressures, compared with grand canonical Monte Carlo simulation data. Overall all the four methods are reasonably accurate in comparison with the simulation results, but our results show for each specific gas and interested condition, there is a best candidate: for DOE's hydrogen storage target condition, it's FMSA, while for ARPA-E's methane storage target condition, it's MFA. Besides, combing with entropy

scaling method, we can also predict the gas self-diffusion coefficient in the same one time calculation with an average calculating time 30 seconds, far beyond the classical molecular simulation ways.

From a computational perspective, the classical density functional theory is shown as a versatile and promising tool for large-scale screening of thousands of chemicals and nanostructured materials with a modern desktop computer.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Background

In this work, we apply many of the modern computer modeling methods, from classical simulation methods to modern liquid state theory, most importantly classical density functional theory (CDFT). So in this chapter we first gives a brief introduction of modern computer modeling methods we covered, then the introduction to the background of hydration free energy and gas storage in nanoporous materials, finally the organization of this PhD thesis. The next chapter we will focus on the basic formulism of CDFT.

1.1.1 Quantum Mechanics Methods

Back to the 1920s, after the establishment of quantum mechanics (QM), as Dirac indicates, "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 that are much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to explanation of the main features of complex atomic systems without too much computation." Since then, lots of people devote themselves to QM methods developments and application in the past 90 years.

I. Wavefunction Method

Before 1980s, most of the effort were spent on the wavefunction approaches, providing approximation solutions to the Schrödinger equation (eq (1)) that describes any given chemical system.

$$i\hbar\frac{\partial}{\partial t}\Psi = \hat{H}\Psi \tag{1}$$

All these methods originates from the work of Hartree and Fock in the 1920s,[1] which assumes that the exact N-body wavefunction $\Psi(r,\mathbf{R})$ of the system can be approximated by a single slater determinant of N spin orbitals. By applying the variational principle, one can derive a set of N-coupled equations for the N spin orbitals. Solution of these equations yields the Hartree–Fock wavefunction and energy of the system, which are upper-bound approximations of the exact ones. The main disadvantage of the original HF method is that it assumes electrons moving independently of each other, without considering electron correlation. So its accuracy was limited. Thus, nowadays, we usually employ more advanced ‘‘post-HF’’ approaches to real chemical systems, from popular Møller–Plesset perturbation theory (MP)[2] to more accurate coupled cluster (CC)[3] or configuration interaction (CI)[4] methods, which provide different ways of counting the correlation effects missing from HF and approximating the exact wavefunction of the system. Those “post-HF” methods are recognized as the most reliable and accurate QM methods, especially when van der Waals interaction plays an important role in the system. However, the high computational cost of those methods are prohibitive for studying molecules containing over 20 atoms, which is an obvious shortcoming when applied to engineering fields.

II. Quantum Density Functional Theory

In 1964, the great work of Hohenberg and Kohn give the bedrock of modern quantum density functional theory (QDFT) methods by three theorems.[5] The first theorem states that the ground-state electron density uniquely determines the electronic

wavefunction and hence all ground-state properties of an electronic system. The second one establishes that the energy of an electron distribution can be described as a functional of the electron density, while the third one indicates this functional is a minimum for the ground state density. The problem of solving the many-body Schrödinger equation is converted to minimize a density functional. After it was born, QDFT shows its great advantages compared to the inaccuracy of HF and the high computational demanding of post-HF methods. Especially after 1998, when Walter Kohn won the Nobel Prize "for his development of the density-functional theory", more and more eminent works are done by QDFT.[6]

Although the Hohenberg–Kohn theorems prove that the density functional is a universal quantity; they do not specify its form. The common realization of modern QDFT method is through the Kohn–Sham (KS) approach.[7] The KS method is on the basis of the construction of a noninteracting reference system yielding the same density as the original problem.

$$E(\rho) = T_0(\rho) + \int v_{ext}\rho + J(\rho) + E_{ex}(\rho) \tag{2}$$

where on the right, the first term is the energy of reference system, the second term is energy due to the external field, the third term is the kinetic energy and the forth term is the exchange-correlation energy contribution. While the form of the first three energy functional parts are known exactly, the only unknown term is the exchange–correlation functional, which is not known except for the free electron gas. So the central theme for modern QDFT developments is to find proper approximations for $E_{ex}(\rho)$ that could permit the calculation of molecular properties at certain level of accuracy.

The most fundamental and simplest approximation for the exchange–correlation functional is the local-density approximation (LDA), in which the energy only depends on the density at the point where the functional is evaluated,[8] corresponding to that of a homogeneous electron gas at the same density (eq (3)).

$$E_{XC}^{LDA} = \int e_{XC}^{UEG}(\rho) \tag{3}$$

While LDA achieves a great success in solid state physics, its strong tendency of over-binding prediction fails a lot in the studying of chemical systems. To better describe the inhomogeneous nature of molecular systems, generalized gradient approximation (GGA) (eq (4)) was introduced by considering the energy depending on both the electron density and its gradient. GGA functionals such as BP86,[9] PBE[10] and BLYP[9] are proved to be very efficient and yielding good results, particularly for molecular structures, but are often less accurate for other properties.

$$E_{XC}^{GGA}\left[n_{\uparrow}, n_{\downarrow}\right] = \int e_{XC}(n_{\uparrow}, n_{\downarrow}, \vec{\nabla} n_{\uparrow}, \vec{\nabla} n_{\downarrow}) n(\mathbf{r}) d\mathbf{r} \tag{4}$$

In 1993, Becke introduces hybrid functionals (eq (5)) to QDFT community, which mix GGA functionals with part of exact Hartree–Fock exchange contribution,[11] which are still the main stream QDFT methods for chemical systems nowadays. Among them, B3LYP functional[9] is one of the most successfully and widely used QDFT methods for chemists in a wide variety of chemical systems and properties[12].

$$E_{XC} = aE_X^{HF} + (1-a)E_X^{LSDA} + bE_X^{GGA} + E_C^{LSDA} + cE_C^{GGA} \tag{5}$$

Following Perdew's Jacob's Ladder,[13] recent theoretical developments include the ''meta-GGA'' functionals, which extend the GGA corrections to higher derivatives, like TPSS[14]

and M062X[15]; and the ‘‘double hybrid’’ functionals, such as B2PLYP[16] and PWPB95,[17] which contain not only a fraction of exact HF exchange contribution but also a fraction of orbital-dependent non-local correlation energy contribution estimated at the level of second-order many-body perturbation theory (MP2). These new functionals yield better reaction energies and other basic properties compared to experimental data, even beyond more expensive MP2 method.[18] Noncovalent interactions is often thought to be a major shortcoming of traditional QDFT methods due to the lack of accurate long-range correlation part of free energy functional. Besides double-hybrid functionals, range-separated functionals are also considered to be promising solutions, like CAM-B3LYP[19] and ωB97X-D.[20] Another more engineering way to solve this problem, we think is semi-empirical London-dispersion correction method, which could be easily applied to current QDFT methods and shows significant improvement on describing van der Waals interactions in recent papers.[18] [21]

Quantum density functional theory methods have already been proved as a valuable research tool both in independent theoretical studying and as a complement of experimental investigations in broad areas, for example, molecular structures, vibrational frequencies, spectroscopic properties and total energies. However, we need to realize that unlike the wavefunction-based methods, there is still no systematic way to give a universal choice for QDFT methods, in other words, the success of certain functional in one case does not guarantee its performance in other situations. Therefore, before applied certain QDFT method to our works, we must do reasonable validation and estimate the error range.

1.1.2. Classical Simulation Methods

As Titus Lucretius says, "The atoms are eternal and always moving. Everything comes into existence simply because of the random movement of atoms, which, given enough time, will form and reform, constantly experimenting with different configurations of matter from which will eventually emerge everything we know...". Given enough time, we could use quantum mechanics to solve any system we are interested, however, even today, it's impossible to use full QM methods to study complicated molecular system as indicated above. People develop lots of semi-empirical methods trying to solving this problem, like AM1[22] and PM3[23]. With the most recent dispersion corrected ones, like PM6-DH+[24] and PM7[25], we can study much larger system, like solvated biomolecular systems and host-guest binding systems[26] in an accepted time and accuracy, however, the accuracy beyond those training set molecules need to be carefully checked.

I. Molecular Mechanic Force Field

A more convenient and efficient way to describe the potential energy surface of complex systems is molecular mechanic force fields. Given the potential $U(\mathbf{r}_1,...,\mathbf{r}_N)$, the force acting upon *ith* atom is determined by the gradient with respect to atomic displacements $\mathbf{r}_i = (x_i, y_i, z_i)$,

$$\mathbf{F}_i = -\nabla_{\mathbf{r}_i} U(\mathbf{r}_1,...,\mathbf{r}_N) = -(\frac{\partial U}{\partial x_i}, \frac{\partial U}{\partial y_i}, \frac{\partial U}{\partial z_i}) \quad (6)$$

Due to the adiabatic (Born–Oppenheimer) approximation, we know that because of the difference in mass between nuclei and electrons, the electron cloud equilibrates much faster than the configuration of the heavy nuclei, which means the electronic and nuclear

problems can be separated. The nuclei could be assumed to move in the field of the averaged electron densities. Then we could describe the potential energy surface by the dynamics of the nuclei without taking explicitly account of the electrons and describes physical systems as collections of atoms kept together by interatomic forces, with empirical analytical functions, instead of solving Schrödinger equations, which could be much more easily applied to study the dynamics of molecular systems following classical mechanics. This starts from the pioneering works in Allinger[27], Lifson[28] and Scheraga[29] groups. After the born of MM2 and CFF[30] force fields, more and more generic force fields are developed, applied to material and biological research.

The most commonly used classical force field models (Class I, e.g. AMBER[31]) incorporate a relatively simple potential energy function:

$$\begin{aligned} E &= \sum_{bond} K_b (b - b_o)^2 + \sum_{angle} K_\theta (\theta - \theta_o)^2 \\ &+ \sum_{dihedral} \frac{K_\varphi}{2} \left[1 + \cos\left(n\varphi - \varphi_o \right) \right] \\ &+ \sum_{impr} \frac{K_\chi}{2} \left[1 + \cos\left(n\chi - \chi_o \right) \right] \\ &+ \sum_{nonbond} \left\{ \varepsilon_{ij} \left[\left(\frac{R_{ij}^o}{R_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^o}{R_{ij}} \right)^{6} \right] + \frac{q_i q_j}{R_{ij}} \right\} \end{aligned} \tag{7}$$

The first four terms describe the covalent structure of the molecular system, by chemical bonds, angles and torsion angles (dihedral and improper dihedral angles), while the last two terms sum the nonbond pair interactions (van der Waals and Coulomb) between atoms. Physically, the first two terms describe energies of deformations of the bond lengths b and bond angles θ from their respective equilibrium values b_0 and θ_0. The common harmonic

form of these terms (with force constants K_b and K_θ) ensures the correct chemical structure, but are not able to describe chemical changes such as bond breaking. The third and fourth terms describe the rotations of connecting chemical bonds characterized by periodic energy terms (with periodicity determined by n and heights of rotational barriers defined by K_φ and K_χ). The fifth term describes the van der Waals interatomic forces in the form of the standard Lennard–Jones 12-6 potential (dispersion), and the Coulomb electrostatic potential.

The task of finding reasonable force field parameters that would adequately mimic the true energy surfaces is nontrivial. It must represent the key physical and chemical properties of the systems of interest correctly. There are many successful force fields already widely used, for example, DREIDING[32] and UFF[33], which essentially covers all the elements in the periodic table and achieve a good quality in material crystal structure predictions, also in gas adsorption applications.[34-37] In biophysics field, AMBER[31], CHARMM[38] and OPLS[39] are proved to have good accuracy in calculating structure related biosystem properties. To further improving the performance of force fields, cross-correlation terms are introduced, like bond-angle correlation, angle-torsion correlation, which are usually called Class II force fields, for example, CFF93[40] and COMPASS[41], perform very well in condensed-phase applications for common organic molecules and polymers.

Thanks to the advancement of quantum mechanics, modern classic force fields could obtain rather accurate intramolecular structure information from first principle calculations, even nonbond interatomic interactions from high accurate QM methods, like

MP2/cc-PVTZ[42,43], and large model systems by periodic QDFT[44,45], less independent of traditional experimental data, and obtain the system thermodynamic properties with high accuracy, such as gas adsorption[46,47] and diffusion[48]. However, most of the first principle force fields are still limited to simple cases, due to the limitation of computer power, which make experimental data is still a valuable reference.

The classic force fields have their own disadvantages: the simple potential functions can't describe neither polarizable effect in condense phase nor chemical bonds breaking cases. So beyond fixed atomic charge models, polarizable force fields are introduced to consider the environmental effects, by fluctuating charge[49,50], drude oscillator[51,52] and induced dipole models[42]. Polarizable force fields often leads to more accurate results compared to traditional ones, however, due to the iteration procedure in the charge calculation part, it is usually much slower than classic ones, also the highly environmental dependence often means the parameters could not be transferred to other cases easily. To describe the chemical reactions, people develop reactive force fields by introducing bond order terms, such as ReaxFF[53] and REBO[54], have been applied to many interesting systems, such as metal catalyzed reactions[55]. The problem here is before using reaction force fields to study certain systems, we must have some pre-knowledge of the reaction path ways, in order to derive proper force field parameter sets. To accelerate the studying of larger complex systems, like protein and polymer, coarse-grained force fields are developed, like MS-CG[56] and MARTINI force field[57]. Since they ignore many atomic details, for structure sensitive problems, the accuracy should be carefully checked.

The results of simulations will be realistic only if the potential energy functions mimic the 'real' forces experienced by the 'real' atoms. On the other hand force fields should have a functional form as simple as possible to speed up the evaluation of the forces. Ideally empirical potentials should also be transferable and applicable to many other interested systems under different conditions.

II. Molecular Dynamics

After obtained the proper force fields, there are usually two ways to study the target system in classical simulation community. One is molecular dynamics (MD), which means the time evolution of a set of interacting particles in the system is following the solution of Newton's equations of motion:

$$\mathbf{F}_i = m_i \frac{d^2\mathbf{r}_i(t)}{dt^2} \tag{8}$$

Where $\mathbf{r}_i(t) = (x_i(t), \mathrm{y}_i(t), \mathrm{z}_i(t))$ is the position vector of *ith* particle and $\mathbf{F}_\mathrm{i}$ is the force acting upon *ith* particle at time t and $\mathrm{m_i}$ is the mass of the particle. Particles usually refer to atoms, but they can represent any species that can be described by certain interaction law, like chemical functional groups in coarse grain force fields. To integrate the above second-order differential equations, the instantaneous forces acting on the particles and their initial positions and velocities need to be specified. The MD trajectories are defined by both particle position and velocity vectors to describe the time evolution of the system in phase space. The positions and velocities are propagated with a finite time interval using numerical integrators, the most common one is the Velocity-Verlet algorithm[58].

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)(\Delta t)^2 \tag{9}$$

$$\mathbf{v}(t+\Delta t) = \mathbf{v}(t) + \frac{1}{2}(\mathbf{a}(t+\Delta t) + \mathbf{a}(t))\Delta t \tag{10}$$

The position of each particle in space is defined by **r**(t), whereas the velocities **v**(t) determines the kinetic energy and temperature in the system at certain time t. From the point of view of statistical mechanics, MD is merely a method of conformational sampling that yields average structural and thermodynamic properties from trajectories. Within infinite simulation time, such averages would converge to the true value of the ensemble averaged thermodynamic properties. However, in reality, the quality of sampling and the accuracy of the force fields used are always limited. The quality of sampling may be very poor, especially for processes with time scale larger than the MD simulations.

III. Monte Carlo Method

Another important classical simulation methods is Monte Carlo (MC) method,[59] first proposed at 1940s,[60] even before MD was invented. The main idea of the MC algorithm is a heuristic prescription for a plausible pattern of changes in the configurations assumed by the system. By carefully designed MC moves, we could generate points in configuration space with a relative probability proportional to desired ensemble distribution.[58] In the standard Metropolis MC algorithm, a move is accepted unconditionally if the new configuration results in a lower potential energy. Otherwise it is accepted with a probability given by the Boltzmann factor,

$$P_{\mathbf{r}\to\mathbf{r}'} \propto \exp(-\frac{\Delta U}{k_B T}) \tag{11}$$

where ΔU denotes the change in the potential energy associated with configuration change by a move $\mathbf{r} \to \mathbf{r}'$ and k_B is the Boltzmann constant. Due to no kinetic concept included,

Monte Carlo method is more convenient than molecular dynamics to be applied to many cases only concerning equilibrium thermodynamic properties, such as gas adsorption, by giving reasonable probability descriptions. However, finding a proper move that would ensure efficient sampling is not a nontrivial problem, due to the generated improper configurations. To solve this problem, configuration-biased Monte Carlo (CBMC) method[61] are introduced by an enhanced probability to 'fit' into the existing configuration to accept or reject moves. A recently developed more efficient MC method is Wang-Landau sampling method.[62] Instead of sampling the probability distribution at a fixed temperature, a random walk is performed in energy space to extract an estimate for the density of states. Then the probability can be computed at any temperature by weighting the density of states by the appropriate Boltzmann factor.

In 1949, Metropolis uses MC simulation to study nuclear physics,[60] later, in 1957, Berni Alder first applies MD simulation to study the phase transition for a hard sphere system,[63] then in 1970s, Levitt, Warshel and Karplus introduce MD to biological field [64,65]. More and more people realize the importance of molecular simulation methods. Computer simulations bring new insights into mechanisms and processes that are not directly accessible through experiment, meanwhile it is faster and less expensive than synthesizing and characterizing the target in real experiments. One big corner stone is Levitt, Warshel and Karplus win the Noble Prize at 2013, as a worldwide recognition of the new era of molecular modelling.

Despite of the many advantages of classic simulation methods, till today the simulated time and space scale is still one of the most severe problems in the community,

for example the real protein folding process may take minutes while current MD simulations still have difficulty to reach microsecond level.[66]. Although MC method doesn't have simulation time concern, but is still limited by the system size, due to the explicitly consideration of large amount of particle-particle interactions, even with more advanced large scale parallel computing, depending on CPU[67] or GPU[68]. Different from classical simulations, classical density functional theory obtain the system equilibrium thermodynamic properties by solving series of math equations instead of brutal forces calculations over tens of thousands particles, which is much more efficient, more details will be given in the following chapters.

1.1.3. Hydration Free Energy

Solvation free energy is one of the most important thermodynamic properties in chemical engineering field, since most of the chemical reactions and processes happen in the solvent environment, which are all related to it, for example why certain nanomaterials could only be synthesized in specific solvent and how to control its morphology by changing the solution conditions. Another direct example is aqueous solubility, which is a major concern in the pharmaceutical development process and in the identification and quantification of potential environmental contaminants, relating to a wide range of applications. The traditional way by doing experiments is very expensive and time consuming, considering over thousands of candidates, so accurate and efficient prediction of solvation free energy are desired in lots of industrial fields.

There have been enormous activities in this field recent years in understanding the microscopic details of solute-solvent interactions and the effects of the local solvent

structure on the chemical and biochemical affinities of dissolved species[69-83]. Generally speaking, current computational methods to investigate the solvent effects could be classified into three categories: continuous approaches[84-89], molecular simulations[90-94], and liquid-state theories[95-101]. A number of hybrid methods also exist by various combinations of these three basic procedures[102-107].

A continuous approach is considering the solvent as a continues isotropic medium, and then the solvent-solute interactions could be described by a number of operators, such as the free energy of cavitation, electrostatic energy, dispersion energy, repulsion energy, thermal fluctuation contribution and so on[108-111]. Due to no explicit solvent is considered during the calculation, this kind of phenomenological approaches are very computationally efficient and widely applied in drug design field to screening promising candidates[112-114], however since lots of semi-empirical parameters are needed to better capture the trend of the initial fitting training sets, it's always doubtful of their reliability beyond the original thermodynamic conditions, that's why usually they could be only applied to similar compounds prediction compared to the original ones[115-117].

Molecular simulation with explicit solvent consideration has become a routine tool and is usually recognized as the most reliable and realistic model for studying the solvent effects over the past few decades[118]. However this kind of calculation intrinsically requires a large number of degrees of freedom, associating with high phase space dimensionality. It makes the free-energy calculation relies on alchemical methods to sample the microstates of the system along various thermodynamic pathways[119,120], instead of one time simple calculation as for the molecular mechanics properties. Due to the high

computational demanding, simulation results compared to experimental data based on a large library of organic compounds have been rarely reported, even for small molecules. An early notable exception is the MC data for the solvation properties of over 200 organic solutes in aqueous and organic solutions reported by Duff and Jorgensen[121]. More recently, Mobley *et al*. [122] calculated the hydration free energies of over 500 small organic molecules using alchemical MD simulations, which is the largest simulation database currently, however compared to the real drug candidates databank containing millions of compounds, such kind of scale and computation cost are still not acceptable.

Modern liquid state theory provides a compromise between conventional semi-empirical methods and molecular simulations. Different from molecular simulations, it allows us to do one time calculation to obtain both thermodynamic properties and the equilibrium configuration without explicit enumeration of microstates. Among them, we think classical density functional theory (CDFT) could be a strong contender as a promising computationally viable hydration free energy prediction method. Since it is based on minimization of the grand potential instead of sampling millions of microstates by brute force calculations. In many cases, the theoretical performance is comparable to molecular simulations but with much less computational burden. More details will be given in the following chapter 3.

II. Gas adsorption in nanoporous materials

Due to the increasing air pollution and global warming, more importantly limited storage amount of fossil fuel, in the last three decades, more and more researches have focused on clean and renewable energy development, for example hydrogen and methane.

However, the energy density of such kind of uncompressed gases energy is too low for practical purpose compared to gasoline. People need to find better gas storage methods, besides traditionally gas tank technology, recent years, great effort has been contributed to the discovery of novel nanoporous materials, in particular metal-organic framework materials (MOFs) as high performance candidates[123]. Unlike conventional adsorbents such as activated carbons or zeolites, MOFs are organic-inorganic hybrid solids with infinite, uniform framework structures built from organic linkers and inorganic metal (or metal containing clusters) nodes that are able to be self-organized into periodic and porous frameworks[124]. The intrinsic nature of MOFs allows us to systematic modulating the pore dimensions, surface areas, topology structures for specific target, based on the vast combinations of available metal ions and organic linkers, unfortunately, in reality, the number of currently synthesized MOF materials is negligible comparing to the total number of possibilities. New efficient computational screening methods for the adsorption properties could be great helpful for experimental scientists.

Huge amount of classical simulation works have been done to studying gas adsorption in nanoporous materials.[46,125-131] Recently many researchers turn to computer-aided materials design, in order to find new promising materials, even ahead of experimental scientists. For example, Han *et al*. combined quantum mechanics the GCMC methods to design the new Li-doped MOFs with much higher hydrogen adsorption amount than undoped ones,[132] indicating that doping with metal ions may be an effective way to enhance the adsorption capability. A more prominent work is done by Wilmer *et al*. to find promising MOFs for methane storage purpose, they first design over 100,000

hypothetical MOFs based on available building blocks and then carry out GCMC calculations screening for the best candidates, also synthesize them.[133] The same simulation method was also used for screening of MOFs toward high-efficacy CO_2 capture[134]. However, in those works, due to the computational demanding with molecular simulation methods, complete simulation calculations for the whole materials database are impossible, even with most advanced computer technology, such as parallel and stream processor computing, many compromises need to be made.

One the other hand, the classical density functional theory (CDFT) has been long recognized as a valuable alternative to simulation methods for gas adsorption in porous materials.[135,136] In fact, CDFT is now the standard method used in physisorption field for the materials characterization of porous materials, integrated in all the experimental measurement equipment.[137,138] As indicated above, CDFT calculations are based on the minimization of the grand potential, making its computational efficiency far superior to molecular simulations. For gas adsorption prediction, common versions of the density functionals typically use the fundamental measure theory (FMT) to account for molecular excluded volume effects,[139-142] and the attractive components are often represented by various forms of weighted density approximations (WDA),[143,144] the van-der-Waals-like mean-field approximation (MFA),[145,146] the density functional expansion method [140,147], and various modifications of these methods.[47,48] But the shortcomings of CDFT calculations in this field have also been well recognized: the predictions are often sensitive to the approximations for the free-energy functionals. Even though all of those density functionals had been calibrated with simulation results for LJ fluids in simple geometries,

they're never tested in real nanoporous materials, except few cases in MOF-5. So for practical purpose, the performance of the free energy functionals must be validated with a large training set of realistic materials over a broad range of conditions pertinent to industrial required conditions, which I will talk about in chapter 4.

1.2 Dissertation organization

This dissertation is focused on the application of density functional theory for high-throughput screening of chemicals and materials, especially at common organic molecular hydration free energy and gas properties in nanoporous materials.

Chapter 2 covers the basic formulism of the classical density functional theory, mainly focus on the construction of excess Helmhotz energy functional, special attention is given to modified fundamental measure theory part.

Chapter 3 focuses on applying classical density functional theory to predict the gas behavior in nanoporous materials. First, four version of common CDFT methods are chosen to examine their performance on hydrogen storage from simple geometry model systems to real nanoporous materials, validating whether it depends on the gas species and examination conditions. Second, based on the original excess entropy scaling relation by Rosenfeld, a new scaling equation for simple fluids in nanoporous materials is built, by taking the contribution from Knudsen diffusion region and free volume effect into consideration. Third, expand the four methods for methane storage purpose, trying to find the key property that could lead to large CH_4 delivery amount to achieve the ARPA-E target. The fourth part is applying quantum density functional theory and Monter Carlo methods to design series of new nitrogen-doped porous aromatic framework (PAF)

materials, to see their performance on CO_2 capture and CO_2/N_2 separation, also explore the storage controlling factors in different conditions.

Chapter 4 focuses on applying molecular density functional theory (MDFT) for large-scale prediction of hydration free energy of organic compounds. First I discuss the results of applying the new developed theory to the 500 solvation database, to valid its accuracy and efficiency compared to experimental and full atomistic simulation data, with standard molecular structure and force field. Then testing the performance of several common generic force fields and molecular structure generation methods for the prediction of small organic molecular drugs hydration free energy in the Sampl4 blind test. Third, based on the extensive 700 organic compounds solvation database, various combination of input structure generation methods, force fields, charge sets and water models are examined to find the generally best choice to apply our MDFT for predicting the water solvation free energy of an unknown organic compound.

Finally Chapter 5 summarizes key conclusions from this dissertation and perspectives for future work.

References:

(1) Roothaan, C. C. J. *Reviews of Modern Physics* **1951**, *23*, 69.

(2) Møller, C.; Plesset, M. S. *Physical Review* **1934**, *46*, 618.

(3) Čárský, P.; Paldus, J.; Pittner, J. *Recent progress in coupled cluster methods: theory and applications*; Springer, 2010; Vol. 11.

(4) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *The Journal of chemical physics* **1987**, *87*, 5968.

(5) Hohenberg, P.; Kohn, W. *Physical Review* **1964**, *136*, B864.

(6) Becke, A. D. *The Journal of Chemical Physics* **2014**, *140*.

(7) Kohn, W.; Sham, L. J. *Physical Review* **1965**, *137*, A1697.

(8) Kohn, W.; Sham, L. J. *Physical Review* **1965**, *140*, A1133.

(9) Becke, A. D. *Physical Review A* **1988**, *38*, 3098.

(10) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Physical Review Letters* **1996**, *77*, 3865.

(11) Becke, A. D. *The Journal of Chemical Physics* **1993**, *98*, 1372.

(12) Niu, S.; Hall, M. B. *Chemical reviews* **2000**, *100*, 353.

(13) Perdew, J. P.; Ruzsinszky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E.; Csonka, G. I. *The Journal of Chemical Physics* **2005**, *123*.

(14) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Physical Review Letters* **2003**, *91*, 146401.

(15) Zhao, Y.; Truhlar, D. *Theor Chem Account* **2008**, *120*, 215.

(16) Grimme, S. *The Journal of Chemical Physics* **2006**, *124*.

(17) Goerigk, L.; Grimme, S. *Journal of Chemical Theory and Computation* **2010**, *7*, 291.

(18) Goerigk, L.; Grimme, S. *Physical Chemistry Chemical Physics* **2011**, *13*, 6670.

(19) Yanai, T.; Tew, D. P.; Handy, N. C. *Chemical Physics Letters* **2004**, *393*, 51.

(20) Chai, J.-D.; Head-Gordon, M. *Physical Chemistry Chemical Physics* **2008**, *10*, 6615.
(21) Thanthiriwatte, K. S.; Hohenstein, E. G.; Burns, L. A.; Sherrill, C. D. *Journal of Chemical Theory and Computation* **2010**, *7*, 88.

(22) Dewar, M. J.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. *Journal of the American Chemical Society* **1985**, *107*, 3902.

(23) Stewart, J. *J. Comp. Chem* **1991**, *12*, 320.

(24) Korth, M. *Journal of Chemical Theory and Computation* **2010**, *6*, 3808.

(25) Stewart, J. P. *J Mol Model* **2013**, *19*, 1.

(26) Muddana, H. S.; Gilson, M. K. *Journal of Chemical Theory and Computation* **2012**, *8*, 2023.

(27) Allinger, N. L. *Journal of the American Chemical Society* **1977**, *99*, 8127.

(28) Hagler, A.; Huler, E.; Lifson, S. *Journal of the American Chemical Society* **1974**, *96*, 5319.

(29) Schrier, E. E.; Pottle, M.; Scheraga, H. A. *Journal of the American Chemical Society* **1964**, *86*, 3444.

(30) Lifson, S.; Warshel, A. *The Journal of Chemical Physics* **1968**, *49*, 5116.

(31) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *Journal of the American Chemical Society* **1995**, *117*, 5179.

(32) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J Phys Chem-Us* **1990**, *94*, 8897.

(33) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J Am Chem Soc* **1992**, *114*, 10024.

(34) Düren, T.; Sarkisov, L.; Yaghi, O. M.; Snurr, R. Q. *Langmuir* **2004**, *20*, 2683.

(35) Walton, K. S.; Snurr, R. Q. *J Am Chem Soc* **2007**, *129*, 8552.

(36) Challa, S. R.; Sholl, D. S.; Johnson, J. K. *The Journal of Chemical Physics* **2002**, *116*, 814.

(37) Keskin, S.; Liu, J.; Rankin, R. B.; Johnson, J. K.; Sholl, D. S. *Industrial & Engineering Chemistry Research* **2009**, *48*, 2355.

(38) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187.

(39) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *Journal of the American Chemical Society* **1996**, *118*, 11225.
(40) Sun, H.; Mumby, S. J.; Maple, J. R.; Hagler, A. T. *Journal of the American Chemical Society* **1994**, *116*, 2978.

(41) Sun, H. *The Journal of Physical Chemistry B* **1998**, *102*, 7338.

(42) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio Jr, R. A. *The journal of physical chemistry B* **2010**, *114*, 2549.

(43) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. *J Comput Chem* **2003**, *24*, 1999.

(44) Cygan, R. T.; Liang, J.-J.; Kalinichev, A. G. *The Journal of Physical Chemistry B* **2004**, *108*, 1255.

(45) Fang, H.; Kamakoti, P.; Zang, J.; Cundy, S.; Paur, C.; Ravikovitch, P. I.; Sholl, D. S. *The Journal of Physical Chemistry C* **2012**, *116*, 10692.

(46) Fu, J.; Sun, H. *The Journal of Physical Chemistry C* **2009**, *113*, 21815.

(47) Lin, L.-C.; Lee, K.; Gagliardi, L.; Neaton, J. B.; Smit, B. *Journal of Chemical Theory and Computation* **2014**, *10*, 1477.

(48) Amirjalayer, S.; Tafipolsky, M.; Schmid, R. *Angewandte Chemie International Edition* **2007**, *46*, 463.

(49) Rick, S. W.; Stuart, S. J.; Berne, B. J. *The Journal of chemical physics* **1994**, *101*, 6141.

(50) Patel, S.; Mackerell, A. D.; Brooks, C. L. *Journal of computational chemistry* **2004**, *25*, 1504.

(51) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D. *Journal of Chemical Theory and Computation* **2005**, *1*, 153.

(52) Lopes, P. E.; Lamoureux, G.; Roux, B.; MacKerell, A. D. *The Journal of Physical Chemistry B* **2007**, *111*, 2873.

(53) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. *The Journal of Physical Chemistry A* **2001**, *105*, 9396.

(54) Brenner, D. W. *Physical Review B* **1990**, *42*, 9458.

(55) Nielson, K. D.; van Duin, A. C.; Oxgaard, J.; Deng, W.-Q.; Goddard, W. A. *The Journal of Physical Chemistry A* **2005**, *109*, 493.

(56) Izvekov, S.; Voth, G. A. *The Journal of Physical Chemistry B* **2005**, *109*, 2469.
(57) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *The Journal of Physical Chemistry B* **2007**, *111*, 7812.

(58) Frenkel, D.; Smit, B. *Understanding molecular simulation: from algorithms to applications*; Academic press, 2001; Vol. 1.

(59) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *The Journal of Chemical Physics* **1953**, *21*, 1087.

(60) Metropolis, N.; Ulam, S. *Journal of the American Statistical Association* **1949**, *44*, 335.

(61) Martin, M. G.; Siepmann, J. I. *The Journal of Physical Chemistry B* **1999**, *103*, 4508.

(62) Landau, D.; Tsai, S.-H.; Exler, M. *American journal of physics* **2004**, *72*, 1294.

(63) Alder, B. J.; Wainwright, T. E. *The Journal of Chemical Physics* **1957**, *27*, 1208.

(64) Karplus, M.; Weaver, D. L. *Nature* **1976**, *260*, 404.

(65) Warshel, A.; Levitt, M. *Journal of molecular biology* **1976**, *103*, 227.

(66) Petrenko, R.; Meller, J. In *eLS*; John Wiley & Sons, Ltd: 2001.

(67) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. *Journal of chemical theory and computation* **2008**, *4*, 435.

(68) Anderson, J. A.; Lorenz, C. D.; Travesset, A. *Journal of Computational Physics* **2008**, *227*, 5342.

(69) Chipot, C.; Pohorille, A. *Free energy calculations : theory and applications in chemistry and biology*; Study ed.; Springer: New York, 2007.

(70) Cramer, C. J.; Truhlar, D. G. *Accounts of Chemical Research* **2008**, *41*, 760.

(71) Zou, X. Q.; Sun, Y. X.; Kuntz, I. D. *Journal of the American Chemical Society* **1999**, *121*, 8033.

(72)Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. *Biochemistry* **1981**, *20*, 849.

(73) Palmer, D. S.; Sergiievskyi, V. P.; Jensen, F.; Fedorov, M. V. *Journal of Chemical Physics* **2010**, *133*, 044104.

(74) Chandler, D. *Nature* **2005**, *437*, 640.

(75) Hirata, F. *Molecular theory of solvation*; Kluwer Academic Publishers: Boston, 2003.

(76) Chang, J.; Lenhoff, A. M.; Sandler, S. I. *Journal of Physical Chemistry B* **2007**, *111*, 2098.

(77) Freedman, H.; Le, L.; Tuszynski, J. A.; Truong, T. N. *Journal of Physical Chemistry B* **2008**, *112*, 2340.

(78) Freedman, H.; Truong, T. N. *Chemical Physics Letters* **2003**, *381*, 362.

(79) Paluch, A. S.; Shah, J. K.; Maginn, E. J. *Journal of Chemical Theory and Computation* **2011**, *7*, 1394.

(80) Jorgensen, W. L.; Ravimohan, C. *Journal of Chemical Physics* **1985**, *83*, 3050.

(81) Marcus, Y. *Chem Rev* **2009**, *109*, 1346.

(82) Ball, P. *Chem Rev* **2008**, *108*, 74.

(83) Senn, H. M.; Thiel, W. *QM/MM Methods for Biological Systems*; Springer-Verlag: Berlin Heidelberg, 2007; Vol. 268.

(84) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L. *Journal of Computational Chemistry* **2004**, *25*, 265.

(85) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2007**, *3*, 2011.

(86) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2008**, *4*, 877.

(87) Tomasi, J. *Theor Chem Acc* **2004**, *112*, 184.

(88) Roux, B.; Simonson, T. *Biophys Chem* **1999**, *78*, 1.

(89) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem Rev* **2005**, *105*, 2999.

(90) Duffy, E. M.; Jorgensen, W. L. *Journal of the American Chemical Society* **2000**, *122*, 2878.

(91) Jang, S.; Kim, E.; Pak, Y. *Proteins-Structure Function and Bioinformatics* **2006**, *62*, 663.

(92) Shirts, M. R.; Pande, V. S. *Journal of Chemical Physics* **2005**, *122*, 134508.

(93) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *Journal of Chemical Physics* **2003**, *119*, 5740.

(94) Straatsma, T. P.; Berendsen, H. J. C. *Journal of Chemical Physics* **1988**, *89*, 5876.

(95) *Molecular Theory of Solvation* Hirata, F., Ed.; Kluwer Academic Publishers: Dordrecht, 2003.

(96) Gray, C. G.; Gubbins, K. E. *Theory of Molecular Fluids: Fundamentals* Oxford University Press: New York, 1985; Vol. 1.

(97) Chandler, D. *Annual Review of Physical Chemistry* **1978**, *29*, 441.

(98) Lue, L.; Blankschtein, D. *J Phys Chem-Us* **1992**, *96*, 8582.

(99) Lue, L.; Blankschtein, D. *Journal of Chemical Physics* **1995**, *102*, 5427.

(100) Rossky, P. J. *Annual Review of Physical Chemistry* **1985**, *36*, 321.

(101) Brooks, C. L.; Karplus, M.; Pettitt, B. M. *Proteins : a theoretical perspective of dynamics, structure, and thermodynamics*; J. Wiley: New York, 1988.

(102) Pratt, L. R.; Pohorille, A. *Chem Rev* **2002**, *102*, 2671.

(103) Lee, M. S.; Salsbury, F. R.; Olson, M. A. *Journal of Computational Chemistry* **2004**, *25*, 1967.

(104) Basdevant, N.; Ha-Duong, T.; Borgis, D. *Journal of Chemical Theory and Computation* **2006**, *2*, 1646.

(105) Bryk, P.; MacDowell, L. G. *Journal of Chemical Physics* **2011**, *135*.

(106) Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. *Journal of Chemical Theory and Computation* **2006**, *2*, 420.

(107) Schmidt, K. F.; Kast, S. M. *Journal of Physical Chemistry B* **2002**, *106*, 6289.

(108) Guo, Z. J.; Li, B.; Dzubiella, J.; Cheng, L. T.; McCammon, J. A.; Che, J. W. *Journal of Chemical Theory and Computation* **2013**, *9*, 1778.

(109) Cramer, C. J.; Truhlar, D. G. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers, New York: 1995; Vol. 6, p 1.

(110) Sandberg, L.; Edholm, O. *Journal of Chemical Physics* **2002**, *116*, 2936.

(111) Jha, A. K.; Freed, K. F. *Journal of Chemical Physics* **2008**, *128*, 034501.

(112) Klamt, A.; Schuurmann, G. *Journal of the Chemical Society, Perkin Transactions 2* **1993**, 799.

(113) Phillips, K. L.; Sandler, S. I.; Greene, R. W.; Di Toro, D. M. *Environmental science & technology* **2008**, *42*, 8412.
(114) Klamt, A.; Eckert, F.; Diedenhofen, M. *Journal of Physical Chemistry B* **2009**, *113*, 4508.

(115) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. *Journal of Chemical Physics* **2002**, *117*, 43.

(116) Boyer, R. D.; Bryan, R. L. *The Journal of Physical Chemistry B* **2012**, *116*, 3772.

(117) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *The Journal of Physical Chemistry B* **2009**, *113*, 4538.

(118) Baron, R.; McCammon, J. A. *Annual Review of Physical Chemistry* **2013**, *64*, 151.

(119) Kollman, P. *Chemical Reviews* **1993**, *93*, 2395.

(120) Mobley, D. L.; Klimovich, P. V. *The Journal of chemical physics* **2012**, *137*, 230901.

(121) Duffy, E. M.; Jorgensen, W. L. *Journal of the American Chemical Society* **2000**, *122*, 2878.

(122) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. *Journal of Chemical Theory and Computation* **2009**, *5*, 350.

(123) Snurr, R. Q. *J Phys Chem Lett* **2011**, *2*, 1842.

(124) Ockwig, N. W.; Delgado-Friedrichs, O.; O'Keeffe, M.; Yaghi, O. M. *Accounts Chem Res* **2005**, *38*, 176.

(125) Nalaparaju, A.; Jiang, J. *Journal of Physical Chemistry C* **2012**, *116*, 6925.

(126) Atci, E.; Erucar, I.; Keskin, S. *J. Phys. Chem. C* **2011**, *115*, 6833.

(127) Guo, H.-c.; Shi, F.; Ma, Z.-f.; Liu, X.-q. *Journal of Physical Chemistry C* **2010**, *114*, 12158.

(128) Keskin, S. *Journal of Physical Chemistry C* **2011**, *115*, 800.

(129) Fairen-Jimenez, D.; Moggach, S. A.; Wharmby, M. T.; Wright, P. A.; Parsons, S.; Dueren, T. *Journal of the American Chemical Society* **2011**, *133*, 8900.

(130) Gallo, M.; Glossman-Mitnik, D. *J. Phys. Chem. C* **2009**, *113*, 6634.

(131) Cao, D. P.; Lan, J. H.; Wang, W. C.; Smit, B. *Angew Chem Int Edit* **2009**, *48*, 4730.

(132) Han, S. S.; Goddard, W. A. *Journal of the American Chemical Society* **2007**, *129*, 8422.

(133) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nat Chem* **2012**, *4*, 83.

(134) Babarao, R.; Jiang, J. W. *Langmuir* **2008**, *24*, 6270.

(135) Ravikovitch, P. I.; Domhnaill, S. C. O.; Neimark, A. V.; Schueth, F.; Unger, K. K. *Langmuir* **1995**, *11*, 4765.

(136) Maddox, M. W.; Olivier, J. P.; Gubbins, K. E. *Langmuir* **1997**, *13*, 1737.

(137) Neimark, A. V.; Ravikovitch, P. I. *Microporous and Mesoporous Materials* **2001**, *44–45*, 697.

(138) Occelli, M. L.; Olivier, J. P.; Petre, A.; Auroux, A. *The Journal of Physical Chemistry B* **2003**, *107*, 4128.

(139) Rosenfeld, Y. *Phys. Rev. Lett.* **1989**, *63*, 980.

(140) Rosenfeld, Y. *Journal of Chemical Physics* **1993**, *98*, 8126.

(141) Yu, Y.-X.; Wu, J. *J. Chem. Phys.* **2002**, *117*, 10156.

(142) Roth, R.; Evans, R.; Lang, A.; Kahl, G. *Journal of Physics: Condensed Matter* **2002**, *14*, 12063.

(143) Tarazona, P.; Evans, R. *Molec. Phys.* **1984**, *52*, 847.

(144) Patra, C. N.; Ghosh, S. K. *Phys. Rev. E* **1993**, *48*, 1154.

(145) Jain, S.; Jog, P.; Weinhold, J.; Srivastava, R.; Chapman, W. G. *J. Chem. Phys.* **2008**, *128*, 154910.

(146) Ravikovitch, P. I.; Vishnyakov, A.; Neimark, A. V. *Phys. Rev. E* **2001**, *64*, 011602.

(147) Zhou, S. Q. *Int. J. Mod. Phys. B* **2005**, *19*, 4701.

Chapter 2. Basic Concepts of Classical Density Functioal Theory

In this chapter we gives a brief introduction of classical density functional theory (CDFT). It is significantly more effective in quantifying the thermodynamic and microscopic structural properties of complex fluid system, relying on variational computation for the equilibrium density profiles and the corresponding thermodynamic properties. In essence, DFT provides quantitative relationships between thermodynamic properties and microscopic structure, with the mathematical procedure naturally applicable to complex molecular systems with intermolecular interactions spanning multiple time and length scales. However, CDFT requires various approximations for the free-energy functionals in terms of the one-body density profiles underlying the microscopic structure of molecular systems.

2.1. Density Profile

Density functional theory expresses the properties of a multi-body system in terms of the density profiles, i.e., the ensemble-averaged spatial-temporal distributions of individual elements (particles). Intuitively, the molecular density profiles represent the average spatial distributions of molecules. To illustrate this mathematical concept, consider a monatomic system with N identical spherical particles. At a given configuration of the multi-body system, the instantaneous particle density at position $\mathbf{r}$ is a summation of the Dirac delta-functions:

$$\hat{\rho}(\mathbf{r}) = \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i), \tag{12}$$

where $\delta(\mathbf{r})$ stands for the 3-dimensional Dirac delta function

$$\delta(\mathbf{r}-\mathbf{r}')=\begin{cases}0, & \mathbf{r}\neq\mathbf{r}'\\ \infty, & \mathbf{r}=\mathbf{r}'\end{cases}. \tag{13}$$

The Dirac function can be understood as a probability density to find the particle at position $\mathbf{r}$; it satisfies the normalization condition

$$\int d\mathbf{r}\,\delta(\mathbf{r}-\mathbf{r}')=1. \tag{14}$$

The density profile is defined as an ensemble average of the instantaneous local density [1]

$$\rho(\mathbf{r})=\hat{\rho}(\mathbf{r})=\left\langle\sum_{i=1}^{N}\delta(\mathbf{r}-\mathbf{r}_i)\right\rangle. \tag{15}$$

The density profile of a multi-body system is intimately related to its thermodynamic properties. For a one-component system with N identical spherical particles considered above, we may write the grand partition function Ξ as

$$\Xi=\sum_{N}\frac{1}{N!\Lambda^{3N}}\int d\mathbf{r}^{N}\exp\left\{-\beta\left[\Phi(\mathbf{r}^{N})+N\mu+\sum_{i=1}^{N}\psi(\mathbf{r}_i)\right]\right\} \tag{16}$$

where Λ stands for the thermal wavelength; $\mathbf{r}^N$ is a short notation for $\mathbf{r}^N=(\mathbf{r}_1,\mathbf{r}_2,\cdots\mathbf{r}_N)$, *i.e.,* the configuration of all particles in the system; $\Phi(\mathbf{r}^N)$ represents the total interaction potential among these particles; μ is the chemical potential; $\psi(\mathbf{r})$ is the one-body external potential, $\beta=1/(k_BT)$, and k_B is the Boltzmann constant. The one-body profile $\rho(\mathbf{r})$ is given by

$$\rho(\mathbf{r})=\frac{1}{\Xi}\sum_{N}\frac{1}{N!\Lambda^{3N}}\int d\mathbf{r}^{N}\sum_{i=1}^{N}\delta(\mathbf{r}_i-\mathbf{r})\exp\left\{-\beta\left[\Phi(\mathbf{r}^{N})+N\mu+\sum_{i=1}^{N}\psi(\mathbf{r}_i)\right]\right\}. \tag{17}$$

We can rewrite Eq. (17) in terms of a functional derivative of Ξ with respect to the one-body external potential

$$\rho(\mathbf{r}) = -\frac{1}{\beta\Xi}\frac{\delta\Xi}{\delta\phi(\mathbf{r})} = -\frac{\delta\ln\Xi}{\beta\delta\phi(\mathbf{r})}. \tag{18}$$

Because the grand potential Ω is defined as

$$\beta\Omega = -\ln\Xi. \tag{19}$$

Substituting Eq. (19) into Eq. (18) yields a simple relation between the one-body density profile and the grand potential

$$\rho(\mathbf{r}) = \frac{\delta\Omega}{\delta\phi(\mathbf{r})}. \tag{20}$$

2. 2 Density Functional Theory

The basic ideas of DFT can be traced back to van der Waals' pioneering work published in the late 19th century that concerned with the surface tension of a vapor-liquid interface. Functional minimization of the free energy was utilized as a criterion of equilibrium for inhomogeneous systems[2]. Modern developments in classical DFT is, however, based on the Hohenberg-Kohn (HK) theorem[3], which states that, in an equilibrium system at a given temperature and the chemical potentials of individual species, the one-body external potential for each species can be uniquely determined by the corresponding one-body density profile. The HK theorem enables the definition of intrinsic Helmholtz free energy, a thermodynamic quantity for inhomogeneous systems that is independent of the one-body external potentials[4]. For a one-component system, we can write the intrinsic Helmholtz energy F as

$$F \circ A - \grave{O} d\mathbf{r} j\left(\mathbf{r}\right) r\left(\mathbf{r}\right), \tag{21}$$

where A denotes the conventional Helmholtz energy. The conventional and intrinsic Helmholtz energies are related to the grand potential

$$\Omega = A - N\mu = F + \int d\mathbf{r}\varphi(\mathbf{r})\rho(\mathbf{r}) - N\mu = F + \int d\mathbf{r}\left[\varphi(\mathbf{r}) - \mu\right]\rho(\mathbf{r}). \tag{22}$$

Because the one-body external potential for each species can be uniquely determined by the corresponding one-body density profile, the intrinsic Helmholtz free energy and subsequently the grand potential are functionals of the one-body density profiles. According to the HK theorem, the equilibrium density profiles minimize the grand potential, i.e., for a one-component system,

$$\frac{dW\left[r\left(\mathbf{r}\right)\right]}{dr\left(\mathbf{r}\right)} = 0. \tag{23}$$

Substituting Eq. (22) into Eq. (23) yields

$$\frac{dF\left[r\left(\mathbf{r}\right)\right]}{dr\left(\mathbf{r}\right)} = m - j\left(\mathbf{r}\right). \tag{24}$$

Given an analytical expression for the intrinsic Helmholtz energy, the HK theorem thus provides a mathematical framework for solving the microscopic structure and corresponding thermodynamic properties, *viz.*, from the functional derivative of the grand potential with respect to the one-body density profiles.

2. 3 Intrinsic Helmholtz Energy Functional

For simple fluid model systems considered in this thesis, the intrinsic Helmholtz energy can be decomposed into two parts: the ideal-gas term F^{id} and the excess term F^{ex}

$$F\left[\rho(\mathbf{r})\right] = F^{id}\left[\rho(\mathbf{r})\right] + F^{ex}\left[\rho(\mathbf{r})\right]. \tag{25}$$

The ideal-gas term can be obtained exactly:

$$\beta F^{id} = \int\left[\ln\rho(\mathbf{r}) - 1\right]\rho(\mathbf{r})\,d\mathbf{r} \quad . \tag{26}$$

The excess intrinsic Helmholtz energy accounts for the thermodynamic non-ideality due to inter- and intra-molecular interactions. Unlike its ideal counterpart, the exact representation for the excess free energy is unknown. Indeed, attaining an accurate formulation of the excess intrinsic Helmholtz energy is critical to all DFT calculations.

2. 4 Euler-Lagrange Equation

For a simple fluid system at equilibrium, the density profiles can be determined from minimization of the grand potential (Eq. (23))

$$\frac{\delta\Omega}{\delta\rho(\mathbf{r})} = 0 \tag{27}$$

The functional derivatives lead to an Euler-Lagrange equation:

$$\rho(\mathbf{r}) = \exp\left[\beta\mu - \beta\Psi(\mathbf{r}) - \frac{\delta\beta F^{ex}}{\delta\rho(\mathbf{r})}\right] \tag{28}$$

The functional derivative of the excess Helmholtz energy with respect to the density profile of the simple fluid gives the local excess chemical potential:

$$\mu^{ex}(\mathbf{r}) = \frac{\delta\beta F^{ex}}{\delta\rho(\mathbf{r})}. \tag{29}$$

2. 5 Excess Helmholtz Free Energy

Now we move on to define the excess Helmholtz free energy F^{ex}. Formulation of this quantity is the key in DFT calculations. For simple fluid considered in this dissertation, the excess Helmholtz free energy can be decomposed into two inter-related contributions:

$$F^{ex} = F_{hs}^{ex} + F_{attr}^{ex} . \tag{30}$$

In Eq.(30), the subscripts indicate contributions to the excess Helmholtz energy due to the hard-sphere repulsion (F_{hs}^{ex}) and the attraction energy part (F_{attr}^{ex}). Each contribution to the thermodynamic potential can be formulated based on a combination of physical insights and mathematic analysis. In the following, we will only focus on the simple fluid system. While in chapter 4 we will focus on more complicated polyatomic system. The mathematical details for derivations of these equations can be found from the literature.

Considering simple LJ fluid system as in Chapter 3, according to the Barker-Henderson (BH) theory of bulk fluids [5]

$$u^{\mathrm{LJ}}(r) = u^{\mathrm{hs}}(r) + u^{\mathrm{attr}}(r) \tag{31}$$

$$u^{\mathrm{hs}}(r) = \begin{cases} \infty & r < d \\ 0 & r > d \end{cases} \tag{32}$$

$$u^{\mathrm{attr}}(r) = \begin{cases} 0 & r < \sigma \\ 4\varepsilon\left[\left(\dfrac{\sigma}{r}\right)^{12} - \left(\dfrac{\sigma}{r}\right)^{6}\right] & r > \sigma \end{cases} \tag{33}$$

where *d* is the hard-sphere diameter[6]

$$d = s\left(\frac{1 + 0.2977T^{*}}{1 + 0.33163T^{*} + 0.0010477T^{*2}}\right) \tag{34}$$

where $T^* = k_B T / \varepsilon$.

For all DFT functionals considered for simple fluid system, the hard-sphere term, F^{hs}, is calculated from the modified fundamental measure theory (MFMT) [7,8]

$$F^{hs} = k_B T \int \Phi^{hs}[n_\alpha(\mathbf{r})] d\mathbf{r} \tag{35}$$

where

$$\begin{aligned} \Phi^{hs} &= -n_0 \ln(1-n_3) + \frac{n_1 n_2 - \mathbf{n}_{V1} \cdot \mathbf{n}_{V2}}{1-n_3} \\ &+ \frac{n_3 + (1-n_3)^2 \ln(1-n_3)}{36\pi n_3^2 (1-n_3)^2} (n_2^3 - 3 n_2 \mathbf{n}_{V2} \cdot \mathbf{n}_{V2}) \end{aligned} \tag{36}$$

$$n_\alpha(\mathbf{r}) = \int \rho(\mathbf{r}') w^{(\alpha)}(|\mathbf{r} - \mathbf{r}'|) d\mathbf{r}' \quad \alpha \in \{0,1,2,3,V1,V2\} \tag{37}$$

and $w^{(\alpha)}$ are the weighting functions related to the hard-sphere geometry[9]

$$\begin{cases} w^{(2)}(r) = \pi d^2 w^{(0)}(r) = 2\pi d w^{(1)}(r) = \delta(d/2 - r) \\ w^{(3)}(r) = \theta(d/2 - r) \\ \mathbf{w}^{(V2)}(\mathbf{r}) = 2\pi d \mathbf{w}^{(V1)}(\mathbf{r}) = w^{(2)}(r) \dfrac{\mathbf{r}}{r} \end{cases} \tag{38}$$

where $\delta(r)$ represents the Dirac delta function, and $\theta(r)$ is the Heaviside step function. In comparison with simulation results, MFMT is extremely accurate for describing the structural and thermodynamic properties of inhomogeneous hard-sphere systems[10].

For the attractive part of the Helmholtz energy functional, we consider the following four different approximations. These functionals have been commonly used in simple fluid systems.

2.5.1. Mean-field approximation (MFA)

The attractive component of the excess Helmholtz energy functional is given by

$$F^{MFA}[\rho(\mathbf{r})] = \frac{1}{2}\iint \rho(\mathbf{r})\rho(\mathbf{r}')u^{\text{attr}}(\mathbf{r}-\mathbf{r}')\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' \tag{39}$$

Like the van der Waals equation of state, MFA ignores correlated distribution of gas molecules inside the material. It has been used since the early application of DFT[11] and remains relevant in recent DFT calculations, in particular for materials characterization. With an accurate equation of state for bulk fluids, the mean-field approximation performs surprisingly well under many circumstances. Because of its simplicity, MFA has been a preferred choice for theoretical investigations of surface phase transitions.[12]

2.5.2. Weight-density approximations

In this version of DFT proposed by Yu[13], here designated as WDA-Y, the attractive part of the excess Helmholtz energy includes a mean-field contribution and an additional term due to correlations:

$$F^{\text{attr}}[\rho(\mathbf{r})] = F^{MFA}[\rho(\mathbf{r})] + F^{\text{cor}}[\rho(\mathbf{r})] \tag{40}$$

The mean-field contribution, F^{MFA} , is the same as that given in Eq.(39), and F^{cor} is defined in terms of a local correlation free energy at an empirically defined weighted density:

$$F^{\text{cor}} = k_{\text{B}}T\int \Phi^{\text{cor}}[\bar{\rho}(\mathbf{r})]\mathrm{d}\mathbf{r} \tag{41}$$

where

$$\bar{\rho}(\mathbf{r}) = \frac{3}{4\pi\sigma^3}\int \rho(\mathbf{r}')\theta(d-|\mathbf{r}-\mathbf{r}'|)\mathrm{d}\mathbf{r}' \tag{42}$$

In Eq.(41), Φ^{cor} is the free energy per volume for the bulk system due to the correlation effect. It can be obtained from

$$\Phi^{\mathrm{cor}}(\rho) = \frac{F_{\mathrm{bulk}}^{\mathrm{LJ}}(\rho) - F_{\mathrm{bulk}}^{\mathrm{hs}}(\rho) - F_{\mathrm{bulk}}^{\mathrm{MFA}}(\rho)}{V} \tag{43}$$

where V is the system volume, $F_{\mathrm{bulk}}^{\mathrm{LJ}}(r)$ is the excess free energy for the bulk LJ fluid at system temperature T and density r, $F_{\mathrm{bulk}}^{\mathrm{hs}}(\rho)$ is the excess free energy for bulk hard-sphere fluid, and $F_{\mathrm{bulk}}^{\mathrm{MFA}}(\rho)$ is the mean-field energy of the bulk system

$$\frac{F_{\mathrm{bulk}}^{\mathrm{MFA}}(\rho)}{V} = -\frac{16}{9}\pi\beta\varepsilon\rho^2\sigma^3 \tag{44}$$

Yu used the analytical results from the first-order-mean-spherical approximation (FMSA) [13,14] for $F_{\mathrm{bulk}}^{\mathrm{LJ}}(\rho)$, and $F_{\mathrm{bulk}}^{\mathrm{hs}}(\rho)$ is calculated from the Carnahan-Starling equation of state.[15] Considering that FMSA is not very accurate for a LJ fluid near the critical condition, in this work we use the modified Benedict-Webb-Rubin (MBWR) equation of state[16] instead of FMSA to calculate $F_{\mathrm{bulk}}^{\mathrm{LJ}}(\rho)$.

Alternatively, we may apply WDA to the correlation part of the excess Helmholtz energy functional per molecule (rather than per volume)

$$F^{cor} = k_B T \int \rho(\mathbf{r}) f^{cor}[\bar{\rho}(\mathbf{r})] d\mathbf{r} \tag{45}$$

This version of free energy functional is designated as WDA-L.

2.5.3. Density expansion method

In this method, we still separate the excess free energy into two parts, F^{hs} and F_{att}^{ex}. The functional expansion is applied only to the attractive part

$$\beta F_{att}^{ex}[\rho(\mathbf{r})] = \beta F_{att}^{ex}[\rho_b] - \int d\mathbf{r} C_{att}^{(1)}(\rho(\mathbf{r}) - \rho_b) - \frac{1}{2}\int\int d\mathbf{r} d\mathbf{r}' C_{att}^{(2)}(|\mathbf{r} - \mathbf{r}'|)(\rho(\mathbf{r}) - \rho_b)(\rho(\mathbf{r}') - \rho_b) \quad (46)$$

where the direct correlation functions (DCF) for the bulk systems are defined as

$$C_{att}^{(1)} = -\partial F_{att}^{ex} / \partial \rho(\mathbf{r})|_b \quad (47)$$

$$C_{att}^{(2)}(|\mathbf{r} - \mathbf{r}'|) = -\partial^2 F_{att}^{ex} / \partial \rho(\mathbf{r}) \partial \rho(\mathbf{r}')|_b \quad (48)$$

Tang proved that the Lennard-Jones potential can be accurately reproduced by a two-Yukawa function,[17] which leads to an analytical expression for $C_{att}^{(1)}$ and $C_{att}^{(2)}(r)$ based on the first-order mean-spherical approximation (FMSA).[18] For convenience, this version of DFT is referred as FMSA. The detail expressions for $C_{att}^{(1)}$ and $C_{att}^{(2)}(r)$ are given elsewhere.[19]

References:

(1) Wu, J. *Molecular Thermodynamics of Complex Systems*; Springer: Berlin, 2009; Vol. 131.

(2) Rowlinson, J. S. *Journal of Statistical Physics* **1979**, *20*, 197.

(3) Hohenberg, P.; Kohn, W. *Physical Review* **1964**, *136*, B864.

(4) Wu, J.; Li, Z. *Annual Review of Physical Chemistry* **2007**, *58*, 85.

(5) Barker, J. A.; Henderso.D *J Chem Phys* **1967**, *47*, 4714.

(6) Cotterman, R. L.; Schwarz, B. J.; Prausnitz, J. M. *AIChE Journal* **1986**, *32*, 1787.

(7) Yu, Y.-X.; Wu, J. *J. Chem. Phys.* **2002**, *117*, 10156.

(8) Roth, R.; Evans, R.; Lang, A.; Kahl, G. *Journal of Physics: Condensed Matter* **2002**, *14*, 12063.

(9) Rosenfeld, Y. *Phys. Rev. Lett.* **1989**, *63*, 980.

(10) Roth, R. *J Phys-Condens Mat* **2010**, *22*.

(11) Tarazona, P.; Evans, R. *Molec. Phys.* **1984**, *52*, 847.

(12) Evans, R.; Henderson, J. R.; Roth, R. *J Chem Phys* **2004**, *121*, 12074.

(13) Yu, Y.-X. *Journal of Chemical Physics* **2009**, *131*, 024704.

(14) Tang, Y.; Lu, B. C. Y. *AIChE Journal* **1997**, *43*, 2215.

(15) Carnahan, N. F.; Starling, K. E. *J. Chem. Phys.* **1969**, *51*, 635.

(16) Johnson, J. K.; John, A. Z.; Keith, E. G. D.-L. D. *Mol. Phys.* **1992**, *78*, 591.

(17) Tang, Y.; Tong, Z.; Lu, B. C. Y. *Fluid Phase Equilibria* **1997**, *134*, 21.

(18) Tang, Y. P. *Journal of Chemical Physics* **2003**, *118*, 4140.

(19) Tang, Y. P.; Wu, J. Z. *Physical Review E* **2004**, *70*, 011201.

Chapter 3. Computer-Aided Materials Design

3.1 Density functional methods for fast screening of metal-organic frameworks for hydrogen storage

Abstract

Classical density functional theory (DFT) has been routinely used in characterization of pore size distributions and specific surface areas of porous materials by physisorption. However, its application to large-scale screening of materials for gas storage has been rather limited because it is commonly believed that the DFT calculations are highly sensitive to the approximations for the free-energy functionals. In this work, we have investigated four representative versions of non-local density functionals for predicting H_2 adsorption in slit pore model and in a large library of metal-organic frameworks (MOFs) under a broad range of temperatures and pressures. The four versions of DFT share a common functional from the modified fundamental measure theory that accounts for the molecular excluded volume effects, while differ in their approximations to represent the intermolecular attractions, *viz*., mean-field approximation, two versions of weighted-density approximations (WDA), and the quadratic functional expansion method. We have tested these functionals with Monte Carlo simulation data for H_2 adsorption at conditions of practical interest. Overall all four versions of DFT are reasonably accurate in comparison with the simulation results. While the density expansion method performs rather well at the DOE target condition for hydrogen storage, the WDA methods are found most accurate at the low-temperature condition typically used in materials characterization. In addition to adsorption isotherms, DFT is able to generate density profiles revealing

microscopic details such as favorable adsorption sites. From a computational perspective, the DFT calculation is at least one order of magnitude faster than conventional simulation methods, rendering it as a promising tool for large-scale screening of nanostructured materials for gas storage.

3.1.1. Introduction

Over the past decades, metal-organic frameworks (MOFs) have attracted much research attention for gas storage purpose.[1-16] Compared to traditional nanoporous materials, MOF materials have well defined crystalline structures with extremely high specific surface areas and superior gas adsorption capabilities. Moreover, the MOF structures can be easily tuned and customized to target specific applications of practical concern. For example, NU-100 [17] and MOF-210 [18] have been identified as two promising MOF materials for H_2 storage, by computational design with particular high surface areas. Because the vast combinations of available metal ions and organic linkers, the number of currently synthesized MOF materials is negligible (less than 1%) comparing to the total number of possibilities. Further experimental exploration of MOFs for gas storage will benefit tremendously from computational screening of the adsorption properties.

At present theoretical modeling of gas adsorption in porous materials is mostly based on molecular simulations.[5-7,13,14,16,19,20] For example, Snurr and coworkers used the grand canonical Monte Carlo (GCMC) simulation to predict hydrogen adsorption in a series of MOFs.[21] It was found that the gas adsorption amount is strongly correlated with the heat of adsorption, the surface area, and the free volume of MOFs at low, intermediate and high pressures, respectively. Han *et al*. used the GCMC method, along with a quantum-

mechanics derived force field, to predict the hydrogen adsorption behavior in Li-doped MOFs.[22] They indicated that doping with metal ions provides an effective way to enhance the adsorption capability. A similar procedure was deployed by Yang *et al*. to study the effect of open metal sites on gas adsorption.[23] Wilmer *et al*. carried large-scale GCMC calculations for over 100,000 MOFs for methane adsorption.[24] The simulation method was also used for screening of MOFs toward high-efficacy CO_2 storage[25]. Despite their wide usage, the molecular simulation methods are still time consuming, even with more advanced technology, such as modern parallel computing.

The classical density functional theory (DFT) has been long recognized as a valuable alternative to simulation methods for gas adsorption in porous materials.[26,27] On the one hand, DFT is now routinely used in physisorption for the characterization of pore size distributions and specifying surface areas of porous materials.[28,29] On the other hand, DFT has been emerging as one of the most versatile theoretical tools in statistical mechanics for describing the interfacial behavior of a broad range of inhomogeneous fluids including gas, liquid and polymeric systems.[30-32] Because DFT calculations are based on the minimization of the grand potential, its computational efficiency is far superior to molecular simulations, in particular for calculating thermodynamic properties such as free energy. Nevertheless, the shortcomings of DFT calculations have also been well recognized: the DFT predictions are often sensitive to the approximations for the free-energy functionals. For gas adsorption, common versions of the density functionals typically use the fundamental measure theory (FMT) to account for molecular excluded volume effects,[33-36] and the attractive components are often represented by various forms

of weighted density approximations (WDA),[37,38] the van-der-Waals-like mean-field approximation (MFA),[39,40] the density functional expansion method [34,41], and different modifications of these methods.[47-48] Most notably, Neimark and coworkers have successfully used MFA to study gas adsorptions and phase transitions of confined fluids in various slit and cylindrical nanopores.[15, 26, 28, 40, 42] They found that in many cases the DFT predictions were in good agreements with Monte Carlo simulation results.[42] Tang *et al.* constructed a free-energy functional using the functional expansion method with the bulk direct correlation function obtained from the first-order mean-spherical approximation (FMSA).[43] The non-mean-field theory was proved to be computationally as efficient as MFA but more accurate in comparison with simulation results for a variety of inhomogeneous Lennard-Jones fluids. While early DFT calculations were mostly concerned with model fluids confined in simple geometries,[40,44-46] applications to more realistic systems have also been considered in recent years.[47-52] For example, Siderius *et al*. successfully used a lattice version of MFA to study H_2 adsorption in MOF-5 at room temperature.[47] As indicated by Liu *et al*.,[49] MFA is not accurate at low temperature, in particular for fluids in contact with a weakly attractive surface. The theoretical results can be much improved by using WDA for both the hard-sphere and attractive parts of the excess free energy functional. The generalized WDA by Liu *et al.* performs well for both low and high temperatures and has been extended to mixtures.[51] Along similar lines, Yu proposed that the WDA methods can be used for both the hard-sphere and correlation parts of the excess free energy functional.[53] To a certain degree, the generalized WDA

approximation introduced by Yu resembles the exchange-correlation functional in electronic DFT.

The density functionals mentioned above had been calibrated with simulation results for either LJ fluids in simple geometries or adsorption of a specific gas in MOF-5. For materials screening, the numerical performance of the free energy functionals should be validated with a large training set of realistic materials over a broad range of conditions pertinent to industrial applications. In this work, we aim to create such a benchmark by carrying out extensive GCMC simulation for H_2 adsorption in a large library of MOF materials and examine the theoretical performances for four typical versions of classical DFT. All these functionals are based on the modified fundamental measure theory to account for molecular excluded volume effects[36,54] but differ in the attraction part of excess free-energy functional, *viz*. MFA, two forms of WDA, and the functional expansion method.

3.1.2. Molecular Model and Methods

Gas adsorption in MOF materials depends on, in addition to the bulk temperature and pressure, interactions between the adsorbed fluid and the host framework. In general, we need to consider both bonded and non-bonded interactions among the MOF atoms, MOF-gas, and gas-gas molecules. In this study, we conjecture that the MOF structures are rigid and unchanged upon gas adsorption. Where structure changes are not uncommon for MOF materials,[55] it has been well recognized, the framework flexibility mainly affect gas transport and has relatively small effect on adsorption isotherms.

In our molecular model, gas molecules can depicted as spherical particles. Inside the MOF materials, each gas molecule is subject to an external field, $V^{ext}(\mathbf{r})$, arising from its interaction with all the fixed framework atoms

$$V^{ext}(\mathbf{r}) = \sum_{i \in \text{MOF}} u_{if}(\mathbf{r} - \mathbf{r}_i) \tag{49}$$

where subscript i represents the ith atom of the MOF framework, $u_{if}(\mathbf{r})$ represents its interaction potential with the gas molecule, $\mathbf{r}$ and $\mathbf{r}_i$ stand for the positions of the gas molecule and the ith MOF atom. The Lennard-Jones (LJ) model is used to describe both the MOF-gas and gas-gas potentials:

$$u_{ij}(r) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right] \tag{50}$$

where parameters ε_{ii} and σ_{ii} are obtained from standard semi-empirical force fields and the Lorentz-Berthelot (LB) mixing rule is employed for different atomic types.

The DFT calculations are based on the minimization of the grand potential functional

$$\Omega[\rho(\mathbf{r})] = F[\rho(\mathbf{r})] + \int [V^{ext}(\mathbf{r}) - \mu]\rho(\mathbf{r})\,d\mathbf{r} \tag{51}$$

where μ denotes the gas chemical potential in the bulk, $F[\rho(\mathbf{r})]$ is the Helmholtz energy functional, and ρ (**r**) is the density profile of gas molecules. The gas chemical potential can be obtained from the equation of state for the corresponding bulk fluid

$$\mu = \mu_{\text{bulk}}(P,T) \tag{52}$$

where P and T represent pressure and temperature, respectively. For all DFT calculations reported in this work, the chemical potential of the bulk system are calculated from the

MBWR equation of state,[56] while the Helmholtz energy functional are constructed following the method in Chapter 2.

We minimize the grand potential functional with respect to the 3-dimensional density profiles of gas molecules by using the conjugate-gradient method (CG Descent).[57] From $r(\mathbf{r})$, we calculate the adsorption amount based on the number of gas molecules in the MOF material and system volume:

$$\Gamma = \frac{k_B T_0}{P_0 V} \int \rho(\mathbf{r}) \mathrm{d}\mathbf{r} \tag{53}$$

where $P_0 = 1\,\mathrm{atm}$ and $T_0 = 25\ ^{\circ}\mathrm{C}$ stand for the standard state (STP). As usual, the adsorption amount is expressed as the volume of gas adsorbed at the standard-state condition in unit volume of material.

To validate the DFT results, we have also calculated H_2 adsorption in slit pores and in a large number of MOF materials using grand canonical Monte Carlo (GCMC) simulation method.[58] In all our simulations, the LJ interactions were evaluated with a spherical cutoff of 12.9 Å. Each simulation run consists of 10^6 trial moves to reach the equilibrium state, and another 10^6 MC moves for calculating ensemble average. Three types of trial moves were attempted in the GCMC simulations: insertion, deletion and displacement of gas molecules. Chemical potentials applied in the simulation works were also calculated from the MBWR equation of state as discussed above. The simulation data for each trajectory were divided into 10 blocks in order to estimate the statistical uncertainties. Unless specifically mentioned, the statistical uncertainty was generally

smaller than the symbol sizes presented in the figures. All MC simulations were carried out with the Towhee 7.0.4 program.[59]

In both DFT and GCMC predictions of H_2 adsorption in MOFs, we used a 2x2x2 supercell and rigid framework for the porous crystal. The LJ parameters for the MOF frameworks were taken from the UFF force field.[60] The molecular model been proved before to be accurate for predicting gas adsorption in MOFs.[24] H_2 molecules are described with the widely used Buch model, which gives ε_{H2}/k_B=36.7 K and σ_{H2}=2.96 Å.[61] The Buch model successfully reproduces the gas properties in the bulk.

3.1.3. Results and Discussion

The four versions of DFT methods discussed above have been calibrated before with GCMC results for LJ fluids in simple pore geometries.[42,43,62] However, the early comparisons are mostly focused on the DFT performance from a theoretical perspective, in particular on its applicability to describing the interfacial behavior of confined liquids. Toward that end, it has been well known that MFA is rather unsatisfactory for any liquid near the triple point and only those functional taking into account the correlation effects are able to attain quantitative results. The situation is quite different though for gas adsorption because the thermodynamic conditions of practical concern are rarely close to the triple point. In most cases, adsorption occurs way beyond the critical temperature such that the intermolecular attraction has little influence on the fluid structure. If the local density inside the pore is not significantly different from that in the bulk, the density profile is primarily determined by the external potential, little affected by the excess Helmholtz

energy. In other words, different DFT methods should yield similar results if the same bulk density and the external potential are used in the calculations.

Figure 3-1 shows the absolute amount of H_2 adsorption in three slit pores with slightly different widths. In DFT and MC calculations, the Steele's 10-4-3 potential is used to represent the interaction between gas molecules and each slit wall:

$$V_{sf}(z) = \varepsilon_w \left[\frac{2}{5}\left(\frac{\sigma_w}{z}\right)^{10} - \left(\frac{\sigma_w}{z}\right)^{4} - \frac{\sigma_w^4}{3\Delta(0.61\Delta + z)^3} \right] \tag{54}$$

where z is the perpendicular distance from the surface, $\sigma_w = \sigma$, $\Delta = 0.7071\sigma$, and $\varepsilon_w = 6.283\varepsilon$. Because the local density of gas molecules inside the slit pore is not drastically different from the bulk density, the adsorption isotherms from different versions of DFT are virtually identical and all agree well with the GCMC data. In other words, the insensitivity of gas adsorption to thermodynamic non-ideality is mainly due to the cancellation of errors. Different DFT functionals differ only in the local excess chemical potential relative to that of the bulk value. Such difference diminishes if the local density becomes identical to the bulk density.

Application of DFT to gas adsorption in realistic materials is more challenging than that for simple pores not only due to the computational problems affiliated with solving 3-dimensional density profiles but also due to much more pronounced local density inhomogeneity. Consider, as an example, H_2 adsorption in MOF-5 at 77K. Figure 3-2 presents the framework structure as well as the local density profiles of gas molecules predicted from DFT (WDA-L). At low pressure (P=1 bar, Figure 2b), the gas molecules in MOF-5 are mainly concentrated around the strong binding sites (*i.e.* metal clusters). The

local gas density, $\rho(\mathbf{r})$, can be higher than the bulk density, ρ_b, by over three orders of magnitude (shown as the yellow iso-surface in Figure 2b). As the bulk pressure increases, adsorption of gas molecules around the organic linkers becomes more significant (Figure 2c). At sufficiently high pressure (~20 bar), all favorable binding sites are saturated by surface adsorption and a further increase in pressure results in more gas molecules filling in the pore. At this point, the free volume takes in charge of the gas adsorption and increasing the pressure reduces the local density inhomogeneity (Figure 2d). The strong inhomogeneity of local density profile predicted by DFT is consistent with previous simulation[63] and experimental results[64]. From practical perspectives, the 3-dimensional density profile is useful not only for generating the adsorption isotherms but also for revealing the microscopic structure of adsorbed guest molecules in nanoporous materials. It contains essential information required for understanding specific binding, diffusion and chemical reactions.

Figure 3-3 presents a comparison of typical adsorption isotherms predicted from GCMC simulation with those from four versions of DFT methods discussed above. All DFT methods show excellent agreement with simulations for hydrogen adsorption in MOF-5 at high (298 K) and intermediate (243 K) temperatures. The difference is noticeable only for the MFA version of DFT, which slightly underestimates the adsorption amount at high pressure (>300 bar). At low temperature (77K), however, the differences between GCMC and different DFT methods become more distinctive. As expected, MFA yields the poorest results in comparison with the simulation data: the deviation from GCMC simulation is more than 10 percent at high pressure (>20 bar). The poor performance of

MFA is consistent with earlier calibrations for LJ fluids near weakly attractive surfaces.[43] Because the correlation effect magnifies at low temperature and high density, it appears that MFA is not an ideal tool for predicting adsorption amount at high loadings. Most surprisingly, the functional expansion method performs poorly as well even at intermediate pressure, which is in stark contrast to earlier validations for simple pores.[43] For this case, the error is probably introduced by the FMSA equation for representing fluid properties at supercritical temperatures and in the application of density expansion for highly inhomogeneous 3-D systems. Conversely, the two versions of WDA show superior performance over the entire ranges of temperatures and pressures examined in this work. We have also tested the two versions of WDA in conjunction with both FMSA and MBWR equations of state. We find that the accuracy of equation of state plays an important role in DFT predictions. For LJ systems, the MBWR equation, obtained by fitting extensive simulation data, is more accurate than FMSA in particular at supercritical conditions. The WDA versions of DFT with FMSA equation, show a much higher RMSD compared to MBWR one with GCMC simulation results and clearly overestimation trend for MOF hydrogen adsorption capability as shown in Figure 4e.

A major objective of this work is to test the numerical performance of DFT for gas adsorption in a large library of MOF materials. Toward that end, we adopt the Northwestern hypothetic MOF database.[24] We picked up 1,200 MOF materials from the library based on the following four characteristic properties: 1) excess CH_4 adsorption in weight (cm^3 (STP)/g), 2) excess CH_4 adsorption in volume (cm^3 (STP)/cm^3), 3) void fraction, and 4) surface area (m^2/cm^3). For each category, we use the top 300 MOFs as

examples (in descending order). We assume that these MOFs consist of a good training set for testing the performance of different DFT methods for large-scale screening of porous materials for H_2 storage.

For potential industrial applications, we consider the thermodynamic condition according to the new H_2 storage target set by the U.S. Department of Energy (DOE) for 2015,[65] *viz*., 5.5 wt% and 40 g L^{-1} H_2 (444 cm^3 (STP)/cm^3) at T = 243K and P = 100 bar. Figure 3-4 compares the results from GCMC simulation and various DFT methods for H_2 adsorption in 1,200 MOFs selected according to the criteria discussed above. Here materials selected accordingly different criteria are coded in different colors. Specifically, blue represents top 300 MOFs in the database based on the excess amount of CH_4 adsorption in weight, red for top 300 MOFs in terms of the volume, purple for top 300 MOFs in void fraction, and teal for top 300 in surface area. It's interesting to see the effects of void fraction and surface area to the H_2 adsorption at this condition. Figure 3-4 shows that materials with highest void fractions give the lowest adsorption amounts for hydrogen storage. Materials in the highest surface areas and the largest excess amount of CH_4 adsorption in weight categories gives higher H_2 adsorption amounts in volume than those with large void fractions, while the latter are better. MOFs with the largest volume adsorption capability for CH_4 are also the best for H_2. Regrettably, none of these materials meet the DOE 2015 volume adsorption target. The highest one among 1,200 MOFs considered in this work is below 250 cm^3(STP)/cm^3, still remote from the 440 cm^3(STP)/cm^3. It appears that stronger guest-framework interactions are imperative to reach the DOE target.

Overall, all DFT predictions agree well with the simulation results for most materials. In particular, FMSA yields the best performance, with the root-mean-square deviation (RMSD) as low as 2.35 cm^3 (STP)/cm^3. This number is remarkable from both practical and theoretical perspectives considering the large target storage capacity (e.g., ~400 cm^3 (STP)/cm^3) and approximations affiliated with the molecular model, e.g. force fields. Surprisingly, MFA performs almost equally as well as FMSA, with RMSD of 2.54 cm^3 (STP)/cm^3, which is even better than those from WDA methods, 6.02 and 3.16 cm^3 (STP)/cm^3 for WDA-Y and WDA-L, respectively. The main reason that FMSA and MFA perform well is that, at this condition, the reduced temperature, $T^* = k_B T / e = 6.60$, is much higher than the reduced critical temperature of the bulk LJ fluid (T*=1.31), and the local density inside the pores are relatively uniform. As discussed before, the intermolecular attraction makes little contributions to correlation effects at high temperature and the density expansion method works well for weakly inhomogeneous systems.

Table 3-1 presents the rankings of top 10 MOFs for H_2 adsorption at the DOE target condition according to different computational procedures. All versions of DFT rank No.36 as the best candidate and predict similar adsorption amount. While the predictions for the top material are in excellent agreement with the GCMC results, the specific ranking varies with the computational methods. For example, No.11 is ranked as 10th candidate and No. 37 as 15th according to GCMC, but all DFT methods rank No.37 instead of No. 11 as a top 10 candidate. Considering 3% statistical uncertainty in GCMC simulations and 1%

difference between No.37 and No.11 for the adsorption amount, we maintain that all DFT methods are sufficiently accurate for identifying best candidates for H_2 storage.

Since experimental characterization of new materials is mainly based on gas adsorption at low temperature (77 K) and atmospheric pressure (1 bar), we have also tested the performances of the four versions of DFT at this condition. In this case, the excess Helmholtz energy functional plays a more important role in gas adsorption and, because of strong density inhomogeneity inside the pores, the correlation effects introduced by intermolecular attractions are non-negligible even at low bulk pressure. Figure 3-5 shows that the differences among the four versions of DFT are much larger in comparison to those at the DOE target condition (Figure 3-4). At low temperature, the two WDA methods are slightly better than FMSA or MFA in comparison with the simulation data. Overall WDA-Y gives the best RMSD, 6.87 cm^3 (STP)/cm^3. While WDA-L and MFA result in slightly larger deviations, with RMSD=8.38 and 11.3 cm^3(STP)/cm^3, respectively. The worst performance by FMSA, with RMSD=27.00 cm^3(STP)/cm^3, was rather unexpected. Apparently, the density expansion method is not ideal for systems with strong density inhomogeneity. By comparing the results shown in Figures 3-4 and 3-5, we find that the adsorption at low temperature alter the materials ranking according, since more binding sites are occupied at this condition along with higher H_2 loadings, besides those most favorite ones acting on high temperatures.

Figure 3-6 shows a further comparison of DFT and GCMC calculations at 77K but higher pressure (50 bar). Similar to the low-pressure case (1 bar), the two WDA methods yield the best agreement with GCMC simulation, with WDA-L slightly better than WDA-

Y. Clearly, both FMSA and MFA are unsatisfactory at low temperature, both significantly underestimating the simulation results. At low temperature and high pressure, materials with the best weight or volume CH_4 adsorption have similar capability for H_2 storage, implying that ranking according to the adsorption in weight provides a better criterion here.

For large-scale materials screening, it is important to have computational methods that are both reliable and fast. In general, different DFT methods do not show significant difference in computational speed, usually much faster than molecular simulations. To further reduce the computational cost, we have employed the Fast Fourier Transformation (FFT) techniques for evaluating the convolutions such as

$$\int f_1(\mathbf{r}) f_2(\mathbf{r}-\mathbf{r}')\mathrm{d}\mathbf{r}' = \mathcal{F}^{-1}\{\mathcal{F}[f_1(\mathbf{r})]\mathcal{F}[f_2(\mathbf{r})]\} \tag{55}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ represent the forward and backward FFT, respectively. In this work, all FFT calculations were based on the FFTW package.[66] Figure 3-7 shows the distribution of computational time for typical DFT calculations on a single desktop PC with one Intel E-1230 CPU core. For the DOE target condition at 243 K and 100 bar, the average time for each MOF material is less than 15 seconds. Such a speed is about two orders of magnitude faster than conventional simulation methods. Even at situation of extremely high gas loading (e.g. 77 K and 50 bar), the average DFT time is still less than 2 minutes per MOF material.

3.1.4. Conclusions

We have calibrated four representative versions of classical density functional theory (DFT) that all use the fundamental measure theory but differ in the attractive part of the Helmholtz energy functional for their potential applications to large-scale screening

of nanostructured porous materials for hydrogen storage. These functionals have been tested before with simulation data for Lennard-Jones (LJ) fluids in simple confined geometries but often with an emphasis on their theoretical performance near liquid conditions. Because gas adsorption typically occurs at high temperature much above the critical point, the thermodynamic properties are less sensitive to the correlation effects due to intermolecular attractions. As a result, the different versions of DFT yield similar results in comparison with grand canonical Monte Carlo (GCMC) simulation for LJ fluids in slit pores. Nevertheless, significant differences are noticeable for their application to realistic pore materials because of the highly inhomogeneous distributions of gas molecules. By extensive comparison of DFT predictions and simulation data for H_2 adsorption in 1,200 MOF materials, we find that the density expansion method (FMSA) performs best under the DOE target condition for H_2 storage. For the purpose of materials screening, it seems that other versions of DFT are also sufficiently accurate at the DOE condition. At the condition typically used in experimental characterization of porous materials (77K and 1 bar), however, both FMSA and mean-field approximations are unsatisfactory. Much more reliable predictions can be achieved with two versions of weighted density approximation (WDA). For all four versions of DFT considered in this work, the computation time is less than 2 minutes for each material, typically on the order of seconds. The computational efficiency makes DFT an ideal tool for large scale screening of promising nanostructured materials for gas storage purpose.

Figure 3-1. H_2 adsorption in slit pores at 243K with reduced width H/s =2,3,4 from top to bottom calculated from GCMC and four different versions of DFT for the attraction part of the excess Helmholtz energy functional, *i.e.*, WDA-Y[53], WDA-L [51], FMSA, and MFA. Γ^* and ρ_b^* means H_2 adsorption amount and bulk density in dimensionless unit.

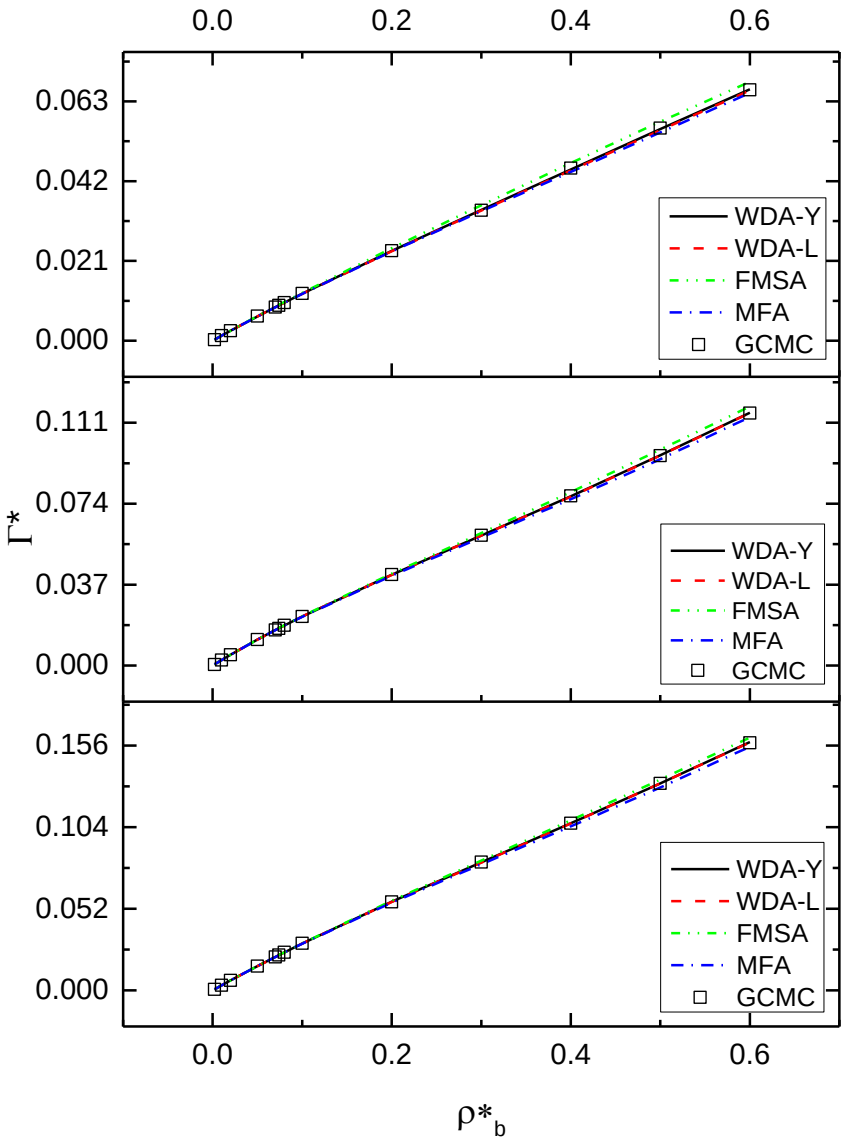


Figure 3-2. Density profile of H_2 molecules adsorbed in MOF-5 at 77K and various pressures according to WDA-L: (a) structure of MOF-5; (b) P=1 bar; (c) P=20 bar; (d) P= 90 bar. Color code: (a) red: Zn, blue: O, purple: C, white: H; (b) purple: $r(\mathbf{r})/r_b$=1, translucent blue: $r(\mathbf{r})/r_b$ = 100, yellow: $r(\mathbf{r})/r_b$ =1000; (c) purple: $r(\mathbf{r})/r_b$ =1, translucent blue: $r(\mathbf{r})/r_b$ =50, yellow: $r(\mathbf{r})/r_b$ =200; (d) purple: $r(\mathbf{r})/r_b$ =0.2, translucent blue: $r(\mathbf{r})/r_b$=5, yellow: $r(\mathbf{r})/r_b$= 40.

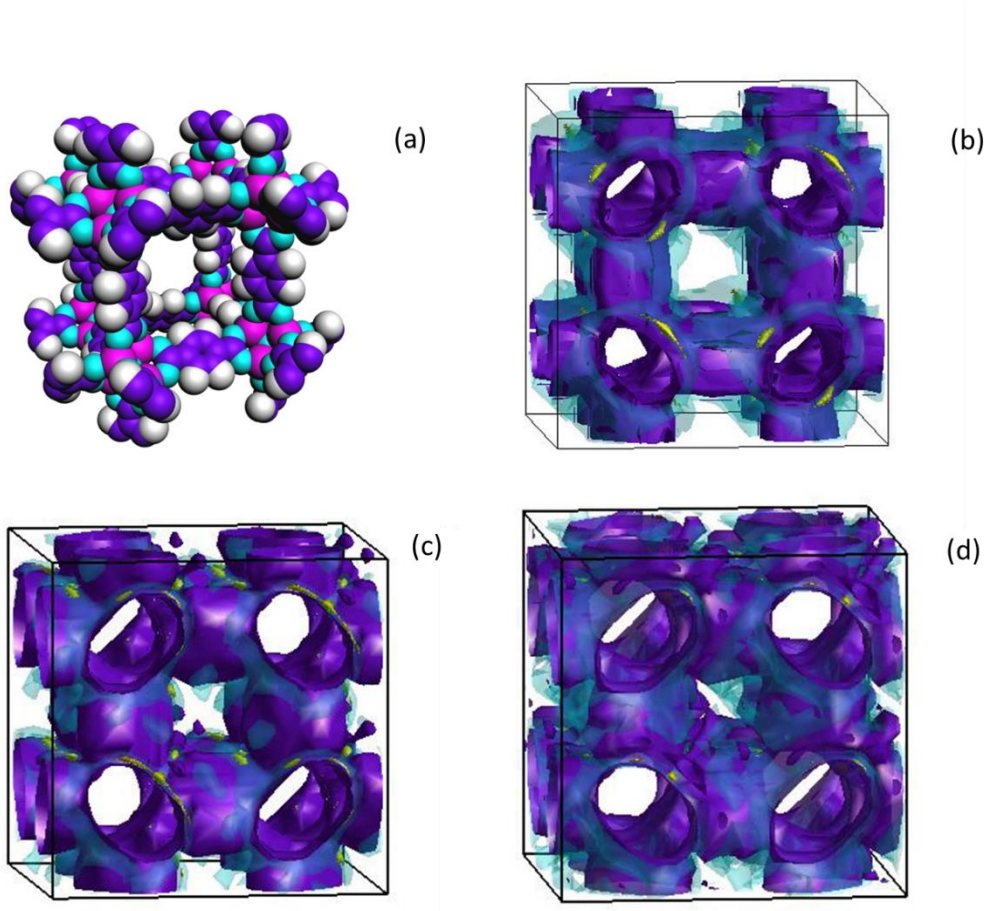
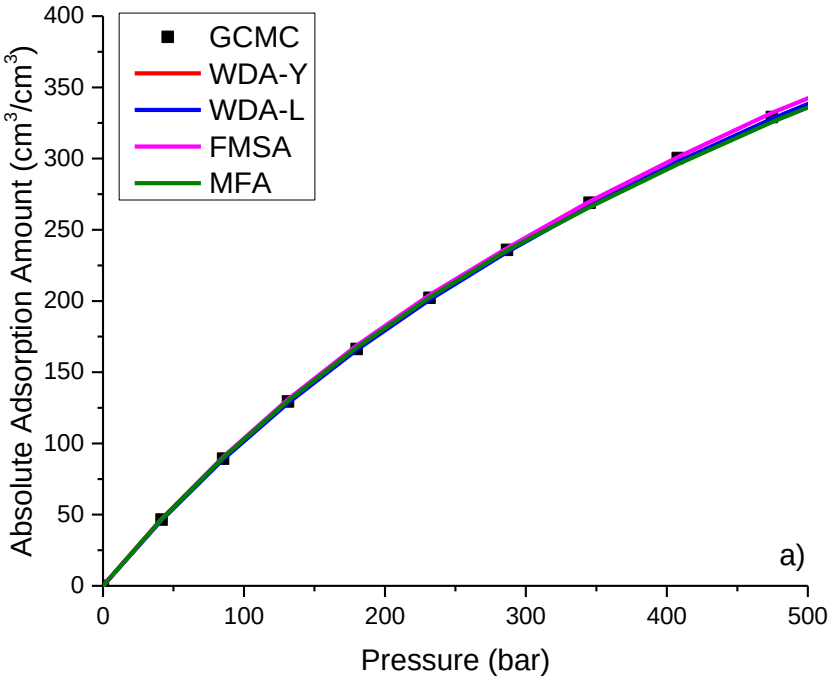
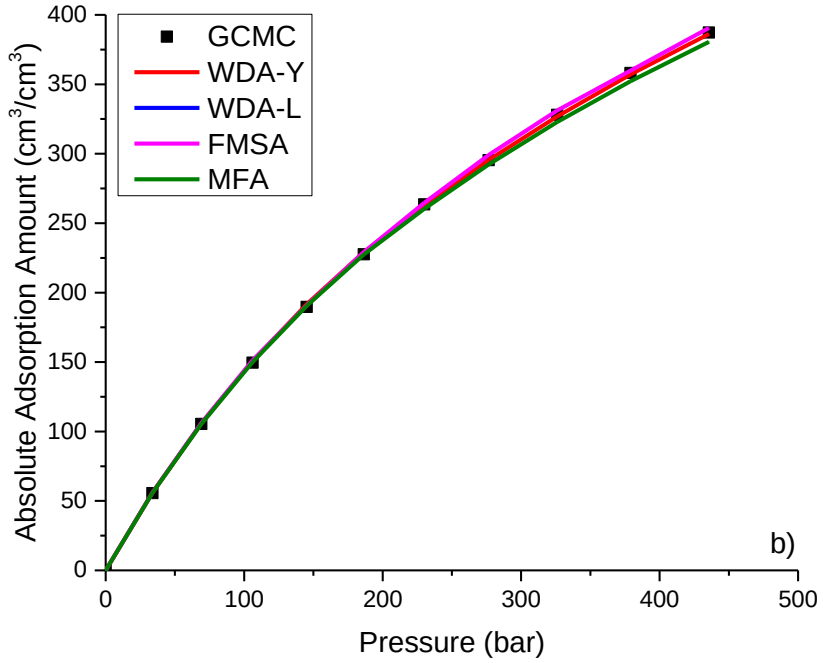


Figure 3-3 Adsorption isotherms for H_2 adsorption in MOF-5 calculated from GCMC and from four different versions of DFT for the attraction part of the excess Helmholtz energy functional. The system temperature is a) 298 K, b) 243 K, and c) 77 K.




GCMC
WDA-Y
WDA-L
FMSA
MFA
Absolute Adsorption Amount (cm3/cm3)
Pressure (bar)
b)
0
50
100
150
200
250
300
350
400
0
100
200
300
400
500

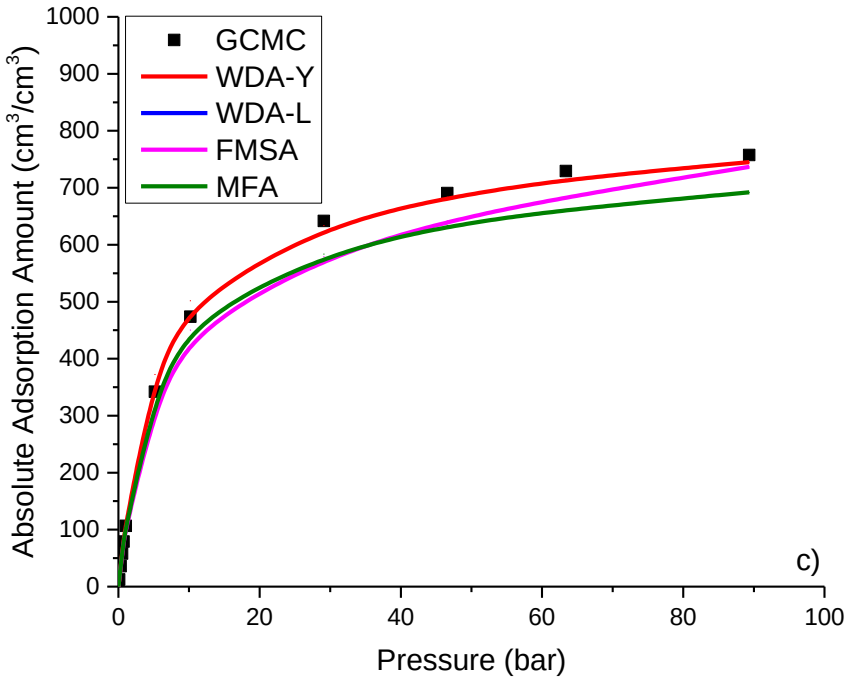
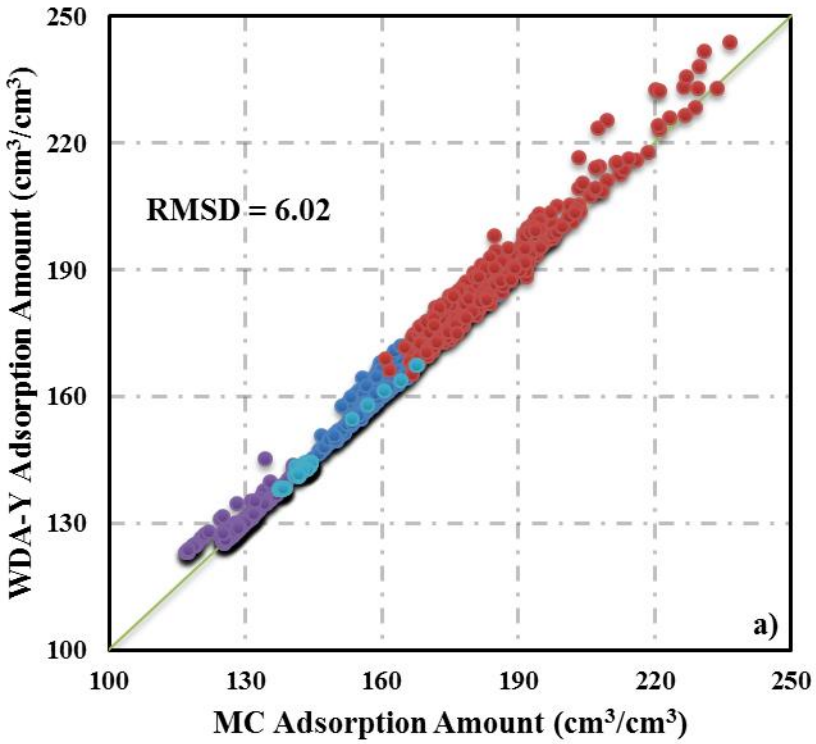
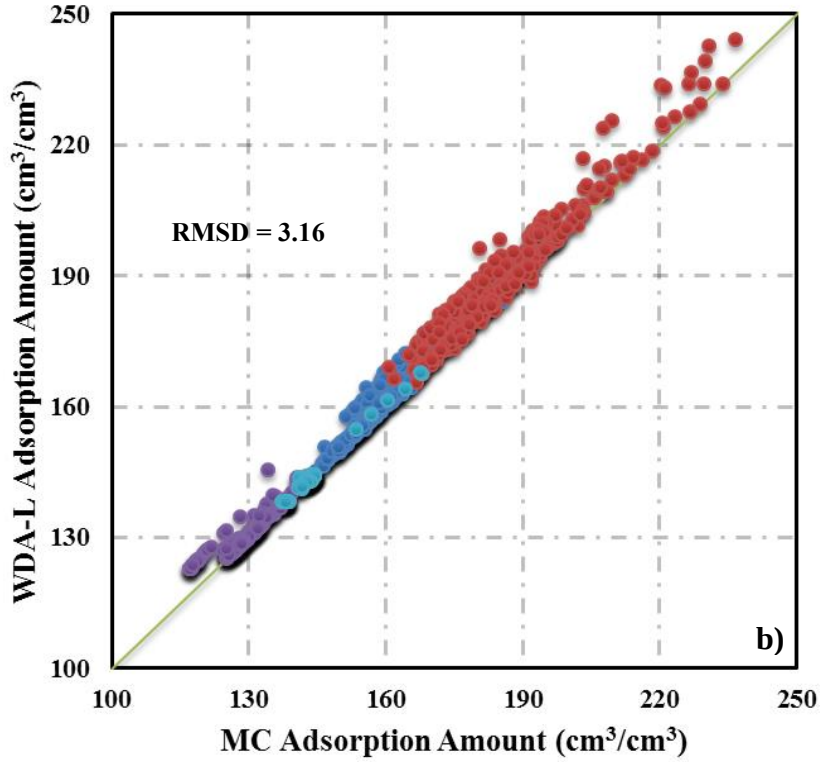
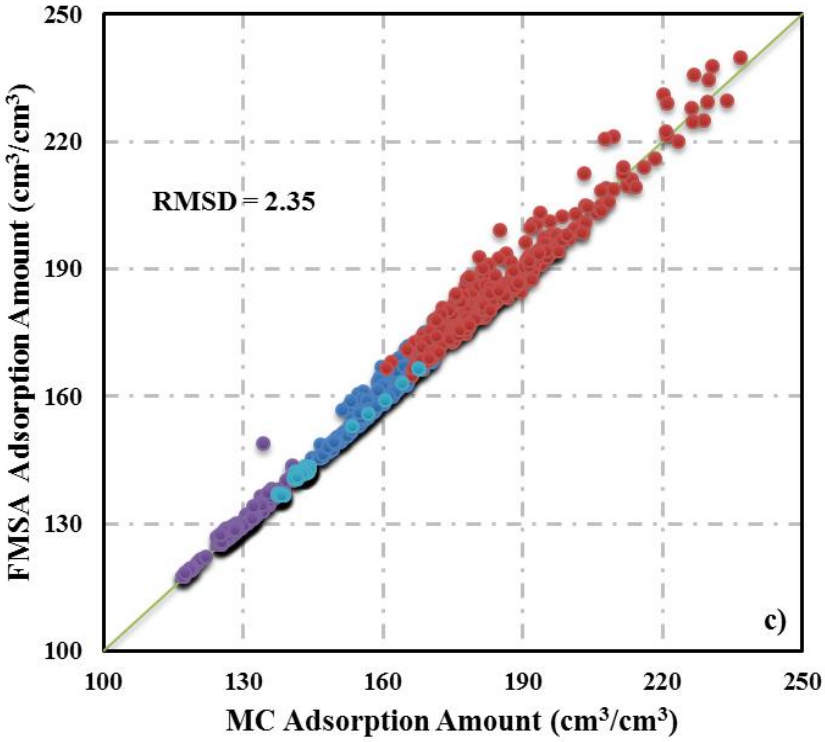
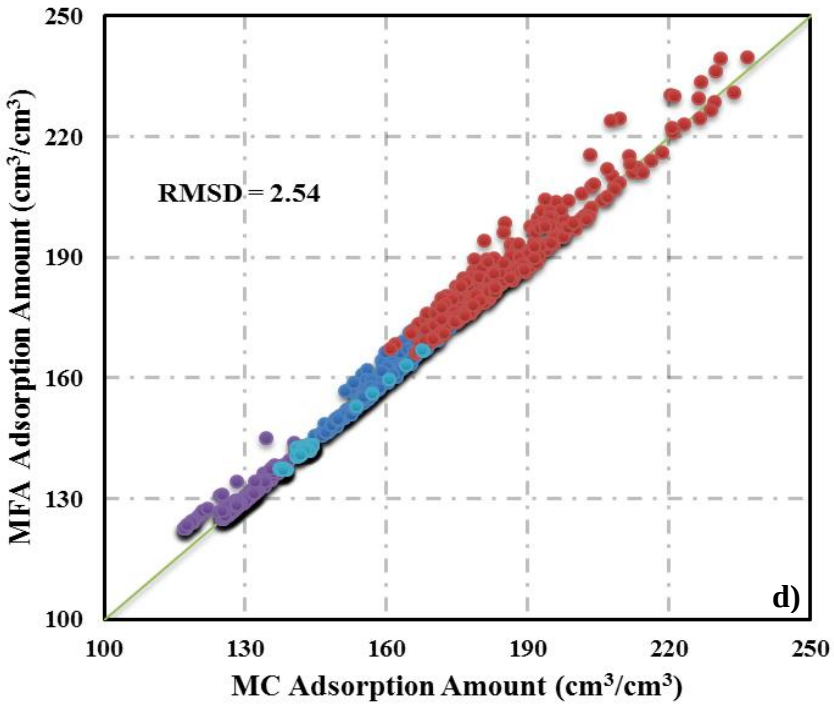

GCMC
WDA-Y
WDA-L
FMSA
MFA
Absolute Adsorption Amount (cm3/cm3)
Pressure (bar)
c)
0
100
200
300
400
500
600
700
800
900
1000
0
20
40
60
80
100

Figure 3-4 Absolute H_2 adsorption amounts in 1,200 MOFs calculated from various DFT methods in comparison with GCMC simulations. Here the temperature and pressure of hydrogen gas in the bulk are fixed at T=243K and P=100 bar, respectively. The DFT functionals are a) WDA-Y, b) WDA-L, c) FMSA, d) MFA and e) WDA-Y with FMSA EOS. Color code: Blue, top 300 from excess CH_4 adsorption in weight category; Red, top 300 from excess CH_4 adsorption in volume category; Purple, top 300 from void fraction category; Teal, top 300 from surface area category.




250
220
190
160
130
100
WDA-L Adsorption Amount (cm³/cm³)
RMSD = 3.16
b)
100
130
160
190
220
250
MC Adsorption Amount (cm³/cm³)


250
220
190
160
130
100
FMSA Adsorption Amount (cm³/cm³)
RMSD = 2.35
c)
100
130
160
190
220
250
MC Adsorption Amount (cm³/cm³)


RMSD = 2.54
MFA Adsorption Amount (cm³/cm³)
MC Adsorption Amount (cm³/cm³)
d)
100
130
160
190
220
250

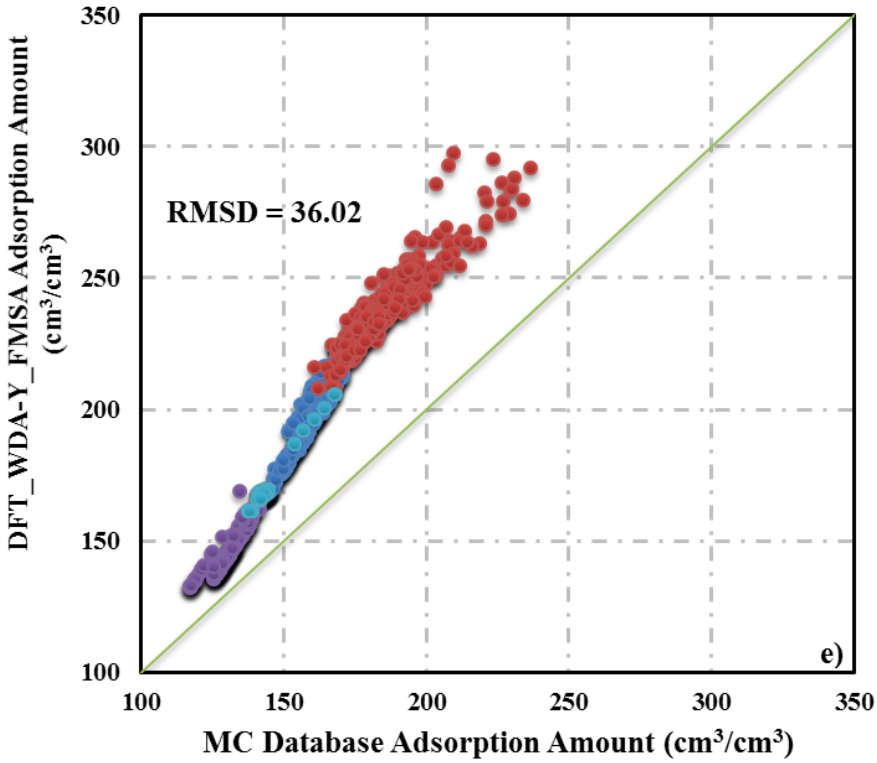
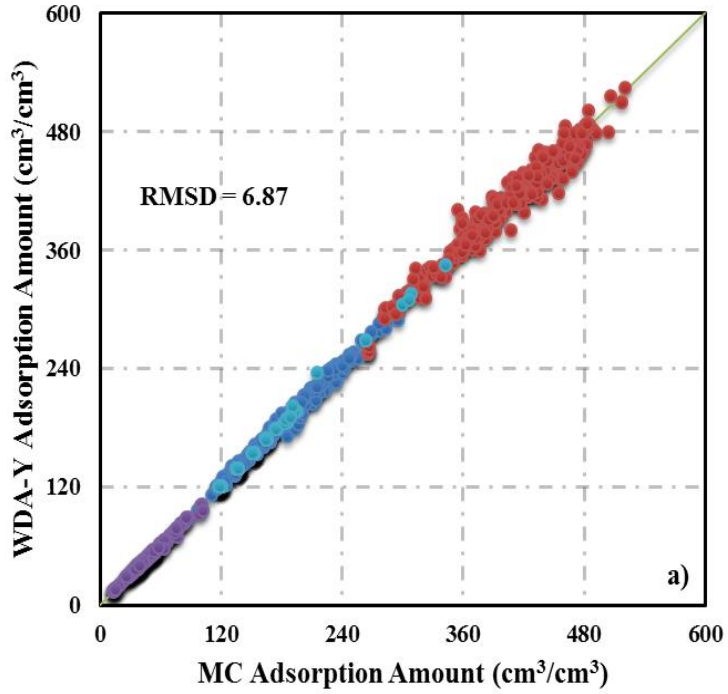
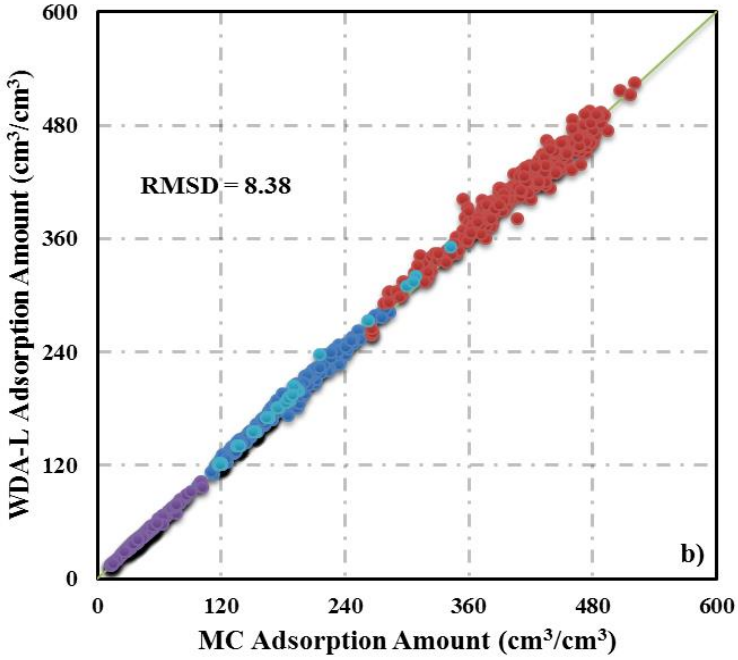
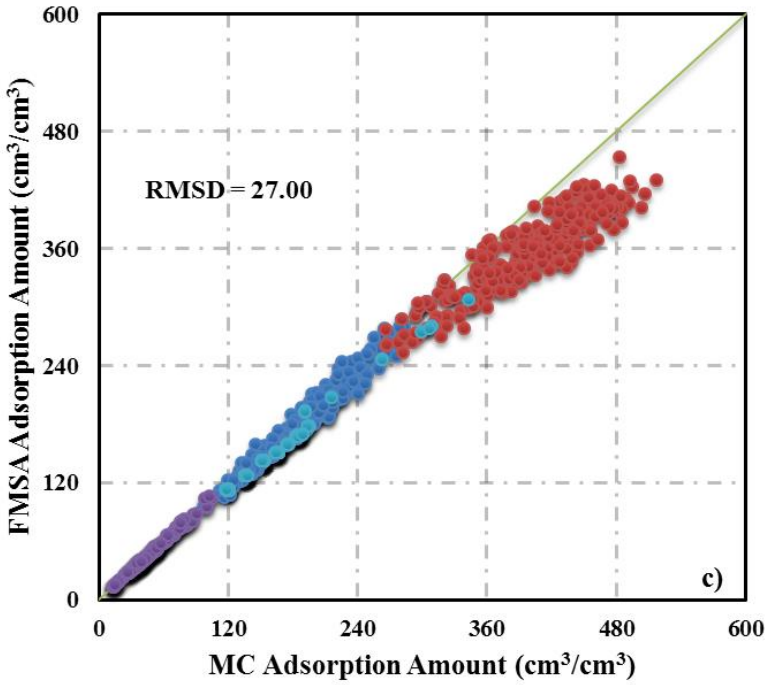

RMSD = 36.02
DFT_WDA-Y_FMSA Adsorption Amount (cm³/cm³)
MC Database Adsorption Amount (cm³/cm³)
e)
100
150
200
250
300
350

Figure 3-5 The same as Figure 3 but at T=77K and P=1bar. The DFT functionals are a) WDA-Y, b) WDA-L, c) FMSA, and d) MFA.






RMSD = 27.00
c)
FMSA Adsorption Amount (cm³/cm³)
MC Adsorption Amount (cm³/cm³)
0
120
240
360
480
600

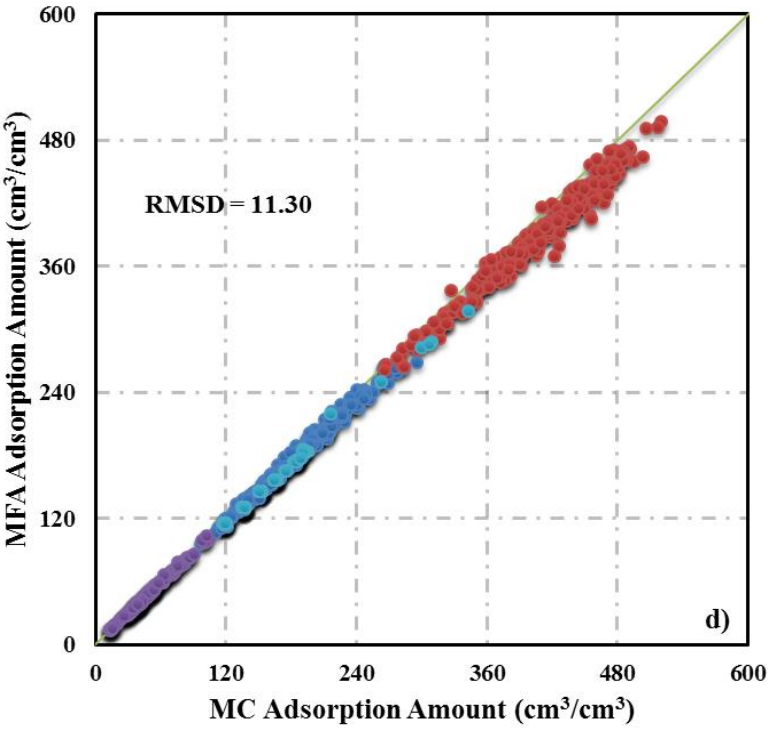
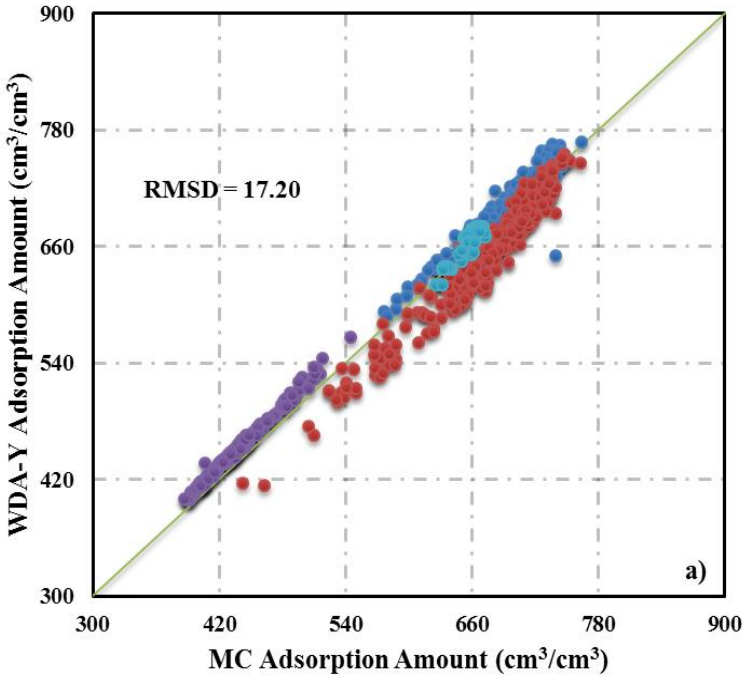
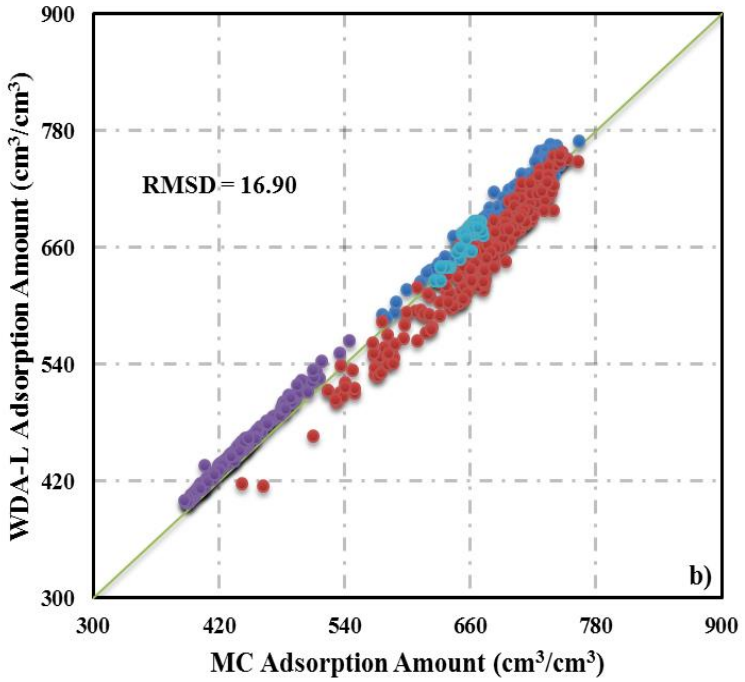
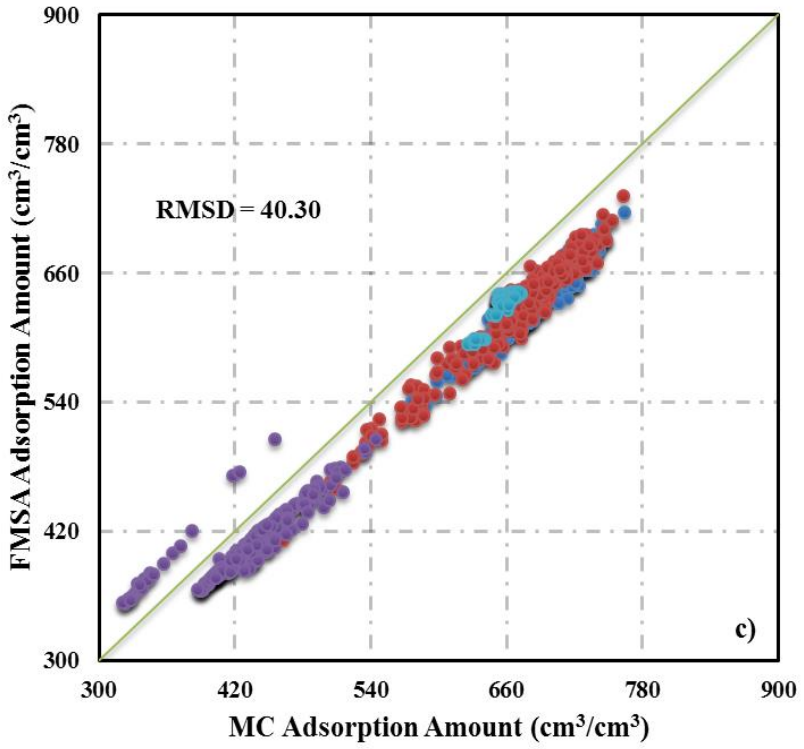

RMSD = 11.30
d)
MFA Adsorption Amount (cm³/cm³)
MC Adsorption Amount (cm³/cm³)
0
120
240
360
480
600

Figure 3-6 The same as Figure 3 but at T=77K and P=50bar. The DFT functionals are a) WDA-Y, b) WDA-L, c) FMSA, and d) MFA.






900
780
660
540
420
300
RMSD = 40.30
c)
FMSA Adsorption Amount (cm³/cm³)
300
420
540
660
780
900
MC Adsorption Amount (cm³/cm³)

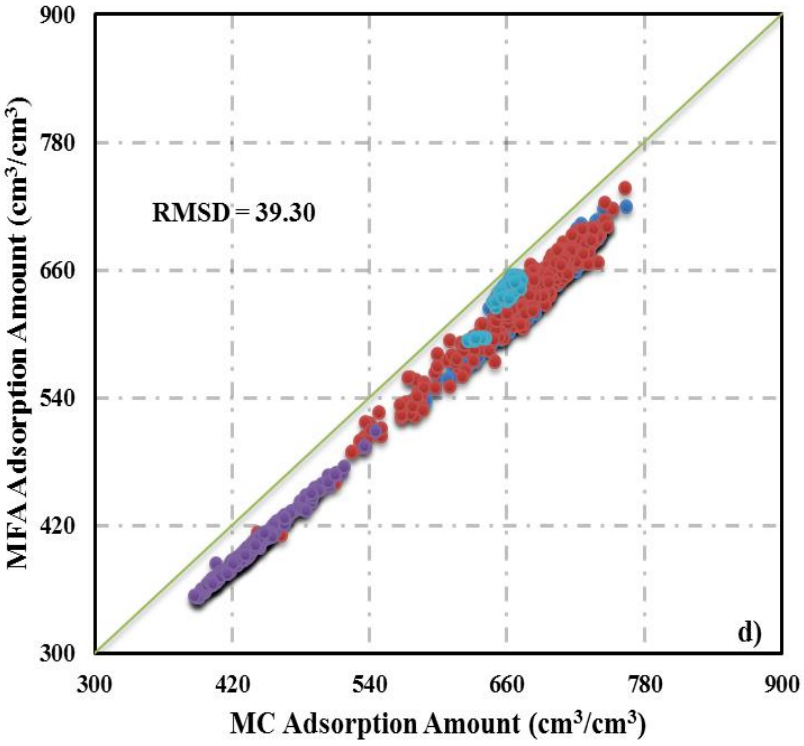

900
780
660
540
420
300
RMSD = 39.30
d)
MFA Adsorption Amount (cm³/cm³)
300
420
540
660
780
900
MC Adsorption Amount (cm³/cm³)

Figure 3-7 The distribution of computational time (unit: second) in DFT calculations for hydrogen adsorption in 1,200 MOFs with WDA–Y at T=243K and P=100bar. The computational time is based on a desktop PC with one Intel E-1230 CPU core.

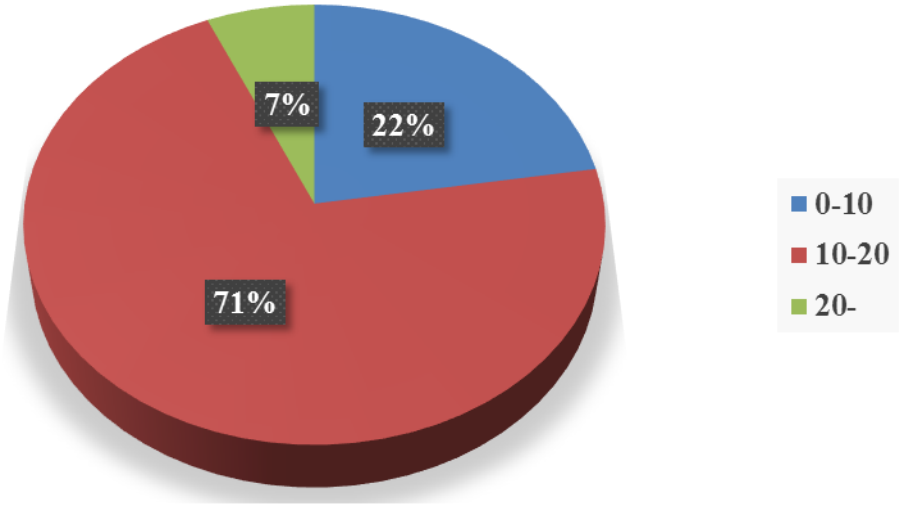


Table 3-1 Top 10 MOF candidates for H_2 adsorption at T=243 K and P=100 bar according to GCMC and four versions of DFT. Here "No." means the ranks of MOFs in the Northwestern hypothetic MOF database among the category of excess CH_4 adsorption in volume, Γ is H_2 adsorption amount in units of $cm^3(STP)/cm^3$.

Table 1

GCMC		WDA-Y		WDA-L		FMSA		MFA	
No.	**Γ**	**No.**	**Γ**	**No.**	**Γ**	**No.**	**Γ**	**No.**	**Γ**
36	236.6	36	243.8	36	244.1	36	239.8	36	239.9
153	233.8	71	241.7	71	242.7	71	237.7	71	239.5
71	230.9	27	238.2	27	239.3	189	235.8	27	236.2
27	229.9	189	235.7	189	236.5	27	234.5	189	233.5
22	229.7	2	233.3	2	234.1	37	231.2	153	231.0
131	228.9	22	233.1	153	234.0	153	229.6	37	230.3
189	226.9	153	233.0	22	233.9	22	229.4	77	230.2
285	226.6	37	232.6	37	233.6	77	229.0	2	229.5
2	226.4	77	232.2	77	233.2	2	227.8	22	228.8
11	223.3	131	228.5	131	229.4	131	224.9	131	226.5

3.2 An extension of Rosenfeld's excess-entropy scaling method for atomic diffusion in nanoporous materials

Abstract

We present an efficient computation procedure for the rapid calculation of the self-diffusivity atomic fluids in nanoporous materials by a combination of the Knudsen model, Rosenfeld's excess entropy scaling method, and a classical density functional theory (DFT). The self-diffusivity conforms to the Knudsen model at low density, and the effects of intermolecular interactions at higher densities are accounted for by Rosenfeld's excess-entropy scaling method. The classical DFT predicts the excess entropy used in the scaling analysis. The hybrid computational procedure has been calibrated with MD simulation for the adsorption of H_2, He, Ne and Ar gases in several nanoporous materials over a broad range of pressure. It predicts adsorption isotherms and different types of diffusion behavior in excellent agreement with the simulation results. While simulation of gas diffusion in nanoporous materials is time consuming, the new procedure is computationally very efficient. The theoretical method is extendable to more complicated systems including diffusion of water and polymers in confined geometry.

3.2.1. Introduction

Gas transport in porous materials is conventionally characterized by two types of diffusivities namely transport diffusivity and self-diffusivity[67,68]. The former is defined from a macroscopic perspective within the generic framework of transport phenomenon, and the latter in terms of molecular motions under equilibrium conditions. The transport diffusivity, here denoted as D_t, is introduced according to Fick's law,

$$\mathbf{J} = -D_t \nabla c \tag{56}$$

where **J** is the diffusion flux, and c stands for the local gas concentration. In general, D_t depends on both the confining environment and the gas concentration and should be considered as a tensor instead of a scalar parameter. For convenience, however, D_t is often simply referred to as an average of the three diagonal elements of the diffusion tensor, i.e., $D_t = (D_{t,xx} + D_{t,yy} + D_{t,zz})/3$. The self-diffusivity, designated as D_s, is defined in terms of the molecular trajectories according to Einstein's equation,

$$D_s = \lim_{t\to\infty} \frac{1}{2Dt} \left\langle |\mathbf{r}_i(t) - \mathbf{r}_i(0)|^2 \right\rangle_i \tag{57}$$

where D represents the spatial dimensionality, $\mathbf{r}_i(t)$ is the position of molecule i at time t, and the angle brackets denote the ensemble average over the trajectories of all molecules. Although there is no apparent connection between the two diffusivities, a common approximation is given by [68]

$$D_t \approx D_s \left(\frac{\partial \ln f}{\partial \ln c} \right)_T \tag{58}$$

where f represents the gas fugacity, and T is the absolute temperature. In the limit of low gas concentration, $c \to 0$, Eq. (58) indicates that the two diffusivities become identical, $D_t(0) = D_s(0)$. The approximate connection between the transport and self-diffusivities enables the broad applications of various statistical-mechanical methods to studying transport processes that often depart from equilibrium conditions.

Theoretical predictions of self-diffusivities from a molecular perspective are mostly based on molecular dynamics (MD) simulations[69,70]. Within the framework of a semi-empirical force field, the simulation method is in principle exact and yields ultimate microscopic details. However, diffusion under confinement often entails quasi-activation processes that make direct simulation extremely time consuming. While the numerical efficiency may not be a major concern from a scientific perspective, it becomes an issue for *in silico* screening of a big library of premeditated materials for gas storage and separation[71-74]. Towards rapid predictions of both thermodynamic and transport properties, the classical density functional theory (DFT) provides an excellent alternative to the simulation methods[30]. Whereas the usage of DFT for gas adsorption in porous materials has already been well documented[35,47,49,53,75-77], its application to investigating the transport properties of inhomogeneous fluids is normally considered in terms of the time-dependent or dynamic DFT[78]. As recently suggested by Truskett *et al.*[79,80], the thermodynamic properties predicted by the equilibrium DFT can be directly connected with transport coefficients by excess entropy scaling. Such a connection was discovered first by Rosenfeld many years ago[81]. Through the excess entropy scaling, the equilibrium DFT

alone can be used to predict both the thermodynamic properties and transport coefficients of confined fluids without explicit consideration of molecular motions.

The purpose of the present work is to extend Rosenfeld's excess entropy scaling method, which was originally proposed for correlating the thermodynamic and transport coefficients of bulk fluids, to predict gas diffusion under confined geometry. As detailed below, the new computational method combines the Knudsen model for gas diffusion at low density and the excess entropy scaling at high density inside the pores. While the extension of the excess entropy scaling method to inhomogeneous has been attempted before[79,80], we are unaware of its combination with the Knudsen model and the application of classical DFT for predicting the excess entropy.

3.2.2. Theoretical development

I. Rosenfeld's scaling relation between self-diffusivity and excess entropy

The scaling relation between the self-diffusivity and the excess entropy of a bulk simple fluid was discovered first by Rosenfeld in 1977 based on earlier simulation results for the transport coefficients of a wide variety of one-component systems including those containing hard spheres, soft spheres, or plasma[81,82]. Rosenfeld demonstrated that, in dimensionless units, the self-diffusivity of a bulk fluid is well correlated with the excess entropy in terms of an exponential relation:

$$D_R^* = A\exp(Bs^{ex}) \tag{59}$$

where D_R^* is the reduced self-diffusivity of the bulk fluid, defined by $D_R^* = D_s\rho^{1/3}(m/k_BT)^{1/2}$; ρ is the molecular number density, m is the molecular mass, k_B is

the Boltzmann constant, and $s^{ex} = S^{ex}/(Nk_B)$ is the reduced excess entropy per molecule. By correlation of extensive simulation results with Eq.(59), Rosenfeld conjectured that parameters A = 0.585 and B = 0.788 are quasi-universal constants, independent of the specific forms of the intermolecular potentials. The universality of these constants may be rationalized by the fact that the microscopic structure of a simple fluid can be well represented by the hard-sphere model and that diffusion of spherical molecules is dominated by intermolecular collisions.

While *a posteriori* arguments had been given to justify for the excess-entropy scaling method, the validity of Rosenfeld's relations was established by essentially empirical means. From a theoretical perspective, Dzugutov showed that the exponential dependence of the self-diffusivity on the excess entropy could be derived on the ground of a statistical-mechanical analysis[83,84]. According to the theoretical derivation, the exponential form holds when the reduced self-diffusivity is defined as $D_R^* = D_s \Gamma_E^{-1} \sigma^{-2}$, where $\Gamma_E = 4\sigma^2 g(\sigma)\rho\sqrt{\pi k_B T/m}$ is the collision frequency, $g(r)$ is the radial distribution function of the bulk fluid, and σ is the molecular collision diameter. In terms of the dimensionless self-diffusivity introduced by Dzugutov, the universal constants in Eq. (59) should be modified as A = 0.049 and B = 1. Dzugutov indicated that B = 1 arises from the ergodic assumption, which is usually not satisfied in practical molecular simulations due to the finite time scale. The intrinsic error introduced in the simulation explains why $B < 1$ in the original scaling relation proposed by Rosenfeld[84]. Several alternative expressions have also been proposed relating transport coefficients with the reduced excess

entropies[79,85,86]. These later developments were mostly based on semi-empirical modifications of the original scaling relations by Rosenfeld or Dzugutov rather than on rigorous theoretical derivations. Nevertheless, they provide significant improvements over the numerical accuracy and extend the applicability scope of the excess entropy scaling method.

Neither the scaling relation proposed by Rosenfeld or by Dzugutov is strictly valid for diffusion in confined fluids. For example, He *et al*. [85] used Eq. (59) to predict the diffusion of methane in silica nanopores by treating parameter *A* as a function of the free volume. While satisfactory results were obtained for gas diffusion at the high-density limit, the scaling relation exhibits large errors at the low density. Unlike that in bulk systems, the self-diffusivity is not an intrinsic property of the confined fluid but depends also on the confining geometry as well as the surface energy. Besides, it has been shown that the configuration of a confined fluid is sensitive to the details of non-hard-sphere interactions[87]. In particular, Truskett and coworkers explored systematically possible correlations between the self-diffusion coefficients of various fluids under confinement with different thermodynamic measures[79,80,87,88]. While the scaling relations established for bulk systems were in general not applicable under confinement, excellent correlations were identified between the self-diffusivities of confined hard-sphere and Lennard-Jones fluids and the corresponding excess entropies over a broad range of thermodynamic conditions[80,89].

II. Extension of the excess entropy scaling to confined fluids

Gas diffusion in a nanoporous system is dominated by two types of collisions: gas-gas collisions and gas-surface collisions. In the limit of high packing density, gas diffusivity is dominated by collisions among gas molecules rather than gas-surface collisions. In this case, we expect that the entropy-scaling rule should be applicable because the mechanism of gas-gas collisions in a confined geometry is similar to that in the bulk phase. In the low-density limit, however, collisions between gas molecules and the material surface is more important than the gas-gas collisions. As a result, Eq. (59), which works well for bulk systems, is qualitatively inaccurate. At the extremely low density, the collision frequency approaches a non-zero constant in a confined space but zero in the bulk system. In the former case, gas diffusion follows the Knudsen model instead of the excess entropy scaling.

Traditionally, the modified free-volume theories (MFV) are often used for predicting the diffusivity of gas molecules in confined geometry[90-92]. While these methods are computationally efficient, their practical applications rely on a set of empirical parameters that must be obtained by fitting to the experimental data. For example, the Cohen-Turnbull (CT) theory [93], which is one of the most popular version of the MFV models, represents the self-diffusivity in terms of the free volume of a porous material:

$$D_s = ka^* u \exp(-\gamma v^* / v_f) \tag{60}$$

where k represents a geometric factor, a^* is a parameter related to the pore size, u stands for the mean speed of the gas molecules in the bulk, γ is the free-volume overlap parameter,

v^* is the critical volume facilitating the diffusion process, and v_f is the molecular free volume.

The CT theory suggests a linear relationship between $\ln D_s$ and $-1/v_f$ [Eq.(60)]. The larger the molecular free volume, the larger the diffusion coefficient. Because the diffusion of gas molecules in a confined geometry reflects a competition between the gas-surface and the gas-gas interactions, we assume that the overall diffusivity may be determined by a linear combination of the Knudsen diffusion model in the limit of large molecular free volume and the excess-entropy scaling in the opposite limit:

$$\ln D_s = \left(1 - \frac{\lambda}{v_{free}}\right) \ln D_K + \frac{\lambda}{v_{free}} \ln D_E \tag{61}$$

where v_{free} is defined by the total accessible volume of the material divided by the number of confined gas molecules [85], and λ is a preset factor weighting the different contributions of the gas-surface and gas-gas interactions to the overall diffusivity. In the Knudsen model, the self-diffusivity is independent of the gas density. In that case, we have

$$D_K = D_s(0) \tag{62}$$

where $D_s(0)$ can be obtained by a single run of molecular dynamics simulation. For the contribution of gas-gas interactions, D_E, we follow Rosenfeld's excess entropy scaling rule:

$$D_E = \frac{0.585}{\rho^{1/3}} \sqrt{\frac{k_B T}{m}} \exp\left(0.788 s^{ex}\right) \tag{63}$$

where s^{ex} is calculated from the classical density functional theory (CDFT). The detail procedure for calculation of s^{ex} is given in the Supplementary Materials.

In Eq.(61), $\lambda / v_{\text{free}}$ can be understood as the fraction of the free space inside the pore occupied by gas molecules. To make it dimensionless, we assume that the volume of gas molecules is given by

$$\lambda = \alpha\sigma^3 \tag{64}$$

where σ is the diameter of gas molecules, and α is a constant reflecting the molecular packing density. A similar approach has also been used by Mittal *et al*. for hard-sphere quenched-annealed (HS QA) models.[94] In the limit of the closest packing for spherical particles as in a face-centered-cubic lattice, we have

$$\alpha_{FCC} = \sqrt{2}/2\,. \tag{65}$$

From a physics point of view, we expect that $\alpha \approx \alpha_{FCC}$. For best fitting of the simulation or experimental data, however, α may also be used as an adjustable parameter. While this parameter may be slightly different for different systems, it should be independent of thermodynamic conditions such as temperature and pressure.

3.2.3. Results and discussion

To test the numerical performance of the extended excess entropy scaling method, we first consider the adsorption isotherms and self-diffusion coefficients of H_2 gas in four types of nanoporous materials at room temperature: MOF5, MFI, CuBTC and ZIF8. The details of the molecular models and simulation methods are described in the Supplementary Materials. Figure 3-8 compares the adsorption isotherms predicted from the CDFT with those from Monte Carlo simulation. As expected, the CDFT predictions agree well with the simulation results for all four types of nanoporous materials. The good agreement is

somewhat expected because the CDFT has already been demonstrated to be an excellent computational method for predicting a wide variety of thermodynamic properties pertinent to gas adsorptions[35,49,53,77].

Prediction of the gas diffusivity inside the porous materials depends on parameter α in Eq.(64). Without any additional input, we may treat α as a universal constant that is the same as α_{FCC}. A more quantitative prediction can be achieved by using α as an adjust parameter for each material. Figure 3-9 shows the self-diffusivities of H_2 molecules in CuBTC and MOF5. For CuBTC, the extended excess entropy scaling method reproduces the whole diffusion curve very well with $\alpha = \alpha_{FCC}$. For MOF5, however, it captures the diffusion curve semi-quantitatively with $\alpha = \alpha_{FCC}$. The slight overestimation can be readily corrected by using α as an adjustable parameter ($\alpha = 0.89$).

Figure 3-10 shows the hydrogen self-diffusivities in MFI and ZIF8 also at room temperature. Similar to Figure 3-9, $\alpha = \alpha_{FCC}$ yields semi-quantitative predictions and the theoretical results can be further improved by using α as an adjustable parameter for each material. The change of α value in these materials is small, approximately in the range between 0.7 and 0.9. The fluctuation of α for different materials may arise from the inhomogeneous density distribution of gas molecules. For materials such as MOF5, gases are nearly concentrated at the surface of the framework and collisions are essentially occurred near the surface instead of the whole free volume. In this case, a smaller free volume or equivalently, a larger α is required to bridge the Knudsen and bulk behavior. Comparing hydrogen diffusion in MOF5, MFI, CuBTC and ZIF8, we find that the effect of pressure on the self-diffusion coefficient goes in a diverse way: while the diffusivity of

hydrogen molecules in MOF5 and CuBTC attenuates linearly with the pressure; it declines with pressure in MFI in a nonlinear way. Interestingly, the diffusivity is almost independent of pressure in ZIF8. Qualitatively, the extended excess entropy scaling method captures all these trends even with the simple approximation $\alpha \approx \alpha_{FCC}$.

We consider also the diffusion of other gas molecules in nanoporous materials. Figure 3-11 shows the diffusivities of He and Ne in MFI. The structure of MFI and all the related force-field parameters are the same as those given in the literature[95]. It seems that the self-diffusion coefficients of both gases decline linearly with the pressure. As for hydrogen diffusion, the extended excess entropy scaling method yields a semi-quantitative prediction of the diffusion curve with $\alpha = \alpha_{FCC}$. While the theory overestimates the simulation results, its performance can be much improved for both cases by using a fitted value of α, *viz.*, α_{He}= 1.21 and α_{Ne}= 0.93.

Similar conclusions can be reached by considering diffusion of argon molecules in MOF5. As shown in Figure 3-12, the extended excess entropy scaling method underestimates the diffusivity with $\alpha = \alpha_{FCC}$ but excellent results can be achieved by using α = 0.57. Comparing the values of α for different cases, it seems that α depends on the size of the gas molecules (σ_{He} = 2.28 Å, σ_{Ne} = 2.789 Å and σ_{Ar} = 3.42 Å), i.e., α decreases with the increasing of the gas diameter. At the same molecular packing density, the larger molecular diameter implies a lower number density of gas molecules. As the gas-material collision becomes more important, or namely, the Knudsen diffusion gives more contributions, parameter α, which is proportional to the entropy scaling contribution, should be reduced.

In comparison with molecular dynamic simulations, the extended excess entropy scaling method is computationally very efficient. Instead of hours of computation time in a supercomputer, the calculation of each curve for the self-diffusivity as a function of pressure can be finished within just one minute in a desktop computer.

3.2.4. Conclusions

We have proposed an efficient computational procedure for predicting the self-diffusivity of gas molecules in confined geometry by the semi-empirical combination of the entropy-scaling method and the Knudsen model. The classical density functional theory (CDFT) is introduced to calculate the excess entropy, and the Knudsen diffusivity is calculated from regular molecular dynamics simulation. The theoretical predictions agree well with the simulation results for the diffusivities as well as the adsorption isotherms of several light gases in different types of nanoporours materials. In comparison with conventional simulation methods, a key advantage of the theoretical procedure proposed in this work is its computational efficiency. The time cost of one CDFT calculation is less than 1 minute in a single 3.0 GHz CPU, which is negligible comparing to molecular simulations.

As demonstrated before, the CDFT predictions of gas adsorption isotherms accord well with the simulation results. By using a universal constant, $\alpha = \alpha_{FCC}$, the extended excess entropy scaling method yields semi-quantitative results for the self-diffusion coefficients of various gases in confined geometry, and the theoretical predictions can be further improved by treating α as an adjustable parameter for each gas/material. We find

that parameter α seems to depend on the size of the guest molecules: the larger diameter of gas molecules, the smaller value for α.

An important distinction of this work from the previous publications is that we provide an explicit expression to correlate the excess entropy with the diffusion coefficient. Such expression was not given before for inhomogeneous systems. While an adjustable parameter is used to incorporate the Knudsen model into the excess entropy scaling, this parameter can be reasonably estimated such that the hybrid method can be used as a useful predictive tool. It has been reported recently that similar correlations may exist between the excess entropies and transport coefficients for more complicated molecular systems including water, ionic fluids and polymeric fluids[88,96-98]. Because the reliable DFT formulations have been well established for such systems[99], we expect that the computational procedure proposed in this work can be similarly applicable to predicting the thermodynamic properties and self-diffusivities of non-spherical molecules under confinement.

Figure 3-8 Comparison of absolute adsorption isotherms predicted by the DFT and Monte Carlo simulation for H_2 adsorption in 4 types of nanoporous materials at 298K.

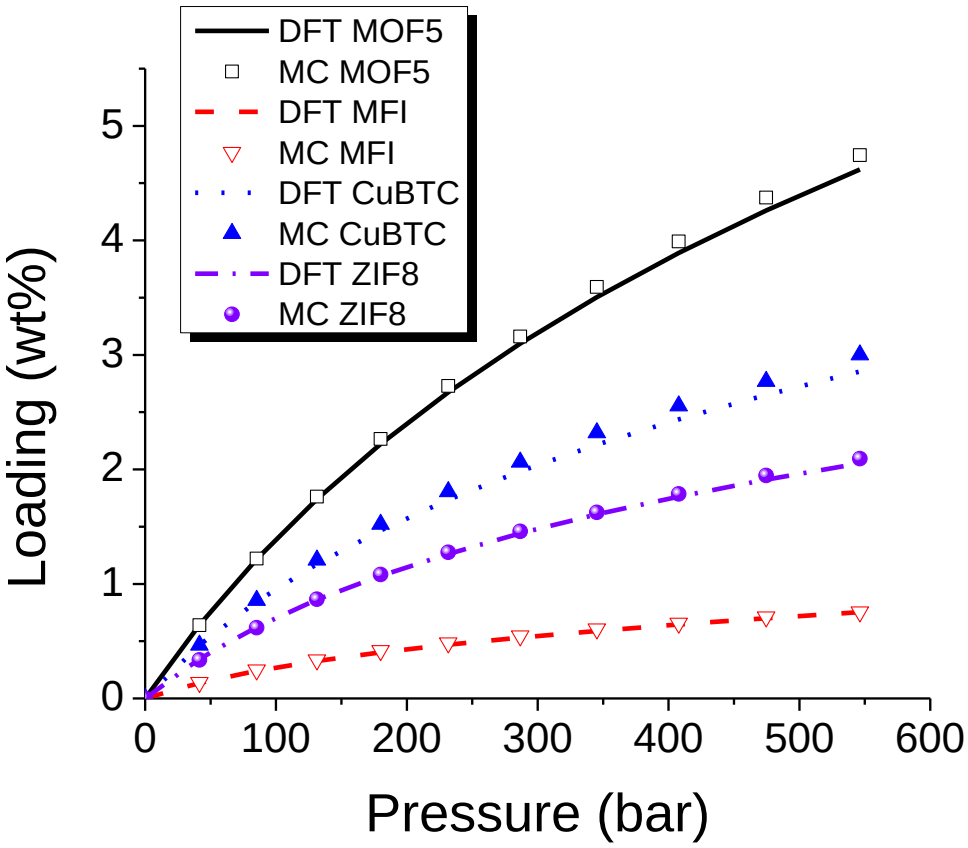


Figure 3-9 Self-diffusivities of H_2 in CuBTC and MOF5 at 298K predicted by MD simulation and the extended excess entropy scaling method, respectively. Simulation results for MOF5 are from ref [69].

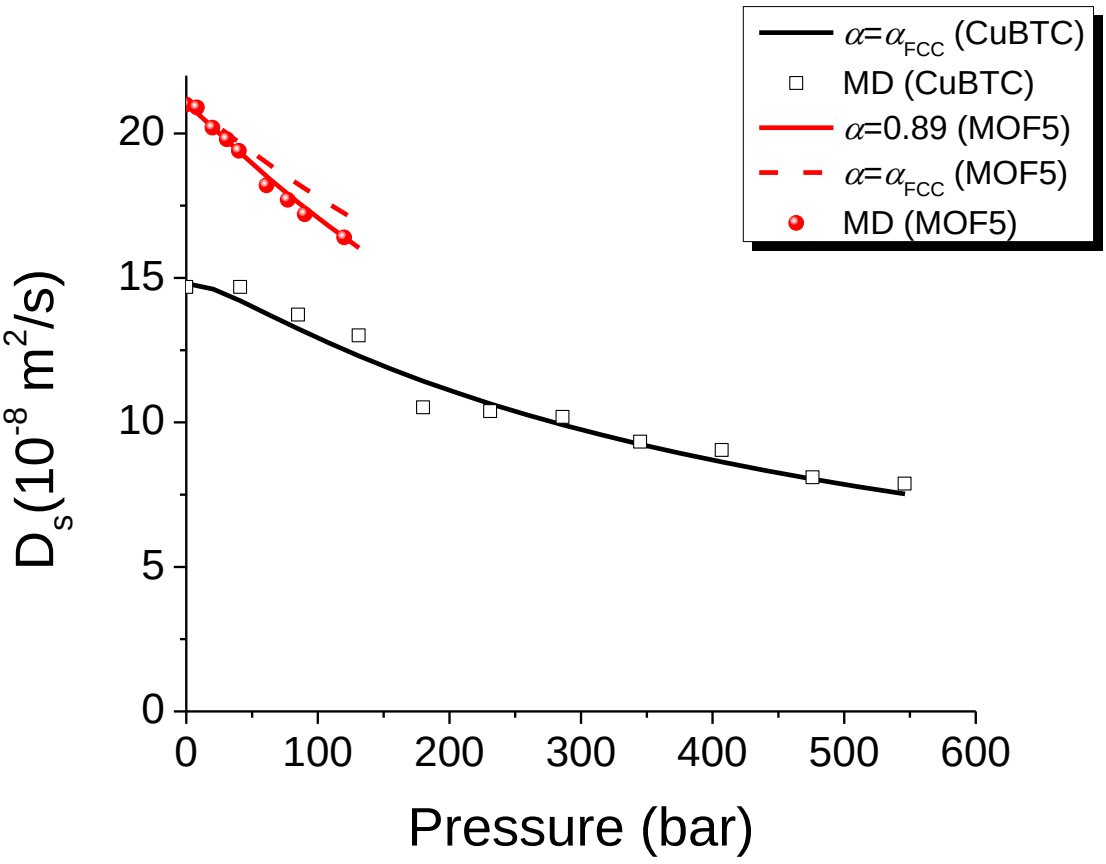


Figure 3-10 Self-diffusivities of H_2 in MFI and ZIF8 at 298K, predicted by MD simulation and the extended excess entropy scaling method, respectively.

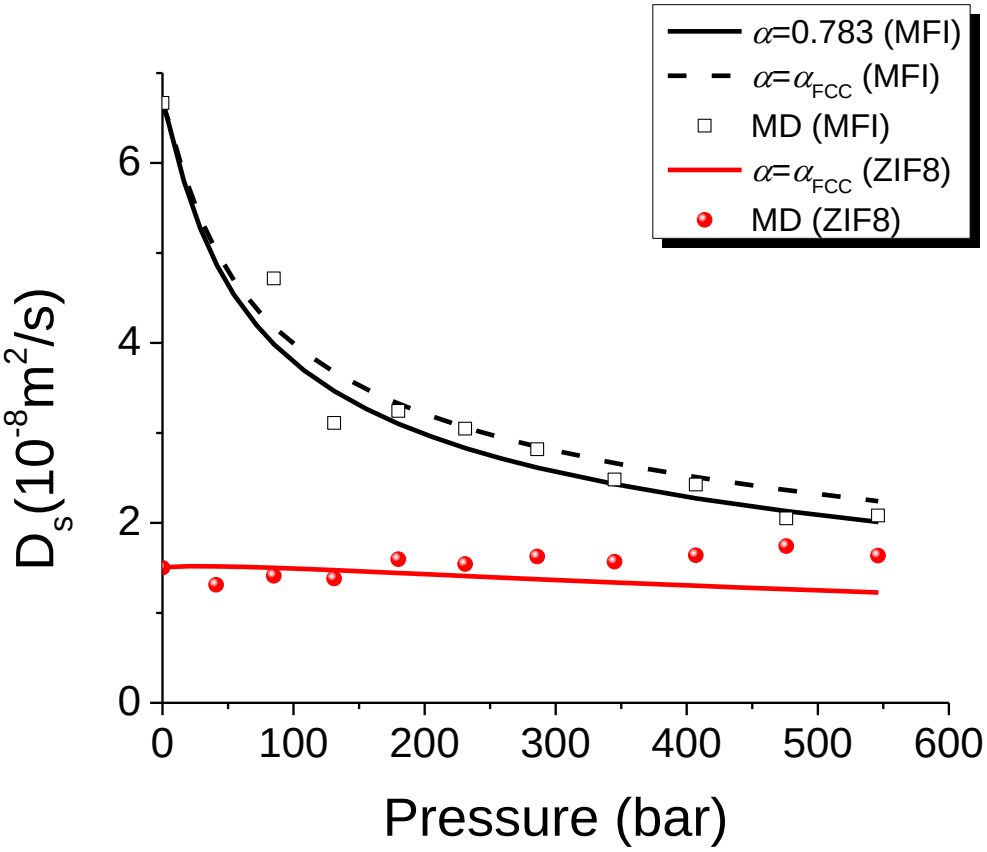


Figure 3-11 Self-diffusivities of He and Ne in MFI, predicted by MD simulation and the extended excess entropy scaling method, respectively. Simulation results are from ref [95].

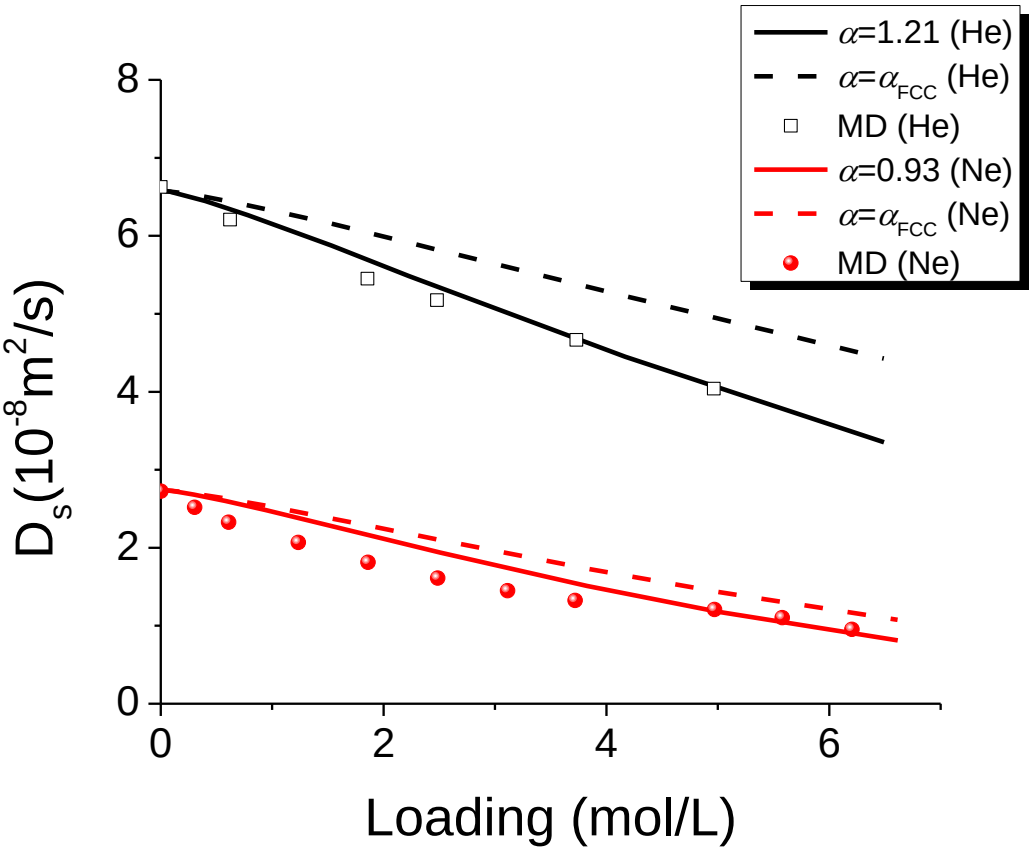
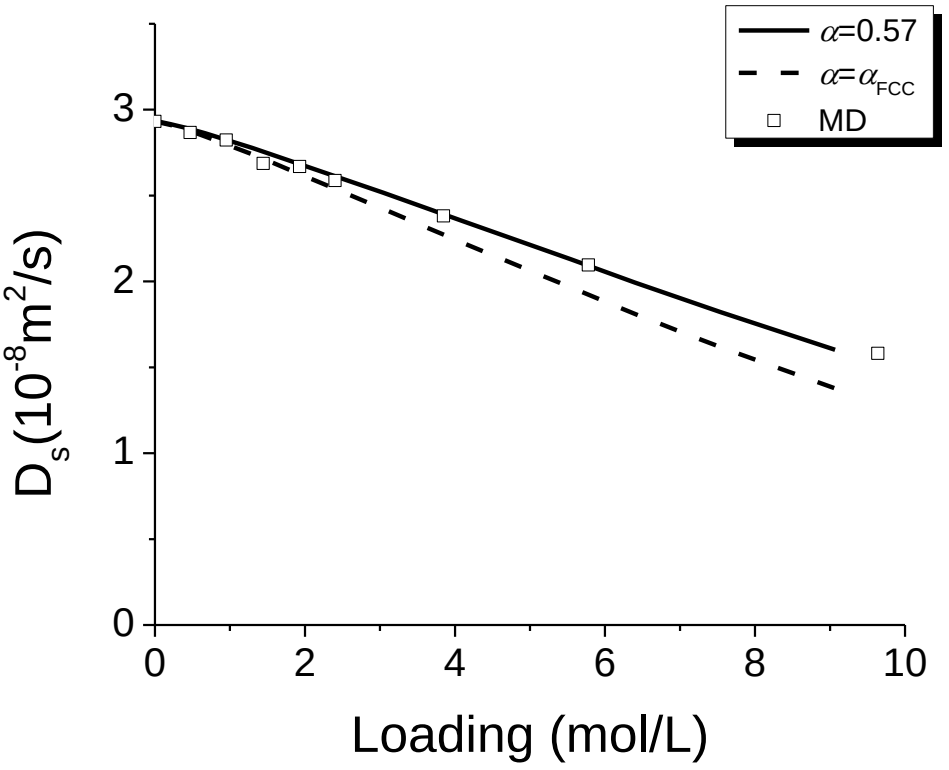


Figure 3-12 Self-diffusivity of Ar adsorbed in MOF5 predicted by MD simulation and the extended excess entropy scaling method, respectively. Simulation results are from ref [69].



Supporting Information

This supplementary material details the Monte Carlo (MC) and molecular dynamics (MD) simulations employed in this work as well as the density functional theory (DFT) equations for calculation of the excess entropy.

I. Molecular Models and Simulation Methods

We used the united atom model for all gas molecules and the rigid molecular model for the nanoporous materials. The Lennard-Jones (LJ) potential is used to describe all atomic pair interaction energies:

$$u_{ij}(\mathbf{r}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right] \tag{66}$$

where the cross parameters are given by Lorentz-Berthelot combination rule

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j} \qquad \sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \tag{67}$$

Table S1 gives the LJ parameters for the gas molecules studied in this work.

Gas	σ(Å)	ε(K)
H_2	2.958	36.7
He	2.28	10.223
Ne	2.789	35.7
Ar	3.42	124.07

Table S3-1. Lennard-Jones Potential Parameters for H_2, He, Ne, Ar[61,95]

The LJ parameters for MOF-5, Cu-BTC and ZIF-8 were from universal force field (UFF), while MFI was from Hirotani et al. The experimental determined crystal structures were

used in all theoretical calculations.[100-103] In both simulation and DFT calculations, we did not consider the flexibility of the naonporous materials.

Adsorption isotherms of gas molecules in different porous materials were predicted using the conventional GCMC simulation method, as illustrated in 3.1. To calculate the self-diffusivities of the gas molecules, we have used equilibrium molecular dynamics (EMD) simulations. The simulations were performed in the canonical ensemble (NVT) using the Nosé-Hoover thermostat. In anisotropic three-dimensional materials, the self-diffusivity is calculated from the Einstein relation

$$D_s(c) = \lim_{t\to\infty} \frac{1}{6t} \left\langle \left|\mathbf{r}(t) - \mathbf{r}(0)\right|^2 \right\rangle \tag{68}$$

where $\mathbf{r}(t)$ stands for the position of a tagged particle at time t, and the angular brackets indicate an ensemble average. At a given thermodynamic condition, we performed 5 independent MD simulations for each loading, with a simulation length 4.0 ns after 1.0 ns EMD simulations to equilibrate the system. The standard deviation is about 5% ~ 20% times of D_s. The time step for the MD simulation is 1.0 fs. All MD simulations were carried out using LAMMPS simulation package.[104]

II. Excess entropy

The classical density functional theory (DFT) used in this work has been reported in 3.1. Once we have the density profile $\rho(\mathbf{r})$, the absolute adsorption amount is calculated from

$$N = \int \rho(\mathbf{r}) \mathrm{d}\mathbf{r} \tag{69}$$

The excess entropy is obtained from temperature derivative of the excess free energy functional:

$$S^{\mathrm{ex}} = -\left(\frac{\partial F^{\mathrm{ex}}}{\partial T}\right)_{\rho(\mathbf{r})} \tag{70}$$

The DFT calculation is performed in a 3D cubic box with grid resolution of 0.2 Å. We have also tested higher resolutions such as 0.1 Å and 0.15 Å for a few cases and the results are very similar. The system is considered to converge when

$$|\rho^{(n)}(\mathbf{r}) - \rho^{(n+1)}(\mathbf{r})| < 10^{-5}\rho_{\mathrm{av}}^{(n)} \tag{71}$$

where $\rho_{\mathrm{av}}^{(n)}$ is the average density of step *n*, then the excess entropy is calculated from numerical derivation:

$$S^{\mathrm{ex}} = -\frac{F^{\mathrm{ex}}[\rho(\mathbf{r}), T+\Delta T] - F^{\mathrm{ex}}[\rho(\mathbf{r}), T]}{\Delta T} \tag{72}$$

where ΔT = 0.001 K.

3.3 Classical Density Functional Theory for Methane Adsorption in Metal-Organic Framework Materials

Abstract

Natural gas is being considered as a promising alternative to petroleum as the next generation of primary transportation fuel owing to relatively smaller carbon footprint and lower SO_x/NO_x emissions and to fast shale-gas developments in recent years. Since the volumetric energy density of methane amounts to only about 1% of that of gasoline at ambient conditions, natural gas storage represents one of the key challenges for prevalent deployment of natural gas vehicles (NGVs). In this work, we present a molecular thermodynamic model that is potentially useful for high-throughput screening of nanoporous materials for natural gas storage. Specifically, we have investigated methane adsorption in a large library of metal-organic frameworks (MOFs) using four versions of classical density functional theory (DFT) and calibrated the theoretical predictions with extensive simulation data for both the total uptakes and the delivery capacities. In combination with an extended excess entropy scaling method, the DFT is also used to predict the self-diffusion coefficients of the confined gas in several top-ranked MOFs. The thermodynamic model has been used to identify promising MOF materials and possible alternations of operation parameters to meet the ARPA-E target set by the US Department of Energy for natural gas storage.

3.3.1. Introduction

Adsorption thermodynamics plays a pivotal role in traditional as well as modern chemical engineering ranging from separation processes and heterogeneous catalysis to

various tools of chemical analyses such as chromatography[105,106]. From a practical point of view, one of the most useful thermodynamic models for gas adsorption was established by Professor John Prausnitz and one of his PhD students, Alan Myers, published in the *AIChE Journal* about 50 years ago[107]. We are pleased to dedicate this article to the special issue honoring Professor John Prausnitz.

According to conventional wisdom, adsorption is a surface phenomenon related to adhesion of molecules, atoms or ions at an interface. Whereas concepts such as surface area and binding energy are indispensable in classical thermodynamics to describe the interfacial phenomena, Myers and Prausnitz took a radically different approach by extending the ideal solution model for vapor-liquid equilibria (VLE) to adsorption of gas mixtures. The "solution model" is remarkably accurate for predicting the adsorption isotherm of a gas mixture from those corresponding pure components at moderate pressure, yet it is computationally convenient for practical applications. Because the number of pure species is rather limited in comparison to that for mixtures, the solution model has been invaluable for design and optimization of industrial adsorption processes and remains relevant today[108].

Driven by increasing concerns over global climate change and rapid developments in renewable energy, research in gas adsorption has experienced considerable growth in recent years. However, the main focus of current literatures has been shifted from adsorption equilibria *per se* to discovery of novel nanoporous materials, in particular open framework materials promising for large-scale hydrogen/methane storage or for selective adsorption of carbon dioxide from gas mixtures[109]. Unlike conventional adsorbents such as

activated carbons or silica gels, open framework materials are porous crystalline solids viable for solution synthesis by coordinate bonding of secondary building units (SBUs), i.e., libraries of organic linkers and organometallic/nonmetallic nodes that are able to be self-organized into periodic and porous frameworks[110]. Theoretical investigations are useful to acquire a better knowledge of adsorbate-adsorbent interactions and, perhaps more important, for discovery and rational design of nanoporous materials tailored to specific adsorbates. The modular nature of the building blocks makes the framework structures predictable on the basis of the topology and geometry of specific link-node coordination complexes, rendering unprecedented opportunities for computational discovery of crystalline porous materials to have a precise control over the porosity and atomic composition (size, shape, and function). Molecular modeling is useful to identify specific open framework materials from virtually infinite numbers of SBU combinations to meet specific practical needs[111-114].

Whereas there have been impressive theoretical developments toward the discovery of better adsorbents for gas storage and separation[73,74,115,116], computational materials design remains an enormous challenge from practical perspectives. To predict the material performance from the chemistry of its building blocks, we need efficient multiscale modeling methods that account for not only its intrinsic properties in vacuum but also its interactions with a working environment under diverse thermodynamic conditions. As illustrated schematically in Figure 3-13, computational materials design (or discovery) entails quantum mechanical (QM) calculations for determining electronic properties as well as molecular and crystal structures, and statistical mechanical (SM) calculations for

predicting the physiochemical properties of materials under macroscopic environments. A large number of computational methods are available for predicting electronic, molecular and thermodynamic properties[117]. Typically, these computational methods are established in the context of drastically different theoretical frameworks, and approximations are inevitable at each level. As a result, multiscale modeling requires a judicious selection of QM and SM methods, and an optimal choice reflects not necessarily only in terms of the theoretical rigor at individual scales but, perhaps more important, a reliable connection between different methods and the combined computational efficiency to attain pragmatic goals.

Amid numerous computational tools in QM and SM, density functional theory (DFT) provides a unified mathematical framework to describe the properties of many-body systems using the one-body density profiles as the fundamental variables. Whereas the original concepts, as first introduced many years ago by Pierre Hohenberg and Walter Kohn[118], were intended to provide an alternative to the Schrödinger equation for predicting ground-state electronic properties, the mathematical framework have been extended to electronic systems at finite temperature and to statistical-mechanical systems including those consisting of classical particles and polymers[119]. While nowadays DFT is best known for its applications to electronic systems at 0 K due to its extreme popularity in computational chemistry and materials science[120], the practicality of classical DFT for studying adsorption and other interfacial phenomena has been well established[30]. For example, classical DFT programs are routinely used to analyze experimental data for characterization of porous materials by gas adsorption[121]. Because DFT calculations avoid

explicit multi-body wave functions for quantum systems or microstates for thermodynamic systems containing classical particles, its computational efficiency is by far superior to wave-function-based QM methods and conventional molecular simulations in SM. The computational efficiency and flexibility in density functionals make DFT an ideal choice in multiscale modeling for both QM and SM calculations[122].

In a previous work[123], we demonstrated that the classical DFT could be used for rapid screening a large library of metal-organic frameworks (MOFs) potentially useful for H_2 storage. Using extensive Monte Carlo (MC) simulation data as the benchmark, we tested the theoretical performance of four versions of free-energy functionals that are commonly used in the literature to describe the thermodynamic properties of inhomogeneous simple fluids[43,49,75,124]. All these functionals incorporate the modified fundamental-measure theory (MFMT)[35] to account for molecular excluded volume effects but they are different in representing the attraction part of the excess free energy. In stark contrast to previous calibrations with model systems containing the Lennard-Jones (LJ) fluids in idealized pores, we find that the theoretical predictions are sensitive to specific gas components (e.g., N_2 adsorption versus to hydrogen storage) and thermodynamic conditions. In this work, we test the performance of different classical DFT methods for predicting methane adsorption in model slit pores as well as in a large library of hypothetical nanostructured materials. In addition to the adsorption isotherms, we consider methane diffusivity in top-performed MOFs using an extended excess-entropy scaling scheme proposed in our earlier work[125]. After extensive calibrations with simulation results, also performed in this work, the DFT methods have been applied to identify promising MOF materials for methane storage and

possible alternations of the operation conditions to attain the adsorption target set by the U. S. Department of Energy (DOE) for in vehicle nature gas storage.

3.3.2. Molecular Model and Theory

The theoretical models for predicting adsorption isotherms and diffusivity have been reported in our previous publications[123,125]. For completeness, here we recapitulate only the basic ideas of classical DFT and the computational procedures. Supplementary Materials present the key equations from the four different versions of the excess free energy functional used in this work and additional results from the classical DFT predictions.

I. Molecular model

To describe gas adsorption in MOF materials, we need to define, in addition to temperature T and pressure P of the gas phase, both bonded and non-bonded interactions among the MOF atoms, and pair interactions between MOF-gas and gas-gas molecules. As in our previous works, we assume that all MOF materials considered in this work are sufficiently rigid so that their crystalline structures are unchanged by gas adsorption. The crystallographic information files (CIFs) of MOF materials are adopted from the hypothetical library constructed by Wilmer et al.[73]

To calculate methane adsorption in a specific MOF framework, we use a $2\times2\times2$ supercell and consider its interaction with methane molecules with the periodic boundary conditions (PBC). As usual, methane molecules are represented by the Lennard-Jones (LJ) particles, with parameters, ε_{CH4}/k_B=148.0 K and σ_{CH4}=3.73 Å, obtained from the TraPPE

force field[126]. For each gas molecule inside the MOF material, the potential energy at position **r** is approximated by pairwise-additive interactions with all solid atoms

$$V^{\text{ext}}(\mathbf{r}) = \sum_{i \in \text{MOF}} u_{if}(\mathbf{r} - \mathbf{r}_i) \tag{73}$$

where subscript i represents the i^{th} atom from the MOF framework, and $\mathbf{r}_i$ stands for the position of atom i. The pair potential between atom i and gas molecule f, $u_{if}(\mathbf{r})$, is also represented by the LJ potential

$$u_{if}(r) = 4\varepsilon_{if}\left[\left(\frac{\sigma_{if}}{r}\right)^{12} - \left(\frac{\sigma_{if}}{r}\right)^{6}\right] \tag{74}$$

where parameters ε_{if} and σ_{if} are calculated from the Lorentz-Berthelot (LB) mixing rule. In both simulations and classical DFT calculations performed in this work, the LJ parameters for the MOF atoms are taken from the universal force field (UFF)[60].

II. Adsorption thermodynamics

Classical DFT allows us to predict the local number density of gas molecules within the adsorbent phase, *viz.*, inside each MOF material[127]

$$\rho(\mathbf{r}) = \rho_b \exp\left[-\beta V^{\text{ext}}(\mathbf{r}) - \beta\Delta\mu^{\text{ex}}(\mathbf{r})\right] \tag{75}$$

where ρ_b stands for the number density of gas molecules in the bulk, $\beta = 1/(k_B T)$, and k_B is the Boltzmann constant. The last term on the right side of Eq. (75) is defined as $\Delta\mu^{\text{ex}}(\mathbf{r}) = \mu^{\text{ex}}(\mathbf{r}) - \mu_b^{ex}$, which represents the deviation of local excess chemical potential $\mu^{\text{ex}}(\mathbf{r})$ from that corresponding to the bulk phase, μ_b^{ex}. Thermodynamic properties of the

gas phase, including ρ_b and μ_b^{ex} as functions of T and P, are calculated from the modified Benedict-Webb-Rubin (MBWR) equation of state[56]. In general, the local chemical potential depends on local density profile $\rho(\mathbf{r})$ and must be determined from Eq.(75) self-consistently by iterations.

Intuitively, Eq.(75) may be understood as the Boltzmann law for the distribution of gas molecules in the presence of an external field except that, within the framework of classical DFT, this equation is formally exact. Approximations are introduced only in calculating the local excess chemical potential, which is derived from the functional derivative of excess Helmholtz energy $F^{ex}(\mathbf{r})$

$$\mu^{ex}(\mathbf{r}) = \delta F^{ex}(\mathbf{r}) / \delta\rho(\mathbf{r}). \tag{76}$$

Like the excess properties of a uniform thermodynamic system, $F^{ex}(\mathbf{r})$ accounts for the thermodynamic non-ideality arising from intermolecular interactions. The excess Helmholtz energy is an intrinsic property of the gas system independent of the external potential, i.e., its interaction with the adsorbent atoms. If the intermolecular interaction between gas molecules is negligible, the excess Helmholtz energy disappears; and Eq.(75) corresponds to the distribution of ideal-gas molecules inside the porous material.

Whereas accurate expressions have been developed for $F^{ex}(\mathbf{r})$ of inhomogeneous simple fluids below the vapor-liquid critical temperature[43,128], little is known about their performance at high temperatures and pressures, i.e., conditions relevant to gas adsorption in typical industrial processes. In this work, we consider four versions of classical DFT that have been used before for predicting hydrogen adsorption in various MOF materials. All

these functionals are based on the modified fundamental measure theory (MFMT)[35] to describe molecular excluded volume effects but differ in the contribution to account for intermolecular attractions. The attractive part of the excess Helmholtz energy functional is distinguished in terms of the mean-field approximation (MFA)[75], the first-order mean-spherical approximation (FMSA)[43], and two slightly different forms of weighted density approximations (WDA-Y[124] and WDA-L[49]). The equations for these functionals are explained in Supplementary Materials.

The total amount of adsorption is conventionally expressed as the volume of adsorbed gas molecules per unit volume of the porous material at the standard state (STP). The adsorption amount can be calculated from the molecular density profile by integration over the supercell

$$G = \frac{k_B T_0}{P_0 V} \int \rho(\mathbf{r}) \, d\mathbf{r} \tag{77}$$

where $P_0 = 1\,\text{atm}$, $T_0 = 25\,^{\circ}\text{C}$, and V is the supercell volume. For applications related to gas storage, we are more interested in the delivery amount, G_{del}, which corresponds to the change in total gas uptake by the material at compression and release conditions. The total amount of gas adsorption or delivery amount should not be confused with the surface excess, defined as

$$G^{ex} = \frac{k_B T_0}{P_0 V} \int [\rho(\mathbf{r}) - \rho_b] \, d\mathbf{r} \tag{78}$$

In general, the local number density of gas molecules inside a nanostructured material is highly inhomogeneous due to extreme confinement and attraction from adsorbent atoms.

Therefore, the difference between absolute and excess adsorption is negligible only when the pressure of the bulk gas is at very low.

Given an expression for the excess Helmholtz energy functional, we can readily derive the excess entropy of gas molecules inside the porous material from

$$S^{\mathrm{ex}} = -\left(\frac{\partial F^{\mathrm{ex}}}{\partial T}\right)_{\beta\mu,V} \tag{79}$$

and the heat of adsorption from

$$Q = -T\left(\frac{\partial \Delta\Omega}{\partial T}\right)_{\beta\mu,V} \tag{80}$$

In Eq.(80), $\Delta\Omega$ correspond to the change in the grand potential for the adsorption of gas at constant temperature and the chemical potential:

$$\Delta\Omega = k_B T \int d\mathbf{r}\, \rho(\mathbf{r})\{\ln[\rho(\mathbf{r})/\rho_b] - 1 - \beta\mu_b^{ex} + Z + \beta V_{ext}(\mathbf{r})\} + F^{ex}[\rho(\mathbf{r})] \tag{81}$$

where $Z = Z(T,P)$ stands for the compressibility factor of the bulk gas. While heat of adsorption is not studied in this work, we note in passing that gas pressure inside a porous material is a second-order tensor and various with the position. As a result, the adsorbed gas does not have an enthalpy as that typically defined in classical thermodynamics of bulk fluids. Confusion might be inevitable if one calculates the heat of adsorption from enthalpy changes.

III. Excess entropy scaling for gas diffusivity

As in our previous work[125], we assume that self-diffusivity of gas molecules in a porous medium D_s may be determined from a linear combination of those from the Knudsen diffusion model and the excess-entropy scaling

$$\ln D_{\mathrm{s}} = \left(1 - \frac{\alpha\sigma^3}{v_{\mathrm{free}}}\right) \ln D_K + \frac{\alpha\sigma^3}{v_{\mathrm{free}}} \ln D_{\mathrm{E}} \tag{82}$$

where v_{free} is the total accessible volume of the porous material divided by the number of confined gas molecules[85], α is a constant reflecting the maximum molecular packing density, and σ is the LJ diameter of the gas molecules. In Eq.(82), $\alpha\sigma^3 / v_{\mathrm{free}}$ may be understood as the fraction of the free space inside the pore occupied by gas molecules. In the limit of the closest packing, we may estimate α from that for spherical particles as in a face-centered-cubic lattice

$$\alpha_{FCC} = \sqrt{2} / 2 \tag{83}$$

Eq.(83) provides a first-order estimate for α if no additional information is available. For best prediction of the simulation or experimental data, however, α should be used as an adjustable parameter. While this parameter may be slightly different for different systems depending on the crystal structures, it is independent of the thermodynamic conditions of the bulk gas including temperature and pressure.

The Knudsen model applies to gas diffusion at infinite dilution

$$D_K = \lim_{r_b \to 0} D_{\mathrm{s}} \tag{84}$$

where the limiting value can be obtained from MD simulation at low gas pressures. Excess entropy scaling accounts for contributions due to gas-gas interactions inside the pore, which can be described by Rosenfeld's scaling rule[81]

$$D_{\mathrm{E}} = \frac{0.585}{r_{av}^{1/3}} \sqrt{\frac{k_{\mathrm{B}} T}{m}} \exp\left(0.788 S^{\mathrm{ex}} / N k_B\right) \tag{85}$$

where m represents the molecular mass, r_{av} is the average number density of gas molecules inside the porous material, and N represents the total number of adsorbed gas molecules. With the excess entropy calculated from the classical DFT, the hybrid method provides an efficient procedure to predict the self-diffusion coefficient of gas molecules in nanoporous materials as a function temperature and bulk pressure[125].

3.3.3. Results and Discussion

I. Adsorption in slit pores

The four classical DFT methods considered in this work have been calibrated in previous publications by extensive comparison with simulation data for inhomogeneous Lennard-Jones (LJ) fluids in model systems such as near a planar wall or in slit pores. At conditions near vapor-liquid equilibrium, it has been shown that the mean-field approximation (MFA) fails to reproduce the depletion of a LJ liquid near a hard or weakly attractive surface while the first-order means-spherical approximation (FMSA) performs rather well in comparison with simulation[43,128]. Surprisingly, all free-energy functionals including MFA are highly accurate for predicting the adsorption of LJ fluids in slit pores. Regrettably, previous publications were mostly focused on comparisons for model systems at low temperatures (<200 K) that have little relevance to methane storage.

In this work, the classical DFT methods are first tested with simulation results for methane adsorption at room temperature (298 K) in slit pores of different widths and surface energies. Figure 3-14 shows various theoretical predictions for the total amount of adsorption and comparison with simulation data. We find that FMSA overestimates the total adsorption amount if the surface is only weakly attractive, while the opposite is true

for the MFA predictions. Conversely, both weight-density approximations (WDA-L and WDA-Y) agree excellent with the simulation data at all conditions (Figure 3-14a). For adsorption in more attractive slit pores (Figure 3-14b), the three versions of non-mean-field methods give virtually identical results, but MFA underestimates the total adsorption amount in comparison to the simulation data. The poor performance of MFA is somewhat anticipated because it neglects the correlation effects due to van der Waals attractions; inclusion of the correlation contribution would increase the total adsorption amount.

II. Adsorption and delivery capacities

The performance of different DFT methods for methane adsorption in porous materials may not be fully consistent with that for simple systems due to extreme confinements. To illustrate, we consider methane adsorption in two representative MOF materials, MOF-5[129] and Cu-BTC[130], again at room temperature. Figure 3-15 compares the total adsorption amount predicted from different classical DFT methods with GCMC simulation. Overall, the four versions of classical DFT methods show good agreement with the simulation data. Nevertheless, their accuracies are slightly different for different materials. In stark contrast to their performance in slit pores, MFA and FMSA are more accurate than WDA-L and WDA-Y for gas adsorption in MOF-5 at room temperature over the entire range of testing pressures. Both WDA methods overestimate the total adsorption amount especially at intermediate pressures. For methane adsorption in Cu-BTC, however, WDA-Y reproduces GCMC data very well. While WDA-L and FMSA slightly underestimate the storage capacity at moderate pressures, the total adsorption amount predicted by MFA is noticeably smaller than those form other DFT methods at high gas

loadings. As indicated in our previous study[123], the performance of classical DFT methods for gas adsorption in nanoporous materials depends also on the testing conditions.

For potential applications to materials discovery, the classical DFT methods should be calibrated in the context of multiscale modeling and benchmarked with a large set of experimental data. While novel computational methods for predicting the crystalline structures of framework materials and their interactions with gas molecules are fast emerging[131], much work remains to be done for calibration of multiscale modeling methods. In this work, we confine our interest in comparing the theoretical predictions with GCMC data for methane adsorption in over 1,000 MOF materials at various thermodynamic conditions. The crystal structures of these materials are from the Northwestern hypothetic MOF database[73], generated according to the topology of some common SBUs. These materials have been studied before for methane storage with GCMC simulations and thus provide a good training ground to test the performance of different classical DFT methods for materials screening.

Figure 3-16 shows the theoretical results predicted by MFA in comparison with the GCMC data. Similar comparisons for other classical DFT methods are presented in Supplementary Materials. In addition to the total adsorption amount at 298 K and 35 bar, we consider also the methane delivery amount for each material when the pressure is reduced to 5 bar at the same temperature. The thermodynamic conditions for the bulk gas are in line with the ARPA-E target for methane storage. Table 3-2 summarizes the root-mean-square deviations (RMSD) and theory-simulation correlation coefficients (the R values) for different classical DFT methods. In general, all four classical DFT methods

show rather good agreements with the GCMC data for the total amount of methane adsorption at the ARPA-E target condition. Surprisingly, MFA gives the best results in comparison to the simulation data, with RMSD=7.90 cm^3(STP)/cm^3 and R^2=0.99. Because this functional is mathematically simpler in comparison to FMSA and weighted-density approximations, it is numerically more stable and easier to converge in solving for the molecular density profiles. Although other methods show slightly larger deviations, the RMSD values are still within about 6% of the total adsorption amount.

The original target set by ARPA-E for CH_4 storage states that, at 298 K and 35 bar, the volumetric storage capacity should exceed 180 v/v for qualified nanoporous materials. Under such condition, the adsorbed natural gas (ANG) will have an energy density comparable to that of the compressed natural gas (CNG) at the same temperature and 250 bar[73]. While that target can be easily reached by many MOF candidates, none of the existing materials meet the new ARPA-E target for methane storage in transportation vehicles[132]. The ARPA-E target specifies not only the total uptake amount but also the methane delivery amount by nanoporous materials between 35 bar and 5 bar at room temperature. In Figure 3-16b, we compare the methane delivery amount predicted by MFA and GCMC methods. Generally speaking, the MFA predictions agree very well with the simulation data, especially for medium and high loading MOF materials. However, deviation becomes more noticeable for materials with high total amount of adsorption but low delivery amount. Considering the accuracy of MFA prediction for the total methane storage capacity at 298 K and 5 bar (Figure 3-16a), the large discrepancy may be attributed to numerical issues in our classical DFT calculations. Because those MOFs have large

adsorption capabilities at both 35 and 5 bar, the delivery amount is small and sensitive to numerical details.

III. Diffusivity

Classical DFT gives the density profile of gas molecules in a porous material that reflects the microscopic details of adsorbent-adsorbate interactions. Not only is the density profile related to the adsorption isotherm, it can also be used to predict other thermodynamic properties including heat of adsorption and excess entropy. While the heat of adsorption is of tremendous importance for practical applications in its own right, the excess entropy provides a convenient starting point for the quantitative prediction of transport coefficients over a broad range of thermodynamic conditions.

In a previous work[125], we have demonstrated that Rosenfeld's excess-entropy scaling method can be combined with the Knudsen model to predict the self-diffusivity of gas molecules in various nanoporous materials. Because the excess entropy is readily available from adsorption calculations and the Knudsen model applies to gas diffusion at infinite dilution, the new computational procedure is much faster than traditional MD simulation methods for calculating the gas self-diffusivity over a broad range of thermodynamic conditions.

In Figure 3-17, we present the self-diffusion coefficients of CH_4 in two MOF materials, MOF-5 and Cu-BTC, at 298 K. The theoretical results are calculated according to Eq.(82). With α treated as an adjustable parameter ($a = 0.42$), the extended excess entropy scaling method is able to reproduce the simulation data near quantitatively. Semi-quantitative results are attainable with approximation $a \approx a_{FCC}$.

Figure 3-18 shows the self-diffusivities of CH_4 in top five MOFs that show the highest volumetric delivery amount as identified below (see Table 3-4). Here temperature is fixed at 298 K, and α is treated as an adjustable parameter for each material. Again, the extended excess entropy scaling method reproduces the simulation data very well over the entire range of pressures, and approximation $a \gg a_{FCC}$ leads to a semi-quantitative agreement between theory and MD simulation predictions. Interestingly, the MOF materials with top delivery amounts show similar gas diffusivity at both low and high loadings, but their relative ranks are slightly different.

IV. Characteristics of "good" adsorbents

Conventional thermodynamic models for gas adsorption such as Langmuir and BET isotherms presume that, at a given thermodynamic condition, the total gas uptake is linearly proportional to the surface area (or accessible surface area) of the adsorbent. Whereas ambiguity might arise in its definition (and measurement) for a particular porous material, the surface area has been routinely used as an important measure for the identification of promising MOFs with high adsorption capability. The significance of surface area is rarely challenged even with the widespread applications of simulation methods.

To find possible correlations between gas adsorption capability and materials surface area, we plot the methane adsorption amounts versus the accessible surface areas[133,134] of 1,200 MOFs materials at 298 K and 35 bar. While Figure 3-19a shows some positive correlation for the total adsorption amount, no trend could be observed from Figure 3-19b between the accessible surface areas and the delivery amounts. More importantly, if

we focus on the top MOF candidates for methane storage as shown in Figure 3-20, neither the simulation results nor the classical DFT predictions exhibit any convincing relationship between those two properties. The decoupling of the delivery amount from the (accessible) surface area is somewhat anticipated. Although the geometric accessible surface areas are calculated based on the locus of the surface potential minimum, it accounts for only one gas molecule a time. The surface area does not take into consideration the correlation between adsorbed gas molecules, which is important especially for candidates with high loading capacities. The BET surface area is correlated with the adsorption amount because it is obtained from fitting to the gas adsorption data. In other words, the good correlation between BET surface area and adsorption capability should not be considered surprising.

According to Figure 3-16 and S3-1, a material with high methane adsorption capacity does not necessarily lead to a large delivery amount at the ARPA-E conditions. While the adsorption capacity appears in good correlation with the void fraction, such correlation does not exist for the top candidates. Conversely, Figure 3-16b shows that MOFs with high CH_4 weight adsorption amount coincide with those with high delivery amount, suggesting that it may serve as a good indicator to identify promising materials.

V. ARPA-E MOVE target

The methane delivery amounts for all MOF materials considered in this work are below 160 cm^3/cm^3, far from the ARPA-E target 315 cm^3/cm^3. To achieve the volumetric delivery amount, we could either increase the gas uptake at the compressed stage or reduce the remnant amount upon release. Table 3-3 summarizes the performance of the top 10 MOFs if we set the initial pressure to 250 bar, the same as that in a compressed natural gas

tank (CNG). According to this table, the top materials identified from GCMC are in good agreement with those from various classical DFT methods. As far as the top candidates are of concern, the list generated from FMSA agrees the best with that from the simulation, with 9 out of 10 promising MOFs identical.

If the initial compression pressure is set to 250 bar, many MOFs meet the target for the weight delivery amount of 0.5 g/g, and the best material is able to achieve up to 95% of the volumetric target value. If we raise the release temperature to 358 K, the ARPA-E volumetric target may also be reached (see Table S3-2). Alternatively, the ARPA-E target can also be reached by reducing the initial storage temperature to 233 K, the lowest operation temperature according to the DOE instruction. In that case, many MOFs could also reach the ARPA-E target by compression at 75 bar and release at 298 K and 5 bar (see Table S3-3). At higher release temperature, all the top 10 MOFs achieve the ARPA-E methane storage target in terms of both gravimetric and volumetric measures (see Table 3-3).

Unlike the results shown in Table 3-2, the DFT and GCMC methods predict rather different top 10 candidates to meet the ARPA-E target (see Table 3-3). Nevertheless, all these materials exhibit similar adsorption capabilities. Besides, the GCMC candidates are among the top 50 materials identified by the classical DFT methods. In both Tables 3-2 and 3-3, there are about 3% numerical discrepancies between DFT and GCMC data. Considering usually 5% statistical error in GCMC simulations, we believe the classical DFT methods should be sufficiently accurate to identify promising MOF materials for further experimental investigations.

3.3.4. Conclusions

We have calibrated four versions of classical density functional theory (DFT) for potential applications to the discovery of novel metal-organic frameworks (MOFs) for methane storage. Overall the classical DFT methods show good agreement with the simulation results for methane adsorption in slit pores as well as in MOF-5 and Cu-BTC. Nevertheless, their relative theoretical performance is sensitive to thermodynamic conditions and depends on a specific material under consideration. Using the ARPA-E methane delivery capabilities of over 1,200 MOFs as the benchmark, we find that, surprisingly, the results from the mean-field approximation (MFA) agree best with the simulations data, with a root-mean-square deviation (RMSD) of only 7.90 $cm^3(STP)/cm^3$. Considering its theoretical simplicity, we conclude that MFA is probably the best choice for screening nanostructured materials for CH_4 storage.

Whereas the BET surface area is conventionally used as a benchmark to identify the total adsorption capacity of a specific material, we find that, in general, there is no convincing correlation between the surface area and the net delivery amount at conditions relevant to methane storage. While none of the existing materials satisfies the ARPA-E target, many MOFs can have a weight delivery amount exceeding 0.5 g/g if the initial compression pressure is modified to 250 bar, the same as that in a compressed natural gas tank. The volumetric target may be reached if the release temperature is raised to 358 K. Alternatively, the ARPA-E target can be fulfilled by reducing the initial storage temperature to 233 K, the lowest operation temperature according to the DOE instruction.

In comparison to conventional molecular simulations, a key advantage of the classical DFT methods is that they are extremely efficient for calculating a large library of materials over a broad parameter space. Figure 9 shows the statistics of computational time for implementing the classical DFT calculations on a single desktop computer (i.e., one 3.0 GHz Sandy Bridge CPU core). The average computational time is within 2 minutes for each sample, which is faster than that of conventional GCMC simulation by more than one order of magnitude. More importantly, DFT provides quantitative information on the excess entropy that can be used for predicting gas self-diffusivities, while simulations of the transport properties of confined gases are extremely time-consuming with the conventional methods. The good theoretical performance and computational efficiency make classical DFT methods an ideal choice for high-throughput calculations.

Figure 3-13. Computational materials design involves quantum mechanical (QM) calculations for understanding the chemistry of the building blocks and microscopic structure and statistical-mechanical (SM) calculations to predict their performance under diverse thermodynamic conditions. Because multiple approaches are available for both QM and SM calculations, multiscale modeling should be calibrated with extensive experimental data in terms of both theoretical accuracy and computational efficiency.

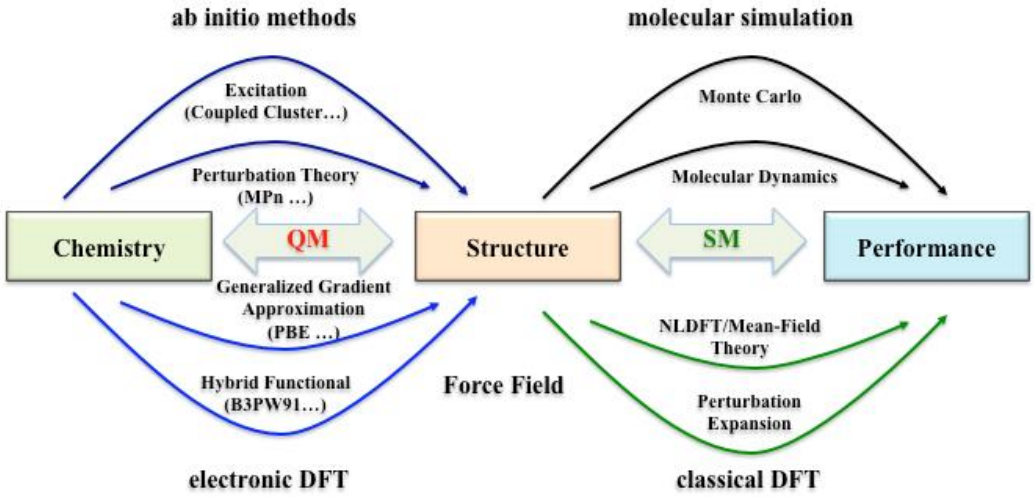
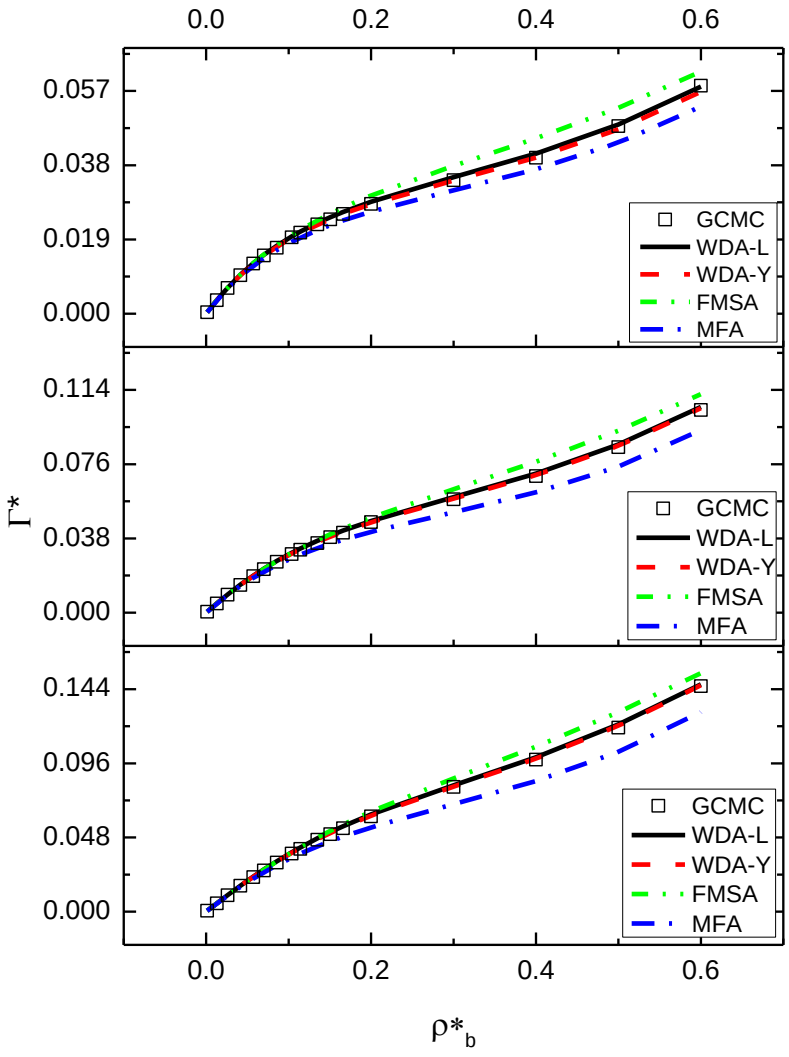
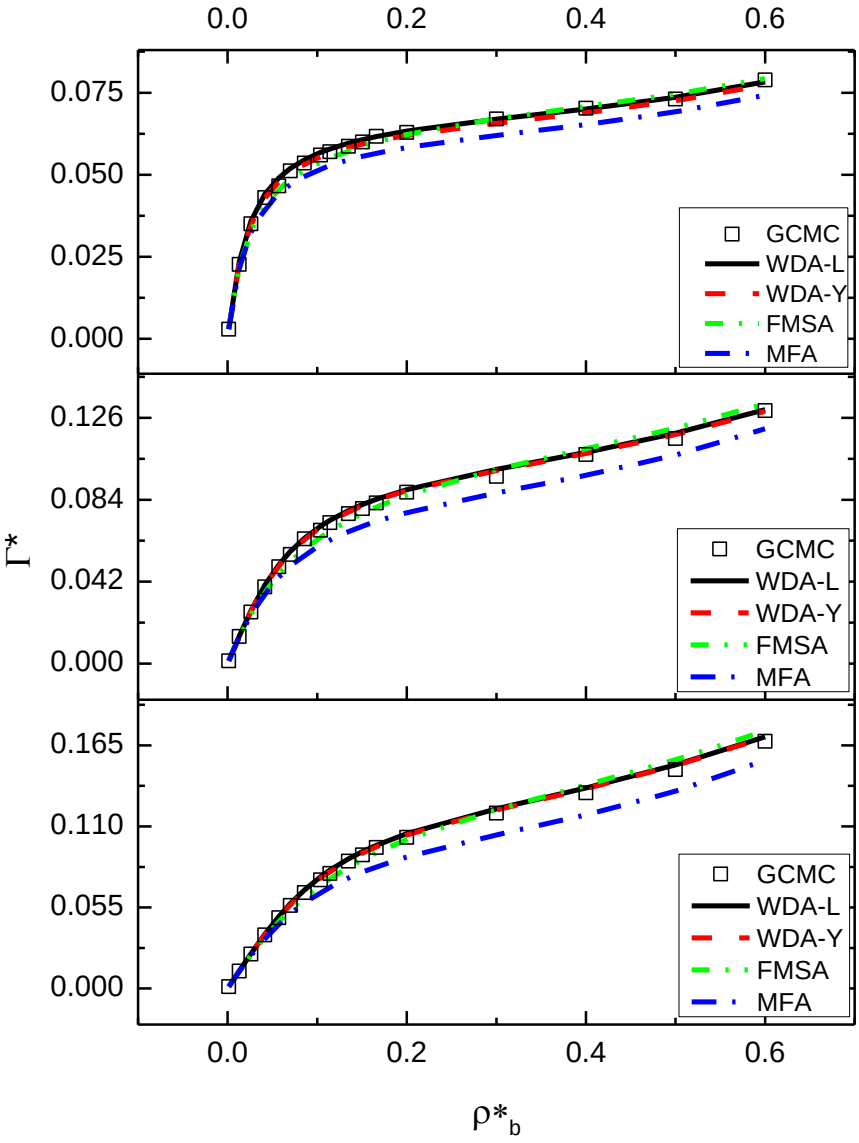


Figure 3-14. a) Methane adsorption in slit pores of reduced width H/σ =2, 3, 4 from top to bottom at 298K. The interaction between the gas molecule and each wall is described with Steele's 10-4-3 potential, where $\sigma_w = \sigma$, $\varepsilon_w = 3.142\varepsilon$ and $\Delta = 0.7071\sigma$; b) The same as a) except $\varepsilon_w = 6.283\varepsilon$.



a)


0.0
0.2
0.4
0.6
0.075
0.050
0.025
0.000
0.126
0.084
0.042
0.000
0.165
0.110
0.055
0.000
Γ*
GCMC
WDA-L
WDA-Y
FMSA
MFA
ρ*b

b)

Figure 3-15. Methane adsorption isotherms at 298K for a) MOF-5 and b) Cu-BTC framework materials calculated from GCMC simulation and from different classical DFT methods.

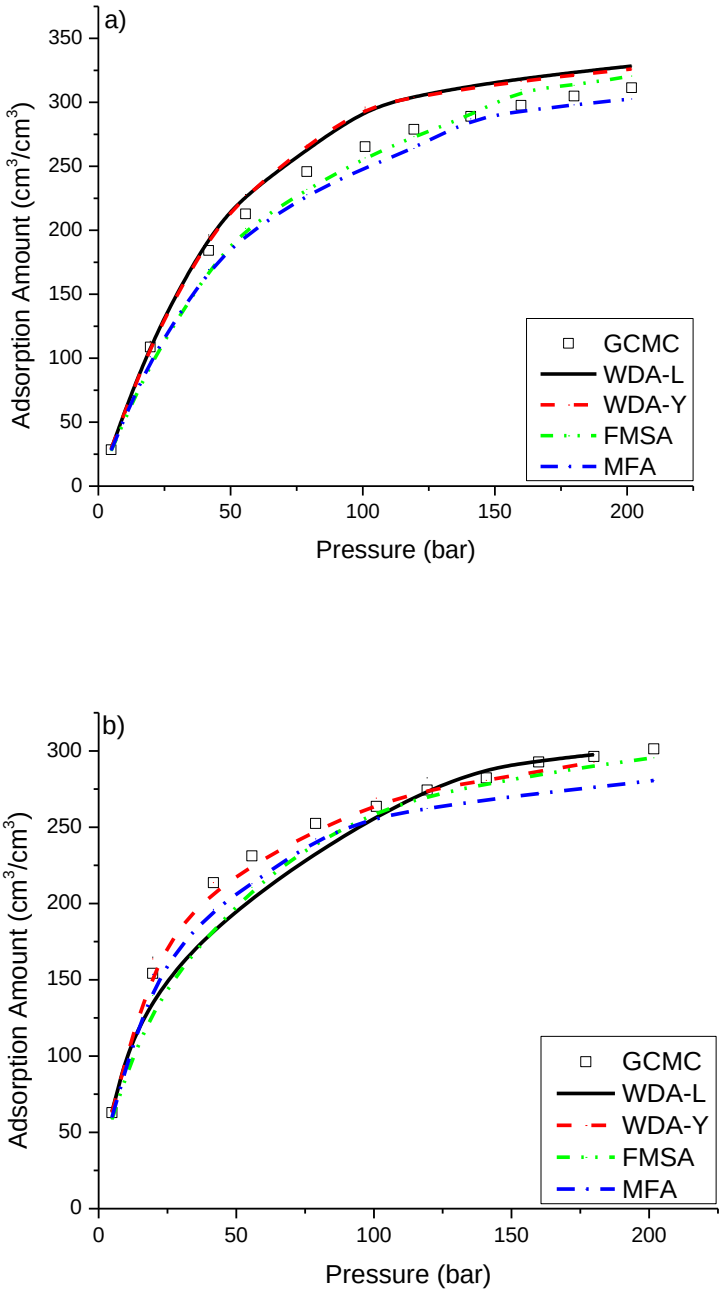
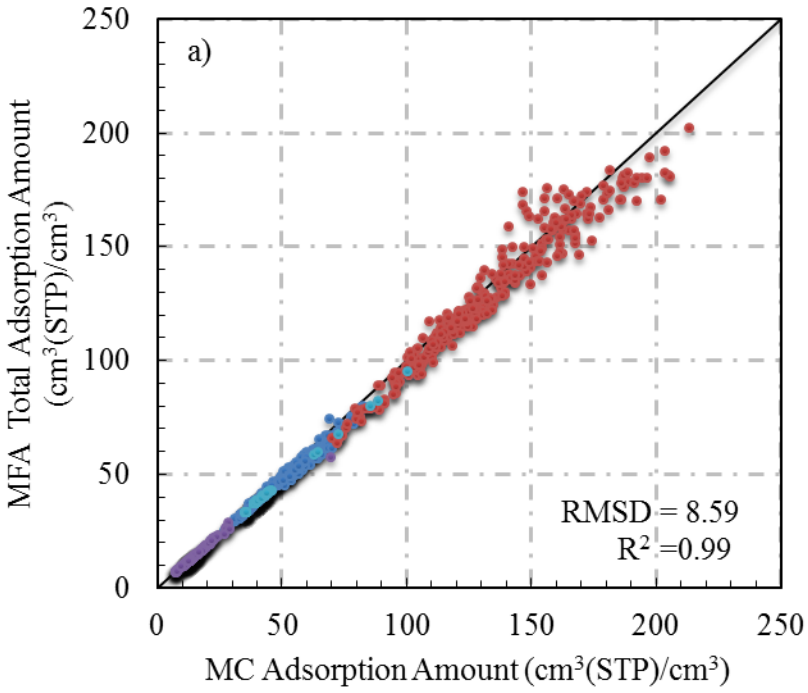


Figure 3-16. Comparison of methane adsorption at 298 K in 1,200 MOFs calculated from GCMC and from MFA methods. a) Total adsorption amount at 5 bar; b) Delivery amount between 35 bar and 5 bar. Color code: Navy blue, top 300 from excess CH_4 adsorption in weight category; Red, top 300 from excess CH_4 adsorption in volume category; Purple, top 300 from void fraction category; Sky blue, top 300 from surface area (m^2/cm^3) category from the Northwestern Hypothetical MOF Database.



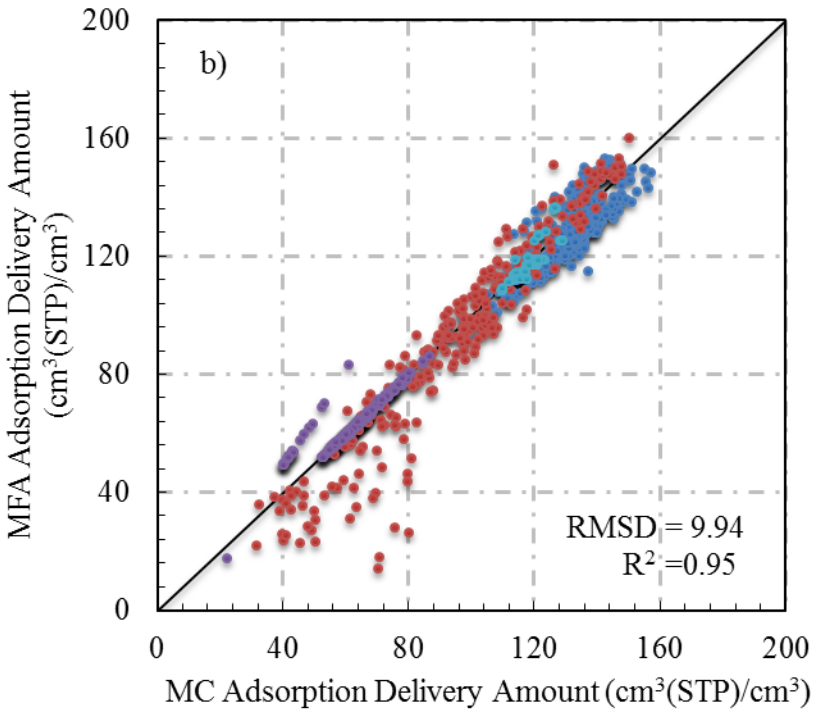

b)
MFA Adsorption Delivery Amount (cm³(STP)/cm³)
MC Adsorption Delivery Amount (cm³(STP)/cm³)
0
40
80
120
160
200
RMSD = 9.94
R² =0.95

Figure 3-17. Self-diffusivities of CH_4 molecules in MOF-5 and Cu-BTC at 298 K predicted by MD simulation and by the extended excess-entropy scaling method, respectively. Filled dots are results from MD simulations, while the lines are obtained from the extended excess-entropy scaling method.

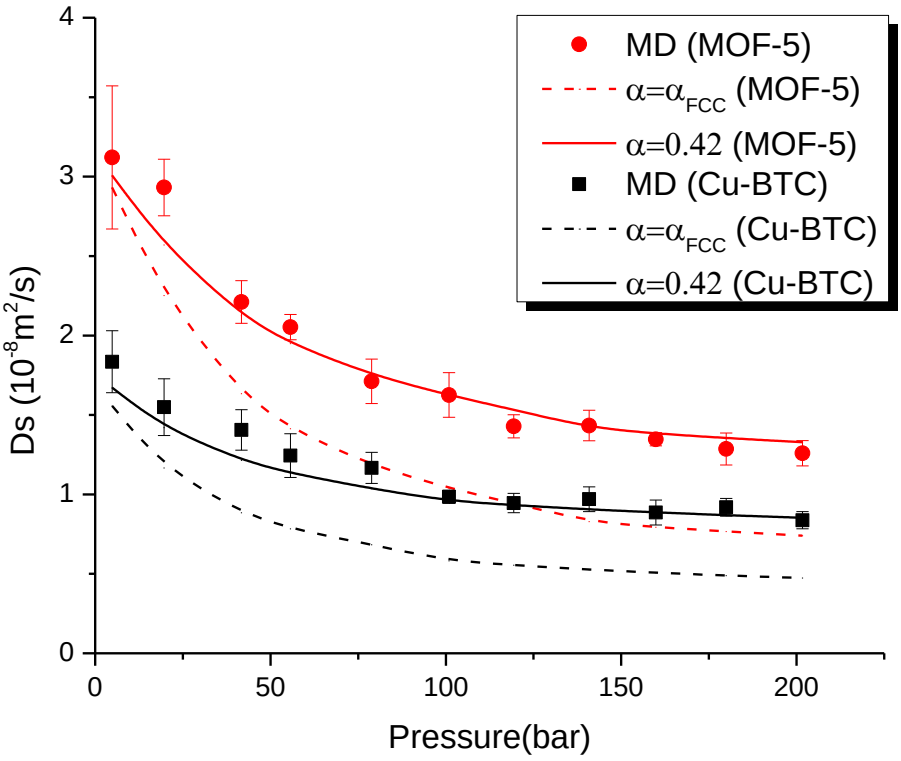


Figure 3-18. Self-diffusivities of CH_4 in top 5 MOFs with the highest volumetric delivery amount for compression at 233 K and 75 bar and release at 358 K and 5 bar predicted by FMSA method. The symbols are self-diffusion coefficients predicted by MD simulation, and the lines are from the extended excess-entropy scaling method.

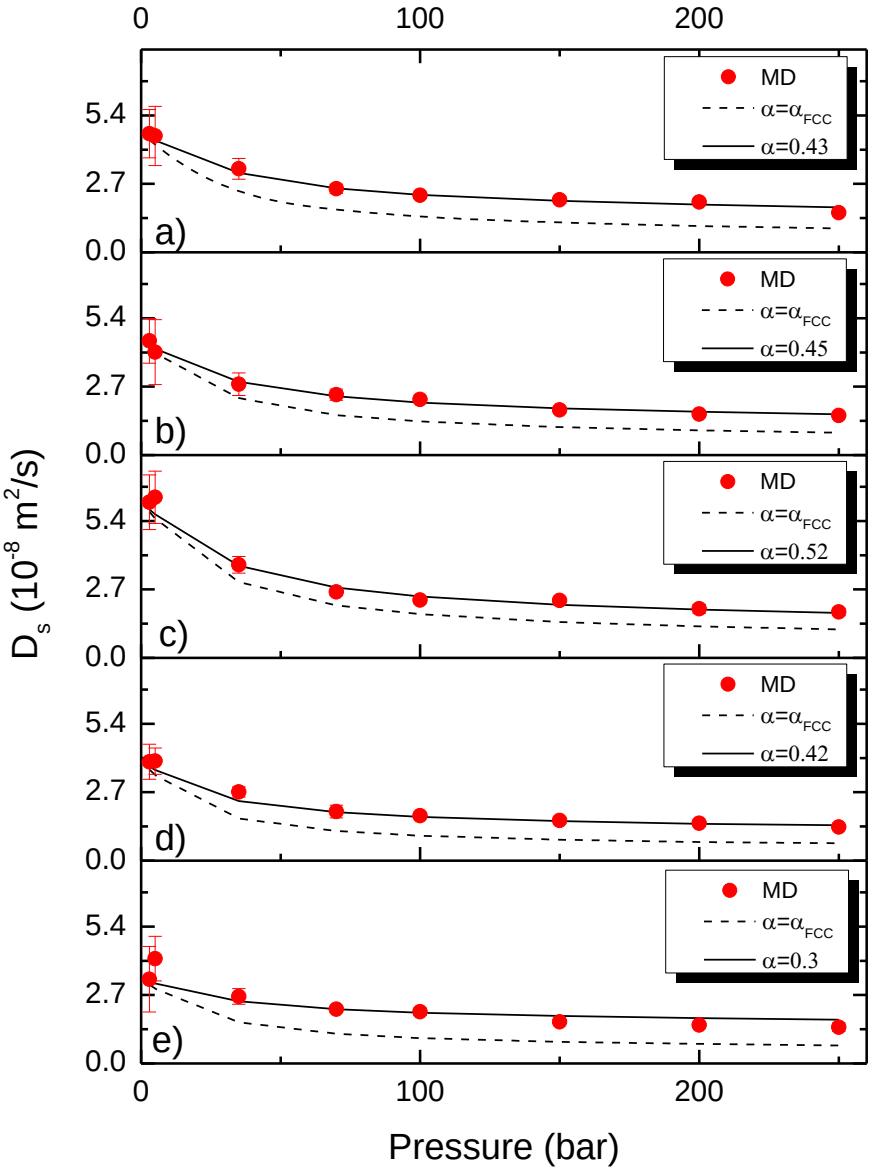


Figure 3-19. The adsorption amounts of methane in 1,200 MOFs at 298 K and 35 bar as a function of the accessible surface area calculated according to the MFA and GCMC methods. a) Total adsorption amount; b) Delivery amount. Green dots: DFT data; Red dots: GCMC data.

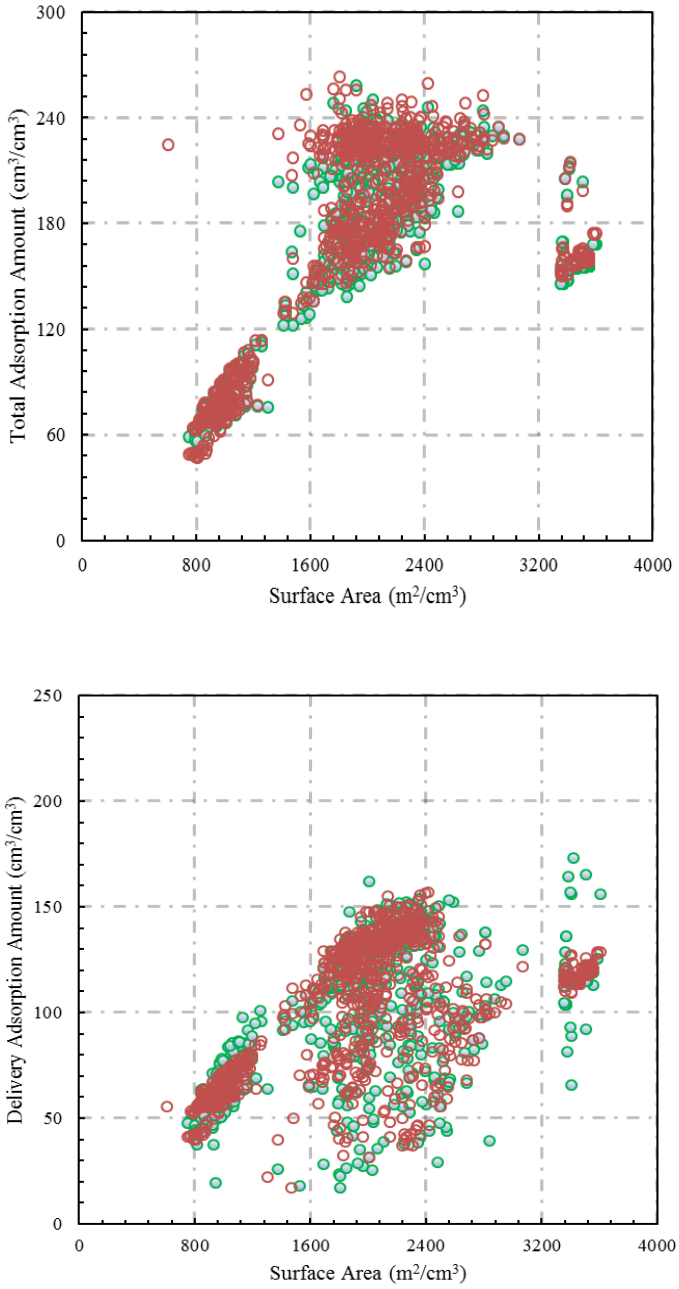


Figure 3-20. Top 10 MOF candidates with the highest adsorption amounts of methane at 298 K and 35 bar as a function of the accessible surface area. Cycles: predicted by GCMC; Diamonds: predicted by MFA, while filled ones stand for total adsorption amounts and blank ones for delivery amounts between 298 K, 35 bar and 298 K, 5 bar

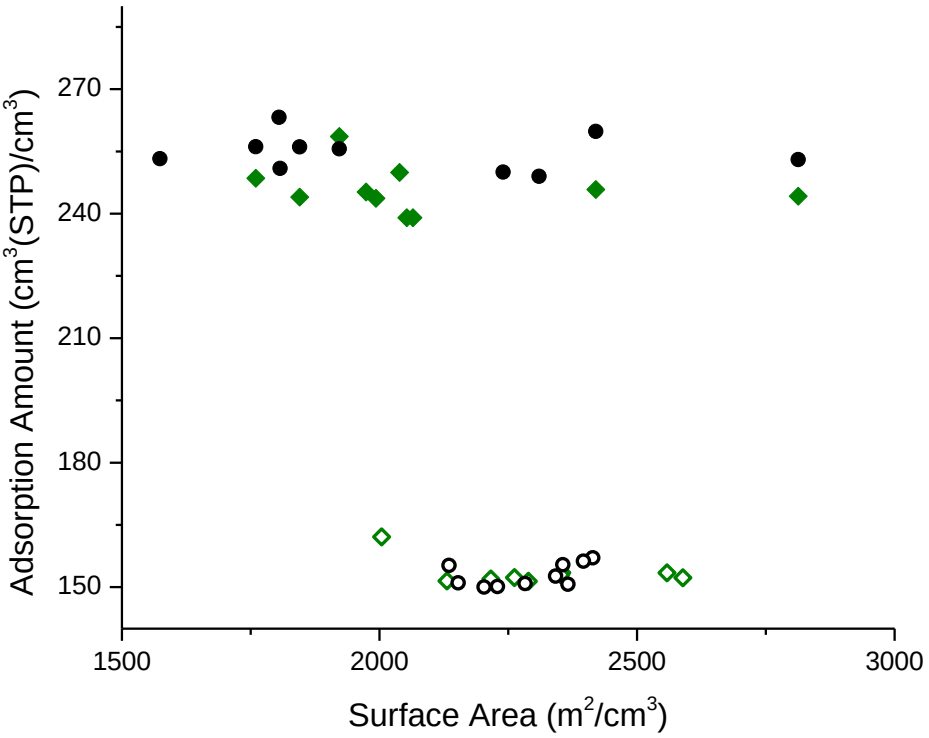


Figure 3-21. The distribution of computational time for predicting methane adsorption in 1,200 MOFs with MFA.

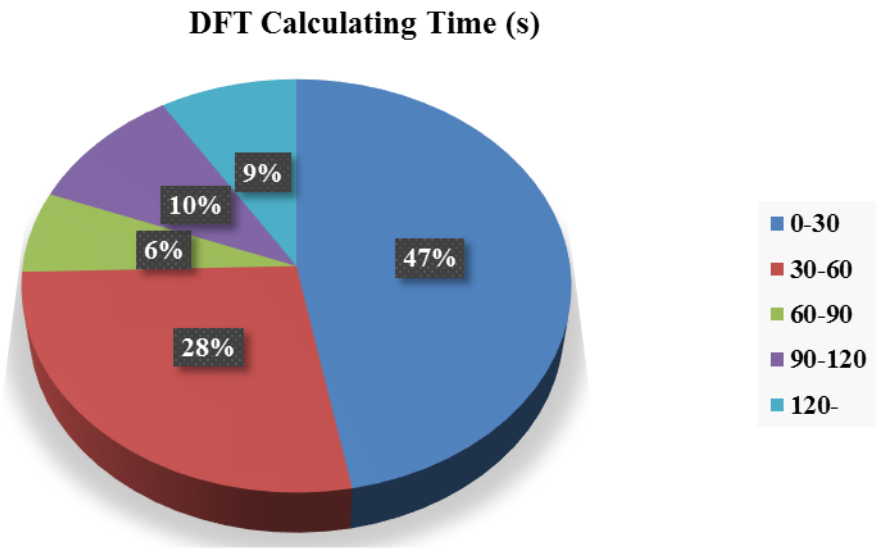


Table 3-2 Root-mean-square deviation (RMSD) and correlation coefficients (R-value) for various classical DFT methods in comparison with GCMC simulation data for methane adsorption in 1,200 MOFs at 298 K and 35 bar.

	FMSA	**MFA**	**WDA-Y**	**WDA-L**
RMSD cm^3(STP)/cm^3	16.45	7.90	15.76	16.58
R^2	0.94	0.99	0.98	0.97

Table 3-3 Top 10 MOF candidates identified by GCMC and by various classical DFT methods according to the volumetric delivery amount. The methane delivery amount is calculated between compression at *T*=298 K, *P*=250 bar and release at T=298 K, P=5 bar.

	FMSA			**MFA**			**WDA-Y**[1]			**GCMC**		
Rank	**No.**	**$\Gamma_{del,V}$**	**$\Gamma_{del,M}$**	**No.**	**$\Gamma_{del,V}$**	**$\Gamma_{del,M}$**	**No.**	**$\Gamma_{del,V}$**	**$\Gamma_{del,M}$**	**No.**	**$\Gamma_{del,V}$**	**$\Gamma_{del,M}$**
1	1101	300.5	0.63	1033	272.1	0.45	1213	307.4	0.66	1031	290.1	0.57
2	1089	299.1	0.62	1101	271.8	0.57	1101	306.7	0.64	1089	290.1	0.60
3	1213	298.7	0.64	1031	271.6	0.53	1089	305.2	0.64	1101	290.0	0.61
4	1031	297.5	0.58	1213	271.1	0.58	1031	304.1	0.59	1213	289.8	0.62
5	1111	295.7	0.58	1089	270.9	0.56	1111	303.2	0.59	1111	287.1	0.56
6	1052	295.6	0.60	1026	269.7	0.42	1025	301.9	0.58	1034	287.0	0.56
7	1145	295.2	0.60	1111	269.5	0.53	1145	301.8	0.61	1086	286.8	0.58
8	1025	294.5	0.57	1064	268.7	0.50	1052	300.6	0.61	1030	286.5	0.55
9	1086	294.2	0.59	1025	268.7	0.52	1233	300.4	0.73	1145	286.5	0.58
10	1034	294.1	0.57	1052	268.6	0.54	1086	299.9	0.60	1052	286.2	0.58

*: The No. represents the serial number in the Northwestern Hypothetical MOF Database, while the 1st digit means the category (1 for excess CH_4 adsorption in weight category; 2 for excess CH_4 adsorption in volume category; 3 for void fraction category; 4 for surface area (m^2/cm^3) category), the last 3 digits means the rank in the category. Γ_{del} means delivery amount, where V stands for adsorption amount in units of $cm^3(STP)/cm^3$ and M stands for adsorption amount in units of g/g.

[1]: Since the two WDA methods give nearly the same results, we only give WDA-Y data here.

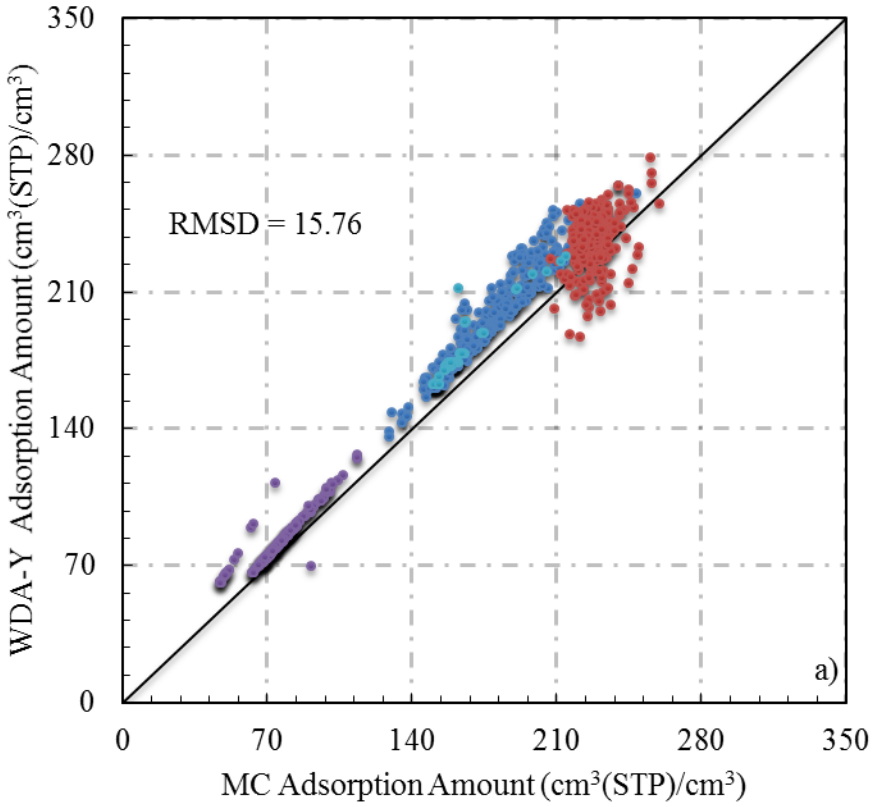
Table 3-4 Top 10 MOF candidates identified by GCMC and by various classical DFT methods according to the volumetric delivery amount between compression at T=233 K, P=75 bar and release at T=358 K, P=5 bar.

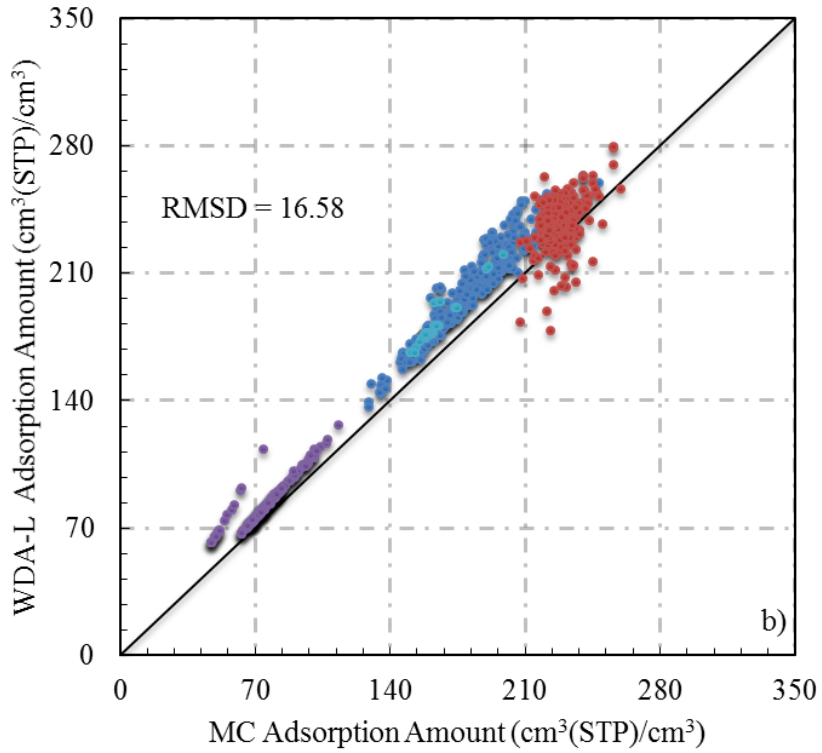
	FMSA			MFA			WDA-Y			GCMC		
Rank	**No.**	**$\Gamma_{del,V}$**	**$\Gamma_{del,M}$**	**No.**	**$\Gamma_{del,V}$**	**$\Gamma_{del,M}$**	**No.**	**$\Gamma_{del,V}$**	**$\Gamma_{del,M}$**	**No.**	**$\Gamma_{del,V}$**	**$\Gamma_{del,M}$**
1	1031	321.4	0.63	1033	315.5	0.53	1111	353.9	0.69	1162	336.1	0.66
2	1037	318.7	0.59	1011	315.2	0.49	1025	352.1	0.68	1026	335.8	0.53
3	1111	318.4	0.62	1022	314.4	0.50	1064	349.2	0.65	1011	335.5	0.52
4	1002	318.2	0.55	1026	313.9	0.49	1034	348.9	0.68	1007	335.3	0.59
5	1007	318.2	0.56	1013	312.7	0.50	1041	348.2	0.68	1145	335.3	0.68
6	1055	318.1	0.57	1046	312.2	0.45	1030	347.9	0.67	1031	334.7	0.65
7	1025	317.7	0.61	1098	311.7	0.47	1055	346.0	0.62	1019	334.6	0.51
8	1033	317.6	0.53	1044	311.7	0.46	1071	345.9	0.60	1001	334.2	0.58
9	1010	317.5	0.55	1019	311.0	0.47	1086	345.8	0.70	1071	334.1	0.58
10	1011	317.3	0.49	1197	310.5	0.48	1033	345.1	0.57	1087	333.9	0.59

Support Information

I. Methane Adsorption in MOFs

Figure S3-1 Comparison of CH_4 absolute adsorption amount at 298 K and 35 bar in 1,200 MOFs calculated from GCMC and from different classical DFT methods: a) WDA-Y b) WDA-L c) FMSA. Color code: Navy blue, top 300 from excess CH_4 adsorption in weight category; Red, top 300 from excess CH_4 adsorption in volume category; Purple, top 300 from void fraction category; Sky blue, top 300 from surface area (m^2/cm^3) category from the Northwestern Hypothetical MOF Database.




RMSD = 16.58
b)
WDA-L Adsorption Amount (cm³(STP)/cm³)
MC Adsorption Amount (cm³(STP)/cm³)
0
70
140
210
280
350

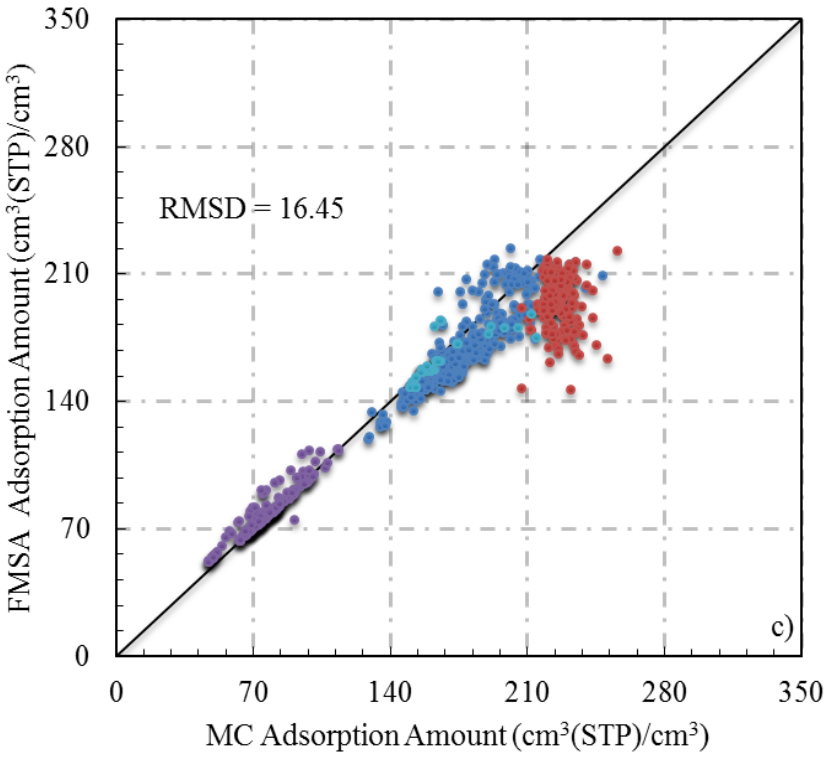

RMSD = 16.45
c)
FMSA Adsorption Amount (cm³(STP)/cm³)
MC Adsorption Amount (cm³(STP)/cm³)
0
70
140
210
280
350

Table S3-2 Top 10 MOF candidates identified by GCMC and by various classical DFT methods according to the CH_4 volumetric delivery amount between compression at T=298 K, P=250 bar and release at T=358 K, P=5 bar.

	FMSA			MFA			WDA-Y			GCMC		
Rank	**No.**	**$\Gamma_{del,V}$**	**$\Gamma_{del,M}$**	**No.**	**$\Gamma_{del,V}$**	**$\Gamma_{del,M}$**	**No.**	**$\Gamma_{del,V}$**	**$\Gamma_{del,M}$**	**No.**	**$\Gamma_{del,V}$**	**$\Gamma_{del,M}$**
1	1101	314.8	0.66	1026	293.7	0.46	1101	322.3	0.68	1031	309.0	0.60
2	1089	313.6	0.65	1033	292.9	0.49	1213	322.1	0.69	1111	307.0	0.60
3	1031	313.6	0.61	1013	292.6	0.46	1111	322.0	0.63	1089	307.0	0.64
4	1033	312.4	0.52	1011	292.5	0.46	1031	321.7	0.63	1030	306.8	0.59
5	1111	312.2	0.61	1022	292.4	0.47	1089	321.1	0.67	1032	306.7	0.52
6	1213	311.7	0.67	1019	291.3	0.44	1025	321.0	0.62	1101	306.7	0.64
7	1025	311.2	0.60	1197	290.7	0.45	1064	319.9	0.60	1011	306.4	0.48
8	1026	311.2	0.49	1028	289.7	0.44	1145	318.6	0.65	1034	306.2	0.60
9	1064	310.9	0.58	1046	289.4	0.42	1053	317.9	0.60	1086	305.5	0.62
10	1011	310.8	0.48	1098	289.2	0.44	1011	317.6	0.49	1213	305.2	0.66

*: The No. stands for the serial number in the Northwestern Hypothetical MOF Database, while the 1st digit means the category (1 for excess CH_4 adsorption in weight category; 2 for excess CH_4 adsorption in volume category; 3 for void fraction category; 4 for surface area (m^2/cm^3) category), the last 3 digits means the rank in the category. Γ_{del} means delivery amount, where V stands for adsorption amount in units of $cm^3(STP)/cm^3$ and M stands for adsorption amount in units of g/g

Table S3-3 Top 10 MOF candidates predicted by GCMC and various DFT methods according to CH_4 volumetric delivery amount between compression at T=233 K, P=75 bar and release at T=298 K, P=5 bar.

	FMSA			MFA			WDA-Y			GCMC		
Rank	**No.**	**$\Gamma_{del,V}$**	**$\Gamma_{del,M}$**	**No.**	**$\Gamma_{del,V}$**	**$\Gamma_{del,M}$**	**No.**	**$\Gamma_{del,V}$**	**$\Gamma_{del,M}$**	**No.**	**$\Gamma_{del,V}$**	**$\Gamma_{del,M}$**
1	1031	305.3	0.60	1033	294.7	0.49	1111	335.2	0.66	1145	317.1	0.64
2	1101	302.0	0.63	1031	292.3	0.57	1025	333.0	0.64	1089	317.0	0.66
3	1111	301.8	0.59	1055	290.5	0.52	1034	331.1	0.65	1162	317.0	0.62
4	1025	301.0	0.58	1022	290.2	0.46	1041	330.8	0.65	1213	316.8	0.68
5	1089	300.9	0.63	1111	289.9	0.57	1064	329.0	0.62	1101	316.3	0.66
6	1213	300.5	0.65	1026	289.9	0.45	1030	328.9	0.63	1031	315.7	0.62
7	1037	300.1	0.55	1011	289.8	0.45	1086	328.3	0.66	1297	312.4	0.53
8	1052	299.7	0.60	1098	287.6	0.44	1055	324.6	0.58	1064	312.1	0.59
9	1055	299.5	0.54	1064	286.8	0.54	1188	324.4	0.67	1111	311.7	0.61
10	1034	299.3	0.58	1071	286.8	0.50	1297	323.7	0.55	1055	311.6	0.56

3.4 Seeking Metal-Organic Frameworks (MOFs) for Methane Storage in Natural Gas Vehicles

Concerns over the national energy security and global climate change are calling for the development of alternative fuels that have minimal carbon footprint yet reliable supplies. The low cost and huge reserve of natural gas (NG) make it an attractive choice as a midtermsolution. NG is mainly composed of methane, which has a high research octane number (RON=107) and contributes no SO_x and NO_x emissions.[135] One of the key challenges for its broader applications to natural gas vehicles (NGV) lies in its storage and delivery at ambient conditions. Conventional methods, including Liquefied Natural Gas (LNG) and Compressed Natural Gas (CNG)[136], involve either cryogenic cooling (112 K) or high pressure (>200 bar) operations with heavy-duty tanks that are expensive and hazardous. Recent years have witnessed tremendous developments in Adsorbed Natural Gas (ANG) method based on activated carbons and nanostructured porous materials such as metal-organic frameworks (MOFs). In particular, ANG method is most promising for NGV applications owing to its excellent performance at moderate operating conditions. [130,136-146]

Unlike amorphous porous materials such as activated carbons, MOFs have crystalline structures that are predicable on the basis of the modular building blocks. [129,147-149] Besides, these materials can be synthesized at the industrial scale. [150] A wide variety of MOFs have been reported potentially useful for CH_4 storage. Prominent examples include NU-111, [151] NU-125, [152] UTSA-20, [153] PCN-14, [154] NiMOF-74. [155] Recently, Yaghi's group demonstrated that MOF-519[156] provides an impressive methane delivery amount of

230 cm^3/cm^3 when it is operating between 80 bar and 5 bar at 298 K. While that capacity is higher than all MOFs presently known, it does not yet meet the ARPA-E target,[132] which requires volumetric delivery amount of 315 $cm^3(STP)/cm^3$ (corresponding to energy density 12.5 MJ/L, the same as CNG does at 250 bar 298K) and weight delivery amount of 0.5 g/g, considering 25% packing loss.

The number of MOFs that have been tested for methane storage is trivially small in comparison to the huge variety of the nanostructures that could be synthesized by changing the metal-organic building blocks. A key question of practical concern is whether it is possible to identify potential nanoporous materials to meet the DOE target. Because experimental synthesis and test for a large number of materials is tedious and time consuming, computational methods such as the classical molecular simulation is a welcome choice in the materials community.[63,138,157-161] For example, Wilmer *et al.*[24,162] used grand canonical Monte Carlo (GCMC) simulation to predict methane adsorption in about 100,000 hypothetical MOFs and identified a group of promising MOF materials for tailored synthesis. By studying the structure-property relationships, Wilmer *et al.* also explored the limits of these nanoporous materials for methane storage and delivery. Regrettably, it appears that none of the possible MOFs meets the ARPA-E target, at least for gas storage at ambient temperature (298 K) and moderate pressure (< 65 bar). In this work, we use GCMC to explore possible alternatives to reach the ARPA-E target by changing the operational temperatures and pressures. Our goal is to identify important indicators from the perspective of methane delivery and optimal MOFs for practical NGV applications.

We consider over 1,000 MOF candidates from the Northwestern Hypothetical MOF Database that were previously identified as promising for methane storage.[24] These MOF materials have been categorized according to their rankings with different criteria. As in a previous work,[163] we start with GCMC calculations for methane adsorption at 298 K and 35 bar specified in the ARPA-E target. The computational details are given in Support Information (SI). Figure 3-22 presents both the total and delivery adsorption capacity (between compression at 298 K and 35 bar and release at 298K and 5bar). In the volumetric adsorption categories (see Figure 3-22a and 3-22c), our GCMC results indicate that none of those MOF materials meet the ARPA-E target of 260 cm^3(STP)/cm^3 (9.2 MJ/L, corresponding to CNG at 250 bar without considering the 25% packing loss) even in terms of the total storage capacity. Ironically, MOFs with high total volumetric adsorption often have small deliverable amount, typically less than 20% of the total uptake at 35 bar. Due to the strong attraction of these materials with CH_4 molecules, the large adsorption amount persists at both high and low pressures. By contrast, MOF materials with moderate interaction strength with methane molecules are most likely to have high delivery amount. The simulation results show no clear correlation between the weight adsorption capacity and total volumetric storage (see also Tables S3-4~S3-13 in SI).

We find that MOFs with the highest volumetric delivery amount (Figure 3-22c) mostly coincide with those having the largest weight adsorption capacity (Figure 3-22b), even though the total adsorption amount may not necessarily be at the top. In terms of the gravimetric adsorption amount (Figure 3-22b and 3-22d), the total and delivery capacities follow similar trends. In particular, the MOF with the highest total gas uptake coincides

that with the best delivery capacity. However, the top candidates in the weight category rank poorly in terms of the volumetric adsorption, less than 40% in comparison to the top-ranked MOFs shown in Figure 3-22a and 3-22c. Surprisingly, those top candidates have very close total and delivery amounts in the volumetric category. It appears that the top-ranked materials in the weight category do not have many favorable binding sites, acting as if gas molecules were in a usual compressed tank. Considering all possible measures, we conclude that the volumetric deliverable amount is the most important indicator for materials selection for NGV applications. A similar conclusion was reached from recent experimental observations, from the perspective of driving range.[146] In the following discussions, we thus focus on the materials performance in terms of the volumetric deliverable amount.

The volumetric delivery amount can be raised either by increasing the gas uptake at the compression stage or reducing the remnant amount upon release. Figure 3-23 shows the performance of some top ranked MOFs in the volumetric category if we set the initial pressure higher than 35 bar. More extensive results are given in Tables S3-14~S3-23. Figure 3-23a shows that if we double the initial pressure (i.e., from 35 bar to 70 bar), both the volumetric and weight deliverable amounts increase by nearly 30%. However, the delivery capacity is only about 60% of the ARPA-E target, even for the top ranked MOF materials. The volumetric target (i.e., 260 $cm^3(STP)/cm^3$) could be reached only if the pressure is further increased to 170 bar. While the volumetric target appears unattainable, the weight deliverable amount has already exceeded the 0.5 g/g target, suggesting that 170 bar could be taken as the upper limit for studying the adsorbed methane storage capacity.

At this pressure, most MOFs approach the saturation platform of their CH_4 storage capacity, i.e., further increasing the pressure does not have much effect on the volumetric delivery amount.

Figure 3-23c shows that when the pressure is increased to 250 bar, the same as the common CNG gas tank, the total adsorption amount is 50% higher than that in Figure 3-23b. However, the volumetric delivery amount is increased less than 5%. Snurr *et al*. suggested that the MOF materials could reach the ARPA-E target only if the attractive energy between the gas and substrate is quadrupled and if the release temperature is increased to 398 K.[24] Considering that the tolerated temperate range for CH_4 operation is between -40-85℃ in the ARPA-E project, we have investigated also the effect of two temperature boundaries on methane adsorption. In one case, we increase the adsorption amount by lowering the temperature at the compression stage to 233 K. In another case, we raise the delivery temperature to 358 K so that the remnant gas adsorption could be minimized. Figure 3-23d shows the top 10 MOF candidates with the new temperature (233 K) and pressure (75 bar) at the compression stage. Over half of these materials achieve the ARPA-E target in both volumetric and weight delivery categories. The reduction in compression temperature greatly improves the methane volumetric adsorption performance, much better than that at 298 K and 250 bar.

Table 3-5 gives a list of MOF candidates that yield the maximum deliverable amount at various operational conditions (more data details could be found in Tables S3-14~S3-33). We may attain the weight delivery target by either increasing the pressure or decreasing the temperature at the compression stage. However, the volumetric target

cannot be reached solely by increasing the delivery temperature. The volumetric delivery amount is sensitive to the delivery temperature at moderate pressure (<100 bar). Although none of the MOF materials we studied could meet the ARPA-E requirement, over 90% of the demand could be satisfied at 298 K and 170 bar if we increase the delivery temperature to 358 K. We can further increase the volumetric delivery capacity by decreasing the compression temperature. For example, many MOF materials could achieve the ARPA-E target at a relatively low pressure (<50 bar) if the compression and delivery temperatures are set at 233 K and 358 K, respectively.

In summary, we have investigated a large number of promising MOF materials and operational conditions that may meet the ambitious DOE ARPA-E target for methane storage. We propose that new materials development should be focused on improving the volumetric delivery amount rather than the weight delivery or other criteria that often quoted in the literature. While none of MOF materials known today meets the DOE standards, the ARPA-E target could be reached by modifying the operational condition for gas compression and delivery. If methane is compressed at 298 K and 170 bar and released at 358 K and 5 bar, the top ranked MOF is able to deliver over 90% of the required methane amount. A large number of MOF materials may achieve the ARPA-E target at higher delivery and lower compression temperatures.

Figure 3-22 Top 10 MOF materials accordingly to their performance at 298K and 35bar in terms of a) Total volumetric adsorption amount; b) Total weight adsorption amount; c) Volumetric delivery amount (released at the same temperature and 5bar); d) Weight delivery amount. In each panel, the horizontal axis gives the serial numbers in the Northwestern Hypothetical MOF Database[24]; the 1st digit represents the category of ranking, i.e., 1 for excess CH_4 adsorption in the weight category; 2 for excess CH_4 adsorption in the volume category; and 5 for the void fraction category. The last 3 digits of each serial number denote the ranking within the individual category. Here black and red column bars stand for the total and delivery volumetric adsorption amounts, blue and pink lines correspond to the total and delivery weight adsorption amounts, respectively.

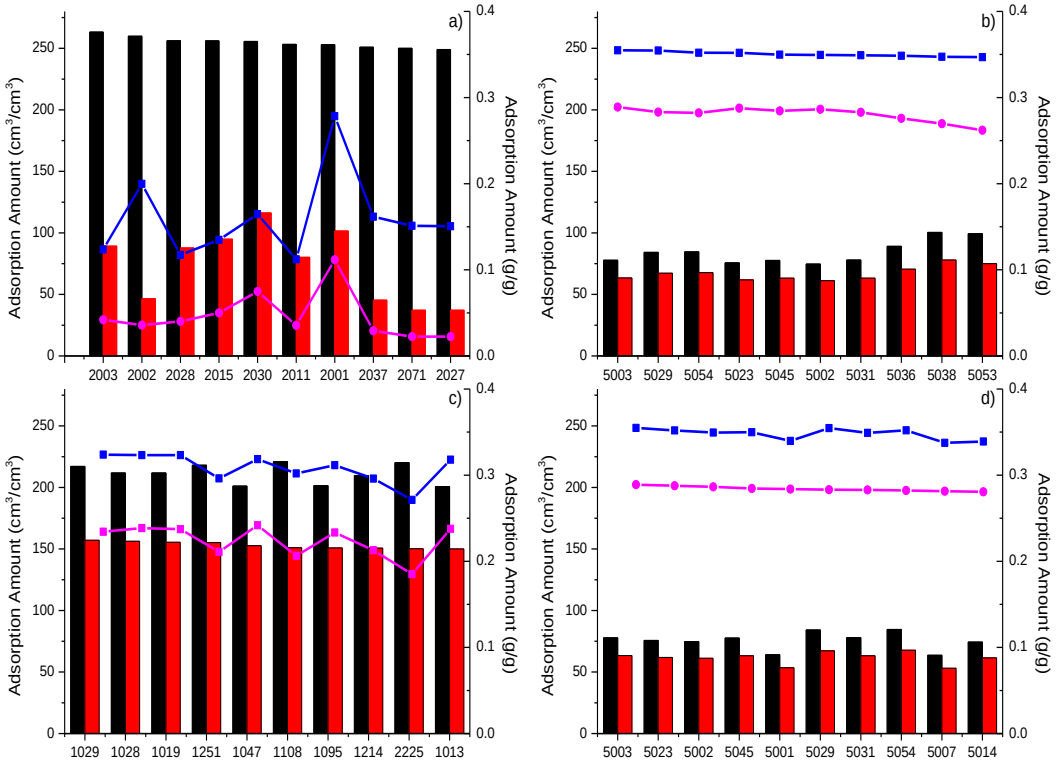

a)
b)
c)
d)
Adsorption Amount (cm3/cm3)
Adsorption Amount (g/g)
2003 2002 2028 2015 2030 2011 2001 2037 2071 2027
5003 5029 5054 5023 5045 5002 5031 5036 5038 5053
1029 1028 1019 1251 1047 1108 1095 1214 2225 1013
5003 5023 5002 5045 5001 5029 5031 5054 5007 5014

Figure 3-23 Top 10 MOF materials in terms of the volumetric delivery amount at 298K and 5 bar. The compressed pressures are a) 70bar; b) 170bar; c) 250bar; d) 75bar (233K). Dash lines correspond to the ARPA-E target, the golden line stands for the volumetric delivery target 315 $cm^3(STP)/cm^3$, while green line stands for the weight delivery target 0.5 g/g.

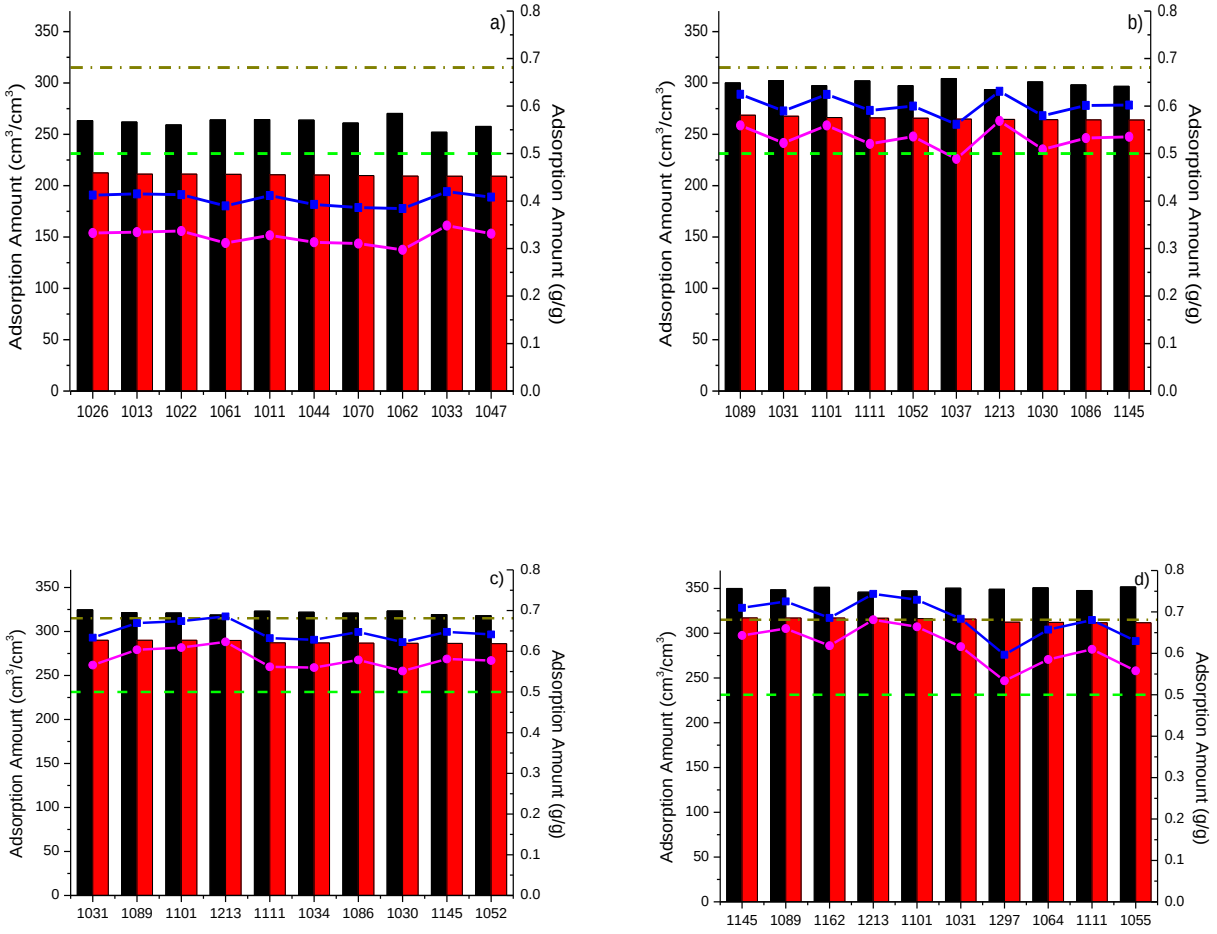


Table 3-5 MOF materials with the maximum adsorption capacity in different categories.

	298K				358K			
P[a)]	**ID**	**$\Gamma_{Del,V}$**[b)]	**ID**	**$\Gamma_{Del,M}$**[c)]	**ID**	**$\Gamma_{Del,V}$**[d)]	**ID**	**$\Gamma_{Del,M}$**[e)]
35	1029	157	5003	0.29	1226	206	5003	0.32
70	1026	212	5001	0.59	1062	246	5001	0.61
100	1033	234	5001	0.79	1046	264	5001	0.81
150	1089	260	5077	1.08	1026	286	5077	1.09
170	1089	268	5077	1.19	1197	290	5077	1.21
200	1031	277	5124	1.34	1011	297	5124	1.35
250	1031	290	5124	1.52	1031	309	5124	1.53
35[f)]	1033	280	5003	0.64	1028	311	5003	0.67
45[f)]	1026	291	5002	0.71	1028	321	5001	0.81
75[f)]	1145	317	5007	1.16	1162	336	5007	1.18

a) Pressure, in the units of bar; b) $\Gamma_{del,V}$, the delivery amount between this condition and 298K and 5bar, in the units of cm^3/cm^3; c) $\Gamma_{del,M}$, in the units of g/g; d) $\Gamma_{del,V}$, delivery amount between this condition and 358K and 5bar, in the units of cm^3/cm^3; e) $\Gamma_{del,M}$, in the units of g/g; f) the testing temperature is 233K.

Support Information

Table S3-4 Top 10 MOF materials with highest total volumetric adsorption amount at 298K and 35bar.

Rank	ID	Γ_{Tot}[2](cm^3/cm^3)	Γ_{Del}[3](cm^3/cm^3)	Γ_{Tot}[2](g/g)	Γ_{Del}[3](g/g)
1	2003[1]	263.20	89.27	0.12	0.04
2	2002	259.83	46.55	0.20	0.04
3	2028	256.14	87.82	0.12	0.04
4	2015	256.08	94.92	0.13	0.05
5	2030	255.60	116.23	0.16	0.07
6	2011	253.25	80.22	0.11	0.04
7	2001	253.01	101.56	0.28	0.11
8	2037	250.91	45.37	0.16	0.03
9	2071	250.02	37.17	0.15	0.02
10	2027	249.01	37.19	0.15	0.02

1: The No. means the serial number in the Northwestern Hypothetical MOF Database, while the 1st number means the category (1 for excess CH_4 adsorption in weight category; 2 for excess CH_4 adsorption in volume category; 3 for void fraction category; 4 for surface area (m^2/cm^3) category), the last 3 digits means the rank in the category.

2: Γ_{tot} means total adsorption amount at this condition

3:Γ_{del} means delivery amount between this condition and 298K and 5bar, so as the following tables.

Table S3-5 Top 10 MOF materials with highest total volumetric adsorption amount at 298K and 70bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	2003	292.26	118.33	0.14	0.06
2	2030	287.21	147.84	0.18	0.10
3	2015	286.54	125.38	0.15	0.07
4	2011	285.96	112.92	0.13	0.05
5	1226	281.71	171.31	0.33	0.20
6	2010	281.71	171.31	0.33	0.20
7	2001	281.52	130.06	0.31	0.14
8	2006	280.78	164.10	0.32	0.19
9	2028	280.23	111.92	0.13	0.05
10	1116	280.04	174.51	0.35	0.22

Table S3-6 Top 10 MOF materials with highest total volumetric adsorption amount at 298K and 100bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	2003	301.20	127.28	0.14	0.06
2	2015	300.34	139.17	0.16	0.07
3	2030	296.55	157.18	0.19	0.10
4	2006	295.77	179.10	0.34	0.20
5	1116	294.27	188.73	0.36	0.23
6	2013	294.27	188.73	0.36	0.23
7	2001	293.56	142.11	0.32	0.16
8	1226	293.22	182.82	0.35	0.22
9	2010	293.22	182.82	0.35	0.22
10	1156	290.66	186.24	0.35	0.23

Table S3-7 Top 10 MOF materials with highest total volumetric adsorption amount at 298K and 150bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	2003	311.99	138.06	0.15	0.06
2	2015	310.74	149.58	0.16	0.08
3	1023	308.40	234.08	0.43	0.33
4	2157	308.40	234.08	0.43	0.33
5	2030	307.51	168.14	0.20	0.11
6	1226	307.49	197.09	0.36	0.23
7	2010	307.49	197.09	0.36	0.23
8	1116	307.33	201.79	0.38	0.25
9	2013	307.33	201.79	0.38	0.25
10	1011	307.17	253.40	0.48	0.39

Table S3-8 Top 10 MOF materials with highest total volumetric adsorption amount at 298K and 170bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	2003	318.31	144.38	0.15	0.07
2	1048	315.04	250.58	0.44	0.35
3	1023	313.73	239.42	0.44	0.34
4	2157	313.73	239.42	0.44	0.34
5	1226	312.62	202.22	0.37	0.24
6	2010	312.62	202.22	0.37	0.24
7	1116	312.30	206.77	0.38	0.25
8	2013	312.30	206.77	0.38	0.25
9	2030	311.76	172.39	0.20	0.11
10	1108	311.54	241.60	0.43	0.33

Table S3-9 Top 10 MOF materials with highest total volumetric adsorption amount at 298K and 200bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1023	319.97	245.66	0.45	0.35
2	2157	319.97	245.66	0.45	0.35
3	1017	319.80	253.95	0.47	0.37
4	1011	319.00	265.22	0.50	0.41
5	2003	318.66	144.73	0.15	0.07
6	1020	316.85	261.65	0.48	0.40
7	1019	316.73	260.41	0.48	0.40
8	1108	316.46	246.51	0.43	0.34
9	1029	316.42	256.42	0.47	0.38
10	1036	316.18	255.45	0.46	0.37

Table S3-10 Top 10 MOF materials with highest total volumetric adsorption amount at 298K and 250bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1023	330.46	256.14	0.46	0.36
2	2157	330.46	256.14	0.46	0.36
3	1011	328.23	274.45	0.51	0.43
4	1019	327.31	270.99	0.50	0.41
5	1020	326.32	271.11	0.50	0.41
6	2215	326.02	250.07	0.41	0.32
7	1048	326.01	261.55	0.46	0.37
8	1032	325.69	281.31	0.56	0.48
9	1004	325.69	260.69	0.52	0.41
10	1026	325.46	274.58	0.51	0.43

Table S3-11 Top 10 MOF materials with highest total volumetric adsorption amount at 233K and 35bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	2003	336.80	162.88	0.16	0.08
2	2015	335.10	173.94	0.18	0.09
3	2011	334.55	161.51	0.15	0.07
4	1023	334.38	260.06	0.47	0.37
5	2157	334.38	260.06	0.47	0.37
6	1028	333.78	278.31	0.51	0.42
7	1043	333.69	261.14	0.46	0.36
8	2229	333.69	261.14	0.46	0.36
9	1144	332.52	238.01	0.43	0.30
10	2088	332.52	238.01	0.43	0.30

Table S3-12 Top 10 MOF materials with highest total volumetric adsorption amount at 233K and 45bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1028	343.96	288.48	0.53	0.43
2	2225	342.62	272.64	0.42	0.33
3	1023	342.56	268.24	0.48	0.37
4	2157	342.56	268.24	0.48	0.37
5	1017	342.42	276.56	0.50	0.40
6	1026	342.07	291.18	0.54	0.44
7	1048	341.13	276.67	0.48	0.38
8	2003	340.95	167.02	0.16	0.08
9	1116	340.56	235.02	0.42	0.28
10	2013	340.56	235.02	0.42	0.28

Table S3-13 Top 10 MOF materials with highest total volumetric adsorption amount at 233K and 75bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1019	357.54	301.21	0.55	0.46
2	1011	357.40	303.62	0.56	0.47
3	1026	356.78	305.90	0.56	0.48
4	1029	356.46	296.46	0.53	0.44
5	1048	356.20	291.74	0.50	0.41
6	1023	355.95	281.63	0.50	0.40
7	2157	355.95	281.63	0.50	0.40
8	1001	355.91	304.86	0.62	0.53
9	2226	355.84	255.95	0.41	0.29
10	1021	355.51	275.44	0.49	0.38

Table S3-14 Top 10 MOF materials with highest delivery volumetric adsorption amount between 298K, 35bar and 298K, 5bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1029	217.07	157.07	0.32	0.23
2	1028	211.75	156.27	0.32	0.24
3	1019	211.75	155.43	0.32	0.24
4	1251	218.05	155.23	0.30	0.21
5	1047	201.11	152.65	0.32	0.24
6	1108	220.97	151.02	0.30	0.21
7	1095	201.43	150.88	0.31	0.23
8	1214	209.60	150.71	0.30	0.21
9	2225	220.14	150.16	0.27	0.19
10	1013	200.71	150.03	0.32	0.24

Table S3-15 Top 10 MOF materials with highest delivery volumetric adsorption amount between 298K, 70bar and 298K, 5bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1026	263.27	212.38	0.41	0.33
2	1013	261.96	211.28	0.42	0.33
3	1022	259.22	211.21	0.41	0.34
4	1061	264.08	211.06	0.39	0.31
5	1011	264.42	210.64	0.41	0.33
6	1044	263.86	210.41	0.39	0.31
7	1070	261.05	209.87	0.39	0.31
8	1062	270.37	209.30	0.38	0.30
9	1033	252.08	209.22	0.42	0.35
10	1047	257.66	209.21	0.41	0.33

Table S3-16 Top 10 MOF materials with highest delivery volumetric adsorption amount between 298K, 100bar and 298K, 5bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1033	277.54	234.68	0.46	0.39
2	1026	284.87	233.99	0.45	0.37
3	1013	284.62	233.94	0.45	0.37
4	1197	281.07	232.80	0.44	0.36
5	1025	268.97	232.70	0.52	0.45
6	1031	266.69	232.12	0.52	0.45
7	1011	285.65	231.87	0.44	0.36
8	1047	280.14	231.68	0.44	0.37
9	1007	277.47	231.47	0.49	0.41
10	1061	284.22	231.20	0.42	0.34

Table S3-17 Top 10 MOF materials with highest delivery volumetric adsorption amount between 298K, 150bar and 298K, 5bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1089	291.76	260.31	0.61	0.54
2	1031	294.21	259.64	0.57	0.51
3	1101	290.33	259.29	0.61	0.54
4	1111	294.96	258.91	0.58	0.51
5	1025	294.13	257.87	0.56	0.50
6	1052	288.80	257.03	0.58	0.52
7	1034	291.78	256.78	0.57	0.50
8	1086	290.75	256.62	0.59	0.52
9	1030	292.80	256.03	0.56	0.49
10	1026	306.80	255.92	0.48	0.40

Table S3-18 Top 10 MOF materials with highest delivery volumetric adsorption amount between 298K, 170bar and 298K, 5bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1089	300.11	268.65	0.62	0.56
2	1031	302.22	267.65	0.59	0.52
3	1101	297.40	266.36	0.62	0.56
4	1111	302.01	265.96	0.59	0.52
5	1052	297.40	265.64	0.60	0.54
6	1037	304.33	264.84	0.56	0.49
7	1213	293.51	264.50	0.63	0.57
8	1030	301.08	264.32	0.58	0.51
9	1086	298.17	264.05	0.60	0.53
10	1145	296.75	263.94	0.60	0.54

Table S3-19 Top 10 MOF materials with highest delivery volumetric adsorption amount between 298K, 200bar and 298K, 5bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1031	312.35	277.78	0.61	0.54
2	1101	308.35	277.31	0.65	0.58
3	1213	305.84	276.83	0.66	0.60
4	1052	307.99	276.23	0.62	0.56
5	1089	306.75	275.29	0.64	0.57
6	1111	311.19	275.14	0.61	0.54
7	1041	309.61	275.08	0.61	0.54
8	1025	310.70	274.43	0.60	0.53
9	1145	306.82	274.01	0.62	0.56
10	1034	308.47	273.47	0.60	0.53

Table S3-20 Top 10 MOF materials with highest delivery volumetric adsorption amount between 298K, 250bar and 298K, 5bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1031	324.64	290.08	0.63	0.57
2	1089	321.52	290.06	0.67	0.60
3	1101	321.07	290.03	0.67	0.61
4	1213	318.79	289.78	0.69	0.62
5	1111	323.16	287.11	0.63	0.56
6	1034	322.00	287.00	0.63	0.56
7	1086	320.95	286.83	0.65	0.58
8	1030	323.31	286.54	0.62	0.55
9	1145	319.26	286.45	0.65	0.58
10	1052	317.94	286.18	0.64	0.58

Table S3-21 Top 10 MOF materials with highest delivery volumetric adsorption amount between 233K, 35bar and 298K, 5bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1033	322.62	279.75	0.54	0.47
2	1028	333.78	278.31	0.51	0.42
3	1093	318.78	276.58	0.52	0.45
4	1022	324.47	276.47	0.52	0.44
5	1011	330.22	276.44	0.51	0.43
6	1111	312.41	276.37	0.61	0.54
7	1047	324.10	275.65	0.51	0.44
8	1013	325.92	275.24	0.52	0.44
9	1280	316.85	275.21	0.55	0.47
10	1197	323.34	275.08	0.50	0.43

Table S3-22 Top 10 MOF materials with highest delivery volumetric adsorption between 233K, 45bar and 298K, 5bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1026	342.07	291.18	0.52	0.44
2	1197	339.13	290.87	0.51	0.44
3	1031	325.40	290.83	0.62	0.55
4	1111	325.98	289.93	0.62	0.55
5	1025	325.33	289.06	0.61	0.54
6	1028	343.96	288.48	0.51	0.43
7	1037	327.77	288.29	0.60	0.52
8	1055	328.16	288.10	0.57	0.50
9	1007	333.60	287.61	0.57	0.48
10	1280	329.06	287.43	0.55	0.48

Table S3-23 Top 10 MOF materials with highest delivery volumetric adsorption between 233K, 75bar and 298K, 5bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1145	349.93	317.12	0.71	0.64
2	1089	348.47	317.02	0.73	0.66
3	1162	351.32	316.95	0.69	0.62
4	1213	345.82	316.80	0.74	0.68
5	1101	347.31	316.27	0.73	0.66
6	1031	350.31	315.74	0.68	0.62
7	1297	349.10	312.37	0.60	0.53
8	1064	350.68	312.14	0.66	0.59
9	1111	347.75	311.70	0.68	0.61
10	1055	351.69	311.62	0.63	0.56

Table S3-24 Top 10 MOF materials with highest volumetric delivery amount at 298K and 35bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}^{*}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}^{*}(g/g)$
1	1226	248.54	205.71	0.29	0.24
2	2030	255.60	204.80	0.16	0.13
3	2006	244.94	203.55	0.28	0.23
4	2010	245.07	202.24	0.29	0.24
5	2004	247.30	201.24	0.26	0.21
6	1116	240.07	200.88	0.30	0.25
7	2013	240.07	200.88	0.30	0.25
8	1156	239.85	200.69	0.29	0.24
9	2017	239.85	200.69	0.29	0.24
10	1152	234.74	198.94	0.29	0.25

*:Γ_{del} means delivery amount between this condition and 358K and 5bar, so as the following tables.

Table S3-25 Top 10 MOF materials with highest volumetric delivery amonut at 298K and 70bar

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1062	270.37	246.24	0.38	0.35
2	1050	274.03	245.23	0.37	0.33
3	2119	274.03	245.23	0.37	0.33
4	2225	271.27	245.08	0.33	0.30
5	1023	274.00	244.77	0.39	0.34
6	2157	274.00	244.77	0.39	0.34
7	1048	270.22	244.75	0.38	0.35
8	1036	268.40	244.35	0.39	0.35
9	1108	271.09	243.89	0.37	0.33
10	1072	269.32	243.39	0.37	0.34

Table S3-26 Top 10 MOF materials with highest volumetric delivery amount at 298K and 100bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1046	287.00	264.24	0.42	0.38
2	1026	284.87	263.91	0.45	0.41
3	1011	285.65	263.78	0.44	0.41
4	1013	284.62	263.71	0.45	0.42
5	1017	289.22	263.16	0.42	0.38
6	1029	287.08	263.04	0.43	0.39
7	1020	285.50	262.89	0.43	0.40
8	1036	286.88	262.83	0.41	0.38
9	1072	288.46	262.53	0.40	0.36
10	1061	284.22	262.50	0.42	0.39

Table S3-27 Top 10 MOF materials with highest volumetric delivery amonut at 298K and 150bar

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1026	306.80	285.84	0.48	0.45
2	1011	307.17	285.30	0.48	0.44
3	1028	306.69	283.90	0.47	0.43
4	1020	306.07	283.46	0.47	0.43
5	1013	303.68	282.77	0.48	0.45
6	1019	305.35	282.44	0.47	0.43
7	1046	304.16	281.41	0.44	0.41
8	2225	307.15	280.97	0.38	0.35
9	1045	302.26	280.71	0.45	0.42
10	1033	298.57	280.33	0.50	0.47

Table S3-28 Top 10 MOF materials with highest volumetric delivery amonut at 298K and 170bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1197	310.00	289.66	0.48	0.45
2	1048	315.04	289.58	0.44	0.41
3	1011	310.58	288.70	0.48	0.45
4	1028	311.20	288.41	0.48	0.44
5	1019	311.00	288.09	0.47	0.44
6	1026	308.58	287.61	0.48	0.45
7	1013	308.27	287.36	0.49	0.46
8	1037	304.33	287.06	0.56	0.53
9	1031	302.22	286.56	0.59	0.56
10	1003	304.17	286.46	0.56	0.53

Table S3-29 Top 10 MOF materials with highest volumetric delivery amount at 298K and 200bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1011	319.00	297.12	0.50	0.46
2	1032	316.08	297.12	0.54	0.51
3	1031	312.35	296.69	0.61	0.58
4	1197	315.71	295.38	0.49	0.46
5	1111	311.19	295.05	0.61	0.58
6	1007	314.32	294.65	0.55	0.52
7	1025	310.70	294.64	0.60	0.57
8	1064	311.57	294.58	0.58	0.55
9	1020	316.85	294.24	0.48	0.45
10	1041	309.61	294.01	0.61	0.58

Table S3-30 Top 10 MOF materials with highest volumetric delivery amount at 298K and 250bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1031	324.64	308.98	0.63	0.60
2	1111	323.16	307.02	0.63	0.60
3	1089	321.52	306.96	0.67	0.64
4	1030	323.31	306.78	0.62	0.59
5	1032	325.69	306.73	0.56	0.52
6	1101	321.07	306.67	0.67	0.64
7	1011	328.23	306.36	0.51	0.48
8	1034	322.00	306.18	0.63	0.60
9	1086	320.95	305.46	0.65	0.62
10	1213	318.79	305.21	0.69	0.66

Table S3-31 Top 10 MOF materials with highest volumetric delivery amount at 233K and 35bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1028	333.78	310.98	0.51	0.47
2	1011	330.22	308.34	0.51	0.48
3	1019	329.79	306.88	0.50	0.47
4	1046	329.38	306.62	0.48	0.45
5	1017	332.23	306.17	0.49	0.45
6	1048	331.16	305.69	0.47	0.43
7	1029	329.70	305.66	0.49	0.46
8	2225	331.69	305.50	0.41	0.38
9	1045	327.02	305.47	0.49	0.45
10	1043	333.69	305.31	0.46	0.42

Table S3-32 Top 10 MOF materials with highest volumetric delivery amount at 233K and 45bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1028	343.96	321.16	0.53	0.49
2	1026	342.07	321.10	0.54	0.50
3	1197	339.13	318.80	0.53	0.50
4	1011	340.06	318.19	0.53	0.50
5	1061	338.77	317.05	0.50	0.47
6	2225	342.62	316.44	0.42	0.39
7	1046	339.18	316.43	0.49	0.46
8	1017	342.42	316.36	0.50	0.46
9	1019	338.66	315.76	0.52	0.48
10	1048	341.13	315.66	0.48	0.44

Table S3-33 Top 10 MOF materials with highest volumetric delivery amount at 233K and 75bar.

Rank	ID	$\Gamma_{Tot}(cm^3/cm^3)$	$\Gamma_{Del}(cm^3/cm^3)$	$\Gamma_{Tot}(g/g)$	$\Gamma_{Del}(g/g)$
1	1162	351.32	336.14	0.69	0.66
2	1026	356.78	335.82	0.56	0.53
3	1011	357.40	335.52	0.56	0.52
4	1007	354.96	335.29	0.62	0.59
5	1145	349.93	335.26	0.71	0.68
6	1031	350.31	334.65	0.68	0.65
7	1019	357.54	334.63	0.55	0.51
8	1001	355.91	334.21	0.62	0.58
9	1071	352.18	334.07	0.61	0.58
10	1087	352.23	333.95	0.62	0.59

3.5 Nitrogen-Doped Porous Aromatic Frameworks for Enhanced CO_2 Adsorption

Abastract

Recently synthesized porous aromatic frameworks (PAFs) exhibit extremely high surface areas and exceptional thermal and hydrothermal stabilities. Using computer-aided design, we propose new PAFs, designated as NPAFs, by introducing nitrogen-containing groups to the biphenyl unit and predict their CO_2 adsorption capacities with grand canonical Monte Carlo (GCMC) simulations. Among various NPAFs considered, one with imidazole groups shows the highest adsorption capacity for CO_2 (11.5 wt% at 1 bar and 298 K), in comparison with 5 wt% for the parent PAF (PAF-1) at the same condition. At higher pressures (around 10 bar), however, another NPAF with pyridinic N groups performs much better than the rest. This study suggests that adding N functionality to the organic linkers is a promising way to increase CO_2 adsorption capacity of PAFs at ambient condition.

3.5.1. Introduction

CO_2 capture and sequestration (CCS) is pitovally important in addressing the issue of global warming[164,165]. The U.S. Department of Energy (DOE) launched a carbon sequestration program in 2009 aiming to achieve 90% CO_2 capture at an increase in the cost of electricity of no more than 35% for the post-combustion process by 2020. One promising way to meet the DOE target is by finding proper porous materials that offer easy release of CO_2 compared to the conventional reactive sorbents such as aqueous ammonia and amine functionalized solids. Porous materials with a wide variety of topologies, pore sizes, and functionalities have been explored for CO_2 capture, including porous carbons,

zeolites, metal-organic frameworks, and zeolitic-imidazolate frameworks[142,164,166,167]. For example, Yaghi et al. reported ultra-high porosity MOF-210 which has a BET surface area (6240 m^2/g), one of the highest among all reported MOF materials, and exhibits a CO_2 uptake of 2396 mg/g at 50 bar and 298K[168]. Long et al. reported Mg-MOF-74, which shows high $CO_2/CH_4/H_2$ selectivity, even surpassing zeolite 13X – a commonly available and used microporous material for pressure swing adsorption[169].

One of the major drawbacks of existing nanoporous MOFs is their limited physicochemical stability. Recently, Ben et al. developed a new class of stable porous materials, porous aromatic frameworks (PAFs), with diamond-like structures held together by strong C-C covalent bonds[170]. Unlike most MOFs, PAF-1 shows exceptional thermal and hydrothermal stabilities. More interestingly, the PAF-1 framework provides a perfect platform for introducing functionalities on its biphenyl units. Zhou et al. synthesized a series of porous polymer networks (PPNs) [171-173], and among them, sulfonate-functionalized PPN-6, which is the same as PAF-1, shows high CO_2 uptake (~13 wt% at 298 K and 1 bar) [174]. Babarao et al. investigated PAFs functionalized with polar organic groups for CO_2 adsorption and separation, and found that the tetrahydrofuran- and ether-functionalized PAF-1 structures have high adsorption capacity for CO_2 at ambient conditions[175]. Their simulation also suggested that imine-linked interpenetrated frameworks with dimaondoid strucutres would have high CO_2 uptake[176]. Cao et al. investigated the CO_2 storage and separation performance of new PAFs with longer phenyl ring chains, and found the selectivity property is closely related to the difference of isosteric heats of the gas components but independent with the molar fraction at zero pressure[177].

The rich N-containing groups, however, have not been systematically investigated as potential functional groups in PAF frameworks for enhancing CO_2 adsorption, even though N-doped carbons have shown great potential for CO_2 capture[178-181]. For instance, newly synthesized microporous N-doped carbon materials exhibit a high CO_2 uptake capacity up to 6.9 mmol/g at 273 K/1 bar and 4.4 mmol/g at 298 K/1 bar[182] and a CO_2/N_2 selectivity of 16 at 298 K and 1 bar[183]. How N-functionalized PAFs will perform against these reported N-doped carbons would be very interesitng to find out. To that end, in this work we use PAF-1 as a template to examine systematically the effect of the N-containing functional groups on CO_2 adsorption capabilities of this class of materials.

3.5.2. Molecular models and simulation methods

To construct new NPAFs, we propose a large number of organic linkers as shown in Figure 3-24. All these organic linkers contain a biphenyl framework with different nitrogen functional groups. The NPAF structures were constructed by inserting those target linkers between the C−C covalent bonds in diamond structure as shown in Figure 3-25. The moelcular framework is then structurally pre-optimized by applying the UFF force field[60]. The final structures were obtained by geometry optimization with VASP[184].

Moelcular simulations of gas adsorption in various NPAFs were based on a combination of the pairwise site-site Lennard-Jones (LJ) 12-6 terms and Coulombic potentials for the interactions between gas-adsorbent and gas-gas molecules:

$$E(r_{ij}) = E_{LJ} + E_{Coul} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{r_{ij}} \quad (86)$$

The Lorentz-Berthelot combination rule was used for constructing parameters for different atom pairs.

In simulations of CO_2 adsorption in NPAF frameworks, we used a 2x2x2 supercell and a rigid model for the solid crystal atoms. Since the flexibility of the molecular framework has only a marginal effect on the adsorption of small gas molecules. The Lennard-Jones potential parametes of NPAF frameworks were from the DREIDING force field[185], which has been proved before to be accurate for predicting gas adsorption in MOFs and PAFs[9, 18]. Recently, periodic DFT calculation based parial charge derivation methods are also introduced, for example, REPEAT[186], but haven't shown superior performance compared to common cluster fragment models in porous materials cases, like IRMOF-1. So here because of the large number of atoms in a NPAF unit cell, we still calculated the atomic partial charges from the fragmental models cleaved from the NPAF crystal structure as shown in Figure S3-2 in Supporting Information. The dangling bonds were saturated by hydrogen atoms. The fragment structure optimization was based on quantum-mechnical calculations using Gaussian 09 with the PW91 functional and the 6-31G(d,p) basis set[187]. Meanwhile the atomic charges were derived by fitting to the electrostatic potential using the ChelpG scheme[188], which has been sucessfully used in the previous simulation work[19]. Table S3-34 in Supporting Information shows all atomic charges of the proposed NPAF frameworks. To describe the interactions between the gases and adsorbents, we used the well-established potentials in previous work[175]. For both CO_2 and N_2 molecules, 3-site models were used. The C-O bond length in CO_2 is 1.18 Å and the bond angle O-C-O is 180 º. The charges on C and O atoms were +0.576e and -0.288e, respectively, while

ε_C/k_B=29.66 K, σ_C=2.789 Å and ε_O/k_B=82.96 K, σ_C=3.011 Å. The model reproduced the isosteric heat and isotherm of CO_2 adsorption in silicate. N-N bond length in N_2 is 1.10 Å, and a -0.482e charge on each N atom while +0.964e charge on the center-of-mass, which were fitted to the experimental bulk properties of N_2. Lennard-Jones parameter is only valid on N atoms, ε_N/k_B=39.89 K, σ_N=3.385 Å.

The gas adsorption in various NPAFs was simulated by the grand canonical Monte Carlo (GCMC) method[58]. The non-bonded interactions were evaluated with a spherical cutoff of 12.9 Å (or half of the simulation box if r_{cut} is less than 12.9 Å), and the Coulombic interactions were calculated using the Ewald method[189]. $3x10^6$ trial moves were used to reach the equilibrium state in a typical GCMC simulation, and subsequent $3x10^6$ moves were collected for doing ensemble averages. Three types of trial moves were attempted in the GCMC simulations: insertion/deletion, displacement and rotation of gas molecules in the simulation box. Chemical potentials applied in the simulation works were calculated by Widom insertion method, carried under isothermal-isobaric (NPT) ensemble. The data of each simulation trajectory was divided into 10 blocks in order to estimate statistical uncertainties. Unless specifically mentioned, the statistical uncertainty was generally smaller than the symbol sizes presented in the figures. All MC simulations were carried out by the Towhee 7.0.4 program[59].

3.5.3. Results and discussion

We first validated the above simulation procedure by comparing the predicted and experimental CO_2 adsorption data for PAF-1 at 298 K below 10 bar[170]. Figure 3-26 shows that the simulation data from our charge model reproduce the experimental results almost

quantitatively, especially at low pressures (below 4 bar). At high pressures, our model tends to slightly overestimate the amount of CO_2 adsorption. Because a precise measurement of the adsorption isotherm is experimentally challenging and CO_2 capture is usually performed at ambient conditions, our method should be sufficiently accurate to predict the relative performance of different porous materials.

Figure 3-27 and 3-28 shows the GCMC results for CO_2 adsorption isotherms in various NPAFs at 298K. At low pressure, the total amount of gas adsorption is mainly determined by the guest-framework interaction energy. As a result, N-containing functional groups with more polarizable electron density distributions are expected to provide stronger interaction with CO_2. This is why all NPAFs show better CO_2 adsorption capacity than the parent PAF (PAF-1). It should be noticed that according to recently quantum mechanics study by Zhang et al.[190], among all the nitrogen doped 6-atom heterocyclic rings, considering one CO_2 molecular, pyridine has the strongest interaction with CO_2; when the number of CO_2 molecules increases, due to the CO_2-CO_2 interaction, pyrimidine shows the larger binding energy, while the rings with three nitrogen atoms are the least preferred. This suggests heterocyclic rings with two N atoms may be an optimal choices for CO_2 capture purpose, which could be seen in our results.

To explain this trend, in Table 3-6 we explore the correlation between NPAFs structure properties and CO_2 adsorption amount. NPAF-11 displays the best adsorption capacity under 1 bar (11.5 wt% compared to 5.1 wt% of PAF-1), since it has the largest heat of adsorption (22.1 kJ/mol) and second smallest free volume (66.2%) amount all the NPAFs. At ambient condition, it is a clearly interaction strength controlling process, the

CO_2 storage capacity is increasing with heat of adsorption, while decreasing with void fraction, since larger free volume means less binding sites per unit space. In the similar way, along with void fraction, the adsorption amount increases with both decreasing pore limiting diameter (PLD) and maximum pore diameter (MPD), but decreases with decreasing density, which also means less amount of favorite binding sites. One can see that NPAF-5, 11, 12 take the largest amount of CO_2, suggesting that the imidazole and tetrazole groups are the most efficient with respect to enhancing the CO_2 adsorption capacity. This conclusion is also supported by previous quantum mechanical calculations for the interaction between CO_2 and N-containing heterocycles[191]. It is also known that adding tetrazole groups helps increase CO_2 adsorption capacity of polymers with intrinsic micropores[192]. Moreover, recently Zhu et al. synthesized a nitrogen-doped porous material called triazole-functionalized triazine framework and obtained an experimental CO_2 adsorption capacity of 6.0 wt% at 298 K and 1 bar[179]. This triazine framework looks similar to our NPAF-2.

When the pressure increases, the total amount of gas adsorption depends on both the guest-framework interaction energy and the free volume. The latter becomes a dominant factor at high pressure. As shown in Figure 3-28, NPAF-11 provides the highest absolute adsorption amount up to 6 bar due to strong guest-framework interactions. But as the pressure further increases, most of the favorable binding sites are saturated and the free volume takes over as the dominant factor, which could be seen from Table 3-6. We could see that different from low pressure (1bar), at high pressure (10bar), the CO_2 adsorption amount is decreasing with heat of adsorption and density but increasing with void fraction,

PLD and MPD. That's why NPAF-2 shows the best performance at higher pressures (126 wt% at 10 bar), because it has one of the largest void fraction (86.1%) and the N polar functional groups provide it a much larger heat of adsorption (21.7 kJ/mol) compared to PAF-1 (15.9 kJ/mol) and NPAF-1 (19.4 kJ/mol), which increases the interaction strength with CO_2.

Since the ambient conditions (298 K and 1 bar) are most relevant to post-combustion carbon capture, our simulation suggests that the best NPAFs can offer CO_2 adsorption capacities of about 10-11 wt% at such conditions. How does this performance of NPAFs compare with recently reported N-doped porous materials for carbon capture? Sevilla et al. prepared N-doped polypyrrole-based porous carbons and found that the 600^{o}C-activated sample has a N content of 10 wt% and CO_2 uptake of 15 wt% at 298 K and 1 bar[193]. Zhu et al. synthesized N-doped porous carbonaceous membranes (N-PCM) from trimerization reactions of acetyl compounds and found that their 550^{o}C-activated material has a N content of 3.4 at% and a CO_2 uptake of 7.5 wt% at 298 K and 1 bar[179]. El-Kaderi and coworkers have synthesized porous benzimidazole-linked polymers (BILP)[194,195]. Their BILP-4 material is also based on a tetrahedral model structure for 3D linking, similar to the NPAFs in this study. BILP-4 displays an impressive CO_2 uptake of 16 wt% at 1 bar and 298 K[194]. We found that our predicted CO_2 capacities for NPAFs are in line with these experimental CO_2 uptakes of various N-containing porous materials, confirming the role of N in enhancing CO_2 uptake.

For post-combustion CO_2 capture purpose, we also studied the gas mixture CO_2/N_2 adsorption selectivity S(i/j) at 298K shown in Figure 3-29. Here, according to previous

work, we defined $S_{ij} = \frac{x_i}{x_j}$, where x_i is the gas compound mole density in adsorption phase, where the mole ratio of CO_2:N_2 is 15:85. All the materials had nearly constant selectivity ration under 1 bar. Since the new organic linkers provided stronger interaction with CO_2 which lead to higher selectivity than original PAF-1. In general, they have the same relative order as in pure CO_2 storage prediction under 1 bar, since this is still in the interaction strength controlling area, those having larger heat of adsorption of carbon dioxide would have better selectivity. Another reason is those with stronger interaction with CO_2 also have larger difference between PLD and MPD, for example, the PLD of NPAF-11 is only 5.9Å while the PLD of NPAF-2 is 10.7Å, even much larger than the MPD of NPAF-11 (9.6Å), which means there is nearly no geometric barrier for N_2 and CO_2 diffuse in the pore channels. NPAF-11 is still the best one, beside one of the largest heat of adsorption in this condition, it also has the smallest PLD (5.9 Å) and second largest difference between PLD and MPD (3.5 Å).

3.5.4. Conclusion

We proposed a series of new porous aromatic frameworks (NPAFs) with N-doped heterocyclic functional groups for enhancing CO_2 adsorption at ambient conditions. A reasonably good agreement was obtained between the simulation and experimental data for CO_2 adsorption in crystalline PAF-1. Among those new hypothetic functional NPAFs, NPAF-11 with imidazolic groups shows the highest adsorption capacity for CO_2 (11.6 wt%) at 1 bar and 298 K, a 130% increase upon that of the parent PAF-1. This study reveals that

adding N-doped heterocyclic groups to a porous framework such as PAF-1 greatly enhances its CO_2 uptake.

Figure 3-24. N-containing organic linkers used to construct NPAFs (Figure 3-25).

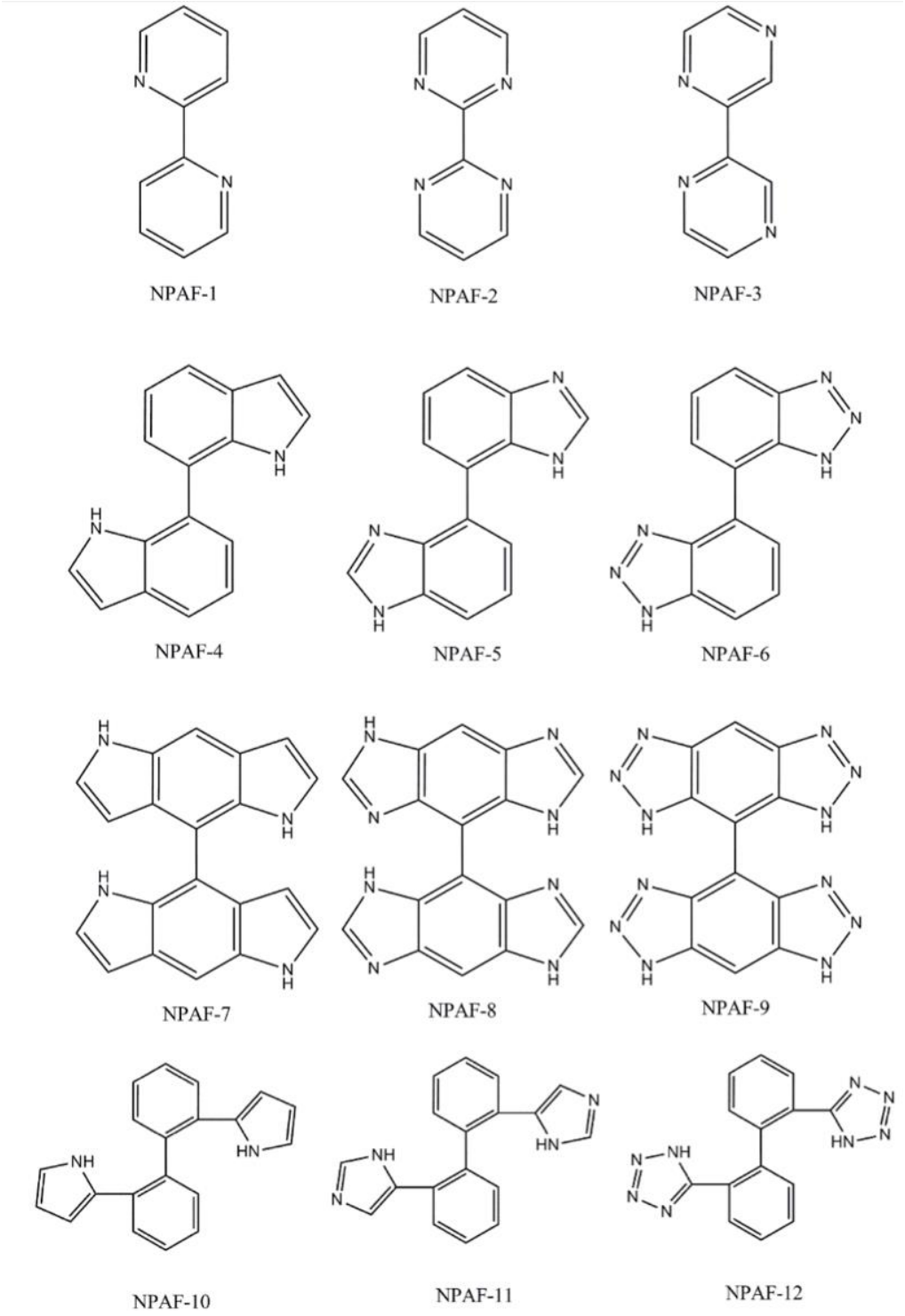


Figure 3-25. Schematic of the NPAF-generation strategy: Every C-C bond in the diamond structure (left) is replaced by a N-containing biphenyl-like linker (middle) to arrive at the desired NPAF structure (right).

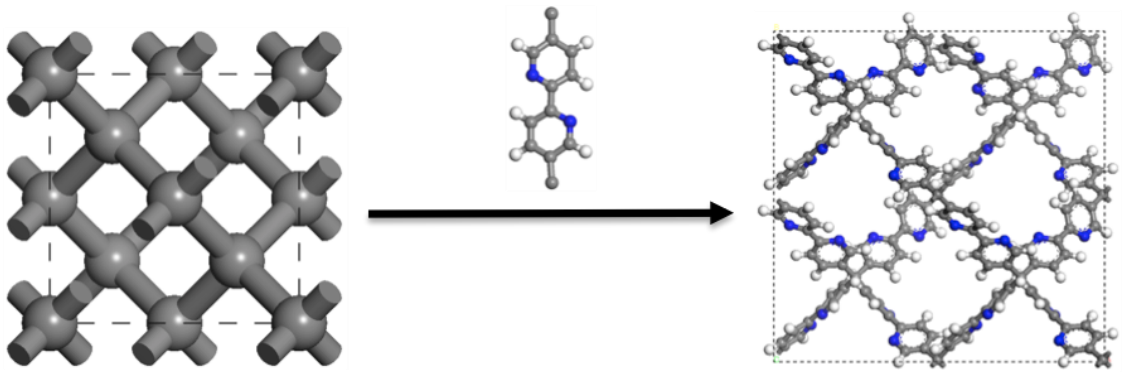
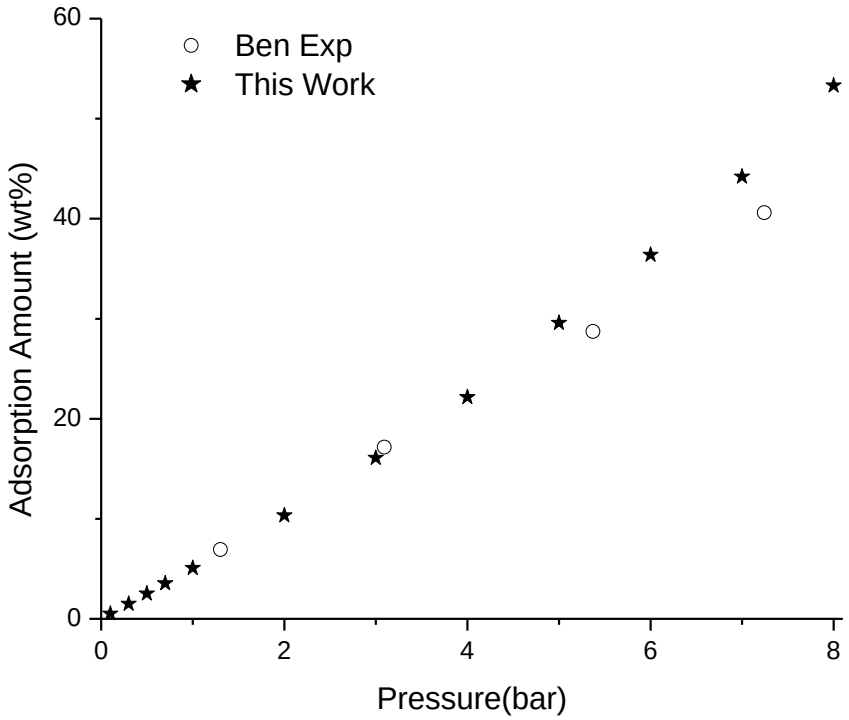


Figure 3-26. Comparison of experimental and simulation data (this work) for CO_2 adsorption isotherms at 298K under 10 bar[170]



.

Figure 3-27. Comparison of NPAFs CO_2 adsorption isotherms at 298K under 1 bar; the results for PAF-1 and NPAF-1 almost exactly overlap each other.

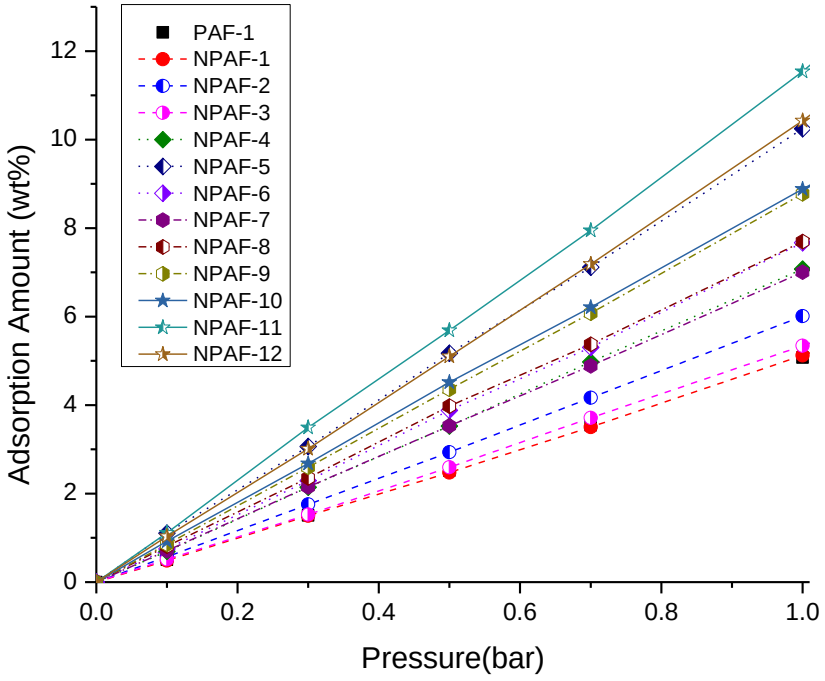


Figure 3-28. Comparison of NPAFs CO_2 adsorption isotherms at 298K and up to 10 bar.

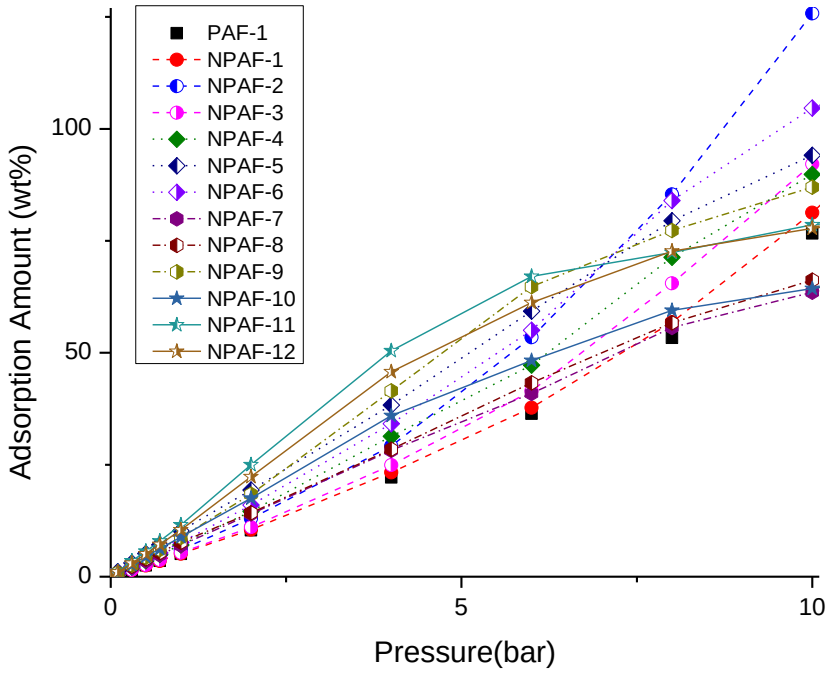


Figure 3-29. Comparison of NPAFs gas mixture adsorption selectivity S(i/j) at 298K, the mole ratio between CO_2 and N_2 is 15:85.

*:The curves of PAF-1 and NPAF-1 overlap with each other.

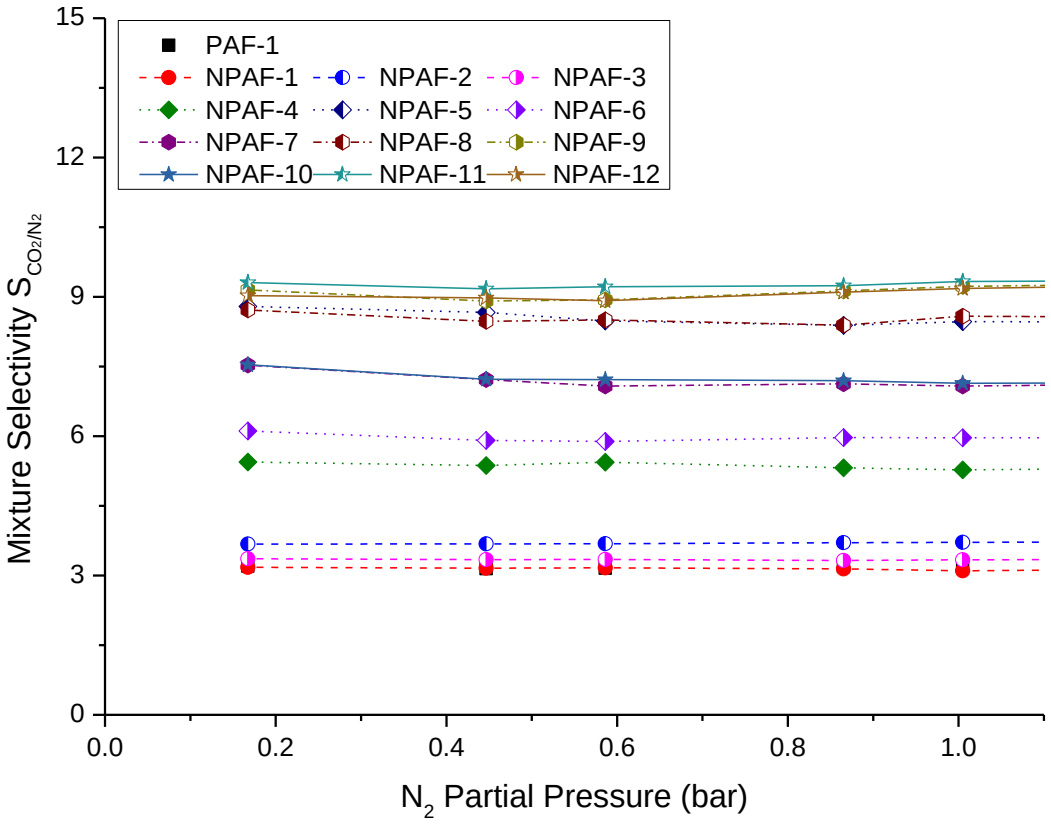
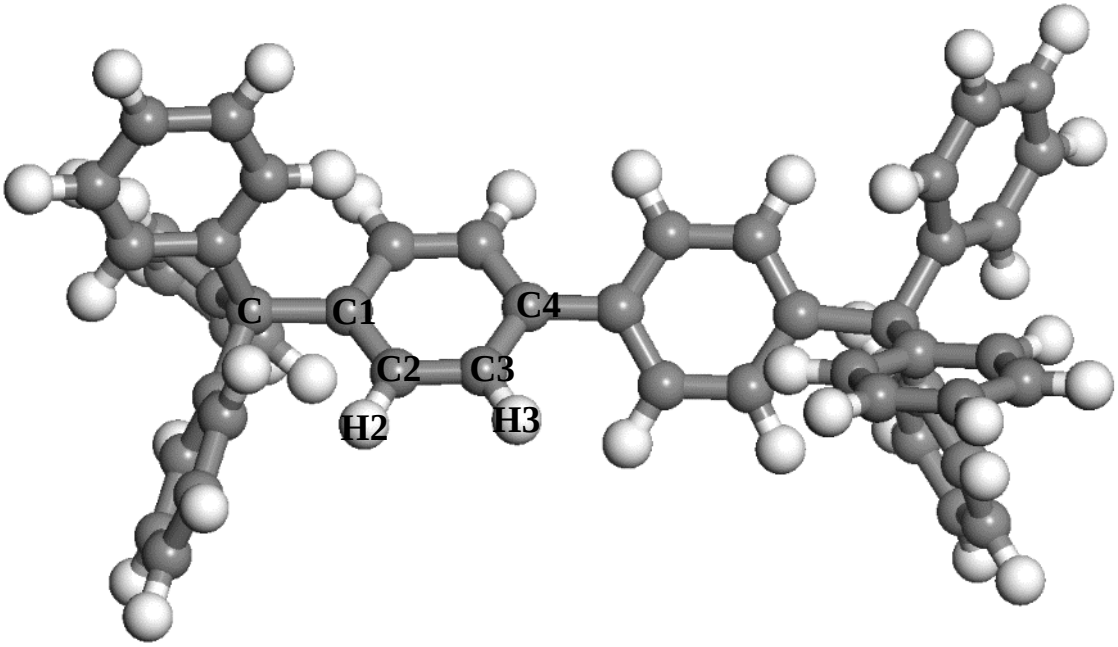


Table 3-6. Correlation between NPAFs structure properties and CO_2 adsorption performance at 298K and different pressures. a) Crystal density in g/cm^3; b) Void fraction in percentage; c) Pore limiting diameter in Å; d) Maximum pore diameter in Å; e) Heat of adsorption in kJ/mol; f) Adsorption amount in wt%. b,c and d used helium as probe.[134]

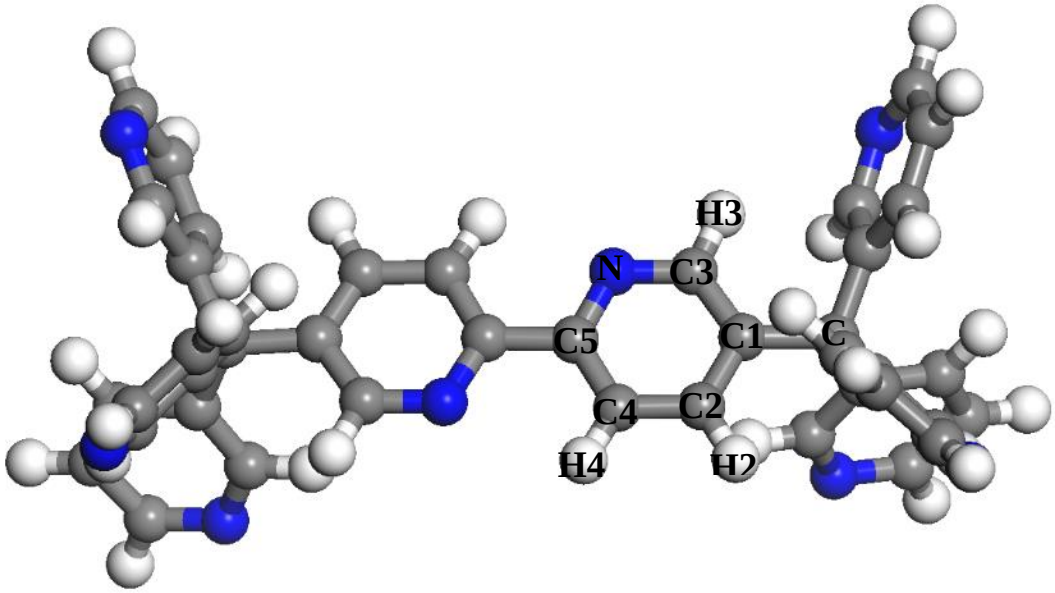
Name	Density[a]	Void[b]	PLD[c]	MPD[d]	Qst_{1bar}[e]	Γ_{1bar}[f]	Qst_{10bar}[e]	Γ_{10bar}[f]
PAF-1	0.312	0.866	10.1	12.9	14.04	5.08	15.88	76.68
NPAF-1	0.332	0.864	10.1	12.8	13.94	5.12	19.42	81.29
NPAF-2	0.354	0.861	10.7	12.5	14.60	6.01	21.70	125.77
NPAF-3	0.354	0.856	10.2	12.3	14.32	5.34	20.05	92.12
NPAF-4	0.454	0.789	7.9	10.5	19.15	7.07	24.38	89.83
NPAF-5	0.447	0.800	8.2	10.9	21.46	10.25	23.98	94.09
NPAF-6	0.450	0.812	9.1	11.8	17.55	7.67	23.56	104.67
NPAF-7	0.589	0.695	7.2	9.2	20.80	7.01	25.92	63.53
NPAF-8	0.596	0.708	7.5	9.3	20.95	7.70	24.32	66.24
NPAF-9	0.589	0.742	8.6	10.4	19.16	8.77	24.82	86.99
NPAF-10	0.568	0.653	5.9	9.4	21.81	8.88	29.21	64.35
NPAF-11	0.572	0.662	5.9	9.6	22.12	11.54	29.87	78.59
NPAF-12	0.580	0.684	6.7	10.2	20.80	10.43	27.44	77.79

Support Information

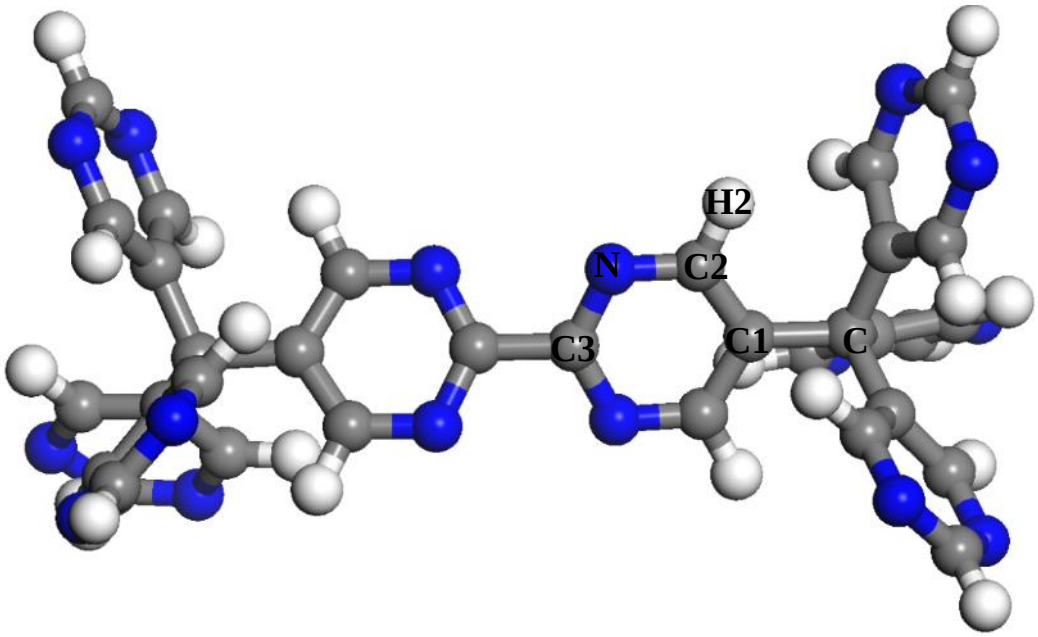
Figure S3-2. Charge models for NPAFs in GCMC adsorption, a) PAF-1; b) NPAF-1; c) NPAF-2; d) NPAF-3; e) NPAF-4; f) NPAF-5; g) NPAF-6; h) NPAF-7; i) NPAF-8; j) NPAF-9; k)NPAF-10; l) NPAF-11; m) NPAF-12



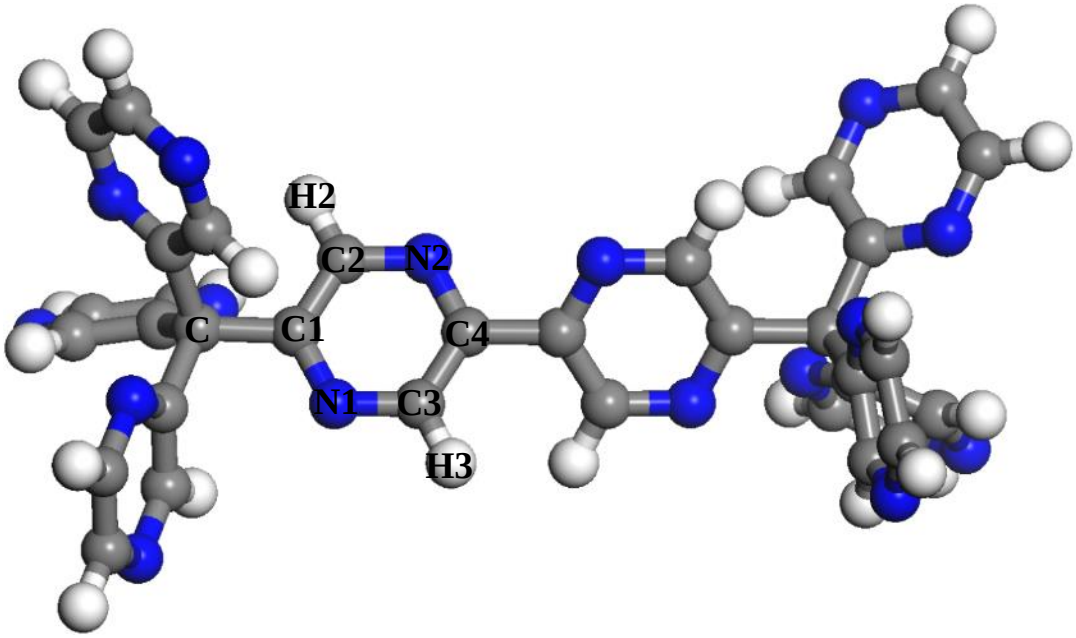
a)



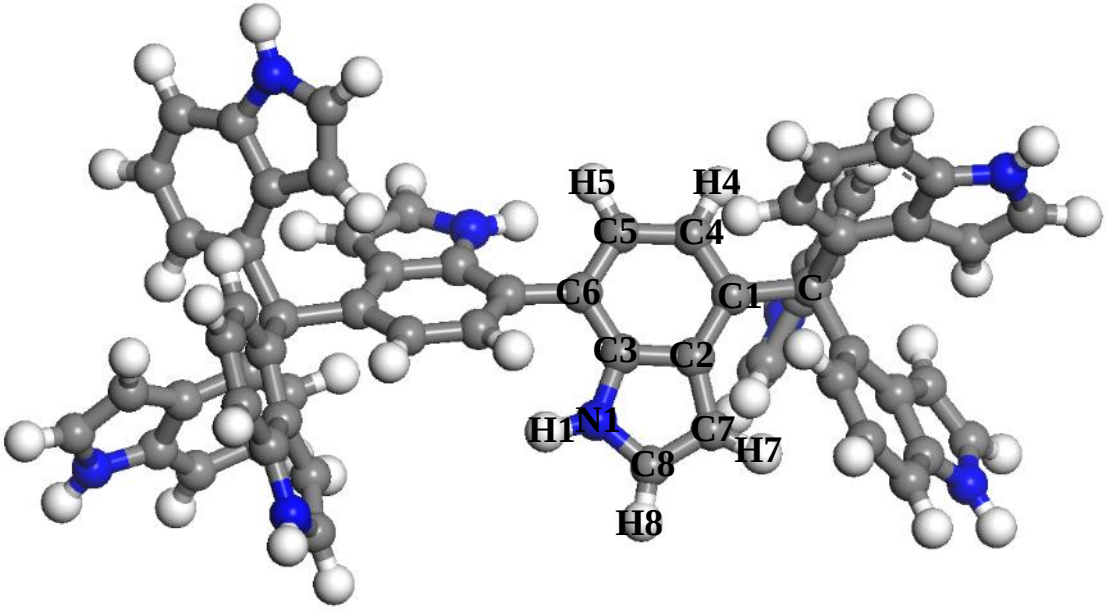
b)



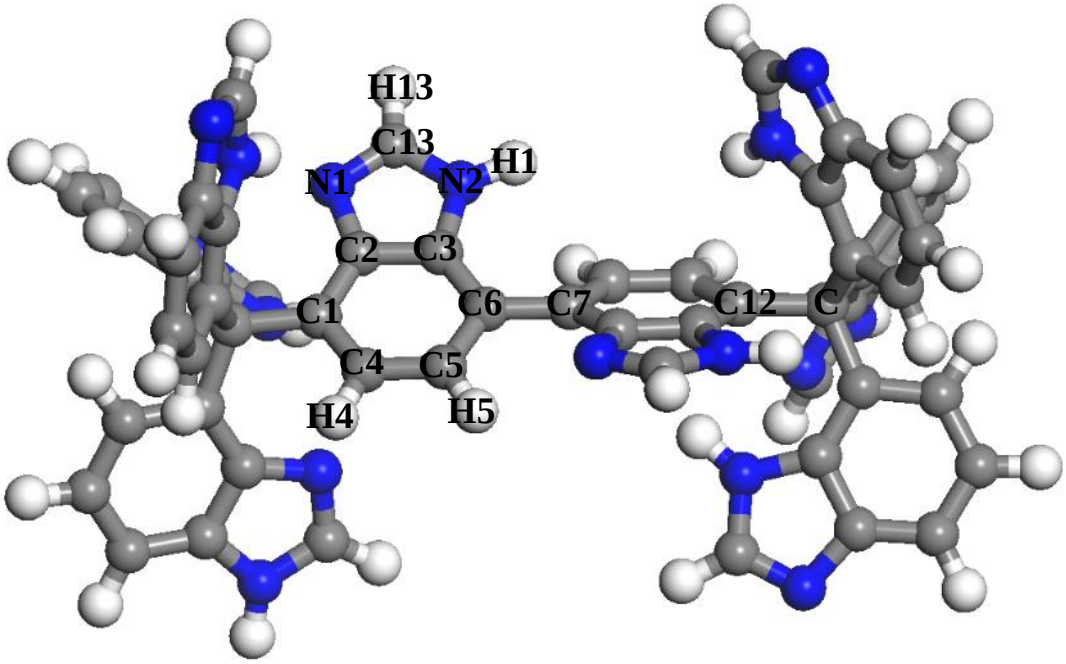
c)



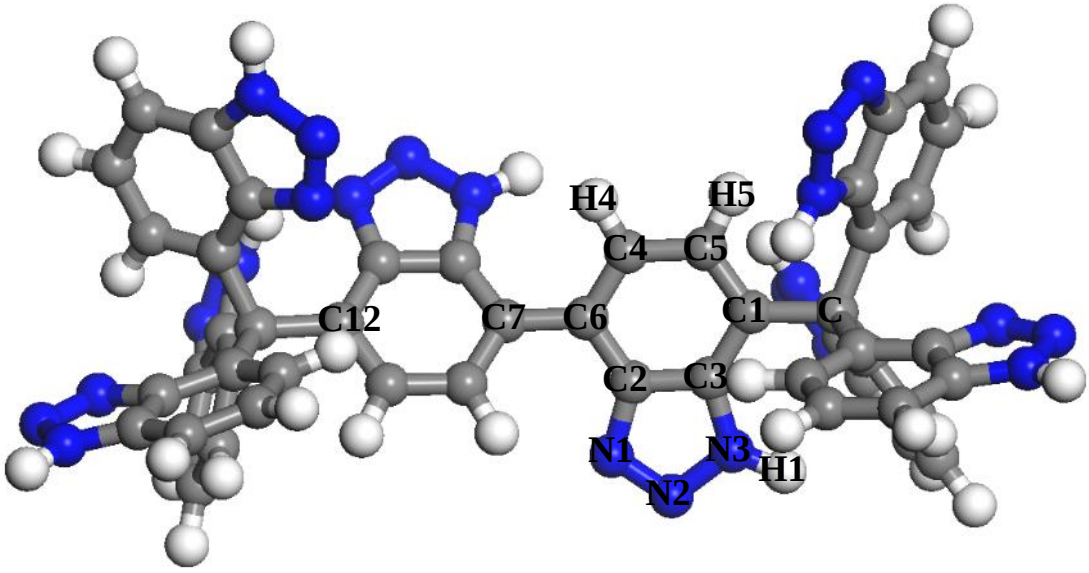
d)



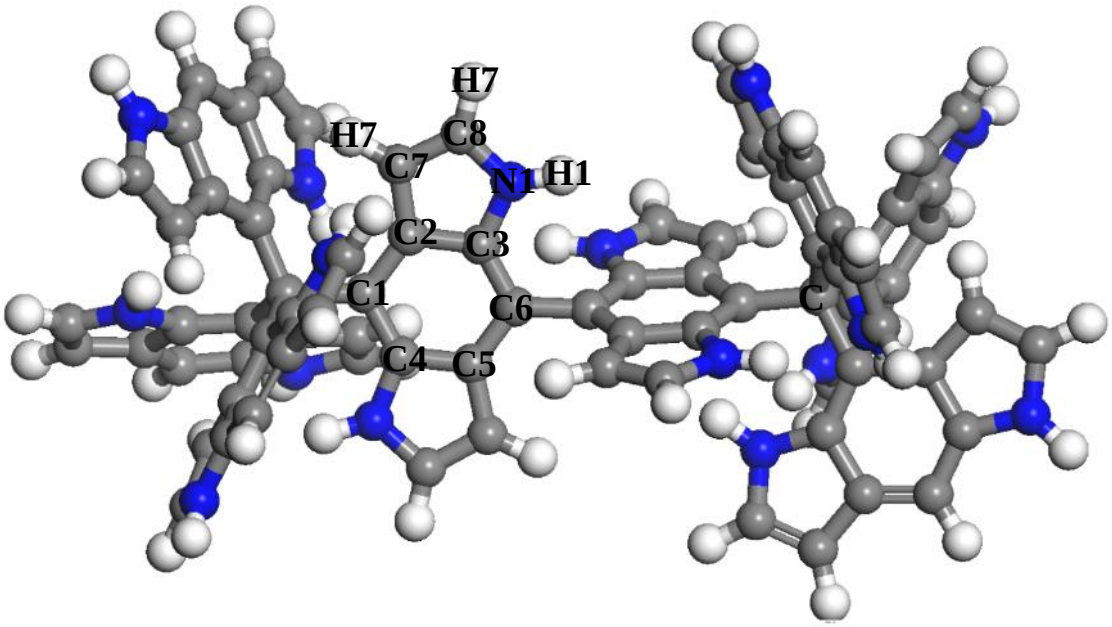
e)


H13
C13
N1
N2
H1
C2
C3
C1
C6
C7
C12
C
C4
C5
H4
H5

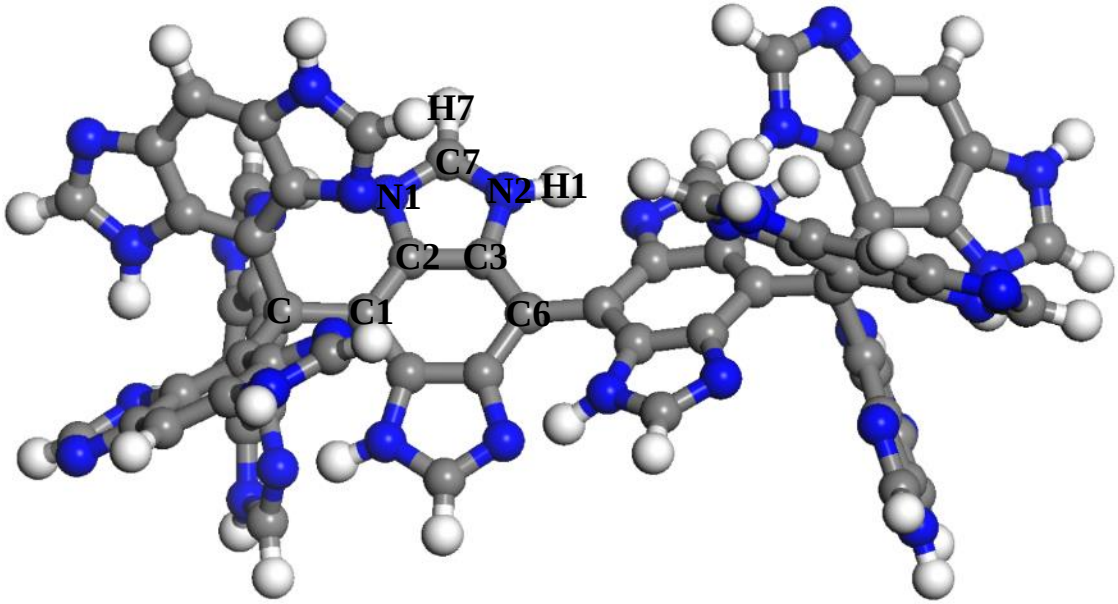
f)


H4
H5
C4
C5
C12
C7
C6
C1
C
C2
C3
N1
N3
H1
N2

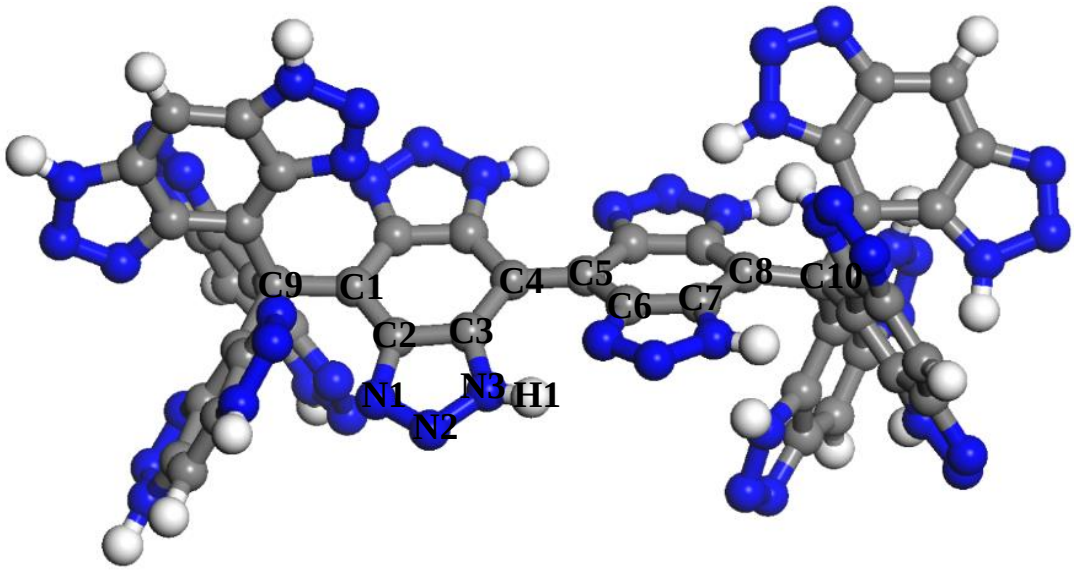
g)



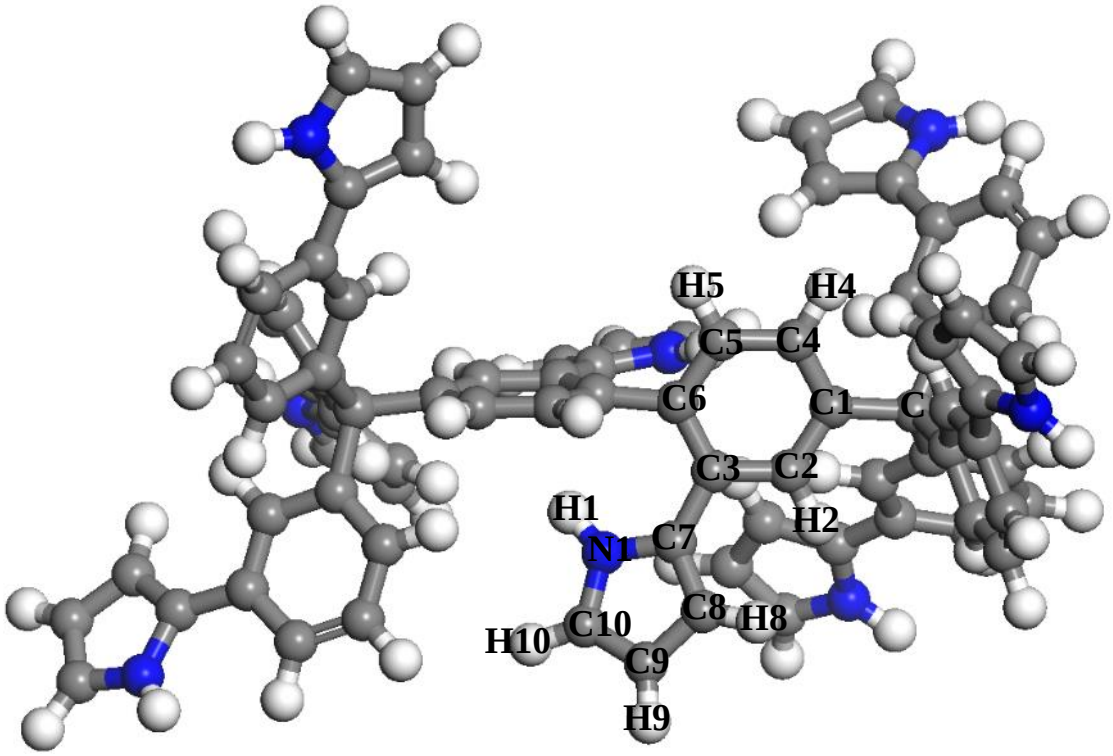
h)



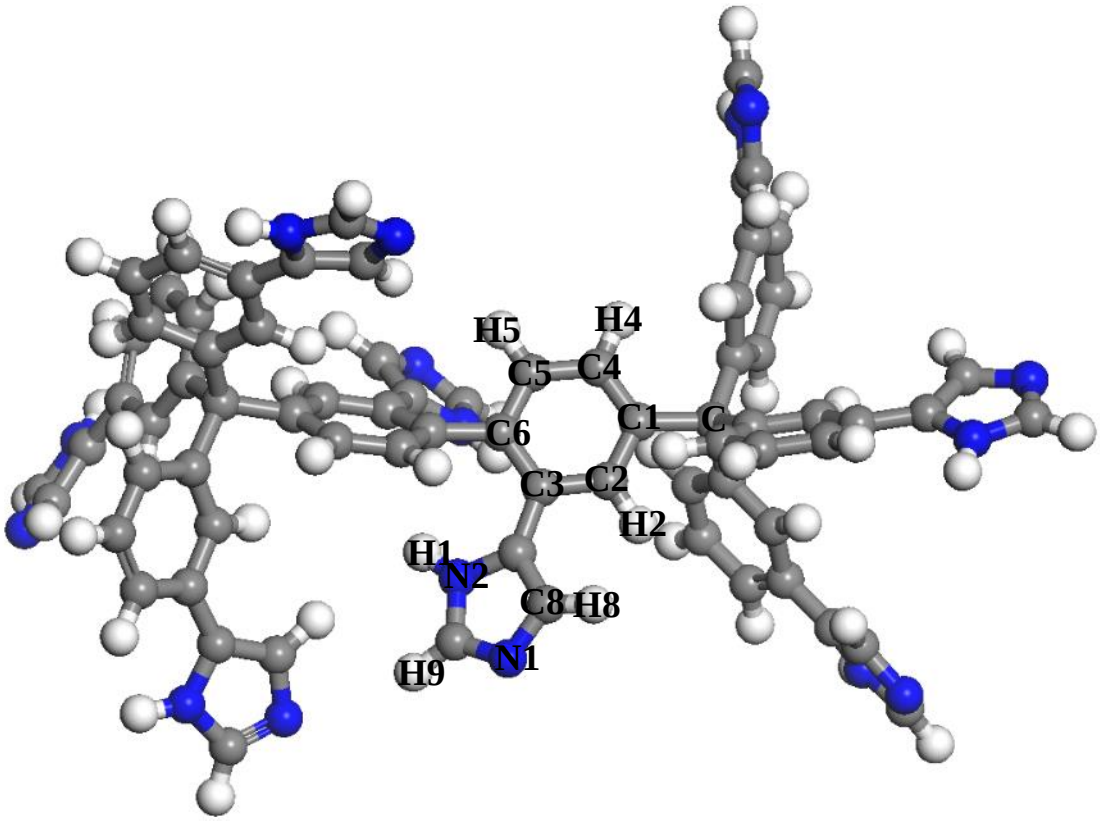
i)


C9
C1
C4
C5
C8
C10
C6
C7
C2
C3
N1
N3
H1
N2

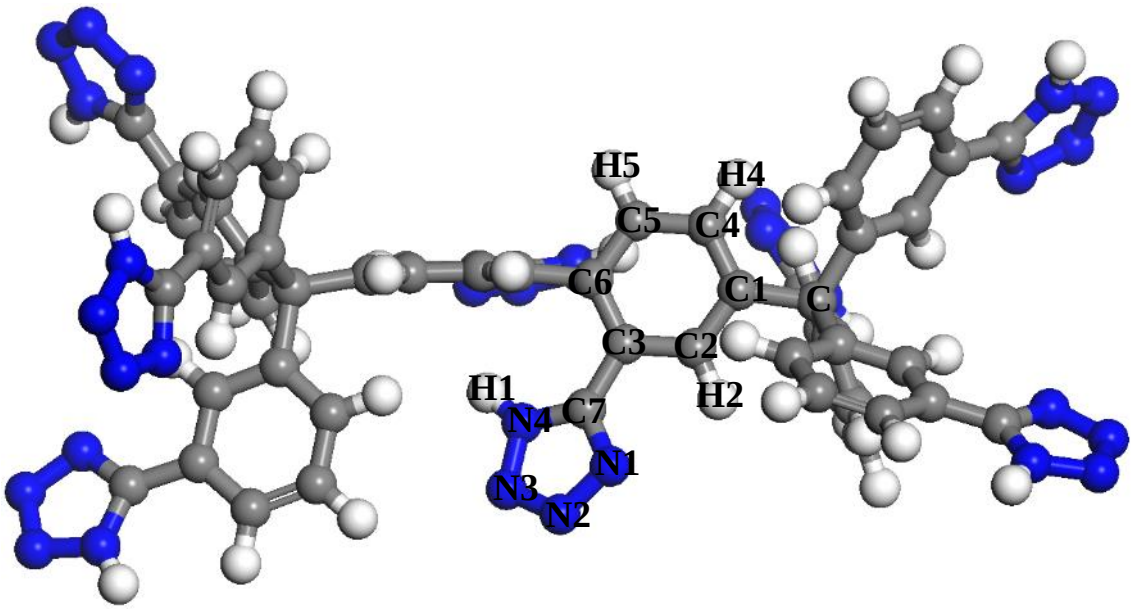
j)


H5
H4
C5
C4
C6
C1
C
C3
C2
H1
H2
N1
C7
C8
H8
H10
C10
C9
H9

k)


H5
H4
C5
C4
C1
C
C6
C3
C2
H2
H1
N2
C8
H8
N1
H9

l)


H5
H4
C5
C4
C6
C1
C
C3
C2
H1
H2
N4
C7
N1
N3
N2

m)

Table S3-34. Atomic charge parameters for GCMC simulation, obtained with PW91/6-31G* and ChelpG method in Gaussian09

PAF-1		NPAF-1		NPAF-2		NPAF-3	
Type	**Charge (e)**	**Type**	**Charge (e)**	**Type**	**Charge (e)**	**Type**	**Charge (e)**
C	-0.11	**C**	0.02	**C**	0.37	**C**	-0.67
C1	0.16	**C1**	-0.04	**C1**	-0.39	**C1**	0.45
C2	-0.12	**C2**	-0.02	**C2**	0.47	**C2**	0.15
H2	0.06	**H2**	0.06	**H2**	0.00	**H2**	0.05
C3	-0.13	**C3**	0.27	**N**	-0.63	**N1**	-0.39
H3	0.10	**H3**	0.01	**C3**	0.61	**C3**	0.03
C4	0.04	**C4**	-0.16			**H3**	0.08
		H4	0.10			**N2**	-0.42
		N	-0.50			**C4**	0.23
		C5	0.28				

NPAF-4		NPAF-5		NPAF-6	
Type	**Charge (e)**	**Type**	**Charge (e)**	**Type**	**Charge (e)**
C	0.62	**C**	1.18	**C13**	0.78
C1	-0.20	**C1**	-0.40	**C1**	-0.41
C2	0.20	**C4**	-0.12	**C4**	-0.01
C3	0.03	**H4**	0.12	**H4**	0.06
C4	-0.09	**C5**	-0.15	**C5**	-0.13
H4	0.05	**H5**	0.14	**H5**	0.10
C5	-0.15	**C2**	0.44	**C2**	0.37
H5	0.09	**C3**	0.05	**C3**	-0.08
C7	-0.26	**N1**	-0.56	**N1**	-0.16
H7	0.09	**C13**	0.34	**N2**	-0.26
C8	-0.05	**H13**	0.11	**N3**	0.09
H8	0.13	**N2**	-0.41	**H1**	0.20
N1	-0.28	**H1**	0.31	**C6**	0.15
H1	0.28	**C6**	0.15	**C7**	-0.20
C6	0.01	**C7**	-0.10	**C12**	-0.26
		C12	-0.80		

NPAF-7		NPAF-8		NPAF-9	
Type	**Charge (e)**	**Type**	**Charge (e)**	**Type**	**Charge (e)**
C	1.17	**C**	1.64	**C9**	2.44
C1	-0.38	**C1**	-0.59	**C1**	-1.27
C2	0.24	**C2**	0.23	**C2**	0.73
C3	-0.12	**C3**	0.05	**C3**	-0.20
C4	0.00	**N1**	-0.46	**C4**	0.12
C5	0.14	**C7**	0.17	**C5**	-0.18
C7	-0.24	**H7**	0.09	**C6**	0.25
H7	0.09	**N2**	-0.20	**C7**	-0.07
C8	-0.07	**H1**	0.20	**C8**	-0.34
H8	0.13	**C6**	-0.01	**C10**	1.05
N1	-0.24			**N1**	-0.19
H1	0.22			**N2**	-0.27
C6	0.03			**N3**	0.14
				H1	0.16

NPAF-10		NPAF-11		NPAF-12	
Type	**Charge (e)**	**Type**	**Charge (e)**	**Type**	**Charge (e)**
C	0.08	**C**	-0.03	**C**	0.04
C1	0.00	**C1**	0.11	**C1**	-0.05
C4	0.00	**C4**	-0.08	**C4**	0.03
H4	0.05	**H4**	0.07	**H4**	0.04
C5	-0.23	**C5**	-0.14	**C5**	-0.19
H5	0.10	**H5**	0.09	**H5**	0.10
C2	-0.10	**C2**	-0.17	**C2**	0.16
H2	0.04	**H2**	0.03	**H2**	0.01
C3	-0.04	**C3**	0.02	**C3**	-0.33
C7	0.09	**C7**	0.03	**C7**	0.49
C8	-0.17	**C8**	0.12	**N1**	-0.25
H8	0.12	**H8**	0.09	**N2**	-0.15
C9	-0.18	**N1**	-0.51	**N3**	-0.13
H9	0.12	**C9**	0.27	**N4**	-0.01
C10	-0.01	**H9**	0.07	**H1**	0.13
H10	0.09	**N2**	-0.32	**C6**	0.13
N1	-0.28	**H1**	0.28		
H1	0.29	**C6**	0.06		
C6	0.11				

References:

(1) Msayib, K. J.; Book, D.; Budd, P. M.; Chaukura, N.; Harris, K. D. M.; Helliwell, M.; Tedds, S.; Walton, A.; Warren, J. E.; Xu, M. C.; McKeown, N. B. *Angew. Chem. Int. Ed.* **2009**, *48*, 3273.

(2) Choi, S.; Watanabe, T.; Bae, T.-H.; Sholl, D. S.; Jones, C. W. *Journal of Physical Chemistry Letters* **2012**, *3*, 1136.

(3) Do, X. D.; Vinh-Thang, H.; Kaliaguine, S. *Micropor. Mesopor. Mat.* **2011**, *141*, 135.

(4) Henninger, S. K.; Habib, H. A.; Janiak, C. *J. Am. Chem. Soc.* **2009**, *131*, 2776.

(5) Fairen-Jimenez, D.; Moggach, S. A.; Wharmby, M. T.; Wright, P. A.; Parsons, S.; Dueren, T. *Journal of the American Chemical Society* **2011**, *133*, 8900.

(6) Nalaparaju, A.; Jiang, J. *Journal of Physical Chemistry C* **2012**, *116*, 6925.

(7) Gallo, M.; Glossman-Mitnik, D. *J. Phys. Chem. C* **2009**, *113*, 6634.

(8) Wei, Z.; Hui, W.; Michael, R. H.; Taner Yildirim, D.-L. D. *J. Phys. Chem. C* **2007**, *111*, 16131.

(9) Sillar, K.; Hofmann, A.; Sauer, J. *J. Am. Chem. Soc.* **2009**, *131*, 4143.

(10) Han, S. S.; Goddard, W. A. *J. Am. Chem. Soc.* **2007**, *129*, 8422.

(11) Babarao, R.; Hu, Z.; Jiang, J.; Chempath, S.; Sandler, S. I. *Langmuir* **2006**, *23*, 659.

(12) Jasuja, H.; Zang, J.; Sholl, D. S.; Walton, K. S. *Journal of Physical Chemistry C* **2012**, *116*, 23526.

(13) Guo, H.-c.; Shi, F.; Ma, Z.-f.; Liu, X.-q. *Journal of Physical Chemistry C* **2010**, *114*, 12158.

(14) Keskin, S. *Journal of Physical Chemistry C* **2011**, *115*, 800.

(15) Boutin, A.; Coudert, F. X.; Springuel Huet, M. A.; Neimark, A. V.; Ferey, G.; Fuchs, A. H. *J. Phys. Chem. C* **2010**, *114*, 22237.

(16) Atci, E.; Erucar, I.; Keskin, S. *J. Phys. Chem. C* **2011**, *115*, 6833.

(17) Farha, O. K.; Yazaydin, A. O.; Eryazici, I.; Malliakas, C. D.; Hauser, B. G.; Kanatzidis, M. G.; Nguyen, S. T.; Snurr, R. Q.; Hupp, J. T. *Nature Chemistry* **2010**, *2*, 944.

(18) Furukawa, H.; Ko, N.; Go, Y. B.; Aratani, N.; Choi, S. B.; Choi, E.; Yazaydin, A. O.; Snurr, R. Q.; O'Keeffe, M.; Kim, J.; Yaghi, O. M. *Science* **2010**, *329*, 424.

(19) Fu, J.; Sun, H. *The Journal of Physical Chemistry C* **2009**, *113*, 21815.

(20) Cao, D. P.; Lan, J. H.; Wang, W. C.; Smit, B. *Angew Chem Int Edit* **2009**, *48*, 4730.

(21) Frost, H.; Düren, T.; Snurr, R. Q. *The Journal of Physical Chemistry B* **2006**, *110*, 9565.

(22) Han, S. S.; Goddard, W. A. *Journal of the American Chemical Society* **2007**, *129*, 8422.

(23) Yang, Q.; Zhong, C. *The Journal of Physical Chemistry B* **2005**, *110*, 655.

(24) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nat Chem* **2012**, *4*, 83.

(25) Babarao, R.; Jiang, J. W. *Langmuir* **2008**, *24*, 6270.

(26) Ravikovitch, P. I.; Domhnaill, S. C. O.; Neimark, A. V.; Schueth, F.; Unger, K. K. *Langmuir* **1995**, *11*, 4765.

(27) Maddox, M. W.; Olivier, J. P.; Gubbins, K. E. *Langmuir* **1997**, *13*, 1737.

(28) Neimark, A. V.; Ravikovitch, P. I. *Microporous and Mesoporous Materials* **2001**, *44–45*, 697.

(29) Occelli, M. L.; Olivier, J. P.; Petre, A.; Auroux, A. *The Journal of Physical Chemistry B* **2003**, *107*, 4128.

(30) Wu, J. Z. *Aiche Journal* **2006**, *52*, 1169.

(31) Wu, J. Z. In *Molecular Therodynamics of Complex Systems*; Lu, X., Hu, Y., Eds. 2009; Vol. 131, p 1.

(32) Wu, J. Z.; Li, Z. D. In *Annual Review of Physical Chemistry*; Annual Reviews: Palo Alto, 2007; Vol. 58, p 85.

(33) Rosenfeld, Y. *Phys. Rev. Lett.* **1989**, *63*, 980.

(34) Rosenfeld, Y. *Journal of Chemical Physics* **1993**, *98*, 8126.

(35) Yu, Y.-X.; Wu, J. *J. Chem. Phys.* **2002**, *117*, 10156.

(36) Roth, R.; Evans, R.; Lang, A.; Kahl, G. *Journal of Physics: Condensed Matter* **2002**, *14*, 12063.

(37) Tarazona, P.; Evans, R. *Molec. Phys.* **1984**, *52*, 847.

(38) Patra, C. N.; Ghosh, S. K. *Phys. Rev. E* **1993**, *48*, 1154.

(39) Jain, S.; Jog, P.; Weinhold, J.; Srivastava, R.; Chapman, W. G. *J. Chem. Phys.* **2008**, *128*, 154910.

(40) Ravikovitch, P. I.; Vishnyakov, A.; Neimark, A. V. *Phys. Rev. E* **2001**, *64*, 011602.
(41) Zhou, S. Q. *Int. J. Mod. Phys. B* **2005**, *19*, 4701.

(42) Ravikovitch, P. I.; Vishnyakov, A.; Neimark, A. V. *Physical Review E* **2001**, *64*, 011602.

(43) Tang, Y. P.; Wu, J. Z. *Physical Review E* **2004**, *70*, 011201.

(44) Zhou, S. Q. *J. Chem. Phys.* **1999**, *110*, 2140.

(45) Li, Z.; Wu, J. *Phys. Rev. E* **2004**, *70*, 031109.

(46) Yu, Y. X.; Wu, J.; Gao, G. H. *J. Chem. Phys.* **2004**, *120*, 7223.

(47) Siderius, D. W.; Gelb, L. D. *Langmuir* **2009**, *25*, 1296.

(48) Xu, X.; Cao, D. P. *J. Chem. Phys.* **2009**, *130*, 164901.

(49) Liu, Y.; Liu, H. L.; Hu, Y.; Jiang, J. W. *J. Phys. Chem. B* **2009**, *113*, 12326.

(50) Zhao, S.; Ramirez, R.; Vuilleumier, R.; Borgis, D. *J Chem Phys* **2011**, *134*.

(51) Liu, Y.; Liu, H. L.; Hu, Y.; Jiang, J. W. *J. Phys. Chem. C* **2010**, *114*, 2820.

(52) Meng, D.; Hjelm, Rex P.; Hu, J.; Wu, J. *Biophysical Journal* **2011**, *101*, 2476.

(53) Yu, Y.-X. *Journal of Chemical Physics* **2009**, *131*, 024704.

(54) Yu, Y.-X.; Wu, J. *The Journal of Chemical Physics* **2002**, *117*, 10156.

(55) Fletcher, A. J.; Thomas, K. M.; Rosseinsky, M. J. *J Solid State Chem* **2005**, *178*, 2491.

(56) Johnson, J. K.; John, A. Z.; Keith, E. G. D.-L. D. *Mol. Phys.* **1992**, *78*, 591.

(57) Hager, W.; Zhang, H. *SIAM Journal on Optimization* **2005**, *16*, 170.

(58) Frenkel, D.; Smit, B. *Understanding molecular simulation: from algorithms to applications*; Academic press, 2001; Vol. 1.

(59) Martin, M. G. http://towhee.sourceforge.net, 2013.

(60) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *Journal of the American Chemical Society* **1992**, *114*, 10024.

(61) Buch, V.; Devlin, J. P. *The Journal of Chemical Physics* **1993**, *98*, 4195.

(62) Yu, Y.-X. *The Journal of Chemical Physics* **2009**, *131*.

(63) Frost, H.; Duren, T.; Snurr, R. Q. *J Phys Chem B* **2006**, *110*, 9565.

(64) Zhou, W.; Wu, H.; Hartman, M. R.; Yildirim, T. *Journal of Physical Chemistry C* **2007**, *111*, 16131.

(65) Mavrandonakis, A.; Klopper, W. *Journal of Physical Chemistry C* **2008**, *112*, 3152.

(66) Frigo, M.; Johnson, S. G. *Proceedings of the IEEE* **2005**, *93*, 216.

(67) Keil, F. J.; Krishna, R.; Coppens, M. O. *Rev Chem Eng* **2000**, *16*, 71.

(68) Kärger, J.; Ruthven, D. M.; Theodorou, D. N. *Diffusion in nanoporous materials*; Wiley-VCH: Weinheim, Germany, 2012.

(69) Skoulidas, A. I.; Sholl, D. S. *J Phys Chem B* **2005**, *109*, 15760.

(70) Smit, B.; Krishna, R. *Chem Eng Sci* **2003**, *58*, 557.

(71) Deem, M. W.; Pophale, R.; Cheeseman, P. A.; Earl, D. J. *J Phys Chem C* **2009**, *113*, 21353.

(72) Haldoupis, E.; Nair, S.; Sholl, D. S. *Phys Chem Chem Phys* **2011**, *13*, 5053.

(73) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nat Chem* **2012**, *4*, 83.

(74) Wu, D.; Wang, C. C.; Liu, B.; Liu, D. H.; Yang, Q. Y.; Zhong, C. L. *Aiche J* **2012**, *58*, 2078.

(75) Ravikovitch, P. I.; Neimark, A. V. *Langmuir* **2006**, *22*, 11171.

(76) Cao, D.; Wu, J. *J. Chem. Phys.* **2004**, *121*, 4210.

(77) Yu, Y. X.; You, F. Q.; Tang, Y. P.; Gao, G. H.; Li, Y. G. *J. Phys. Chem. B* **2006**, *110*, 334.

(78) Marconi, U. M. B.; Melchionna, S. *Journal of Chemical Physics* **2011**, *134*, 064118.

(79) Goel, G.; Krekelberg, W. P.; Pond, M. J.; Mittal, J.; Shen, V. K.; Errington, J. R.; Truskett, T. M. *Journal of Statistical Mechanics-Theory and Experiment* **2009**, P04006.

(80) Mittal, J.; Errington, J. R.; Truskett, T. M. *Physical Review Letters* **2006**, *96*, 177804.

(81) Rosenfeld, Y. *Physical Review A* **1977**, *15*, 2545.

(82) Rosenfeld, Y. *J Phys-Condens Mat* **1999**, *11*, 5415.

(83) Dzugutov, M. *Nature* **1996**, *381*, 137.

(84) Dzugutov, M. *Physical Review E* **2002**, *65*, 032501.

(85) He, P.; Liu, H.; Zhu, J.; Li, Y.; Huang, S.; Wang, P.; Tian, H. *Chemical Physics Letters* **2012**, *535*, 84.

(86) Vaz, R. V.; Magalhaes, A. L.; Fernandes, D. L. A.; Silva, C. M. *Chem Eng Sci* **2012**, *79*, 153.

(87) Carmer, J.; Goel, G.; Pond, M. J.; Errington, J. R.; Truskett, T. M. *Soft Matter* **2012**, *8*, 4083.

(88) Chopra, R.; Truskett, T. M.; Errington, J. R. *Physical Review E* **2010**, *82*, 041201.

(89) Mittal, J.; Errington, J. R.; Truskett, T. M. *J Phys Chem B* **2007**, *111*, 10054.

(90) Rah, K.; Eu, B. C. *Journal of Chemical Physics* **2001**, *115*, 2634.

(91) Laghaei, R.; Eskandari Nasrabad, A.; Eu, B. C. *Journal of Chemical Physics* **2006**, *124*, 154502.

(92) Qin, Y.; Eu, B. C. *J Phys Chem B* **2009**, *113*, 4751.

(93) Cohen, M. H.; Turnbull, D. *Journal of Chemical Physics* **1959**, *31*, 1164.

(94) Mittal, J.; Errington, J. R.; Truskett, T. M. *Physical Review E* **2006**, *74*, 040102.

(95) Skoulidas, A. I.; Sholl, D. S. *J Phys Chem B* **2002**, *106*, 5058.

(96) Chopra, R.; Truskett, T. M.; Errington, J. R. *J Phys Chem B* **2010**, *114*, 10558.

(97) Agarwal, M.; Singh, M.; Jabes, B. S.; Chakravarty, C. *Journal of Chemical Physics* **2011**, *134*, 014502

(98) Goel, T.; Patra, C. N.; Mukherjee, T.; Chakravarty, C. *Journal of Chemical Physics* **2008**, *129*, 164904.

(99) Wu, J. Z.; Li, Z. D. *Annu Rev Phys Chem* **2007**, *58*, 85.

(100) Chui, S. S.-Y.; Lo, S. M.-F.; Charmant, J. P. H.; Orpen, A. G.; Williams, I. D. *Science* **1999**, *283*, 1148.

(101) Eddaoudi, M.; Li, H.; Yaghi, O. M. *Journal of the American Chemical Society* **2000**, *122*, 1391.

(102) Park, K. S.; Ni, Z.; Côté, A. P.; Choi, J. Y.; Huang, R.; Uribe-Romo, F. J.; Chae, H. K.; O'Keeffe, M.; Yaghi, O. M. *Proceedings of the National Academy of Sciences* **2006**, *103*, 10186.

(103) Baerlocher, C.; McCusker, L. 2013.

(104) Plimpton, S. *Journal of Computational Physics* **1995**, *117*, 1.

(105) Yang, R. T. *Gas separation by adsorption processes*; World Scientific: Singapore ; River Edge, N.J., 1997.

(106) Duong, D. D. *Adsorption analysis : equilibria and kinetics*; Imperial College Press: London, 1998.

(107) Myers, A. L.; Prausnitz, J. M. *Aiche Journal* **1965**, *11*, 121.

(108) Myers, A. L.; Monson, P. A. *Adsorption* **2014**, *20*, 591.

(109) Snurr, R. Q. *J Phys Chem Lett* **2011**, *2*, 1842.

(110) Ockwig, N. W.; Delgado-Friedrichs, O.; O'Keeffe, M.; Yaghi, O. M. *Accounts Chem Res* **2005**, *38*, 176.

(111) Keskin, S.; Liu, J.; Rankin, R. B.; Johnson, J. K.; Sholl, D. S. *Ind Eng Chem Res* **2009**, *48*, 2355.

(112) Jiang, J. W.; Babarao, R.; Hu, Z. Q. *Chem Soc Rev* **2011**, *40*, 3599.

(113) Yang, Q. Y.; Liu, D. H.; Zhong, C. L.; Li, J. R. *Chem Rev* **2013**, *113*, 8261.

(114) Getman, R. B.; Bae, Y. S.; Wilmer, C. E.; Snurr, R. Q. *Chem Rev* **2012**, *112*, 703.

(115) Colon, Y. J.; Snurr, R. Q. *Chem Soc Rev* **2014**, *43*, 5735.

(116) Sun, W. Z.; Lin, L. C.; Peng, X.; Smit, B. *Aiche J* **2014**, *60*, 2314.

(117) Tadmor, E. B.; Miller, R. E. *Modeling materials: continuum, atomistic, and multiscale techniques*; Cambridge University Press: Cambridge ; New York, 2011.

(118) Hohenberg, P.; Kohn, W. *Phys Rev B* **1964**, *136*, B864.

(119) Evans, R. *Adv Phys* **1979**, *28*, 143.

(120) Parr, R. G.; Yang, W. *Density-functional theory of atoms and molecules*; Oxford University Press New York, 1989.

(121) Lastoskie, C.; Gubbins, K. E.; Quirke, N. *The Journal of Physical Chemistry* **1993**, *97*, 4786.

(122) Jia Fu, Y. L., and Jianzhong Wu *Chemical Engineering Science* **2014**, *accepted*.

(123) Jia Fu, Y. L., Yun Tian and Jianzhong Wu *Journal of Physical Chemistry C* **2014**, *DOI: 10.1021/jp505963m*.

(124) Tsao, C. S.; Yu, M. S.; Wang, C. Y.; Liao, P. Y.; Chen, H. L.; Jeng, U. S.; Tzeng, Y. R.; Chung, T. Y.; Wu, H. C. *Journal of the American Chemical Society* **2009**, *131*, 1404.

(125) Liu, Y.; Fu, J.; Wu, J. *Langmuir* **2013**, *29*, 12997.

(126) Martin, M. G.; Siepmann, J. I. *The Journal of Physical Chemistry B* **1998**, *102*, 2569.

(127) Wu, J. In *Molecular thermodynamics of complex systems*; Lu, X., Hu, Y., Chen, H., Eds.; Springer: Berlin, 2009, p 1.

(128) Tang, Y. P.; Wu, J. Z. *Journal of Chemical Physics* **2003**, *119*, 7388.

(129) Eddaoudi, M.; Kim, J.; Rosi, N.; Vodak, D.; Wachter, J.; O'Keeffe, M.; Yaghi, O. M. *Science* **2002**, *295*, 469.

(130) Chui, S. S. Y.; Lo, S. M. F.; Charmant, J. P. H.; Orpen, A. G.; Williams, I. D. *Science* **1999**, *283*, 1148.

(131) Addicoat, M. A.; Coupry, D. E.; Heine, T. *J Phys Chem A* **2014**, *118*, 9607.
(132)

(133) Walton, K. S.; Snurr, R. Q. *Journal of the American Chemical Society* **2007**, *129*, 8552.

(134) Sarkisov, L.; Harrison, A. *Mol. Simul.* **2011**, *37*, 1248.

(135) Lozano-Castelló, D.; Alcañiz-Monge, J.; de la Casa-Lillo, M. A.; Cazorla-Amorós, D.; Linares-Solano, A. *Fuel* **2002**, *81*, 1777.

(136) He, Y.; Zhou, W.; Qian, G.; Chen, B. *Chemical Society Reviews* **2014**, *43*, 5657.

(137) Rosi, N. L.; Eckert, J.; Eddaoudi, M.; Vodak, D. T.; Kim, J.; O'Keeffe, M.; Yaghi, O. M. *Science* **2003**, *300*, 1127.

(138) Düren, T.; Sarkisov, L.; Yaghi, O. M.; Snurr, R. Q. *Langmuir* **2004**, *20*, 2683.

(139) Celzard, A.; Fierro, V. *Energy & Fuels* **2005**, *19*, 573.

(140) Ma, S.; Sun, D.; Simmons, J. M.; Collier, C. D.; Yuan, D.; Zhou, H. C. *J Am Chem Soc* **2008**, *130*, 1012.

(141) Suh, M. P.; Park, H. J.; Prasad, T. K.; Lim, D.-W. *Chemical Reviews* **2011**, *112*, 782.

(142) Sumida, K.; Rogow, D. L.; Mason, J. A.; McDonald, T. M.; Bloch, E. D.; Herm, Z. R.; Bae, T.-H.; Long, J. R. *Chemical Reviews* **2011**, *112*, 724.

(143) Liu, J.; Thallapally, P. K.; McGrail, B. P.; Brown, D. R.; Liu, J. *Chemical Society Reviews* **2012**, *41*, 2308.

(144) Makal, T. A.; Li, J.-R.; Lu, W.; Zhou, H.-C. *Chemical Society Reviews* **2012**, *41*, 7761.

(145) Peng, Y.; Krungleviciute, V.; Eryazici, I.; Hupp, J. T.; Farha, O. K.; Yildirim, T. *Journal of the American Chemical Society* **2013**, *135*, 11887.

(146) Mason, J. A.; Veenstra, M.; Long, J. R. *Chemical Science* **2014**, *5*, 32.
(147) Eddaoudi, M.; Moler, D. B.; Li, H.; Chen, B.; Reineke, T. M.; O'keeffe, M.; Yaghi, O. M. *Accounts of Chemical Research* **2001**, *34*, 319.

(148) Eddaoudi, M.; Kim, J.; Vodak, D.; Sudik, A.; Wachter, J.; O'Keeffe, M.; Yaghi, O. M. *Proc Natl Acad Sci U S A* **2002**, *99*, 4900.

(149) Yaghi, O. M.; O'Keeffe, M.; Ockwig, N. W.; Chae, H. K.; Eddaoudi, M.; Kim, J. *Nature* **2003**, *423*, 705.

(150)
(151) Peng, Y.; Srinivas, G.; Wilmer, C. E.; Eryazici, I.; Snurr, R. Q.; Hupp, J. T.; Yildirim, T.; Farha, O. K. *Chem. Commun.* **2013**, *49*, 2992.

(152) Wilmer, C. E.; Farha, O. K.; Yildirim, T.; Eryazici, I.; Krungleviciute, V.; Sarjeant, A. A.; Snurr, R. Q.; Hupp, J. T. *Energy & Environmental Science* **2013**, *6*, 1158.

(153) Guo, Z.; Wu, H.; Srinivas, G.; Zhou, Y.; Xiang, S.; Chen, Z.; Yang, Y.; Zhou, W.; O'Keeffe, M.; Chen, B. *Angewandte Chemie International Edition* **2011**, *50*, 3178.

(154) Ma, S.; Sun, D.; Simmons, J. M.; Collier, C. D.; Yuan, D.; Zhou, H.-C. *Journal of the American Chemical Society* **2007**, *130*, 1012.

(155) Wu, H.; Zhou, W.; Yildirim, T. *Journal of the American Chemical Society* **2009**, *131*, 4995.

(156) Gándara, F.; Furukawa, H.; Lee, S.; Yaghi, O. M. *Journal of the American Chemical Society* **2014**, *136*, 5271.

(157) Garberoglio, G.; Skoulidas, A. I.; Johnson, J. K. *J Phys Chem B* **2005**, *109*, 13094.

(158) Skoulidas, A. I.; Sholl, D. S. *The Journal of Physical Chemistry B* **2005**, *109*, 15760.

(159) Yang, Q.; Zhong, C. *J Phys Chem B* **2006**, *110*, 17776.

(160) Walton, K. S.; Snurr, R. Q. *J Am Chem Soc* **2007**, *129*, 8552.

(161) Liu, B.; Smit, B. *The Journal of Physical Chemistry C* **2010**, *114*, 8515.

(162) Gómez-Gualdrón, D. A.; Wilmer, C. E.; Farha, O. K.; Hupp, J. T.; Snurr, R. Q. *The Journal of Physical Chemistry C* **2014**, *118*, 6941.

(163) Jia Fu, Y. L., Yun Tian and Jianzhong Wu *Journal of Physical Chemistry C* **2014**, *accepted*.

(164) D'Alessandro, D. M.; Smit, B.; Long, J. R. *Angewandte Chemie International Edition* **2010**, *49*, 6058.

(165) Metz, B.; Davidson, O.; De Coninck, H.; Loos, M.; Meyer, L. *IPCC Special Report on Carbon Dioxide Capture and Storage*, 2005.

(166) Yang, H.; Xu, Z.; Fan, M.; Gupta, R.; Slimane, R. B.; Bland, A. E.; Wright, I. *Journal of Environmental Sciences* **2008**, *20*, 14.

(167) Phan, A.; Doonan, C. J.; Uribe-Romo, F. J.; Knobler, C. B.; O'keeffe, M.; Yaghi, O. M. *Acc. Chem. Res* **2010**, *43*, 58.

(168) Furukawa, H.; Ko, N.; Go, Y. B.; Aratani, N.; Choi, S. B.; Choi, E.; Yazaydin, A. Ö.; Snurr, R. Q.; O'Keeffe, M.; Kim, J.; Yaghi, O. M. *Science* **2010**, *329*, 424.

(169) Herm, Z. R.; Krishna, R.; Long, J. R. *Microporous and Mesoporous Materials* **2012**, *151*, 481.

(170) Ben, T.; Ren, H.; Ma, S.; Cao, D.; Lan, J.; Jing, X.; Wang, W.; Xu, J.; Deng, F.; Simmons, J. M.; Qiu, S.; Zhu, G. *Angewandte Chemie* **2009**, *121*, 9621.

(171) Lu, W.; Yuan, D.; Zhao, D.; Schilling, C. I.; Plietzsch, O.; Muller, T.; Bräse, S.; Guenther, J.; Blümel, J.; Krishna, R.; Li, Z.; Zhou, H.-C. *Chemistry of Materials* **2010**, *22*, 5964.

(172) Yuan, D.; Lu, W.; Zhao, D.; Zhou, H.-C. *Advanced Materials* **2011**, *23*, 3723.

(173) Lu, W.; Verdegaal, W. M.; Yu, J.; Balbuena, P. B.; Jeong, H.-K.; Zhou, H.-C. *Energy & Environmental Science* **2013**, *6*, 3559.

(174) Lu, W.; Sculley, J. P.; Yuan, D.; Krishna, R.; Wei, Z.; Zhou, H.-C. *Angewandte Chemie International Edition* **2012**, *51*, 7480.

(175) Babarao, R.; Dai, S.; Jiang, D.-e. *Langmuir* **2011**, *27*, 3451.

(176) Babarao, R.; Custelcean, R.; Hay, B. P.; Jiang, D.-e. *Crystal Growth & Design* **2012**, *12*, 5349.

(177) Yang, Z.; Peng, X.; Cao, D. *The Journal of Physical Chemistry C* **2013**, *117*, 8353.

(178) Islamoglu, T.; Gulam Rabbani, M.; El-Kaderi, H. M. *Journal of Materials Chemistry A* **2013**, *1*, 10259.

(179) Zhu, Y.; Long, H.; Zhang, W. *Chemistry of Materials* **2013**, *25*, 1630.

(180) Zhu, X.; Tian, C.; Mahurin, S. M.; Chai, S.-H.; Wang, C.; Brown, S.; Veith, G. M.; Luo, H.; Liu, H.; Dai, S. *Journal of the American Chemical Society* **2012**, *134*, 10478.

(181) Liu, L.; Deng, Q.-F.; Ma, T.-Y.; Lin, X.-Z.; Hou, X.-X.; Liu, Y.-P.; Yuan, Z.-Y. *Journal of Materials Chemistry* **2011**, *21*, 16001.

(182) Xia, Y.; Mokaya, R.; Walker, G. S.; Zhu, Y. *Advanced Energy Materials* **2011**, *1*, 678.

(183) Chandra, V.; Yu, S. U.; Kim, S. H.; Yoon, Y. S.; Kim, D. Y.; Kwon, A. H.; Meyyappan, M.; Kim, K. S. *Chemical Communications* **2012**, *48*, 735.
(184) Kresse, G.; Hafner, J. *Physical Review B* **1993**, *47*, 558.

(185) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J Phys Chem-Us* **1990**, *94*, 8897.

(186) Campañá, C.; Mussard, B.; Woo, T. K. *Journal of Chemical Theory and Computation* **2009**, *5*, 2866.

(187) M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox; Gaussian, Inc.: Wallingford CT, 2009.

(188) Breneman, C. M.; Wiberg, K. B. *J Comput Chem* **1990**, *11*, 361.

(189) Ewald, P. P. *Annalen der Physik* **1921**, *369*, 253.

(190) Li, W.; Shi, H.; Zhang, J. *ChemPhysChem* **2014**, *15*, 1772.

(191) Vogiatzis, K. D.; Mavrandonakis, A.; Klopper, W.; Froudakis, G. E. *ChemPhysChem* **2009**, *10*, 374.

(192) Du, N.; Park, H. B.; Robertson, G. P.; Dal-Cin, M. M.; Visser, T.; Scoles, L.; Guiver, M. D. *Nat Mater* **2011**, *10*, 372.

(193) Sevilla, M.; Valle-Vigón, P.; Fuertes, A. B. *Advanced Functional Materials* **2011**, *21*, 2781.

(194) Rabbani, M. G.; El-Kaderi, H. M. *Chemistry of Materials* **2012**, *24*, 1511.

(195) Rabbani, M. G.; Sekizkardes, A. K.; El-Kadri, O. M.; Kaafarani, B. R.; El-Kaderi, H. M. *Journal of Materials Chemistry* **2012**, *22*, 25409.

Chapter 4. Fast Screening Method for Predicting Hydration Free Energy

4.1 High-throughput prediction of the hydration free energies of small molecules from a classical density functional theory

ABSTRACT

The classical density functional theory (DFT) is proposed as an efficient computational tool for accurate prediction of the solvation free energies of small molecules in liquid water at the ambient condition. With the solute molecules represented by the AMBER force field and the TIP3P model for the solvent, the new theoretical method predicts the hydration free energies of 500 neutral molecules with average unsigned errors of 0.96 kcal/mol and 1.04 kcal/mol in comparison with the experimental and simulation data, respectively. The DFT predictions are orders of magnitude faster than conventional molecular dynamics simulations and the numerical performance can be further improved by taking into account the molecular flexibility of large solutes.

Solvation free energy plays a key role in solution chemistry and its theoretical prediction often represents a bottleneck in understanding important chemical and biological processes in water including protein-ligand bindings[1,2]. The solvent-solute interactions entail complex microscopic details that make a reliable prediction of solvation properties a serious computational challenge. The conventional methods for solvation free energy calculations are based on either the solvent-implicit models or the solvent-explicit models. The former are usually constructed from knowledge-based macroscopic considerations, while the latter often start with a semi-empirical force field to represent the solvent-solvent and solvent-solute interactions[3]. The continuous and solvent-explicit methods are complementary and their choices for practical applications often reflect a compromise of the computational cost and the precision in microscopic details.

With the solvent molecules depicted as a dielectric continuum, an implicit-solvent model describes the solvation free energy in terms of the geometric measures of the solute-solvent boundary, such as the solvent-accessible surface (SAS), and various energetic contributions due to the solute-solvent electrostatic and van der Waals interactions[4]. A recent example for the implicit-solvent approach was provided by Boyer and Bryan.[5] who correlated the hydration free energy with the SAS and the partial charges of the solute molecules. Impressive fitting of experimental results was achieved with a mean absolute error of 0.513 kcal/mol for the hydration free energies of diverse chemical species. While application of a continuous model drastically oversimplifies solvent-solute interactions and thus reduces the computational cost for predicting the solvation free energy and solvent-mediated interactions, it neglects the local solvent inhomogeneity and the steric effects

affiliated with individual solvent molecules. Recently, Nakamura *et al.*[6] introduced an elegant procedure to account for the local dielectric inhomogeneity near the solute using field-theoretic techniques. The new theoretical method predicts ionic solvation free energies in both single-component liquids and binary liquid mixtures in excellent agreement with the experimental data. The general applicability of this method to more complicated chemical systems is yet to be established.

The explicit-solvent models, which have been widely used in molecular simulation and theoretical investigation of solvation, provide a more realistic representation of the solute-solvent interactions on the length scale pertinent to the solvent-solute interactions. Recent years have witnessed rapid progresses in development of novel theoretical and simulation methods for efficient predictions of solvation free energies with explicit solvents. Analytical predictions of solvation free energies are mostly based on the integral-equation theories [7,8] or the classical density functional methods[9-15]. In particular, the molecular density functional theory (MDFT) of solvation recently proposed by Borgis and coworkers offers an excellent platform to study solvation and solvent-mediated interactions[16-18]. However, the MDFT as well as molecular Ornstein-Zernike (MOZ) theory in the integral-equation theories requires multi-dimensional density profiles of the solvent molecules which is computationally demanding. And there have been several reports on large-scale simulation of the solvation free energies. One of the early examples was given by Duff and Jorgensen[19] who calculated the solvation properties of over 200 organic solutes in aqueous and organic solutions using Monte Carlo (MC) simulation. The simulation results were used to establish empirical correlations between the solvation properties and

a few "descriptors" similar to those used in a continuous model. More recently, Mobley *et al*. [20] calculated the hydration free energies for 504 small organic molecules from molecular dynamics (MD) simulations. The simulation results yield a mean absolute error of 1.24 kcal/mol in comparison with the experimental data. Different from the popular continuous model proposed by Chothia[21], the simulation studies indicate that the hydration free energies of a nonpolar solute is not simply proportional to the surface area or the solvent-excluded volume. While simulation of hydration free energies with an explicit solvent is in general extremely time consuming, advanced simulation techniques have been developed to reduce the computational cost. For example, Okur *et al*. [22] proposed a hybrid procedure to implement the replica exchange molecular dynamics (REMD). The modified parallel tempering method is able to reproduce the explicit-solvent results but with much reduced simulation time.

In comparison with the simulation methods, a major advantage of liquid-state theories is that it is able to predict thermodynamic properties including solvation free energy with little computational cost. While previous publications are often focused on calibration of theoretical developments, yet to be established is their capability for high throughput computations where the numerical efficiency becomes utmost importance. In this work, we demonstrate such capability by predicting the solvation free energies for 500 small molecules in water using a recently developed classical density functional theory (DFT)[14]. We show that the theoretical method is able to reproduce the simulation data with the accuracy comparable to that inherited by the molecular models.

To predict hydration free energies at an ambient condition, we consider a large set of solute molecules that were studied before by Mobley *et al*. with both continuous and solvent-explicit models[20,23]. As in the solvent-explicit MD simulation, the DFT calculations are based on the AMBER force field[24] for the solute molecules and the solvent is represented by the TIP3P water[25]. We use the conventional Coulomb and the Lennard-Jones (LJ) potentials with the Lorentz-Berthelot mixing rule to describe the non-bonded interactions. All solute molecules are assumed to have a rigid conformation with the atomic coordinates obtained from the MD simulation[20]. The DFT calculations were performed in a $(30\ \text{Å})^3 \sim (40\ \text{Å})^3$ cubic box with the atomic density profiles of the solvent molecules discretized onto a 128^3 lattice. (we have also test a few cases by using larger number of discretization points and the results are similar).The computational domain varies slightly for different solutes in accordance with their size. The supporting material provides the details of the theoretical and numerical methods.

Figure 4-1(a) compares the DFT predictions with the experimental results for the hydration free energies of 500 small molecules. The numerical data are listed in Table S4-1 of Supporting Information. Overall the theoretical predictions agree well with the experimental data, in particular for hydrophobic solutes, i.e., those molecules with a positive or small hydration free energy. For all systems considered, the average unsigned error (Δ) is 0.96 kcal/mol and the root mean square (RMS) deviation is 1.29 kcal/mol. The numbers are very close to those from the MD simulations, Δ = 1.03 kcal/mol and RMS = 1.26 kcal/mol, respectively[20].

The good performance of the theoretical method is in part due to the cancelation of errors introduced by in the AMBER force field and approximations for formulation of the DFT functional. The latter can be examined by a direct comparison of the DFT and MD simulation results. As shown in Figure 4-1(b), the DFT accords with the simulation[20] to an extent similar to that with the experimental data. The average unsigned error is 1.04 kcal/mol (RMS = 1.47 kcal/mol), indicating that the numerical error introduced by the DFT is statistically unsubstantial in comparison to that intrinsically affiliated with the solvent-explicit molecular force field.

A closer examination of Figures 4-1(a) and 4-1(b) suggests that the theoretical performance deteriorates slightly as the solvation free energy becomes more negative. We suspect that the increased discrepancy for the hydrophilic solutes is related to the approximations used in the DFT calculations. All procedures in DFT are exact except the bridge functional, which may be the key reason that generated the deviations. In formulation of the free-energy density functional, we utilize a hard-sphere model to represent the bridge functional for water wherein the hard-sphere diameter of the reference system was calibrated with the solvation free energy for methane[14]. Because the local structure of water molecules is sensitive to the solute hydrophobicity, the parameter obtained from a hydrophobic solute could become less reliable for hydrophilic solutes. As the bridge functional usually contributes as a negative correction of the solvation free energy,[14] a smaller hard-sphere diameter for hydrophilic solutes would result in a higher solvation free energies. As the effective hard-sphere diameter for a hydrophilic solute is smaller than that for a hydrophobic solute due to the enhanced solute-solvent attraction,

the approximate bridge functional from the hard-sphere reference system explains at least in part why the DFT underestimates the solvation free energies of most hydrophilic solutes in comparison with the simulation data.

We should indicate that the theoretical performance of the DFT is not entirely related to the polarity or hydrophobicity of solute molecules. To examine the effect of the solute polarity, we present in Figure 4-2 the distributions of the numerical errors affiliated with the DFT predictions relative to the experiment and MD simulation. Neither case shows a significant correlation between the DFT performance and the dipole moments of the solute molecules. Apparently, some solutes have a large dipole moment but a positive hydration free energy. In other words, the dipole moment is not a good indication of the hydrophobicity of a solute molecule.

In addition to the hydrophobicity of solute molecules, we have inspected possible correlations between the numerical errors and the solute flexibility. The hydration free energy of a flexible molecule, F_s, depends on its bond flexibility both in the solution and in vacuum:

$$\beta F_s = -\ln \left\langle \exp[-\beta F(X)] \right\rangle_0 \tag{87}$$

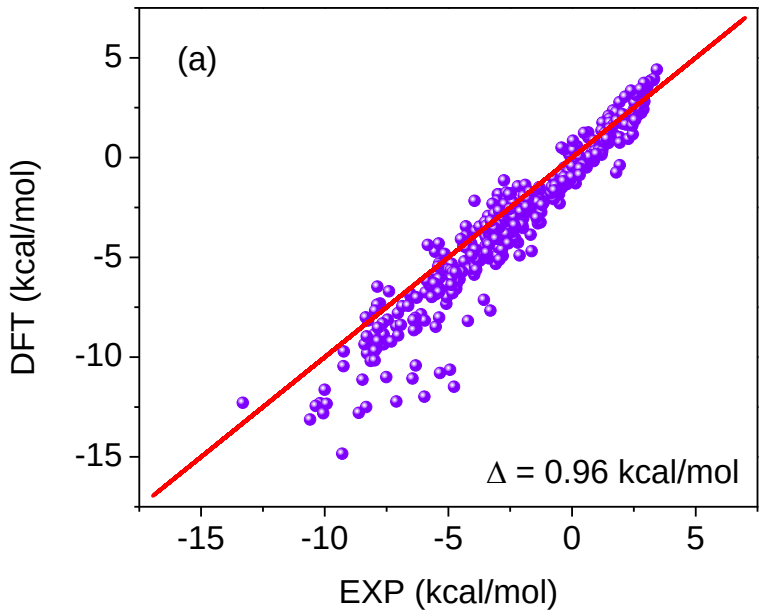
where $\beta = 1/(k_B T)$, k_B stands for the Boltzmann constant, T is the absolute temperature, $F(X)$ is the solvation free energy of a solute conformation X, and $<\cdots>_0$ denotes the ensemble average of the solute conformations in vacuum. As mentioned above, the DFT calculations are based on rigid conformations of the solute molecules generated by the MD simulation in water. To examine the effect of the molecular flexibility on the solvation free

energy, we picked 33 solutes corresponding to the worst DFT predictions in comparison with the experimental data and recalculated their solvation free energies using Eq.87. The ensemble average was estimated from 12 random conformations for each solute sampled from MD simulation in vacuum. Although 12 random conformations may be not enough to get the conformational averaged solvation free energy precisely, it at least gives us the insight of dependence between solvation free energy and flexibility. Figure 4-3 shows the DFT results for 33 outlying solutes in comparison with the experiment and the MD simulation. The numerical values for the solvation free energies are listed in Table S4-2 of the supporting information. We see that the DFT results can be significantly improved by sampling the solute flexibility even with a relatively small number of equilibrium conformations in vacuum. The improvement is around 0 ~ 4 kcal/mol in most cases but for some special ones such as "2_ethoxyethanol", it can be as large as 6 kcal/mol. For a large hydrophilic solute, such as "2_ethoxyethanol", "2_methoxy_111_trimethoxyethane" and "hydrazine", the DFT predictions indicate that the solvation free energy is highly sensitive to the molecular conformation, with a variation in the range of 8 ~ 9 kcal/mol. By contrast, the variation is much smaller for weakly hydrophilic solutes such as "acetonitrile" and "propyne", within the range of 1 kcal/mol. If a single conformation must be used for each solute, the theoretical results based on the conformations in vacuum are slightly better than those from the solvated conformations. Additionally, other undetermined elements such as polarizabilities may also lead to the inaccuracy of the theory. And to implement these elements will be our future focus.

In comparison with MD simulations, one key advantage of DFT is its numerical efficiency for high throughput calculations. For the systems considered in this work, each DFT calculation can be readily processed in a desktop computer with a single CPU (Nehalem 2.67GHz). The time cost varies from 15 to 120 minutes, depending on the solute type. For the 500 solutes shown in Figure 4-1, the average time cost is 42 minutes. The computational cost can be further reduced by optimization of the DFT programs. Up to now, any molecular simulation (usually required hundreds of CPU hours [26]) cannot match such a speed for calculations of hydration free energies with an explicit solvent model.

In conclusion, we have demonstrated that the density functional theory (DFT) provides a powerful alternative to conventional simulation methods for high throughput predictions of the hydration free energies of small molecules. Testing over 500 neutral solutes indicates that the DFT predictions yield an average unsigned error of 0.96 kcal/mol in comparison to the experiment. The mean deviation is slightly lower than that predicted by the MD simulation (1.03 kcal/mol). The theoretical results are most accurate for solutes with slightly positive hydration free energies but become less precise for large hydrophilic solutes. By inspecting possible causes of error for 33 worst performed solutes, we found that the theoretical performance of the DFT can be further improved by considering the solute flexibility. If a rigid model is used for the solutes, the DFT yields better results based on the equilibrium conformations in vacuum. In comparison to solvent-explicit simulations, the DFT calculations are extremely fast. The average time cost is only 42 minutes in just a single CPU. Such computational efficiency is highly desirable for high throughput solvation free energy calculations.

Figure 4-1 Hydration free energies for 500 solutes (a) comparison of DFT and experiment; (b) comparison of DFT and MD simulation.



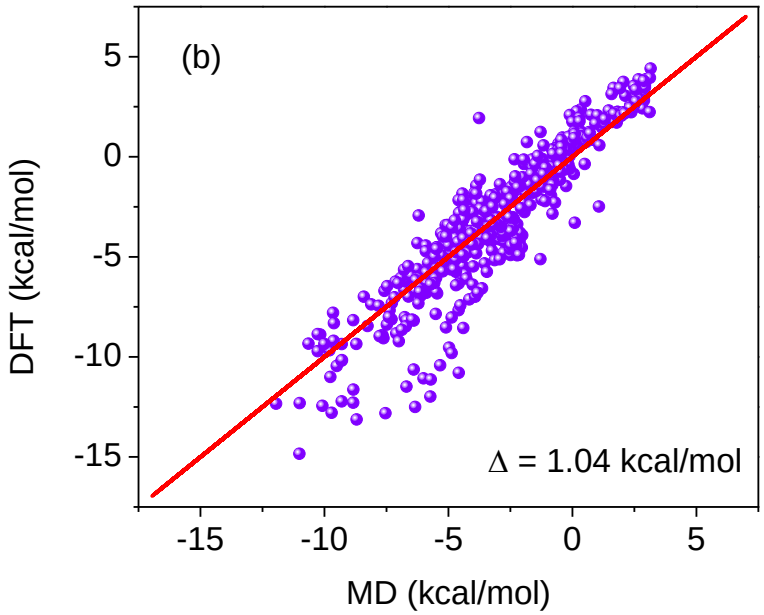


Figure 4-2 Distributions of the numerical errors versus the solute dipole moments relative to experiment (a) and to MD simulation (b).

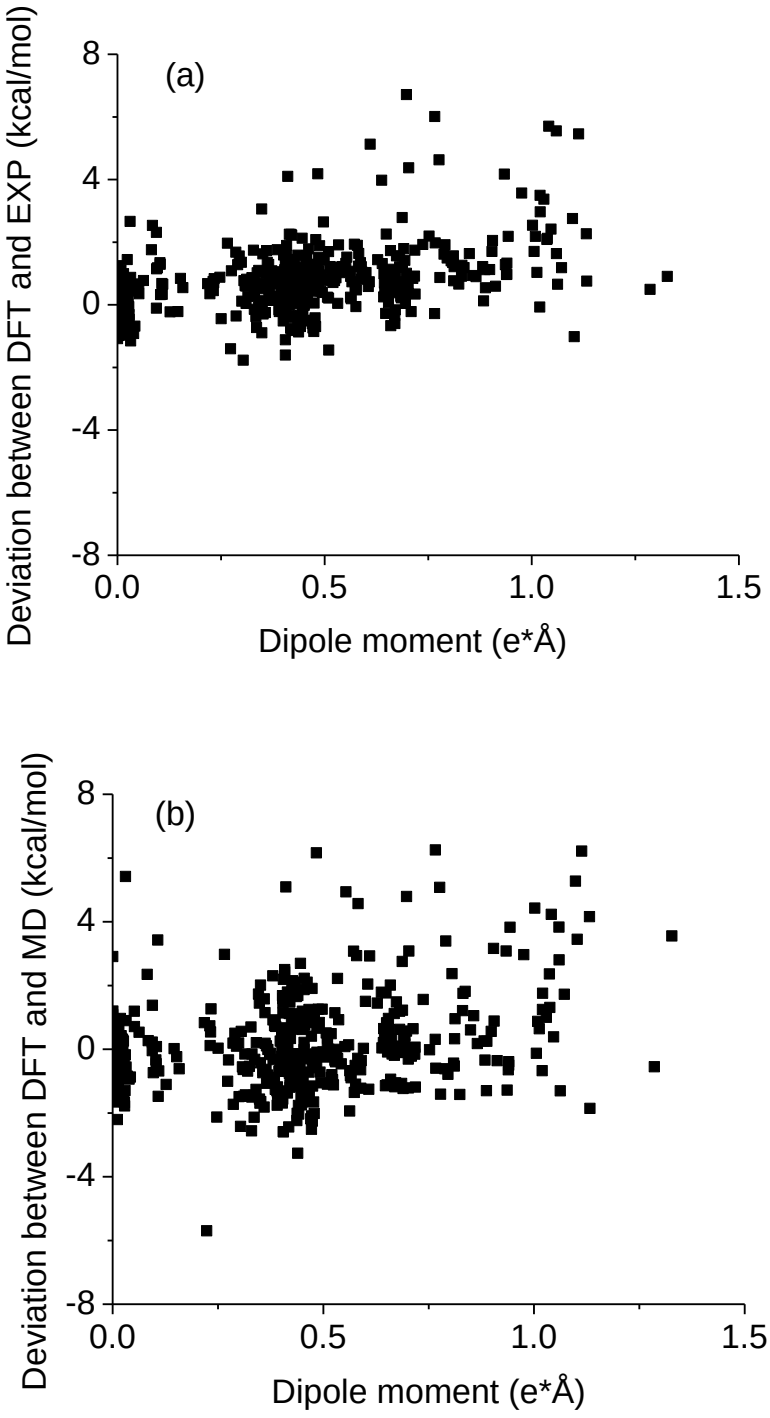
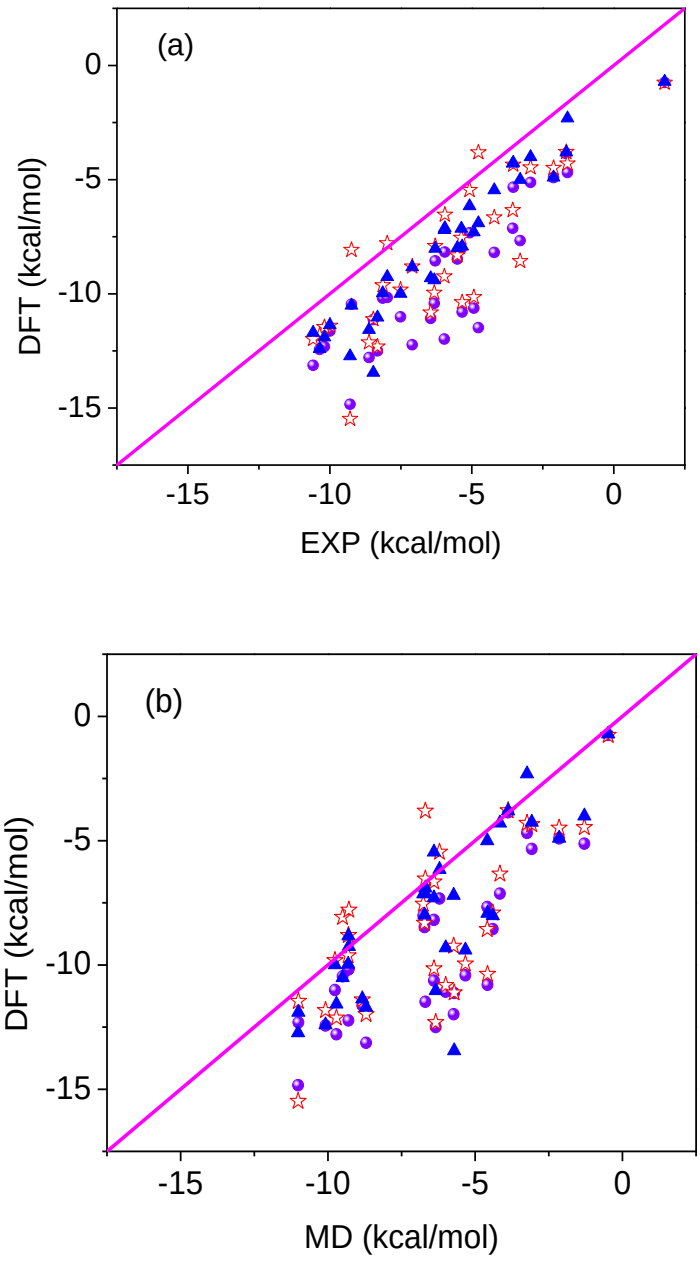


Figure 4-3 Comparisons of the hydration free energies predicted from the DFT for 33 "worst" solutes with experiment (a) and MD simulation (b). Spheres: initial conformation generated from AMBER; triangles: average of 12 random conformations generated from simulation in vacuum; stars: equilibrium conformation in vacuum.



Support Information

1. Numerical data for solvation free energy

Table S4-1: Comparing solvation free energy of 500 molecules (kcal/mol)

Name	DFT	Experiment	MD
1_bromo_2_chloroethane	-0.36	0.02	-1.95
1_bromo_2_methylpropane	0.74	1.24	-0.03
1_bromobutane	0.96	1.68	-0.4
1_bromoheptane	2.36	1.66	0.34
1_bromohexane	2.27	1.78	0.18
1_bromooctane	2.78	1.92	0.52
1_bromopentane	2.08	1.48	-0.1
1_bromopropane	0.66	1.29	-0.56
1_chloro_222_trifluoroethane	-0.86	0.16	0.06
1_chlorobutane	0.56	0.99	-0.16
1_chloroheptane	1.95	1.53	0.29
1_chlorohexane	1.75	1.23	0
1_chloropentane	0.82	1.18	-0.07
1_chloropropane	0.36	0.92	-0.33
1_ethylnaphthalene	-1.84	-3	-2.4
1_iodobutane	-0.06	0.58	-0.25
1_iodoheptane	1.36	1.19	0.27
1_iodohexane	1.26	0.66	0.08
1_iodopentane	1.01	0.66	-0.14
1_iodopropane	-0.38	0	-0.53
1_methyl_imidazole	-6.99	-6.33	-8.41
1_methyl_pyrrole	-3.63	-2.59	-2.89
1_methylcyclohexene	1.23	1.33	0.67
1_methylnaphthalene	-3.4	-3.24	-2.44
1_naphthylamine	-7.31	-7.76	-7.28
1_nitrobutane	-1.87	-1.51	-3.09
1_nitropentane	-1.63	-1.29	-2.82
1_nitropropane	-2.11	-1.38	-3.34
11_diacetoxyethane	-9.54	-7.9	-4.97
11_dichloroethane	-0.59	0.17	-0.84
11_dichloroethene	0.47	1.24	0.25
11_diethoxyethane	-4.26	-2.18	-3.28
11_difluoroethane	-0.6	0.19	-0.11

111_trichloroethane	0.12	0.75	-0.19
111_trifluoro_222_trimethoxyethane	-2.84	-2.26	-0.8
111_trifluoropropan_2_ol	-7.13	-3.56	-4.16
111_trimethoxyethane	-5.45	-3.89	-4.42
1112_tetrachloroethane	-0.44	0.11	-1.28
112_trichloro_122_trifluoroethane	1.75	1.53	1.77
112_trichloroethane	-2.29	-0.5	-1.99
1122_tetrachloroethane	-1.52	-0.41	-2.47
12_diacetoxyethane	-12.5	-8.32	-6.34
12_dibromoethane	-0.12	0.19	-2.33
12_dichlorobenzene	-1.27	-0.64	-1.36
12_dichloroethane	-1.83	-0.52	-1.79
12_dichloropropane	-1.33	-0.2	-1.27
12_diethoxyethane	-5.33	-3.08	-3.54
12_dimethoxyethane	-5.06	-3.36	-4.84
12_ethanediol	-9.36	-7.62	-9.3
123_trichlorobenzene	-0.93	-0.38	-1.24
123_trimethylbenzene	-1.33	-0.65	-1.21
1234_tetrachlorobenzene	-0.51	-0.32	-1.34
1235_tetrachlorobenzene	-0.13	-0.02	-1.62
124_trichlorobenzene	-0.69	-0.18	-1.12
124_trimethylbenzene	-1.16	-0.36	-0.86
1245_tetrachlorobenzene	-0.22	0.25	-1.34
13_dichlorobenzene	-0.95	0.01	-0.98
13_dichloropropane	-1.09	-0.38	-1.89
13_dimethylnaphthalene	-1.86	-2.79	-2.47
135_trichlorobenzene	-0.19	0.45	-0.78
135_trimethylbenzene	-0.84	0.04	-0.9
14_dichlorobenzene	-1.16	-0.1	-1.01
14_dichlorobutane	-1.26	-0.27	-2.32
14_dimethyl_piperazine	-6.7	-7.4	-7.58
14_dimethylnaphthalene	-3.09	-3.29	-2.82
14_dioxane	-5.2	-4.35	-5.06
2_bromo_2_methylpropane	0.84	0.89	0.84
2_bromopropane	0.58	0.97	-0.48
2_butoxyethanol	-4.31	-4.07	-6.25
2_chloro_111_trimethoxyethane	-7.67	-3.3	-4.59
2_chloro_2_methylpropane	0.58	0.82	1.09
2_chloroaniline	-5.63	-4.96	-4.91

2_chlorobutane	0.43	1.21	0
2_chlorophenol	-3.86	-3.23	-4.55
2_chloropropane	0.28	0.82	-0.25
2_chloropyridine	-4.99	-3.36	-4.39
2_chlorotoluene	-1.19	-0.51	-1.14
2_ethoxyethanol	-11.48	-4.77	-6.69
2_ethylpyrazine	-6.83	-5.68	-5.45
2_ethylpyridine	-4.27	-3.21	-4.33
2_ethyltoluene	-1.29	-0.66	-1.04
2_fluorophenol	-3.83	-3.24	-5.29
2_iodophenol	-2.94	-2.91	-6.2
2_iodopropane	-0.37	0.29	-0.46
2_isobutylpyrazine	-4.99	-5.1	-5.04
2_methoxy_111_trimethoxyethane	-11.98	-5.97	-5.73
2_methoxyaniline	-6.96	-6.53	-6.12
2_methoxyethanamine	-6.69	-5.62	-6.55
2_methoxyethanol	-8.01	-5.37	-6.76
2_methoxyphenol	-5.56	-4.74	-5.57
2_methyl_but_2_ene	1.65	2.38	1.31
2_methylbut_2_ene	1.65	2.28	1.31
2_methylbuta_13_diene	1.03	1.81	0.68
2_methylbutan_1_ol	-2.76	-2.78	-4.42
2_methylbutan_2_ol	-1.84	-2.96	-4.43
2_methylbutane	2.22	2.52	2.38
2_methylhexane	3.59	3	2.93
2_methylpent_1_ene	2.31	2.75	1.47
2_methylpentan_2_ol	-2.49	-2.79	-3.92
2_methylpentan_3_ol	-1.44	-2.19	-3.88
2_methylpentane	2.59	2.78	2.51
2_methylpropan_1_ol	-2.74	-2.96	-4.5
2_methylpropan_2_ol	-3.35	-3.09	-4.47
2_methylpropane	2.24	2.74	2.32
2_methylpropene	1.77	2.34	1.16
2_methylpyrazine	-7.86	-6.1	-5.51
2_methylpyridine	-4.63	-3.41	-4.63
2_methyltetrahydrofuran	-1.92	-1.95	-3.3
2_methylthiophene	-0.77	-0.22	-1.38
2_naphthol	-7.38	-7.88	-8.11
2_naphthylamine	-6.46	-7.87	-7.47

2_nitroaniline	-7.68	-7.96	-7.37
2_nitrophenol	-10.8	-5.34	-4.58
2_nitropropane	-2.05	-1.61	-3.13
2_nitrotoluene	-3.91	-3.14	-3.58
2_phenylethanol	-6.62	-5.33	-6.79
2_propoxyethanol	-8.19	-4.21	-6.4
22_dimethylbutane	2.47	2.53	2.51
22_dimethylpentane	2.96	2.9	2.88
22_dimethylpropane	2.52	2.6	2.51
222_trifluoroethanol	-5.87	-3.95	-4.31
224_trimethylpentane	2.81	2.93	2.89
225_trimethylhexane	3.46	2.86	2.93
23_dimethylbuta_13_diene	1.16	1.69	0.4
23_dimethylbutane	2.59	2.69	2.34
23_dimethylnaphthalene	-1.91	-2.59	-2.78
23_dimethylpentane	2.76	2.66	2.52
23_dimethylphenol	-6.35	-4.67	-6.16
23_dimethylpyridine	-4.47	-3.27	-4.82
234_trimethylpentane	3.05	2.86	2.56
24_dimethylpentan_3_one	-1.8	-2.47	-2.74
24_dimethylpentane	2.42	2.89	2.83
24_dimethylphenol	-6.11	-4.55	-6.01
24_dimethylpyridine	-4.4	-3.18	-4.86
25_dimethylphenol	-4.43	-4.39	-5.91
25_dimethylpyridine	-3.6	-2.83	-4.72
25_dimethyltetrahydrofuran	-1.37	-1.89	-2.92
26_dimethylaniline	-5.32	-4.87	-5.21
26_dimethylnaphthalene	-1.75	-2.65	-2.63
26_dimethylphenol	-4.93	-4.09	-5.26
26_dimethylpyridine	-4.01	-3.36	-4.59
3_acetylpyridine	-8.46	-7.1	-8.26
3_chloroaniline	-5.31	-5.37	-5.82
3_chlorophenol	-5.94	-5.28	-6.62
3_chloroprop_1_ene	-0.17	0.88	-0.57
3_chloropyridine	-3.92	-2.5	-4.01
3_cyanophenol	-7.79	-7.04	-9.65
3_cyanopyridine	-6.22	-4.66	-6.75
3_ethylphenol	-5.61	-5.02	-6.25
3_ethylpyridine	-3.69	-2.96	-4.59

3_formylpyridine	-8.82	-7.64	-7.1
3_hydroxybenzaldehyde	-10.46	-9.24	-9.5
3_methoxyaniline	-8.12	-7.45	-7.29
3_methoxyphenol	-8.91	-7.03	-7.66
3_methyl_1h_indole	-6.8	-6.55	-5.88
3_methyl_but_1_ene	2.24	2.8	1.83
3_methylbut_1_ene	2.24	2.67	1.82
3_methylbutan_1_ol	-3.57	-2.92	-4.42
3_methylbutan_2_one	-3.4	-2.93	-3.24
3_methylbutanoic_acid	-6.85	-5.51	-6.09
3_methylheptane	3.87	3.29	2.97
3_methylhexane	2.89	2.76	2.71
3_methylpentane	2.62	2.72	2.51
3_methylpyridine	-4.1	-3.19	-4.77
3_nitroaniline	-8.17	-8.24	-8.84
3_nitrophenol	-8.32	-7.66	-9.62
3_nitrotoluene	-3.7	-3.16	-3.45
3_phenylpropanol	-6.06	-5.5	-6.92
33_dimethylbutan_2_one	-3.75	-2.91	-3.11
33_dimethylpentane	2.76	2.53	2.56
333_trimethoxypropionitrile	-10.63	-4.93	-6.4
34_dimethylphenol	-5.94	-5.33	-6.5
34_dimethylpyridine	-3.99	-3.18	-5.22
35_dimethylphenol	-6.07	-5.12	-6.27
35_dimethylpyridine	-3.48	-2.94	-4.84
4_acetylpyridine	-9.07	-7.63	-7.62
4_bromophenol	-6.24	-5.47	-7.13
4_bromotoluene	-1.12	0.04	-1.39
4_chloro_3_methylphenol	-5.64	-4.9	-6.79
4_chloroaniline	-5.99	-5.28	-5.9
4_chlorophenol	-6.31	-5.36	-7.03
4_cyanophenol	-8.89	-7.61	-10.17
4_cyanopyridine	-6.71	-4.97	-6.02
4_ethylphenol	-6.08	-5.08	-6.13
4_ethylpyridine	-4.08	-2.87	-4.73
4_ethyltoluene	-0.5	-0.05	-0.95
4_fluorophenol	-6.07	-4.99	-6.19
4_formylpyridine	-9.22	-7.31	-7
4_hydroxybenzaldehyde	-11.63	-10	-8.83

4_isopropyltoluene	0.43	0.2	-0.68
4_methoxyacetophenone	-8.56	-6.29	-4.4
4_methoxyaniline	-8.39	-6.92	-7.48
4_methyl_1h_imidazole	-8.86	-7.99	-10.27
4_methylacetophenone	-5.75	-4.82	-4.7
4_methylbenzaldehyde	-6.09	-4.9	-4.27
4_methylpentan_2_ol	-1.14	-2.75	-3.73
4_methylpentan_2_one	-3	-2.96	-3.05
4_methylpyridine	-4.64	-3.41	-4.93
4_n_propylphenol	-5.78	-5.5	-5.9
4_nitroaniline	-9.72	-9.23	-10.27
4_nitrophenol	-9.34	-8.22	-10.64
4_tert_butylphenol	-4.72	-5.54	-5.91
acenaphthene	-3.86	-3.51	-3.15
acetaldehyde	-4.11	-3.39	-3.5
acetic_acid	-8.17	-5.95	-6.69
acetonitrile	-3.86	-1.67	-3.88
acetophenone	-6.33	-5.07	-4.58
alpha_methylstyrene	-2.14	-1.26	-1.24
ammonia	-4.55	-4.04	-4.29
aniline	-6.76	-5.92	-5.49
anisole	-4.03	-2.3	-2.45
anthracene	-4.31	-5.39	-3.95
azetidine	-4.99	-3.41	-5.56
benzaldehyde	-6.39	-4.99	-4.02
benzamide	-12.31	-10.19	-11
benzene	-1.94	-0.7	-0.86
benzonitrile	-3.71	-2.74	-4.21
benzotrifluoride	-1.46	-0.55	-0.25
benzyl_alcohol	-5.47	-5.41	-6.62
benzyl_bromide	-2.23	-1.14	-2.38
benzyl_chloride	-2.77	-1.56	-1.93
biphenyl	-2.93	-3.39	-2.66
bromobenzene	-1.44	-0.37	-1.46
bromoethane	0.31	0.93	-0.74
bromomethane	0.36	1.09	-0.82
bromotrifluoromethane	1.71	2.03	1.79
but_1_ene	1.92	2.48	1.38
but_1_yne	-0.38	1.93	-0.16

buta_13_diene	1.02	1.93	0.61
butan_1_ol	-4.04	-3.14	-4.72
butan_2_ol	-4.26	-3.12	-4.62
butanenitrile	-3.03	-1.41	-3.64
butanoic_acid	-6.69	-5.46	-6.35
butanone	-4.01	-2.97	-3.71
butyraldehyde	-4.01	-3.01	-3.18
chlorobenzene	-1.49	-0.6	-1.12
chlorodifluoromethane	-1.15	-0.04	-0.5
chloroethane	-0.04	0.78	-0.63
chloroethylene	0.3	1.22	-0.59
chlorofluoromethane	-1.18	-0.22	-0.77
chloromethane	0.09	0.81	-0.55
cis_12_dimethylcyclohexane	2.05	2.05	1.58
cyanobenzene	-3.71	-2.38	-4.1
cyclohepta_135_triene	-1.29	0.15	-0.99
cycloheptanol	-4.54	-4.15	-5.48
cyclohexane	1.63	1.67	1.23
cyclohexanol	-4.52	-4.26	-5.46
cyclohexanone	-4.75	-3.84	-4.91
cyclohexene	1.03	1.56	0.37
cyclohexylamine	-2.17	-3.94	-4.59
cyclopentane	1.45	1.53	1.2
cyclopentanol	-5.59	-4.15	-5.49
cyclopentanone	-5.13	-3.71	-4.7
cyclopentene	0.74	1.38	0.56
cyclopropane	2.09	2.64	0.75
decan_1_ol	-2.33	-2.49	-3.64
decan_2_one	-3.53	-1.98	-2.34
di_isopropyl_sulfide	-0.11	0.1	-1.21
di_n_butyl_ether	-0.49	0.59	-0.83
di_n_butylamine	-4.69	-1.63	-3.24
di_n_propyl_ether	0.57	0.21	-1.16
di_n_propyl_sulfide	1.23	0.49	-1.28
di_n_propylamine	-2.39	-2.26	-3.65
dibromomethane	-0.14	0.87	-1.96
dichloromethane	-0.77	0.23	-1.31
diethoxymethoxybenzene	-5.73	-4.44	-5.23
diethyl_disulfide	-0.54	-0.07	-1.64

diethyl_ether	-1.79	-0.7	-1.59
diethyl_malonate	-11.08	-6.45	-6
diethyl_succinate	-11.13	-8.47	-5.71
diethyl_sulfide	-0.31	0.21	-1.46
diethylamine	-2.43	-2.72	-4.07
diiodomethane	-2.08	-0.74	-2.49
diisopropyl_ether	0.14	-0.11	-0.53
diisopropylamine	-3.05	-1.97	-3.22
dimethoxymethane	-3.46	-2.69	-2.93
dimethyl_disulfide	0.73	1.48	-1.83
dimethyl_ether	-1.78	-0.85	-1.91
dimethyl_sulfate	-8.53	-7.85	-5.1
dimethyl_sulfide	-0.55	0.26	-1.61
dimethyl_sulfone	-12.44	-10.36	-10.08
dimethyl_sulfoxide	-9.35	-8.32	-8.71
dimethylamine	-4.4	-3.11	-4.29
E_12_dichloroethene	0.14	1.2	-0.78
E_but_2_enal	-4.78	-3.66	-4.22
E_hept_2_ene	3.34	2.83	1.68
E_hex_2_enal	-3.34	-3.22	-3.68
E_oct_2_enal	-3.69	-2.47	-3.43
ethanamide	-12.79	-8.62	-9.71
ethane	2.08	2.58	1.83
ethanethiol	-1.35	-0.4	-1.14
ethanol	-4.43	-3.45	-5
ethene	1.75	2.36	1.28
ethyl_acetate	-4.13	-3.23	-2.94
ethyl_benzoate	-6.58	-4.68	-3.64
ethyl_butanoate	-3.42	-3.04	-2.49
ethyl_formate	-4.01	-2.99	-2.56
ethyl_hexanoate	-4.92	-2.8	-2.23
ethyl_pentanoate	-3.37	-3.09	-2.49
ethyl_phenyl_ether	-3.37	-2.01	-2.22
ethyl_propanoate	-4.83	-3.39	-2.68
ethylamine	-3	-3.14	-4.5
ethylbenzene	-1.21	-0.59	-0.79
fluorene	-3.45	-4.29	-3.35
fluorobenzene	-1.12	-0.07	-0.8
fluoromethane	0.11	0.79	-0.22

formaldehyde	-4.25	-3.22	-2.75
halothane	0.04	0.7	-0.11
hept_1_ene	3.44	2.97	1.66
hept_1_yne	0.94	2.28	0.6
heptan_1_ol	-2.2	-2.72	-4.21
heptan_2_one	-3.07	-2.8	-3.04
heptan_4_one	-3.74	-2.68	-2.92
heptanal	-2.88	-2.63	-2.67
hex_1_ene	3.14	2.59	1.58
hex_1_yne	1.77	2.1	0.29
hexa_15_diene	2.07	2.44	1.01
hexafluoropropene	1.94	2.29	-3.76
hexan_1_ol	-2.16	-3.03	-4.4
hexan_2_one	-3.18	-2.77	-3.28
hexan_3_ol	-2.67	-2.63	-4.06
hexanal	-3.3	-2.96	-2.81
hexanoic_acid	-7.33	-5.08	-6.21
hydrazine	-12.23	-7.1	-9.3
hydrogen_sulfide	-2.29	-1.17	-0.7
imidazole	-9.2	-7.85	-9.63
indane	-1.71	-1.7	-1.46
iodobenzene	-1.38	-0.34	-1.74
iodoethane	-0.79	-0.1	-0.74
iodomethane	-0.92	0.03	-0.89
isoamyl_acetate	-4.08	-3.05	-2.21
isoamyl_formate	-4.23	-3.39	-2.13
isobutyl_acetate	-3.12	-2.8	-2.36
isobutyl_formate	-3.87	-3.21	-2.22
isobutyl_isobutanoate	-2.41	-2.55	-1.69
isobutylbenzene	0.55	0	0.16
isobutyraldehyde	-3.83	-2.93	-2.86
isoflurane	-3.3	-1.38	0.1
isopropyl_acetate	-3.78	-2.92	-2.64
isopropyl_formate	-3.93	-2.64	-2.02
isopropylbenzene	-0.64	-0.31	-0.3
m_bis_trifluoromethyl?_benzene	-2.48	-1.58	1.07
m_cresol	-6.05	-5.28	-5.49
m_xylene	-1.28	-0.17	-0.83
methane	2.07	2.54	1.99

methanesulfonyl_chloride	-8.03	-6.33	-4.87
methanethiol	-1.1	-0.26	-1.24
methanol	-3.98	-3.48	-5.1
methoxyflurane	-1.24	-0.71	-1.12
methyl_acetate	-5.17	-3.73	-3.13
methyl_benzoate	-7	-5.06	-3.92
methyl_butanoate	-3.97	-3.32	-2.83
methyl_chloroacetate	-5.48	-3.92	-4
methyl_cyanoacetate	-8.48	-5.51	-6.72
methyl_cyclohexanecarboxylate	-3.93	-4.29	-3.3
methyl_cyclohexyl_ketone	-3.68	-3.9	-3.9
methyl_cyclopropanecarboxylate	-4.09	-4.49	-4.1
methyl_cyclopropyl_ketone	-3.46	-3.74	-4.61
methyl_ethyl_ether	-1.47	-0.82	-2.1
methyl_ethyl_sulfide	-0.37	0.34	-1.5
methyl_formate	-4.79	-3.17	-2.78
methyl_hexanoate	-4.91	-3.02	-2.49
methyl_isopropyl_ether	-1.02	-0.75	-2.01
methyl_methanesulfonate	-9.81	-8.29	-4.87
methyl_octanoate	-4.54	-2.94	-2.04
methyl_p_methoxybenzoate	-10.42	-6.32	-5.33
methyl_p_nitrobenzoate	-8.65	-6.39	-6.88
methyl_pentanoate	-3.23	-3.51	-2.56
methyl_propanoate	-5.12	-3.83	-2.93
methyl_propyl_ether	-0.85	-0.41	-1.66
methyl_t_butyl_ether	-1.52	-0.6	-2.21
methyl_tert_butyl_ether	-1.53	-0.68	-2.21
methyl_trifluoroacetate	-2.22	-1.39	-1.1
methyl_trimethylacetate	-3.02	-2.99	-2.4
methylamine	-3.58	-3.44	-4.55
methylcyclohexane	1.85	1.82	1.7
methylcyclopentane	1.74	2.12	1.59
morpholine	-7.02	-6.28	-7.17
N_acetylpyrrolidine	-9.67	-7.97	-9.8
n_butane	2.22	2.54	2.07
n_butanethiol	-0.69	-0.12	-0.99
n_butyl_acetate	-3.89	-3.17	-2.64
n_butylacetamide	-10.19	-8.14	-9.31
n_butylamine	-2.87	-2.82	-4.24

n_butylbenzene	0.4	-0.3	-0.4
n_decane	4.41	3.43	3.16
n_heptane	3.87	3.2	2.67
n_heptylamine	-2.34	-2.72	-3.79
n_hexane	3.54	3.05	2.48
n_hexyl_acetate	-4.49	-2.97	-2.26
n_hexylamine	-1.81	-2.54	-3.95
n_hexylbenzene	0.3	0.24	-0.04
N_methylacetamide	-9.36	-8.39	-10
N_methylaniline	-6.42	-5.74	-4.69
N_methylmorpholine	-6.17	-5.87	-6.32
N_methylpiperazine	-8.96	-8.3	-7.77
N_methylpiperidine	-2.31	-3.21	-3.88
n_nonane	3.95	3.32	3.13
n_octane	3.8	3.13	2.88
n_octylamine	-2.24	-2.38	-3.65
n_pentane	2.46	2.67	2.32
n_pentyl_acetate	-4.61	-2.83	-2.51
n_pentyl_propanoate	-3.79	-3.36	-2.11
n_pentylamine	-2.78	-2.99	-4.09
n_pentylbenzene	0.37	0.05	-0.23
n_pentylcyclopentane	3.35	2.39	2.55
n_propanethiol	-0.96	-0.31	-1.06
n_propyl_acetate	-3.82	-3.33	-2.79
n_propyl_butyrate	-4.11	-2.88	-2.28
n_propyl_formate	-4.43	-3.5	-2.48
n_propyl_propanoate	-3.55	-2.95	-2.44
n_propylamine	-4.07	-3.05	-4.39
n_propylbenzene	-0.44	0.01	-0.53
n_propylcyclopentane	3.05	2.15	2.13
naphthalene	-3.6	-3.34	-2.4
nitrobenzene	-4.3	-3.4	-4.12
nitroethane	-2.59	-1.73	-3.71
nitromethane	-2.76	-2.03	-4.02
NN_dimethyl_p_methoxybenzamide	-14.84	-9.29	-11.01
NN_dimethyl_p_methylbenzamide	-11.01	-7.51	-9.76
NN_dimethyl_p_nitrobenzamide	-12.34	-9.92	-11.95
NN_dimethylaniline	-5.75	-4.7	-3.45
NN_dimethylbenzamide	-10.16	-7.98	-9.29

NN_dimethylformamide	-7.45	-6.86	-7.81
non_1_ene	3.74	2.91	2.06
nonan_1_ol	-1.84	-2.54	-3.88
nonan_2_one	-3.72	-2.51	-2.49
nonan_5_one	-2.96	-2.34	-2.64
nonanal	-2.57	-2.51	-2.07
o_cresol	-6.37	-5.09	-5.87
o_toluidine	-6.07	-5.37	-5.53
o_xylene	-1.51	-0.52	-0.9
oct_1_ene	3.43	2.77	1.92
oct_1_yne	1.17	2.46	0.71
octan_1_ol	-3.83	-2.65	-4.09
octan_2_one	-3.5	-2.3	-2.88
octanal	-4.3	-2.57	-2.29
p_cresol	-6.58	-5.36	-6.13
p_dibromobenzene	-0.98	-0.01	-2.3
p_toluidine	-6.3	-5.56	-5.57
p_xylene	-1.4	-0.67	-0.8
pent_1_ene	2.15	2.44	1.68
pent_1_yne	0.75	1.93	0.01
penta_14_diene	1.66	2.18	0.93
pentachloroethane	0.04	0.31	-1.39
pentan_1_ol	-2.8	-3.14	-4.57
pentan_2_ol	-3.13	-2.87	-4.39
pentan_2_one	-3.51	-3.29	-3.52
pentan_3_ol	-2.66	-2.95	-4.35
pentan_3_one	-3.5	-2.94	-3.41
pentanal	-3.37	-2.96	-3.03
pentanenitrile	-2.73	-1.24	-3.52
pentanoic_acid	-6.74	-5.37	-6.16
phenanthrene	-4.83	-5.15	-3.88
phenol	-6.98	-5.67	-6.61
phenyl_formate	-6.8	-4.83	-3.82
phenyl_methyl_sulfide	-2.38	-1.21	-2.73
phenyl_trifluoroethyl_ether	-5.12	-2.93	-1.29
piperazine	-8.01	-8.34	-7.4
piperidine	-3.42	-3.46	-5.11
prop_2_en_1_ol	-3.41	-3.23	-5.03
propan_1_ol	-3.53	-3.12	-4.85

propan_2_ol	-3.17	-3.28	-4.74
propane	2.04	2.56	1.96
propanenitrile	-3.25	-1.27	-3.84
propanoic_acid	-8.14	-6.41	-6.46
propanone	-4.37	-3.36	-3.8
propene	1.6	2.44	1.32
propionaldehyde	-4.07	-3.08	-3.43
propyne	-0.75	1.79	-0.48
pyrene	-7.43	-6.62	-4.52
pyridine	-4.77	-3.45	-4.69
pyrrole	-5.62	-3.87	-4.78
pyrrolidine	-4.61	-3.91	-5.48
quinoline	-6.38	-4.87	-5.72
sec_butylbenzene	0.85	0.04	-0.45
styrene	-1.49	-1.32	-1.24
teflurane	-0.36	0.46	0.5
tert_butylbenzene	0.5	-0.42	-0.44
tetrachloroethene	1.14	1.41	0.1
tetrachloromethane	0.94	1.41	0.08
tetrafluoromethane	2.23	2.42	3.12
tetrahydrofuran	-2.71	-2.07	-3.47
tetrahydropyran	-2.05	-1.78	-3.12
thiophene	-1.18	-0.34	-1.42
thiophenol	-3.02	-1.43	-2.55
toluene	-1.65	-0.71	-0.89
trans_14_dimethylcyclohexane	2.17	2.05	2.11
triacetyl_glycerol	-12.29	-13.31	-8.84
trichloroethene	0.18	0.99	-0.44
trichloromethane	-0.54	0.32	-1.08
triethyl_phosphate	-12.81	-10.06	-7.54
triethylamine	-2.48	-1.83	-3.22
trimethoxy_methane	-4.91	-4.06	-4.42
trimethoxymethylbenzene	-4.38	-5.83	-4.04
trimethyl_phosphate	-13.13	-10.59	-8.7
trimethylamine	-3.39	-2.32	-3.2
undecan_2_one	-4.91	-2.12	-2.15
Z_12_dichloroethene	-0.87	0.31	-1.17
Z_pent_2_ene	1.88	2.55	1.31

Table S4-2: Comparison of solvation free energy calculated from different conformations (kcal/mol)

Name	DFT[a]	DFT[b]	DFT	DFT	EXP	MD
111_trifluoropropan_2_ol	-7.13	-7.13 ~ - 2.4	-	-	-3.56	-4.16
12_diacetoxyethane	-12.5	-12.5 ~ -9	-	-	-8.32	-6.34
12_diethoxyethane	-5.33	-7.47 ~ -0.86	-	-	-3.54	-3.08
2_chloro_111_trimethoxyethane	-7.67	-7.8 ~ -2.93	-	-5	-3.3	-4.59
2_ethoxyethanol	-	-11.63 ~ -	-	-6.9	-4.77	-6.69
2_methoxy_111_trimethoxyetha	-	-12.24 ~ -	-	-7.2	-5.97	-5.73
2_methoxyethanol	-8.01	-10.43 ~ -	-	-	-5.37	-6.76
2_nitrophenol	-10.8	-10.8 ~ -4.58	-	-	-5.34	-4.58
2_propoxyethanol	-8.19	-8.32 ~ -3.04	-	-	-4.21	-6.4
333_trimethoxypropionitrile	-	-10.63 ~ -	-	-7.3	-4.93	-6.4
3_hydroxybenzaldehyde	-	-9.92 ~ -	-	-	-9.24	-9.5
4_hydroxybenzaldehyde	-	-12.03 ~ -	-	-	-10	-8.83
4_methoxyacetophenone	-8.56	-8.56 ~ -7.41	-	-	-6.29	-4.4
acetic_acid	-8.17	-8.67 ~ -6.32	-	-	-5.95	-6.69
acetonitrile	-3.86	-3.9 ~ -3.72	-	-3.8	-1.67	-3.88
benzamide	-	-14.49 ~ -	-	-	-	-11
di_n_butylamine	-4.69	-4.95 ~ 0.55	-	-	-1.63	-3.24
diethyl_malonate	-	-11.08 ~ -	-	-	-6.45	-6
diethyl_succinate	-	-17.72 ~ -	-	-	-8.47	-5.71
dimethyl_sulfone	-	-14.64 ~ -	-	-	-	-
ethanamide	-	-12.91 ~-	-	-	-8.62	-9.71
hexanoic_acid	-7.33	-7.33 ~ -5.29	-	-	-5.08	-6.21
hydrazine	-	-12.23 ~ -6.4	-	-	-7.1	-9.3
methyl_cyanoacetate	-8.48	-8.79 ~ -7.06	-	-	-5.51	-6.72
methyl_p_methoxybenzoate	-	-10.42 ~ -	-	-9.4	-6.32	-5.33
n_butylacetamide	-	-11.86 ~ -	-	-	-8.14	-9.31
NN_dimethyl_p_methoxybenza	-	-14.96 ~ -	-	-	-9.29	-
NN_dimethyl_p_methylbenzami	-	-12.76 ~ -	-	-	-7.51	-9.76
NN_dimethylbenzamide	-	-10.71 ~ -	-	-	-7.98	-9.29
phenyl_trifluoroethyl_ether	-5.12	-5.57 ~ -1.53	-	-	-2.93	-1.29
propyne	-0.75	-0.77 ~ -0.64	-	-	1.79	-0.48
trimethyl_phosphate	-	-13.13 ~ -	-	-	-	-8.7
undecan_2_one	-4.91	-6.82 ~ -2.84	-	-4.9	-2.12	-2.15

a. Calculated from the initial conformation.
b. Calculated from 13 different conformations including the initial one and 12 random ones.
c. Calculated from the equilibrium conformation in vacuum.
d. Average of the 13 different conformations.

4.2 Fast prediction of hydration free energies for SAMPL4 from a classical density functional theory

ABSTRACT

We report the performance of a classical density functional theory (DFT) in the competition for solvation free-energy category of the SAMPL4 blind prediction event. The influence of molecular force fields on the DFT predictions is analyzed by carrying out calculations with both Amber and OPLS force fields. In comparison with the experimental data, the DFT method shows the numerical accuracy similar to that with explicit-solvent MD simulations. However, it reduces the computational cost by almost 3 orders of magnitude. The SAMPL4 blind test indicates that the DFT method is promising for large-scale screening applications such as small molecular drug design.

4.2.1. Introduction

Hydration free energy plays a key role in solution chemistry and its theoretical prediction often represents a bottleneck in understanding important chemical and biological processes[1,2]. In particular, solvent-solute interactions in an aqueous environment entail complex microscopic details that make a reliable prediction of solvation properties a serious computational challenge.

The conventional methods for solvation free energy calculations are typically based on either implicit or explicit solvent models. Implicit-solvent methods are usually constructed from knowledge-based macroscopic considerations, augmented with the Poisson-Boltzmann equation[27] or semi-empirical models such as AMSOL[28]. Conversely, most explicit-solvent models employ a semi-empirical force field to represent the solvent-solvent and solvent-solute interactions[3]. The implicit models have been popular in the past few years, since they are faster than explicit-solvent models by orders of magnitude. However, implicit models usually ignore the hydrogen-bonding network among water molecules and other solvent structures[29] that are essential for understanding hydration properties from a microscopic perspective. Besides, the accuracy of explicit models is normally higher than that of implicit models. From a practical point of view, the continuous and solvent-explicit methods are complementary and their choices for particular applications often reflect a compromise of the computational cost and the accuracy in microscopic details.

Classical density functional theory (DFT) provides a compromise between conventional semi-empirical methods and molecular simulations. . From a mathematical

perspective, classical DFT closely resembles electronic DFT except that in the former case, the density functional applies to the structure of atoms or coarse-grained elements of a polymeric molecule, whereas the latter applies to electrons. Motivated by the need for high-performance computational methods, we recently developed a three-dimensional classical density functional theory (DFT) method based on the site-site intermolecular interactions[14]. Different from molecular simulations, DFT allows us to do one time calculation to obtain both thermodynamic properties, including the hydration free energy, and the equilibrium configuration without explicit enumeration of microstates. In other words, DFT calculation is based on minimization of the grand potential instead of sampling millions of microstates by brute force calculations. A key advantage of DFT is that it generates thermodynamic properties such as entropy and free energy without evoking multiple state thermodynamic integral or potential perturbation calculations. The accuracy of DFT has been well examined in the past decades for both simple and complex fluids[30]. In many cases, the theoretical performance is comparable to molecular simulations but with much less computational cost. Another theoretical method we should mention here is the integral-equation theories within the framework of the Reference Interaction Site Model (RISM)[31,32]. Similar to DFT, the RISM calculation is based on the numerical solution of a set of integral-differential equations in order to obtain thermodynamic properties. Different from DFT, RISM solve the direct correlation function self-consistently but the closure used in RISM is not rigorous. Although RISM calculations are also much faster than conventional molecular simulations, the more complicated multidimensional integral still leads to higher computation burden than the DFT method.

In this paper, we report a test of the DFT method through participation in the solvation-free-energy category of the SAMPL4 blind prediction challenge[33]. The sample includes hydration free energies for 47small organic molecules prepared by Guthrie[34]. It provides a good benchmark for validating the quantitative performance of our method. As a hallmark of SAMPL4 competition, we did not have or use any *a priori* information about the experimental data of those compounds in all the calculations. Because the experimental data are not included in the training sets of those two force fields we apply, the theoretical predictions reflect a truly blind test of the molecular model and the DFT calculations.

4.2.2. Model and Theory

We consider a rigid solute model composed of *M* atoms dissolved in an aqueous solution at 298K and 1 atm. The solute-solvent interactions are represented by the Lennard-Jones 12-6 plus Columbic potential:

$$u_{ij}(\mathbf{r}) = 4e_{ij}\left[\left(\frac{s_{ij}}{r}\right)^{12} - \left(\frac{s_{ij}}{r}\right)^{6}\right] + \frac{Z_i Z_j e^2}{4pe_0 r} \tag{88}$$

where $\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}$, $\sigma_{ij} = (\sigma_{ii} + \sigma_{jj})/2$, ε_{ii} and σ_{ii} are the energy well depth and the diameter of the *i*th atom, Z_i is the (fractional) valence of atom *i*, *e* is the elementary charge, and ε_0 is the vacuum permittivity.

Near a rigid solute molecule with its center of mass fixed at the origin, the external potential on each solvent site *i* can be calculated from:

$$V_i^{\text{ext}}(\mathbf{r}) = \sum_{j=1}^{M} u_{ij}(\mathbf{r} - \mathbf{r}_j) \tag{89}$$

where $\mathbf{r}_j$ is the coordinate of the jth atom of the solute molecule. The grand potential $\Omega$of the solvent molecules can be expressed as:

$$\Omega = F^{\mathrm{id}} + F^{\mathrm{ex}} + F^{\mathrm{ext}} - \mu N \tag{90}$$

where F^{id}, F^{ex} and F^{ext} represent the ideal (kinetic energy contribution), excess (solvent-solvent interaction contribution) and external (solute-solvent interaction contribution) free energy functional, respectively; μ is the chemical potential, and N denotes the number of solvent molecules.

The ideal-gas term F^{id} can be formally expressed as a functional of the molecule density profile $\rho(\mathbf{R})$:

$$F^{\mathrm{id}} = k_{\mathrm{B}}T \int \rho(\mathbf{R})[\ln \rho(\mathbf{R}) - 1 - V_{\mathrm{intra}}(\mathbf{R})]\mathrm{d}\mathbf{R} \tag{91}$$

where k_{B} is the Boltzmann constant, T is the temperature, $V_{\mathrm{intra}}(\mathbf{R})$ is the intramolecular potential of the solvent molecule, $\mathbf{R} = \{\mathbf{r}_i\}$ is a multidimensional coordinate with $\mathbf{r}_i$ being the three-dimensional coordinate of the ith atom in the solvent molecule. In this work, three-site rigid water model is used. In that case, $\mathbf{R}$ can be expressed as a six-dimensional vector, and $V_{\mathrm{intra}}(\mathbf{R})$ is represented by 3 Dirac delta functions due to the fixed molecular conformation.

By minimizing the grand potential

$$\frac{\delta \Omega}{\delta \rho(\mathbf{R})} = 0 \tag{92}$$

we have the Euler-Lagrange equation

$$r(\mathbf{R}) = \exp\left[bm - \frac{db(F^{\mathrm{ex}} + F^{\mathrm{ext}})}{dr(\mathbf{R})} - bV_{\mathrm{intra}}(\mathbf{R}) \right] \tag{93}$$

where $\beta = 1/(k_B T)$. Given explicit expressions for F^{ex} and F^{ext}, equation (6) can be solved self-consistently. However, direct calculation of the configurational density is computationally intensive because **R** is a multidimensional variable. The computational cost can be drastically reduced by using the site density profiles:

$$\rho_i(\mathbf{r}) = \int \delta(\mathbf{r} - \mathbf{r}_i)\rho(\mathbf{R})\mathrm{d}\mathbf{R} \tag{94}$$

where $\delta(\mathbf{r}\text{-}\mathbf{r}_i)$ is the Dirac function, and *i* represent the *i*th atom of the solvent molecule. By inserting equation 94 into 93, we have:

$$\rho_i(\mathbf{r}) = \int \delta(\mathbf{r} - \mathbf{r}_i)\exp\left[\beta\mu - \sum_j \frac{\delta\beta(F^{ex} + F^{ext})}{\delta\rho_j(\mathbf{r}')}\delta(\mathbf{r}' - \mathbf{r}_j) - \beta V_{intra}(\mathbf{R})\right]\mathrm{d}\mathbf{R} \tag{95}$$

For rigid 3-site water models (SPC, SPC/E and TIP3P), Eq. 95 can be simplified to:

$$\begin{cases} \rho_O(\mathbf{r}_O) = \exp[\lambda_O(\mathbf{r}_O)]\left\langle \exp\{\lambda_H[\mathbf{r}_{H1}(\mathbf{r}_O, \omega)] + \lambda_H[\mathbf{r}_{H2}(\mathbf{r}_O, \omega)]\}\right\rangle_\omega \\ \rho_{H1}(\mathbf{r}_{H1}) = \exp[\lambda_H(\mathbf{r}_{H1})]\left\langle \exp\{\lambda_O[\mathbf{r}_O(\mathbf{r}_{H1}, \omega)] + \lambda_H[\mathbf{r}_{H2}(\mathbf{r}_{H1}, \omega)]\}\right\rangle_\omega \\ \rho_{H2}(\mathbf{r}) = \rho_{H1}(\mathbf{r}) \end{cases} \tag{96}$$

where $<\cdots>_\omega$ represents an average over the angular vector $w = (q, f, f')$. As shown in Figure 4-4, (θ, ϕ) are the Euler angles of the dipole direction, and ϕ' is the rotation angle of two H atoms around the dipole direction; $\mathbf{r}_O$, $\mathbf{r}_{H1}$, $\mathbf{r}_{H2}$ are the corresponding atomic coordinates of water molecules.

In Eq.96, the effective one-body potential is represented by:

$$\lambda_i(\mathbf{r}) = \beta\mu_i - \frac{\delta\beta(F^{ex} + F^{ext})}{\delta\rho_i(\mathbf{r})} \tag{97}$$

where μ_i is the nominal chemical potential of site *i*, which satisfies $\sum_i \mu_i = \mu$. For a given solute with a fixed position and orientation, the external potential F^{ext} can be calculated exactly:

$$F^{\text{ext}} = \sum_i \int V_i^{\text{ext}}(\mathbf{r})\rho_i(\mathbf{r})\mathrm{d}\mathbf{r} \tag{98}$$

The excess free energy F^{ex}, however, in general unknown. In this work, we approximate F^{ex} with the ansatz of the universality functional[35]

$$\begin{aligned} F^{\text{ex}} &= F_{\text{bulk}}^{\text{ex}} + \sum_i \mu_i^{\text{ex}} \int \Delta\rho_i(\mathbf{r})\mathrm{d}\mathbf{r} \\ &\quad - \frac{k_{\text{B}}T}{2} \sum_{i,j} \iint c_{ij}^{(2)}(|\mathbf{r}-\mathbf{r}'|)\Delta\rho_i(\mathbf{r})\Delta\rho_j(\mathbf{r}')\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' + F^{\text{B}} \end{aligned} \tag{99}$$

where $F_{\text{bulk}}^{\text{ex}}$ is the excess free energy of the bulk system, $\Delta\rho_i(\mathbf{r}) = \rho_i(\mathbf{r}) - \rho_i^{\text{bulk}}$ and ρ_i^{bulk} is the bulk density of atom *i*; $\mu_i^{\text{ex}} = \mu_i - \mu_i^{\text{id}}$ is the nominal excess chemical potential, μ_i^{id} is the nominal ideal chemical potential which satisfies $\sum_i \mu_i^{\text{id}} = \mu^{\text{id}} = k_{\text{B}}T \ln \rho_O^{\text{bulk}}$; $c_{ij}^{(2)}(r)$ is the second order direct correlation function between site *i* and *j* of the bulk system, which is obtained from MD simulation;[36] and F^{B} is the bridge functional which accounts for higher-order contributions in the functional Taylor expansion.

Following Rosenfeld,[35] we approximate the bridge functional by a reference hard-sphere (HS) system:

$$\begin{aligned} F^{\text{B}} &= F_{\text{HS}}^{\text{B}}[\rho_{\text{O}}(\mathbf{r})] \\ &= F_{\text{HS}}^{\text{ex}}[\rho_{\text{O}}(\mathbf{r})] - F_{\text{HS}}^{\text{ex}}[\rho_{\text{O}}^{\text{bulk}}] - \mu_{\text{HS-b}}^{\text{ex}} \int \Delta\rho_{\text{O}}(\mathbf{r})\mathrm{d}\mathbf{r} \\ &\quad + \frac{k_{\text{B}}T}{2} \iint c_{\text{HS}}^{(2)}(|\mathbf{r}-\mathbf{r}'|)\Delta\rho_{\text{O}}(\mathbf{r})\Delta\rho_{\text{O}}(\mathbf{r}')\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \end{aligned} \tag{100}$$

where $F_{\mathrm{HS}}^{\mathrm{ex}}[\rho_{\mathrm{O}}(\mathbf{r})]$ is the HS excess free energy functional, $\mu_{\mathrm{HS\text{-}b}}^{\mathrm{ex}}$ and $c_{\mathrm{HS}}^{(2)}(r)$ are the excess chemical potential and direct correlation function of the bulk HS system, respectively. In this work, we use the modified fundamental measure theory to calculate $F_{\mathrm{HS}}^{\mathrm{ex}}[\rho_{\mathrm{O}}(\mathbf{r})]$, $\mu_{\mathrm{HS\text{-}b}}^{\mathrm{ex}}$ and $c_{\mathrm{HS}}^{(2)}(r)$. [14,37]

By inserting Eqs.98, 99 and 100 into 96, we have:

$$\lambda_i(\mathbf{r}) = \beta\mu_i^{\mathrm{id}} - \beta V_i^{\mathrm{ext}}(\mathbf{r}) + \sum_j \int \Delta\rho_j(\mathbf{r}')c_{ij}^{(2)}(|\mathbf{r}-\mathbf{r}'|)\mathrm{d}\mathbf{r}' + B_i(\mathbf{r}) \tag{101}$$

where $B_i(\mathbf{r})$ is the bridge function of atom *i* defined as:

$$B_i(\mathbf{r}) = \frac{\delta\beta F^{\mathrm{B}}}{\delta\rho_i(\mathbf{r})} \tag{102}$$

Eqs.101 and 96 allow us to solve the site density profiles $\rho_i(\mathbf{r})$ by numerical iterations. Subsequently, we can calculate the solvation free energy[5,14]

$$\begin{aligned} F_{\mathrm{s}} = &-\Delta N k_{\mathrm{B}}T + \frac{k_{\mathrm{B}}T}{2}\sum_{ij}\iint c_{ij}^{(2)}(|\mathbf{r}-\mathbf{r}'|)[\rho_i(\mathbf{r})\rho_j(\mathbf{r}') - \rho_i^{\mathrm{bulk}}\rho_j^{\mathrm{bulk}}]\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \\ &+ F^{\mathrm{B}} - k_{\mathrm{B}}T\sum_i \int \rho_i(\mathbf{r})B_i(\mathbf{r})\mathrm{d}\mathbf{r} \end{aligned} \tag{103}$$

All DFT calculations in this work were carried out in a (30 Å)^3 ~ (40 Å)^3 cubic box depending on the size of solute. The density profiles are discretized into 128^3 lattice grids. In comparison to solvent-explicit simulations, the DFT calculations are extremely fast. The average time cost is less than 30 minutes on a single CPU core. More detailed numerical procedures are listed in our previous publications.[5,14]

As in MD simulations, the classical DFT calculations hinge on molecular models and force field parameters. In this work, the water molecules are described by the TIP3P

model[38], while all van der Waals parameters for solute molecules are from GAFF[39] and OPLS-AA[40] force fields. We have tested two atomic charge sets, one is AM1-BCC charge[41], as the common choice for GAFF, another is ChelpG charge[42], which is often used by CHARMM[43] and OPLS force fields. Special attention is given to the effects of force field parameters and solute configurations on the DFT predictions. All the initial structures of those solute compounds are optimized using HF/B3LYP/MP2 methods and 6-31+G* basis set (QM structures).Then based on the initial molecular structures, the AM1-BCC charge sets are obtained from Antechamber-1.27[44] while ChelpG charge sets are obtained with the same QM calculations. All the quantum mechanics calculations are carried by Gaussian 09[45]. With different combinations of the charge sets and force field van der Waals parameters, we also apply molecular mechanics to optimize the solute structures (MM structures) in vacuum environment using the conjugated-gradient method available from GROMACS 4.6.2[46].

4.2.3. Results and Discussion

Figure 4-5 compares the DFT predictions with the experimental results for the hydration free energies of 47 compounds posted at the SAMPL4 website. We should mention that different from simulations, DFT solves a set of equations instead of sampling, which means there is no uncertainty (error bar) of DFT calculation. Here the two force field van der Waals parameter sets and the two charge sets are employed to describe solute-solvent interactions. All the numerical values are listed in Supporting Information. For the SAMPL4 challenge, we submitted the results based on the OPLS-AA van der Waals parameters and ChelpG charge set, while the solute structures were obtained from MM

optimization in vacuum (Figure 4-5a). The submitted results (submission ID 548) yield an average unsigned error (AUE) of 2.38 kcal/mol and root mean square (RMS) of 2.99 kcal/mol, which ranks in the middle among 49 submissions from 19 groups. After comparing to the released experimental data, we find that the theoretical results can be drastically improved by several other combinations of the charge sets and van der Waals parameters. In Figure 4-5, for the charge part, we tested OPLS-AA with two common electrostatic potential fitting (ESP) charge sets, ChelpG (Figure 4-5a) and AM1-BCC (Figure 4-5b). Besides, we used both B3LYP/6-31+G*, the most widely used QM method, and molecular mechanics to determine solute structures by energy minimization in vacuum.

To see our method performance in details, we show the RMS data of our first stage calculation by Table 4-1. It indicates that different structure optimization methods contributes to around 10% RMS deviation on the solvation free energy. One possible reason is that we use relatively cheap QM methods, considering our large-scale screening target. More expensive QM calculations, like MP2/6-311G(d,p) would increase the computational time by more than two order of magnitude. Besides, more accurate vacuum calculation may be unnecessary because we consider only a single conformation for each solute. More important factor to the DFT predictions is the selection of the charge sets, in general, AM1-BCC yields better results than ChelpG charge set, which makes the difference of RMS more than 20% in our validation. One possible reason is while ChelpG charge set is highly affected by the molecular structures, AM1-BCC charge set is made to

make less configuration dependable, importing some flexible structure effect, more proper for our explicit considering water model condition here.

In Figure 4-5c, we show the DFT results with General Amber Force Field van der Waals parameters and AM1-BCC charge set. Again, we see only negligible influence of the solute structure differences. It should be noticed that hydrophilic compounds (Fs < -10 kcal/mol) show large errors in comparison to the experimental values. One possible reason is that our DFT calculation ignores the contribution of flexible solute structures. For example, molecule 001 has six hydroxyl groups, leading to complicated hydrogen bonding network in water, which cannot be captured by a single conformation. The DFT results can be significantly improved by sampling the solute flexibility with a relatively small number of equilibrium conformations in vacuum and doing the ensemble average. The improvement is around 0 ~ 4 kcal/mol[47]. Instead of choosing those solute configurations randomly, in this work, we try to give reasonable prediction by single input solute configurations for fast prediction purpose through comparing various QM optimization solute structures and charge sets, which will be discussed in Table 4-2. Another possible reason lies in the force field parameters. As suggested by Mobley et al.[33], the AUE value could be reduced to 1 kcal/mol (submission ID 005) by introducing new hydroxyl group parameters, which give the best MD simulation results in the challenge.

Table 4-1 shows that overall the theoretical predictions agree well with the experimental data, in particular for those molecules with a positive or small hydration free energy (>-5 kcal/mol). With AM1-BCC charge set, the RMS is less than 1.50 kcal/mol, even in the top tier of hydration free energy challenge players. But when solutes turn more

hydrophilic, the situation becomes worse, as we discussed above. Another possible reason for the increased discrepancy may be attributed to the approximations used in the DFT calculations. In comparison with MD simulation, the error of DFT calculations is mainly introduced in the approximations for the bridge functional. In formulation of the free-energy density functional, we utilize a hard-sphere model to represent the bridge functional for water wherein the hard-sphere diameter of the reference system was calibrated with the solvation free energy for methane[14]. Because the local structure of water molecules is sensitive to the solute hydrophobicity, the parameter obtained from a hydrophobic solute may become less reliable for hydrophilic solutes. Since the effective hard-sphere diameter for a hydrophilic solute is smaller than that for a hydrophobic solute due to the enhanced solute-solvent attraction.

We have also explored the way of improving our DFT calculations by optimizing the input solute structures with IEFPCM[48] implicit solvent model under B3LYP/HF/MP2 methods with 6-31+G* basis set. The final results are shown in Table 4-2. Due to the consideration of solvent effect, the gap between AM1-BCC and ChelpG charge sets becomes more apparent, even over 200% in RMS. AM1-BCC remains the better choice. With this charge set, solute structures obtained from various QM methods do not show significant differences, similar to the above vacuum configuration ones. Again, we did not use higher-level QM methods, like M06X or MP2 with Aug-cc-PVTZ basis set[49,50], because of the high computation cost and rigid solute models used in those DFT calculations. However, as discussed in the previous work[47], there is no significant correlation between the DFT performance and the dipole moments of the solute molecules.

From the data of Table 4-2, we find the best combination choice for our current purpose, (OPLS-AA van der Waals parameters/AM1-BCC charge set/HF/IEFPCM optimization method) to reduce the AUE and RMS to 1.54 kcal/mol and 2.08 kcal/mol separately, compared to experimental data, which is very similar to the MD simulation by Yang's group (submission ID 575, AUE=1.59 kcal/mol, RMS= 2.10 kcal/mol) and by Gilson's group (submission ID 542, AUE= 1.30 kcal/mol, RMS=1.89 kcal/mol).

4.2.4. Conclusions

We have demonstrated that the density functional theory (DFT) provides a powerful alternative to conventional simulation methods for fast prediction of small molecular hydration free energy. We find that with proper input parameters combination (OPLS-AA/AM1-BCC/HF/IEFPCM), we could reduce the AUE and RMS to 1.54 kcal/mol and 2.08 kcal/mol separately, compared to our original submission result 2.38 kcal/mol and 2.99 kcal/mol, on a par with classical MD simulations. The theoretical results are most accurate for solutes with slightly positive hydration free energies but become less precise for large hydrophilic solutes.

In comparison with MD simulations, one key advantage of DFT is its numerical efficiency. For the systems considered in this work, each DFT calculation can be readily processed with a single CPU core (Nehalem Xeon 2.67GHz), within one hour. The computational cost could be further reduced by optimization of the DFT programs. In SAMPL4 challenge, there are also other fast and accurate hydration free energy calculation methods, like SEA[51] considering explicit solvent effect and QM with implicit solvent model[52], our method provides another way by modern liquid theory. After the input of

direct correlation function from explicit bulk water simulation, it is ready to give water structure information and other important thermodynamic beside solvation free energy in one time calculation.

Figure 4-4. The molecular frame used in the DFT calculations for the solvation of a tagged molecule (solute) placed at the origin. The y-axis is perpendicular to the paper inward; the location of a neighboring water molecule (solvent) is described by the oxygen atom position $\mathbf{r} = (x, y, z)$ and three Euler angles (q, f, f') .

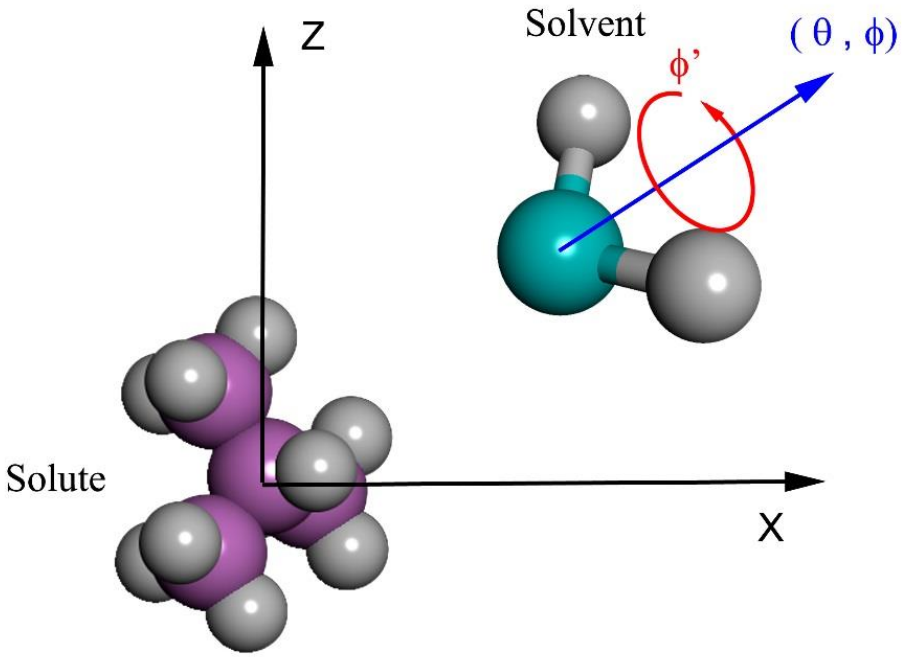
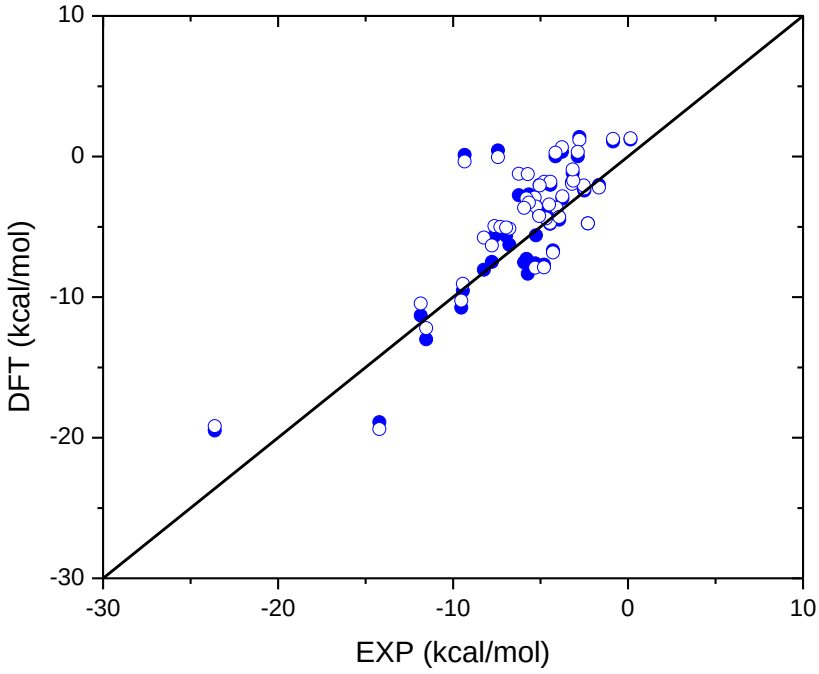
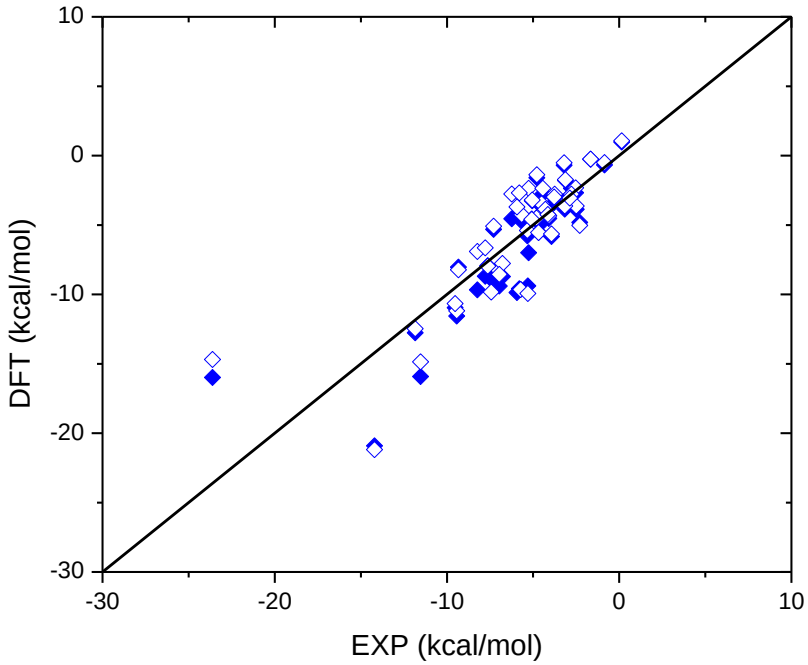


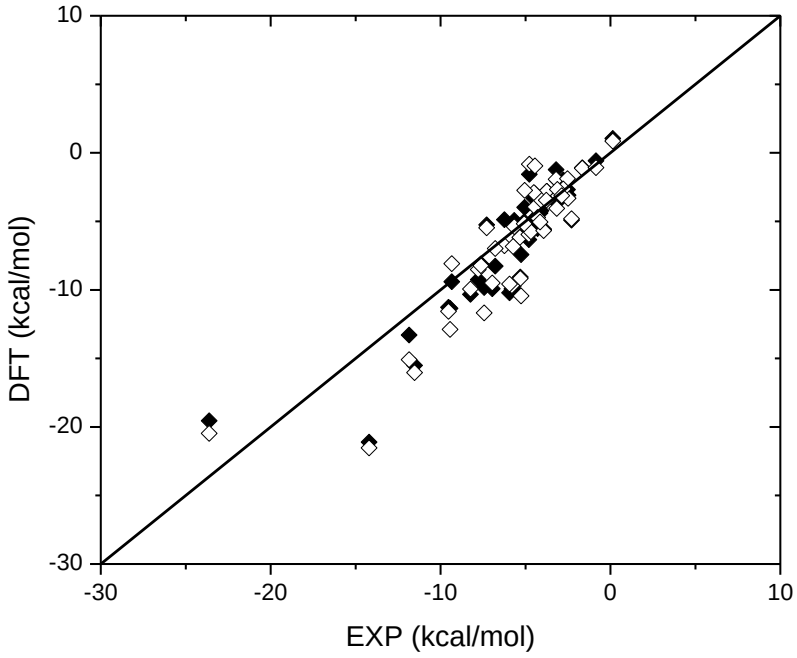
Figure 4-5 Comparison of hydration free energies from DFT calculations with experimental data, a) OPLS-AA force field van der Waals parameters with AM1-BCC charge set; b) OPLS-AA force field van der Waals parameters with ChelpG charge set; c) GAFF van der Waals parameters with AM1-BCC charge set. Filled symbols are calculated on the basis of solute structures obtained by B3LYP/6-31+G* optimization, while blank symbols are from solute structures obtained by molecular mechanic optimization, all in vacuum.



a)


10
0
-10
-20
-30
DFT (kcal/mol)
-30
-20
-10
0
10
EXP (kcal/mol)

b)


10
0
-10
-20
-30
DFT (kcal/mol)
-30
-20
-10
0
10
EXP (kcal/mol)

c)

Table 4-1 The Root Mean Square (RMS) of the DFT calculations in comparison to the experimental data. Here QM means solute structures obtained by B3LYP/6-31+G* optimization, while MM means by molecular mechanic optimization, all in vacuum.

RMS (kcal/mol)	OPLS-AA				GAFF	
	QM		MM		QM	MM
	AM1-BCC	ChelpG	AM1-BCC	ChelpG	AM1-BCC	ChelpG
ALL	2.36	2.79	2.46	2.99	2.20	2.40
Fs >-5	1.30	2.17	1.38	2.29	1.27	1.47
-10 < Fs < -5	2.23	3.28	2.24	3.55	2.32	2.52
Fs < -10	5.55	3.22	5.91	3.50	4.53	4.85

Table 4-2 Comparison of DFT prediction results and experimental data, with various QM optimization solute structures using IEFPCM implicit solvent model, combining with different force field van der Waals parameters and charge sets.

RMS(kcal/mol)	B3LYP			
	GAFF		OPLS-AA	
	AM1-BCC	ChelpG	AM1-BCC	ChelpG
ALL	2.36	5.51	2.24	5.37
Fs >-5	1.15	3.81	1.18	3.48
-10 < Fs < -5	2.59	4.87	2.47	4.90
Fs < -10	4.80	12.14	4.38	11.92

RMS(kcal/mol)	HF			
	GAFF		OPLS-AA	
	AM1-BCC	ChelpG	AM1-BCC	ChelpG
ALL	2.15	3.37	2.08	3.09
Fs >-5	1.27	2.51	1.30	2.17
-10 < Fs < -5	2.32	3.15	2.23	3.28
Fs < -10	4.26	6.91	4.03	5.61

RMS(kcal/mol)	MP2			
	GAFF		OPLS-AA	
	AM1-BCC	ChelpG	AM1-BCC	ChelpG
ALL	2.38	7.93	2.24	7.73
Fs >-5	1.10	5.63	1.11	5.33
-10 < Fs < -5	2.34	6.44	2.29	6.45
Fs < -10	5.28	17.38	4.68	16.90

Support Information

Table S4-3: Comparing solvation free energy of SAMPL4 47 compounds with OPLS-aa force field and optimization solute structures in vacuum environment

ID	EXP(kcal/mol)	Err	$OPLS_{(QM\_bcc)}$	$OPLS_{(MM\_bcc)}$	$OPLS_{(QM\_ChelpG)}$	$OPLS_{(MM\_ChelpG)}$
1	-23.62	0.32	-15.99	-14.69	-19.48	-19.18
2	-2.49	0.85	-3.87	-3.63	-2.42	-2.28
3	-4.78	0.25	-1.59	-1.37	-1.88	-1.80
4	-4.45	0.24	-4.77	-4.04	-4.79	-4.73
5	-5.33	0.1	-5.78	-5.39	-3.39	-2.93
6	-5.26	0.18	-6.99	-2.36	-5.61	-3.56
9	-8.24	0.76	-9.67	-6.91	-8.05	-5.77
10	-6.24	0.38	-4.55	-2.76	-2.76	-1.23
11	-7.78	0.77	-8.69	-6.64	-7.49	-6.33
12	-3.75	0.21	-3.05	-2.78	-3.00	-2.83
13	-4.44	0.43	-2.76	-2.36	-2.02	-1.80
14	-4.09	0.17	-4.52	-4.33	-3.84	-3.65
15	-4.51	0.1	-3.65	-3.47	-3.68	-3.41
16	-3.2	0.27	-0.70	-0.51	-1.82	-1.94
17	-2.53	0.25	-2.67	-2.32	-2.28	-2.06
19	-3.78	0.1	-3.38	-2.98	0.33	0.66
20	-2.78	0.1	-2.36	-2.77	1.38	1.19
21	-7.63	0.12	-7.93	-7.96	-5.66	-4.95
22	-6.78	0.1	-8.72	-7.78	-6.28	-5.13
23	-9.34	0.62	-8.03	-8.23	0.12	-0.35
24	-7.43	0.6	-8.90	-9.83	0.43	-0.04
25	-5.73	0.15	-9.64	-9.64	-8.34	-1.25
26	-5.31	0.1	-9.38	-9.93	-7.60	-7.91
27	-4.8	0.39	-4.67	-4.49	-7.69	-7.88
28	-4.29	0.39	-4.02	-3.87	-6.67	-6.82
29	-1.66	0.1	-0.26	-0.24	-2.03	-2.20
30	-2.29	0.12	-4.82	-5.03	-4.76	-4.76
32	-7.29	0.1	-5.31	-5.09	-5.11	-5.00
33	-6.96	0.1	-9.39	-8.55	-5.64	-5.05
34	-5.8	0.1	-9.59	-2.69	-7.28	-2.98
35	-4.68	0.1	-5.52	-5.54	-4.40	-4.40
36	-5.66	0.1	-4.73	-4.24	-2.70	-3.28
37	-5.94	0.1	-9.87	-3.70	-7.54	-3.65

38	-3.93	0.1	-5.83	-5.65	-4.49	-4.30
39	-0.85	0.1	-0.67	-0.50	1.07	1.25
41	-5.05	0.1	-3.25	-3.20	-2.02	-2.06
42	-3.13	0.1	-1.79	-1.73	-1.80	-1.72
43	0.14	0.1	0.98	1.07	1.20	1.30
44	-5.08	0.1	-4.56	-4.60	-4.16	-4.23
45	-11.53	0.29	-15.92	-14.86	-13.01	-12.21
46	-9.44	0.74	-11.56	-11.18	-9.55	-9.05
47	-14.21	1.1	-20.91	-21.18	-18.90	-19.39
48	-11.85	0.35	-12.76	-12.46	-11.30	-10.46
49	-3.16	0.1	-3.85	-3.63	-1.17	-0.92
50	-4.14	0.1	-4.44	-4.24	0.02	0.27
51	-9.53	0.28	-10.96	-10.66	-10.75	-10.25
52	-2.87	0.69	-2.89	-3.06	0.01	0.33

Table S4-4: Comparing solvation free energy of SAMPL4 47 compounds with GAFF force field and optimization solute structures in vacuum environment

ID	EXP(kcal/mol)	Err	AMBER(QM)	AMBER(MM)
1	-23.62	0.32	-19.54	-20.47
2	-2.49	0.85	-3.09	-3.32
3	-4.78	0.25	-1.57	-0.83
4	-4.45	0.24	-4.78	-0.95
5	-5.33	0.1	-5.38	-6.17
6	-5.26	0.18	-7.42	-10.43
9	-8.24	0.76	-10.33	-9.94
10	-6.24	0.38	-4.87	-6.77
11	-7.78	0.77	-9.25	-8.51
12	-3.75	0.21	-2.81	-2.80
13	-4.44	0.43	-2.99	-3.95
14	-4.09	0.17	-4.22	-3.44
15	-4.51	0.1	-3.32	-2.90
16	-3.2	0.27	-1.22	-1.93
17	-2.53	0.25	-2.68	-1.89
19	-3.78	0.1	-3.43	-3.45
20	-2.78	0.1	-1.98	-2.63
21	-7.63	0.12	-8.11	-8.24
22	-6.78	0.1	-8.27	-6.98
23	-9.34	0.62	-9.41	-8.08
24	-7.43	0.6	-9.82	-11.68
25	-5.73	0.15	-9.75	-6.84
26	-5.31	0.1	-9.06	-9.16
27	-4.8	0.39	-6.33	-5.95
28	-4.29	0.39	-5.53	-4.72
29	-1.66	0.1	-1.09	-1.13
30	-2.29	0.12	-4.91	-4.81
32	-7.29	0.1	-5.25	-5.47
33	-6.96	0.1	-9.91	-9.49
34	-5.8	0.1	-9.92	-9.69
35	-4.68	0.1	-5.56	-5.75
36	-5.66	0.1	-4.93	-5.39
37	-5.94	0.1	-10.21	-9.57
38	-3.93	0.1	-5.57	-5.68

39	-0.85	0.1	-0.58	-1.08
41	-5.05	0.1	-3.98	-2.74
42	-3.13	0.1	-2.01	-2.67
43	0.14	0.1	1.05	0.85
44	-5.08	0.1	-5.02	-5.17
45	-11.53	0.29	-15.51	-16.04
46	-9.44	0.74	-11.35	-12.89
47	-14.21	1.1	-21.10	-21.52
48	-11.85	0.35	-13.29	-15.09
49	-3.16	0.1	-3.64	-4.05
50	-4.14	0.1	-4.39	-5.05
51	-9.53	0.28	-11.27	-11.56
52	-2.87	0.69	-2.67	-3.12

Table S4-5: Comparing solvation free energy of SAMPL4 compounds with various force field combinations and QM optimization solute structures under B3LYP/6-31+G*/IEFPCM

ID	EXP(kcal/mol)	Error	GAFF_BCC	GAFF_ChelpG	OPLS_BCC	OPLS_ChelpG
1	-23.62	0.32	-27.89	-39.14	-25.14	-36.00
2	-2.49	0.85	-3.20	-5.26	-3.97	-5.98
3	-4.78	0.25	-1.66	-4.53	-1.67	-4.50
4	-4.45	0.24	N/A	N/A	N/A	N/A
5	-5.33	0.1	-6.02	-6.08	-5.81	-6.10
6	-5.26	0.18	-7.46	-10.04	-7.03	-9.64
9	-8.24	0.76	-10.62	-15.09	-10.17	-14.75
10	-6.24	0.38	-5.16	-5.88	-4.83	-5.63
11	-7.78	0.77	-9.78	-13.12	-9.23	-12.63
12	-3.75	0.21	-2.88	-6.74	-3.13	-6.92
13	-4.44	0.43	-3.00	-4.57	-2.87	-4.39
14	-4.09	0.17	-4.34	-9.17	-4.65	-9.33
15	-4.51	0.1	-3.41	-9.02	-3.76	-9.29
16	-3.2	0.27	-1.31	-4.63	-0.78	-4.06
17	-2.53	0.25	-2.77	-6.17	-2.76	-6.12
19	-3.78	0.1	-3.43	-0.39	-3.38	-0.53
20	-2.78	0.1	-1.99	1.47	-2.37	0.84
21	-7.63	0.12	-10.05	-13.40	-9.77	-13.17
22	-6.78	0.1	-8.43	-7.94	-8.88	-8.57
23	-9.34	0.62	-10.27	-3.75	-8.94	-3.19
24	-7.43	0.6	-10.01	-1.94	-9.07	-1.59
25	-5.73	0.15	-10.68	-14.38	-10.60	-14.46
26	-5.31	0.1	-9.37	-12.63	-9.72	-13.04
27	-4.8	0.39	-6.41	-13.91	-4.74	-12.33
28	-4.29	0.39	-5.61	-12.54	-4.10	-11.12
29	-1.66	0.1	-1.33	-5.02	-0.48	-4.23
30	-2.29	0.12	-3.94	-7.34	-3.97	-7.22
32	-7.29	0.1	-5.29	-6.95	-5.34	-7.04
33	-6.96	0.1	-9.97	-10.36	-9.45	-9.79
34	-5.8	0.1	-9.95	-11.91	-9.62	-11.51
35	-4.68	0.1	-5.60	-7.05	-5.56	-6.97
36	-5.66	0.1	-4.98	-4.27	-4.78	-4.14
37	-5.94	0.1	-10.25	-11.84	-9.91	-11.53

38	-3.93	0.1	-5.19	-6.66	-5.48	-6.98
39	-0.85	0.1	-0.57	0.85	-0.66	0.65
41	-5.05	0.1	-3.45	-3.74	-2.85	-3.12
42	-3.13	0.1	-2.04	-3.48	-1.82	-3.50
43	0.14	0.1	1.05	1.17	0.98	1.10
44	-5.08	0.1	-5.04	-6.57	-4.58	-6.30
45	-11.53	0.29	-15.55	-19.75	-15.99	-20.91
46	-9.44	0.74	-12.12	-15.07	-12.44	-15.95
47	-14.21	1.1	-21.63	-29.43	-21.48	-30.65
48	-11.85	0.35	-13.52	-18.87	-13.04	-19.37
49	-3.16	0.1	-3.64	-1.74	-3.86	-1.88
50	-4.14	0.1	-4.37	-0.47	-4.43	-0.76
51	-9.53	0.28	-10.79	-14.88	-10.38	-15.13
52	-2.87	0.69	-2.67	-0.60	-2.89	-1.08

*: N/A means direct QM optimization fails in short time (<5min)

Table S4-6: Comparing solvation free energy of SAMPL4 compounds with various force field combinations and QM optimization solute structures under HF/6-31+G*/IEFPCM

ID	EXP(kcal/mol)	Error	GAFF_BCC	GAFF_ChelpG	OPLS_BCC	OPLS_ChelpG
1	-23.62	0.32	-26.31	-36.72	-23.61	-33.70
2	-2.49	0.85	-3.09	-1.64	-3.87	-2.42
3	-4.78	0.25	-1.57	-1.84	-1.59	-1.88
4	-4.45	0.24	-4.78	-4.77	-4.77	-4.79
5	-5.33	0.1	-6.00	-3.53	-5.78	-3.39
6	-5.26	0.18	-7.42	-6.02	-6.99	-5.61
9	-8.24	0.76	-10.12	-8.43	-9.67	-8.05
10	-6.24	0.38	-4.87	-3.09	-4.55	-2.76
11	-7.78	0.77	-9.25	-8.02	-8.69	-7.49
12	-3.75	0.21	-2.81	-2.78	-3.05	-3.00
13	-4.44	0.43	-2.89	-2.14	-2.76	-2.02
14	-4.09	0.17	-4.21	-3.50	-4.52	-3.84
15	-4.51	0.1	-3.30	-3.33	-3.65	-3.68
16	-3.2	0.27	-1.22	-2.36	-0.70	-1.82
17	-2.53	0.25	-2.68	-2.28	-2.67	-2.28
19	-3.78	0.1	-3.43	0.51	-3.38	0.33
20	-2.78	0.1	-1.98	2.04	-2.36	1.38
21	-7.63	0.12	-8.16	-5.72	-7.93	-5.66
22	-6.78	0.1	-8.27	-5.66	-8.72	-6.28
23	-9.34	0.62	-9.41	-0.42	-8.03	0.12
24	-7.43	0.6	-9.82	0.26	-8.90	0.43
25	-5.73	0.15	-9.75	-8.52	-9.64	-8.34
26	-5.31	0.1	-9.06	-7.35	-9.38	-7.60
27	-4.8	0.39	-6.33	-9.24	-4.67	-7.69
28	-4.29	0.39	-5.53	-8.11	-4.02	-6.67
29	-1.66	0.1	-1.09	-2.81	-0.26	-2.03
30	-2.29	0.12	-4.89	-4.81	-4.82	-4.76
32	-7.29	0.1	-5.25	-5.03	-5.31	-5.11
33	-6.96	0.1	-9.91	-6.29	-9.39	-5.64
34	-5.8	0.1	-9.92	-7.71	-9.59	-7.28
35	-4.68	0.1	-5.56	-4.39	-5.52	-4.40
36	-5.66	0.1	-4.93	-2.79	-4.73	-2.70
37	-5.94	0.1	-10.21	-7.93	-9.87	-7.54

38	-3.93	0.1	-5.57	-4.17	-5.83	-4.49
39	-0.85	0.1	-0.58	1.30	-0.67	1.07
41	-5.05	0.1	-3.98	-2.69	-3.25	-2.02
42	-3.13	0.1	-2.01	-1.96	-1.79	-1.80
43	0.14	0.1	1.05	1.29	0.98	1.20
44	-5.08	0.1	-5.02	-4.65	-4.56	-4.16
45	-11.53	0.29	-15.51	-12.55	-15.92	-13.01
46	-9.44	0.74	-11.35	-9.28	-11.56	-9.55
47	-14.21	1.1	-21.10	-18.47	-20.91	-18.90
48	-11.85	0.35	-13.29	-11.49	-12.76	-11.30
49	-3.16	0.1	-3.64	-1.07	-3.85	-1.17
50	-4.14	0.1	-4.39	0.35	-4.44	0.02
51	-9.53	0.28	-11.27	-10.86	-10.96	-10.75
52	-2.87	0.69	-2.67	0.45	-2.89	0.01

Table S4-7: Comparing solvation free energy of SAMPL4 compounds with various force field combinations and QM optimization solute structures under MP2/6-31+G*/IEFPCM

ID	EXP(kcal/mol)	Error	GAFF_BCC	GAFF_ChelpG	OPLS_BCC	OPLS_ChelpG
1	-23.62	0.32	-28.11	-45.17	-25.37	-41.83
2	-2.49	0.85	-3.03	-8.15	-3.80	-8.71
3	-4.78	0.25	-3.40	-11.54	-3.35	-11.34
4	-4.45	0.24	0.00	N/A	N/A	N/A
5	-5.33	0.1	-6.12	-8.48	-5.91	-8.64
6	-5.26	0.18	-7.67	-13.33	-7.25	-12.91
9	-8.24	0.76	-10.70	-17.96	-10.35	-17.64
10	-6.24	0.38	-5.32	-7.20	-5.02	-7.00
11	-7.78	0.77	-9.85	-16.08	-9.33	-15.57
12	-3.75	0.21	-3.05	-9.83	-3.32	-10.04
13	-4.44	0.43	-3.10	-5.86	-2.96	-5.61
14	-4.09	0.17	-3.63	-10.30	-3.96	-10.49
15	-4.51	0.1	-3.62	-12.30	-3.97	-12.52
16	-3.2	0.27	-1.56	-6.04	-1.03	-5.45
17	-2.53	0.25	-3.06	-9.28	-3.04	-9.18
19	-3.78	0.1	-3.64	-1.69	-3.55	-1.76
20	-2.78	0.1	-1.86	0.46	-2.24	-0.11
21	-7.63	0.12	-10.27	-16.80	-9.97	-16.36
22	-6.78	0.1	-7.87	-9.39	-8.26	-9.81
23	-9.34	0.62	0.00	N/A	N/A	N/A
24	-7.43	0.6	-5.25	-1.67	-4.88	-1.54
25	-5.73	0.15	0.00	N/A	N/A	N/A
26	-5.31	0.1	-9.55	-16.59	-9.88	-17.00
27	-4.8	0.39	-6.38	-16.25	-4.72	-14.66
28	-4.29	0.39	-5.57	-14.63	-4.07	-13.23
29	-1.66	0.1	-1.37	-6.08	-0.52	-5.30
30	-2.29	0.12	-5.14	-12.38	-5.06	-11.96
32	-7.29	0.1	-5.38	-7.87	-5.44	-7.95
33	-6.96	0.1	-10.09	-13.37	-9.59	-12.86
34	-5.8	0.1	-7.75	-11.82	-7.31	-11.65
35	-4.68	0.1	-5.56	-9.84	-5.55	-9.74
36	-5.66	0.1	-5.01	-5.47	-4.81	-5.30
37	-5.94	0.1	-10.41	-14.97	-10.08	-14.74

38	-3.93	0.1	-5.70	-10.90	-5.96	-10.93
39	-0.85	0.1	-0.59	0.13	-0.68	-0.02
41	-5.05	0.1	-4.40	-5.17	-3.65	-4.34
42	-3.13	0.1	-2.13	-4.51	-1.91	-4.65
43	0.14	0.1	1.02	0.96	0.94	0.90
44	-5.08	0.1	-5.08	-8.30	-4.62	-8.23
45	-11.53	0.29	-16.35	-23.30	-16.77	-24.67
46	-9.44	0.74	-13.23	-18.26	-13.58	-19.26
47	-14.21	1.1	-20.93	-35.70	-20.27	-36.41
48	-11.85	0.35	-16.64	-23.82	-16.36	-23.90
49	-3.16	0.1	-3.61	-2.92	-3.84	-3.11
50	-4.14	0.1	-4.39	-1.85	-4.45	-2.08
51	-9.53	0.28	-12.49	-19.06	-12.20	-19.61
52	-2.87	0.69	-2.58	-2.18	-2.83	-2.74

*: N/A means direct QM optimization fails in short time (<10min)

4.3 Molecular density functional theory for multiscale modeling of hydration free energy

ABSTRACT

Recent developments in physical and computer sciences enable quantitative predictions of chemical reactions and thermodynamic data from first principles by multiscale modeling. The hierarchical approach integrates different theoretical frameworks ranging from those describing phenomena at the electronic length and time scales to those pertinent to complex biomolecular systems and macroscopic phase transitions, promising broad applications to problems of practical concern. Whereas multiscale modeling has been emerging as a popular computational tool for engineering applications, the connection between calculations at different scales is far from being coherent, and the multiple choices of quantum/classical methods at each scale renders numerous combinations that have been rarely calibrated against extensive experimental data. In this work, we have examined a multiscale procedure for predicting the solvation free energies of a large set of small molecules in liquid water at ambient conditions. Using the experimental data for the hydration free energies as the benchmark, we find that the theoretical results are sensitive to the selection of quantum-mechanical methods for determining atomic charges and solute configurations, the assignment of the force-field parameters in particular the atomic partial charges, and approximations in the statistical-mechanical calculations. Because of significant uncertainties in quantum-mechanical calculations and the semi-empirical nature of force-field models, computational efficiency makes the classical density functional

theory a valuable alternative to molecular simulations for future development and application of multiscale modeling methods.

4.3.1. Introduction

Chemical processes of practical concern mostly take place at multiple spatiotemporal scales. While chemical reactions and intermolecular interactions are determined by electronic activities at atomic levels, such interactions lead to collective thermodynamic behavior such as structure formation and phase transitions at macroscopic scales. As shown schematically in Figure 4-6, conventional multiscale methods to describe multiscale phenomena hinge on different theoretical frameworks including quantum-mechanical (QM) calculations of electronic properties, molecular dynamics (MD) or Monte Carlo (MC) simulations for dynamic and equilibrium properties determined by small length-scale correlations, coarse-grained models and continuum-level phenomenological equations for describing macroscopic phenomena and high-throughput computations[53]. Whereas recent theoretical developments make it possible for a quantitative connection of different quantum and classical approaches *via* hierarchical multiscale modeling, the integration between different models at various length and time scales is far from being perfect. Because of the uncertainties in the computational methods at each scale, a judicious combination of different methods is critically important for practical applications of multiscale modeling methods. In this work, we exam how the theoretical performance of multiscale modeling depends on different combinations of quantum and classical methods for predicting the solvation free energies of a large number of chemicals in liquid water at ambient conditions.

Solvent effects play a central role in solution thermodynamics. Despite a long history of studies, theoretical prediction of solvation free energy remains a bottleneck in understanding important chemical and biological processes in liquid surroundings. There have been enormous activities in this field, and the literature is vast. Recent interest has been mostly devoted to understanding the microscopic details of solute-solvent interactions and the effects of the local solvent structure on the chemical and biochemical affinities of dissolved species[1,54-67]. Such information is indispensible for studying molecular events in organic or aqueous systems including chemical reactions and relaxation dynamics[68], stability of biomacromolecules[69], and host-guest interactions for rational drug design [70-74].

Approximately, existing computational methods to investigate the solvent effects may be classified into three categories: continuous approaches[3,75-79], molecular simulations[80-84], and liquid-state theories[2,85-90]. A number of hybrid methods also exist by various combinations of these three basic procedures[22,91-95]. While continuous approaches are usually constructed from the knowledge-based, macroscopic properties of the solvent, molecular simulations and liquid-state theories rely on semi-empirical force fields to account for the solvent-solvent and the solute-solvent interactions. In principle, the solvation free energy can be directly predicted from a combination of quantum mechanical/molecular mechanical (QM/MM) calculations [96], but the first principles methods are computationally prohibitive for large-scale calculations[97-101]. To a certain degree, the "top-down" and "bottom-up" approaches are complementary, and a practical choice often reflects a compromise of the computational cost and the precision in the microscopic details. Because of the complexity of solute-solvent interactions at the atomic

scale, development of computational methods for *fast yet reliable* predictions of solvent effects remains a daunting theoretical challenge.

With solvent molecules depicted as a dielectric continuum, a continuous approach is able to describe the solvation free energy in terms of the geometric measures of the solute-solvent boundary, such as the solute size, solvent-accessible surface area, surface curvature and various energetic contributions due to the solute-solvent electrostatic and van der Waals interactions[4,102-104]. Neglecting the microscopic details of the solvent molecules makes the phenomenological approach computationally extremely efficient and thus convenient for practical applications[105-107]. The so-called primitive model of electrolyte solutions provides a prime example: with ions represented by spherical particles and the solvent as a dielectric continuum, it is able to reproduce both the solvation free energies and the thermodynamic properties of many simple electrolytes with remarkable accuracy[108-111]. The continuous approach is also applicable to a wide class of chemical systems containing neutral inorganic atoms, large and possibly polar organic molecules, and polyatomic ions[112,113]. Despite the great success, a major issue with the continuous methods is that they typically deploy semi-empirical parameters that are valid for a narrow range of thermodynamic conditions[5,49,114]. Besides, a continuous model neglects the local solvent inhomogeneity and the steric effects affiliated with individual solvent molecules. Recently, Nakamura *et al.* proposed an elegant procedure to account for the variation of the dielectric inhomogeneity near the solute using field-theoretic techniques[6]. The new theoretical method predicts ionic solvation free energies in both single-component liquids and binary liquid mixtures in excellent agreement with the experimental data. Other recent

developments include the solvent polarizable pseudo-particle approach[17] and the use of the frequency dependent multipolar polarizability to account for the dispersion contributions [115]. The applicability of these interesting developments to complicated chemical systems is yet to be tested.

Molecular simulation has been a popular choice to study solvent effects over the past few decades[116]. Excellent reviews are available focused on molecular simulations for computation of solvation free energy in liquid water[117,118]. Unlike those properties directly related to molecular mechanics and configurations, the free-energy calculation relies on alchemical methods to sample the microstates of the system along various thermodynamic pathways[119,120]. Although advanced simulation techniques have been developed to reduce the computational cost, the accelerated methods are not always robust [121-130]. Besides, simulation results have been rarely reported for the solvation free energies of a large library of chemical systems [117]. A notable exception is the MC data for the solvation properties of over 200 organic solutes in aqueous and organic solutions reported by Duff and Jorgensen[19]. More recently, Mobley *et al.* [20] calculated the hydration free energies of over 500 small organic molecules using MD simulations. These large-scale simulation studies provide fresh insights into the microscopic details of the solvent structure ignored in continuous models and, perhaps more importantly, invaluable benchmark data for calibrations of new theoretical developments.

Historically, most liquid-state methods were established in the context of simple fluids and coarse-grained models of molecular systems[131]. Over the past decade there have been tremendous developments in both integral-equation and density functional theories for an

atomistic description of molecular systems[15,16,18,47,132-140]. To a certain degree, liquid-state theories represent a good compromise between continuous approaches and molecular simulations. On the one hand, liquid-state theories retain the microscopic details of the solvent molecules that are accounted for in molecular simulations. Instead of sampling the microstates, the theoretical methods provide analytical expressions for correlation functions and thermodynamic properties including the solvation free energy and thus are computationally much more efficient than molecular simulations. On the other hand, liquid-state theories use continuous functions to describe the local structure of solvent molecules in an extended chemical environment. Like that for continuous approaches, the system size is much less an issue in numerical implementation of liquid-state theories. In a recent work[14], we proposed a molecular density functional theory (MDFT) based on the universality hypothesis of the bridge functional originally proposed by Rosenfeld[141]. With the bridge functional represented by the modified fundamental measure theory[141,142], MDFT allows us to predict the solvation structure and the solvation free energy in excellent agreement with molecular simulations[14,47,140].

The main purpose of this work is to investigate how the theoretical predictions of the hydration free energy are influenced by different first principles calculations for the solute configurations and by the selections of different solvent models and the solute parameters in the molecular force fields. Because MDFT is much faster than molecular simulations and yet provides fully atomistic descriptions of molecular systems, it renders an excellent avenue to investigating the connections between different theoretical procedures commonly used in multiscale modeling. While the theoretical methods within their specific

scales have been routinely examined, the existence of multiple choices in each scale leads to a large number of combinations. Calibration of different multiscale modeling methods has been rarely reported in the literature, in particular by comparison with extensive experimental data of practical concern.

This rest of this article is organized as follows: In the next section, we introduce the basic statistical-mechanical equations for calculating the solvation free energy. The force-field models used in this work are reviewed briefly in Section III. Section IV recapitulates the MDFT equations for predicting hydration free energy. Section V presents the numerical results and discussions. Finally, Section VI summarizes the main conclusions with prospects for future developments.

4.3.2. Statistical thermodynamics of solvation

We consider dissolution of a single solute molecule in liquid water at fixed temperature and pressure. The configuration of *M* atoms in the solute molecule can be defined by its atomic positions, $\mathbf{X} \equiv (\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_M)$. Given an atomic configuration of the solute, its interaction with a solvent molecule of atomic configuration $\mathbf{x}^{o}$ $(\mathbf{r}_{H1}, \mathbf{r}_O, \mathbf{r}_{H2})$ can be calculated from the pairwise additive Lennard-Jones (LJ) and Coulomb potentials:

$$\Upsilon(\mathbf{X}, \mathbf{x}) = \sum_{k=1}^{M} \sum_{i=\mathrm{H1,O,H2}} \left\{ 4\varepsilon_{ik} \left[(\sigma_{ik} / r_{ik})^{12} - (\sigma_{ik} / r_{ik})^{6} \right] + \frac{Z_i Z_k e^2}{4\pi\varepsilon_0 r_{ik}} \right\} \qquad (104)$$

where subscripts *k* and *i* denote atoms from the solute and solvent molecules, respectively, r_{ik} represents the center-to-center distance between *i* and *k* atoms, Z stands for (fractional) valence, *e* is the elementary charge, and ε_0 is the vacuum permittivity. For the water

models considered in this work, the LJ potential applies only to the interaction of the oxygen atom with the solute molecule. As in typical molecular simulations, we use the Lorentz-Berthelot combining rules for the LJ parameters, $\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}$, $\sigma_{ij} = (\sigma_{ii} + \sigma_{jj})/2$, where ε_{ii} and σ_{ii} are the energy and size parameters for atom i. For each configuration of the solvent molecules, the total solute-solvent energy is given by a summation of all solute-solvent interactions

$$\Phi(\mathbf{X}) = \sum_{j=1} \Psi(\mathbf{X}, \mathbf{x}_j) \tag{105}$$

The solvation free energy is defined by the reversible work to transfer the solute molecule from a vacuum into pure solvent

$$F_s \equiv -k_B T \ln \langle \exp[-\beta\Phi(\mathbf{X})] \rangle_{0+\mathbf{X}} \tag{106}$$

where k_B represents the Boltzmann constant, $\beta = 1/(k_B T)$, and $\langle \cdots \rangle_{0+\mathbf{X}}$ stands for the ensemble average over the microstates of the solute and solvent molecules, with subscripts 0 and $\mathbf{X}$ denoting the atomic configurations of the pure solvent and the solute, respectively. In experiment and most molecular simulations, the solvation free energy is evaluated at fixed temperature, pressure, and total volume. In other words, the reversible work corresponds to change in the Gibbs energy, ΔG.

Molecular simulation of the solvation free energy is commonly carried out in an isobaric-isothermal ensemble using the free-energy perturbation, thermodynamic integration or Bennett acceptance ratio methods[117]. However, theoretical predictions of the solvation free energy are mostly based on the grand canonical ensemble, *i.e.*, systems with

fixed solvent chemical potential μ and volume V instead of the number of solvent molecules N and pressure P. As illustrated schematically in Figure 4-7, the solvation free energy can be related to the change in the grand potential for the solvent as an open system, ΔW, and a reversible work due to the volume change ΔV[139]

$$F_s(\mathbf{X}) = \Delta W - P\Delta V \tag{107}$$

Eq.107 is valid only when the solute has a rigid configuration. In general, the solvation free energy must be calculated by considering all possible solute configurations

$$\beta F_s = -\ln\left\langle \exp[-\beta F_s(\mathbf{X})]\right\rangle_{\mathbf{X}} \tag{108}$$

where $\langle\cdots\rangle_{\mathbf{X}}$ denotes an ensemble average over all solute configurations in the vacuum. In this work, we are interested in the solvation free energies of small molecules so that the ensemble average may be estimated from a single or few solute configurations. At fixed T and P, the volume change due to the dissolution of a solute molecule can be identified as the solute partial molar volume at infinite dilution, i.e., $\Delta V = \bar{v}_s^{\infty}$. At ambient conditions, the partial molar volume of a small molecule in liquid water is typically on the order of 100 ml/mol[143,144], which gives $P\bar{v}_s^{\infty}$ 0.1 kJ/mol at P = 1 atm. The volumetric term is negligible in comparison to the typical values of the hydration free energy (~ kcal/mol) and is thus not considered in this work.

4.3.3. Force fields

Predicting the solvation free energy from atomistic models requires a force field to specify the inter- and intra- molecular interactions. The force-field approach was introduced originally for predicting the molecular structures of organic and biological

systems[145-147]. As described in the pioneering work by Lifson *et al.*[148], a conventional force field uses relatively simple mathematical functions to describe bond potentials and non-bonded inter-atomic interactions:

$$
\begin{aligned}
E = & \sum_{bond} K_b(b - b_o)^2 + \sum_{angle} K_q(q - q_o)^2 \\
& + \sum_{dihedral} \frac{K_j}{2}\left[1+\cos\left(nj - j_o\right)\right] \\
& + \sum_{impr} \frac{K_c}{2}\left[1+\cos\left(nc - c_o\right)\right] \\
& + \sum_{nonbond} \left\{ 4e_{ij}\left[\left(\frac{s_{ij}}{r_{ij}}\right)^{12} - \left(\frac{s_{ij}}{r_{ij}}\right)^{6}\right] + \frac{Z_i Z_j e^2}{4pe_0 r_{ij}} \right\}
\end{aligned}
\tag{109}
$$

In Eq.109, the first two terms on the right describe energies related to the deformations of bond length b and bond angle θ from their equilibrium values, b_0 and θ_0, respectively. The harmonic forms (with force constants K_b and K_θ) are used to ensure the correct topological structure for each molecule. The bond potentials are not intended to represent chemical changes such as bond breaking. The third and fourth terms (dihedral and improper angles) are affiliated with the rotations of the neighboring bonds. These terms are characterized by periodic variation energies of the torsion angles (with periodicity determined by n and heights of rotational barriers defined by K_φ and K_χ, respectively). The last two terms arise from pair interactions between non-bonded atoms, typically represented by the Lennard-Jones and the Coulomb potentials. Both the van der Waals and Coulomb potentials depend only on the atomic center-to-center distance $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$.

Semi-empirical force fields provide a convenient starting point for theoretical prediction of the energetic properties of molecular systems. However, a systematic

evaluation of the model parameters is rather challenging because they must reflect both the properties of individual molecules in the vacuum and experimental data for macroscopic systems. From the fundamental point of view, the force-field approach represents only an approximation of the QM descriptions of electrons and nuclei that constitute all molecular systems. Although the essential ingredients of an atom are invariant with the changes in thermo-physical properties including chemical reactions and phase transitions, these particles interact with each other without differentiating bonded and non-bonded interactions. In other words, it is the same electron cloud that determines chemical reactions, hydrogen bonding, and van der Waals forces.

Parameters related to the bond lengths, bond angles, torsional angles and atomic charges are typically determined from the QM calculations. Toward that end, first principles calculations are often based on the electronic DFT, which is computationally efficient but entails many choices to approximate the exchange-correlation functional. Because conventional DFT methods involve various local/semi-local density approximations that are not sufficiently accurate for long-range electrostatic correlations, the DFT methods are rarely used to fix the van der Waals parameters. Besides, the results from QM calculations are not strictly consistent with the pairwise additivity assumption for van der Waals interactions. As a result, the LJ parameters are usually obtained by fitting with experimental data such as liquid density and heat of evaporation. The connection of the macroscopic properties with the molecular parameters inevitably requires statistical-mechanical calculations. While such connection is conventionally established through MD

and MC simulations, the accuracy of a force field depends also on the quality of QM calculations, reliability of experimental data, and the scope of calibrations.

Many successful force fields have already been widely in use for multiscale modeling. For example, DREIDING[149] and UFF[150] cover essentially all elements in the periodic table and are able to achieve good quality for predicting the crystalline structure of nanoporous materials and gas adsorptions.[151-154] For applications to chemical and biophysical systems including solvation, common force fields include AMBER[155], CHARMM[156] and OPLS[40]. The TraPPE force field is a popular choice for phase-equilibrium calculations and for predictions of the thermophysical properties of gases and liquids[157,158]. To further improve the performance, a force field may include correlation terms like bond-angle correlation and angle-torsion correlation. The so-called Class II force fields, such as CFF93[159] and COMPASS[160], perform well for condensed phases containing organic molecules and polymers.

4.3.4. Molecular density functional theory (MDFT)

Theoretical details for the MDFT and the numerical procedure for its implementation can be found in our previous publications [14,47,140]. Here we recapitulate only the key equations and justifications for approximations used in formulation of the free energy functional.

In MDFT calculations, the solvation free energy is defined as the reversible work to transfer a solute molecule from the vacuum into a pure solvent at fixed temperature and the solvent chemical potential. The free energy can be evaluated using an open system consisting of solvent molecules in the presence of an external potential that arises from the

solute-solvent interactions. For a given configuration of the solute molecule, the solvation free energy corresponds to the difference between the grand potential of the pure solvent and that of the solute-solvent system

$$F_s[r(\mathbf{x});\mu,V,T] = \Omega[r(\mathbf{x});\mu,V,T] - \Omega_0[\mu,V,T] \tag{110}$$

where Ω_0 represents the grand potential of the pure solvent. As discussed earlier, the solvation free energy calculated from Eq.110 is virtually identical to the change in Gibbs energy derived from experiment or molecular simulations.

For a one-component system at given temperature T, volume V and chemical potential μ, the grand potential is defined as a functional of the molecular density profile, $r(\mathbf{x})$:

$$\begin{aligned}\Omega[\rho(\mathbf{x});\mu,V,T] = k_B T \int d\mathbf{x}\rho(\mathbf{x})\{\ln[\rho(\mathbf{x})\Lambda^3] - 1 - V_{\text{intra}}(\mathbf{x})\} \\ + \int d\mathbf{x}[\Psi(\mathbf{x}) - \mu]\rho(\mathbf{x}) + F^{ex}[\rho(\mathbf{x})]\end{aligned} \tag{111}$$

where $\mathbf{x}$ is a composite vector specifying the atomic positions of a solvent molecule. For the water models considered in this work, $\mathbf{x} \equiv (\mathbf{r}_{H1}, \mathbf{r}_O, \mathbf{r}_{H2})$ defined the positions of hydrogen and oxygen atoms. In Eq.111, the first term on the right side defines the *intrinsic* Helmholtz energy of an ideal-gas system, *i.e.*, a system with the same molecular density profile but without intermolecular interactions; Λ stands for an *effective* thermal wavelength, which is immaterial in solvation free-energy calculations; and $V_{\text{intra}}(\mathbf{x})$ represents the intramolecular potential. As defined in Eq. (1), $\Psi(\mathbf{x})$ represents the solute-solvent potential. For simplicity, here the solute configuration is not explicitly shown. The

last term, $F^{ex}[\rho(\mathbf{x})]$, denotes the excess intrinsic Helmholtz energy, *i.e.*, deviation from that of the ideal-gas system due to intermolecular interactions.

The density profile of the solvent molecules is calculated by minimizing the grand potential

$$\frac{\delta\Omega}{\delta\rho(\mathbf{x})} = 0 \tag{112}$$

Substituting Eq.111 for the grand potential into Eq.112 leads to the Euler-Lagrange equation

$$r(\mathbf{x}) \sim \exp[\beta\mu - \beta Y(\mathbf{x}) - \beta V_{\text{intra}}(\mathbf{x}) - \delta\beta F^{\text{ex}} / \delta r(\mathbf{x})] \tag{113}$$

where $\beta = 1/(k_B T)$, and the proportionality constant can be fixed by the bulk solvent density[161]. In numerical implementations of MDFT, we use the atomic density profiles[138,162]

$$r_i(\mathbf{r}) = r_i^0 \exp[-\omega_i(\mathbf{r})] < \exp\{-\sum_{j \neq i} \omega_j(\mathbf{r}_j)\} >_{\mathbf{x}'} \tag{114}$$

where r_i^0 is the number density of atom i in the bulk, $<\cdots>_{\mathbf{x}'}$ represents an average over all molecular conformations with the position for atom i fixed at $\mathbf{r}$. The reduced one-body potential, $\omega_i(\mathbf{r})$, is related to the external field for atom i, $\phi_i^{\text{ext}}(\mathbf{r})$, the atomic excess chemical potential μ_i^{ex}, and the deviation of the local excess chemical potential for each atom form the bulk value[163]

$$\omega_i(\mathbf{r}) = \beta\phi_i^{\text{ext}}(\mathbf{r}) + \delta\beta F^{\text{ex}} / \delta r_i(\mathbf{r}) - \beta\mu_i^{\text{ex}} \tag{115}$$

The atomic density as well as the atomic excess chemical potentials can be fixed by the normalization conditions. In writing Eq.115, we have assumed that the external potential for each molecule can be decomposed into contributions from individual atoms

$$Y(\mathbf{x}) = \sum_i j_i(\mathbf{r}_i) \tag{116}$$

The decomposition of the solute-solvent potential is fully consistent with the van der Waals and Coulomb potentials used in conventional force fields.

The physical meaning of Eq.114 is intuitively appealing: it resembles the Boltzmann equation for inhomogeneous atomic distributions. The effective one-body potential reflects the interaction of each atom with the surroundings, *i.e.*, other atoms from the same molecule and all other molecules, and the external field. With an analytical expression for the excess intrinsic Helmholtz energy and specific descriptions of the external and intramolecular potentials as detailed below, we can readily calculate the atomic density profiles, and subsequently the solvation free energy.

The intrinsic Helmholtz energy is an intrinsic property *solely* determined by the microscopic details of the solvent molecules. Without loss of generality, the excess *intrinsic* Helmholtz energy can be expressed relative to that of a bulk system with the same temperature, volume and chemical potential:

$$\begin{aligned} F^{ex}[r_i(\mathbf{r})] = F^{ex}[r_i^{\,0}] + \sum_i m_i^{ex} \int d\mathbf{r} \Delta r_i(\mathbf{r}) \\ - \frac{k_B T}{2} \sum_{i,j} \int d\mathbf{r} \int d\mathbf{r}' \Delta r_i(\mathbf{r}) \Delta r_j(\mathbf{r}') c_{ij}^{(2)}(|\mathbf{r} - \mathbf{r}'|) + F_B[r_i(\mathbf{r})] \end{aligned} \tag{117}$$

In Eq.117, $\Delta r_i(\mathbf{r}) = r_i(\mathbf{r}) - r_i^0$ is the deviation of the local atomic density from the bulk value, and $c_{ij}^{(2)}(r)$ denotes the site-site direct correlation function (DCF) of the bulk system

$$c_{ij}^{(2)}(\mathbf{r},\mathbf{r}') \equiv -\frac{\delta^2 \beta F^{ex}}{\delta r_i(\mathbf{r}) \delta r_j(\mathbf{r}')} \tag{118}$$

The last term on the right size of Eq.117, $F_B[\rho_i(\mathbf{r})]$, defines the bridge functional, which accounts for all contributions to the excess Helmholtz free energy beyond the quadratic expansion, i.e., all high-order terms in the functional Taylor expansion of the excess intrinsic Helmholtz energy.

Eq.117 is formally exact and immaterial to the additivity of the intermolecular potentials. In a previous work[36,164], we have shown that the site-site DCFs for water can be obtained from molecular simulation for the bulk system. Whereas we have no *a priori* knowledge for the bridge functional, it has been well documented in the liquid-state literature that the high-order terms are dominated by short-range interactions, insensitive to the mathematical details of the long-range intermolecular forces[131]. The insensitivity of the bridge functional to the precise form of the intermolecular potential is known as *the bridge universality*, proposed first by Rosenfeld about 20 years ago[165-167]. The *universality ansatz* was supported by the fact that the quadratic approximation (*i.e.*, without the bridge functional) conforms to the exact results at both the "ideal gas" and "ideal liquid" limits for the long-range components of the intermolecular potentials[168]. Using the fundamental measure theory (FMT) as an input, Rosenfeld also demonstrated that the universality hypothesis performs well for a wide variety of liquid systems, including charged Yukawa fluids and plasma mixtures (see ref[167] for an overview).

The universality *ansatz* allows us to approximate the bridge functional with that of a reference hard-sphere (HS) system:

$$F_B[\rho_i(\mathbf{r})] \approx F_{HS}^{ex}[\rho_i(\mathbf{r})] - F_{HS}^{ex}(\rho_i^0) - \sum_i \mu_{i,HS}^{ex} \int d\mathbf{r} \Delta\rho_i(\mathbf{r}) + \frac{k_B T}{2} \sum_{i,j} \int d\mathbf{r}_1 \int d\mathbf{r}_2 \Delta\rho_i(\mathbf{r}_1) \Delta\rho_i(\mathbf{r}_2) c_{i,j}^{HS}(|\mathbf{r}_1 - \mathbf{r}_2|) \quad (119)$$

where $F_{HS}^{ex}[\rho_i(\mathbf{r})]$ represents the excess Helmholtz energy functional of the reference system, $\mu_{i,HS}^{ex}$ and $F_{HS}^{ex}(r_i^0)$ are, respectively, the excess chemical potential and the excess Helmholtz energy of the reference system at bulk density r_i^0; and $c_{ij}^{HS}(|\mathbf{r}_1 - \mathbf{r}_2|)$ is the corresponding direct correlation functions. Accurate expressions are available for the excess Helmholtz energy functional and correlation functions of inhomogeneous hard-sphere systems[142,169,170]. For water models considered in this work, oxygen atoms dominate short-range interactions or the excluded volume effects. In that case, we may assume that the bridge functional depends only on the density profiles of oxygen atoms. This assumption is justified by the small size of hydrogen atoms in comparison to oxygen. Whereas in principle the hard-sphere diameter could be determined self-consistently, we estimate this parameter by simply reproducing the solvation free energy for a single solute (e.g., methane). The approximation yields good results for a large number of small organic solutes but with much improved computational efficiency.

With the atomic density profiles solved from Eq.113, we can readily calculate the solvation free energy from an analytical equation

$$\beta F[\rho_i(\mathbf{r})] = -\frac{1}{M_s}\sum_i \int d\mathbf{r}\Delta\rho_i(\mathbf{r}) + \frac{1}{2}\sum_{i,j}\iint d\mathbf{r}d\mathbf{r}' c_{ij}^{(2)}(|\mathbf{r}-\mathbf{r}'|)[\rho_i(\mathbf{r})\rho_j(\mathbf{r}') - \rho_i^0\rho_j^0] + \beta F^{\mathrm{B}}[\rho_i(\mathbf{r})] - \sum_i \int d\mathbf{r}\rho_i(\mathbf{r})B_i(\mathbf{r}) \quad (120)$$

where $M_s = 3$ for all water models used in this work, and

$$B_i(\mathbf{r}) \equiv dF_B / d\rho_i(\mathbf{r}). \quad (121)$$

We have demonstrated in a previous work that the MDFT is able to reproduce the simulation data for the hydration free energies of hundreds of chemicals [47]. The numerical accuracy is comparable to that inherited from the molecular models for water and for the solute molecules. For the hydration free energies of about 500 small molecules studied by simulation, the MDFT yields an average unsigned error of 1.04 kcal/mol, which is close to the averaged absolute difference between MD simulation and experiment data, 1.03 kcal/mol.

4.3.5. Results and discussion

The multiscale procedure for predicting hydration free energy connects first principles calculations for determining the atomic charge set of each solute molecule, selection of force-field parameters for describing intramolecular and van der Waals interactions, optimization of the solute configuration, and the MDFT calculations for the molecular density profiles and grand potential. To calibrate its performance with experimental data, we consider a number of different combinations of the multiscale methods and compare the theoretical results with the hydration free energies of a large library of chemical compounds. The experimental data for these chemicals (~700) were originally complied

by the groups of Mobley[20] and Truhlar [171]. These compounds cover many functional groups and fragments commonly used in general chemical products and rational drug design. Because conventional force fields were rarely calibrated with a large set of solubility data, this test provides a good benchmark for validating the quantitative performance of multiscale modeling methods. As a hallmark of first principles calculations, we do not use any *a priori* information about the experimental data for these chemicals.

Three common water models were used in our MDFT calculations: TIP3P[38], SPC[172], and SPC/E[173]. For each solute molecule, the van der Waals parameters were obtained from either GAFF[39] or OPLS-AA[40] force fields; the atomic partial charges and the solute configuration were calculated from different QM methods. To predict the atomic configuration of each solute molecule, we have tested a number of optimization methods, including energy minimization in the vacuum or in the solvent with a semi-empirical force field, or by direct QM calculations. In the latter case, the solute structures were determined by using different level of QM methods (HF/HF-3c/BLYP) with two kinds of basis sets (SVP/TZVP). We employed the geometry correction method (gCP) and the D3-BJ correction function, respectively, to correct errors due to intramolecular basis set superposition and dispersion effects improperly handled in the electronic DFT calculations [174]. According to the GMTKN30 benchmark for common DFT calculations based on GGAs and meta-GGAs [175], the dispersion corrected BLYP-D3 functional is among the most reliable methods for predicting noncovalent interactions. The accuracy is comparable to MP2 but with much less computational cost. In our previous work for a small set of solutes (SAMPL4 blind test)[140], we found that the input solute structures from the QM calculations

in the solvent (HF/6-31+G*/IEFPCM) yields the best agreement between MDFT predictions and experimental data. Here additional solute configurations have been generated from alternative QM methods, either with or without using the COSMO solvation model to account for the solvent effect[176].

The atomic partial charges for each solute molecule can be assigned according to the AM1-BCC charge set[41] or the ChelpG charge set[42]. The former is a common choice for the GAFF force field, and the latter is often used in CHARMM[43] and OPLS[40] models. With the QM optimized molecular structures as the input, the AM1-BCC charge sets were obtained from Antechamber-1.27 package[44], while the ChelpG charge sets were obtained directly from the QM structure optimization. To accelerate the computational speed, we used the RIJCOSX acceleration method for Hartree-Fock (HF) calculations and the RIJ method for BLYP calculations. All the quantum mechanics calculations are carried with the ORCA 3.0.1 software package [177].

Tables 4-3~4-5 summarize the overall performance of various combinations of the multiscale methods for predicting the solvation free energies of 700 chemicals in liquid water 25 ^{o}C and 1 atm. The structures of these compounds were obtained from different QM methods as specified in the 2nd and 3rd columns. As shown in Eq.(1), the solute-solvent interaction depends explicitly on the solute structure and such interaction dictates the inhomogeneous distribution of solvent molecules. For easy comparison, Figure 4-8 shows the average unsigned error (AUE) values from the different multiscale procedures. While all combinations perform reasonably well in comparison with the experimental data, noticeable differences can be identified that are of chemical significance (~1 kcal/mol on

average). The best combination is provided by calculations based on the TIP3P water model, the solute configurations generated from HF/SVP optimization, OPLS LJ parameters, and ChelpG partial charges (the cheapest QM combination among our testing sets). For this particular combination, the MDFT predictions give 1.35 kcal/mol for AUE and 2.13 kcal/mol for RMSD (see the row with fold fonts in Table 3). These values are comparable to best hydration free energy predictions using MD simulation (AUE is 0.68 kcal/mol and RMSD is 1.26 kcal/mol), and they are close to the discrepancies between results from MD simulation and MDFT calculations[47]. One possible reason for the good performance of this particular combination is that HF is a simple QM method that generates less environmentally dependent charges. While high-level DFT methods give more accurate charge distribution for a specific molecular configuration, the results are sensitive to the changes in solute configurations and to the solute-solvent interactions. Because MDFT calculations involve only a rigid configuration for each solute molecule, it appears that the simple QM method yields a better-averaged charge distribution and, a certain degree, takes into account an effective solute flexibility. In stark contrast to our previous work for calibration of MDFT with MD simulations [140], QM methods that account for the solvent effect on the solute charge and atomic configuration lead to poor MDFT predictions in comparison with the experimental data. Given the same structure generation method and the solute charge set, however, the MDFT predictions are rather incentives to different sets of van der Waals parameters.

Figure 4-9 shows a comparison of the best theoretical predictions with the experimental data for about 700 small molecules. Overall the good correlation between theory and

experimental data demonstrates the robustness of the multiscale modeling methods used in this work. Unlike simulation or experiment, there are no error bars for the MDFT predictions; the results are obtained by solving a set of equations instead of sampling of different microstates. Figure 4 shows that the multiscale procedure performs better for hydrophobic solutes and less accurate for hydrophilic compounds (F_s < -5 kcal/mol). One possible reason is that the MDFT calculation ignores the flexibility of solute structures and assumes that the solute structure is invariant with the environment. The structure effect is especially important for large hydrophilic molecules that form hydrogen bonds with water molecules. The structures of such molecules are drastically different in the vacuum and in water and cannot be captured by a single conformation. In a previous work[47], we demonstrated that the flexibility effect could be successfully taken into account by sampling the solute flexibility with a relatively small number of equilibrium conformations in the vacuum (the improvement is around 0 ~ 4 kcal/mol[47]). Another possible reason for the deviation between theory and experiment lies in the selection of force-field parameters. Considering most of the compounds here were not considered in the original force fields training set, we expect that better results are attainable by optimizing parameters targeted for hydration free-energy calculations, in particular those related to hydrophilic functional groups like –OH and –O-NO_2. The increased discrepancy may also be attributed to the approximations used in the MDFT calculations. In comparison with MD simulations, the errors introduced in the MDFT calculations are mainly due to the approximations for the bridge functional. In formulation of the free-energy density functional, we have utilized a hard-sphere model to represent the bridge functional for water wherein the hard-sphere

diameter of the reference system was calibrated with the solvation free energy for methane[14]. Because the local structure of water molecules is sensitive to the solute hydrophobicity, the parameter obtained from a small hydrophobic solute may become less reliable for hydrophilic solutes. Since the effective hard-sphere diameter for a single hydrophilic solute is smaller than that for a hydrophobic solute due to the enhanced solute-solvent attraction.

As expected, the multiscale procedure is sensitive to different combinations of first-principle methods to assign the atomic charges and the solute structure. For example, the best combination discussed above may turn into worse predictions if the AM1-BCC charge set is used instead (AUE=1.83, RMSD=2.72 kcal/mol, see Table 4-5). Because these numbers are averaged over 700 chemicals, the difference is more dramatic if one considers specific solutes, as shown by the open circles in Figure 4-9. The hydration free energies of many solutes are overestimated by the AM1-BCC charge set, leading to the much larger AUE.

In our MDFT calculations, structure optimization methods could contribute to more than 100% RMS deviation for the solvation free energy calculations. Besides, the theoretical performance is sensitive to the solvent model. Among three common water models considered in this work, we found that MDFT performs the best with the TIP3P model. Therefore, we consider the effects of solute structure primarily on the base of this solvent model with different QM optimization methods for the solute configurations. As indicated above, among all combinations of solvent models and structure optimization methods tested, the best result was achieved with the solute structures obtained from

HF/SVP calculations in the vacuum, along with OPLS van der Waals parameters, the ChelpG charge set, and the TIP3P solvent model. Ironically, we see no improvement in the MDFT predictions when the BLYP functional is used in electronic DFT calculations, even with the correction for dispersion (D3-BJ) and basis set superposition errors (BSSE). Because the atomic charges and the configuration of a solute molecule is sensitive to the changes in the local solvent environment, more sophisticated QM methods do not warrant better input for the atomistic structure used in our MDFT calculations. A similar conclusion holds for the COSMO solvent model. Interestingly, the HF-3c method [178], which supposes to be used as an efficient method to handle large systems, gives even worse results than the original HF method, indicating that it may not be a good choice for generating the atomic structures of solute molecules.

For systems considered in this work, the MDFT calculation can be readily processed in a modern desktop computer. As shown in Figure 4-10, the computational cost varies from few minutes to about an hour, depending on the solute type. By contrast, explicit-solvent simulation methods require at least hundreds of CPU hours per solute[26]. For predicting the hydration free energies of 700 solutes shown in Figure 4-9, the average computation time is 23 minutes. The computational speed can be further improved by optimizing the MDFT programs. For large-scale calculations, the MDFT calculations can be easily implemented in parallel computers or with modern GPUs.

4.3.6. Conclusions and perspectives

In this work, we have tested various combinations of multiscale methods for predicting the hydration free energies of 700 small molecules at the ambient condition. Although the

systems considered are relatively simple, the computational procedure entails all essential components of multiscale modeling, *i.e.*, predictions of atomic charges and solute configurations from first principles, the selection of force-field parameters, and statistical-mechanical calculations of thermodynamic variables. The application of molecular density functional theory (MDFT) empowers the calibration of different multiscale procedures with experimental data for a large library of systems. Among all solute structures determined from QM calculations in combination with various force-field methods, we find that the best combination is provided by calculations based on the TIP3P model for liquid water, with the solute configurations generated from HF/SVP optimization, the OPLS LJ parameters, and the ChelpG partial charges. For this particular combination, MDFT predict the hydration free energies of 700 small molecules with AUE = 1.35 kcal/mol and RMSD = 2.13 kcal/mol in comparison with experimental data. Considering the computational efficiency of both QM and classical DFT calculations, we expect that this combination will be useful for practical applications, in particular for large-scale screening of the solubility of small molecules.

As in molecular simulations, our MDFT calculations are based on the pairwise additive potentials that do not account for the polarization effects. However, the importance of non-additive interactions in liquid systems has been well recognized [179,180]. Inclusion of the polarization effects may drastically increase the computational costs of traditional simulation methods, typically by one order magnitude in comparison to that with non-polarizable models[181-183]. From a practical point of view, there is yet no convincing procedure to properly account for the effects of polarization on solvation free energy

calculations[184]. The computational efficiency of MDFT makes it an outstanding candidate for development and applications of polarizable force fields.

Whereas conventional multiscale modeling methods use QM calculations for predicting the electronic properties and classical statistical-mechanical methods for integrating the atomic degrees of freedom, a large class of chemical systems consists of light elements with strongly coupled electron-nuclei correlations that are not amenable to cascade calculations[185-191]. The nuclear quantum effects are significant even at ambient conditions for heterogeneous materials containing metallic elements and liquid water or hydrogen gas. From a practical perspective, such systems are particularly important because they are commonplace in electrochemical systems for energy conversion and storage. Besides, it has been well recognized that hydrogen gas exhibits quantum effects at relatively high temperature, and many unique properties of liquid water are intrinsically related to the inseparability of the electronic and nuclei degrees of freedom[192,193]. Understanding the nuclear quantum effects is thus important in theoretical investigations of many chemical reactions in liquid water and in computational design of novel materials for hydrogen storage[194]. Because of the broad implication of water and hydrogen gas in technological applications, theoretical developments beyond the Born-Oppenheim approximation are urgently needed[195].

From a theoretical perspective, both electronic and classical DFT use one-body density profiles as fundamental variables to describe the properties of many-body systems[196-207]. While the original DFT concepts, introduced first by Pierre Hohenberg and Walter Kohn, were intended to provide an alternative to the Schrödinger equation for predicting the

ground-state properties of electronic systems[208], the mathematical framework is applicable to electronic systems at finite temperature[209], to thermodynamic systems of classical particles[210,211], to coarse-grained models of polymeric systems[163], and to multicomponent mixtures of quantum particles[212-215]. Regrettably, none of the generalized versions has received as much attention as the Kohn-Sham (KS) implementation of DFT for inhomogeneous electronic systems at 0 K [216-218]. Because of its extreme popularity in computational chemistry and materials science for predicting the ground-state electronic properties and various static response functions of atoms, molecules, and solids, today DFT becomes almost synonymous to KS-DFT, even though the applicability of DFT for a wide variety of thermodynamic systems has also been well documented[219]. Because the mathematic procedure is similarly applicable to problems at different scales, we expect that DFT will be useful for future developments of multiscale modeling methods without empirical separation and re-combination of quantum and classical calculations.

Figure 4-6 Multiscale modeling aims to predict thermodynamic properties from a cascade of theoretical methods. Because of multiple choices at each scale and inevitable approximations, multiscale modeling may lead to various combinations of a large number of theoretical methods.

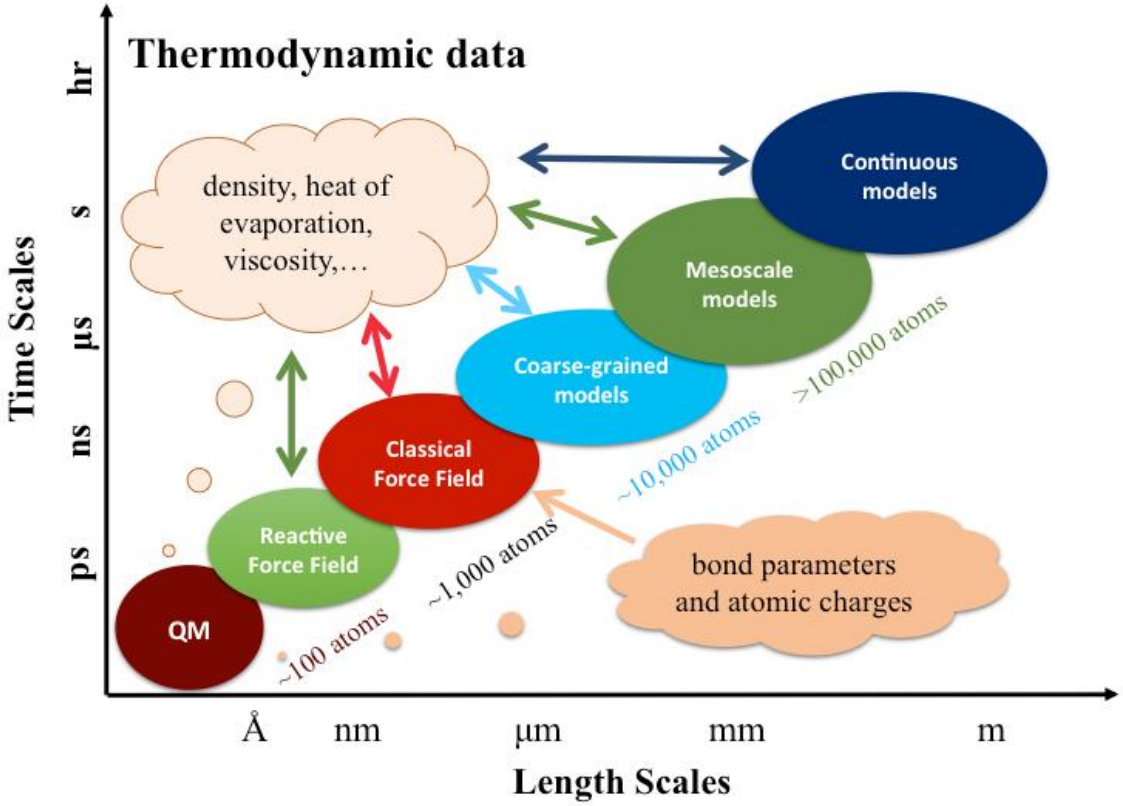


Figure 4-7 In experiment or conventional MD simulations, the solvation free energy is defined in terms of the change in the Gibbs energy, ΔG, at constant temperature T, pressure P and the number of solvent molecules N (top path). The MDFT calculation gives the change in the grand potential, ΔW, at constant temperature, volume V and the solvent chemical potential μ (top path). The difference between ΔG and ΔW can be found by a thermodynamic path that I) defining an open system with the volume and the solvent chemical potential the same as those corresponding to the closed system; II) inserting a solute molecule at constant volume and the solvent chemical potential; ΔN solvent molecules are removed due to the solute insertion; III) expanding the system volume at constant solvent chemical potential. In the thermodynamic limit, the solvent properties in the bulk remain unchanged due to the solvation of a single solute. For small molecules, the reversible work of expansion, $-P\Delta V$, is typically much smaller than other contributions to the solvation free energy.

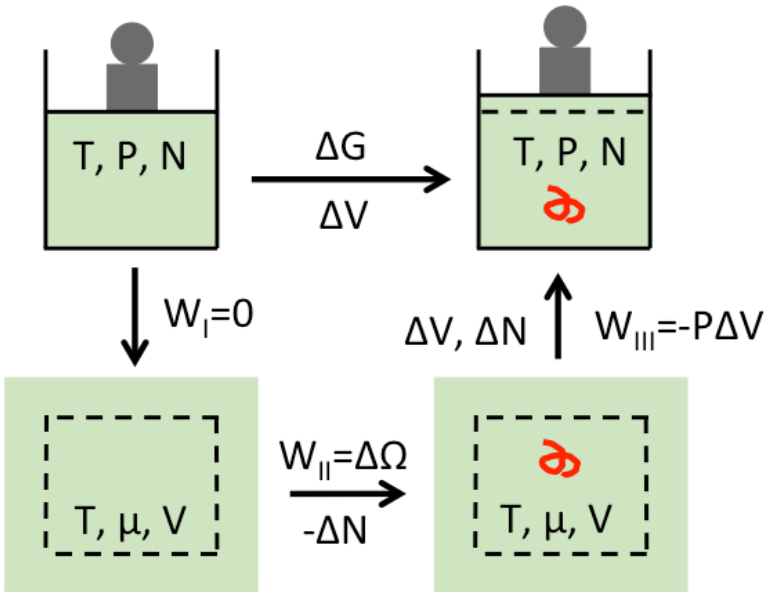
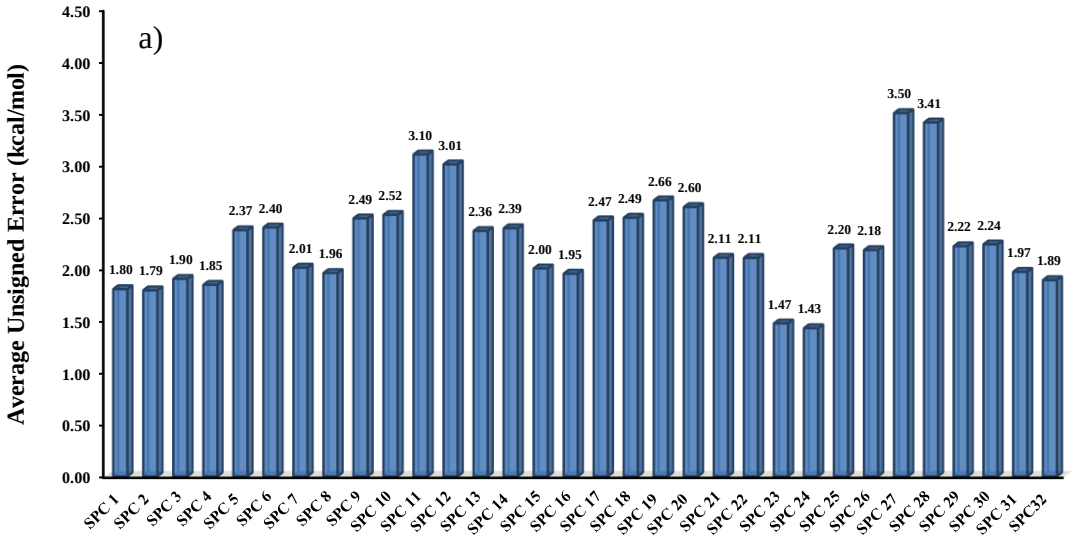
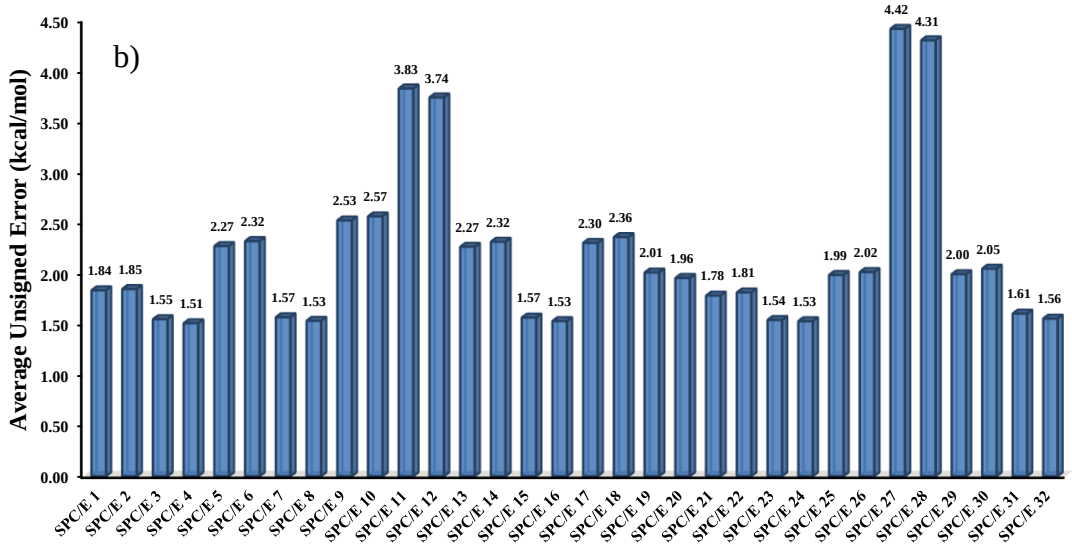


Figure 4-8 Comparison of experimental data with MDFT predictions based on 3 common water models a) SPC, b) SPC/E, and c) TIP3P. The codes at the bottom denote different combination of the solvent model, force field parameters and charge sets for the solute, and optimization methods for the solute structure as defined in Tables 1-3, the number ids correspond to the list order. For all solutes considered in this work, the average AUE is 2.18 kcal/mol and RMSD is 3.58 kcal/mol.




b)
Average Unsigned Error (kcal/mol)
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
1.84 1.85 1.55 1.51 2.27 2.32 1.57 1.53 2.53 2.57 3.83 3.74 2.27 2.32 1.57 1.53 2.30 2.36 2.01 1.96 1.78 1.81 1.54 1.53 1.99 2.02 4.42 4.31 2.00 2.05 1.61 1.56
SPC/E 1 SPC/E 2 SPC/E 3 SPC/E 4 SPC/E 5 SPC/E 6 SPC/E 7 SPC/E 8 SPC/E 9 SPC/E 10 SPC/E 11 SPC/E 12 SPC/E 13 SPC/E 14 SPC/E 15 SPC/E 16 SPC/E 17 SPC/E 18 SPC/E 19 SPC/E 20 SPC/E 21 SPC/E 22 SPC/E 23 SPC/E 24 SPC/E 25 SPC/E 26 SPC/E 27 SPC/E 28 SPC/E 29 SPC/E 30 SPC/E 31 SPC/E 32

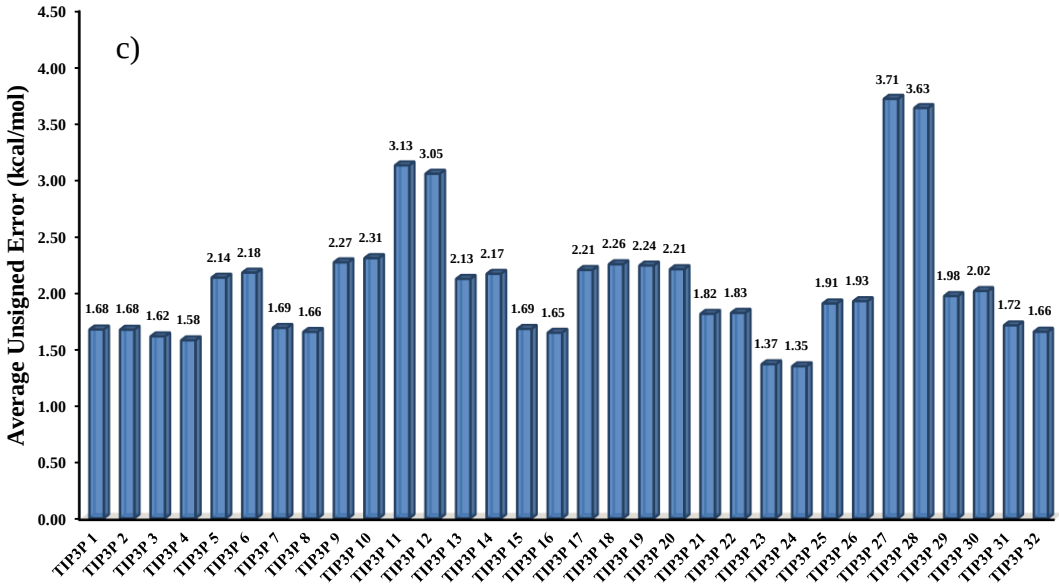

c)
Average Unsigned Error (kcal/mol)
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
1.68 1.68 1.62 1.58 2.14 2.18 1.69 1.66 2.27 2.31 3.13 3.05 2.13 2.17 1.69 1.65 2.21 2.26 2.24 2.21 1.82 1.83 1.37 1.35 1.91 1.93 3.71 3.63 1.98 2.02 1.72 1.66
TIP3P 1 TIP3P 2 TIP3P 3 TIP3P 4 TIP3P 5 TIP3P 6 TIP3P 7 TIP3P 8 TIP3P 9 TIP3P 10 TIP3P 11 TIP3P 12 TIP3P 13 TIP3P 14 TIP3P 15 TIP3P 16 TIP3P 17 TIP3P 18 TIP3P 19 TIP3P 20 TIP3P 21 TIP3P 22 TIP3P 23 TIP3P 24 TIP3P 25 TIP3P 26 TIP3P 27 TIP3P 28 TIP3P 29 TIP3P 30 TIP3P 31 TIP3P 32

Figure 4-9 Comparison of the MDFT calculations with experimental data for hydration free energies of 700 small solutes in liquid water. Filled cycles are calculated from the ChelpG charge set, and open cycles are from the AM1-BCC charge set. In both cases, OPLS van der Waals parameters were used for the solutes, and TIP3P model was used for water.

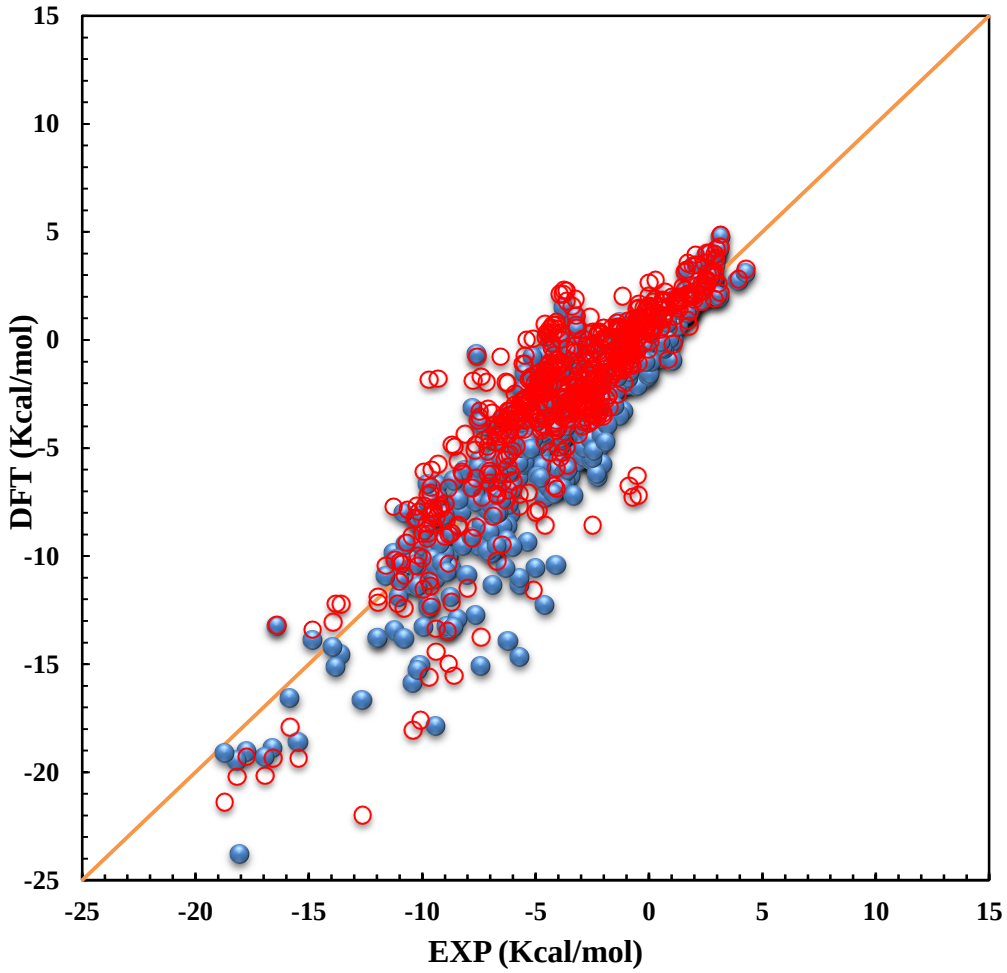


Figure 4-10 The distribution of computational time used in MDFT calculations for predicting the solvation free energies of 700 compounds. Here the numbers are based on HF/SVP optimization solute structures in the vacuum, OPLS-aa van der Waals parameters, the ChelpG charge set, and the TIP3P water model. The average computational time is 23 min, based on MDFT calculation on a desktop PC with single Intel E-1230 CPU core.

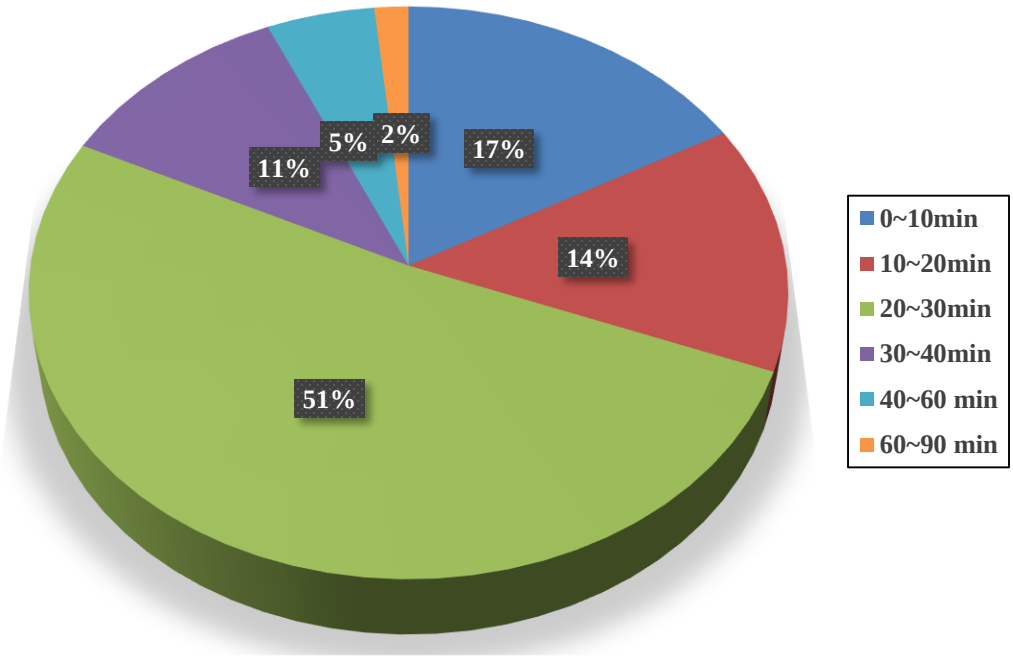


Table 4-3 Average unsigned errors (AUE) and root mean-square deviations (RMSD), both in units of kcal/mol, for different multiscale predictions of the hydration free energies of 700 small molecules in SPC water.

Water Model	QM Method	Basis Set	D3	gCP	COSMO	LJ Parameter	Charge	AUE	MSD
SPC	OMEGA[*1]	N/A	N/A	N/A	N/A	GAFF	BCC	1.80	3.35
SPC	OMEGA	N/A	N/A	N/A	N/A	OPLS	BCC	1.79	3.32
SPC	OMEGA	N/A	N/A	N/A	N/A	GAFF	ChelpG	1.90	2.53
SPC	OMEGA	N/A	N/A	N/A	N/A	OPLS	ChelpG	1.85	2.52
SPC	BLYP	TZVP	Yes	Yes	No	GAFF	BCC	2.37	4.05
SPC	BLYP	TZVP	Yes	Yes	No	OPLS	BCC	2.40	4.12
SPC	BLYP	TZVP	Yes	Yes	No	GAFF	ChelpG	2.01	2.62
SPC	BLYP	TZVP	Yes	Yes	No	OPLS	ChelpG	1.96	2.61
SPC	BLYP	TZVP	Yes	Yes	Yes	GAFF	BCC	2.49	4.47
SPC	BLYP	TZVP	Yes	Yes	Yes	OPLS	BCC	2.52	4.53
SPC	BLYP	TZVP	Yes	Yes	Yes	GAFF	ChelpG	3.10	5.67
SPC	BLYP	TZVP	Yes	Yes	Yes	OPLS	ChelpG	3.01	5.57
SPC	BLYP	TZVP	Yes	No	No	GAFF	BCC	2.36	4.04
SPC	BLYP	TZVP	Yes	No	No	OPLS	BCC	2.39	4.11
SPC	BLYP	TZVP	Yes	No	No	GAFF	ChelpG	2.00	2.61
SPC	BLYP	TZVP	Yes	No	No	OPLS	ChelpG	1.95	2.61
SPC	BLYP	SVP	Yes	Yes	No	GAFF	BCC	2.47	4.18
SPC	BLYP	SVP	Yes	Yes	No	OPLS	BCC	2.49	4.25
SPC	BLYP	SVP	Yes	Yes	No	GAFF	ChelpG	2.66	3.12
SPC	BLYP	SVP	Yes	Yes	No	OPLS	ChelpG	2.60	3.10

SPC	HF	SVP	No	No	No	GAFF	BCC	2.11	2.80
SPC	HF	SVP	No	No	No	OPLS	BCC	2.11	2.87
SPC	HF	SVP	No	No	No	GAFF	ChelpG	1.47	2.08
SPC	HF	SVP	No	No	No	OPLS	ChelpG	1.43	2.08
SPC	HF	SVP	No	No	Yes	GAFF	BCC	2.20	3.25
SPC	HF	SVP	No	No	Yes	OPLS	BCC	2.18	3.28
SPC	HF	SVP	No	No	Yes	GAFF	ChelpG	3.50	5.75
SPC	HF	SVP	No	No	Yes	OPLS	ChelpG	3.41	5.65
SPC	HF-3c[*2]	MINI	Yes	N/A	No	GAFF	BCC	2.22	3.51
SPC	HF-3c	MINI	Yes	N/A	No	OPLS	BCC	2.24	3.60
SPC	HF-3c	MINI	Yes	N/A	No	GAFF	ChelpG	1.97	2.78
SPC	HF-3c	MINI	Yes	N/A	No	OPLS	ChelpG	1.89	2.74
Overall								2.28	3.56

*1: The solute structures were generated using the OMEGA-TK software for 504 molecules provided by Mobley et al[20] and using M062X/MG3S quantum mechanics method optimized in vacuum environment for 196 molecules provided by Truhlar et al.[76]

*2: HF-3c proposed by Grimme et al. corrects effect due to dispersion and basis set superposition errors.[178]

Table 4-4 Average unsigned errors (AUE) and root mean-square deviations (RMSD), both in units of kcal/mol, for different multiscale predictions of the hydration free energies of 700 small molecules in SPC/E water.

Water Model	QM Method	Basis Set	D3	gCP	COSMO	LJ Parameter	Charge	AUE	MSD
SPC/E	OMEGA	N/A	N/A	N/A	N/A	GAFF	BCC	1.84	3.83
SPC/E	OMEGA	N/A	N/A	N/A	N/A	OPLS	BCC	1.85	3.80
SPC/E	OMEGA	N/A	N/A	N/A	N/A	GAFF	ChelpG	1.55	2.37
SPC/E	OMEGA	N/A	N/A	N/A	N/A	OPLS	ChelpG	1.51	2.36
SPC/E	BLYP	TZVP	Yes	Yes	No	GAFF	BCC	2.27	4.31
SPC/E	BLYP	TZVP	Yes	Yes	No	OPLS	BCC	2.32	4.41
SPC/E	BLYP	TZVP	Yes	Yes	No	GAFF	ChelpG	1.57	2.37
SPC/E	BLYP	TZVP	Yes	Yes	No	OPLS	ChelpG	1.53	2.39
SPC/E	BLYP	TZVP	Yes	Yes	Yes	GAFF	BCC	2.53	4.88
SPC/E	BLYP	TZVP	Yes	Yes	Yes	OPLS	BCC	2.57	4.96
SPC/E	BLYP	TZVP	Yes	Yes	Yes	GAFF	ChelpG	3.83	6.94
SPC/E	BLYP	TZVP	Yes	Yes	Yes	OPLS	ChelpG	3.74	6.80
SPC/E	BLYP	TZVP	Yes	No	No	GAFF	BCC	2.27	4.30
SPC/E	BLYP	TZVP	Yes	No	No	OPLS	BCC	2.32	4.39
SPC/E	BLYP	TZVP	Yes	No	No	GAFF	ChelpG	1.57	2.37
SPC/E	BLYP	TZVP	Yes	No	No	OPLS	ChelpG	1.53	2.39
SPC/E	BLYP	SVP	Yes	Yes	No	GAFF	BCC	2.30	4.41
SPC/E	BLYP	SVP	Yes	Yes	No	OPLS	BCC	2.36	4.51
SPC/E	BLYP	SVP	Yes	Yes	No	GAFF	ChelpG	2.01	2.65
SPC/E	BLYP	SVP	Yes	Yes	No	OPLS	ChelpG	1.96	2.66

SPC/E	HF	SVP	No	No	No	GAFF	BCC	1.78	2.64
SPC/E	HF	SVP	No	No	No	OPLS	BCC	1.81	2.74
SPC/E	HF	SVP	No	No	No	GAFF	ChelpG	1.54	2.35
SPC/E	HF	SVP	No	No	No	OPLS	ChelpG	1.53	2.31
SPC/E	HF	SVP	No	No	Yes	GAFF	BCC	1.99	3.22
SPC/E	HF	SVP	No	No	Yes	OPLS	BCC	2.02	3.32
SPC/E	HF	SVP	No	No	Yes	GAFF	ChelpG	4.42	6.94
SPC/E	HF	SVP	No	No	Yes	OPLS	ChelpG	4.31	6.80
SPC/E	HF-3c	MINI	Yes	N/A	No	GAFF	BCC	2.00	3.62
SPC/E	HF-3c	MINI	Yes	N/A	No	OPLS	BCC	2.05	3.72
SPC/E	HF-3c	MINI	Yes	N/A	No	GAFF	ChelpG	1.61	2.70
SPC/E	HF-3c	MINI	Yes	N/A	No	OPLS	ChelpG	1.56	2.67
Overall								2.19	3.75

Table 4-5 Average unsigned errors (AUE) and root mean-square deviations (RMSD), both in units of kcal/mol, for different multiscale predictions of the hydration free energies of 700 small molecules in TIP3P water.

Water Model	QM Method	Basis Set	D3	gCP	COSMO	LJ Parameter	Charge	AUE	MSD
TIP3P	OMEGA	N/A	N/A	N/A	N/A	GAFF	BCC	1.68	3.32
TIP3P	OMEGA	N/A	N/A	N/A	N/A	OPLS	BCC	1.68	3.25
TIP3P	OMEGA	N/A	N/A	N/A	N/A	GAFF	ChelpG	1.62	2.43
TIP3P	OMEGA	N/A	N/A	N/A	N/A	OPLS	ChelpG	1.58	2.43
TIP3P	BLYP	TZVP	Yes	Yes	No	GAFF	BCC	2.14	3.86
TIP3P	BLYP	TZVP	Yes	Yes	No	OPLS	BCC	2.18	3.95
TIP3P	BLYP	TZVP	Yes	Yes	No	GAFF	ChelpG	1.69	2.50
TIP3P	BLYP	TZVP	Yes	Yes	No	OPLS	ChelpG	1.66	2.51
TIP3P	BLYP	TZVP	Yes	Yes	Yes	GAFF	BCC	2.27	4.17
TIP3P	BLYP	TZVP	Yes	Yes	Yes	OPLS	BCC	2.31	4.25
TIP3P	BLYP	TZVP	Yes	Yes	Yes	GAFF	ChelpG	3.13	5.45
TIP3P	BLYP	TZVP	Yes	Yes	Yes	OPLS	ChelpG	3.05	5.34
TIP3P	BLYP	TZVP	Yes	No	No	GAFF	BCC	2.13	3.85
TIP3P	BLYP	TZVP	Yes	No	No	OPLS	BCC	2.17	3.93
TIP3P	BLYP	TZVP	Yes	No	No	GAFF	ChelpG	1.69	2.49
TIP3P	BLYP	TZVP	Yes	No	No	OPLS	ChelpG	1.65	2.51
TIP3P	BLYP	SVP	Yes	Yes	No	GAFF	BCC	2.21	3.95
TIP3P	BLYP	SVP	Yes	Yes	No	OPLS	BCC	2.26	4.04
TIP3P	BLYP	SVP	Yes	Yes	No	GAFF	ChelpG	2.24	2.90
TIP3P	BLYP	SVP	Yes	Yes	No	OPLS	ChelpG	2.21	2.90

TIP3P	HF	SVP	No	No	No	GAFF	BCC	1.82	2.63
TIP3P	HF	SVP	No	No	No	OPLS	BCC	1.83	2.72
TIP3P	HF	SVP	No	No	No	GAFF	ChelpG	1.37	2.15
TIP3P	**HF**	**SVP**	**No**	**No**	**No**	**OPLS**	**ChelpG**	**1.35**	**2.13**
TIP3P	HF	SVP	No	No	Yes	GAFF	BCC	1.91	3.00
TIP3P	HF	SVP	No	No	Yes	OPLS	BCC	1.93	3.09
TIP3P	HF	SVP	No	No	Yes	GAFF	ChelpG	3.71	5.70
TIP3P	HF	SVP	No	No	Yes	OPLS	ChelpG	3.63	5.58
TIP3P	HF-3c	MINI	Yes	N/A	No	GAFF	BCC	1.98	3.40
TIP3P	HF-3c	MINI	Yes	N/A	No	OPLS	BCC	2.02	3.49
TIP3P	HF-3c	MINI	Yes	N/A	No	GAFF	ChelpG	1.72	2.71
TIP3P	HF-3c	MINI	Yes	N/A	No	OPLS	ChelpG	1.66	2.68
Overall								2.08	3.42

References

(1) Chipot, C.; Pohorille, A. *Free energy calculations : theory and applications in chemistry and biology*; Study ed.; Springer: New York, 2007.

(2) *Molecular Theory of Solvation* Hirata, F., Ed.; Kluwer Academic Publishers: Dordrecht, 2003.

(3) Roux, B.; Simonson, T. *Biophys Chem* **1999**, *78*, 1.

(4) Guo, Z. J.; Li, B.; Dzubiella, J.; Cheng, L. T.; McCammon, J. A.; Che, J. W. *J Chem Theory Comput* **2013**, *9*, 1778.

(5) Boyer, R. D.; Bryan, R. L. *The Journal of Physical Chemistry B* **2012**, *116*, 3772.

(6) Nakamura, I.; Shi, A.-C.; Wang, Z.-G. *Physical Review Letters* **2012**, *109*, 257802.

(7) Fries, B. P. H. *Molecular Physics* **1997**, *90*, 841.

(8) Richardi, J.; Fries, P. H.; Krienke, H. *Journal of Chemical Physics* **1998**, *108*, 4079.

(9) Sindhikara, D. J.; Hirata, F. *J Phys Chem B* **2013**, *117*, 6718.

(10) Sergiievskyi, V. P.; Fedorov, M. V. *J Chem Theory Comput* **2012**, *8*, 2062.

(11) Palmer, D. S.; Frolov, A. I.; Ratkova, E. L.; Fedorov, M. V. *Mol Pharmaceut* **2011**, *8*, 1423.

(12) Chuev, G. N.; Valiev, M.; Fedorov, M. V. *J Chem Theory Comput* **2012**, *8*, 1246.

(13) Chuev, G. N.; Chiodo, S.; Erofeeva, S. E.; Fedorov, M. V.; Russo, N.; Sicilia, E. *Chem Phys Lett* **2006**, *418*, 485.

(14) Liu, Y.; Zhao, S.; Wu, J. *Journal of Chemical Theory and Computation* **2013**, *9*, 1896.

(15) Zhao, S.; Jin, Z.; Wu, J. *Journal of Physical Chemistry B* **2011**, *115*, 15445.

(16) Levesque, M.; Vuilleumier, R.; Borgis, D. *Journal of Chemical Physics* **2012**, *137*, 034115.

(17) Masella, M.; Borgis, D.; Cuniasse, P. *J Comput Chem* **2013**, *34*, 1112.

(18) Jeanmairet, G.; Levesque, M.; Vuilleumier, R.; Borgis, D. *J Phys Chem Lett* **2013**, *4*, 619.

(19) Duffy, E. M.; Jorgensen, W. L. *Journal of the American Chemical Society* **2000**, *122*, 2878.

(20) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. *Journal of Chemical Theory and Computation* **2009**, *5*, 350.

(21) Chothia, C. H. *Nature* **1974**, *248*, 338.

(22) Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. *J Chem Theory Comput* **2006**, *2*, 420.

(23) Mobley, D. L.; Dill, K. A.; Chodera, J. D. *J Phys Chem B* **2008**, *112*, 938.

(24) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J Comput Chem* **2005**, *26*, 1668.

(25) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *Journal of Chemical Physics* **1983**, *79*, 926.

(26) Yang, L.; Ahmed, A.; Sandler, S. I. *J Comput Chem* **2013**, *34*, 284.

(27) Ellingson, B. A.; Geballe, M. T.; Wlodek, S.; Bayly, C. I.; Skillman, A. G.; Nicholls, A. *J Comput Aided Mol Des* **2014**.

(28) G. D. Hawkins, D. J. G., G. C. Lynch, ; C. C. Chambers, I. R., J. W. Storer, J. Li, J. D. Thompson, P. ; Winget, B. J. L., D. Rinaldi, D. A. Liotard, C. J. Cramer, and D. G. Truhlar; University of Minnesota, Minneapolis: 2003.

(29) Cramer, C. J.; Truhlar, D. G. *Chemical Reviews* **1999**, *99*, 2161.

(30) Wu, J.; Li, Z. *Annu. Rev. Phys. Chem.* **2007**, *58*, 85.

(31) Singer, S. J.; Chandler, D. *Molecular physics* **1985**, *55*, 621.

(32) Kinoshita, M.; Okamoto, Y.; Hirata, F. *Journal of computational chemistry* **1997**, *18*, 1320.

(33) Mobley, D. L.; Wymer, K. L.; Lin, N. M. *J Comput Aided Mol Des* **2014**.

(34) Guthrie, J. P. *J Comput Aided Mol Des* **2014**.

(35) Rosenfeld, Y. *Journal of Chemical Physics* **1993**, *98*, 8126.

(36) Zhao, S.; Wu, J. *Molecular Physics* **2011**, *109*, 2553.

(37) Yu, Y.-X.; Wu, J. *J. Chem. Phys.* **2002**, *117*, 10156.

(38) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926.

(39) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *Journal of Computational Chemistry* **2004**, *25*, 1157.

(40) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *Journal of the American Chemical Society* **1996**, *118*, 11225.

(41) Jakalian, A.; Jack, D. B.; Bayly, C. I. *Journal of Computational Chemistry* **2002**, *23*, 1623.

(42) Breneman, C. M.; Wiberg, K. B. *Journal of Computational Chemistry* **1990**, *11*, 361.

(43) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187.

(44) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *Journal of Molecular Graphics and Modelling* **2006**, *25*, 247.

(45) M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox; Gaussian, Inc.: Wallingford CT, 2009.

(46) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *Journal of Computational Chemistry* **2005**, *26*, 1701.

(47) Liu, Y.; Fu, J.; Wu, J. *The Journal of Physical Chemistry Letters* **2013**, *4*, 3687.

(48) Scalmani, G.; Frisch, M. J. *The Journal of Chemical Physics* **2010**, *132*.

(49) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *The Journal of Physical Chemistry B* **2009**, *113*, 4538.

(50) Mobley, D. L.; Dumont, É.; Chodera, J. D.; Dill, K. A. *The Journal of Physical Chemistry B* **2007**, *111*, 2242.

(51) Li, L.; Fennell, C. J.; Dill, K. A. *The Journal of Physical Chemistry B* **2013**.

(52) Sandberg, L. *J Comput Aided Mol Des* **2014**.

(53) Grotendorst, J.; Attig, N.; Blügel, S.; Marx, D.; Institute for Advanced Simulation: Forschungszentrum Jülich, Germany, 2009; Vol. 42.

(54) Cramer, C. J.; Truhlar, D. G. *Accounts of Chemical Research* **2008**, *41*, 760.

(55) Zou, X. Q.; Sun, Y. X.; Kuntz, I. D. *Journal of the American Chemical Society* **1999**, *121*, 8033.

(56) Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. *Biochemistry* **1981**, *20*, 849.

(57) Palmer, D. S.; Sergiievskyi, V. P.; Jensen, F.; Fedorov, M. V. *Journal of Chemical Physics* **2010**, *133*, 044104.

(58) Chandler, D. *Nature* **2005**, *437*, 640.

(59) Hirata, F. *Molecular theory of solvation*; Kluwer Academic Publishers: Boston, 2003.

(60) Chang, J.; Lenhoff, A. M.; Sandler, S. I. *Journal of Physical Chemistry B* **2007**, *111*, 2098.

(61) Freedman, H.; Le, L.; Tuszynski, J. A.; Truong, T. N. *Journal of Physical Chemistry B* **2008**, *112*, 2340.

(62) Freedman, H.; Truong, T. N. *Chemical Physics Letters* **2003**, *381*, 362.

(63) Paluch, A. S.; Shah, J. K.; Maginn, E. J. *Journal of Chemical Theory and Computation* **2011**, *7*, 1394.

(64) Jorgensen, W. L.; Ravimohan, C. *Journal of Chemical Physics* **1985**, *83*, 3050.

(65) Marcus, Y. *Chem Rev* **2009**, *109*, 1346.

(66) Ball, P. *Chem Rev* **2008**, *108*, 74.

(67) Senn, H. M.; Thiel, W. *QM/MM Methods for Biological Systems*; Springer-Verlag: Berlin Heidelberg, 2007; Vol. 268.

(68) Voth, G. A.; Hochstrasser, R. M. *J Phys Chem-Us* **1996**, *100*, 13034.

(69) Ren, P. Y.; Chun, J. H.; Thomas, D. G.; Schnieders, M. J.; Marucho, M.; Zhang, J. J.; Baker, N. A. *Q Rev Biophys* **2012**, *45*, 427.

(70) Kinoshita, M. *Front Biosci-Landmrk* **2009**, *14*, 3419.

(71) Klebl, B.; Müller, G.; Hamacher, M. *Protein kinases as drug targets*; Wiley-VCH: Weinheim, 2011.

(72) Truhlar, D. G. *Rational drug design*; Springer: New York, 1999.

(73) Zheng, Y. *Rational drug design : methods and protocols*; Humana Press ; Springer: New York, 2012.

(74) Popelier, P. *Curr Top Med Chem* **2012**, *12*, 1924.

(75) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L. *Journal of Computational Chemistry* **2004**, *25*, 265.

(76) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2007**, *3*, 2011.

(77) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2008**, *4*, 877.

(78) Tomasi, J. *Theor Chem Acc* **2004**, *112*, 184.

(79) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem Rev* **2005**, *105*, 2999.

(80) Duffy, E. M.; Jorgensen, W. L. *Journal of the American Chemical Society* **2000**, *122*, 2878.

(81) Jang, S.; Kim, E.; Pak, Y. *Proteins-Structure Function and Bioinformatics* **2006**, *62*, 663.

(82) Shirts, M. R.; Pande, V. S. *Journal of Chemical Physics* **2005**, *122*, 134508.

(83) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *Journal of Chemical Physics* **2003**, *119*, 5740.

(84) Straatsma, T. P.; Berendsen, H. J. C. *Journal of Chemical Physics* **1988**, *89*, 5876.

(85) Gray, C. G.; Gubbins, K. E. *Theory of Molecular Fluids: Fundamentals* Oxford University Press: New York, 1985; Vol. 1.

(86) Chandler, D. *Annual Review of Physical Chemistry* **1978**, *29*, 441.

(87) Lue, L.; Blankschtein, D. *J Phys Chem-Us* **1992**, *96*, 8582.

(88) Lue, L.; Blankschtein, D. *Journal of Chemical Physics* **1995**, *102*, 5427.

(89) Rossky, P. J. *Annual Review of Physical Chemistry* **1985**, *36*, 321.

(90) Brooks, C. L.; Karplus, M.; Pettitt, B. M. *Proteins : a theoretical perspective of dynamics, structure, and thermodynamics*; J. Wiley: New York, 1988.

(91) Pratt, L. R.; Pohorille, A. *Chem Rev* **2002**, *102*, 2671.

(92) Lee, M. S.; Salsbury, F. R.; Olson, M. A. *Journal of Computational Chemistry* **2004**, *25*, 1967.

(93) Basdevant, N.; Ha-Duong, T.; Borgis, D. *Journal of Chemical Theory and Computation* **2006**, *2*, 1646.

(94) Bryk, P.; MacDowell, L. G. *Journal of Chemical Physics* **2011**, *135*.

(95) Schmidt, K. F.; Kast, S. M. *Journal of Physical Chemistry B* **2002**, *106*, 6289.

(96) Hassanali, A. A.; Cuny, J.; Verdolino, V.; Parrinello, M. *Philos T R Soc A* **2014**, *372*.

(97) Li, G. H.; Zhang, X. D.; Cui, Q. *Journal of Physical Chemistry B* **2003**, *107*, 8643.

(98) Hu, H.; Yang, W. T. *Annual Review of Physical Chemistry* **2008**, *59*, 573.

(99) Lin, H.; Truhlar, D. G. *Theor Chem Acc* **2007**, *117*, 185.

(100) Schwarz, K. A.; Sundararaman, R.; Letchworth-Weaver, K.; Arias, T. A.; Hennig, R. G. *Phys Rev B* **2012**, *85*, Artn 201102.

(101) Sundararaman, R.; Letchworth-Weaver, K.; Arias, T. A. *Journal of Chemical Physics* **2012**, *137*.

(102) Cramer, C. J.; Truhlar, D. G. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers, New York: 1995; Vol. 6, p 1.

(103) Sandberg, L.; Edholm, O. *Journal of Chemical Physics* **2002**, *116*, 2936.

(104) Jha, A. K.; Freed, K. F. *Journal of Chemical Physics* **2008**, *128*, 034501.

(105) Klamt, A.; Schuurmann, G. *Journal of the Chemical Society, Perkin Transactions 2* **1993**, 799.

(106) Phillips, K. L.; Sandler, S. I.; Greene, R. W.; Di Toro, D. M. *Environmental science & technology* **2008**, *42*, 8412.

(107) Klamt, A.; Eckert, F.; Diedenhofen, M. *Journal of Physical Chemistry B* **2009**, *113*, 4508.

(108) Simonin, J. P.; Bernard, O.; Blum, L. *Journal of Physical Chemistry B* **1998**, *102*, 4411.

(109) Simonin, J. P.; Blum, L.; Turq, P. *J Phys Chem-Us* **1996**, *100*, 7704.

(110) Pitzer, K. S. *Activity coefficients in electrolyte solutions*; 2nd ed.; CRC Press: Boca Raton, 1991.

(111) Sitkoff, D.; Sharp, K. A.; Honig, B. *J Phys Chem-Us* **1994**, *98*, 1978.

(112) Cramer, C. J.; Truhlar, D. G. *Chem Rev* **1999**, *99*, 2161.

(113) Bashford, D.; Case, D. A. *Annual Review of Physical Chemistry* **2000**, *51*, 129.

(114) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. *Journal of Chemical Physics* **2002**, *117*, 43.

(115) Duignan, T. T.; Parsons, D. F.; Ninham, B. W. *Journal of Physical Chemistry B* **2013**, *117*, 9412.

(116) Baron, R.; McCammon, J. A. *Annual Review of Physical Chemistry* **2013**, *64*, 151.

(117) Shirts, M. R.; Mobley, D. L. *Methods in molecular biology* **2013**, *924*, 271.

(118) Shirts, M. R.; Mobley, D. L.; Chodera, J. D. *Annual Reports in Computational Chemistry* **2007**, *3*, 41.

(119) Kollman, P. *Chemical Reviews* **1993**, *93*, 2395.

(120) Mobley, D. L.; Klimovich, P. V. *The Journal of chemical physics* **2012**, *137*, 230901.

(121) de Ruiter, A.; Oostenbrink, C. *Curr Opin Chem Biol* **2011**, *15*, 547.

(122) Brandsdal, B. O.; Osterberg, F.; Almlof, M.; Feierberg, I.; Luzhkov, V. B.; Aqvist, J. *Advances in protein chemistry* **2003**, *66*, 123.

(123) Lin, Y.-L.; Aleksandrov, A.; Simonson, T.; Roux, B. *J Chem Theory Comput* **2014**, *in press*.

(124) Beierlein, F. R.; Michel, J.; Essex, J. W. *The journal of physical chemistry. B* **2011**, *115*, 4911.

(125) Singh, N.; Warshel, A. *Proteins* **2010**, *78*, 1705.

(126) Gilson, M. K.; Zhou, H. X. *Annual review of biophysics and biomolecular structure* **2007**, *36*, 21.

(127) Lv, C.; Zheng, L.; Yang, W. *The Journal of chemical physics* **2012**, *136*, 044103.

(128) Lee, S.; Chen, M.; Yang, W.; Richards, N. G. *J Am Chem Soc* **2010**, *132*, 7252.

(129) Zheng, L.; Chen, M.; Yang, W. *The Journal of chemical physics* **2009**, *130*, 234105.

(130) Zheng, L.; Chen, M.; Yang, W. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 20227.

(131) Hansen, J. P.; McDonald, I. R. *Theory of Simple Liquids*; Third ed.; Academic Press: London, 2006.

(132) Wu, J. Z.; Li, Z. D. *Annual Review of Physical Chemistry* **2007**, *58*, 85.

(133) Borgis, D.; Gendre, L.; Ramirez, R. *Journal of Physical Chemistry B* **2012**, *116*, 2504.

(134) Kovalenko, A. *Pure Appl Chem* **2013**, *85*, 159.

(135) Kovalenko, A.; Kobryn, A. E.; Gusarov, S.; Lyubimova, O.; Liu, X. J.; Blinov, N.; Yoshida, M. *Soft Matter* **2012**, *8*, 1508.

(136) Gendre, L.; Ramirez, R.; Borgis, D. *Chemical Physics Letters* **2009**, *474*, 366.

(137) Ramirez, R.; Borgis, D. *Journal of Physical Chemistry B* **2005**, *109*, 6754.

(138) Liu, Y.; Wu, J. Z. *Journal of Chemical Physics* **2013**, *139*, 041103.

(139) Sergiievskyi, V. P.; Jeanmairet, G.; Levesque, M.; Borgis, D. *J Phys Chem Lett* **2014**, *5*, 1935.

(140) Fu, J.; Liu, Y.; Wu, J. *Journal of Computer-Aided Molecular Design* **2014**, *28*, 299.

(141) Rosenfeld, Y. *Phys. Rev. Lett.* **1989**, *63*, 980.

(142) Yu, Y. X.; Wu, J. Z. *Journal of Chemical Physics* **2002**, *117*, 10156.

(143) Shahidi, F. *J Chem Soc Farad T 1* **1981**, *77*, 1511.

(144) Edward, J. T.; Farrell, P. G.; Shahidi, F. *J Chem Soc Farad T 1* **1977**, *73*, 705.

(145) Allinger, N. L. *Journal of the American Chemical Society* **1977**, *99*, 8127.

(146) Hagler, A.; Huler, E.; Lifson, S. *Journal of the American Chemical Society* **1974**, *96*, 5319.

(147) Schrier, E. E.; Pottle, M.; Scheraga, H. A. *Journal of the American Chemical Society* **1964**, *86*, 3444.

(148) Lifson, S.; Warshel, A. *The Journal of Chemical Physics* **1968**, *49*, 5116.

(149) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J Phys Chem-Us* **1990**, *94*, 8897.

(150) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *Journal of the American Chemical Society* **1992**, *114*, 10024.

(151) Düren, T.; Sarkisov, L.; Yaghi, O. M.; Snurr, R. Q. *Langmuir* **2004**, *20*, 2683.

(152) Walton, K. S.; Snurr, R. Q. *J Am Chem Soc* **2007**, *129*, 8552.

(153) Challa, S. R.; Sholl, D. S.; Johnson, J. K. *The Journal of Chemical Physics* **2002**, *116*, 814.

(154) Keskin, S.; Liu, J.; Rankin, R. B.; Johnson, J. K.; Sholl, D. S. *Industrial & Engineering Chemistry Research* **2009**, *48*, 2355.

(155) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *Journal of the American Chemical Society* **1995**, *117*, 5179.

(156) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187
.

(157) Martin, M. G.; Siepmann, J. I. *The Journal of Physical Chemistry B* **1998**, *102*, 2569.

(158) Rai, N.; Siepmann, J. I. *Journal of Physical Chemistry B* **2013**, *117*, 273.

(159) Sun, H.; Mumby, S. J.; Maple, J. R.; Hagler, A. T. *Journal of the American Chemical Society* **1994**, *116*, 2978.

(160) Sun, H. *The Journal of Physical Chemistry B* **1998**, *102*, 7338.

(161) Yu, Y. X.; Wu, J. Z. *Journal of Chemical Physics* **2002**, *117*, 2368.

(162) Seok, C.; Oxtoby, D. W. *Journal of Chemical Physics* **1998**, *109*, 7982.

(163) Chandler, D.; Mccoy, J. D.; Singer, S. J. *Journal of Chemical Physics* **1986**, *85*, 5971.

(164) Zhao, S. L.; Liu, Y.; Liu, H. L.; Wu, J. Z. *Journal of Chemical Physics* **2013**, *139*, Artn 064509.

(165) Rosenfeld, Y. *Journal of Chemical Physics* **1993**, *98*, 8126.

(166) Rosenfeld, Y. *J. Phys.-Condes. Matter* **2002**, *14*, 9141.

(167) Rosenfeld, Y. *Journal of Physics-Condensed Matter* **2002**, *14*, 9141.

(168) Rosenfeld, Y. *Journal of Chemical Physics* **1993**, *98*, 8126.

(169) Yu, Y. X.; Wu, J. Z. *Journal of Chemical Physics* **2002**, *116*, 7094.

(170) Yu, Y. X.; Wu, J. Z.; Xin, Y. X.; Gao, G. H. *Journal of Chemical Physics* **2004**, *121*, 1535.

(171) Marenich, A. V.; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. *University of Minnesota, Minneapolis* **2009**.

(172) H. J. C. Berendsen; J. P. M. Postma; W. F. van Gunsteren; Hermans, J. *Intermolecular Forces*; Reidel, Dordrecht, 1981.

(173) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *The Journal of Physical Chemistry* **1987**, *91*, 6269.

(174) Kruse, H.; Grimme, S. *The Journal of chemical physics* **2012**, *136*, 154101.

(175) Goerigk, L.; Grimme, S. *Physical Chemistry Chemical Physics* **2011**, *13*, 6670.

(176) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *Journal of computational chemistry* **2003**, *24*, 669.

(177) Neese, F. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 73.

(178) Sure, R.; Grimme, S. *Journal of Computational Chemistry* **2013**, *34*, 1672.

(179) Ren, P. Y.; Ponder, J. W. *Journal of Physical Chemistry B* **2003**, *107*, 5933.

(180) Warshel, A.; Kato, M.; Pisliakov, A. V. *Journal of Chemical Theory and Computation* **2007**, *3*, 2034.

(181) Patel, S.; Brooks, C. L. *Mol Simulat* **2006**, *32*, 231.

(182) Bauer, B. A.; Patel, S. *Theor Chem Acc* **2012**, *131*.

(183) Ren, P. Y.; Wu, C. J.; Ponder, J. W. *Journal of Chemical Theory and Computation* **2011**, *7*, 3143.

(184) Schnieders, M. J.; Baltrusaitis, J.; Shi, Y.; Chattree, G.; Zheng, L.; Yang, W.; Ren, P. *J Chem Theory Comput* **2012**, *8*, 1721.

(185) Bowman, J. M. *Science* **2008**, *319*, 40.

(186) Pisana, S.; Lazzeri, M.; Casiraghi, C.; Novoselov, K. S.; Geim, A. K.; Ferrari, A. C.; Mauri, F. *Nature Materials* **2007**, *6*, 198.

(187) Zobov, N. F.; Polyansky, O. L.; LeSueur, C. R.; Tennyson, J. *Chemical Physics Letters* **1996**, *260*, 381.

(188) Schwenke, D. W. *J Phys Chem A* **2001**, *105*, 2352.

(189) Kong, L. Z.; Wu, X. F.; Car, R. *Phys Rev B* **2012**, *86*, Artn 134203.

(190) Gonzalez, B. S.; Noya, E. G.; Vega, C.; Sese, L. M. *Journal of Physical Chemistry B* **2010**, *114*, 2484.

(191) Ando, K. *Journal of Chemical Physics* **2001**, *114*, 9470.

(192) Noya, E. G.; Vega, C.; Sese, L. M.; Ramirez, R. *Journal of Chemical Physics* **2009**, *131*, 124518.

(193) Vega, C.; Abascal, J. L. *Physical Chemistry Chemical Physics* **2011**, *13*, 19663.

(194) Patchkovskii, S.; Tse, J. S.; Yurchenko, S. N.; Zhechkov, L.; Heine, T.; Seifert, G. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 10439.

(195) Hammes-Schiffer, S. *Energ Environ Sci* **2012**, *5*, 7696.

(196) Geerlings, P.; De Proft, F.; Langenaeker, W. *Chem Rev* **2003**, *103*, 1793.

(197) Zhao, Y.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2005**, *1*, 415.

(198) Marques, M. A. L.; Gross, E. K. U. *Annual Review of Physical Chemistry* **2004**, *55*, 427.

(199) Drut, J. E.; Furnstahl, R. J.; Platter, L. *Prog Part Nucl Phys* **2010**, *64*, 120.

(200) Baer, R.; Livshits, E.; Salzner, U. *Annu Rev Phys Chem* **2010**, *61*, 85.
(201) March, N. H. *Phys Chem Liq* **2010**, *48*, 141.

(202) Grafenstein, J.; Cremer, D. *Journal of Chemical Physics* **2009**, *130*.

(203) Wu, J. Z. In *Molecular Therodynamics of Complex Systems*; Lu, X., Hu, Y., Eds. 2009; Vol. 131, p 1.

(204) Ho, G. S.; Huang, C.; Carter, E. A. *Curr Opin Solid St M* **2007**, *11*, 57.

(205) Hafner, J. *Journal of Computational Chemistry* **2008**, *29*, 2044.

(206) Wu, J. Z.; Li, Z. D. *Annual Review of Physical Chemistry* **2007**, *58*, 85.

(207) Botti, S.; Schindlmayr, A.; Del Sole, R.; Reining, L. *Rep Prog Phys* **2007**, *70*, 357.

(208) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864.

(209) Mermin, N. D. *Physical Review* **1965**, *127*, A1441.

(210) Ebner, C.; Saam, W. F.; Stroud, D. *Physical Review A* **1976**, *14*, 2264.

(211) Evans, R. *Advances in Physics* **1979**, *28*, 143.

(212) Capitani, J. F.; Nalewajski, R. F.; Parr, R. G. *Journal of Chemical Physics* **1982**, *76*, 568.

(213) Gidopoulos, N. *Phys Rev B* **1998**, *57*, 2146.

(214) Kreibich, T.; Gross, E. K. U. *Physical Review Letters* **2001**, *86*, 2984.

(215) Chakraborty, A.; Pak, M. V.; Hammes-Schiffer, S. *Physical Review Letters* **2008**, *101*, Artn 153001.

(216) Kohn, W.; Sham, L. J. *Physical Review* **1965**, *140*, A1133.

(217) Burke, K. *Journal of Chemical Physics* **2012**, *136*.

(218) Cohen, A. J.; Mori-Sanchez, P.; Yang, W. T. *Chem Rev* **2012**, *112*, 289.

(219) Parr, R. G.; Yang, W. *Density-functional theory of atoms and molecules*; Oxford University Press New York, 1989.

Chapter 5. Conclusions

In this dissertation, I have developed systematic theoretical approaches to do large-scale screening of chemicals and materials, i.e. hydration free energy of organic compounds and gas storage capabilities in nanoporous materials. The classical density functional theory provides one of the most versatile and powerful modelling tools for material design purpose.

For the computer-aided materials design aspect, I have calibrated four representative versions of classical density functional theory (CDFT) for large-scale screening of the gas storage capabilities in nanoporous materials. Significant differences are observed for their application to realistic pore materials because of the highly inhomogeneous nature of gas molecules distribution. By extensive comparison of CDFT predictions and simulation data in 1,200 MOF materials, we find that the performance of those methods vary with specific gas compounds and testing conditions. For H_2 adsorption studying, the density expansion method (FMSA) performs best under the DOE target condition (243 K and 100 bar), while at low temperature (77 K) reliable predictions only could be achieved with the two versions of weighted density approximation (WDA) methods. For methane storage studying under ARPA-E targeted condition (298 K and 35 bar), mean-field approximation (MFA) is the best candidate for screening purpose. We also demonstrate that, while none of the testing materials satisfies the ARPA-E MOVE target, it could be achieved by changing the initial storage and final release thermodynamic conditions, for example raising the release temperature is raised to 358 K and reducing the initial storage temperature to 233 K, with relative mild pressure requirements.

Unlike common thought that the surface area is a good measurement of the potential gas storage capabilities of nanoporous materials, we found this descriptor is not useful in the delivery amount studying, no clearly correlation could be found. For the natural gas vehicle research, it is found in our studying, the total weight adsorption amount may be a good indicator to find promising materials. Besides the equilibrium thermodynamic properties, by utilizing the new proposed excess entropy scaling method developed in our group, we could also predict the gas transport properties in nanoporous materials by the same one-time calculations, instead of trajectory average method from molecular dynamic simulations, by only one MD data, we could quantitate capture the simulation results . All those make us could easily evaluate the materials real gas storage performance for specific targets.

In my work, I already show the efficiency and accuracy of CDFT for gas storage studying, however here it's still limited to pure simple LJ fluids, while in real chemical engineering field, what we are more interested are polyatomic molecules, like n-alkanes and charged species, like carbon dioxide, or even more complicated mixture system. It's nearly impossible to study such kinds of system in a large scale by classical simulation methods, due to the explicitly consideration of huge amount of particles in the system. While in CDFT aspect, we only need to solve series of mathematical equations. Because the reliable CDFT formulations have been well established for many cases mentioned above, I expect that the computational procedure developed in this work can be similarly applicable.

For the hydration free energy prediction part, I have demonstrated that CDFT provides a powerful alternative to conventional simulation methods for high throughput predictions of the water solvation free energies of small organic molecules. I have tested various combinations of multiscale methods for predicting the hydration free energies of 700 small molecules at the ambient condition, i.e., obtaining atomic charges and solute configurations from first principle calculations and then choosing van der Waals parameters from various organic force-fields. Among all those combinations, I find that the best combination is provided by calculations based on the TIP3P model for liquid water, with the solute configurations generated from HF/SVP optimization, the OPLS LJ parameters, and the ChelpG partial charges. For this particular combination, MDFT predict the hydration free energies of 700 small molecules with AUE = 1.35 kcal/mol and RMSD = 2.13 kcal/mol in comparison with experimental data. The theoretical results are most accurate for solutes with slightly positive hydration free energies but become less precise for large hydrophilic solutes, where the configurations change a lot from vacuum to water environment. This could be further improved by doing ensemble average based on several solute configurations. For small organic compounds, this is usually not a big issue, which makes our methods could be a useful tool in developing more accurate force fields. Compared to traditional classical simulation methods, our method also considers water solvent explicitly, but with a much faster calculation speed, over one magnitude, i.e. tens of minutes compared to hundreds of hours.

The new MDFT methods could be a value tools to pharmaceutical industry where solubility is one of the key problems in drug design, considering its efficiency and accuracy

compared to common quantitative structure activity relationships (QSPR) or general solubility equations (GSE) methods, which highly depend on the training sets quality. This also give us a hint that people in simulation community often come from chemistry or biology field, who haven't realized many important thermodynamic properties could be derived from modern equation of state, which could obtained parameters using solvation free energy data. Combing the classical density functional theory, molecular mechanics and macroscopic thermodynamics together will provide our chemical engineers an invaluable tool.

Currently our MDFT calculations are based on the classical pairwise additive potentials that do not account for the polarization effects, which plays a very important role in many solvent-mediated phenomena. Even though the importance of non-additive interactions in liquid systems has been well recognized, including this effect in classical simulations is not trivial, due to the drastically increasing computational costs. The computational efficiency of MDFT could make it as an outstanding candidate for development and applications of polarizable force fields. A much further step is combine the CDFT and QDFT together, since from a theoretical perspective, both of them use one-body density profiles as fundamental variables to describe the properties of many-body systems within the same mathematical framework. One day, I think the two DFT brothers could finally joint together, making us could study the system under one unified framework, without worrying about the boundary issue between classical and quantum regions, also no empirical parameters bothers.