He inclusion in ice-like and clathrate-like frameworks: a benchmark quantum chemistry study of guest-host interactions

Raquel Yanes-Rodríguez[§], Daniel J. Arismendi-Arrieta^{§,‡}, and Rita Prosmiti^{§*}

[§] *Institute of Fundamental Physics (IFF-CSIC), CSIC, Serrano 123, 28006 Madrid, Spain*

[‡] *Donostia International Physics Center (DIPC), Paseo Manuel de Lardizabal 4, 20018
Donostia-San Sebastián, Gipuzkoa, Spain*

E-mail: rita@iff.csic.es; Phone: +34-91-5616800Ext.941131

Abstract

Energetics and structural properties of selected type and size He@hydrate frameworks, *e.g.* from regular structured ice-channels to clathrate-like cages, are presented from first-principles quantum chemistry methods. The scarcity of information on He@hydrates make such complexes challenging targets, while their computational study entails an interesting and arduous task. Some of them have been synthesized in the laboratory, that motivates further investigations on their stability. Hence, the main focus is to examine the performance and accuracy of different wavefunction-based electronic structure methods, such as MP2, CCSD(T), their explicitly-correlated (F12) and domain-based local pair-natural orbital (DLPNO) analogs, as well as modern and conventional density functional theory (DFT) approaches, and analytical model potentials available. Different structures are considered, starting from the “simplest system” formed by a noble gas atom (such as He) and one water molecule, followed by the study of the “fundamental units” present in all ice-like and clathrate-like frameworks (such as pentamers and hexamers) and finally, the description of interactions in the “building blocks” of 3D ice channels (*e.g.* horizontal and perpendicular ice II and I_h) and clathrate-like cages, such as the 5¹² present in the most common sI, sII and sH clathrate-hydrate structures. The idea is to provide well-converged DLPNO-CCSD(T) and DFMP2/CBS reference datasets, that in turn used to validate how DFT functionals (in total 29 approaches from GGA, meta-GGA, to hybrid and range-separated functionals, including dispersion correction treatments, were checked), and analytical semiempirical /*ab initio*-based potentials perform compared with high-level alternatives. Within all tested approaches, those best-performing were identified and classified. Most of the DFT/DFT-D functionals, as well as available analytical pairwise model potentials face difficulties in describing both hydrogen-bonded water framework and the dispersion bound He-water interaction. Including dispersion corrections yields to an overall well-balance performance for the LC ω PBE-D3BJ and PBE0-D4 functionals. Such benchmark datasets can benefit research into the development of new chem-informatics models, as can serve to guide and cross check methodologies, lending

increased predicted power to future molecular simulations for investigating the role of structures and phase-transitions from nanoscale clusters to macroscopic crystalline structures.

1. INTRODUCTION

Water molecules can easily form complexes either between them or with other compounds, while noble gases are more unlikely to form bonds with nearly any other molecular system. In this sense, noble gas hydrates are of particular interest, due to their ability to keep trapped such inert gases and thus to produce compounds. In order to identify them, it is necessary to understand the nature of bonding of any water-lattice (host) with a noble gas atom (guest), and thus the balance between intermolecular hydrogen-bonded water network and weakly van der Waals (vdW) forces.

Although, noble gas hydrates have been investigated, only few studies are available in the literature on noble gas-water host network interactions.¹⁻¹¹ The crystalline clathrate hydrate and the ice-lattice unit cells are by themselves aggregates of hydrogen-bonded H_2O molecules forming closed cavities (cages) or open polymorphs of ice, respectively, and some of them are stabilized by the presence of the guest gas.^{12,13} It is important to note that there are several analogies and differences between the physical properties of filled ice and those of clathrates.¹⁴ Therefore, fundamental insights into predictive nature of realistic models which describe guest-water interactions of prototypes, may made easier the study of such compounds.

The research activities on noble gas hydrates got the attention after the discovery of Ar hydrate.¹⁵ Since then, a number of spectroscopic and diffraction studies have been conducted to explore at which conditions (pressure, temperature, composition of the aqueous solution) such gas hydrates may be formed or decomposed,^{3,16-18} while the most recent research on the lighter He and Ne clathrates is related with the stabilization of new low-density ice polymorphs,^{14,19-21} such as ice XVI^{22,23} or cubic structure III.²⁴ Although He and Ne atoms were initially considered to be too small for stabilizing clathrate cages, such clathrates have also been observed recently.^{22,23} The Ne atom was able to form a sII clathrate structure, with the possibility of easily traveling between the large $5^{12}6^4$ cages passing through the six membered rings of water molecules, which facilitated their removal in order to isolate

the guest-free cage of ice XVI.²² In the case of He the process was the contrary. The empty XVI structure was refilled with He atoms until the long-sought He-clathrate formed.²³ For the heavier noble gases (Xe, Kr and Ar), it has been found that under pressure they form clathrate hydrates of sI, sII and sH types, although investigations by different set of experiments were not always in accord on such crystal structure transformations.^{3,11}

On the one hand, theoretical and computational approaches could provide useful information, that it could possibly settle disagreements, and/or on the other hand, it could bring new insights into less-explored phase diagram regimes stimulating future experiments. In the present work, the guest atom of noble gas-water complexes is Helium, the lightest noble gas atom; its principal characteristic is the low chemical reactivity due to its high stable electronic configuration. The He-hydrate system has been identified by Londono *et al.*,^{25,26} who have showed that hydrates of He in ice-II structure can be obtained by applying a pressure of 0.28-0.48 GPa on helium gas in water, and since then, it still remains relatively unexplored. In particular, the interactions that govern their structure and dynamics on the microscopic level is needed for explaining a variety of physico-chemical processes.

Numerous quantum chemistry electronic structure methods, ranging from the most accurate wavefunction-based (WF) methods to the computationally viable DFT ones, provide independent and reliable means for describing noncovalent guest-host molecular interactions. Such information is indispensable for unraveling the subtleties of their stability and structural transformations, as well as their storage capacity. In recent years, a tremendous amount of effort has been dedicated to the development of new methodologies/approximations that tend to improve the scalability of the computational cost with respect the size of the molecular system, and benchmark of their performance (*e.g.* see Refs.^{7,27-33} and references therein). In turn, emerging approaches to developing machine-learned potential models offer a promising alternative for the representation/modeling of the molecular interactions, requiring large high-quality reference datasets from electronic structure calculations.³⁴⁻³⁶

Thus, in this study the central idea is to assess the performance and accuracy of different

WF-based and DFT-based electronic structure methods, as well as analytical semiempirical and *ab initio*-based potential models in order to provide reliable description of potential interactions in the He inclusion complexes, such as clathrate-like cavities and filled ice-like channels. Given the delicate balance between the interaction energy components (vdW, hydrogen bonds) present on such systems, the idea is to propose a systematic bottom-up approach, to validate how various density functionals from different families, climbing up to the Jacob’s ladder, with and without dispersion corrections, perform for He-water systems of different types and sizes compared to new reference data from wavefunction-based methods, such as MP2, CCSD(T) and their explicitly-correlated, as well as domain-based local pair-natural orbital approaches depending on the size of the system. Our choice includes the study of the “simplest system” (He atom interacts with a H₂O water-monomer), followed by the “fundamental units” (He atom interacts with (H₂O)₅ water-pentamer and (H₂O)₆ water-hexamer) present in all ices polymorphs and common clathrate structures, up to the “building blocks” corresponding to the close cavity of the pentagonal dodecahedron (5¹² formed by 20 H₂O molecules and common in all sI, sII and sH clathrate structures), as well as the open channels of the perpendicular and horizontal ice II, (formed by 18 and 24 H₂O molecules, respectively), and I_h crystals (formed by 24 H₂O molecules). The study provides well-converged benchmark datasets of energies and He@hydrostructures, that are obtained in a systematic way from high-level quantum chemistry technologies. The accuracy issues to define such reference energies have profound implications on the development of new chem-informatics models based on data-driven techniques or other auxiliary force-fields/potentials for molecular dynamics simulations. Indeed, this work represents the first step towards transferable models based on cluster data, fed with equilibrium and non equilibrium configurations, and trained via active learning algorithms.³⁷

The article is organized as follows: in the “Computational Details” section, we describe the reference set of configuration structures, some pairwise representative potential models, and the electronic structure methods/approaches employed in the study; in the “Results

and Discussion” section, we report and analyze the performance of various WF-based, and DFT-based methods, as well as those of the analytical pairwise semiempirical/*ab initio*-based potential models; and finally, in the “Summary and Conclusions” section we discuss some concluding remarks and future perspectives of the study.

2. COMPUTATIONAL DETAILS

Set of reference structures

In this work we have performed a staggered study on a variety of $\text{He}-(\text{H}_2\text{O})_n$ systems, with $n=1-24$, following a hierarchical structural organization, as shown in Figure 1, from the simplest system to the fundamental units and the building blocks of common water-lattice structures. The He atom interacts with water-monomer, and specific-size and structure water-oligomers, such as the most common structures in the ice polymorphs and sI, sII and sH clathrate-hydrates, *e.g.* water-pentamer, water-hexamer and the 5^{12} cage, respectively. As we will describe later on, by determining interaction forces of the simplest system, pairwise approaches, by summing up the monomer energy interactions, for both fundamental units and building blocks could be employed, while further by taking into account the summation of interaction energies of higher-size clusters, such as pentamers, hexamers, icosamers, *etc*, then higher-order many-body effects could be included in the representation of interaction potentials of larger-size systems. In a similar manner as the many-body expansion approach,^{38–40} one could model large systems in a systematic way by partitioning energies into a hierarchy of decreasingly significant contributions, that it reduces the scaling problem and improves accuracy of electronic structure calculations. Recently, such potential representations have been reported for studying several clathrate hydrate systems.^{41–43}

So, the guest-free 5^{12} cage is a structure composed by 20 water molecules, and its geometry was extracted from the 3D crystalline frameworks of clathrates, as they have been determined in Ref.⁴⁴ satisfying the ice rules, having the lowest potential energy configuration

for the protons, and a net zero dipole moment. In turn, as He hydrates has been reported for the ice I_h and II host lattices,^{3,16,25,26} such ice I_h -like and II-like channels are also considered here. The ice I_h is the most commonly known phase of ice (ordinary ice), which presents a hexagonal structure composed by two kinds of six-membered rings (chair and boat conformation), which are stacked leading to the formation of channels, and we choose a channel along c-axis crystal with 24 water molecules. The ice II contains cavities which are also formed by two six-membered rings, but one of these rings is almost flat and the other is puckered as in ice I_h , and has chair conformation.⁴⁵ Therefore, for the ice II-like we considered both the horizontal (along a-axis of crystal) and perpendicular (along c-axis of crystal) directions, namely II(h) and II(p), respectively. All the 3D-ice structures used in the present study were extracted from Ref.⁴⁶

Figure 1: The set of reference structures considered in this study (see text).

For all of the above mentioned cases, a He atom is added to the specific structure, and

for each of the corresponding $\text{He}-(\text{H}_2\text{O})_n$ system, we have carried out⁴⁷ optimization calculations at CCSD(T)/AVQZ ($n=1$), MP2/AVQZ ($n=5$), MP2/AVTZ ($n=6$) and MP2/AVDZ ($n=18, 20$ and 24 for He@II(p) , He@5^{12} , and $\text{He@II(h)}/\text{He@I}_h$, respectively). We should note that such systems also present numerous local minima on their PESs, many of them lying very close in energy. The resulting reference structures, considered hereafter in this study, are displayed in Figure 1, while their corresponding coordinates are given in the Supporting Information. Notice that in the simplest system the origin of coordinates is placed on the oxygen atom, and the X-axis lies along the C_{2v} -axis of the water molecule, while for the fundamental units and the building blocks the origin of coordinates is at the center of mass of the corresponding hydrostructure, with the Z-axis serving as the reference direction for this study.

Analytical model potentials

The availability of analytical model potentials allows to carry out molecular computer simulations for larger systems and over longer time intervals, as they are computationally much less demanding than direct (on-the-fly) first-principles calculations. Both semiempirical and *ab initio* model potentials are available in the literature for representing the interactions in the simplest $\text{He-H}_2\text{O}$ system. So, we assumed a sum-of-potentials form to express the potential energy surface (PES) of each of the larger-size systems under study. The potential energy of any $\text{He}-(\text{H}_2\text{O})_n$ ($n > 1$) system is given here by considering only the pairwise additive terms of the water-water ($w-w$) and He-water ($\text{He}-w$) interactions,

$$V_{tot} = \sum_{w=1}^n \sum_{w' > w}^n V_{w-w'} + V_{\text{He}-w_n}, \text{ with } V_{\text{He}-w_n} = \sum_{w=1}^n V_{\text{He}-w} . \quad (1)$$

We should note that such potential representation includes only the lowest two-body term of the many-body expansion,³⁸ although such forms have been widely used to a plethora of molecular dynamics simulations.^{5,48} In this work, we use the SPC/E⁴⁹ and TIP4Pice⁵⁰

models for the water–water interactions ($V_{w-w'}$) and three different models for the He–water interactions (V_{He-w}); two are semiempirical (SPCEice and TIP4Pice),⁵ and the other is the *ab initio* potential for the He-H₂O interaction from Ref.⁵¹ *Ab initio* potentials are usually designed to account for the complex short-range interactions and the subtle weak long-range interactions is obtained by fits to large sets of electronic structure data of the system over various configurations, whereas the semiempirical models are adjusted to describe properties of the system at specific thermodynamic conditions.³¹

The semiempirical potentials combine vdW interactions between the He and O atoms of the water molecules and electrostatic interactions between 3-point charges at O and H or M-sites, given by:

$$V_{He-w_n} = \sum_i \sum_{j>i}^n \left\{ \frac{q_i q_j}{4\epsilon_0 r_{ij}} + 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \right\} , \quad (2)$$

where the i and j summation runs over He and all O, H, and M-site (in the case of TIP4Pice) centers on each water molecule, respectively. The standard Lorentz-Berthelot combination rules are used to determine the Lennard-Jones (LJ) parameters between unlike atoms, satisfying that $\epsilon_{ij} = (\epsilon_{ii}\epsilon_{jj})^{1/2}$ and $\sigma_{ij} = (\sigma_{ii} + \sigma_{jj})/2$. The LJ parameters between similar type atoms are listed in Table 1 together with the partial charges for the Coulomb interactions. The *ab initio* He-H₂O PES is based on an analytical many-body (up to 4-body) form,⁵¹ with each term being represented by sums of Morse-type products, which has been corrected asymptotically by introducing an additional dispersion term. This expression has been fitted to CCSD(T) data refined at their complete basis set limit, determined as a sum of accurate values of MP2 extrapolated using large basis sets. The analytical form and its corresponding 42 parameters are given in Ref.⁵¹

Table 1: Partial atomic charges and LJ parameters for the TIP4Pice (SPCE-ice) and the He atom. The LJ potential parameters between He-O atoms are determined by the combination rules (see text).

Atom	q[e]	$\sigma[\text{\AA}]$	$\epsilon[\text{kJ/mol}]$
O(water)	0 (-0.8476)	3.1668 (3.1556)	0.88215 (0.65063)
H(water)	+0.5897 (+0.4238)	0 (0)	0 (0)
M(water)	-1.1794 (0)	0 (0)	0 (0)
He	0 (0)	2.861 (2.855)	0.273 (0.235)

Electronic structure calculations

The WF-based calculations were carried out with Molpro⁵² and ORCA⁵³ software, whereas the DFT calculations were performed using Gaussian⁵⁴ package with the additional dispersion corrections computed by the DFT-D3⁵⁵ and DFT-D4⁵⁶ programs.

Reference interaction energies for all systems are obtained from density fitting DFMP2 computations, as implemented in Molpro⁵² package, with a careful analysis against basis set superposition error (BSSE) for different correlation-consistent basis sets, such as the aug-cc-pVXZ⁵⁷ (namely AVXZ), with X=T, Q and 5. The counterpoise corrected (CP) interaction energies are calculated as the difference between the uncorrected interaction energy and the BSSE correction,⁵⁸ $\Delta E^{CP} = E_{AB}^{AB} - E_A^A - E_B^B - BSSE = E_{AB}^{AB} - E_A^{AB} - E_B^{AB}$, where $BSSE = E_A^{AB} - E_A^A + E_B^{AB} - E_B^B$, E_A^{AB} and E_B^{AB} stand for the monomer energies calculated with the dimer (AB) basis set, E_A^A and E_B^B represent the same energies calculated with the monomer (A or B) basis set and E_{AB}^{AB} corresponds to the dimer energy calculated with the dimer basis set. Here, A corresponds to the He atom, while B to the water-monomer, or water-pentamer/hexamer, or 5¹² cage, or the I_h and II ice-like systems, shown in Fig. 1.

Such water-containing systems suffers from huge BSSE effects, even in large basis sets, and thus complete basis set (CBS) extrapolation schemes have been employed to obtain accurate well-converged results. Various basis set extrapolation schemes were checked and used⁵⁹⁻⁶³ to reach the CBS limit. As the CP correction has a large influence on the in-

teractions energies, we found that weighted average CBS schemes of the CP-corrected and uncorrected energies, such as the so-called half-counterpoise^{62,63} and the two-step scheme by Lee *et al.*,⁶¹ are the most appropriate for obtaining the interaction energies. After a careful analysis of the results, we choose to employ the two-step weighted average scheme by Lee *et al.*,⁶¹ $E_{\text{CBS}} = \frac{1}{2} \frac{(\delta_X \epsilon_{X+1} - \delta_{X+1} \epsilon_X)}{(\delta_X - \delta_{X+1})}$, with $\delta_X = E_X^{CP} - E_X$, $\epsilon_X = E_X^{CP} + E_X$, and E^{CP} being the CP-corrected energy, while E represents the uncorrected one.

Looking for a computationally accessible method to obtain the reference energies, single point calculations using different WF-based methods were carried out, employing the following levels of theory: couple-cluster based treatments, such as CCSD(T), explicitly-correlated CCSD(T)-F12, as well as domain-based local pair-natural orbital DLPNO-CCSD(T) approaches, and Møller-Plesset perturbation theory, such as MP2, density fitting DFMP2 and explicitly-correlated DFMP2-F12, as implemented in the Molpro⁵² and ORCA⁵³ packages of codes. The augmented correlation consistent basis sets AVXZ (X=T, Q, 5) were employed in the conventional methods for orbital expansion, and auxiliary basis sets were employed for the F12, DLPNO and DF variants. All explicitly-correlated CCSD(T)-F12, MP2-F12 and DLPNO-MP2-F12 calculations used the correlation-consistent basis sets in conjunction with the appropriate complementary auxiliary basis set for resolution of the identity (OptRI), and density fitting of the Fock, exchange and other two-electron integrals (JKfit, MP2fit/CCSDfit).^{52,53} We used the F12b energies, as we found that they converge monotonically using both AVTZ and AVQZ basis sets. The F12a/AVQZ values were found to be close to those from F12b/AVQZ, while the F12a/AVTZ overestimate the basis set limit. Also, the correlation fitting basis sets aug-cc-pVTZ/C are necessary for the resolution of identity approximation as part of the DLPNO scheme. In all PNO-based calculations discussed here the tight threshold for both SCF and PNO settings were employed to achieve better converged wave-functions and to reduce any numerical noise, respectively.

In the DFT calculations, different density functionals were chosen from different groups/rungs (according to the Jacob’s ladder classification), such as TPSS⁶⁴ and M06L⁶⁵ (meta-GGAs/3rd

rung), B3LYP,⁶⁶ PBE0⁶⁷ and M06-2X⁶⁸ (hybrid/meta-GGAs/4th rung), and CAMB3LYP,⁶⁹ LC- ω PBE,⁷⁰ and ω B97XD⁷¹ (range-separated hybrid). All DFT computations were carried out with the AVQZ basis set and for the numerical integration the ultrafine grid was used. In turn, the semi-local DFT approaches are corrected including different semi-classical dispersion schemes, such as the D2,⁷² the original zero-damping function D3(0),⁷³ as well as the most popular Becke-Johnson damping function D3BJ,⁷⁴ and the most recent developed DFT-D4^{75,76} model including many-body dispersion interactions beyond the pairwise terms.

3. RESULTS AND DISCUSSION

Performance of wavefunction-based approaches: New reference interaction energies

The total and interaction energies were computed for a series of wavefunction methods at different levels of theory using augmented correlation-consistent basis sets, as well as the above mentioned 2-step extrapolation CBS scheme for all structures mentioned (see Fig. 1) with $n=1, 5, 6$ and $n=18, 20, 24$ water molecules. The obtained interaction energies, corresponding to the simplest system and fundamental units are listed in Table 2. Both CP corrected and uncorrected values are calculated. We found that for the He(H₂O) the CP correction counts $\sim 3.6, 1.3$, and 0.8 cm^{-1} , at the CCSD(T)-F12 level of theory using the AVTZ, AVQZ, and AV5Z basis, respectively. By increasing the number of water molecules due to cooperative effects, this correction increases to 5 cm^{-1} for He-(H₂O)_{5,6} from the DFMP2-F12/AVTZ calculations, corresponding to about 10% of the energy values. This indicates that such corrections should be taken into account for any quantitatively accurate prediction of the interaction energies of these systems.

Table 2: Interaction energies (cm^{-1}) with, ΔE^{CP} , and without, ΔE , counterpoise correction and at their CBS limit, ΔE^{CBS} , computed at the indicated levels of theory for the simplest and fundamental units.

Method/Basis set		He(H ₂ O)	He(H ₂ O) ₅	He(H ₂ O) ₆
CCSD(T)/CBS [56]	ΔE^{CBS}	-34.4	-	-
CCSD(T)/AVQZ	ΔE	-36.2	-64.5	-
	ΔE^{CP}	-32.1	-57.6	-
CCSD(T)/AVTZ	ΔE	-38.8	-73.2	-108.4
	ΔE^{CP}	-28.4	-54.8	-80.2
CCSD(T)-F12/AV5Z	ΔE	-34.9	-	-
	ΔE^{CP}	-34.0	-	-
CCSD(T)-F12/AVQZ	ΔE	-35.0	-61.3	-
	ΔE^{CP}	-33.7	-58.8	-
CCSD(T)-F12/AVTZ	ΔE	-37.5	-64.5	-94.5
	ΔE^{CP}	-33.8	-58.0	-86.0
DLPNO-CCSD(T)/AVTZ	ΔE^{CP}	-26.6	-50.8	-75.2
MP2/CBS [Q5]	ΔE^{CBS}	-25.8	-46.2	-64.7
DLPNO-MP2/AVTZ	ΔE^{CP}	-19.1	-42.2	-57.7
DFMP2-F12/AV5Z	ΔE	-25.4	-47.2	-66.4
	ΔE^{CP}	-24.8	-46.0	-64.6
DFMP2-F12/AVQZ	ΔE	-25.3	-47.1	-66.0
	ΔE^{CP}	-24.6	-45.9	-64.5
DFMP2-F12/AVTZ	ΔE	-27.7	-50.2	-69.3
	ΔE^{CP}	-25.5	-45.9	-64.5
DLPNO-MP2-F12/AVTZ	ΔE^{CP}	-25.0	-44.7	-62.8
DFMP2/CBS [Q5]	ΔE^{CBS}	-25.3	-46.1	-64.7

Conventional MP2/CCSD(T) at CBS limit and their explicitly-correlated variants were used for the smaller He-water clusters, given that for larger systems the couple-clusters type methods are currently beyond the reach of computational hardware resources, and we could only employed them here for the smaller basis sets, *e.g.* AVTZ. The density-fitting and explicitly-correlated DFMP2/DFMP2-F12 variants were also tested, as well as pair natural orbital (PNO) local correlation type approaches, combined with F12 treatments.^{53,77,78} The latter appear as a promising alternative to provide basis set convergence and nearly linear

scaling performance. However, the practical implementation of such approaches involves various thresholds and tuned keywords, so their benchmarking with the CCSD(T)/CBS or CCSD(T)-F12 results, at smaller systems, is necessary in order to validate their reliability.⁷⁹

Figure 2: Comparison between CP interaction energies obtained from DFMP2/CBS[Q5] and each of the indicated level of theories for the He(H₂O), He(H₂O)₅ and He(H₂O)₆ systems. δ represents the energy difference between them. Vertical color dashed lines correspond to data obtained from the highest level calculations performed here for each system.

Thus, in Figure 2 we compare the difference, δ , obtained between the interaction energy values of each method (see Table 2) with those of the DFMP2/CBS[Q5] energies, available in the present study for all systems of interest. The calculated values from the highest level CCSD(T)/CBS[56], CCSD(T)-F12/AVQZ and CCSD(T)-F12/AVTZ calculations available

for each of the He-water, He-pentamer and He-hexamer water clusters, respectively, are also displayed (see vertical dashed lines). As it was expected, the DFMP2/CBS[Q5] values are very close to the MP2/CBS ones, as well as to their F12 analogs, and higher than those estimated by the CSSD(T)/CBS and CCSD(T)-F12 variants by 10–20 cm^{-1} . Also, we found that the DLPNO approaches, such as DLPNO-CCSD(T), DLPNO-MP2 and DLPNO-MP2-F12, provide values that are in reasonable accord with their conventional analogs. In particular, differences of less than 5 cm^{-1} are obtained between DLPNO-CCSD(T)/AVTZ and CCSD(T)/AVTZ, while the results from DLPNO-MP2-F12/AVTZ calculations are very close to those from DFMP2-F12/AVTZ and DFMP2/CBS[Q5]. Further, we checked the computational performance of these approximations, and we should highlight the significant reduction of computing time, in the case of DLPNO-CCSD(T)/AVTZ compared to the CCSD(T)/AVTZ calculations.

In turn, the CP interaction energies of the $\text{He}@5^{12}$, $\text{He}@I_h$ and $\text{He}@II(\text{h/p})$ building blocks (see Fig. 1) are listed in Table 3 from DLPNO-CCSD(T)/AVTZ, DFMP2-F12/AVTZ and DFMP2/CBS[Q5] calculations. The results from the DFMP2/CBS and DFMP2-F12/AVTZ calculations are very similar, indicating the efficiency of the F12 methods to achieve rapid basis set convergence. Taking into consideration the interaction energies values for the $\text{He}@I_h$ and $\text{He}@II(\text{h})$ systems, we can see that they are nonbound structures, although one should consider that the periodicity plays an important role, and therefore, it can contribute to the stabilization of the structures. As regards the results of the interaction energies on these larger-size building blocks systems, similar trends as in the case of the simplest/fundamental units systems were found, with the DLPNO-CCSD(T) calculations being systematic and they should also be considered for benchmark studies. Further, previous data available from MP2/6-31G** calculations have reported⁸⁰ interaction energies of -0.31 kcal/mol (-108.4 cm^{-1}) for the $\text{He}@5^{12}$ cage system. Such results predict a less stable by 60-90 cm^{-1} guest-host complex than all present data listed in Table 3.

Table 3: Interaction energies (cm^{-1}) with, ΔE^{CP} , and without, ΔE , CP correction and their values at CBS limit, ΔE^{CBS} , computed at indicated levels of theory for the He@5^{12} , He@I_h and He@II(h/p) channels) building blocks.

Method/Basis set		He@5^{12}	He@I_h	He@II(h)	He@II(p)
DLPNO-CCSD(T)/AVTZ	ΔE^{CP}	-190.1	11.1	9652.7	-79.7
DFMP2-F12/AVTZ	ΔE	-182.7	32.0	9525.0	-63.6
	ΔE^{CP}	-170.4	48.3	9557.2	-47.4
DFMP2/CBS [Q5]	ΔE^{CBS}	-169.2	49.1	9526.6	-46.3

Summarizing, the coupled cluster methods offer more accurate results, but at higher computational cost, in such a way that using them in molecular aggregates larger than the simplest system is getting a computationally harder task (*e.g.* for the fundamental units), or even impossible for the building blocks systems, depending also on the size of the basis set employed. On the contrary, in spite of the fact that the density fitting Møller-Plesset perturbation theory methods provide rougher results, the computational time and memory space required for the calculations is significantly reduced. As regards the explicitly-correlated methods (F12), the most outstanding feature is that the convergence is produced at the CBS level with smaller basis sets. For example, in the case of the $\text{He-(H}_2\text{O)}_6$, we can see that the interaction energy calculated at the DFMP2-F12/AVTZ level (-64.5 cm^{-1}) is practically equal to those obtained with DFMP2/CBS[Q5] (-64.7 cm^{-1}), although within $\approx 20 \text{ cm}^{-1}$ from the CCSD(T)-F12 value. However, as the number of atoms increases, the difference between the energy values obtained with CCSD(T) and DMFP2 (and its explicitly-correlated analogs) also grows, for example comparing CCSD(T)/AVTZ with DFMP2/CBS the difference is about 3, 9 and 15 cm^{-1} for the simplest and two fundamental units systems, respectively. Furthermore, results corresponding to DFMP2 and DFMP2-F12 calculations are close to that of MP2, what suggests that the minimum loss of accuracy that the approximation of density fitting can suppose is totally compensated by the reduction of computational time.

Thus, in this way we evaluate the effect of the method and the basis set on the interaction energies of the systems under study. Since performing the calculations at the coupled cluster

level is not a possibility for large complexes, we concluded that the most adequate and practical approach to obtain the reference interaction energies is from the DFMP2/CBS[Q5] calculations, considering that an underestimate with a lower error to be given by the DPNLO-CCSD(T)/AVTZ calculations, available for all studied systems. All total and interaction energies (see Tables 2 and 3) for the above mentioned calculations are given in the Supporting Information material.

Performance of pairwise model potentials: semiempirical *vs ab initio* representation

As mentioned in subsection 2.2, we considered three different pairwise V_{He-w_n} interaction model potentials, corresponding to the semiempirical SPCEice and TIP4Pice (see Eq. 2 and Table 1) and *ab initio*-based (see Eq. 1 and Ref.⁵¹) ones, namely PES1, PES2 and PES3, respectively, hereafter.

In Figure 3 we display the potential curves of the PES1, PES2 and PES3 models together with the interaction energies from the corresponding DFMP2/CBS[Q5] and the higher level couple-cluster variant currently affordable for each system, as a function of the He–O distance R for the He–(H₂O) (see upper panel), and Z -coordinate for the He–(H₂O)₅ and He–(H₂O)₆ clusters (see middle and lower panels). One can see certain differences in the potential curves, corresponding at their well-depths and positions, as well as in the repulsive part of them, which are getting larger as the size of the system increases. As it was expected, for the simplest system (see upper panel) the PES3 and CCSD(T) potential curves show very similar behavior especially at equilibrium configurations, while the PES1 and PES2 semiempirical curves show significant differences in these regions. The DLPNO-CCSD(T) value is found between the CCSD(T) and DFMP2 CBS results. We should point out that until very recently, MP2/CBS or CCSD(T) with low basis sets were used to define the upper

Figure 3: Comparison of the DFMP2/CBS[Q5] and CCSD(T)-F12 (* indicates the highest level calculation performed here for each system) interaction energies with the potential curves obtained from the PES1 and PES2 semiempirical and PES3 *ab initio*-based models for the indicated (simplest and fundamental unit) systems (see text). The $\times$ symbol corresponds to the DLPNO-CCSD(T)/AVTZ interaction energy values for the He atom placed at the specific R/Z position.

and lower boundaries of confidences, while now, it is possible to lock-up those boundaries with a much fine tuning from the DLPNO-CCSD(T). For the fundamental unit systems (see middle panel) the curves show two slightly different minima at $+Z$ and $-Z$ directions, influenced by the hydrogen atoms orientation in the pentamer and hexamer (rings) water systems. For the $\text{He}-(\text{H}_2\text{O})_5$ cluster the potential curves obtained from all analytical PES1,

PES2 and PES3 models show deviations from those of the DFMP2/CCSD(T)-F12 calculations, especially for configurations nearby the barrier at $Z=0$. In contrary, for the $\text{He}-(\text{H}_2\text{O})_6$, the PES3 curve lies closer to the DFMP2/CBS[Q5] and CCSD(T)-F12/AVTZ energies, even in the repulsive region up to the barrier corresponding to the “on-the-surface” position of the He atom in the hexamer cluster, where larger deviations between *ab initio*-based results and semiempirical models are found (see lower panel).

Figure 4: Comparison of the DFMP2/CBS[Q5] interaction energies with the potential curves obtained from the PES1 and PES2 semiempirical and PES3 *ab initio*-based models for the indicated (building blocks) systems. The $\times$ symbol corresponds to the DLPNO-CCSD(T)/AVTZ interaction energy values with He atom at $Z=0$ or $2 \text{ \AA}$.

The corresponding potential curves along the Z -axis for the building blocks, $\text{He}@5^{12}$, $\text{He}@I_h$ and $\text{He}@II(h/p)$, are shown in Figure 4. Using such scans one could also corroborate if, by any chance, the diffusion of the He atom is favorable inside the channel or cavity structures, or the only way to release the guest should be via the deformation of the water structures through decomposition. For the $\text{He}@5^{12}$ we observe two global maxima corresponding to the location of the He atom on the surface of the cavity (around the center of the pentamer along the Z -axis at around 3.5 Å), and a broad minimum region between -1.5 and 1.5 Å, corresponding to the configuration with the He atom close to the center of the 5^{12} cavity (see inset panel). In the cases of the $\text{He}@I_h$ and $\text{He}@II(h)$ structures, all potential curves have approximately the same shape, presenting two maxima when the He atom is at the edge of the ice-like channels, and a global minimum when the He atom is located in the center of the structure (see corresponding inset panels). In contrary, for the $\text{He}@II(p)$ structure the potential minima correspond to the location of the He atom near the center of the side cavities, and the maxima at the center of the three parallel each other hexamers along the II structure (see inset panel). By comparing the overall shape of the curves, the PES3 results are found to be closer to the DFMP2/CBS than those obtained from the semiempirical models. However, significant differences are present, such as in the case of $\text{He}@II(h)$, and at specific local configurations of the $\text{He}@I_h$ system.

As we mentioned above, the most problematic regions of the PESs seems to be nearby the barriers, where larger errors are obtained. Conventional analytical PESs, works usually at the specific conditions that they have been constructed, thus in the present case, the PES1 and PES2 represent experimental solid-phase conditions, while PES3 corresponds the gas phase theoretical calculations. We have also note, that for the attractive part of the PESs the errors look less pronounced, with the present models providing just qualitative behavior, and thus requiring more flexible expressions to describe all aspects of the PES's.

Performance of DFT/DFT-D approaches: Jacob’s ladder classification and vdW corrections

In following we proceed with the evaluation of the performance of different DFT/DFT-D functionals. Those considered here for assessment belong to meta-GGA (TPSS and M06L), hybrids (B3LYP, PBE0 and M062X) and range-separated (LC ω PBE, CAMB3LYP and ω B97X) functional rungs. The calculated energies were computed using the AVQZ basis set for all structures, and adding dispersion correction terms (D2, D3(0), D3BJ and D4) on those functionals in which it was available in order to count with long-range London dispersion interactions. The DFT-D4 energies were calculated by using both the third-order Axilrod-Teller-Muto (ATM) term⁸¹ and the coupled-dipole based many-body dispersion (MBD) correction.⁸² We found that the DFT-D4-MBD scheme yields an improved description of the interactions compared to those from the DFT-D4-ATM method. So, DFT-D4 energies from the DFT-D4-MBD were considered hereafter.

Starting with the smaller systems (forming by a monomer/pentamer or hexamer interacting with one He atom), we first compare in Figure 5, the interaction energy values calculated from the DFT/DFT-D functionals with the corresponding DFMP2/CBS[Q5] reference energies. We also indicate with dashed lines in the figure, the value from the CCSD(T)-F12 calculations (red color line), as well as those from the CCSD(T)/CBS[56] ones (pink color line) in the case of the He-H₂O system.

As it can be seen, the DFT functionals show difficulties to determine the interaction energy nearby the equilibrium geometries of each of the complexes. One can see that the addition of dispersion correction terms barely shows any improvement in the results, although some tendencies could be observed. In particular, most of the functional families, *e.g* the TPSS, PBE0, M062X CAMB3LYP, and LC ω PBE, show a consistent behavior for the three systems under study for both the pure functional and dispersion-corrected ones, while for the remaining functionals, like M06L, B3LYP and ω B97X, no clear patterns were found.

Figure 5: Interaction energy difference, δ , between the reference value from the DFMP2/CBS[Q5] and the indicated DFT/DFT-D calculations for the $\text{He-(H}_2\text{O)}_n$ ($n=1, 5$ and 6) systems. Vertical color dashed lines correspond to the values obtained in the present study by the highest level calculation performed here for each system.

Thus, in order to further evaluate the performance of the DFT functionals, in the next step, we select some specific configurations of the systems along the Z coordinate (R in the case of the monomer) to perform the interaction energy calculations. In Figure 6, we show some representative data obtained from the meta-GGA TPSS, the hybrid PBE0 and the range-separated CAMB3LYP and LC ω PBE DFT functionals. By comparing them with the DFMP2/CBS[Q5] data we found that the LC ω PBE functional along with the D3(0) and D3BJ corrections shows the best performance in the monomer and pentamer/hexamer cases,

Figure 6: Interaction energy curves obtained from selected DFT functionals, with and without dispersion corrections, for the He atom moving along the Z -axis (R in the case of the monomer) of the $\text{He}-(\text{H}_2\text{O})_n$ ($n=1, 5$, and 6) systems, and their comparison with the corresponding reference curves computed at DFMP2/CBS[Q5] and CCSD(T)/CCSD(T)-F12 (* indicates the highest level calculation for each system) levels.

respectively. However, by taking into account the predicted higher level CCSD(T)-F12 curves then apart of the LC ω PBE-D3(0) values, one should also consider the PBE0-D3BJ and PBE0-D4-MBD energies or even those from the TPSS-D2 calculations as reasonable estimates. Also we should point out the very close values, or even overlap, of the curves including the D3BJ and D4 dispersion corrections for most functionals depicted in Figure 6.

Once we have completed the analysis of the DFT functionals’ performance for the fundamental units, we continue with the assessment of the selected meta-GGAs, hybrid and range-separated functionals on the larger building block systems with $n=18, 20$, and 24 water molecules. In Figure 7, we display a graphical representation of the individual root-mean-square (RMS) errors of each of the above mentioned DFT/DFT-D functionals, calculated as $RMS = \sqrt{(E_{ref} - E_{DFT})^2}$, where E_{ref} are the DLPNO-CCSD(T)/AVTZ (and the DFMP2/CBS[Q5] see Fig. S1 in the Supporting information) interaction energies at the reference structures of each He-water system (see Fig. 1). Each panel of the figure contains the pure functionals together with the different dispersion schemes employed. In the TPSS panel we clearly see the improvement obtained by including dispersion, with the D2 scheme for all systems. Among the PBE0-D approaches the D4 one shows the better results, although similar to those from the D3BJ scheme, especially for the ice-channel like systems. Next the CAMB3LYP-D3BJ and D4 notably improve the energies of the fundamental unit and He@5¹² systems, while they yield worse results for the He@ice-like systems. In turn, the LC ω PBE-D3BJ data show more balanced behavior, except for the He@II(h) structure, where the LC ω PBE-D3(0) proved somehow better.

In general, the higher the number of molecules, the higher is expected to be the computed RMS error. We should point that the pronounced error values in the He@II(h) for all considered functionals correspond to an error of around 4–5% in the best cases of each functional. We should also highlight the great improvement which supposes the application of dispersion corrections in most functionals, especially in the TPSS and LC ω PBE, while this is not the case for the CAMB3LYP functional including the D3(0) correction. Nonetheless, in a first

Figure 7: RMS errors (in cm^{-1}) of the computed DFT/DFT-D interaction energies against the DLPNO-CCSD(T)/AVTZ (or DFMP2/CBS[Q5], see Fig. S1) benchmark values, calculated at the reference structures of the indicated He-water systems.

glance we may generally say that for the fundamental units and the clathrate-like He@5¹² cage, the best performance functionals are the range-separated CAMB3LYP or LC ω PBE including the D4/D3BJ correction, while for all the He@ice-channel systems the TPSS-D2 and PBE0-D4 ones provide the most accurate results.

Figure 8: Interaction energy curves obtained from the indicated DFT functionals, with and without dispersion corrections, for the He atom moving along the Z-axis in the He@5¹² cage, and compared with the reference DFMP2/CBS[Q5] values. The $\times$ symbol corresponds to the DLPNO-CCSD(T)/AVTZ interaction energy values with He atom at $Z=0$ Å.

In particular, the interaction energies for the He@5¹² complex estimated by the TPSS-D2, PBE0-D4, CAMB3LYP-D3BJ/D4, and LC ω PBE-D3BJ calculations were -250.3, -247.5, -174.7/-161.8, and -221.2 cm⁻¹, respectively, compared to the -190.1 and 169.2 cm⁻¹ from

the DLPNO-CCSD(T)/AVTZ and DFMP2/CBS values, respectively. We should also point out that in a recent study¹⁰ a value of -258.8 cm^{-1} has been reported from B97-D/cc-pVTZ calculations, while for the present benchmark structure such DFT-D computations further overestimate the interaction in the He@5¹² with energy value of -253.8 cm^{-1} .

Further, in Figure 8 we display the interaction energy curve along the Z coordinate for the He@5¹² reference building block system, while such potential curve plots for the He@ice-like building blocks are given in the Supporting Information material (see Figs. S2, S3 and S4, respectively). By checking the performance of each functionals under study at selected representative configurations different that the equilibrium ones, we aim to extract further information on their well-balanced description of the interaction of such He-water complexes. So, by analyzing carefully the obtained curves one may conclude that the LC ω PBE-D3BJ and PBE0-D4 are the most successful approaches in describing the underlying interactions in such He@water systems, while in terms of computational cost TPSS-D2, especially in the ice-like systems could also be considered. In particular, the LC ω PBE-D3BJ provides the most accurate results in the case of He@5¹², while the PBE0-D4 and TPSS-D2 are found to further overestimate interaction energies of the system. For all He@I_h, He@II(p) and He@II(h) cases the PBE0-D4 is the best performed approach followed by the TPSS-D2 one. Finally, in Figure 9 we summarized the overall comparison between the benchmark DFMP2/CBS[Q5] and DLPNO-CCSD(T)/AVTZ data, with the best performed PES3 analytical potential model and the LC ω PBE-D3BJ, PBE0-D4, and TPSS-D2 functionals that were found to provide the most accurate descriptions of the He@water interactions considered here. It is worth pointing out the reasonable performance of the PES3 potential model, especially for the He@5¹² clathrate-like, showing similar behavior as the TPSS-D2 and PBE0-D4 functionals, and He@I_h/He@II(p) ice-like complexes.

Figure 9: Interaction energy curves for the He atom moving along the Z-axis in the He@5¹², He@I_h, He@II(p) and He@II(h) systems, as calculated from the DFMP2/CBS[Q5], the analytical PES model potential and the LCωPBE-D3BJ, PBE0-D4, and TPSS-D2 functionals. The DLPNO-CCSD(T)/AVTZ value (see × symbols) with He atom at Z=0 or 2 Å is also displayed.

4. SUMMARY AND CONCLUSIONS

We have carefully assessed the performance of a variety of DFT-based approaches, as well as analytical semiempirical and *ab initio*-based pairwise potential models available by comparisons with reference data obtained from well-converged wavefunction-based calculations. Initially, a total of seven different He(H₂O)_n reference set of structures, related to the fundamental units and building blocks of clathrate-hydrates and ice-water channels, were considered, and their total and interaction energies were computed. Extrapolation schemes to

the CBS were necessary in order to obtain more accurate results due to the BSSE effects, present even in the large basis sets. Various *ab initio* wavefunction-based methodologies, such as CCSD(T), MP2, as well as their explicitly-correlated (F12) and PNO analogs, were checked. The DLPNO-CCSD(T) and the DFMP2/CBS[Q5] were found as the most adequate and computationally affordable approaches, up to date, to obtain benchmark energies of all He@water systems under study. Thus, we span a broad range of different representative configurations for each benchmark structure, from equilibrium, moderate-binding to non-binding ones to ensure an overall description of the interactions at these levels of theory.

Such reference data were then used to evaluate the performance of the different DFT approaches climbing up the rungs in Jacob’s ladder. In particular, we analyzed different meta-GGA (TPSS and M06L), hybrids (B3LYP, PBE0 and M062X) and range-separated (LC ω PBE, CAMB3LYP and ω B97X) functionals. Pure DFT functionals have been found inadequate to describe the noncovalent interactions (such as hydrogen bonds and vdW interactions playing a relevant role in clathrates), as they do not explicitly describe the dispersion effects, and thus DFT-D methods were then employed, including D2, D3 and the most recent D4 correction schemes. From an initial assessment in the fundamental units (monomer, pentamer and hexamer), we concluded that the functionals within each respective functional rung showing a well-balance performance are the TPSS, PBE0, CAMB3LYP and LC ω PBE functional families, with the LC ω PBE-D3BJ and PBE0-D3BJ/D4 yielding better overall results. So, the present benchmark focused only on the performance of these specific functional families including all corresponding dispersion corrections on the larger He@hydrate systems (clathrate’s cage, Ih and I_{II} channels). Our analysis came to the conclusion that the LC ω PBE and PBE0 functionals including the D3BJ and D4 corrections, respectively, show again better performance, with the PBE0-D4 providing improved results in the He@ice-like systems.

In turn, results obtained from *ab initio*-based and semiempirical model potentials were also compared with both DLPNO-CCSD(T) and DFMP2/CBS reference data, as well as

the best-performed DFT-D approaches. Although, the overall shape of the curves is similar, discordance is encountered in their maxima and the minima regions. As most of the conventional analytical potentials are application-oriented models, the idea of transferability is just the result of capturing mean features or cancellation of errors. Thus, given such limitations, machine learning many-body potentials sounds a promising alternative, especially if one is interested in studying phase transitions (gas–liquid–solid including finite systems, such as clusters and nanostructures). On the one hand, current efforts are oriented to create such models in a general automated fashion by selecting the most representative and diverse set of structures. This suggests that instead of focusing on specific configurational arrangements, high-energy structures should also be included, as they might serve to explore and explain challenging structural problems in heterogeneous systems. On the other hand, the quality of the training energy data also has profound implications for the models development, and thus the outcome of benchmark studies on the availability and choice of well-converged WF and/or reliable DFT approaches, respectively, becomes essential.

In conclusion, among the studied DFT/DFT-D functionals and analytical pairwise potentials, we found difficulties in describing accurately such He@water systems, although, some general trends were observed. Importantly, such systems are large enough to permit checking reliability of DFT on nonstandard noncovalent interactions. Further, we should underline the importance of considering in the benchmark studies non-equilibrium representative geometries, as well as key point of the availability of well-converged wavefunction-based reference data in assessing the quality of DFT approaches and analytical pairwise (normally) potential models. Finally, such investigations provide valuable information that serves to further improve existing approaches for building up predictive (data-driven) models, and employing them in computationally high-demanding dynamics simulations for reliable nanoclusters’ and bulk properties.

Acknowledgement

We thank María Blanco de Paz for sending us He@ice-like configurations from Ref.⁴⁷ The authors thank to the “Centro de Calculo del IFF/SGAI-CSIC” and CESGA-Supercomputing centre for allocation of computer time. This work has been supported by MINECO grants No. FIS2017-83157-P and Comunidad de Madrid grant Ref: IND2018/TIC-9467.

Associated Content

Supporting Information Available

Supplementary files He-W.xyz, He-W5.xyz, He-W6.xyz, He@W20.xyz, He@II-p.xyz, He@II-h.xyz and He@Ih.xyz including reference structures coordinates. Supplementary table Table-SI-reference-RYR.ods containing total and interaction reference energy values. Supplementary S1 figure displays the RMS errors of the DFT/DFT-D functionals with respect the DFMP2/CBS[Q5] data. Supplementary S2, S3 and S4 figures show potential energy curves for the He@ice-like building blocks.

References

- (1) Dyadin, Y. A.; Larionov, E. G.; Manakov, A. Y.; Zhurko, F. V.; Aladko, E. Y.; Mikina, T. V.; Komarov, V. Y. Clathrate Hydrates of Hydrogen and Neon. *Mendeleev Commun.* **1999**, *9*, 209–210.
- (2) Dyadin, Y. A.; Aladko, E. Y.; Manakov, A. Y.; Zhurko, F. V.; Mikina, T. V.; Komarov, V. Y.; Grachev, E. V. Clathrate Formation in Water-Noble Gas (Hydrogen) Systems at High Pressures. *J. Struct. Chem.* **1999**, *40*, 790–795.
- (3) Loveday, J. S.; Nelmes, R. J. High-Pressure Gas Hydrates. *Phys. Chem. Chem. Phys.* **2008**, *10*, 913–1068.

- (4) Mondal, S.; Chattaraj, P. K. Noble Gas Encapsulation: Clathrate Hydrates and their HF Doped Analogues. *Phys. Chem. Chem. Phys.* **2014**, *16*, 17943–17954.
- (5) Belosludov, R. V.; Bozhko, Y. Y.; Subbotin, O. S.; Belosludov, V. R.; Mizuseki, H.; Kawazoe, Y.; Fomin, V. M. Stability and Composition of Helium Hydrates Based on Ices Ih and II at Low Temperatures. *J. Phys. Chem. C* **2014**, *118*, 2587–2593.
- (6) Yu, X.; Zhu, J.; Du, S.; Xu, H.; Vogel, S. C.; Han, J.; Germann, T. C.; Zhang, J.; Jin, C.; Francisco, J. S. et al. Crystal Structure and Encapsulation Dynamics of Ice II-Structured Neon Hydrate. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 10456–10461.
- (7) Teeratchanan, P.; Hermann, A. Computational Phase Diagrams of Noble Gas Hydrates Under Pressure. *J. Chem. Phys.* **2015**, *143*, 154507.
- (8) Vítek, A.; Arismendi-Arrieta, D. J.; Rodríguez-Cantano, R.; Prosimi, R.; Villarreal, P.; Kalus, R.; Delgado-Barrio, G. Computational Investigations of the Thermodynamic Properties of Size-Selected Water and Ar-Water Clusters: High-Pressure Transitions. *Phys. Chem. Chem. Phys.* **2015**, *17*, 8792–8801.
- (9) Arismendi-Arrieta, D. J.; Vítek, A.; Prosimi, R. High Pressure Structural Transitions in Kr Clathrate-Like Clusters. *J. Phys. Chem. C* **2016**, *120*, 26093–26102.
- (10) Kaur, S. P.; Ramachandran, C. N. Host-Guest and Guest-Guest Interactions in Noble Gas Hydrates. *Mol. Phys.* **2018**, *116*, 54–63.
- (11) Takeya, S.; Hachikubo, A. Structure and Density Comparison of Noble Gas Hydrates Encapsulating Xenon, Krypton and Argon. *J. Chem. Phys.* **2019**, *149*, 074502.
- (12) Sloan, E. D. Fundamental Principles and Applications of Natural Gas Hydrates. *Nature* **2003**, *426*, 353–359.
- (13) Sloan, E. D.; Koh, C. A.; Koh, C. *Clathrate Hydrates of Natural Gases*; CRC Press, 2007.

- (14) Matsui, T.; Hirata, M.; Yagasaki, T.; Matsumoto, M.; Tanaka, H. Hypothetical Ultralow-Density Ice Polymorphs. *J. Chem. Phys.* **2017**, *147*, 091101.
- (15) Villard, P. Combinaison de l'Argon avec l'Eau. *C. R. Hebd. Seances Acad. Sci.* **1896**, *123*, 377–379.
- (16) Lobban, C.; Finney, J. L.; Kuhs, W. F. The p-T Dependency of the Ice II Crystal Structure and the Effect of Helium Inclusion. *J. Chem. Phys.* **2002**, *117*, 3928–3934.
- (17) Papadimitriou, N. I.; Tsimpanogiannis, I. N.; Stubos, A. K.; Martin, A.; Rovetto, L. J.; Peters, C. J. Unexpected Behavior of Helium as Guest Gas in sII Binary Hydrates. *J. Phys. Chem. Lett.* **2010**, *1*, 1014–1017.
- (18) Ildyakov, A. V.; Manakov, A.; Aladko, E.; Kosyakov, V. I.; Shestakov, V. A. Solubility of Helium in Ice Ih at Pressures up to 2000 bar: Experiment and Calculations. *J. Phys. Chem. B* **2013**, *117*, 7756–7762.
- (19) Salzmann, C. G.; Radaelli, P. G.; Slater, B.; Finney, J. L. The Polymorphism of Ice: Five Unresolved Questions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18468–18480.
- (20) del Rosso, L.; Celli, M.; Ulivi, L. New Porous Water Ice Metastable at Atmospheric Pressure Obtained by Emptying a Hydrogen-Filled Ice. *Nat. Commun.* **2016**, *7*, 13394.
- (21) Salzmann, C. G. Advances in the Experimental Exploration of Water's Phase Diagram. *J. Chem. Phys.* **2019**, *150*, 060901.
- (22) Falenty, A.; Hansen, T. C.; Kuhs, W. F. Formation and Properties of Ice XVI Obtained by Emptying a Type sII Clathrate Hydrate. *Nature* **2014**, *516*, 231–233.
- (23) Kuhs, W. F.; Hansen, T. C.; Falenty, A. Filling Ices with Helium and the Formation of Helium Clathrate Hydrate. *J. Phys. Chem. Lett.* **2018**, *9*, 3194–3198.

- (24) Huang, Y.; Zhu, C.; Wang, L.; Cao, X.; Su, Y.; Jiang, X.; Meng, S.; Zhao, J.; Zeng, X. C. A New Phase Diagram of Water Under Negative Pressure: The Rise of the Lowest-Density Clathrate s-III. *Sci. Adv.* **2016**, *2*, e1501010.
- (25) Londono, D.; Kuhs, W. F.; Finney, J. L. Enclathration of Helium in Ice II: the First Helium Hydrate. *Nature* **1988**, *332*, 141–142.
- (26) Londono, D.; Kuhs, W. F.; Finney, J. L. Formation, Stability and Structure of Helium Hydrate at High Pressure. *J. Chem. Phys.* **1992**, *97*, 547–552.
- (27) Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. Sparse Maps—A Systematic Infrastructure for Reduced-Scaling Electronic Structure Methods. II. Linear Scaling Domain Based Pair Natural Orbital Coupled Cluster Theory. *J. Chem. Phys.* **2016**, *144*, 024109.
- (28) Mostofi, A. A.; Haynes, P. D.; Skylaris, C. K.; Payne, M. C. ONETEP: Linear-Scaling Density-Functional Theory with Plane-Waves. *Mol. Simul.* **2007**, *33*, 551–555.
- (29) Miliordos, E.; Xantheas, S. S. An Accurate and Efficient Computational Protocol for Obtaining the Complete Basis Set Limits of the Binding Energies of Water Clusters at the MP2 and CCSD(T) Levels of Theory: Application to $(\text{H}_2\text{O})_m$, $m = 2-6, 8, 11, 16$, and 17. *J. Chem. Phys.* **2015**, *142*, 234303.
- (30) Arismendi-Arrieta, D. J.; Valdés, A.; Prosimi, R. A Systematic Protocol for Benchmarking Guest-Host Interactions by First-Principles Computations: Capturing CO₂ in Clathrate Hydrates. *Chem. Eur. J.* **2018**, *24*, 9353–9363.
- (31) León-Merino, I.; Rodríguez-Segundo, R.; Arismendi-Arrieta, D. J.; Prosimi, R. Assessing Intermolecular Interactions in Guest-Free Clathrate Hydrate Systems. *J. Phys. Chem. A* **2018**, *122*, 1479–1487.

- (32) Al-Hamdani, Y. S.; Tkatchenko, A. Understanding Non-Covalent Interactions in Larger Molecular Complexes from First Principles. *J. Chem. Phys.* **2019**, *150*, 010901.
- (33) Kodrycka, M.; Patkowski, K. Platinum, Gold, and Silver Standards of Intermolecular Interaction Energy Calculations. *J. Chem. Phys.* **2019**, *151*, 070901.
- (34) Behler, J. Perspective: Machine Learning Potentials for Atomistic Simulations. *J. Chem. Phys.* **2016**, *145*, 170901.
- (35) Wu, Z.; Ramsundar, B.; Feinberg, E.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: a Benchmark for Molecular Machine Learning. *Chem. Sci.* **2018**, *9*, 513–530.
- (36) Mueller, T.; Hernandez, A.; Wang, C. Machine Learning for Interatomic Potential Models. *J. Chem. Phys.* **2020**, *152*, 050902.
- (37) Loeffler, T.; Patra, T.; Chan, H.; Cherukara, M.; Sankaranarayanan, S. K. R. S. Active Learning the Potential Energy Landscape for Water Clusters from Sparse Training Data. *J. Phys. Chem. C* **2020**, *124*, 4907–4916.
- (38) Elrod, M. J.; Saykally, R. J. Many-Body Effects in Intermolecular Forces. *Chem. Rev.* **1994**, *94*, 1975–1997.
- (39) Yao, K.; Herr, J. E.; Parkhill, J. The Many-Body Expansion Combined with Neural Networks. *J. Chem. Phys.* **2017**, *146*, 014106.
- (40) Qu, C.; Bowman, J. M. A fragmented, permutationally invariant polynomial approach for potential energy surfaces of large molecules: Application to N-methyl acetamide. *J. Chem. Phys.* **2019**, *150*, 141101.
- (41) Ou, C.; Bowman, J. M. Ab Initio, Embedded Local-Monomer Calculations of Methane Vibrational Energies in Clathrate Hydrates. *J. Phys. Chem. C* **2016**, *120*, 3167–3175.

- (42) Wang, Q. K.; Bowman, J. M. Two-Component, Ab Initio Potential Energy Surface for $\text{CO}_2\text{H}_2\text{O}$, Extension to the Hydrate Clathrate, $\text{CO}_2@(\text{H}_2\text{O})_{20}$, and VSCF/VCI Vibrational Analyses of Both. *J. Chem. Phys.* **2017**, *147*, 161714.
- (43) Ou, C.; Bowman, J. M. Assessing the Importance of the $\text{H}_2\text{H}_2\text{OH}_2\text{O}$ Three-Body Interaction on the Vibrational Frequency Shift of H_2 in the sII Clathrate Hydrate and Comparison with Experiment. *J. Phys. Chem. A* **2018**, *123*, 329–335.
- (44) Takeuchi, F.; Hiratsuka, M.; Ohmura, R.; Alavi, S.; Sum, A.; Yasuoka, K. Water Proton Configurations in Structures I, II, and H Clathrate Hydrate Unit Cells. *J. Chem. Phys.* **2013**, *138*, 124504.
- (45) Ball, P. *Life’s Matrix: A Biography of Water*; Perseus UCAL POD, 2001.
- (46) Chaplin, M. Water Structure and Science: Ice Phases. http://www1.lsbu.ac.uk/water/ice_phases.html Last accessed: Dec. 2019.
- (47) Blanco de Paz, M.; Arismendi-Arrieta, D. J.; Prosmiti, R. Benchmarking the Performance of Density-Functional-Based Approaches on Intermolecular Interactions of Helium–Water Complexes. *J. Phys.: Conf. Ser.* **2017**, *875*, 102026.
- (48) Alavi, S.; Woo, T. K. How Much Carbon Dioxide Can be Stored in the Structure H Clathrate Hydrates?: A Molecular Dynamics Study. *J. Chem. Phys.* **2007**, *126*, 044703.
- (49) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
- (50) Abascal, J. L. F.; Sanz, E.; García Fernández, R.; Vega, C. A Potential Model for the Study of Ices and Amorphous Water: TIP4P/Ice. *J. Chem. Phys.* **2005**, *122*, 234511.
- (51) Makarewicz, J. Ab initio Intermolecular Potential Energy Surfaces of the Water-Rare Gas Atom Complexes. *J. Chem. Phys.* **2008**, *129*, 184310.

- (52) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; M. Schütz, M.; et al., MOLPRO, Version 2012.1, A Package of Ab Initio Programs. See <http://www.molpro.net> *Last accessed: Jan. 2020.*
- (53) Neese, F. Software Update: the ORCA Program System, version 4.0. *WIREs Computational Molecular Science* **2018**, 8, e1327.
- (54) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H. et al. Gaussian16 Revision C.01. 2016; Gaussian Inc. Wallingford CT.
- (55) Bonn, U. A dispersion correction for density functionals, Hartree-Fock and semi-empirical quantum chemical methods. <https://www.chemie.uni-bonn.de/pctc/mulliken-center/software/dft-d3/> *Last accessed: Feb. 2020.*
- (56) Bonn, U. A dispersion correction for density functionals, Hartree-Fock and semi-empirical quantum chemical methods. <https://www.chemie.uni-bonn.de/pctc/mulliken-center/software/dftd4> *Last accessed: Feb. 2020.*
- (57) Kendall, R.; Dunning, T. H. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, 96, 6796–6806.
- (58) Boys, S.; Bernardi, F. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, 19, 553–566.
- (59) Schwartz, C. Importance of Angular Correlations between Atomic Electrons. *Phys. Rev.* **1962**, 126, 1015–1019.
- (60) Peterson, K. A.; Woon, D. E.; Dunning, T. H. J. Benchmark Calculations with Correlated Molecular Wave Functions. IV. The Classical Barrier Height of the $\text{H} + \text{H}_2 \rightarrow \text{H}_2 + \text{H}$ Reaction. *J. Chem. Phys.* **1994**, 100, 7410–7415.

- (61) Lee, E. C.; Kim, D.; Jurecka, P.; Tarakeshwar, P.; Hobza, P.; Kim, K. S. Understanding of Assembly Phenomena by Aromatic-Aromatic Interactions: Benzene Dimer and the Substituted Systems. *J. Phys. Chem. A* **2007**, *111*, 3446–3457.
- (62) Halkier, A.; Klopper, W.; Helgaker, T.; Jørgensen, P.; Taylor, P. R. Basis Set Convergence of the Interaction Energy of Hydrogen-Bonded Complexes. *J. Chem. Phys.* **1999**, *111*, 9157–9167.
- (63) Burns, L.; Marshall, M.; Sherrill, C. Comparing Counterpoise-Corrected, Uncorrected, and Averaged Binding Energies for Benchmarking Noncovalent Interactions. *J. Chem. Theory Comput.* **2014**, *10*, 49–57.
- (64) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Non empirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- (65) Zhao, Y.; Truhlar, D. G. A New Local Density Functional for Main-Group Thermochemistry, Transition Metal Bonding, Thermochemical Kinetics, and Noncovalent Interactions. *J. Chem. Phys.* **2006**, *125*, 194101.
- (66) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (67) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- (68) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Non-Covalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

- (69) Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange-Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51–57.
- (70) Vydrov, O. A.; Scuseria, G. E. Assessment of a Long-Range Corrected Hybrid Functional. *J. Chem. Phys.* **2006**, *125*, 234109.
- (71) Chai, J.; Head-Gordon, M. Systematic Optimization of Long-Range Corrected Hybrid Density Functionals. *J. Chem. Phys.* **2008**, *128*, 084106.
- (72) Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.
- (73) Grimme, S.; Jens, A.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
- (74) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 456–1465.
- (75) Caldeweyher, E.; Bannwarth, C.; Grimme, S. Extension of the D3 Dispersion Coefficient Model. *J. Chem. Phys.* **2017**, *147*, 034112.
- (76) Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A Generally Applicable Atomic-Charge Dependent London Dispersion Correction. *J. Chem. Phys.* **2019**, *150*, 154122.
- (77) Neese, F.; Hansen, A.; Liakos, D. G. Efficient and Accurate Approximations to the Local Coupled Cluster Singles Doubles Method Using a Truncated Pair Natural Orbital Basis. *J. Chem. Phys.* **2009**, *131*, 064103.
- (78) Ma, Q.; Werner, H.-J. Explicitly Correlated Local Coupled-Cluster Methods Using Pair Natural Orbitals. *WIREs Computational Molecular Science* **2018**, *8*, e1371.

- (79) Schmitz, G.; Hättig, C. Accuracy of Explicitly Correlated Local PNO-CCSD(T). *J. Chem. Theory Comput.* **2017**, *13*, 2623–2633.
- (80) Kumar, P.; Sathyamurthy, N. Theoretical Studies of Host-Guest Interaction in Gas Hydrates. *J. Chem. Phys.* **2011**, *115*, 14276–14281.
- (81) Axilrod, B. M.; Teller, E. Interaction of the van der Waals Type Between Three Atoms. *J. Chem. Phys.* **1943**, *11*, 299–300.
- (82) DiStasio, R. A.; Gobre, V. V.; Tkatchenko, A. Many-body van der Waals Interactions in Molecules and Condensed Matter. *J. Phys.: Condens. Matter* **2014**, *26*, 213202.

Graphical TOC Entry