JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

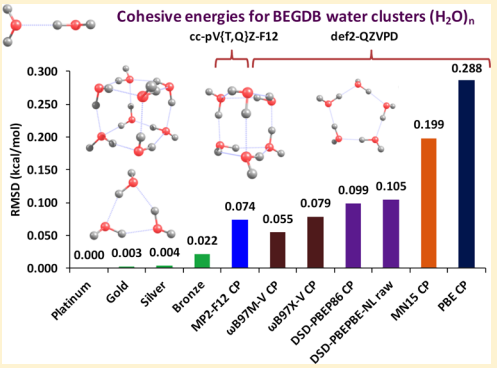
Conventional and Explicitly Correlated ab Initio Benchmark Study on Water Clusters: Revision of the BEGDB and WATER27 Data Sets

Debashree Manna, Manoj K. Kesharwani, Nitai Sylvetsky, and Jan M. L. Martin*

Department of Organic Chemistry, Weizmann Institute of Science, 76100 Rechovot, Israel

Supporting Information

ABSTRACT: Benchmark ab initio energies for BEGDB and WATER27 data sets have been re-examined at the MP2 and CCSD(T) levels with both conventional and explicitly correlated (F12) approaches. The basis set convergence of both conventional and explicitly correlated methods has been investigated in detail, both with and without counterpoise corrections. For the MP2 and CCSD−MP2 contributions, rapid basis set convergence observed with explicitly correlated methods is compared to conventional methods. However, conventional, orbital-based calculations are preferred for the calculation of the (T) term, since it does not benefit from F12. CCSD(F12*) converges somewhat faster with the basis set than CCSD-F12b for the CCSD-MP2 term. The performance of various DFT methods is also evaluated for the BEGDB data set, and results show that Head-Gordon's ωB97X-V and ωB97M-V functionals outperform all other DFT functionals. Counterpoise-corrected DSD-PBEP86 and raw DSD-PBEPBE-NL also perform well and are close to MP2 results. In the WATER27 data set, the anionic (deprotonated) water clusters exhibit unacceptably slow basis set convergence with the regular cc-pV*n*Z-F12 basis sets, which have only diffuse s and p functions. To overcome this, we have constructed modified basis sets, denoted aug-cc-pV*n*Z-F12 or aVnZ-F12, which have been augmented with diffuse functions on the higher angular momenta. The calculated final dissociation energies of BEGDB and WATER27 data sets are available in the Supporting Information. Our best calculated dissociation energies can be reproduced through n-body expansion, provided one pushes to the basis set and electron correlation limit for the two-body term; for the three-body term, post-MP2 contributions (particularly CCSD-MP2) are important for capturing the three-body dispersion effects. Terms beyond four-body can be adequately captured at the MP2-F12 level.



INTRODUCTION

Significant computational effort has been devoted to benchmarking various noncovalently bonded systems (see, e.g., refs 1−8 for recent reviews). Very recently, the general S66x8 benchmark[9,10] has been revised by our group;[11] this benchmark includes hydrogen-bonded species, π-stacking interactions, London dispersion complexes, and mixed-influence situations.

Among hydrogen-bonded systems, water clusters have been the subject of basic scientific interest because they dictate water's bulk properties. Bulk water has many anomalous properties, such as a remarkably high boiling point, low thermal expansion coefficient, unusual density behavior, and so on. Thus, a great many experimental and theoretical studies have been carried out on the structure and properties of water in both the gas and liquid phases.[12−16] Understanding the simplest water clusters is a fundamental step toward understanding complex hydrogen-bonding dynamics, and ab initio quantum mechanical wave function-based methods have provided valuable insight into the structures and energetics of small water clusters.[17−22]

During the last 2 decades, several theoretical studies have been performed on small and medium water clusters, mostly at the Møller−Plesset second-order (MP2) level of perturbation theory, which generally yields good performance for hydrogen-bonded systems (e.g., see ref 11). Some results can also be found at the CCSD(T) level of theory. Extensive theoretical work has been carried out on individual small water clusters.[23−31] At the turn of the millennium, the even-numbered water clusters from $(H_2O)_6$ to $(H_2O)_{20}$ were studied by Day et al.[32] at the MP2 level using a double-ζ basis set. They also reported CCSD(T) energies at MP2 geometries for six structures of the hexamer. Similar quality MP2 binding energies were reported by Kim and co-workers[33,34] and by Maheshwary et al.[35] for $(H_2O)_{2-12}$ and $(H_2O)_{8-20}$, respectively. MP2/CBS (complete basis set) extrapolated limits for the binding energies of the dimer through the ring pentamer and four hexamer isomers using the augmented correlation consistent basis sets (aug-cc-pV*n*Z, *n* = D, T, Q, 5)[36,37] were reported by Xantheas et al.,[38] and B3LYP/aug-cc-pVDZ dissociation energies for clusters up to $(H_2O)_{22}$ were reported by Lenz and co-workers.[39] More recently, Shields et al.[17] also reported MP2/aug-cc-pV*n*Z energies for $(H_2O)_{2-10}$ but with MP2/6-31G* optimized

Received: October 26, 2016
Published: May 22, 2017

DOI: 10.1021/acs.jctc.6b01046
J. Chem. Theory Comput. 2017, 13, 3136−3152

geometries. They extended their study through extrapolation of RI-MP2 energies to their complete basis set limit, included a CCSD(T) correction using a smaller basis set, and added finite temperature corrections within the rigid-rotor-harmonic-oscillator (RRHO) model using scaled and unscaled harmonic vibrational frequencies.[40] Both MP2 and CCSD(T) binding energies with the aug-cc-pVTZ and cc-pVTZ basis sets were reported by Bates et al. for $(H_2O)_{3-10}$.[41] The first CCSD(T) optimized geometries through the hexamer were reported by Xantheas and co-workers.[42] Very recently, Tschumper and co-workers[21] benchmarked the structures and vibrational frequencies for $(H_2O)_n$ (n = 2−6), near the CCSD(T) complete basis set limit. Convergence of the binding energies of water clusters with respect to the level of electron correlation and the size of the basis set has been examined by Xantheas and co-workers[19] for $(H_2O)_n$ (n = 2−6, 8) with basis sets up to quintuple zeta quality at the MP2 and CCSD(T) levels of electron correlation. In addition, Sahu et al. reported CCSD(T)/CBS energies for $(H_2O)_{16-17}$ with MP2/aug-cc-pVTZ optimized geometries.[43]

For the water dimer, Bowman and co-workers[44] reported the accurate HBB2 global potential surface obtained from (a) a fit to almost 30 000 CCSD(T)/AVTZ calculations with counterpoise corrections scaled (by 0.52) to fit the exact zero-point exclusive dissociation energy $D_e[(H_2O)_2]$ = 5.02 kcal/mol for the intermonomer potential; (b) an earlier spectroscopically accurate monomer force field;[45] and (c) asymptotic morphing of the potential into TTM3-F for the long-distance limit.[46] The HBB2 paper also reviews earlier work. The dissociation energy at absolute zero, D_0 = 3.15 kcal/mol, obtained[44] from HBB2 through diffusion Monte Carlo,[47] was later confirmed experimentally by Reisler and co-workers.[48]

For the purpose of training and validation of more approximate methods, such as DFT functionals, two benchmark data sets on water clusters are in use. The first is the water cluster subset of Hobza's BEGDB (benchmark energy and geometry data base);[49] it covers different isomers of neutral water clusters from $(H_2O)_2$ to $(H_2O)_{10}$, with the geometries and energetics having been taken from Shields and co-workers.[11]

The second is the WATER27 component[50] of the GMTKN24[35] and GMTKN30[36] benchmark suites. WATER27 was introduced by Bryantsev et al.;[50] it consists of 10 neutral structures of $(H_2O)_n$ (n = 2−8), 4 isomers of $(H_2O)_{20}$, 5 protonated water clusters $H_3O^+(H_2O)_n$ (n = 1−3, 6), 7 hydrated hydroxide clusters $OH^-(H_2O)_n$ (n = 1−6), and 1 hydroxonium−hydroxide zwitterion, $H_3O^+(H_2O)_6OH^-$. All geometries were optimized at the B3LYP/6-311++G(2d,2p) level of theory. The original benchmark calculations involved MP2/CBS energies, corrected with CCSD(T)/aug-cc-pVDZ high level corrections (HLC) for all isomers except $(H_2O)_{20}$.[50]

In the present study, both the BEGDB and WATER27 data sets have been subjected to an explicitly correlated benchmarking study near the basis set limit. For comparison, conventional CCSD(T) calculations have also been carried out. The effect of the interaction energies (MP2 correlation, CCSD-MP2, and (T) contributions) with and without Boys−Bernardi[53−55] counterpoise corrections on the different components has also been studied in this work. In addition, we will consider n-body expansion schemes, analyze the various contributions in them, and explore the requirements for such schemes to accurately reproduce whole-cluster calculations. Furthermore, we will evaluate the performance of DFT methods for these archetypical H-bonded systems.

COMPUTATIONAL DETAILS

All conventional and explicitly correlated ab initio calculations were carried out using MOLPRO 2015.1.[56] Most DFT and double-hybrid DFT calculations were performed using the ORCA program package,[57] with additional calculations being performed using Q-CHEM 4.4[58] and MRCC.[59] All calculations were performed on the Faculty of Chemistry cluster at the Weizmann Institute of Science.

Explicitly correlated CCSD(T)-F12b and CCSD(T)(F12*) single-point energy calculations were performed using the cc-pVnZ-F12 (where n = D, T, Q, 5) basis sets together with the associated auxiliary and complementary auxiliary (CABS) basis sets.[60,61] The cc-pVnZ-F12 basis set family was specifically developed for explicitly correlated calculations by Peterson et al.[62] As prescribed in ref 63, the geminal exponent (β) values were set to 0.9 for cc-pVDZ-F12 and 1.0 for both the cc-pVTZ-F12 and cc-pVQZ-F12 basis sets. For the cc-pV5Z-F12 basis set, we used β = 1.2, as recommended in ref 64. The SCF component was improved through the "CABS correction".[65,66]

Three different corrections were considered for the (T) term (which does not benefit from F12):

(a) the Marchetti−Werner approximation,[67,68] denoted (T*), in which the (T) contribution is scaled by the MP2-F12/MP2 correlation energy ratio;
(b) analogues, denoted (Tb) and (Tc), in which the (T) contribution is scaled by the respective CCSD-F12b/CCSD and CCSD(F12*)/CCSD correlation energy ratios; and
(c) uniform scaling of the (T) contributions, denoted (Ts),[64] in which the (T) contributions are multiplied by constant scaling factors of 1.1413 for cc-pVDZ-F12 and 1.0527 for cc-pVTZ-F12.

(T*) and (Tb)/(Tc) are not exactly size-consistent, which can be corrected by applying the respective dimer correlation energy ratios to the monomers as well. This is indicated by the notations (T^*_{sc}), (Tb_{sc}), and (Tc_{sc}), respectively.

MP2-F12 results were also obtained as byproducts of the explicitly correlated coupled cluster calculations.

For conventional ab initio calculations, we used Dunning's correlation consistent cc-pVnZ (n = D, T, Q, 5) basis sets and their diffuse-function augmented counterparts (aug-cc-pVnZ).[36,37,69,70] Similar to our previous work,[11] we combined regular cc-pVnZ basis sets on hydrogen with the corresponding diffuse-function augmented basis sets on all other atoms. We denote this haVnZ for short.

Basis set extrapolations were carried out using the two-point formula

$$E_\infty = E(L) + [E(L) - E(L-1)]/\left[\left(\frac{L}{L-1}\right)^\alpha - 1\right] \quad (1)$$

where L is the angular momentum in the basis set and α is an exponent specific to the level of theory and basis set pair. In the present study, basis set extrapolation exponents (α) were taken from the compilation in Table 2 of ref 11. For CCSD(F12*)/cc-pV{D,T}Z, the extrapolation exponent 3.0598 was obtained by following the procedure described in ref 63.

All DFT calculations were performed using the def2-QZVPD basis set,[71] in which diffuse functions are included to the def2-

DOI: 10.1021/acs.jctc.6b01046
J. Chem. Theory Comput. 2017, 13, 3136−3152

QZVP. Respective auxiliary basis sets[72] for simultaneously fitting Coulomb and exchange were used for DFT calculations, and for double-hybrid DFT calculations, additionally associated RI-MP2 auxiliary basis sets[73] were used.

The DFT functionals evaluated include (grouped by rung on the "Jacob's Ladder" of Perdew[74])

- GGAs (second rung): PBE,[75] BP86,[76,77] BLYP[76,78]
- meta-GGAs (third rung): TPSS,[79] TPSS0[80,81]
- hybrid GGAs and hybrid meta-GGAs (fourth rung): B3LYP,[78,82,83] B3PW91,[82,84] PBE0,[85] M06,[86] M06-2X[86]
- double hybrid (fifth rung): B2GP-PLYP[87] and the spin component scaled double hybrids DSD-PBEP86,[88] DSD-PBEPBE,[89] and DSD-PBEB95[89]

We have also assessed the performance of Head-Gordon's range-separated hybrid functionals ωB97X,[90] ωB97X-D3,[91] ωB97X-V,[92] and ωB97M-V.[93] Calculations with Head-Gordon's newly developed functionals ωB97X-V and ωB97M-V were carried out using QChem 4.4.[58]

Further, we have considered empirical dispersion corrections for DFT energies, particularly, Grimme's DFT-D3[94,95] version with Becke–Johnson damping, denoted by the suffix "-D3BJ".

In addition, we have also examined the Vydrov–Van Voorhis (VV10) "nonlocal" (NL) dispersion functional.[96] The required short-range attenuation parameters, b, were obtained from ref 97 for the conventional DFT functionals and from ref 98 for DSD double-hybrid methods. These calculations were carried out using its implementation in ORCA.

The dissociation energies of water clusters were investigated with and without Boys–Bernardi[54] counterpoise corrections. The counterpoise-corrected dissociation energy D_{cp} of dimer AB is defined as

$$D_{cp} = E[A(B)] + E[B(A)] - E[AB] \tag{2}$$

The uncorrected ("raw") dissociation energy D_{raw} is found as

$$D_{raw} = E[A] + E[B] - E[AB] \tag{3}$$

In these equations, $E[A(B)]$ represents the total energy of monomer A in the presence of the basis functions on monomer B, and conversely for $E[B(A)]$. As D_{cp} tends to converge to the basis set limit from the underbinding direction and D_{raw} tends to converge to the basis set limit from the overbinding direction, the average

$$D_{half-half} = (D_{cp} + D_{raw})/2 \tag{4}$$

(i.e., so-called "half-counterpoise"[5,99]) suggests itself as an alternative that exhibits more rapid basis set convergence in both conventional[5] and explicitly correlated[100] calculations. (We note that the empirical CP scaling factor of 0.52, used for the HBB2 surface[44] in order to reproduce $D_e[(H_2O)_2]$, effectively amounts to half-CP.)

The generalization of the counterpoise method to three- and more-body systems is not unique;[101] the most commonly employed generalization, which we shall also apply in the present work, appears to be the site–site function counterpoise method of Wells and Wilson, in which, for a trimer

$$D_{cp} = E[A(B)(C)] + E[(A)B(C)] + E[(A)(B)C] - E[ABC] \tag{5}$$

and similarly for tetramers and larger.

In a recent paper,[11] we proposed the correlation spin polarization index (CSPI) as an indicator of the type of noncovalent bonding character.

$$\mathrm{CSPI} = \frac{D_{e,ab}^{(2)} - D_{e,SS}^{(2)}}{D_{e,ab}^{(2)} + D_{e,SS}^{(2)}} \tag{6}$$

where $D_e^{(2)}$ is the MP2 correlation component of the dissociation energy and $D_{e,ss}^{(2)}$and $D_{e,ab}^{(2)}$are the same-spin and opposite-spin components, respectively. Systems dominated by dispersion have CSPI values close to zero, whereas those with significant electrostatic and induction character will have CSPI values significantly different from zero. Calculated values at the MP2-F12/cc-pVDZ-F12 level can be found in Table S1. The percentages of the dissociation energy accounted for at the Hartree–Fock level are also given there.

$$\%\mathrm{HF} = \frac{100\% \times D_{e,SCF}}{D_{e,SCF} + D_{e,SS}^{(2)} + D_{e,ab}^{(2)} + D_{e,HLC}} \approx \frac{100\% \times D_{e,SCF}}{D_{e,SCF} + D_{e,SS}^{(2)} + D_{e,ab}^{(2)}} \tag{7}$$

Both indices indicate that all water clusters are dominated by electrostatic interactions rather than dispersion.

RESULTS AND DISCUSSION

BEGDB Data Set. Thirty-eight water clusters ranging from water dimers $(H_2O)_2$ to water decamers $(H_2O)_{10}$ were considered.[40] All reference geometries (which were originally obtained[40] at the RI-MP2/aug-cc-pVDZ level) were downloaded from the BEGDB website and used without further optimization; the structures are depicted in Figure 1. It should be noted that the monomer structures in this database are the unrelaxed structures of the monomers within the cluster; hence, these dissociation energies do not include monomer relaxation, and for consistency, neither do ours unless it is explicitly noted otherwise.

For explicitly correlated calculations, due to hardware limitations, CCSD(T)-F12b/cc-pV5Z-F12 calculations were limited to $(H_2O)_{n\leq4}$, CCSD-F12b/cc-pV5Z-F12 and CCSD-(T)-F12b/cc-pVQZ-F12 calculations were limited to $(H_2O)_{n\leq6}$, and CCSD-F12b/cc-pVQZ-F12 calculations were limited to $(H_2O)_{n\leq8}$.

For conventional CCSD(T) calculations, we were unable to compute beyond $(H_2O)_7$ for the haVQZ basis set and beyond $(H_2O)_6$ for the haV5Z basis set. However, we were able to perform conventional CCSD calculations for the two $(H_2O)_8$ structures with the haVQZ basis set.

Additionally, we also performed CCSD(T)(F12*) calculations for all 38 water clusters with the cc-pVDZ-F12 and cc-pVTZ-F12 basis sets. We also considered $(H_2O)_{n\leq4}$ at the CCSD(F12*)/cc-pVQZ-F12 level, but we found that the CCSD-F12b and CCSD(F12*) results are functionally equivalent with this basis set. This is consistent with earlier findings[64,102] that the CCSD(F12*)–CCSD-F12b difference converges rapidly with the basis set.

RMSDs (root-mean-square differences) for the dissociation energies, by their very nature, will be dominated by the largest species. In order to avoid skewing of this nature, we will instead consider the RMSD on twice the cohesive energy, $2(nE[H_2O] - E[(H_2O)_n])/n = 2D_e/n$, which brings everything on the same scale as the water dimer.

DOI: 10.1021/acs.jctc.6b01046
J. Chem. Theory Comput. 2017, 13, 3136–3152

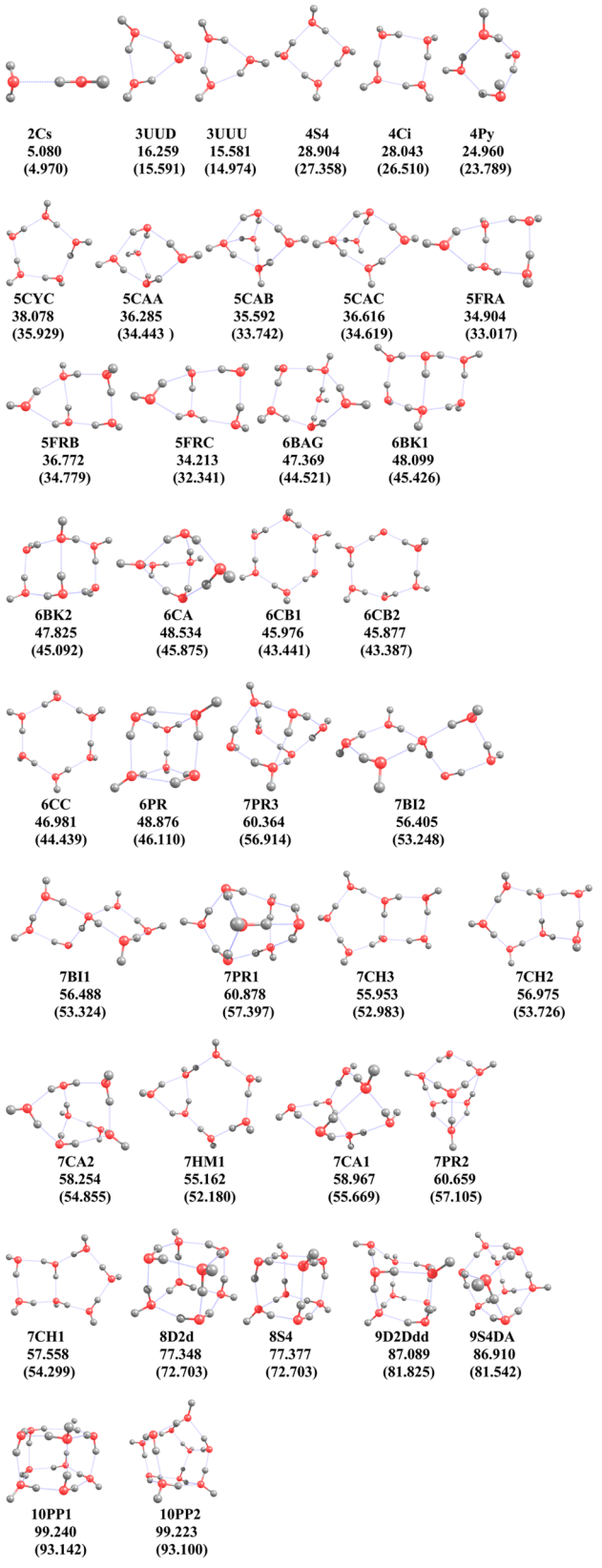


Figure 1. RI-MP2/aug-cc-pVDZ optimized geometries of $(H_2O)_{2-10}$ isomers (BEGDB data set) reported by Shields and co-workers.[11] Dissociation energies (kcal/mol) with our best level [PLATINUM: through $(H_2O)_6$; GOLD: $(H_2O)_7$; and SILVER: $(H_2O)_{n=8-10}$] are also given exclusive (and, in parentheses, inclusive) of monomer relaxation energies.

SCF and MP2 Components. First, we shall discuss the Hartree−Fock contribution including the CABS correction. Through $(H_2O)_6$, counterpoise-corrected HF+CABS/cc-pVQZ-F12 yields results essentially identical to the corresponding HF+CABS/cc-pV5Z-F12 values; hence, we used the former level of theory as the reference, as results are available through $(H_2O)_{10}$. With CABS corrections, basis set convergence is much faster than in conventional HF calculations (Table S2). Moreover, for all HF+CABS/cc-pV*n*Z-F12 (n = D, T, Q) calculations, full counterpoise correction is the closest to the basis set limit, as expected, and as also found for S66x8.[11] However, conventional half-counterpoise corrected haVTZ values are quite close to the basis set limit.

Let us next consider the MP2-F12 correlation component. RMSD values are reported in Table 1. As the reference, we have chosen half-counterpoise corrected cc-pV{Q,5}Z-F12, which is available through $(H_2O)_6$. RMS deviations for cc-pV{Q,5}Z-F12 raw and counterpoise-corrected results are just 0.004 kcal/mol. Counterpoise-corrected and half-counterpoise corrected MP2-F12/cc-pV{T,Q}Z-F12 results are close to the reference, with RMSD values of 0.006 and 0.007 kcal/mol, respectively. With half-counterpoise, even MP2-F12/cc-pVTZ-F12 performs very well relative to our reference (RMSD = 0.002 kcal/mol).

Table 1. RMS Deviations (kcal/mol) for the MP2 Correlation Component of Water Cluster (BEGDB) Cohesive Energies[a] Calculated with Various Basis Sets Using Conventional and Explicitly Correlated Methods

	through $(H_2O)_6$ relative to MP2-F12/cc-pV{Q,5}Z-F12 half		
	CP	raw	half
MP2/haVDZ	1.851	0.615	1.230
MP2/haVTZ	0.816	0.034	0.421
MP2/haVQZ	0.374	0.013	0.183
MP2/haV5Z	0.203	0.009	0.098
MP2/haV{T,Q}Z	0.039	0.042	0.040
MP2/haV{Q,5}Z	0.004	0.007	0.002
MP2-F12/cc-pVDZ-F12	0.142	0.068	0.038
MP2-F12/cc-pVTZ-F12	0.051	0.049	0.002
MP2-F12/cc-pVQZ-F12	0.019	0.027	0.004
MP2-F12/cc-pV5Z-F12	0.009	0.012	0.001
MP2-F12/cc-pV{D,T}Z-F12	0.015	0.041	0.013
MP2-F12/cc-pV{T,Q}Z-F12	0.006	0.019	0.007
MP2-F12/cc-pV{Q,5}Z-F12	0.004	0.004	REF

[a]The cohesive energy in this and the following tables is defined as $\frac{2}{n}(\sum_{i=1}^{n} E(\text{monomer})_i - E(\text{cluster}))$.

Performance of conventional MP2/haV*n*Z is somewhat disappointing: The fact that the RMSD is smallest for raw, rather than counterpoise or half−half, can be seen[100] as indicating that one is still some distance away from basis set convergence. Extrapolation from at least MP2/haV{T,Q}Z results is required for acceptable performance.

Post-MP2 Corrections. The CCSD(T)−MP2 difference is usually discussed as a single "high-level correction" (HLC). In

DOI: 10.1021/acs.jctc.6b01046
J. Chem. Theory Comput. 2017, 13, 3136−3152

fact, in the context of explicitly correlated calculations, it may be advantageous to partition the HLC into two terms, HLC = [CCSD−MP2] + (T); the former benefits directly from F12, unlike the latter.

We will first focus on the CCSD−MP2 term. We were able to complete CCSD-F12b/cc-pV5Z-F12 calculations with counterpoise corrections through $(H_2O)_4$, and without counterpoise corrections through $(H_2O)_6$. (The latter required NFS-mounting the scratch file system of a second node over Infiniband, as the scratch files exceeded the 6TB SSD available on our largest nodes.) Uncorrected and counterpoise-corrected [CCSD-F12b−MP2-F12]/cc-pV{Q,5}Z-F12 results are effectively indistinguishable (RMS difference = 0.001 kcal/mol; Table 2), so we considered raw [CCSD-F12b−MP2-F12]/cc-pV{Q,5}Z-F12 as the reference.

Table 2. RMS Deviations (kcal/mol) for the CCSD−MP2 Components from the Basis Set Limit Values of the Water Cluster (BEGDB) Cohesive Energies Calculated with Various Basis Sets

	relative to [CCSD-F12b−MP2-F12]/cc-pV{Q,5}Z-F12 raw		
through $(H_2O)_6$	CP	raw	half
CCSD − MP2 (orbital based)			
haVDZ	0.042	0.133	0.087
haVTZ	0.067	0.021	0.025
haVQZ	0.083	0.018	0.033
haV5Z	0.058	0.022	0.018
haV{T,Q}Z	0.009	0.026	0.018
haV{Q,5}Z	0.010	0.025	0.017
CCSD-F12b−MP2-F12			
cc-pVDZ-F12	0.160	0.130	0.145
cc-pVTZ-F12	0.052	0.036	0.044
cc-pVQZ-F12	0.008	0.022	0.015
cc-pV5Z-F12		0.007	
cc-pV{D,T}Z-F12	0.011	0.002	0.005
cc-pV{T,Q}Z-F12	0.007	0.016	0.005
cc-pV{Q,5}Z-F12		REF	
CCSD(F12*)−MP2-F12			
cc-pVDZ-F12	0.080	0.072	0.076
cc-pVTZ-F12	0.023	0.031	0.027
cc-pVQZ-F12	0.010	0.026	0.018
cc-pV{D,T}Z-F12	0.002	0.015	0.007
through $(H_2O)_4$			
CCSD-F12b−MP2-F12			
cc-pV{Q,5}Z-F12	0.001	REF	0.000

As can be seen in Table 2, for [CCSD-F12b−MP2-F12], the cc-pV{Q,5}Z-F12 and cc-pV{T,Q}Z-F12 extrapolated values are very close to each other, differing by just 0.007 kcal/mol RMS with full counterpoise and 0.005 kcal/mol RMS with half-counterpoise.

CCSD(F12*)/cc-pV*n*Z-F12 for *n* = D, T cuts the RMSD in about half compared to CCSD-F12b; particularly for cc-pVDZ-F12, CCSD(F12*) is recommended.

Both CCSD(F12*)/cc-pV{D,T}Z-F12 and CCSD-F12b/cc-pV{D,T}Z-F12 extrapolations come very close to the basis set limit, especially with half-counterpoise. For cc-pVQZ-F12, the two approximations, F12b and (F12*), yield functionally equivalent results.

For conventional calculations with the haVTZ and haVQZ basis sets, the raw results are slightly better than half- or full-counterpoise corrected ones, owing to error compensation between BSSE and IBSI (intrinsic basis set insufficiency).[100]

At the request of a reviewer, we investigated the performance of an intermediate between CCSD(F12*) and CCSD-F12b. The latter introduces two approximations over CCSD(F12*): (a) neglect of the CABS contribution in the projector, which occurs in the dominant CCSD-F12 coupling terms, and (b) neglect of coupling terms that require double RI approximations.

Approximation (a) can be suppressed in MOLPRO by combining CCSD-F12b with the option IXPROJ = 1, leaving only (b) in place. The same option can, in fact, be applied to CCSD-F12a,[65] which omits additional terms compared to CCSD-F12b. The claim has in fact been made (see, e.g., ref 103 and references therein) that CCSD-F12a with small basis sets is more accurate than the more rigorous CCSD-F12b or even CCSD(F12*) (a.k.a., CCSD-F12c) models. In an attempt to address both questions, we calculated CCSD-F12a and CCSD-F12b dissociation energies, with IXPROJ = 1 as well as the default of IXPROJ = 0, for $(H_2O)_n$ (n = 2−10) with the cc-pVDZ-F12 basis set, for n = 2−6 with the cc-pVTZ-F12 basis set, and for n = 2−4 with the cc-pVQZ-F12 basis sets. The RMS and mean signed differences in the cohesive energies are given in Tables S3−S6.

First, the difference between IXPROJ = 1 and IXPROJ = 0 decays very rapidly with the basis set, dropping to 0.003 kcal/mol RMS for cc-pVTZ-F12 and becoming essentially nil for cc-pVQZ-F12. For cc-pVDZ-F12, however, the difference is fairly significant, particularly for CCSD-F12b (RMSD 0.059 kcal/mol). Indeed, IXPROJ = 1 bridges most of the gap between CCSD-F12b/cc-pVDZ-F12 and CCSD(F12*)/cc-pVDZ-F12 for the present systems (RMSD 0.067 kcal/mol for IXPROJ = 0 and 0.013 kcal/mol for IXPROJ = 1). For the cc-pVTZ-F12 basis set, IXPROJ = 1 still brings the CCSD-F12b results closer to the most rigorous CCSD(F12*) model, but the improvement is an order of magnitude smaller than the residual error (RMSD 0.006 kcal/mol for IXPROJ = 0 and 0.004 kcal/mol for IXPROJ = 1). With the cc-pVQZ-F12 basis set, CCSD-F12b and CCSD(F12*) yield nearly identical results in any case; note that, even for such large basis sets, CCSD-F12a still displays a small but (compared to the residual basis set incompleteness error) significant difference with the two other approaches (Table S3).

Considering the mean signed differences (MSD) from CCSD(F12*) with the various basis sets, it can be seen that CCSD-F12a systematically overbinds, whereas CCSD-F12b underbinds (Table S4). For cc-pVDZ-F12, IXPROJ = 1 actually degrades CCSD-F12a further (MSD goes from +0.056 to +0.068 kcal/mol), whereas it largely corrects the problem for CCSD-F12b: F12b/(IXPROJ = 0) = −0.065; F12b/(IXPROJ = 1) = −0.007 kcal/mol.

For cc-pVTZ-F12, CCSD-F12a on average overbinds and CCSD-F12b underbinds by similar amounts (MSD = +0.037 and −0.029 kcal/mol, respectively), whereas for cc-pVQZ-F12, as already mentioned, CCSD-F12a still overbinds (MSD = +0.035 kcal) and CCSD-F12b becomes functionally equivalent to CCSD(F12*).

It is thus seen that the claim that CCSD-F12a is superior for small basis sets rests on a fortunate error compensation with

DOI: 10.1021/acs.jctc.6b01046
J. Chem. Theory Comput. 2017, 13, 3136−3152

Table 3. RMS Deviations (kcal/mol) for the (T) Term of Conventional CCSD(T) Calculated for Water Cluster (BEGDB) Cohesive Energies with Various Basis Sets

through $(H_2O)_6$	relative to haV{Q,5}Z half						
	haVDZ	haVTZ	haVQZ	haV5Z	haV{D,T}Z	haV{T,Q}Z	haV{Q,5}Z
CP	0.252	0.088	0.034	0.015	0.020	0.006	0.000
raw	0.031	0.009	0.006	0.003	0.003	0.005	0.000
half	0.136	0.048	0.020	0.009	0.011	0.001	REF

Table 4. RMS Deviations (kcal/mol) for the (T) Term of Explicitly Correlated CCSD(T)-F12 Calculated for Water Cluster (BEGDB) Cohesive Energies with Various Basis Sets

	CP					raw					half				
	(T*)	(T)	(Ts)	(T^*_{sc})	($T(b/c)_{sc}$)	(T*)	(T)	(Ts)	(T^*_{sc})	($T(b/c)_{sc}$)	(T*)	(T)	(Ts)	(T^*_{sc})	($T(b/c)_{sc}$)
through $(H_2O)_6$ (relative to [CCSD(T)–CCSD]/haV{Q,5}Z half CP)															
F12b/cc-pVDZ-F12	0.081	0.232	0.153	0.106	0.124	0.088	0.071	0.036	0.095	0.071	0.010	0.151	0.060	0.011	0.029
F12b/cc-pVTZ-F12	0.004	0.073	0.035	0.006	0.015	0.061	0.005	0.042	0.075	0.064	0.032	0.037	0.005	0.035	0.025
F12b/cc-pVQZ-F12	0.009	0.029	0.011	0.006	0.001	0.035	0.007	0.026	0.043	0.036	0.022	0.011	0.007	0.024	0.018
(F12*)/cc-pVDZ-F12	0.095	0.245	0.166	0.121	0.134	0.069	0.086	0.021	0.076	0.059	0.017	0.165	0.076	0.025	0.040
(F12*)/cc-pVTZ-F12	0.007	0.081	0.044	0.014	0.024	0.050	0.011	0.033	0.065	0.054	0.022	0.046	0.007	0.025	0.015
through $(H_2O)_4$ (relative to [CCSD(T)–CCSD]/haV{Q,5}Z half CP)															
F12b/cc-pV5Z-F12	0.007	0.011	0.002	0.006	0.002	0.014	0.003	0.011	0.019	0.016	0.010	0.004	0.005	0.012	0.009
(F12*)/cc-pVQZ-F12	0.003	0.028	0.014	0.001	0.005	0.025	0.002	0.017	0.031	0.026	0.014	0.013	0.002	0.016	0.011

Table 5. RMS Deviations (kcal/mol) for the High-Level Corrections (HLC = [(CCSD(T)−MP2)/haV*n*Z)] Components of the Water Cluster (BEGDB) Cohesive Energies

through $(H_2O)_6$	(relative to PLATINUM reference[a])						
	haVDZ	haVTZ	haVQZ	haV5Z	haV{D,T}Z	haV{T,Q}Z	haV{Q,5}Z
CP	0.289	0.021	0.049	0.043	0.091	0.100	0.038
raw	0.148	0.025	0.024	0.025	0.028	0.023	0.025
half	0.218	0.023	0.013	0.009	0.059	0.039	0.006

[a][CCSD-F12b−MP2-F12]/cc-pV{Q,5}Z-F12 raw + [CCSD(T)−CCSD]/haV{Q,5}Z half CP: "PLATINUM".

basis set incompleteness error. While the authors acknowledge that some reliance on error compensation (a.k.a. "Pauling points", a.k.a. "right answer for the wrong reason") is a quantum chemical fact of life, we would advocate seeking "the right answer for the right reason" where possible.[104] In fact, it was previously noted[105] that, with small basis sets, CCSD-F12a performs better for hydrogen-bonded noncovalent interactions and CCSD-F12b performs better for dispersion-dominant cases; in that paper,[105] a "dispersion-weighted" approach was proposed in which the mixing coefficient is made a parametric function of the electrostatic vs dispersive character of the complex. Clearly, reliance on such artifices blurs the line between ab initio and semiempirical approaches and obviates the chief advantage of the former.

We now turn to the connected triple excitations contribution (T). RMS deviations for conventional (T) calculations are given in Table 3.

It is seen immediately that haV{T,Q}Z with half-counterpoise and haV{Q,5}Z with or without counterpoise yield essentially indistinguishable results for $(H_2O)_{2-6}$. The insensitivity toward counterpoise for haV{Q,5}Z is especially reassuring concerning the quality of the extrapolation when considering that BSSE should vanish at the one-particle basis set limit.

We have chosen the half-counterpoise haV{Q,5}Z data as the reference. For raw and full-counterpoise (T)/haV{T,Q}Z, the RMSDs are just 0.005 and 0.006 kcal/mol, respectively. Perhaps most noteworthy among the other results is the excellent performance of raw (T)/haV{D,T}Z (RMSD 0.003 kcal/mol), which is fairly inexpensive and can easily be carried out for larger clusters as well. Counterpoise correction is actually counterproductive for this contribution, at least with smaller basis sets; for larger basis sets, its inclusion or lack thereof does not seem to matter.

The (T) contribution does not benefit from F12 directly. A number of scaling schemes for the connected triples have been introduced.[11,64,67] Somewhat surprisingly, owing to a felicitous error compensation, unscaled CCSD(T)-F12b/cc-pVTZ-F12 without counterpoise yields excellent results, comparable in quality to conventional (T)/haV{D,T}Z (Table 4). While the cost scaling is roughly the same, (T)-F12b can be a byproduct of the CCSD(T)-F12b calculation, rather than requiring a

DOI: 10.1021/acs.jctc.6b01046
J. Chem. Theory Comput. 2017, 13, 3136−3152

Table 6. RMS Deviations (kcal/mol) for the High-Level Correction (HLC = [(CCSD(T)-F12−MP2-F12)/cc-pV*n*Z-F12)]) Components of the Water Cluster (BEGDB) Cohesive Energies

	CP					raw					half				
	HLC (T*)	HLC (T)	HLC (Ts)	HLC (T^*_{sc})	HLC ($T(b/c)_{sc}$)	HLC (T*)	HLC (T)	HLC (Ts)	HLC (T^*_{sc})	HLC ($T(b/c)_{sc}$)	HLC (T*)	HLC (T)	HLC (Ts)	HLC (T^*_{sc})	HLC ($T(b/c)_{sc}$)
***through* $(H_2O)_6$**	(relative to PLATINUM reference[a])														
F12b/cc-pVDZ-F12	0.240	0.392	0.312	0.265	0.284	0.043	0.199	0.097	0.037	0.060	0.141	0.295	0.204	0.151	0.172
F12b/cc-pVTZ-F12	0.049	0.125	0.087	0.057	0.067	0.025	0.036	0.007	0.039	0.028	0.013	0.083	0.041	0.010	0.020
F12b/cc-pVQZ-F12	0.002	0.036	0.019	0.002	0.009	0.013	0.015	0.004	0.021	0.014	0.007	0.026	0.008	0.010	0.003
(F12*)/cc-pVDZ-F12	0.175	0.325	0.246	0.201	0.214	0.020	0.158	0.060	0.021	0.025	0.091	0.241	0.152	0.100	0.115
(F12*)/cc-pVTZ-F12	0.029	0.104	0.066	0.037	0.046	0.020	0.041	0.007	0.034	0.024	0.007	0.072	0.033	0.005	0.012
***through* $(H_2O)_4$**	(relative to PLATINUM reference[a])														
F12b/cc-pV5Z-F12	0.006	0.011	0.002	0.005	0.002	0.007	0.004	0.005	0.013	0.010	0.007	0.007	0.001	0.009	0.006
(F12*)/cc-pVQZ-F12	0.005	0.035	0.021	0.007	0.012	0.004	0.019	0.004	0.010	0.005	0.001	0.027	0.013	0.002	0.004

[a][CCSD-F12b−MP2-F12]/cc-pV{Q,5}Z-F12 raw + [CCSD(T)−CCSD]/haV{Q,5}Z half CP: "PLATINUM".

separate conventional CCSD(T) calculation. The cheapest F12-scaled alternative available appears to be raw (Ts)/cc-pVDZ-F12, although half−half (T*)/cc-pVDZ-F12 and (T^*_{sc})/cc-pVDZ-F12 perform still better, at the expense of requiring the additional "ghost" calculations (Table 4).

There is little to choose between the CCSD-F12b and CCSD(F12*) ansätze for the triples. If one wishes to rely entirely on F12 calculations rather than a separate conventional (T) step, CCSD(F12*) may be preferable overall because of its faster basis set convergence on the CCSD−MP2 difference.

What if we instead consider the combined CCSD(T)−MP2 difference, i.e., the HLC? Water clusters appear to be a very good test system for this. As the reference for $(H_2O)_{n\leq 6}$, we will combine the CCSD-F12b/cc-pV{Q,5}Z-F12 raw results for CCSD−MP2 with the half-counterpoise corrected (T)/haV-{Q,5}Z contribution. The results are presented in Table 5.

It would appear that in the conventional calculations, half-counterpoise correction is better for large basis sets (haV5Z and haVQZ). However, counterpoise correction seems ineffective for haVTZ, whereas counterpoise-corrected haV-{D,T}Z, which was used in the S66 paper,[9] is actually worse than its uncorrected values. In the S66x8 paper, the HLC was obtained from [CCSD(T)-MP2]/haVDZ, which here is clearly "weighed in the balance and found wanting".[10]

Similar error compensation is also at work in the explicitly correlated HLCs. Raw HLC(F12*) (Tc_{sc})/cc-pVDZ-F12, at RMSD = 0.025 kcal/mol, yields better results (Table 6); it was recently used[11] in the re-evaluation of the S66x8 data set. (The selection of that level was based there on comparison with cc-pVTZ-F12 HLCs for a subset of systems.) Here, we find that for the cc-pVTZ-F12 basis set, there is relatively little variation between different (T) F12 scaling strategies, but the small RMSDs for half-counterpoise and HLC(T^*_{sc}) or HLC(Tb_{sc}) stand out. For cc-pVQZ-F12, all counterpoise and (T) F12 scaling choices yield excellent RMSDs, with the exception of unscaled (T) with full counterpoise. Finally, for the limited cc-pV5Z-F12 data, even unscaled (T) can be used.

We have also considered a composite HLC strategy in which conventional (T) is combined with CCSD−F12b or CCSD-(F12*) for the CCSD−MP2 part. RMSDs are reported in Table 7. The references used here are same as in Tables 5 and 6. Table 7 indicates that the RMS deviations for [CCSD-F12b−MP2-F12]/cc-pV{D,T}Z-F12 (both half-CP and raw) and [CCSD(F12*)−MP2-F12]/cc-pV{D,T}Z-F12 (half-CP) combined with orbital-based (T)/haV{D,T}Z (raw) are close to the best available results, and any of these combinations can be used as a cost-effective and accurate option for HLC. However, here we prefer CCSD(F12*) over CCSD-F12b because of the former's faster basis set convergence and more rigorous character, except for the very largest basis set, cc-pV5Z-F12, where the two approaches are expected to yield functionally identical results and even the CCSD-F12b calculations for $(H_2O)_6$ already exceeded the 6TB local scratch storage limits on our largest compute nodes.

All of this suggests a hierarchy of four approaches, ordered by decreasing computational cost:

Table 7. RMS Deviations (kcal/mol) for the High-Level Correction (HLC) Components from the Basis Set Limit Values of the Water Cluster (BEGDB) Cohesive Energies Calculated from Combining an Explicitly Correlated (CCSD−MP2 part) with an Orbital-Based (T) Component

through $(H_2O)_6$			
[CCSD-F12−MP2-F12] half CP	**Orbital based (T)**	**(T) half CP**	**(T) raw**
(F12b)/cc-pVDZ-F12	haVDZ	0.279	0.162
(F12b)/cc-pVTZ-F12	haVTZ	0.091	0.051
(F12b)/cc-pVQZ-F12	haVQZ	0.035	0.021
(F12b)/cc-pV{D,T}Z-F12	haV{D,T}Z	0.016	0.008
(F12b)/cc-pV{T,Q}Z-F12	haV{T,Q}Z	0.004	0.010
(F12*)/cc-pVDZ-F12	haVDZ	0.211	0.098
(F12*)/cc-pVTZ-F12	haVTZ	0.075	0.034
(F12*)/cc-pV{D,T}Z-F12	haV{D,T}Z	0.018	0.009
[CCSD-F12−MP2-F12] raw	**Orbital based (T)**	**(T) half**	**(T) raw**
(F12b)/cc-pVDZ-F12	haVDZ	0.264	0.148
(F12b)/cc-pVTZ-F12	haVTZ	0.084	0.043
(F12b)/cc-pVQZ-F12	haVQZ	0.042	0.028
(F12b)/cc-pV5Z-F12	haV5Z	0.016	0.010
(F12b)/cc-pV{D,T}Z-F12	haV{D,T}Z	0.011	0.003
(F12b)/cc-pV{T,Q}Z-F12	haV{T,Q}Z	0.016	0.022
(F12b)/cc-pV{Q,5}Z-F12	haV{Q,5}Z	REF	0.000
(F12*)/cc-pVDZ-F12	haVDZ	0.208	0.095
(F12*)/cc-pVTZ-F12	haVTZ	0.079	0.039
(F12*)/cc-pV{D,T}Z-F12	haV{D,T}Z	0.026	0.017

DOI: 10.1021/acs.jctc.6b01046
J. Chem. Theory Comput. 2017, 13, 3136−3152

Table 8. RMSDs (kcal/mol) of the DFT[a] Cohesive Energies Relative[b] to the Best Available Reference[c] for the (BEGDB) Water Clusters

	No DISP			D3BJ			NL		
	raw	CP	half	raw	CP	half	raw	CP	half
B2GPPLYP	0.16	0.42	0.25	0.65	0.27	0.45	1.02	0.63	0.82
DSD-PBEP86	0.36	0.10	0.19	0.53	0.18	0.36	0.49	0.14	0.32
DSD-PBEPBE	0.40	0.76	0.58	0.16	0.51	0.33	0.10	0.32	0.15
DSD-PBEB95	0.70	1.02	0.86	0.41	0.72	0.56	0.16	0.62	0.24
BP86	1.99	2.01	2.00	0.31	0.30	0.30	1.48	1.47	1.47
PBE	0.28	0.29	0.28	1.33	1.31	1.32	1.93	1.91	1.92
BLYP	2.89	2.91	2.90	0.26	0.27	0.26	1.07	1.05	1.06
TPSS	1.52	1.56	1.54	0.38	0.34	0.36	1.66	1.62	1.64
TPSS0	1.18	1.21	1.20	0.62	0.59	0.60	1.32	1.28	1.30
B3LYP	1.59	1.61	1.60	0.63	0.61	0.62	1.45	1.43	1.44
B3PW91	2.58	2.60	2.59	0.34	0.35	0.34	0.71	0.69	0.70
PBE0	0.26	0.27	0.26	1.22	1.20	1.21	1.73	1.71	1.72
M06	0.49	0.70	0.59						
M06-2X	0.49	0.45	0.47						
	ωB97X			ωB97X-D3			ωB97X-V		
	2.06	2.08	2.07	0.53	0.50	0.52	0.10	0.08	0.09
	MN15						ωB97M-V		
	0.38	0.20	0.24				0.15	0.05	0.10
							dRPA75		
							0.17	0.70	0.27

[a]The def2-QZVPD basis set was used throughout, except for ωB97M-V and dRPA75, where the aug-cc-pVQZ and haVQZ basis sets were used, respectively. [b]By way of comparison, the statistics for MP2-F12/cc-pVQZ-F12 are raw = 0.13 kcal/mol, CP = 0.07 kcal/mol, half−half = 0.10 kcal/mol; these errors reflect the neglect of HLC. [c]PLATINUM: through $(H_2O)_6$; GOLD: $(H_2O)_7$; and SILVER: $(H_2O)_{n=8-10}$

- PLATINUM: CCSD-F12b/cc-pV{Q,5}Z raw combined with conventional (T)/haV{Q,5}Z half CP (results available through $(H_2O)_6$)
- GOLD: CCSD-F12b/cc-pV{T,Q}Z half-CP combined with conventional (T)/haV{T,Q}Z half CP (results available through $(H_2O)_7$)
- SILVER: MP2-F12/cc-pV{T,Q}Z-F12 half-CP combined with HLC from [CCSD(F12*)−MP2-F12]/cc-pV{D,T}Z-F12 half-CP and conventional (T)/haV-{D,T}Z raw (results available through $(H_2O)_{10}$)
- BRONZE: MP2-F12/cc-pV{T,Q}Z-F12 half-CP combined with HLC from CCSD(F12*)(Tc_{sc})/cc-pVDZ-F12 raw; this is the proposed approximation for larger species when n-body decomposition is impractical (see below).

Comparison with Earlier Benchmark Studies. Here, it would be appropriate to compare our ab initio results with the most recent available literature values of Shields and co-workers.[40] For their dissociation energy calculations, they used the optimum geometry of water monomer rather than the monomer-in-cluster geometries; this means that we have to consider monomer relaxation energies to bring everything on the same scale. Our MP2 limit dissociation energies (MP2-F12/cc-pV{Q,5}Z-F12 raw through $(H_2O)_6$; MP2-F12/cc-pV{T,Q}Z-F12 half-CP for $(H_2O)_{n=7-10}$) differ from their reported MP2 limits by 0.180 kcal/mol RMS; their values were obtained from RI-MP2/aug-cc-pV{D,T,Q}Z results using an $A + B/L^4 + C/L^5$ extrapolation formula, without employing any counterpoise correction. Further, their HLCs were obtained at the conventional CCSD(T)/aug-cc-pVDZ level, whereas our best HLCs are ([CCSD-F12b−MP2-F12]/cc-pV{Q,5}Z-F12 raw + (T)/haV{Q,5}Z half-CP through $(H_2O)_6$, [CCSD-F12b−MP2-F12]/cc-pV{T,Q}Z-F12 half-CP + (T)/haV-{T,Q}Z half-CP for $(H_2O)_7$, and [CCSD(F12*)−MP2-F12]/cc-pV{D,T}Z-F12 half-CP + (T)/haV{D,T}Z raw for $(H_2O)_{n=8-10}$). Their HLCs differ from ours by 0.119 kcal/mol, which is actually slightly more than the discrepancy of 0.091 kcal/mol RMS between the dissociation energies themselves. Another paper was published by the same group,[18] in which they have further discussed the HLCs obtained from variants of explicitly correlated CCSD(T)-F12 with the cc-pVDZ-F12 basis set. Their reported $HLC_{(T^*)\text{-F12b}}$ differs from our best HLCs by 0.033 kcal/mol RMS and the corresponding $HLC_{(T^*)\text{-F12a}}$ correction differs by 0.468 kcal/mol RMS, which clearly shows the advantage of the F12b approximation over F12a (see above).

One might question the adequacy of the level of theory originally used (RI-MP2/aug-cc-pVDZ) for the optimization of water clusters. Therefore, we have reoptimized all geometries at the RI-MP2/haVTZ level using the RI-MP2 analytical gradient code in ORCA. All reoptimized geometries are almost similar to the original geometries, except for 4Py, 5CAA, and 7CH3, which converged to 4Ci, 5CAC, and 7CH2, respectively. The RMS deviation for cohesive energies calculated with the RI-MP2/haVTZ method is just 0.073 kcal/mol relative to the RI-MP2/haVTZ//RI-MP2/aug-cc-pVDZ calculated values, which suggests that conclusions concerning the basis set convergence, n-particle convergence, or performance of DFT functionals will not be materially affected by the use of reoptimized geometries.

Performance of DFT Functionals. We will now discuss the performance of various density functionals. The best available data (PLATINUM: through $(H_2O)_6$; GOLD: $(H_2O)_7$;

DOI: 10.1021/acs.jctc.6b01046
J. Chem. Theory Comput. 2017, 13, 3136−3152

Table 9. n-Body Dissociation Energies (kcal/mol) of Water Clusters through $(H_2O)_6$ with the Largest Basis Sets Considered[a]

	MP2-F12				CCSD(F12*)−MP2-F12			CCSD(T)−CCSD		
	cc-pV5Z-F12 two-body	cc-pVQZ-F12 three-body	cc-pVQZ-F12 four-body	cc-pVQZ-F12 five-body	cc-pV5Z-F12 two-body	cc-pVQZ-F12 three-body	cc-pVTZ-F12 four-body	haV5Z two-body	haVQZ three-body	haVTZ four-body
$(H_2O)_2$										
2Cs	5.072				−0.245			0.251		
$(H_2O)_3$					0.000			0.000		
3UUD	13.789	2.534			−0.942	−0.060		0.923	0.015	
3UUU	13.386	2.325			−0.964	−0.049		0.869	0.013	
$(H_2O)_4$										
4S4	22.338	6.274	0.561		−1.677	−0.095	−0.003	1.446	0.038	0.022
4Ci	21.782	5.997	0.545		−1.692	−0.085	−0.004	1.436	0.038	0.022
4Py	21.029	3.848	0.059		−1.351	−0.144	−0.001	1.507	0.005	0.002
$(H_2O)_5$										
5CYC	28.051	9.149	1.208	0.091	−2.236	−0.065	−0.017	1.798	0.052	0.045
5CAA	29.807	6.188	0.270	0.001	−1.947	−0.220	−0.009	2.152	0.019	0.006
5CAB	29.178	6.041	0.290	−0.015	−1.846	−0.224	−0.007	2.138	0.019	0.010
5CAC	29.494	6.639	0.458	−0.030	−1.891	−0.239	−0.007	2.150	0.013	0.018
5FRA	28.346	6.408	0.262	−0.022	−2.025	−0.170	0.002	2.054	0.047	0.008
5FRB	28.446	7.711	0.737	0.010	−1.963	−0.167	−0.016	1.946	0.031	0.030
5FRC	27.729	6.369	0.222	−0.022	−2.005	−0.148	0.005	2.016	0.046	0.008
$(H_2O)_6$										
6BAG	36.196	10.423	1.095	0.019	−2.832	−0.193	−0.023	2.585	0.048	0.046
6BK1	36.970	10.439	1.022	0.038	−2.821	−0.167	−0.015	2.519	0.063	0.043
6BK2	37.054	10.171	0.958	0.015	−2.843	−0.184	−0.016	2.549	0.066	0.040
6CA	39.059	9.161	0.490	−0.011	−2.792	−0.282	−0.015	2.861	0.030	0.015
6CB1	33.454	11.317	1.582	0.156	−2.703	−0.062	−0.026	2.124	0.064	0.060
6CB2	33.322	11.349	1.563	0.157	−2.656	−0.062	−0.024	2.106	0.059	0.056
6CC	33.832	11.754	1.738	0.186	−2.688	−0.050	−0.030	2.103	0.063	0.062
6PR	39.221	9.049	0.645	−0.076	−2.595	−0.323	−0.023	2.898	0.026	0.018

[a]For results with smaller basis sets please see Table S9.

and SILVER: $(H_2O)_{n=8-10}$) were used as reference. Statistics are given in Table 8. For some perspective, it needs to be kept in mind that even straight MP2 can achieve RMSD = 0.07 kcal/mol: of the many functionals we considered, only Mardirossian and Head-Gordon's ωB97M-V[93] outperform MP2, whereas counterpoise corrected ωB97X-V,[92] DSD-PBEP86, and raw DSD-PBEPBE-NL at least come close. The very recent MN-15 functional,[106,107] which is a range-separated hybrid meta-GGA like ωB97M-V but does not include any explicit dispersion term and involves a much larger number of emipirical parameters, performs reasonably well (RMS = 0.20 kcal/mol with counterpoise correction). It is well-known (see, e.g., Table 10 of ref 11) that hydrogen bonds are a particularly favorable usage scenario for MP2, so one might expect this to carry over to water clusters. The dRPA75 functional,[108] which includes dRPA correlation, also performs resonably with RMSD 0.17 kcal/mol without any counterpoise correction.

n-Body Decomposition. Further, we have analyzed the n-body interactions. Electronic structure calculations for water clusters provide details of their structures and energetics, whereas decomposition of the dissociation energies of water clusters into n-body contributions will help to understand the importance of many-body interactions[109] in water clusters, which is essential to develop accurate analytical representations of the water potential energy surface.[38,110] Moreover, they may offer a route to accurate energies for large water clusters where the cost of all-atom HLCs would be prohibitive.

In a many-body expansion, the total energy of a cluster containing n individual fragments $(f_i f_j \ldots f_n)$ takes the form of the equation written below.

$$E[f_i f_j \ldots f_n] = E_1 + E_2 + E_3 + \ldots + E_n$$

Where

$$E_1 = \sum_{i=1}^{n} E[f_i]$$

$$E_2 = \sum_{i=1}^{n-1} \sum_{j>1}^{n} E[f_i f_j] - (E[f_i] + E[f_j])$$

$$E_3 = \sum_{i=1}^{n-2} \sum_{j>i}^{n-1} \sum_{k>j}^{n} [E[f_i f_j f_k] - (E[f_i f_j] + E[f_i f_k] + E[f_j f_k]) + (E[f_i] + E[f_j] + E[f_k])$$

E_1 is the one-body energies, E_2 is the pairwise interactions term, E_3 are the triad or three-body interactions, and so on.

This approach was previously considered in the work of Tschumper and co-workers[41,111] where they demonstrate efficient CCSD(T) optimized geometries and harmonic vibrational frequencies for molecular clusters with the n-body:many-body QM:QM technique.[41,111] For ring water clusters, Xantheas also found that four-body and higher terms were negligible[112] on the accuracy scale they were considering.

MP2, CCSD-MP2, and (T) contributions to various n-body terms for $(H_2O)_2$ through $(H_2O)_6$ are summarized in Table 9. A number of salient points come to the fore.

First of all, five-body and higher terms are recovered well at the MP2-F12/cc-pVTZ-F12 level; for the $(H_2O)_6$ structures,

DOI: 10.1021/acs.jctc.6b01046
J. Chem. Theory Comput. 2017, 13, 3136−3152

Table 10. Total Dissociation Energies (kcal/mol) from n-Body Analysis and Cumulative n-Body Contributions for Water Cluster Isomers through $(H_2O)_6$ Computed in This Work[b,c] and Reported by Paesani and Co-workers[13] (in Italic)[a]

	best value	∑ n-body	two-body	three-body	four-body	five-body and up		best value	∑ n-body	two-body	three-body	four-body	five-body and up
$(H_2O)_2$	5.08	5.08	5.08				$(H_2O)_6$						
$(H_2O)_3$							6BAG	47.37	47.37	35.95	10.28	1.12	0.02
3UUD	16.26	16.26	13.77	2.49					*46.80*	*35.28*	*10.35*	*1.16*	*0.01*
3UUU	15.58	15.58	13.29	2.29					46.98	35.39	10.43	1.13	0.02
$(H_2O)_4$							6BK1	48.10	48.10	36.67	10.34	1.05	0.04
4S4	28.90	28.90	22.11	6.22	0.58				*47.52*	*36.02*	*10.38*	*1.08*	*0.04*
		28.81	*22.05*	*6.17*	*0.59*				47.63	36.10	10.43	1.06	0.05
4Ci	28.04	28.04	21.53	5.95	0.56		6BK2	47.83	47.82	36.76	10.05	0.98	0.02
		27.94	*21.47*	*5.91*	*0.56*				*47.26*	*36.13*	*10.11*	*1.00*	*0.02*
4Py	24.96	24.95	21.19	3.71	0.06				47.42	36.24	10.17	0.99	0.02
		24.88	*21.13*	*3.68*	*0.07*		6CA	48.53	48.52	39.13	8.91	0.49	0.00
$(H_2O)_5$									*47.96*	*38.47*	*8.97*	*0.53*	*−0.01*
5CYC	38.08	38.08	27.61	9.14	1.24	0.09			48.13	38.58	9.06	0.49	0.00
		37.93	*27.53*	*9.07*	*1.24*	*0.09*	6CB1	45.98	45.98	32.88	11.32	1.62	0.17
5CAA	36.28	36.27	30.01	5.99	0.27	0.00			*45.44*	*32.30*	*11.34*	*1.63*	*0.17*
		36.17	*29.94*	*5.94*	*0.29*	*0.00*			45.62	32.40	11.43	1.62	0.17
5CAB	35.59	35.59	29.47	5.84	0.29	−0.01	6CB2	45.88	45.88	32.77	11.35	1.60	0.17
		35.49	*29.39*	*5.78*	*0.31*	*0.01*			*45.36*	*32.24*	*11.34*	*1.61*	*0.17*
5CAC	36.62	36.61	29.75	6.41	0.47	−0.03			45.55	32.34	11.44	1.60	0.17
		36.54	*29.67*	*6.35*	*0.49*	*0.03*	6CC	46.98	46.98	33.25	11.77	1.77	0.20
5FRA	34.90	34.91	28.37	6.28	0.27	−0.02			*46.47*	*32.71*	*11.78*	*1.78*	*0.20*
		34.83	*28.30*	*6.23*	*0.28*	*0.02*			46.65	32.81	11.87	1.78	0.20
5FRB	36.77	36.77	28.43	7.57	0.75	0.01	6PR	48.88	48.86	39.52	8.75	0.64	−0.06
		36.64	*28.35*	*7.52*	*0.76*	*0.01*			*48.24*	*38.94*	*8.70*	*0.66*	*−0.06*
5FRC	34.21	34.22	27.74	6.27	0.24	−0.02			48.38	39.05	8.76	0.63	-0.06
		34.14	*27.67*	*6.20*	*0.25*	*0.02*							

[a]For $(H_2O)_6$, our n-body results at Paesani's reference geometries[c] are underlined. Our best whole-system benchmark values are also given for comparison to show that our n-body expansion is adequately converged. [b]Two-body: CCSD(F12*)/cc-pV5Z-F12 + [CCSD(T)−CCSD]/haV5Z; three-body: CCSD(F12*)/cc-pVQZ-F12 + [CCSD(T)−CCSD]/haVQZ; four- and five-body: MP2-F12/cc-pVQZ-F12 + [CCSD(F12*)−MP2-F12]/cc-pVTZ-F12 + [CCSD(T)−CCSD]/haVTZ; and six-body: MP2-F12/cc-pVQZ-F12. [c]For $(H_2O)_{4,5}$, both studies employ the BEGDB reference geometries; for $(H_2O)_6$, Paesani and co-workers[13] employed MP2/haVTZ optimized geometries (see text). Our n-body results at those different geometries are underlined.

HLCs with the haVTZ basis set (see Table S7) are found to be 0.01 kcal/mol or less, whereas further basis set expansion to MP2-F12/cc-pVQZ-F12 has an effect of 0.005 kcal/mol or less.

At the other extreme, the two-body term is exquisitely sensitive to both basis set and electron correlation level, with even cc-pV5Z-F12 (albeit in the absence of counterpoise correction) still showing a nontrivial basis set increment for MP2-F12; granted, this is compensated to a large extent by an opposite-sign basis set increment in CCSD(F12*)−MP2-F12. As for the (T) term, even the haV5Z basis set increment still adds up to 0.025 kcal/mol over haVQZ. Fortunately, the calculation of two-body terms is relatively inexpensive in terms of resources, requiring just $n(n-1)/2$ water dimer calculations.

The three-body term, which requires $n(n-1)(n-2)/6$ water trimer calculations, is of some interest. Basis set expansion effects beyond cc-pVTZ-F12 are nontrivial in both MP2-F12 and CCSD(F12*)−MP2-F12, but they largely compensate; consequently, most of the three-body term is recovered at the CCSD(F12*)/cc-pVTZ-F12 level. (For conventional CCSD vs MP2 calculations, on the other hand, we see a noticeable basis set increment from haVTZ to haVQZ.) Note that for $(H_2O)_6$, for instance, the post-MP2 correction is highly structure-dependent; notably, the prism and cage have CCSD−MP2 differences in the −0.25 kcal/mol range, whereas those for the three ring structures are only in the −0.05 kcal/mol range. One explanation that suggests itself are Axilrod−Teller−Muto (ATM)[113−116] terms; these three-body dispersion terms can be attractive or repulsive and first show up at third order in a many-body perturbation theory expansion. Indeed, the (MP3−MP2)/haVTZ contribution (see Table S8) has three-body terms for the $(H_2O)_6$ isomers that qualitatively display the same pattern as the CCSD-MP2 difference. Furthermore, we obtained approximate ATM terms using the expressions of Grimme and co-workers as implemented in their DFTD3 empirical dispersion model[94] and again found the same qualitative pattern of repulsive ATM terms destabilizing the more compact structures. It stands to reason that the latter would be disfavored, since the leading distance dependence of ATM terms is[116] R^{-9}.

The four-body terms are fairly well described at the MP2-F12/cc-pVTZ-F12 level, although they are reduced by up to 0.03 kcal/mol when the basis set is expanded to cc-pVQZ-F12. CCSD-MP2 contributions here are slightly antibonding, partially compensating for attractive (T) contributions that can reach 0.06 kcal/mol for the $(H_2O)_6$ ring structures.

The cumulative n-body contributions, and their sums, can be found in Table 10, together with our best computed values without resorting to n-body decomposition. The two sets of values are seen there to agree to within a few hundredths of a kcal/mol, showing that a hierarchical n-body scheme indeed permits this level of accuracy. This offers a path to benchmark

DOI: 10.1021/acs.jctc.6b01046
J. Chem. Theory Comput. 2017, 13, 3136−3152

data for larger clusters, provided the following particulars are adhered to

- the two-body term needs to be calculated as accurately as possible, using basis sets all the way through V5Z-F12 and haV5Z;
- for the three-body term, CCSD−MP2 is essential;
- post-MP2 corrections for the four-body term are desirable;
- five-body terms and up are adequately captured at the MP2-F12/cc-pV*n*Z-F12 level (*n* = T, or definitely for *n* = Q).

From a computational cost point of view, our two-body calculations are relatively inexpensive, whereas the three-body calculations with cc-pVQZ-F12 and haVQZ basis sets are somewhat costlier. Still, the cost and resource requirements of the individual calculations are independent of cluster size, making it practical to converge these terms to the basis set and correlation limit.

After the original version of the present article was submitted for publication, a paper was published[13] in which a new force field for liquid water was derived from conventional and explicitly correlated ab initio calculations using an n-body decomposition scheme, but considerably smaller basis sets were used than those in the present work. For the $(H_2O)_{4,5}$ isomers, the values given in the electronic Supporting Information (ESI) of said paper agree to 0.10 kcal/mol RMS with the present work, but for the isomers of $(H_2O)_6$, a fairly substantial discrepancy (up to 0.6 kcal/mol) is seen. About 0.4 kcal/mol of this difference is due to different reference geometries; it can be verified by comparison between the ESIs of refs 117 and 118 and the ESI of ref 13 that the calculations in ref 13 were carried out at MP2/haVTZ reference geometries just for $(H_2O)_6$, unlike those for $(H_2O)_4$ and $(H_2O)_5$, which employed the same RI-MP2/AVDZ reference geometries from BEGDB as the present work. Upon recalculation of our $(H_2O)_6$ n-body data at ref 13's MP2/haVTZ reference geometries, the remaining discrepancy falls in line with the two other species and scales roughly linearly with the number of water molecules. It is essentially concentrated in the two-body and three-body terms; four-body and higher terms are in very good agreement. While the levels of theory used for these latter terms in ref 13 and in the present work are comparable, for the three-body term ref 13 employed full-CP conventional CCSD(T)/AVTZ+bond functions, compared to CCSD(F12*)/cc-pVQZ-F12+(T)/haVQZ here. The two-body term in ref 13 was calculated at the CCSD(T)/AV{T,Q}Z + bond functions level (no counterpoise correction specified), compared to CCSD(F12*)/cc-pV5Z-F12 + (T)/haV5Z here. Moreover, the small residual discrepancy between, on one hand, our *n*-body summations and, on the other hand, the whole-system benchmark results presented in the preceding sections is an order of magnitude smaller than that between either of these and ref 13.

As an aside, we note (Table S10) that the effect of the different reference geometries on the $(H_2O)_6$ energetics is reduced by almost an order of magnitude if monomer relaxation is taken into account; the RMS difference between the dissociation energies at both sets of geometries drops from 0.40 kcal/mol without monomer relaxation to 0.05 with it.

WATER27 Data Set. The B3LYP/6-311++G(2d,2p) reference geometries were downloaded from the GMTKN30 website[52] and used without further optimization. The four $(H_2O)_{20}$ structures are too large for a convergence study of the post-MP2 corrections; the remaining subset of 23 structures (denoted WATER23) includes 10 structures of neutral $(H_2O)_n$ (n = 2−8); 5 structures of $H_3O^+(H_2O)_n$ (n = 1−3, 6), 7 structures of $OH^-(H_2O)_n$ (n = 1−6), and 1 zwitterion, $H_3O^+(H_2O)_6OH^-$. The original study[50] presented MP2/CBS energies, corrected for WATER23 with CCSD(T)/aug-cc-pVDZ high level corrections.[50] To our knowledge, no other systematic benchmark data have thus far been reported for WATER23 or WATER27. However, a few extensive reports are available on the performance of various DFT functionals for this data set.[50−52] Here, we have reported revised high-level ab initio calculated benchmark data for WATER27. Similar to the BEGDB data set, here we are also reporting dissociation energies for WATER27 (Figure 2), both with and without monomer relaxation terms. (The original WATER27 data[50] were reported including monomer relaxation.) However, cohesive energies are also provided in the Supporting Information.

At the MP2-F12 level, we were able to perform calculations with the cc-pV5Z-F12 basis set for WATER23; for the $(H_2O)_{20}$ isomers, the largest basis set with which we were able to perform MP2-F12 calculations was cc-pVQZ-F12. For the WATER23 subset, we have considered the counterpoise corrected cc-pV{Q,5}Z-F12 extrapolated energies as reference for the MP2-F12 correlation component, which agrees with raw and half-counterpoise corrected values to 0.022 and 0.011 kcal/mol RMSD, respectively (Table 11). The counterpoise-corrected MP2-F12/cc-pV{T,Q}-F12 extrapolation performs quite well with RMS deviation 0.007 kcal/mol. Unlike BEGDB, however, raw and half-counterpoise corrected values exhibit RMSDs an order of magnitude larger, at 0.117 and 0.061 kcal/mol, respectively. This is almost entirely due to convergence issues with the anionic species in WATER23 (cf. Table 11); it should be noted that cc-pV*n*Z-F12 basis sets carry s and p diffuse functions only on the heavy atoms, not for higher angular momenta.

In order to cope with the basis set convergence problem of anionic water clusters, we constructed modified basis sets, denoted aug-cc-pV*n*Z-F12 or aVnZ-F12, which have been augmented with diffuse functions on the higher angular momenta. For oxygen, the additional diffuse function exponents were obtained by minimizing the MP2-F12 energy for oxygen anion; full details of the aug-cc-pV*n*Z-F12 basis sets for the first two rows and selected heavy p-block elements will be published elsewhere.[119] Geminal exponent (β) values were set to 0.9 for aug-cc-pVDZ-F12 and 1.0 for both aug-cc-pVTZ-F12 and aug-cc-pVQZ-F12 basis sets. Results with aV{T,Q}Z-F12 extrapolation seems to be converged with RMS deviations of 0.015 (raw) and 0.006 (CP) kcal/mol. Thus, we may eliminate counterpoise corrections by using aV{T,Q}Z-F12 for the MP2-F12 correlation component. Further, counterpoise corrected aVTZ-F12 performs quite well, with 0.026 kcal/mol RMS deviation. For the deprotonated clusters, counterpoise corrected cc-pV{T,Q}Z-F12 and aV{T,Q}Z-F12 also perform quite well, with RMSD values 0.012 and 0.011 kcal/mol, respectively.

For the CCSD(F12*)−MP2-F12 difference term, we have performed the calculations up to the cc-pVTZ-F12 basis set for WATER23. Calculations with the corresponding aug-cc-pV*n*Z-F12 basis sets were also performed. Counterpoise corrected aV{D,T}Z-F12 extrapolation closely matches raw and half-CP results with the same basis sets, with just 0.007 and 0.004 kcal/

DOI: 10.1021/acs.jctc.6b01046
J. Chem. Theory Comput. 2017, 13, 3136−3152

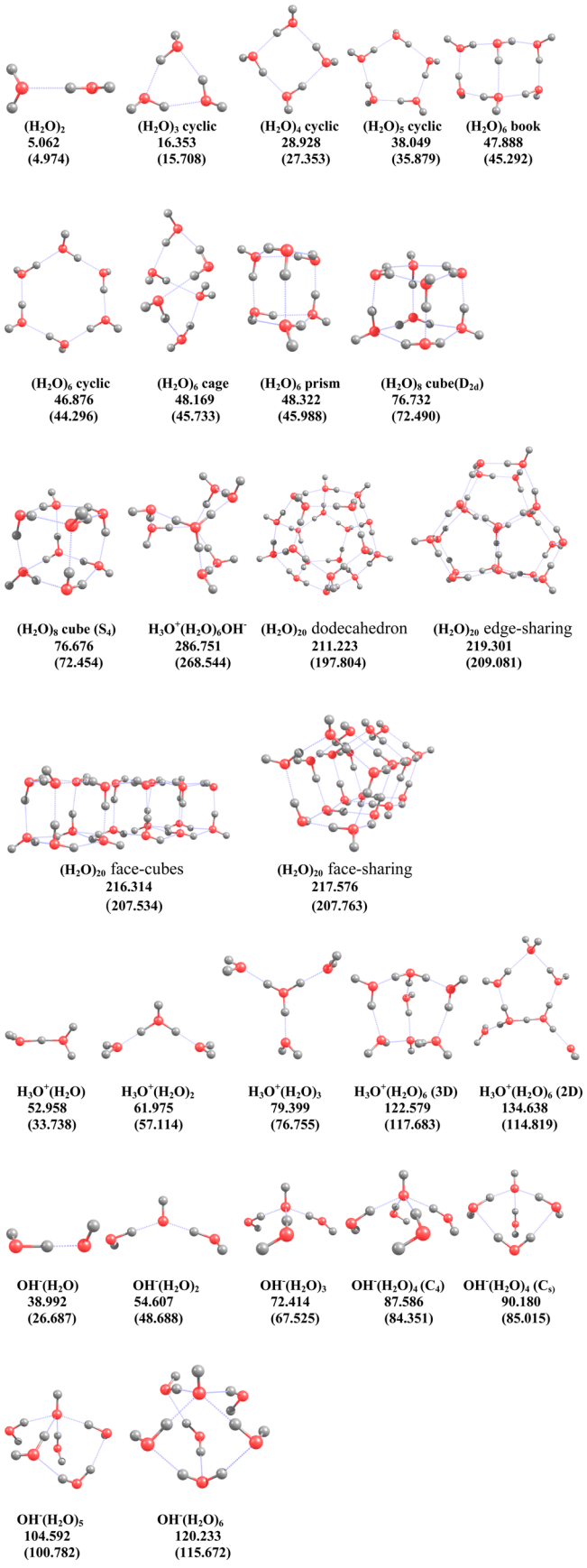


Figure 2. B3LYP/6-311++G(2d,2p) optimized geometries of the WATER27 data set. For WATER23, dissociation energies (kcal/mol) calculated with our best level are also given exclusive (and, in parentheses, inclusive) of monomer relaxation energies. For $(H_2O)_{20}$ isomers, the dissociation energies (kcal/mol) provided here are the combination of our best MP2-F12 [MP2-F12/cc-pV{T,Q}Z-F12 CP] results with the high-level correction obtained from n-body analysis.

Table 11. RMS Deviations (kcal/mol) for the MP2-F12 Correlation Component of Cohesive Energies Calculated with Various Basis Sets for the WATER23 Data Set

	WATER23			Only for deprotonated water clusters		
	raw	CP	half	raw	CP	half
cc-pVDZ-F12	0.225	0.132	0.085	0.379	0.138	0.149
cc-pVTZ-F12	0.195	0.044	0.090	0.333	0.044	0.160
cc-pVQZ-F12	0.139	0.015	0.069	0.241	0.017	0.122
cc-pV5Z-F12	0.060	0.005	0.030	0.103	0.006	0.052
cc-pV{D,T}Z-F12	0.184	0.018	0.096	0.315	0.027	0.166
cc-pV{T,Q}Z-F12	0.117	0.007	0.061	0.204	0.012	0.106
cc-pV{Q,5}Z-F12	0.022	REF	0.011	0.037	REF	0.019
aVDZ-F12	0.056	0.111	0.054	0.100	0.114	0.048
aVTZ-F12	0.065	0.026	0.031	0.107	0.023	0.056
aVQZ-F12	0.028	0.010	0.016	0.047	0.013	0.028
aV{D,T}Z-F12	0.072	0.022	0.046	0.111	0.038	0.074
aV{T,Q}Z-F12	0.015	0.006	0.010	0.026	0.011	0.018
aVTZ	0.162	0.023	0.072	0.224	0.024	0.104
aVQZ	0.086	0.017	0.036	0.115	0.015	0.050
aV5Z	0.035	0.007	0.014	0.039	0.007	0.016
awCVQZ	0.093	0.017	0.043	0.149	0.020	0.073
awCV5Z	0.041	0.007	0.019	0.066	0.008	0.033

mol RMSD, respectively (Table 12). Thus, we have considered counterpoise corrected aV{D,T}Z-F12 extrapolated values as our reference. Further, to validate the basis set convergence for deprotonated water clusters, additional calculations were performed with cc-pVQZ-F12 and aVQZ-F12 basis sets for $OH^-(H_2O)_n$ (n = 1−4). Relative to aV{T,Q}Z-F12, performance of aV{D,T}Z-F12 is very satisfactory, with RMS deviation less than 0.009 kcal/mol. This also clearly shows that counterpoise-corrected cc-pV{D,T}Z-F12 performs reasonably well for all of the systems including deprotonated water clusters.

Similar to the BEGDB benchmark study (vide supra), for the WATER23 data set, the (T) contributions were taken from conventional CCSD(T) calculations. For these calculations, haVDZ and haVTZ basis sets were used throughout. Considering 23 systems, raw and counterpoise corrected results with haV{D,T}Z extrapolation differ by 0.042 kcal/mol RMSD; however, for neutral and protonated water clusters, this difference is much smaller (0.015 kcal/mol) (Table 13). This distinctly indicates a basis set convergence problem for (T) contributions for deprotonated water clusters. Hence, additional calculations were performed with the haVQZ basis set for deprotonated species only; the minute difference (0.005 kcal/

DOI: 10.1021/acs.jctc.6b01046
J. Chem. Theory Comput. 2017, 13, 3136−3152

Table 12. RMS Deviations (kcal/mol) for the CCSD(F12*)−MP2-F12 Components of Cohesive Energies Calculated with Various Basis Sets for the WATER23 Data Set

	WATER23			Only for neutral and protonated water clusters			Only for deprotonated water clusters up to $OH^-(H_2O)_4$		
	raw	CP	half	raw	CP	half	raw	CP	half
cc-pVDZ-F12	0.128	0.111	0.118	0.077	0.099	0.087	0.194	0.128	0.161
cc-pVTZ-F12	0.068	0.033	0.050	0.033	0.029	0.031	0.110	0.036	0.072
cc-pVQZ-F12							0.083	0.011	0.046
cc-pV{D,T}Z-F12	0.045	0.004	0.023	0.015	0.001	0.008	0.077	0.014	0.037
cc-pV{T,Q}Z-F12							0.068	0.003	0.034
aVDZ-F12	0.051	0.090	0.069	0.026	0.081	0.053	0.078	0.097	0.087
aVTZ-F12	0.016	0.025	0.019	0.005	0.022	0.013	0.024	0.023	0.023
aVQZ-F12							0.012	0.006	0.009
aV{D,T}Z-F12	0.007	REF	0.004	0.007	REF	0.003	0.006	0.009	0.006
aV{T,Q}Z-F12							0.007	REF	0.004

Table 13. RMS Deviations (kcal/mol) for the (T) Term of Conventional CCSD(T) Calculated for Cohesive Energies of the WATER23 Data Set with Various Basis Sets

	WATER23			Only for neutrals and protonated water clusters			Only for deprotonated water clusters		
	raw	CP	half	raw	CP	half	raw	CP	half
haVDZ	0.047	0.322	0.178	0.051	0.290	0.168	0.049	0.374	0.203
haVTZ	0.028	0.095	0.042	0.007	0.086	0.045	0.039	0.123	0.050
haVQZ							0.018	0.052	0.022
haV{D,T}Z	0.042	REF	0.021	0.015	REF	0.007	0.057	0.021	0.024
haV{T,Q}Z							0.005	REF	0.002

mol) between raw and counterpoise-corrected haV{T,Q}Z suggests that the basis set is properly converged.

Thus, our final reference level to calculate the cohesive energies for WATER23 data set (excluding $(H_2O)_{20}$ isomers) is as follows:

- For neutral and protonated water clusters: MP2-F12/cc-pV{Q,5}Z-F12 CP + [CCSD(F12*)−MP2-F12]/aV{D,T}Z-F12 CP + [CCSD(T)−CCSD]/haV{D,T}Z CP
- For anionic (deprotonated) water clusters: MP2-F12/cc-pV{Q,5}Z-F12 CP + [CCSD(F12*)−MP2-F12]/aV{D,T}Z-F12 CP + [CCSD(T)−CCSD]/haV{T,Q}Z CP

The final dissociation energies calculated at these levels differ from the previous reference values presented by Goddard and co-workers[50] by RMSD = 0.164 kcal/mol.

Because of their large size, $(H_2O)_{20}$ clusters have always been challenging systems for higher level ab initio calculations. Xantheas and co-workers[120] reported MP2/CBS binding energies for $(H_2O)_{20}$ isomers, which were further used for the evaluation of DFT functionals by Goddard and co-workers.[50] In this study, we report MP2-F12/cc-pV{T,Q}Z-F12 extrapolated results for $(H_2O)_{20}$ isomers, which differ from previously reported MP2/CBS results[120] by 6.18 kcal/mol RMS. Since our attempts at performing CCSD(T)/haVTZ level calculations were successful only for the edge sharing $(H_2O)_{20}$ and face cube $(H_2O)_{20}$, we have combined MP2-F12/cc-pV{T,Q}Z-F12 CP dissociation energies for the $(H_2O)_{20}$ isomers with just two-body and three-body HLC corrections, which are the most important; the smaller four-body HLC would already require 20!/(16!4!) = 4845 water tetramer calculations per isomer. The two-body terms (190 dimer calculations per isomer) were obtained at the [CCSD(F12*)−MP2-F12]/cc-pVQZ-F12 level combined with orbital-based (T)/haV{T,Q}Z energies, whereas the three-body terms (entailing 1140 water trimer calculations per isomer) were obtained at the [CCSD(F12*)−MP2-F12]/cc-pVTZ-F12 + (T)/haV{D,T}Z level. Recently, Friedrich and co-workers[121] have reported explicitly correlated MP2-F12/cc-pVQZ-F12, as well as the incremental CCSD(T)(F12*)|MP2-F12 interaction energies for $(H_2O)_{20}$ isomers. Their reported dissociation energies[121] differ from our final values by 0.51 kcal/mol RMS.

We have also performed dRPA75/haVTZ and dRPA75/haVQZ calculations for $(H_2O)_{20}$, as dRPA75/haV{T,Q}Z was suggested in a recent application by Csonka and co-workers[122] of their dRPA75 methods to water clusters. Our raw dRPA75/haV{T,Q}Z results are close to their counterpoise-corrected dRPA75/AVTZ values,[122] both overbinding by a few kcal/mol compared to our reference level reported here.

DOI: 10.1021/acs.jctc.6b01046
J. Chem. Theory Comput. 2017, 13, 3136−3152

CONCLUSIONS

BEGDB and WATER27 data sets have been considered in this work for the conventional and explicitly correlated ab initio benchmark study. The cohesive energies of water clusters have been investigated with and without Boys−Bernardi counterpoise corrections, along with the " half-counterpoise" method.

For the MP2 correlation component, results with conventional calculations are to some extent disappointing, and basis set convergence of the explicitly correlated methods is considerably better for both data sets. For CCSD−MP2, CCSD(F12*) appears to be closer to the basis set limit than CCSD-F12b, with RMS deviations being almost halved with the same basis set; for cc-pVQZ-F12, both yield functionally identical results. For the calculation of the (T) term, conventional, orbital-based calculations are preferred. The high-level correction should preferably be assembled from conventional (T) and explicitly correlated CCSD−MP2. Among the DFT functionals considered, only ωB97X-V and ωB97M-V outperform MP2, although the double-hybrid DFT functionals DSD-PBEP86 and DSD-PBEPBE-NL also perform reasonably well.

Our best calculated BEGDB dissociation energies can be reproduced through n-body expansion, provided one is willing to push to the basis set and electron correlation limit for the two-body term; for the three-body term, post-MP2 contributions (particularly CCSD-MP2) are important for capturing the three-body dispersion effects. Terms beyond four-body can be adequately captured at the MP2-F12 level. In this manner, larger water clusters can be treated fairly accurately, as we demonstrate for $(H_2O)_{20}$ isomers. n-body analysis broken down by correlation energy component also reveals that post-MP2 correlation contributions to the cage-ring equilibrium in $(H_2O)_6$ result primarily from three-body ATM terms.

In the WATER27 data set, the anionic clusters exhibit unacceptably slow basis set convergence with the cc-pV*n*Z-F12 basis sets. Using modified basis sets, denoted aug-cc-pV*n*Z-F12, in which diffuse functions are also added to the higher angular momenta, we were able to remedy this problem.

The calculated final dissociation energies of the BEGDB and WATER27 data sets differ nontrivially from the previous reference values used by Shields and co-workers[17] for BEGDB and Goddard and co-workers[50] for WATER23, by 0.091 and 0.164 kcal/mol RMS, respectively.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b01046.

Spreadsheets containing calculated total and association and cohesive energies for the various data sets discussed in the article (ZIP)

AUTHOR INFORMATION

Corresponding Author

*E-mail: gershom@weizmann.ac.il. Fax: +972 8 934 3029.

ORCID

Jan M. L. Martin: 0000-0002-0005-5074

Funding

D.M. and M.K.K. acknowledge postdoctoral fellowships and N.S. acknowledges a graduate fellowship from the Feinberg Graduate School. This research was supported by the Israel Science Foundation (grant 1358/15), the Minerva Foundation, the Lise Meitner-Minerva Center for Computational Quantum Chemistry, and the Helen and Martin Kimmel Center for Molecular Design (Weizmann Institute of Science).

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Řezáč, J.; Hobza, P. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. *Chem. Rev.* **2016**, *116*, 5038−5071.

(2) Hobza, P. Calculations on Noncovalent Interactions and Databases of Benchmark Interaction Energies. *Acc. Chem. Res.* **2012**, *45*, 663−672.

(3) Grimme, S. Density Functional Theory with London Dispersion Corrections. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 211−228.

(4) Dubecký, M.; Derian, R.; Jurečka, P.; Mitas, L.; Hobza, P.; Otyepka, M. Quantum Monte Carlo for Noncovalent Interactions: An Efficient Protocol Attaining Benchmark Accuracy. *Phys. Chem. Chem. Phys.* **2014**, *16*, 20915−20923.

(5) Burns, L. A.; Marshall, M. S.; Sherrill, C. D. Comparing Counterpoise-Corrected, Uncorrected, and Averaged Binding Energies for Benchmarking Noncovalent Interactions. *J. Chem. Theory Comput.* **2014**, *10*, 49−57.

(6) Gillan, M. J.; Alfè, D.; Bygrave, P. J.; Taylor, C. R.; Manby, F. R. Energy Benchmarks for Water Clusters and Ice Structures from an Embedded Many-Body Expansion. *J. Chem. Phys.* **2013**, *139*, 114101.

(7) Gillan, M. J.; Alfè, D.; Michaelides, A. Perspective: How Good Is DFT for Water? *J. Chem. Phys.* **2016**, *144*, 130901.

(8) Mardirossian, N.; Head-Gordon, M. How Accurate Are the Minnesota Density Functionals for Noncovalent Interactions, Isomerization Energies, Thermochemistry, and Barrier Heights Involving Molecules Composed of Main-Group Elements? *J. Chem. Theory Comput.* **2016**, *12*, 4303−4325.

(9) Rezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-Balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427−2438.

(10) Řezáč, J.; Riley, K. E.; Hobza, P. Extensions of the S66 Data Set: More Accurate Interaction Energies and Angular-Displaced Nonequilibrium Geometries. *J. Chem. Theory Comput.* **2011**, *7*, 3466−3470.

(11) Brauer, B.; Kesharwani, M. K.; Kozuch, S.; Martin, J. M. L. The S66 × 8 Benchmark for Noncovalent Interactions Revisited: Explicitly Correlated Ab Initio Methods and Density Functional Theory. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20905−20925.

(12) Ludwig, R. Water: From Clusters to the Bulk. *Angew. Chem., Int. Ed.* **2001**, *40*, 1808−1827.

(13) Reddy, S. K.; Straight, S. C.; Bajaj, P.; Huy Pham, C.; Riera, M.; Moberg, D. R.; Morales, M. A.; Knight, C.; Götz, A. W.; Paesani, F. On the Accuracy of the MB-Pol Many-Body Potential for Water: Interaction Energies, Vibrational Frequencies, and Classical Thermodynamic and Dynamical Properties from Clusters to Liquid Water and Ice. *J. Chem. Phys.* **2016**, *145*, 194504.

(14) Góra, U.; Cencek, W.; Podeszwa, R.; van der Avoird, A.; Szalewicz, K. Predictions for Water Clusters from a First-Principles Two- and Three-Body Force Field. *J. Chem. Phys.* **2014**, *140*, 194101.

(15) Jankowski, P.; Murdachaew, G.; Bukowski, R.; Akin-Ojo, O.; Leforestier, C.; Szalewicz, K. Ab Initio Water Pair Potential with Flexible Monomers. *J. Phys. Chem. A* **2015**, *119*, 2940−2964.

(16) Willow, S. Y.; Zeng, X. C.; Xantheas, S. S.; Kim, K. S.; Hirata, S. Why Is MP2-Water "Cooler" and "Denser" than DFT-Water? *J. Phys. Chem. Lett.* **2016**, *7*, 680−684.

(17) Shields, R. M.; Temelso, B.; Archer, K. A.; Morrell, T. E.; Shields, G. C. Accurate Predictions of Water Cluster Formation, $(H_2O)_n$, n = 2−10. *J. Phys. Chem. A* **2010**, *114*, 11725−11737.

(18) Temelso, B.; Renner, C. R.; Shields, G. C. Importance and Reliability of Small Basis Set CCSD(T) Corrections to MP2 Binding and Relative Energies of Water Clusters. *J. Chem. Theory Comput.* **2015**, *11*, 1439−1448.

DOI: 10.1021/acs.jctc.6b01046
J. Chem. Theory Comput. 2017, 13, 3136−3152

(19) 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 $(H_2O)_m$, m = 2−6, 8, 11, 16, and 17. *J. Chem. Phys.* **2015**, *142*, 234303.
(20) Howard, J. C.; Tschumper, G. S. Wavefunction Methods for the Accurate Characterization of Water Clusters. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 199−224.
(21) Howard, J. C.; Tschumper, G. S. Benchmark Structures and Harmonic Vibrational Frequencies Near the CCSD(T) Complete Basis Set Limit for Small Water Clusters: $(H_2O)_n$, n = 2, 3, 4, 5, 6. *J. Chem. Theory Comput.* **2015**, *11*, 2126−2136.
(22) Gadre, S. R.; Yeole, S. D.; Sahu, N. Quantum Chemical Investigations on Molecular Clusters. *Chem. Rev.* **2014**, *114*, 12132−12173.
(23) Nielsen, I. M. B.; Seidl, E. T.; Janssen, C. L. Accurate Structures and Binding Energies for Small Water Clusters: The Water Trimer. *J. Chem. Phys.* **1999**, *110*, 9435−9442.
(24) Pérez, J. F.; Hadad, C. Z.; Restrepo, A. Structural Studies of the Water Tetramer. *Int. J. Quantum Chem.* **2008**, *108*, 1653−1659.
(25) Ramírez, F.; Hadad, C. Z.; Guerra, D.; David, J.; Restrepo, A. Structural Studies of the Water Pentamer. *Chem. Phys. Lett.* **2011**, *507*, 229−233.
(26) Acelas, N.; Hincapié, G.; Guerra, D.; David, J.; Restrepo, A. Structures, Energies, and Bonding in the Water Heptamer. *J. Chem. Phys.* **2013**, *139*, 044310.
(27) Xantheas, S. S.; Aprà, E. The Binding Energies of the D_{2d} and S_4 Water Octamer Isomers: High-Level Electronic Structure and Empirical Potential Results. *J. Chem. Phys.* **2004**, *120*, 823−828.
(28) Kim, J.; Kim, K. S. Structures, Binding Energies, and Spectra of Isoenergetic Water Hexamer Clusters: Extensive Ab Initio Studies. *J. Chem. Phys.* **1998**, *109*, 5886−5895.
(29) Pedulla, J. M.; Vila, F.; Jordan, K. D. Binding Energy of the Ring Form of $(H_2O)_6$: Comparison of the Predictions of Conventional and Localized-orbital MP2 Calculations. *J. Chem. Phys.* **1996**, *105*, 11091−11099.
(30) Lenz, A.; Ojamäe, L. On the Stability of Dense versus Cage-Shaped Water Clusters: Quantum-Chemical Investigations of Zero-Point Energies, Free Energies, Basis-Set Effects and IR Spectra of $(H_2O)_{12}$ and $(H_2O)_{20}$. *Chem. Phys. Lett.* **2006**, *418*, 361−367.
(31) Wang, Y.; Babin, V.; Bowman, J. M.; Paesani, F. The Water Hexamer: Cage, Prism, or Both. Full Dimensional Quantum Simulations Say Both. *J. Am. Chem. Soc.* **2012**, *134*, 11116−11119.
(32) Day, P. N.; Pachter, R.; Gordon, M. S.; Merrill, G. N. A Study of Water Clusters Using the Effective Fragment Potential and Monte Carlo Simulated Annealing. *J. Chem. Phys.* **2000**, *112*, 2063−2073.
(33) Lee, H. M.; Suh, S. B.; Lee, J. Y.; Tarakeshwar, P.; Kim, K. S. Structures, Energies, Vibrational Spectra, and Electronic Properties of Water Monomer to Decamer. *J. Chem. Phys.* **2000**, *112*, 9759−9772.
(34) Lee, H. M.; Suh, S. B.; Kim, K. S. Structures, Energies, and Vibrational Spectra of Water Undecamer and Dodecamer: An Ab Initio Study. *J. Chem. Phys.* **2001**, *114*, 10749−10756.
(35) Maheshwary, S.; Patel, N.; Sathyamurthy, N.; Kulkarni, A. D.; Gadre, S. R. Structure and Stability of Water Clusters $(H_2O)_N$, N = 8−20: An Ab Initio Investigation. *J. Phys. Chem. A* **2001**, *105*, 10525−10537.
(36) Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(37) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96*, 6796−6806.
(38) Xantheas, S. S.; Burnham, C. J.; Harrison, R. J. Development of Transferable Interaction Models for Water. II. Accurate Energetics of the First Few Water Clusters from First Principles. *J. Chem. Phys.* **2002**, *116*, 1493−1499.
(39) Lenz, A.; Ojamäe, L.; Schindler, T.; Beauchamp, J. L.; Ojamäe, L.; Klein, M. L. A Theoretical Study of Water Clusters: The Relation between Hydrogen-Bond Topology and Interaction Energy from Quantum-Chemical Computations for Clusters with up to 22 Molecules. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1905−1911.
(40) Temelso, B.; Archer, K. A.; Shields, G. C. Benchmark Structures and Binding Energies of Small Water Clusters with Anharmonicity Corrections. *J. Phys. Chem. A* **2011**, *115*, 12034−12046.
(41) Bates, D. M.; Smith, J. R.; Janowski, T.; Tschumper, G. S. Development of a 3-Body:many-Body Integrated Fragmentation Method for Weakly Bound Clusters and Application to Water Clusters (H2O)n = 3 − 10, 16, 17. *J. Chem. Phys.* **2011**, *135*, 044123.
(42) Miliordos, E.; Aprà, E.; Xantheas, S. S. Optimal Geometries and Harmonic Vibrational Frequencies of the Global Minima of Water Clusters $(H_2O)_n$, n = 2−6, and Several Hexamer Local Minima at the CCSD(T) Level of Theory. *J. Chem. Phys.* **2013**, *139*, 114302.
(43) Sahu, N.; Singh, G.; Nandi, A.; Gadre, S. R. Toward an Accurate and Inexpensive Estimation of CCSD(T)/CBS Binding Energies of Large Water Clusters. *J. Phys. Chem. A* **2016**, *120*, 5706−5714.
(44) Shank, A.; Wang, Y.; Kaledin, A.; Braams, B. J.; Bowman, J. M. Accurate Ab Initio and "hybrid" Potential Energy Surfaces, Intramolecular Vibrational Energies, and Classical Ir Spectrum of the Water Dimer. *J. Chem. Phys.* **2009**, *130*, 144314.
(45) Partridge, H.; Schwenke, D. W. The Determination of an Accurate Isotope Dependent Potential Energy Surface for Water from Extensive Ab Initio Calculations and Experimental Data. *J. Chem. Phys.* **1997**, *106*, 4618−4639.
(46) Fanourgakis, G. S.; Xantheas, S. S. Development of Transferable Interaction Potentials for Water. V. Extension of the Flexible, Polarizable, Thole-Type Model Potential (TTM3-F, v. 3.0) to Describe the Vibrational Spectra of Water Clusters and Liquid Water. *J. Chem. Phys.* **2008**, *128*, 074506.
(47) Anderson, J. B. Quantum Chemistry by Random Walk. *J. Chem. Phys.* **1976**, *65*, 4121−4127.
(48) Rocher-Casterline, B. E.; Ch'ng, L. C.; Mollner, A. K.; Reisler, H. Communication: Determination of the Bond Dissociation Energy (D_0) of the Water Dimer, $(H_2O)_2$, by Velocity Map Imaging. *J. Chem. Phys.* **2011**, *134*, 211101.
(49) Řezáč, J.; Jurečka, P.; Riley, K. E.; Černý, J.; Valdes, H.; Pluháčková, K.; Berka, K.; Řezáč, T.; Pitoňák, M.; Vondrášek, J.; Hobza, P. Quantum Chemical Benchmark Energy and Geometry Database for Molecular Clusters and Complex Molecular Systems. *Collect. Czech. Chem. Commun.* **2008**, *73*, 1261−1270.
(50) Bryantsev, V. S.; Diallo, M. S.; van Duin, A. C. T.; Goddard, W. A. Evaluation of B3LYP, X3LYP, and M06-Class Density Functionals for Predicting the Binding Energies of Neutral, Protonated, and Deprotonated Water Clusters. *J. Chem. Theory Comput.* **2009**, *5*, 1016−1026.
(51) Goerigk, L.; Grimme, S. A General Database for Main Group Thermochemistry, Kinetics, and Noncovalent Interactions − Assessment of Common and Reparameterized (Meta-)GGA Density Functionals. *J. Chem. Theory Comput.* **2010**, *6*, 107−126.
(52) Goerigk, L.; Grimme, S. Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals—Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2011**, *7*, 291−309.
(53) Boys, S. F.; Bernardi, F. (Reprinted) The Calculation of Small Molecular Interactions by the Differences of Separate Total EnergiesSome Procedures with Reduced Errors. *Mol. Phys.* **2002**, *100*, 65−73.
(54) Boys, S. F.; 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.
(55) Handy, N. C. Comment on "The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. Mol. Phys. 1970, 19, 553-566". *Mol. Phys.* **2002**, *100*, 63.
(56) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.;

DOI: 10.1021/acs.jctc.6b01046
J. Chem. Theory Comput. 2017, 13, 3136−3152

Goll, E.; Hampel, C.; Hesselman, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.; Palmieri, P.; Peng, D.; Pflüger, K.; Pitzer, R. M.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M. *MOLPRO*, version 2015.1, a Package of Ab Initio Programs; University of Cardiff Chemistry Consultants (UC3): Cardiff, Wales, UK, 2012.

(57) Neese, F. The ORCA Program System. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 73−78.

(58) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. a.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. a.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio, R. A.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. a.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. a.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. a.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. a.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K. L.; Chipman, D. M.; Cramer, C. J.; Goddard, W. A.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer, H. F.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. a.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. a.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2015**, *113*, 184−215.

(59) Rolik, Z.; Szegedy, L.; Ladjánszki, I.; Ladóczki, B.; Kállay, M. An Efficient Linear-Scaling CCSD(T) Method Based on Local Natural Orbitals. *J. Chem. Phys.* **2013**, *139*, 094105.

(60) Yousaf, K. E.; Peterson, K. A. Optimized Auxiliary Basis Sets for Explicitly Correlated Methods. *J. Chem. Phys.* **2008**, *129*, 184108.

(61) Yousaf, K. E.; Peterson, K. A. Optimized Complementary Auxiliary Basis Sets for Explicitly Correlated Methods: Aug-Cc-pVnZ Orbital Basis Sets. *Chem. Phys. Lett.* **2009**, *476*, 303−307.

(62) Peterson, K. A.; Adler, T. B.; Werner, H.-J. Systematically Convergent Basis Sets for Explicitly Correlated Wavefunctions: The Atoms H, He, B-Ne, and Al-Ar. *J. Chem. Phys.* **2008**, *128*, 084102.

(63) Hill, J. G.; Peterson, K. A.; Knizia, G.; Werner, H.-J. Extrapolating MP2 and CCSD Explicitly Correlated Correlation Energies to the Complete Basis Set Limit with First and Second Row Correlation Consistent Basis Sets. *J. Chem. Phys.* **2009**, *131*, 194105.

(64) Peterson, K. A.; Kesharwani, M. K.; Martin, J. M. L. The cc-pV5Z-F12 Basis Set: Reaching the Basis Set Limit in Explicitly Correlated Calculations. *Mol. Phys.* **2015**, *113*, 1551−1558.

(65) Adler, T. B.; Knizia, G.; Werner, H.-J. A Simple and Efficient CCSD(T)-F12 Approximation. *J. Chem. Phys.* **2007**, *127*, 221106.

(66) Noga, J.; Šimunek, J. On the One-Particle Basis Set Relaxation in R12 Based Theories. *Chem. Phys.* **2009**, *356*, 1−6.

(67) Marchetti, O.; Werner, H.-J. Accurate Calculations of Intermolecular Interaction Energies Using Explicitly Correlated Wave Functions. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3400−3409.

(68) Marchetti, O.; Werner, H.-J. Accurate Calculations of Intermolecular Interaction Energies Using Explicitly Correlated Coupled Cluster Wave Functions and a Dispersion-Weighted MP2Method. *J. Phys. Chem. A* **2009**, *113*, 11580−11585.

(69) Woon, D. E.; Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. III. The Atoms Aluminum through Argon. *J. Chem. Phys.* **1993**, *98*, 1358−1371.

(70) Peterson, K. A.; Woon, D. E.; Dunning, T. H. Benchmark Calculations with Correlated Molecular Wave Functions. IV. The Classical Barrier Height of the $H+H_2 \rightarrow H_2+H$ Reaction. *J. Chem. Phys.* **1994**, *100*, 7410−7415.

(71) Rappoport, D.; Furche, F. Property-Optimized Gaussian Basis Sets for Molecular Response Calculations. *J. Chem. Phys.* **2010**, *133*, 134105.

(72) Weigend, F. Hartree-Fock Exchange Fitting Basis Sets for H to Rn. *J. Comput. Chem.* **2008**, *29*, 167−175.

(73) Hättig, C. Optimization of Auxiliary Basis Sets for RI-MP2 and RI-CC2 Calculations: Core−valence and Quintuple-ζ Basis Sets for H to Ar and QZVPP Basis Sets for Li to Kr. *Phys. Chem. Chem. Phys.* **2005**, *7*, 59−66.

(74) Perdew, J. P.; Schmidt, K. Jacob's Ladder of Density Functional Approximations for the Exchange-Correlation Energy. In *AIP Conference Proceedings*; Van Doren, V., Van Alsenoy, C., Geerlings, P., Eds.; AIP Conference Proceedings; AIP: Antwerp, Belgium, 2001; Vol. *577*, pp 1−20.

(75) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

(76) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098−3100.

(77) Perdew, J. P. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8822−8824.

(78) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.

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

(80) Grimme, S. Accurate Calculation of the Heats of Formation for Large Main Group Compounds with Spin-Component Scaled MP2Methods. *J. Phys. Chem. A* **2005**, *109*, 3067−3077.

(81) Quintal, M. M.; Karton, A.; Iron, M. A.; Boese, A. D.; Martin, J. M. L. Benchmark Study of DFT Functionals for Late-Transition-Metal Reactions. *J. Phys. Chem. A* **2006**, *110*, 709−716.

(82) Becke, A. D. Density-functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(83) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623−11627.

(84) Wang, Y.; Perdew, J. P. Correlation Hole of the Spin-Polarized Electron Gas, with Exact Small-Wave-Vector and High-Density Scaling. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *44*, 13298−13307.

(85) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110*, 6158−6170.

(86) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Function. *Theor. Chem. Acc.* **2008**, *120*, 215−241.

(87) Karton, A.; Tarnopolsky, A.; Lamère, J.-F.; Schatz, G. C.; Martin, J. M. L. Highly Accurate First-Principles Benchmark Data Sets for the Parametrization and Validation of Density Functional and Other Approximate Methods. Derivation of a Robust, Generally

DOI: 10.1021/acs.jctc.6b01046
J. Chem. Theory Comput. 2017, 13, 3136−3152

Applicable, Double-Hybrid Functional for Thermochemistry and Thermochemical. *J. Phys. Chem. A* **2008**, *112*, 12868−12886.
(88) Kozuch, S.; Martin, J. M. L. DSD-PBEP86: In Search of the Best Double-Hybrid DFT with Spin-Component Scaled MP2 and Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20104−20107.
(89) Kozuch, S.; Martin, J. M. L. Spin-Component-Scaled Double Hybrids: An Extensive Search for the Best Fifth-Rung Functionals Blending DFT and Perturbation Theory. *J. Comput. Chem.* **2013**, *34*, 2327−2344.
(90) Chai, J.-D.; Head-Gordon, M. Systematic Optimization of Long-Range Corrected Hybrid Density Functionals. *J. Chem. Phys.* **2008**, *128*, 084106.
(91) Lin, Y.-S.; Li, G.-D.; Mao, S.-P.; Chai, J.-D. Long-Range Corrected Hybrid Density Functionals with Improved Dispersion Corrections. *J. Chem. Theory Comput.* **2013**, *9*, 263−272.
(92) Mardirossian, N.; Head-Gordon, M. ωB97X-V: A 10-Parameter, Range-Separated Hybrid, Generalized Gradient Approximation Density Functional with Nonlocal Correlation, Designed by a Survival-of-the-Fittest Strategy. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9904−9924.
(93) Mardirossian, N.; Head-Gordon, M. ωB97M-V: A Combinatorially Optimized, Range-Separated Hybrid, Meta-GGA Density Functional with VV10 Nonlocal Correlation. *J. Chem. Phys.* **2016**, *144*, 214110.
(94) Grimme, S.; Antony, J.; 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.
(95) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(96) Vydrov, O. A.; Van Voorhis, T. Nonlocal van Der Waals Density Functional: The Simpler the Better. *J. Chem. Phys.* **2010**, *133*, 244103.
(97) Hujo, W.; Grimme, S. Performance of the van Der Waals Density Functional VV10 and (hybrid)GGA Variants for Thermochemistry and Noncovalent Interactions. *J. Chem. Theory Comput.* **2011**, *7*, 3866−3871.
(98) Kesharwani, M. K.; Karton, A.; Martin, J. M. L. Benchmark Ab Initio Conformational Energies for the Proteinogenic Amino Acids through Explicitly Correlated Methods. Assessment of Density Functional Methods. *J. Chem. Theory Comput.* **2016**, *12*, 444−454.
(99) Schwabe, T.; Grimme, S. Double-Hybrid Density Functionals with Long-Range Dispersion Corrections: Higher Accuracy and Extended Applicability. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397−3406.
(100) Brauer, B.; Kesharwani, M. K.; Martin, J. M. L. Some Observations on Counterpoise Corrections for Explicitly Correlated Calculations on Noncovalent Interactions. *J. Chem. Theory Comput.* **2014**, *10*, 3791−3799.
(101) Martin, J. M. L.; François, J.-P.; Gijbels, R. Combined Bond-Polarization Basis Sets for Accurate Determination of Dissociation Energies. *Theor. Chim. Acta* **1989**, *76*, 195−209.
(102) Martin, J. M. L.; Kesharwani, M. K. Assessment of CCSD(T)-F12 Approximations and Basis Sets for Harmonic Vibrational Frequencies. *J. Chem. Theory Comput.* **2014**, *10*, 2085−2090.
(103) Sirianni, D. A.; Burns, L. A.; Sherrill, C. D. Comparison of Explicitly Correlated Methods for Computing High-Accuracy Benchmark Energies for Noncovalent Interactions. *J. Chem. Theory Comput.* **2017**, *13*, 86−99.
(104) Yau, A. D.; Perera, S. A.; Bartlett, R. J. Vertical Ionization Potentials of Ethylene: The Right Answer for the Right Reason? *Mol. Phys.* **2002**, *100*, 835−842.
(105) Marshall, M. S.; Sherrill, C. D. Dispersion-Weighted Explicitly Correlated Coupled-Cluster Theory [DW-CCSD(T**)-F12]. *J. Chem. Theory Comput.* **2011**, *7*, 3978−3982.
(106) Yu, H. S.; He, X.; Li, S. L.; Truhlar, D. G. MN15: A Kohn−Sham Global-Hybrid Exchange−correlation Density Functional with Broad Accuracy for Multi-Reference and Single-Reference Systems and Noncovalent Interactions. *Chem. Sci.* **2016**, *7*, 5032−5051.
(107) Yu, H. S.; He, X.; Li, S. L.; Truhlar, D. G. Correction: MN15: A Kohn−Sham Global-Hybrid Exchange−correlation Density Functional with Broad Accuracy for Multi-Reference and Single-Reference Systems and Noncovalent Interactions. *Chem. Sci.* **2016**, *7*, 6278−6279.
(108) Mezei, P. D.; Csonka, G. I.; Ruzsinszky, A.; Kállay, M. Construction and Application of a New Dual-Hybrid Random Phase Approximation. *J. Chem. Theory Comput.* **2015**, *11*, 4615−4626.
(109) Hankins, D.; Moskowitz, J. W.; Stillinger, F. H. Water Molecule Interactions. *J. Chem. Phys.* **1970**, *53*, 4544−4554.
(110) Burnham, C. J.; Xantheas, S. S. Development of Transferable Interaction Models for Water. I. Prominent Features of the Water Dimer Potential Energy Surface. *J. Chem. Phys.* **2002**, *116*, 1479−1492.
(111) Howard, J. C.; Tschumper, G. S. N-body:Many-Body QM:QM Vibrational Frequencies: Application to Small Hydrogen-Bonded Clusters. *J. Chem. Phys.* **2013**, *139*, 184113.
(112) Xantheas, S. S. Ab Initio Studies of Cyclic Water Clusters $(H_2O)_n$, n = 1−6. II. Analysis of Many-body Interactions. *J. Chem. Phys.* **1994**, *100*, 7523−7534.
(113) Axilrod, B. M.; Teller, E. Interaction of the van Der Waals Type Between Three Atoms. *J. Chem. Phys.* **1943**, *11*, 299−300.
(114) Muto, Y. Force between Nonpolar Molecules. *Proc. Phys.-Math. Soc. Jpn.* **1943**, *17*, 629−631.
(115) Řezáč, J.; Huang, Y.; Hobza, P.; Beran, G. J. O. Benchmark Calculations of Three-Body Intermolecular Interactions and the Performance of Low-Cost Electronic Structure Methods. *J. Chem. Theory Comput.* **2015**, *11*, 3065−3079.
(116) Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016**, *116*, 5105−5154.
(117) Dahlke, E. E.; Olson, R. M.; Leverentz, H. R.; Truhlar, D. G. Assessment of the Accuracy of Density Functionals for Prediction of Relative Energies and Geometries of Low-Lying Isomers of Water Hexamers. *J. Phys. Chem. A* **2008**, *112*, 3976−3984.
(118) Bates, D. M.; Tschumper, G. S. CCSD(T) Complete Basis Set Limit Relative Energies for Low-Lying Water Hexamer Structures. *J. Phys. Chem. A* **2009**, *113*, 3555−3559.
(119) Sylvetsky, N.; Kesharwani, M. K.; Martin, J. M. L. Aug-Cc-pVnZ-F12 Basis Sets for Explicitly Correlated Calculations on Anionic Species. *To be submitted for publication*.
(120) Fanourgakis, G. S.; Aprà, E.; Xantheas, S. S. High-Level Ab Initio Calculations for the Four Low-Lying Families of Minima of $(H_2O)_{20}$. I. Estimates of MP2/CBS Binding Energies and Comparison with Empirical Potentials. *J. Chem. Phys.* **2004**, *121*, 2655−2663.
(121) Anacker, T.; Friedrich, J. New Accurate Benchmark Energies for Large Water Clusters: DFT Is Better than Expected. *J. Comput. Chem.* **2014**, *35*, 634−643.
(122) Mezei, P. D.; Ruzsinszky, A.; Csonka, G. I. Application of a Dual-Hybrid Direct Random Phase Approximation to Water Clusters. *J. Chem. Theory Comput.* **2016**, *12*, 4222−4232.

DOI: 10.1021/acs.jctc.6b01046
J. Chem. Theory Comput. 2017, 13, 3136−3152