AIP | The Journal of Chemical Physics

Consistent structures and interactions by density functional theory with small atomic orbital basis sets

Stefan Grimme, Jan Gerit Brandenburg, Christoph Bannwarth, and Andreas Hansen

Citation: The Journal of Chemical Physics **143**, 054107 (2015); doi: 10.1063/1.4927476
View online: http://dx.doi.org/10.1063/1.4927476
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/143/5?ver=pdfcov
Published by the AIP Publishing

Consistent structures and interactions by density functional theory with small atomic orbital basis sets

Stefan Grimme,[a] Jan Gerit Brandenburg, Christoph Bannwarth, and Andreas Hansen
Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms Universität Bonn, Beringstraße 4, 53115 Bonn, Germany

(Received 28 May 2015; accepted 15 July 2015; published online 5 August 2015)

A density functional theory (DFT) based composite electronic structure approach is proposed to efficiently compute structures and interaction energies in large chemical systems. It is based on the well-known and numerically robust Perdew-Burke-Ernzerhoff (PBE) generalized-gradient-approximation in a modified global hybrid functional with a relatively large amount of non-local Fock-exchange. The orbitals are expanded in Ahlrichs-type valence-double zeta atomic orbital (AO) Gaussian basis sets, which are available for many elements. In order to correct for the basis set superposition error (BSSE) and to account for the important long-range London dispersion effects, our well-established atom-pairwise potentials are used. In the design of the new method, particular attention has been paid to an accurate description of structural parameters in various covalent and non-covalent bonding situations as well as in periodic systems. Together with the recently proposed three-fold corrected (3c) Hartree-Fock method, the new composite scheme (termed PBEh-3c) represents the next member in a hierarchy of "low-cost" electronic structure approaches. They are mainly free of BSSE and account for most interactions in a physically sound and asymptotically correct manner. PBEh-3c yields good results for thermochemical properties in the huge GMTKN30 energy database. Furthermore, the method shows excellent performance for non-covalent interaction energies in small and large complexes. For evaluating its performance on equilibrium structures, a new compilation of standard test sets is suggested. These consist of small (light) molecules, partially flexible, medium-sized organic molecules, molecules comprising heavy main group elements, larger systems with long bonds, 3d-transition metal systems, non-covalently bound complexes (S22 and S66×8 sets), and peptide conformations. For these sets, overall deviations from accurate reference data are smaller than for various other tested DFT methods and reach that of triple-zeta AO basis set second-order perturbation theory (MP2/TZ) level at a tiny fraction of computational effort. Periodic calculations conducted for molecular crystals to test structures (including cell volumes) and sublimation enthalpies indicate very good accuracy competitive to computationally more involved plane-wave based calculations. PBEh-3c can be applied routinely to several hundreds of atoms on a single processor and it is suggested as a robust "high-speed" computational tool in theoretical chemistry and physics. [http://dx.doi.org/10.1063/1.4927476]

I. INTRODUCTION

Kohn-Sham density functional theory (KS-DFT, or simply DFT in the following)[1,2] is nowadays the most widely used method for electronic structure calculations of larger molecules and solids. DFT is considered as the natural theory for extended systems but its partial semi-empirical character requires extensive benchmarking on theoretical or experimental reference data. In recent years, such benchmark studies have mainly concentrated on energetic (thermochemical) properties (see e.g., Refs. 3–6) because energies are important for chemistry but difficult to compute accurately. In practice, however, DFT is even more important for routine computations of equilibrium structures (R_e) as starting point for the investigation of various other spectroscopic or thermochemical properties[7] or for higher-level wave function theory (WFT). Even when considering the recent progress in the WFT based computation of large systems (see e.g., Refs. 8–10), it can be expected that geometry optimizations, vibrational frequency calculations, or molecular dynamics applications will be dominated by DFT in the foreseeable future. With this in mind, it seems appropriate to reconsider its accuracy for molecular structures in particular for extended systems and non-covalently bound systems (e.g., biomolecular or molecular crystals).[11,12] The development of a simplified, "low-cost" DFT based compound method which yields primarily good structures but yet reasonable energetic properties is the topic of this work. In addition, an overview on the performance of a few well established and modern density functionals for the computation of a broad range of molecular structures is provided.

With computationally demanding coupled-cluster WFT approaches, relative deviations between theory and experimental data for rotational constants B_e (which overall represent size and shape) of small molecules are tiny[13] (<0.1%) and computed covalent bond lengths are normally

[a]Electronic mail: grimme@thch.uni-bonn.de

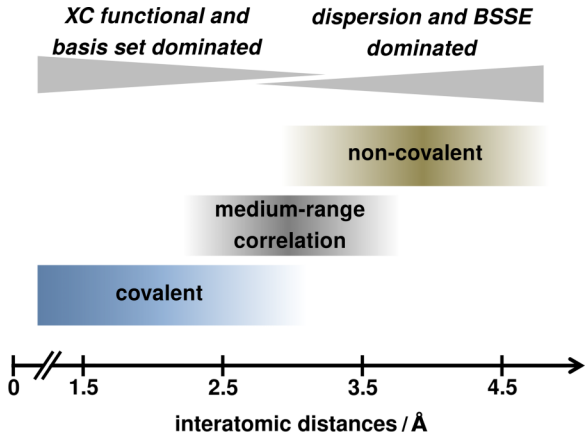

FIG. 1. The three different distance regimes in molecules and associated problems in small basis set DFT calculations (XC: "exchange-correlation," BSSE: "basis set superposition error").

accurate to better than 0.1 pm (0.001 Å).[14] In a recent comprehensive study on the B_e values of nine medium-sized organic molecules in the gas phase (ROT25 benchmark set[15]), it was found that the accurate description of entire molecular structures as reflected in the B_e values is quite challenging. Already in medium-sized systems (20-30 atoms), three basic and relevant inter-atomic distance regimes can be classified as shown schematically in Figure 1.

Short-range electron exchange-correlation (XC) effects, which are covered reasonably accurately by most density functionals, mainly influence covalent bonding, i.e., the distances R_{cov} which are typically <2.5 Å. For large inter-atomic distances >4.5 Å, the interaction is dominated by long-range correlation effects, normally called dispersion (or better London dispersion (LD) to emphasize their long-range character). They are not included in any semi-local (hybrid) density functional and require so-called dispersion corrections (e.g., by our D3 scheme,[16] for reviews, see Refs. 17–19). The effects in the distance regime between these two extrema are nowadays referred to as medium-range correlation[20] (MRC, also called "short-range dispersion") and may be included in modern density functionals to a strongly varying degree. The overall structure of larger molecules is influenced by all distances R_{cov}, R_{MRC}, and R_{LD}. Dispersion corrections consider mainly the latter two regimes while the functionals themselves differ more in the description of the R_{cov} (and partially the R_{MRC}) range. For example, the functional design strategy of the Truhlar group[3,4] is to accurately model the R_{cov} and R_{MRC} distance regimes by flexible and highly parametrized meta generalized-gradient-approximation (GGA) functional forms and to neglect the LD part.

Already early DFT investigations on the molecular structure of electronically simple and small systems revealed a systematic overestimation of covalent bond lengths at the GGA level.[21,22] For heavier main group elements, this effect can be more pronounced (see Section IV B). Hybrid functionals, which replace semi-local DFT XC part by non-local Fock-exchange,[23] suffer less from this problem since Hartree-Fock (HF) generally yields too short bonds and due to the Fock exchange included, this behavior is transferred to a hybrid functional. However, a consistent description of short bonds in light molecules, partially multiply bonded, and long ones (mostly single bonds) in heavier or metallic molecules still seems challenging for DFT.

In simplified DFT schemes which should be routinely applicable to hundreds or even thousands of atoms, two other issues are important (Figure 1). Quantum chemically computed super-molecular interaction energies and molecular structures are subject to the basis set superposition error (BSSE) in an incomplete, atom-centered, one-particle basis set.[24] The BSSE arises due to an unbalanced basis set expansion in the individual (separated) monomers and in the more dense multimer complex. In a dimer complex of the moieties A and B, the basis of the complex is larger than the individual bases of A and B because the unoccupied orbital space from A can be used by B to lower the energy (and vice versa). The well established counterpoise (CP) scheme by Boys and Bernardi[25] covers the intermolecular BSSE but in fact, any close lying molecular assembly (another monomer, alkyl side-chains or bulky substituents) "offers" its unoccupied (virtual) basis functions to another part of the molecule, reducing the energy of the whole assembly. This extends the idea of BSSE to the intramolecular, structurally important case (IBSSE), although a uniform, clear definition is missing. It has to be taken into account when one is aiming at globally consistent molecular structures. For a discussion of different strategies to treat BSSE in mean-field theories, see, e.g., Ref. 26. At the other end, we need to consider the related basis set incompleteness error (BSIE). Normally, it leads to too long covalent bonds because the various bonding effects (electron sharing, polarization, and charge-transfer) are less well described in a small basis than the classical Coulomb and non-classical Pauli repulsion between electrons. The tendency of the BSIE to lengthen bonds is systematic[27] and hence, empirical schemes can be used to compensate for such methodological deficiencies.

Large part of the success of standard quantum chemistry originates from systematic error compensation that allows routine calculations with finite basis sets and other truncated wave function expansions. This is also the idea behind the present approach where we try to maximize the accuracy-computational effort ratio for large systems without introducing too much empiricism. The present work was triggered by the good performance of the recently proposed minimal basis set Hartree-Fock approach with atom-pairwise dispersion and basis set corrections (HF-3c)[28] as well as the performance of small basis set DFT for structural data of bio-molecular systems.[11] In our case, we choose a variant of the PBE0 global hybrid density functional[29,30] (GHDF) (dubbed PBEh here) which yields reasonable bond lengths in large basis set calculations (see below). Empirically, it is observed that typical covalent bond lengths are shortened with increasing amount of Fock exchange admixture in common global hybrid functionals. As will be shown here, a modified version of PBEh with increased amount of Fock-exchange (which "shortens" bonds) together with a small, polarized valence-double-ζ (DZ) basis set (which "lengthens" bonds due to BSIE) leads to distances that are on average correct. In order to account for the remaining (I)BSSE as well as the missing London dispersion, the well-established atom-pairwise corrections schemes gCP

(geometrical counter-poise)[31] and D3 for London dispersion (with Becke-Johnson damping),[16,32] respectively, are used.

The idea to use low-level quantum chemical methods for structure optimizations and to benefit from error compensation between a small atomic orbital (AO) basis and an incomplete (or even absent) treatment of electron correlation is rather old. Several years ago, Pople noted that HF/STO-3G optimized geometries for small molecules are excellent, even better than HF is inherently capable of yielding in a complete basis.[33,34] Similar observations were made by Kołos in 1979, who obtained good interaction energies for a HF/minimal-basis method together with a counterpoise-correction as well as a correction to account for the London dispersion energy.[35] Brothers and Scuseria furthermore showed that the computation of enthalpies of formation with DFT and minimal basis sets can be largely improved by treating atomic energies as empirical parameters.[36] Regarding energies, we mention the EDF1 method of Adamson, Gill, and Pople which specifically was designed to work with the 6-31+G* AO basis set.[37] For previous composite DFT methods that were specifically designed to work with medium-sized basis sets, see, e.g., Ref. 38. Also related is the approach of DiLabio and co-workers which employs a standard functional in a small basis (B3LYP/6-31++G**) with BSSE correction and London dispersion fitted into effective core potentials (ECPs). This method combination is specifically designed to work for bond dissociation enthalpies and non-covalent binding energies.[39,40] In the context of correlated wave function methods that strongly rely on error cancellation, we mention the Coulomb attenuated variants of MP2 in combination with an augmented double-zeta basis.[41] Here, the long-range attenuation of the Coulomb operator cancels most of the BSSE and, in combination with the small basis set, it reduces some intrinsic MP2 errors in the description of London dispersion interactions. For a recent study on the balance of BSIE and BSSE in explicitly correlated WFT methods, see Ref. 42.

Based on the observations mentioned above and the fact that the D3 scheme provides easy access to relatively accurate dispersion energies (typical relative error of 5%-10%), we recently proposed the HF-3c method.[28] It performs well for structures and non-covalent interaction (NCI) energies and can be applied to thousands of atoms routinely. However, it lacks accurate thermochemistry and requires, in addition to D3 and gCP, a further BSIE correction for bonds between electronegative atoms. Since it is based on HF, it should not be applied to electronically more complicated (open-shell or metallic) systems.

The main aim of the present study is to fill the gap between existing semi-empirical methods or HF-3c and large basis set DFT in terms of the cost-accuracy ratio with a physically sound and numerically well-behaved approach. It is developed along similar lines as HF-3c but with the following conceptual and technical changes:

1. Replacing HF by hybrid DFT introduces short-range and static electron correlation effects and allows treatment of electronically more complicated situations like transitions states, open-shell species, or transition metal complexes.
2. Replacing the minimal by a polarized valence-double-zeta basis set significantly improves the energetic description without sacrificing the computational efficiency too much.
3. The method is less empirical and contains only two instead of three atom-pair wise corrections.

The method is dubbed PBEh-3c from now on to indicate its origin and the GGA components as usual. The abbreviation highlights the relation to HF-3c and more specifically, "3c" stands for the (slightly modified, see below) gCP (for (I)BSSE), the D3 (for dispersion) corrections, and minor modifications of the basis set to ensure consistent bond lengths across the periodic table. Similar to HF-3c, it does not suffer significantly from the self-interaction error (SIE) (also called delocalization error[43]) and should represent an alternative to semi-local functionals in such problematic cases. Note that we are not recommending the use of small AO basis sets in general. Whenever computationally feasible, expansions close to the complete basis set (CBS) limit should be the preferred procedure, e.g., the Ahlrichs-type QZVP sets[44,45] are used routinely for energy calculations in our group for some years. Different from other common density functional developments, the present work is primarily focused on a good description of molecular structures and their aggregates in the condensed phase. Additionally, the performance for standard thermochemistry is considered in some detail with the perspective of being reasonably accurate in typical applications and much better than it is possible with HF-3c. We prefer to use well-tested standard functional components instead of sometimes numerically problematic[46] Taylor-series expansions as, e.g., used in the Minnesota functionals[4] or the B97 class which has recently been investigated in great detail.[47]

The M06-2X[3] and B3LYP[23,48,49] functionals together with small DZ type valence basis sets (and corresponding ECPs for heavier elements) as competitors are used for comparison to the proposed PBEh-3c method. They are evaluated "plain" without any correction because they are still used in this form in many applications. Although B3LYP can be improved by adding D3 and gCP corrections[50] and M06-2X through the D3(zero-damping) scheme,[51,52] we think that a comparison to the plain functionals also provides insight how large dispersion and BSSE effects (and their mutual compensation) are in typical systems. A separate and detailed analysis of these effects is, however, beyond the scope of the present study which focuses on the new approach. In order to put the results into some broader perspective, the dispersion corrected TPSS-D3 functional[53] and the PBE0-D3 hybrid functional[30] in a larger Ahlrichs-type triple-ζ (def2-TZVP) basis set[45] are tested. TPSS-D3/def2-TZVP has similar (slightly larger) computational costs compared to PBEh-3c and has been our standard density functional/basis set combination for structure optimizations for some years.

II. THEORY

The starting point for calculating the electronic energy is a standard hybrid density functional treatment with a small Gaussian AO basis set. Two terms are added to correct the DFT energy E_{tot}^{PBEh} in order to include long-range dispersion

interactions and to account for the BSSE. We modify the electronic part of the energy (functional and basis set) for optimum performance. Hence, we use the abbreviation "3c" for the entire scheme in analogy to our previous HF-3c approach[28] which includes an additional atom pair-wise term (but no changes in the electronic HF part).

The total energy is calculated as

$$E_{tot}^{PBEh\text{-}3c} = E_{tot}^{PBEh} + E_{disp} + E_{BSSE}^{damped\ gCP}, \tag{1}$$

where PBEh denotes the density functional used (PBEh in the double-zeta basis set, see below) and the first correction term E_{disp} is the London dispersion energy from the D3 scheme.[16] It consists of the leading atom-pair wise (indicated by superscript (2)) term with Becke-Johnson (BJ) damping,[32,54,55]

$$E_{disp}^{(2)} = -\frac{1}{2}\sum_{A}^{atoms}\sum_{B}^{atoms}\left(\frac{C_6^{AB}}{R_{AB}^6 + \left(a_1 R_{AB}^0 + a_2\right)^6} + s_8\frac{C_8^{AB}}{R_{AB}^8 + \left(a_1 R_{AB}^0 + a_2\right)^8}\right). \tag{2}$$

Here, C_n^{AB} denotes the nth-order dispersion coefficient (orders = 6, 8) for each atom pair AB, R_{AB} is their internuclear distance, and s_n are the order-dependent scaling factors. The cutoff radii $R_{AB}^0 = \sqrt{C_8^{AB}/C_6^{AB}}$ and the fitting parameters a_1, and a_2 are used as introduced in the original works.[54,55] For the present method, the three usual parameters s_8, a_1 and a_2 were re-fitted using reference interaction energies of the S22 set.[56] During the fitting procedure linear dependencies between the D3 and gCP parameters appeared and hence s_8 was set to zero for PBEh-3c, i.e., the second term in Eq. (2) is dropped. Cross-checks were conducted for interaction energies for the 528 complexes of the S66×8 set[57] as well as on the L7[58] and the S12L[59] NCI energies benchmark sets.

Already in the original D3 publication, inclusion of the Axilrod-Teller-Muto (ATM) three-body dispersion was considered. The corresponding C_9 coefficients are approximated from the C_6 coefficients. The ATM contribution $E_{disp}^{(3)}$ is evaluated in an atom-triple-wise scheme and efficient analytical gradients were implemented recently by our group. For the energy expression of $E_{disp}^{(3)}$, we refer to the original paper.[16] Here, we propose for the first time to include it by default for the computation of the total dispersion energy (and its geometrical derivatives) according to

$$E_{disp} = E_{disp}^{(2)} + a_3 E_{disp}^{(3)}. \tag{3}$$

The factor a_3 (which is unity by default) is simply used to switch the ATM term on (or off if $a_3 = 0$) for a particular functional but could be used in principle for further empirical modification. The $E_{disp}^{(3)}$ correction is relatively small for medium-sized molecules but should be taken into account for larger, more dense systems (see Refs. 60–62 for a detailed discussion of many-body dispersion effects). Note, that the neglected R^{-8} two-body and newly introduced R^{-9} three-body dispersion terms do not depend on (or compensate) each other. Moreover, they are of very different magnitude, i.e., higher-order multipole contributions, which are effectively contained in the D3-R^{-8} term, can reach up to 50% of the dispersion energy in DFT-D3 (strongly depending on the density functional) while the ATM term typically amounts to only 2%–3% of the two-body dispersion energy.

The second correction term $E_{BSSE}^{damped\ gCP}$ denotes a slight modification of our recent geometrical CP scheme.[31] Here, the difference in atomic energy E_A^{miss} between a large (nearly complete) basis set and the target basis set (here def2-mSVP, see below) for each free atom A is calculated at the Hartree-Fock level and used as a measure for basis incompleteness. By cross checking protein ligand binding affinities and solid state geometries of bromine containing systems, we detected some inconsistencies. Thus, we replaced the fourth-row E_A^{miss} with the corresponding DZP values. While the non-covalent binding energies and geometries for heavy element containing systems improved significantly, covalent bonds are not affected. The E_A^{miss} term is multiplied with a decay function depending on the inter-atomic distances R_{AB}. The BSSE in self-consistent-field (SCF) methods strongly depend on the charge density overlap. The density has an exponential tail, which is adapted in the decay function. The sum over all atom pairs reads

$$E_{BSSE}^{gCP} = \sigma\sum_{A}^{atoms}\sum_{A\neq B}^{atoms} E_A^{miss}\frac{\exp\left(-\alpha(R_{AB})^{\beta}\right)}{\sqrt{S_{AB}N_B^{virt}}}, \tag{4}$$

where α, β, and σ are fitting parameters, S_{AB} is a Slater-type overlap integral, and N_B^{virt} is the number of virtual orbitals on atom B in the target basis. The S_{AB} is evaluated over a single s-type orbital centered on each atom and using optimized Slater exponents weighted by the fourth fitting parameter η. In combination with the D3 correction, the various DFT-gCP-D3/"small basis" methods perform well for NCIs but can still be recommended for accurate binding energies. For further details and recent applications see Refs. 31, 50, and 63. Because the method proposed here puts more emphasis on accurate thermochemistry, we investigated the effect of gCP on this property in more detail. It turned out that in a few cases, the above correction introduces artifacts in the short-range inter-atomic part and furthermore adds too much repulsive force to short covalent bonds. In HF-3c, this could be partially compensated by the third correction (which is omitted here) and thermochemistry was not in the focus.

Hence, we decided to keep the basic form of gCP but to damp its short-range part according to

$$E_{BSSE}^{damped\ gCP} = E_{BSSE}^{gCP}\frac{1}{1 + k_{dmp}^1\left(R_{AB}/(R_0'^{AB})\right)^{-k_{dmp}^2}}, \tag{5}$$

where $R_0'^{AB}$ are the D3 damping radii.[16] This damping function was proposed first by Chai and Head-Gordon[64] and already used in the zero-damping version of the D3 method. The two parameters in Eq. (5) ($k^1_{dmp} = 4$ and $k^2_{dmp} = 6$) were adjusted by "inspection" to exclude the above mentioned bonded region from the gCP contribution. They are kept fixed in the least-squares fitting of the D3/gCP parameters.

In order to achieve more flexibility with the D3/gCP corrections, the fitting procedure was applied simultaneously to the D3 part (a_1, a_2, and initially also s_8) and three of the gCP parameters (α, β, η; the global scaling factor σ was fixed to unity). All optimum parameters have reasonable values similar to those obtained previously for similar density functionals and basis sets.[31,32] During the cross-check phase of PBEh-3c, we observed some inconsistencies for the NCIs involving hydrogen and the rare gas elements and consequently, we set the corresponding E_A^{miss} values to zero which solved these problems.

Various combinations of GGA exchange and correlation components were tested in combination with various small basis sets (minimal to simplified triple-zeta). We exclusively employed well-tested standard GGA components which are available in most common quantum chemistry codes. More specifically, we tried the B97[65] GGA and B95[66] meta-GGA forms, PBE exchange (with modified enhancement factor, parameters μ and κ, see below) in combination with LYP[67] correlation, functionals with modified local density approximation (LDA) components, and even variants with a range-separated hybrid part. For reasons of computational efficiency as well as numerical stability and according to "Ockham's razor," we finally chose one of the simplest possibilities, a modified form of PBEh which is discussed in greater detail below. A summary of the construction principle and components of the 3c-methods is given in Table I.

A. PBEh-3c construction

The global hybrid density functional part is based on PBE GGA exchange and correlation[29] according to

$$E_{XC}^{PBEh} = (1 - a_x)E_X^{PBE} + a_x E_X^{HF} + E_C^{PBE}, \tag{6}$$

where the parameter a_x controls the admixture of non-local HF exchange. This combination was first proposed by Adamo and Barone[30] with $a_x = 0.25$. We tested a similar composite scheme based on a range-separated PBE hybrid (RSH) functional[68] but this variant turned out to be less robust for thermochemistry. Because RSH implementations are furthermore slower than GHDFs and less widespread available in typical quantum chemistry codes, a GHDF was preferred. As mentioned above, different GGA components were tested but not found to be superior to PBE in general. In order to have some flexibility in the electronic part, it was decided to change the parameters appearing in the PBE expressions.

The enhancement factor F_X for PBE exchange is given by

$$F_X^{PBE} = 1 + \frac{\mu s^2}{1 + \frac{\mu s^2}{\kappa}}. \tag{7}$$

The parameters μ and κ have been modified several times in the literature and we mention here only the versions with $\kappa = 1.245$ and $\mu = 0.219\,514\,9$ (original μ value) termed revPBE[69] as well as $\kappa = 0.804$ (original κ value) and $\mu = 0.123\,456\,79$ (PBEsol[70]). Because revPBE yields better atomization energies than PBE but worse bond lengths,[71] we propose here a compromise, i.e., the average of both values of $\kappa = 1.0245$. Because PBEsol describes certain difficult electron correlations better than PBE,[72] we chose the PBEsol value of $\mu = 0.123\,456\,79$ in PBEh-3c. The PBE correlation functional E_C^{PBE} contains a parameter β_{PBE} which controls the amount of GGA correction to the LDA correlation energy. Originally, it reads $\beta_{PBE} = 0.066\,72$ but is reduced in PBEsol to a value of 0.046, i.e., increasing the LDA character for the correlation energy. We treat this parameter empirically here in order to compensate for basis set incompleteness effects for thermochemistry (atomization energies). A "hand-made" adjustment to atomization and total energies of a few small molecules and reaction energies from the GMTKN30 data base led to a value of $\beta_{PBE} = 0.03$ for PBEh-3c.

Because we want to develop a general purpose functional, the choice of the critical Fock-exchange admixture parameter a_x is non-trivial.[73] A further complication arises from our requirement to yield accurate structures (bond lengths). In the chosen relatively small AO basis set, this can only be achieved with a_x values between about 0.3 and 0.5. A relatively large value of $a_x = 0.375$ has been used successfully in recent works by us.[16,74] Values of this magnitude were found to be sufficient to avoid artificial charge-transfer over long distances in charged systems, yield good reaction barriers, and reasonable orbital energy gaps, excitation energies, and dynamic polarizabilities in time-dependent DFT treatments. After some testing, a_x values around 0.4 appeared appropriate and we settled for $a_x = 0.42$ as suggested first in the Boese-Martin functional for kinetics (BMK).[75] The parameters defining the new functional are given in Table II. In total 12, empirical parameters were considered but only seven of them were actually adjusted to theoretical reference data.

The last modification concerns the AO basis set. In general, the Ahlrichs-type split valence double-zeta basis sets def2-SV(P)[45,76] (together with the corresponding ECPs[77] for heavier elements) is employed. It has the advantage of being consistently available for almost all elements of the periodic

TABLE I. Overview about the 3c-methods in comparison to our previous standard DFT optimization level.

	HF-3c[a]	PBEh-3c	TPSS-D3/def2-TZVP
Adjusted electronic part	No	Yes	No
Atom-pairwise dispersion	Yes	Yes	Yes
Three-body ATM dispersion	No	Yes	No
Atom-pairwise BSSE correction	Yes	Yes	No[b]
Atom-pairwise BSIE correction	Yes	No	No
AO basis set level	Minimal[c]	DZ	TZ

[a]Reference 28.
[b]The typical BSSE for NCI is 5%-10% of ΔE.
[c]def2-SV(P) basis for elements $Z > 18$.

TABLE II. Empirical parameters of the PBEh-3c method.

	Parameter			
Contribution	κ	μ	β_{PBE}	a_x
E_{xc}	1.0245[a]	0.123 456 79[a]	0.03	0.42
	a_1	a_2	s_8	a_3
E_{disp}	0.4860	4.5	0[a]	1[a]
	α	β	σ	η
E_{gCP}	0.276 49	1.9560	1[a]	1.324 92

[a]Constrained value.

table and that it contains a smaller number of primitive Gaussian functions compared to other DZ sets thus improving computational efficiency. For the elements B–Ne, the similar DZ valence sets[76] are used which contain one uncontracted valence s-function more than the corresponding split-valence (SV) sets leading to slightly improved (shortened) bond lengths. Furthermore, it was found that the standard exponents of d-polarization functions are not optimum. We took values resembling those from the 6-31G* basis[78] set, i.e., $\alpha(d) = 0.8$ for carbon to fluorine. The standard exponents of the polarization functions in def2-SV(P) are employed for boron ($\alpha(d) = 0.5$) and neon ($\alpha(d) = 1.8$). A weakness of the def2-SV(P) set is the systematic overestimation of bond distances involving hydrogen mainly caused by the missing p-polarization functions on hydrogen atoms. This effect could be largely compensated by scaling of all its s-function exponents by a value of 1.2^2 which accounts for increased effective nuclear charge in typical molecular environments. This factor has been adjusted manually to reproduce X–H bond lengths and atomization energies of the methane, ammonia, and water molecules. Note that this scaling increases the energy of the hydrogen atom by only about 1 mE_h. For all other elements, the def2-SV(P) basis is used without any changes and the entire new set is termed def2-mSVP ("m" stands for modified) from now on (see also the supplementary material[79]).

The implementation of PBEh-3c into any quantum chemistry code is straightforward if the software infrastructure for a hybrid functional is available. For the D3-gCP, freely available codes can be downloaded from our website.[80]

III. TECHNICAL DETAILS OF THE CALCULATIONS

The molecular calculations were carried out using TURBOMOLE 6.6[81,82] or a development version of the ORCA 3.0 program.[83,84] The employed Gaussian AO basis sets are def2-SV(P) and def2-mSVP as described above. Some test calculations using the computationally more costly 6-31G* Pople basis gave similar results compared to def2-mSVP and hence are not reported. Furthermore, the def2-SV(P) basis set was chosen instead of the more popular 6-31G* because it is consistently available for all employed elements. Note that def2-SV(P) and def2-mSVP involve practically the same computational costs. Large basis set results for comparison were taken from previous work and usually refer to def2-QZVP[44] or def2-QZVPD (def2-QZVP with additional diffuse functions[85]) sets which provide results close to the CBS limit in mean field treatments. For the larger molecules, the RI-J approximation was used[86–88] with default auxiliary basis sets.[89]

For the semi-local exchange-correlation part, the numerical quadrature grid m4[90] (m5 for M06-2X) was used (grid5 in ORCA) and tight SCF convergence criteria were applied in geometry optimizations. For the D3-gCP corrections, our freely available codes have been used.[80] The new method was implemented into the ORCA and TURBOMOLE codes and will be available in the next releases. In ORCA, it is invoked with the compound keyword "pbeh-3c" while in TURBOMOLE, it is used as functional name.

The periodic calculations were carried out with a development version of CRYSTAL14.[91] The basis sets and the compound DFT keyword PBEh-3c have been implemented and will be available in the next release. Note that due to its numerical complexity, some M06-2X SCF calculations did not converge and are omitted.[79] The results for the estimated CBS were taken from previous work and were conducted with a projector-augmented plane wave (PAW[92,93]) basis set with large energy cutoff of 1000 eV as implemented in VASP 5.3.[94,95]

In some cases (e.g., for anions), it is necessary to add diffuse functions for specific atoms as usual. According to some test calculations for PBEh-3c, it seems to be the most robust strategy to keep the gCP parameters in this case. The gCP program prints out a warning about this basis set mismatch and there are small inconsistencies in the treatment of BSSE. In particular, NCIs may be described less well for such choices because of the subtle balance of gCP and D3 correction terms (which are numerically of different sign).

For the computation of thermodynamic properties, we propose to use PBEh-3c frequencies scaled by a factor of 0.95. The scaling was adjusted to reproduce the experimental heat capacities of ten small organic molecules (formic acid, butane, ethyne, ammonia, tetramethylsilane, benzene, acetone, and neopentane).[96]

Throughout the paper, we will sometimes employ the short-hand notations "/L" for the large def2-QZVP(D) AO basis set, "/M" (medium) for def2-TZVP, and "/S" for def2-SV(P). The term "3c" always implies the def2-mSVP basis set. The quality of def2-SV(P) and def2-mSVP is very similar in typical applications and we will not distinguish them in the discussion of the statistical data. As noted above, many of our conclusions apply similarly to the 6-31G* basis set.

IV. RESULTS AND DISCUSSION

The proposed composite method shall be applicable to all kinds of organic and organo-metallic systems with finite orbital gap. Due to its low-cost character, the main focus is on structural properties and non-covalent bond energies in large systems. At the same time, kinetic and thermochemical properties shall be within reasonable accuracy as tested with the huge general main-group, thermochemistry, kinetics, and non-covalent interactions (GMTKN30) database.[5] Before this is discussed, first, two new molecular structure benchmarks will be introduced.

TABLE III. Experimental[a] reference bond distances (in pm) for 11 molecules containing third-row or higher main group elements (HMGB11 set).

Molecule	Bond	Experimental
Cl_2	Cl–Cl	198.8
S_2H_2	S–S	205.5
P_2Me_4	P–P	221.2
Br_2	Br–Br	228.1
Se_2H_2	Se–Se	234.6
Ge_2H_6	Ge–Ge	241.0
As_2Me_4	As–As	242.9
Te_2Me_2	Te–Te	268.6
Sn_2Me_6	Sn–Sn	277.6
Sb_2Me_4	Sb–Sb	281.8
Pb_2Me_6	Pb–Pb	288.0

[a]Reference 97 and references therein.

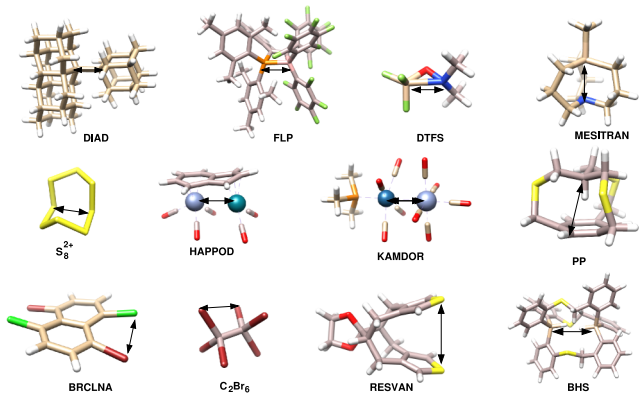

FIG. 2. Structures of the investigated larger molecules in the LB12 test set. The arrows indicate the considered interatomic distances.

A. HMGB11 and LB12 benchmark

Typical geometry benchmark sets focus on bond lengths in light (first and second row) molecules. In order to evaluate the PBEh-3c functional more broadly, we compiled two new molecular structure benchmark sets. The HMGB11 (heavy main group bond) set complements the standard equilibrium bond lengths set of small first and second row molecules with small to medium-sized molecules consisting of third-row or higher main group elements. These elements are typically not in the fit sets of empirical density functionals and their careful benchmarking seems necessary. As shown below, the HMGB11 error spread of the different methods (differences in the mean absolute deviations (MADs)) is significantly larger compared to the first and second row molecules and even larger than for 3d-transition metal complexes. The HMGB11 molecules with the experimental reference data are given in Table III.

As a further cross-check, we compiled a set of larger systems with chemically interesting bonding situations (LB12, for 12 long bonds). The examples were taken in part from previous works[32,98–101] where they appeared as critical test cases for electronic structure methods. The geometries and considered interatomic distances are shown in Figure 2 and explicitly given in Table IV. Because not all experimental values refer to isolated molecule (gas phase) conditions and crystal packing effects may still be present, and furthermore no vibrational (thermal) corrections are made, we consider deviations between theory and experiment of 1-2 pm as negligible. Similar considerations apply to HMGB11.

So-called Frustrated Lewis Pairs (FLP) have attracted a lot of attention in the chemical community for the ability to activate small molecules at ambient conditions.[112] The investigated system has a typically weak P–B donor-acceptor type interaction augmented by intramolecular non-covalent stacking interactions between the aryl rings. The (*N,N*-dimethylaminoxy)trifluorosilane (DTFS) and methyl-silatrane (MESITRAN) molecules show unusual N → Si donor-acceptor bonds with a distance well below that of typical van der Waals (vdW) interactions but longer than covalent first-second row bonds.[104,105] Two larger organometallic complexes with data base codes KAMDOR and HAPPOD are considered as examples for 3d-4d closed-shell transition metal interactions. These bonds are rather long and involve low electron densities and significant dispersion contributions. The interaction between two formally non-bonded sulphur atoms in S_8^{2+} was found to be badly described by many DFT methods[32,106] and can be considered as a borderline case between covalent and non-covalent bonding. Pascal's cyclophane (PP) and the in,in-bis(hydrosilane) (BHS) are examples for hydrogen atoms pointing inwards which creates short non-bonded contacts well below typical vdW distances (e.g., a H···H distance of only about 160 pm in BHS). The sulphur-sulphur, bromine-bromine, and bromine-chlorine distances in RESVAN, C_2Br_6, and the 1,2,5,6,dibrom-dichlornaphthalene, respectively, are examples for intramolecular non-bonded contacts between heavier elements. In all cases, covalent as well

TABLE IV. Experimental references for ground state bond distances (in pm) of 12 molecules with unusual bonds or intramolecular, non-bonded interactions (LB12 set).

Molecule/code	Bond	Experimental
DIAD	C–C	171[a]
FLP	P–B	212[b]
DTFS	Si–N	227[c]
MESITRAN	Si–N	245[d]
S_8^{2+}	S–S	286[e]
HAPPOD	Rh–Cr	308[f]
KAMDOR	Os–Cr	310[g]
PP	C–C	312[h]
BRCLNA	Br–Cl	313[i]
C_2 Br_6	Br–Br	342[j]
RESVAN	S–S	419[k]
BHS	Si–Si	443[l]

[a]X-ray value from Ref. 102.
[b]X-ray value from Ref. 103.
[c]Electron diffraction value from Ref. 104.
[d]Electron diffraction value from Ref. 105.
[e]X-ray value from Ref. 106.
[f]X-ray value from Ref. 107.
[g]X-ray value from Ref. 99.
[h]X-ray value from Ref. 108.
[i]X-ray value from Ref. 109.
[j]X-ray value from Ref. 110.
[k]Estimated gas phase value from Ref. 100.
[l]X-ray value from Ref. 111.

as MRC and London dispersion effects are in detailed balance. Actually, the shortest distance considered is a covalent single bond in the 2-(1-diamantyl)[121]tetramantane (DIAD) system of Schreiner *et al.*[102] which represents the "world-record" for the longest C–C bond. Cartesian coordinates at the PBEh-3c level are given in the supplementary material.[79]

B. Molecular structures

The structural analysis is ordered as follows. We first investigate equilibrium bond distances of various relatively small molecules (first and second row molecules, heavy main group covalent bonds, and 3d-transition metal complexes). Second, we test the correct balance between the bonding interactions and the medium- and long-range NCIs in medium sized molecules. Results for the recently established set of rotational constants (ROT34[113]), a newly compiled set of unusually long bonds (LB12), and structures and conformations of small peptides (P26[114]) are presented. Noncovalent bond distances are judged via the well known S22[56] and S66[57] molecular dimer test sets. We always consider the new compound method PBEh-3c compared to the M06-2X and B3LYP functionals evaluated in the similar def2-SV(P) basis set. Results for our current default routine geometry optimization level TPSS-D3/M and the hybrid PBE0-D3/M are given as well to show the typical good accuracy of dispersion corrected functionals evaluated in larger basis sets. Further increase of the basis set to a quadruple-zeta level normally has only minor effects on bond distances.[113,115] Note that the latter two methods (especially the hybrid PBE0) are computationally more expensive than the three small basis set approaches. For instance, the PBE0/M energy evaluation for a NCI complex with about 100 atoms is slower by approximately a factor of 20 compared to the PBEh-3c calculation.

The set of first and second row molecules (dubbed LMGB35, light main group bonds) contain the systems H_2, HF, H_2O, HOF, OH, NH_3, OH^+, NH, C_2H_2, NO^+, HCN, NH^+, C_2H_4, CH_4, N_2, CH_2O, N_2^+, O_2^+, CH, CO, HCN, CO_2, C_2H_2, CH_2O, BO, O_2, BH, BF, CF, NF, F_2^+, C_2H_4, F_2, HOF, and B_2.[116] As benchmark for 3d-transition metal systems, we use the well established set of 32 complexes with 50 analyzed bond distances compiled by Bühl and Kabrede (dubbed TMC32 in the following).[117] The set consists of the complexes $Sc(acac)_3$, $TiCl_4$, $TiMeCl_3$, $TiMe_2Cl_2$, $Ti(BD_4)_3$, VOF_3, VF_5, $VOCl_3$, $V(NMe_2)_4$, $V(Cp)(CO)_4$, CrO_2F_2, CrO_2Cl_2, $CrO_2(NO_3)_2$, $Cr(C_6H_6)_2$, $Cr(C_6H_6)(CO)_3$, $Cr(O)_4$, MnO_3F, $MnCp(CO)_3$, $Fe(CO)_5$, $Fe(CO)_3$, $Fe(CO)_2(NO)_2$, $FeCp_2$, $Fe(C_2H_4)(CO)_4$, $Fe(C_5Me_5)$, $CoH(CO)_4$, $Co(CO)_3(NO)$, $Ni(CO)_4$, $Ni(acac)_2$, $Ni(PF_3)_4$, $CuCH_3$, CuCN, and $Cu(acac)_2$.

1. Equilibrium bond distances of small molecules

We separate the analysis of bond distances into light main group (LMGB35, first and second row), heavy main group (HMGB11, third row and higher, compare Section IV A), and 3d-transition metal systems (TMC32). The individual data are given in the supplementary material[79] (reference values and results for all considered functionals). The statistical evaluation of all three sets is summarized in Table V.

TABLE V. Comparison of experimental and calculated ground state equilibrium bond distances R_e (in pm) for 35 small first and second row molecules,[a] third-row or higher main group molecules, and 3d-Transition metal complexes.[b] The two smallest MADs in each subset are written in bold letters.

Measure	PBEh-3c	M06-2X/S[c]	B3LYP/S[c]	TPSS-D3/M	PBE0-D3/M
LMGB35 (and second row molecules)					
MD[d]	−0.6	−0.2	0.7	0.8	−0.4
MAD[e]	**0.9**	1.5	1.2	**0.9**	1.0
SD[f]	1.5	2.0	1.4	0.7	1.4
MAX[g]	5.3	6.2	3.6	3.0	4.5
HMGB11 (third-row or higher main group molecules)					
MD	−0.2	2.0	5.1	1.9	−0.3
MAD	**0.8**	2.2	5.1	1.9	**0.9**
SD	1.2	1.4	1.9	1.3	1.1
MAX	2.2	4.1	7.9	3.8	2.0
TMC32 (3d-transition metal complexes)					
MD	−2.7	0.4	−0.5	−1.0	−0.5
MAD	3.3	3.2	**1.7**	1.9	**1.4**
SD	2.7	4.1	2.2	3.1	1.9
MAX	6.0	8.3	7.7	12.4	7.0

[a]Reference 116.
[b]Reference 97 and references therein.
[c]def2-SV(P) basis set and the corresponding ECP was used.
[d]Mean deviation.
[e]Mean absolute deviation.
[f]Standard deviation.
[g]Maximum absolute deviation.

For the light molecules, all methods give reasonable results. TPSS-D3/M yields systematically too long bonds but has a low error spread (standard deviation, SD) and a MAD below 1 pm. While the MAD of PBE0-D3/M is similar, the mean deviation (MD) is smaller with slightly too small distances. Out of the small basis set methods, only PBEh-3c has a similarly small MAD below 1 pm while M06-2X/S and B3LYP/S are slightly worse. Part of the relatively small MD and large MAD for PBEh-3c (and this also holds for M06-2X/S) can be attributed to the outliers F_2^+ and F_2 which are badly described with large amounts of Fock-exchange in the functional. Excluding these two cases yields very small MD and MAD values of −0.3 pm and 0.7 pm, respectively, for PBEh-3c (0.2 and 1.3 pm for M06-2X/S). Despite its systematic bond length overestimation, the semi-local TPSS-D3 functional has a very small SD because it implicitly accounts for static correlation effects and hence is not affected by these outliers.

For the heavier main group molecules, the deviations are more pronounced. The results for the low-cost methods and TPSS-D3/M for comparison are graphically shown as deviations in Figure 3. PBEh-3c performs excellently with a MAD below 1 pm and as good as the more expensive PBE0-D3/M method. It is significantly better compared to TPSS-D3/M and the other small basis set methods. Since these molecules exhibit no significant static electron correlation effects, the semi-local functional TPSS does not perform better than hybrids. Obviously, the bonding between the larger, more polarizable atoms requires inclusion of non-local Fock-exchange as well as a consistent treatment of MRC effects as indicated by the very bad B3LYP results.

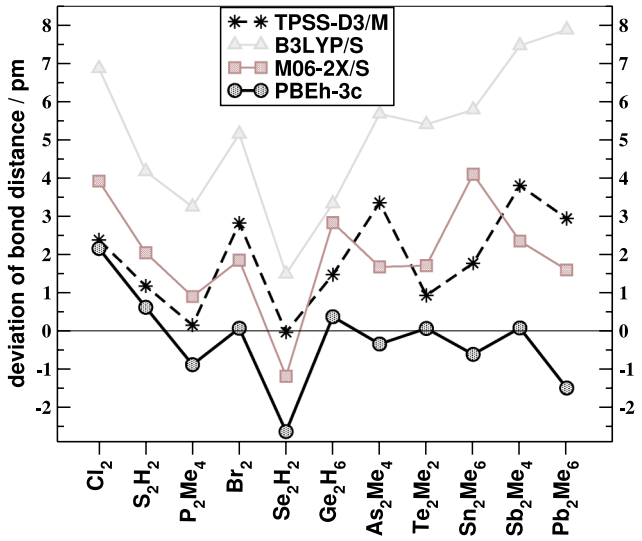

FIG. 3. Deviations of bond distances from experimental reference values for the HMGB11 test set.

For 3d-transition metal complexes, a large amount of Fock-exchange is expected to deteriorate the results. To some extent, this is also observed for the complexes tested here (PBEh-3c and M06-2X have the largest MAD). For those systems, GGAs or hybrids with only small portions of Fock-exchange are the best choice. However, the obtained bond distances are not unreasonably bad with our new method and we can still recommend it for organometallic chemistry. Part of its relatively large MAD results from a systematic underestimation of M=O distances in some oxo-complexes due to the elevated HF character of PBEh-3c. The typical carbonyl or cyclopentadienyl complexes on the other hand are well described. Overall, PBE0-D3/M performs very well and can be generally recommended for such systems. PBEh-3c seems to be an ideal alternative if the larger basis set is not computationally feasible or SIE errors play a role so that semi-local functionals cannot be used.

2. Structure of medium sized molecules and small peptides

Having analyzed the covalent bond distances, we test the correct balance between the bonding interactions and the medium- and long-range NCIs. This is important to get accurate structures of larger molecules and for "soft" biochemical systems in particular. The statistical evaluation of all sets considered here is given in Table VI. First, we investigate the rotational constants of the recently introduced ROT34 benchmark set. The molecules in the ROT34 set are ethynyl-cyclohexane, isoamyl-acetate, diisopropyl-ketone, bicyclo[2.2.2]octadiene, triethylamine, vitamin C, serotonine, aspirin, the natural product cassyrane, proline, lupinene, and limonene. The last three molecules were added to the ROT25 set in Ref. 113 yielding the ROT34 benchmark. The conformational species investigated in each case (as taken from the experimental studies) are shown in the inset of Figure 4. For reference to the experimental data and more details see Refs. 15 and 113. Typical medium-sized organic molecules with various functional groups and some conformational flexibility are included which allows us to draw general conclusions. We consider the equilibrium rotational constants B_e as a measure of molecular structure. Its accurate computation requires a consistently good description

TABLE VI. Comparison of experimental and calculated geometries of medium sized molecules as judged by deviations in the rotational constants (ROT34[a]), deviations in the bond lengths of unusual bonds or intramolecular, non-bonded interactions (LB12 set[b]), and heavy atom Cartesian root-mean square deviation (RMSD)[118] from reference structures of small peptides (P26). The two smallest MADs in each subset are written in bold letters.

Measure	PBEh-3c	M06-2X/S[c]	B3LYP/S[c]	TPSS-D3/M	PBE0-D3/M
		ROT34 (rotational constants, deviations in %)[a]			
MD[d]	0.2	0.2	1.8	1.2	−0.2
MAD[e]	**0.4**	0.5	1.8	1.2	**0.2**
SD[f]	0.5	0.6	0.5	0.4	0.3
MAX[g]	1.2	0.9	2.6	2.9	0.8
		LB12 (long bonds, deviations in pm)[b]			
MD	−0.3	−0.4	13.2	0.7	0.3
MAD	**3.9**	7.8	13.2	4.8	**3.1**
SD	6.1	11.6	11.5	8.0	5.1
MAX	14	21	39	18	13
		P26 (small peptides, heavy atom RMSD in pm)			
MAD	**9.7**	**7.6**	43.3	14.7	12.6
SD	5.2	4.0	23.5	12.5	17.7
MAX	30.9	19.5	107	66.1	64.2

[a]Reference 113 with reference error estimate of 0.2%.
[b]LB12 set shown in Fig. 2, reference uncertainty of 2 pm due to possible crystal packing and anharmonic vibrational effects.
[c]def2-SV(P) basis set and the corresponding ECP was used.
[d]Mean deviation.
[e]Mean absolute deviation.
[f]Standard deviation.
[g]Maximum absolute deviation.

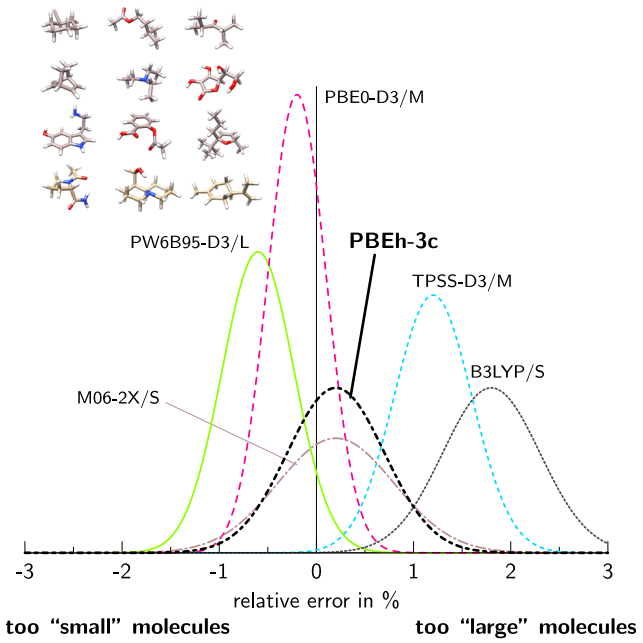

FIG. 4. Normal distribution of the relative errors in the computed rotational constants B_e for the ROT34 benchmark set with various theoretical methods. Large basis set results ("L") are taken from Ref. 113 and compared to small/medium basis set PBEh-3c, M06-2X, and B3LYP methods.

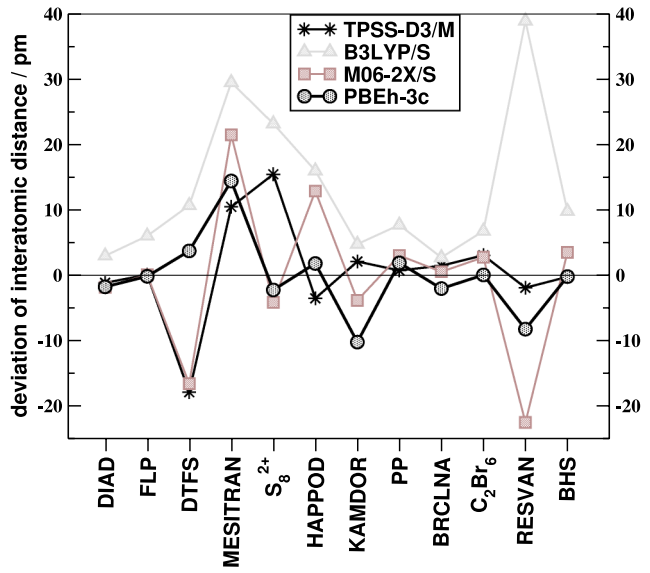

FIG. 5. Deviations of interatomic distances from experimental reference values for the LB12 test set.

of covalent bond lengths as well as non-bonded distances. Furthermore, small changes of internal rotational degrees of freedom result in large deviations (a few percent) of the B_e values from the experimental references so that this test represents a true challenge for any approximate method. As discussed in Ref. 15, dispersion corrections to DFT increase B_e values (shrink molecular size) significantly by about 0.5%-1.5%, thereby in general improving agreement with the reference data. Note that accurate molecular sizes directly transfer to the important property of mass density in a molecular crystal where this can be the key quantity for energetic polymorph ranking (see below). In Figure 4, statistical performance measures for various DFT methods for the 12 molecules with 34 values are given as normal distribution of the relative errors in the computed rotational constants B_e. Here, positive relative deviations correspond to a theoretically overestimated molecular size. Values obtained with large Gaussian basis sets (label "L") from Ref. 113 are given for comparison. PBEh-3c performs best among the small basis set approaches with a tiny mean absolute relative error of 0.4%. This is basically MP2/def2-TZVPP quality[113] but obtained at a very small fraction of the computational cost. M06-2X/S also performs very well with only slightly larger errors indicating fortuitous error compensation between missing long-range dispersion and artificial BSSE effects. B3LYP/S is significantly worse yielding overall too "large" molecules. The TPSS-D3/M functional yields a SD comparable to PBEh-3c but systematically results in too large molecules. This is consistent with the overestimated bond lengths as noted in the previous paragraph for the small molecules. Because it was not included in the original study, the PBE0-D3/M functional is highlighted here. It has a tiny MAD of only 0.2% which is within the uncertainty of the reference values and represents the best global hybrid functional on this test set so far. Only the double hybrid B2PLYP and some range-separated hybrid functionals yield comparably good structures.[15,113]

As a further cross-check, we conducted full geometry optimizations on the newly compiled LB12 set (larger systems with chemically interesting bonding situations, compare Section IV A). The results for the low-cost methods and TPSS-D3/M for comparison are graphically shown as deviations in Figure 5.

Note that many of the considered interactions are rather weak and hence an accurate computation of these distances represents a true challenge for any method. PBEh-3c performs best among the low-cost approaches with a MAD of only 4 pm. The largest deviation of 14 pm occurs for the Si-N distance in the MESITRAN molecule which seems to be problematic for all methods. However, PBEh-3c is still reasonably good and only by 4 pm worse than the best performing PBE0-D3/M method. On average, M06-2X/S is worse by a factor of two compared to PBEh-3c and has a large error spread (SD of roughly 12 pm). The largest errors of M06-2X/S and B3LYP/S occur for DTFS, MESITRAN, and RESVAN (compare Figure 5) for which the PBEh-3c results are significantly better. This indicates that in the PBEh-3c method, all different kinds of short-, medium-, and long-range interactions are treated properly. M06-2X/S and B3LYP/S rely on error compensation of dispersion and BSSE which intrinsically decay differently with interatomic distance. While this can work for some systems, the individual treatment of London dispersion (D3 scheme) and BSSE effects (gCP scheme) seems to be superior. A cross-check with B3LYP/S including D3 dispersion correction (B3LYP-D3/S) revealed that the missing dispersion is a major source of the error in B3LYP/S. Including D3 improves the bad performance on LB12 significantly (MD = 4.6 pm, MAD = 6.7 pm, SD = 9.3 pm, MAX = 26 pm). TPSS-D3/M is slightly worse compared to PBEh-3c while PBE0-D3/M is again the best performing method for this set.

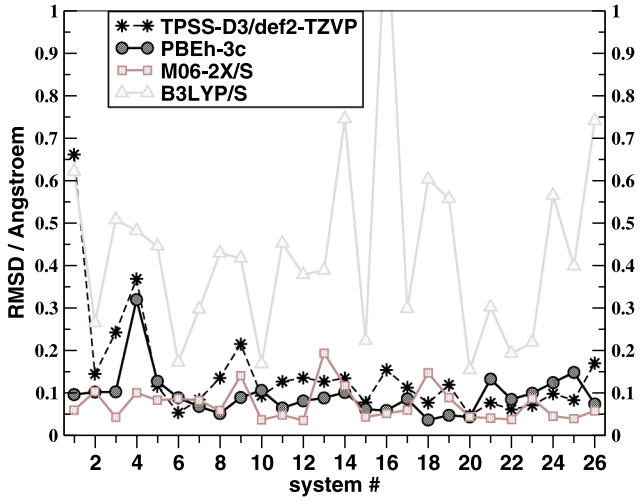

FIG. 6. Root-mean-square-deviations for all heavy atom fits to the MP2/aug-cc-pVTZ reference structures for the P26 peptide benchmark set.

Finally, we consider a set of 26 small peptides with MP2/aug-cc-pVTZ reference geometries.[114] Because MP2 may give a somewhat unbalanced description of intramolecular NCI, these reference structures are certainly not highly accurate. Nevertheless, they should be of sufficient quality for the low-cost methods and the results for these are visualized in Figure 6. PBEh-3c and M06-2X/S perform very well producing essentially no outliers. The average heavy atom root-mean square deviation (RMSD) is only 0.1 Å meaning that overlayed structure plots visually coincide. RMSD values of 0.1-0.2 Å are usually considered to correspond to almost identical structures in biochemical applications so that in conclusion, PBEh-3c performs excellently. Our previous default approach for optimizations (i.e., TPSS-D3/M) and in particular B3LYP/S shows larger deviations from the reference.

3. Non-covalently bound complexes

We now focus on the non-covalent distances in weakly bound dimers. This part of our study is based on the well-known S22 and S66×8 benchmark sets, which were originally designed as interaction energy benchmarks. The statistical evaluation of the tested methods is given in Table VII. The structures of complexes in the S22 set have been optimized by Hobza and co-workers at the triple-ζ MP2 level.[56] These reference data are used here to assess the quality of the DFT methods. The results are shown graphically in Figure 7 where the difference of the center-of-mass inter-fragment distance ΔR_{CMA} with respect to the reference is plotted for all 22 complexes. Note that MP2 overbinds most of the dispersion-dominated complexes 9-15 compared to accurate results from singles and doubles coupled cluster theory with perturbative triples (CCSD(T)).[56] For these cases, one can expect that the reference R_{CMA} values are too small which consequently must be considered when the quality of the DFT methods is evaluated. As can be seen from Figure 7, the R_{CMA} values for PBEh-3c agree best with the MP2 reference, the MAD being only 8 pm with a MD of 3 pm meaning that the PBEh-3c distances are on average slightly too long as

TABLE VII. Deviations of intramolecular center-of-mass distances R_{CMA} to the MP2 reference values for the S22 and to the CCSD(T) references for the S66×8 NCI benchmark set. The two smallest MADs in each subset are written in bold letters.

Measure	PBEh-3c	M06-2X/S	B3LYP/S	TPSS-D3/M	PBE0-D3/M
			S22 (CMA distance in pm)[a]		
MD[b]	3	−8	79	4	1
MAD[c]	8	8	83	**7**	**7**
SD[d]	12	13	115	9	9
MAX[e]	40	40	390	22	23
			S66×8 (CMA distance in %)		
MD	0.8	−2.6	10.5	1.2	0.3
MAD	**1.5**	2.6	11.3	1.6	**1.0**
SD	1.8	1.6	22.2	1.9	1.3
MAX	4.6	7.5	96.2	5.9	3.4

[a]The error of the MP2 reference values is estimated to be about 5 pm from MP2 deviations to the CCSD(T) values for the similar S66×8 set.
[b]Mean deviation.
[c]Mean absolute deviation.
[d]Standard deviation.
[e]Maximum absolute deviation.

expected regarding the systematic error of the MP2 reference. Differences of about 10 pm for R_{CMA} values are considered insignificant due to the weak binding (compared to covalent bonds) and the resulting shallow potentials. The MAD for M06-2X/S is similar but with a negative MD value. B3LYP does not account for London dispersion effects and hence the dispersion dominated and mixed systems often dissociate to some artificial, unphysical BSSE minimum. If the D3 correction is added, quite reasonable structures are obtained (MAD = 7 pm) which, however, have too small intermolecular distances (MD = −5 pm) similar to M06-2X due to the remaining BSSE.

We also considered the RMSD between DFT and reference structure for the entire complex. For the S22 set values of 8 pm (PBEh-3c), 20 pm (M06-2X/S), and 64 pm (B3LYP/S), respectively, were obtained. An analogous treatment of all S66 structures yields for PBEh-3c an average RMSD of only 5 pm meaning that the structures are practically identical to the reference as noted above. Both RMSD and R_{CMA} based

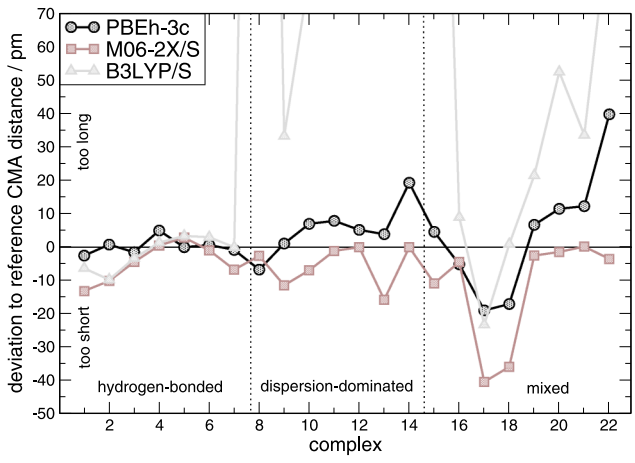

FIG. 7. Deviations of intramolecular center-of-mass distances R_{CMA} to the MP2 reference values for the S22 NCI benchmark set.

analysis show that the PBEh-3c method yields consistently accurate structures for typical vdW and hydrogen-bonded complex geometries of MP2 or better (for aromatic systems) quality and that it outperforms the other low-cost methods. Again, this is attributed to balanced description of dispersion (missing in B3LYP) and BSSE effects (not accounted for in M06-2X/S).

The S66×8 database gives rise to reference potentials beyond MP2 quality (as also noted by Head-Gordon and coworkers recently[119]). For each dimer, geometries at eight different CMA distances with energies at the CCSD(T) level are available. The intramolecular structure is kept fixed at the MP2 equilibrium geometry to exclusively test the intermolecular potential against the CCSD(T) reference. We fit interpolating splines to the energy data and extract the minimum distance. The CCSD(T) reference minimum deviates from the MP2 equilibrium distance and this difference has been used above to judge the quality of MP2 equilibrium geometries. The agreement is reasonable with a MAD of 0.9%. However, we can recognize some systematic MP2 errors. The mainly hydrogen bonded distances (1–23) are by 0.8% systematically too long, while the pure π stacked systems (24–29) are too close by 1.3%. One of the largest deviations is the parallel stacked benzene dimer, where the MP2 CMA distance is by 3.4% too short. This confirms that some of the larger PBEh-3c deviations for S22 originate from bad reference values.

In Figure 8, we show the relative deviation from the CCSD(T) reference intermolecular distance. For the mostly vdW bonded systems, B3LYP/S only shows artificial BSSE minima. Including the D3 correction leads to much improved CMA distances (MD = −1.6%, MAD = 1.6%, SD = 1.3%, MAX = 4.8%). However, due to the large BSSE, the binding energies of the S66 set are drastically overestimated by B3LYP-D3/S (MD = MAD = 2.5 kcal/mol) which corresponds to a relative overestimation by about 50%. Thus, treating both effects (dispersion and BSSE) consistently in a physically sound manner is mandatory. The dispersion vs. BSSE error compensation in B3LYP/S seems to work only for the strongly hydrogen bonded systems, where the results are reasonable. M06-2X/S performs well yielding too short distances by 2.6%. However, out of the small basis set methods, PBEh-3c gives the best agreement with the reference with slightly too large intermolecular distances by 0.8%. This behavior is similar to that of TPSS-D3/M. Only PBE0-D3/M performs significantly better with MD and MAD values of only 0.3% and 1.0%, respectively, reaching or even surpassing a MP2 quality level.

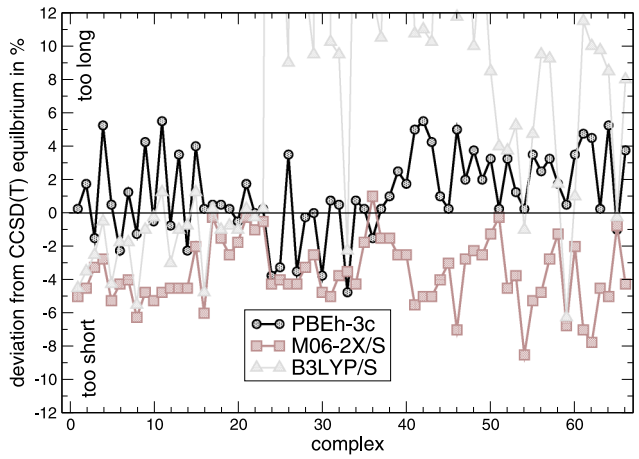

FIG. 8. Deviations of intramolecular distances to the CCSD(T) reference values for the S66×8 benchmark set as obtained from interpolated potential energy curves. A deviation of 1% from the reference corresponds on average to a value of about 2.5 pm for the R_{CMA}.

C. Molecular crystals

The def2-mSVP basis set is rather compact lacking (semi)diffuse functions and hence it can be applied in periodic boundaries without any modification. While larger Gaussian basis sets can be constructed also for the solid state, their numerical complexity (especially with Fock-exchange) is high and near-linear dependencies in the basis can hamper or even destroy SCF convergence.[120,121] Therefore, a small basis set approach is ideally suited for crystalline systems. Other basis set options are available and we typically use projector augmented PAWs to approach the basis set limit. However, the number of PAW basis functions can become very large (e.g., 10^5 functions for a slightly larger organic crystal[122]) and especially hybrid functionals are computationally demanding in this technical setup.

We have tested gCP in combination with the Ahlrichs def2-SVP basis set for organic crystals recently.[63] The lattice energy could be significantly improved, but the cell geometries showed some inconsistencies. Now, we have incorporated the gCP and D3 in the functional design with emphasis on molecular structures. This should lead to a more consistent treatment of energies and geometries in the solid state which is the focus of this section. Especially, their correct long-range limits are of utmost importance. A sufficiently fast, and at the same time, reasonably accurate electronic structure method is mandatory for the fast growing field of organic crystal structure prediction.[123–125]

In 2012, Johnson and co-workers compiled a set of organic crystals from available low-temperature X-ray structures and experimental sublimation enthalpies.[126] Reilly and Tkatchenko refined and extended this set, which will be dubbed X23 in the following.[127] It has been used by various groups to test electronic structure methods.[128–131] Zero-point vibrational energy (ZPVE) and thermodynamic contributions (harmonic and anharmonic) have been explicitly removed from the sublimation enthalpy to yield purely electronic lattice energies.

Johnson used an artificial negative pressure to include volume expansions due to vibrational effects in the optimization procedure. We have done this in an inverse procedure. We estimate the ZPVE and thermal contribution to the unit cell volume and transform the experimental volume V_0 into a back-corrected reference equilibrium volume V_e. These can then be directly compared to optimizations on the electronic energy surface. We perform constraint (constant volume) optimizations for scaled unit cell volumes. Vibrational frequencies are computed in the harmonic approximation. The ZPVE and Bose-Einstein occupied phonon modes finally lead to volume dependent free energies $F(V)$. To the data points

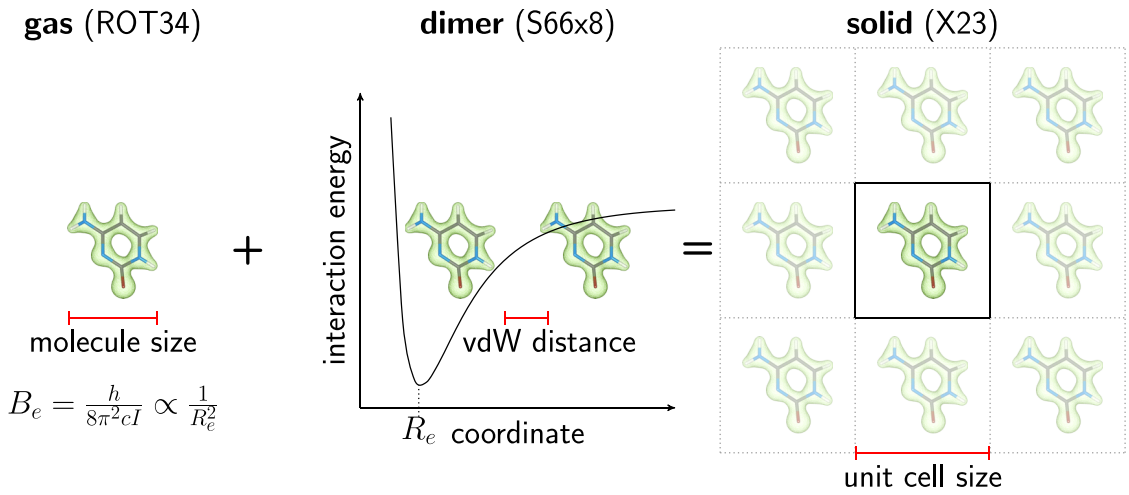

FIG. 9. Important contributions to computed unit cell volumes in an additive scheme.

$E(\{V_i\})$ and $F(\{V_i\})$, a Murnaghan equation of state is adjusted and the equilibrium volume V_e and the free energy volume V_0, respectively, are extracted. The ratio $\Delta V/V_0$ can be used to back-correct the X-ray volumes. This procedure is conducted at the semi-empirical density functional tight-binding DFTB3-D3 level and yields estimated "experimental" reference equilibrium volumes. The reasonably accurate energies and thermal corrections of this method are well documented.[132]

An additional benchmark set of ten ice polymorphs (ICE10) was introduced by us recently.[133] This shall complement the X23 set regarding stronger hydrogen bonds and more emphasis on polymorph ranking. Furthermore, ice (and water) is relevant for various applications in biological processes and for material science at surfaces. The experimental sublimation energies and X-ray geometries have been back-corrected as described above yielding purely electronic lattice energies and ZPVE-exclusive unit cell geometries.

1. Unit cell geometries

The packing of organic molecules in crystal structures is very sensitive to an accurate treatment of NCIs. The size of a crystal unit cell is determined by the molecular size and the intermolecular potential as sketched in Figure 9. The molecular size and the intermolecular potentials both were analyzed in Sec. IV B (ROT34 and S66×8). All geometries are optimized without symmetry constraints at the PBEh-3c, M06-2X/S, and B3LYP/S levels. TPSS-D3 results at the estimated plane wave basis set limit are taken from previous work.[129] They are substantially more expensive compared to the small basis set approaches considered here (about one order of magnitude). The three-body dispersion was shown to be important for the lattice energy with contributions of 4%–7%, while its effect on the structure was minor. To have a consistent description of geometry and energy, the three-body dispersion is always included in the PBEh-3c method. For TPSS-D3 in contrast, the three-body dispersion term is only used for final single point energies. For two systems (anthracene and naphthalene), the SCF of PBEh-3c does not converge within standard numerical thresholds. These crystals have the smallest band gap (some metallic behavior) and hence a large amount of Fock exchange is problematic. For M06-2X/SV(P), seven additional systems without any structural resemblance and even chemically saturated systems showed convergence problems. We attribute this failure to the intrinsic numerical complexity of the Minnesota functional family.

The statistical performance of the different methods is summarized in Table VIII. Explicit energies and volumes are

TABLE VIII. Statistical evaluation of the analyzed methods for geometries and lattice energies of the X23 and ICE10 benchmark sets.[a] The two smallest MADs in each subset are written in bold letters.

	PBEh-3c	M06-2X/S	B3LYP/S	TPSS-D3/L
		X23 unit cell volume in %		
MD[b]	1.8	−12.5	22.1	1.0
MAD[c]	**2.7**	12.5	22.1	**2.8**
SD[d]	3.2	4.4	15.5	4.0
MAX[e]	10.2	23.3	57.3	15.0
		ICE10 unit cell volume in %		
MD	2.5	−15.4	−5.7	−3.8
MAD	**5.0**	15.4	5.7	**3.8**
SD	7.7	2.2	1.4	1.3
MAX	16.6	17.3	8.3	5.5
		X23 lattice energies in kcal/mol		
MD	0.1	5.5	−5.9	−0.6
MAD	**1.3**	5.5	6.5	**1.1**
SD	1.7	4.3	6.7	1.2
MAX	3.7	13.6	25.5	2.7
		ICE10 lattice energies in kcal/mol		
MD	2.6	11.1	9.1	1.0
MAD	**2.6**	11.1	9.1	**1.1**
SD	0.6	1.3	0.6	0.7
MAX	3.7	13.6	10.0	2.3

[a]The error of the lattice energies is estimated to about 1 kcal/mol and for the unit cell volume (mainly from the back-correction scheme) to 1%-1.5%.[133,134]
[b]Mean deviation.
[c]Mean absolute deviation.
[d]Standard deviation.
[e]Maximum absolute deviation. A positive MRD denotes a too large unit cell.

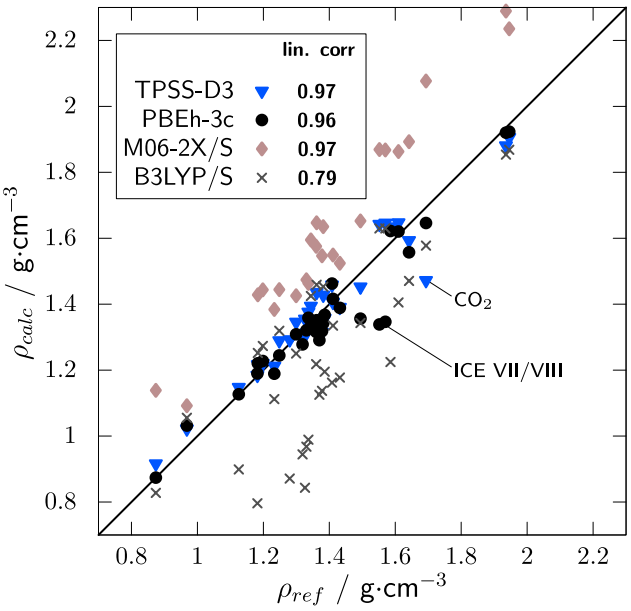

FIG. 10. Correlation between computed and reference mass densities of the benchmark sets X23 and ICE10. One outlier for TPSS-D3 and two for PBEh-3c are highlighted and the overall linear correlation coefficient is given.

given in the supplementary material[79] and are shown in Figures 10 and 11, respectively. As pointed out in the Introduction, good geometries are the main area of application of the new hybrid functional. The unit cell volume is a very sensitive quality criterion for which PBEh-3c performs excellently. The MAD from the X23 reference volumes is 2.7% and even smaller by 0.1% compared to the TPSS-D3/PAW results. Both alternative small basis set approaches do not provide reasonable results. M06-2X/S unit cell volumes are too small by 12.5% and B3LYP/S ones are too large by 22.1%. While some strong hydrogen bonded crystals (oxalic acid)

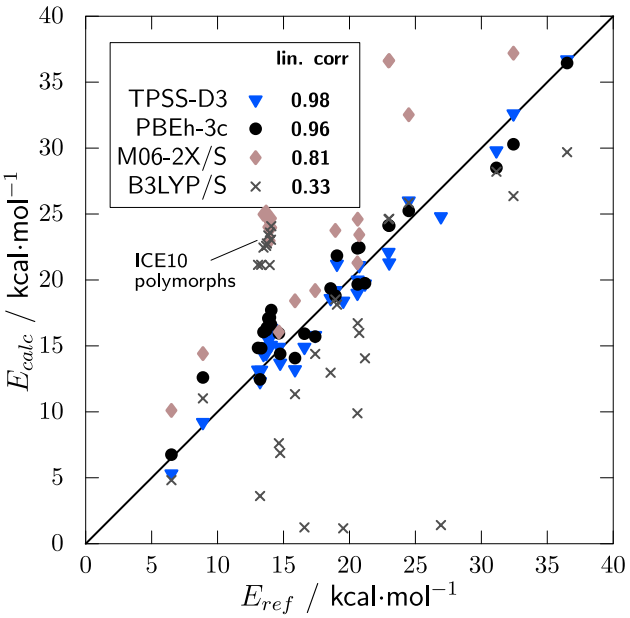

FIG. 11. Correlation between computed and reference lattice energies of the benchmark sets X23 and ICE10. The ICE10 polymorphs are highlighted and the overall linear correlation coefficient is given.

TABLE IX. Systematic deviations of molecular sizes (ROT34, rotational constant), intermolecular distances (S66×8, center of mass distances), and crystal sizes (X23, unit cell volumes).

	ROT34 MD	S66×8 MD	X23 MD (MAD)[a]
PBEh-3c	0.2	0.8	1.8 (2.7)
M06-2X/S	0.2	−2.6	−12.5 (12.5)
B3LYP/S	1.8	10.5	22.1 (22.1)
TPSS/M	1.2	1.2	1.0 (2.8)

[a]Mean deviation (MD) and mean absolute deviation (MAD) are given in %, a positive MD denotes too large structures.

are described reasonably within 2%–3% deviation, others are far off by more than 40% (e.g., adamantane). In B3LYP/S, the error originates from too large molecular sizes and too long intermolecular distances as summarized in Table IX. M06-2X/S yields good molecular structures but too short intermolecular distances which translates into systematically too small unit cell volumes.

Because of the complex networks of strong hydrogen bonds, the ICE10 set is more challenging for the small basis set approaches. The compensation between electrostatic and induction effects is more difficult in very polar systems and the BSSE is substantial in water aggregates. In a comparable gas phase cluster set (WATER27[135]), typical semi-local density functionals are not converged even in the huge def2-QZVP basis set and additional diffuse functions have to be included. This explains the larger error spread in the ICE10 benchmark set. Nevertheless, the PBEh-3c functional yields reasonable geometries with MAD from the reference unit cell volumes of 5.0%. However, especially the high density phases have systematically too large unit cells. If the two high density phases (VII and VIII) are excluded, the MAD drops to 2.2%. TPSS-D3 in a huge PAW basis set yields good structures. Even better results can be obtained by BLYP-D3/PAW, as demonstrated in the original work.[133] M06-2X/S and B3LYP/S perform again not satisfactorily and cannot be recommended for such crystals. The calculated geometries are translated into mass densities in Figure 10. The agreement between TPSS-D3/PAW and PBEh-3c and the reference data is very good with linear correlation coefficients of 0.97 and 0.96, respectively. The systematic errors of M06-2X/S and B3LYP/S with too high and too low computed mass density, respectively, are apparent.

2. Lattice energies

In agreement with all results presented above for molecules, PBEh-3c yields very good interaction energies for the molecular crystals and is considered as energy-structure-consistent. The MAD of the X23 set is small and close to chemical accuracy (1.3 kcal/mol). The performance is equally good for the dispersion and hydrogen bonded systems as shown in Figure 11. This is competitive to the more expansive TPSS-D3/PAW approach which only has a slightly smaller SD value. M06-2X/S and B3LYP/S are far off with MADs of 5.5 and 6.6 kcal/mol, respectively. M06-2X/S systematically overbinds all systems, which is probably due to

TABLE X. Mean absolute (unsigned) deviations (MAD, in kcal/mol) for 28 out of 30 subsets of the GMTKN30 database and S66, S12L, and L7 NCI sets. Values in parentheses are mean (signed) deviations in kcal/mol (negative values indicate underbinding).

Subset	PW6B95-D3/L[a]	PBEh-3c	M06-2X/S[b]	B3LYP/S[b]
		Basis properties		
MB08-165	4.7	17.4	8.36	15.2
W4-08	2.4	10.4	8.07	8.40
G21IP	1.3	5.40	3.77	4.30
PA	2.5	5.06	3.25	3.78
SIE11	7.5	5.70	7.05	9.96
BHPERI	2.3	1.86	1.99	2.97
BH76	3.5	4.99	4.03	7.32
		Reaction energies		
BH76RC	1.6	6.27	6.49	6.99
RSE43	2.4	2.69	1.26	2.61
O3ADD6	4.3	7.45	8.99	4.14
G2RC	3.4	10.4	11.4	12.8
AL2X	1.3	2.14	4.35	6.57
NBPRC	1.8	5.24	4.68	3.43
ISO34	1.2	1.97	2.80	3.15
ISOL22	4.6	3.09	4.04	7.60
DC9	7.0	12.5	8.64	12.8
DARC	3.6	9.17	4.89	7.51
ALK6	4.7	1.40	1.97	7.71
BSR36	4.1	1.34	6.52	14.8
		Non-covalent interactions		
IDISP	3.5	2.92	1.52	15.4
S22	0.34	0.40	1.58	2.82
ADIM6	0.58	0.09	0.54	4.21
RG6	0.03	0.20	0.25	0.68
HEAVY28	0.13	0.57	0.76	1.20
PCONF	0.51	0.79	3.08	1.70
ACONF	0.15	0.26	0.36	0.76
SCONF	0.31	0.41	2.57	2.10
CYCONF	0.31	1.45	0.81	0.67
S66	0.18	0.46 (0.22)	1.53 (1.43)	2.59 (−1.09)
S30L[c]	2.5	3.4 (0.1)	7.4 (7.4)	25.9 (−25.4)
L7[d]	1.3	1.6 (−0.5)	3.1 (2.5)	16.2 (−15.9)

[a]Extended quadruple-zeta basis set results from Ref. 5.
[b]def2-SV(P) basis set, mSVP for S30L and L7.
[c]Back-corrected experimental reference data, see Ref. 138 for details.
[d]Unpublished DLPNO-CCSD(T)/CBS* reference data[139] because of inconsistencies of the original values in Ref. 58.

the unbalanced treatment of BSSE and dispersion. In contrast, B3LYP underbinds most systems, especially the non-hydrogen bonded ones. Interestingly, some systems are even bound too strongly. This very clearly demonstrates that artificial BSSE cannot be used generally to compensate missing dispersion effects. Again, the ICE10 set is more challenging. The MAD of PBEh-3c is 2.6 kcal/mol with systematically too large lattice energies. Though the errors are larger compared to the X23 set, the results are still very reasonable when the computational speed-up is considered. The non-local density functional optPBE-vdW and RPA correlation on converged PBE orbitals (RPA@PBE) are only slightly more accurate compared to PBEh-3c (MADs of 1.4 kcal/mol and 1.6 kcal/mol) with a substantial increase in computational costs.[136,137] M06-2X/S and B3LYP/S show an even stronger overbinding tendency which clearly demonstrates the large BSSE. With increasing crystal density, the errors also increase which indicates its BSSE origin (the linear correlation between the M06-2X/S error and the corresponding mass density is 0.6).

The results presented in this section show that the PBEh-3c functional can be successfully applied to organic solids yielding binding energies and geometries of a quality that is competitive to the computationally more demanding TPSS-D3/L method. Especially for systems where GGAs cannot be applied, PBEh-3c is a promising alternative. We like to point out that the asymptotically correct treatment of London dispersion on the one hand and correcting for BSSE on the other is extremely important for periodic systems. While the other low-cost methods M06-2X/S and B3LYP/S perform satisfactorily in some cases for structures of isolated molecules (due to fortuitous error cancellation, see Section IV B), this is not the case in many crystals.

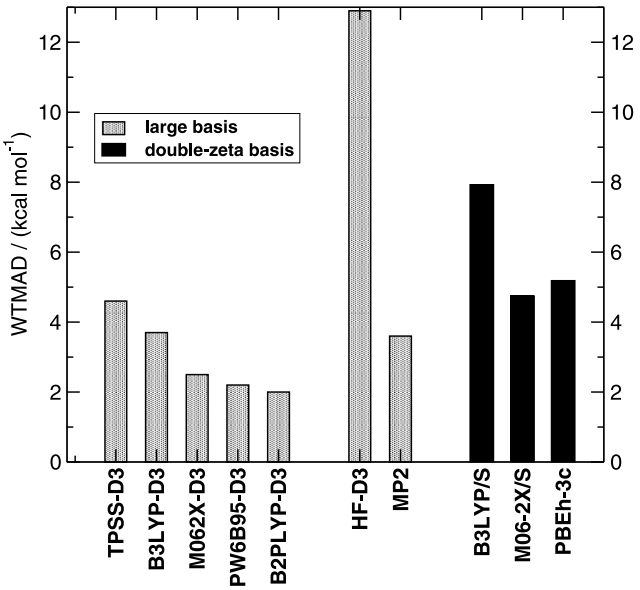

FIG. 12. Weighted MAD values for GMTKN30 with some large basis set DFT-D3 methods, HF-D3, and MP2 in comparison to the small/medium basis set methods considered here (excluding the G21EA and WATER27 subsets for the small basis set methods in the right part of the plot).

D. General interaction and reaction energies

Though the main purpose of PBEh-3c is the computation of accurate structures, a correct treatment of standard thermochemistry, barrier heights, and especially NCIs is important as well. We evaluate the large GMTKN30 database[5] and additionally test the NCIs in larger systems (S66,[57] S30L,[138] L7[58]). We give the PW6B95-D3/L results for comparison since it is the best global hybrid functional for this data base. The results of 28 out of 30 subsets are given in Table X and the weighted MAD is shown in Figure 12. The WATER27 and G21EA benchmarks cannot be treated properly in the small basis set and hence have been omitted from the analysis.

The performance of PBEh-3c is overall very reasonable and comparable to M06-2X/S. No true outlier can be detected, and only some reactions (atomization energies and those in MB08-165) of small molecules are described with only medium accuracy. It performs very well (as expected for a high-Fock-exchange functional) for barriers and SIE related problems (BHPERI, BH76, SIE11, ALK6) as well as for chemically relevant thermochemistry (ISO34, ISOL, BSR36). The lower accuracy for a few other reaction subsets (G2RC, BH76RC, DC9) is in these cases similar to that of M06-2X and B3LYP so that it seems to be caused by the small AO basis. Notable is the very good performance for NCIs. The MADs of the S22 and ADIM subsets are tiny with 0.4 kcal/mol and 0.1 kcal/mol, respectively. Conformational problems (the CONF subsets) are also described well. Because the main targets of all low-cost methods are large systems, we investigate the performance for them in detail. The L7 set was compiled by Hobza and co-workers.[58] We use new DLPNO-CCSD(T)/CBS*[139] reference values which are more consistent than the original data. The S30L test set consists of 30 "real" supramolecular host-guest complexes.[138] The reference energies of the S30L complexes vary between 20 kcal/mol and 150 kcal/mol. Because multiple hydrogen bonds are present and nine systems are charged, the S30L benchmark is challenging for any electronic structure method. Interestingly, PW6B95-D3/L is the best performing hybrid functional on this test set, which is consistent with the GMTKN30 evaluation.

The PBEh-3c results on these large complexes are very promising. The MAD for L7 and S30L are as low as 1.6 kcal/mol and 3.4 kcal/mol, respectively, and hence only slightly worse compared to PW6B95-D3/L. While M06-2X/S performs satisfactorily for L7, it is far off for the S30L complexes. B3LYP/S has huge errors (MAD of 26 kcal/mol and 16 kcal/mol) and cannot be recommended without D3/gCP corrections. This indicates, in agreement with the conclusions from the organic crystals, that an asymptotically correct and consistent treatment of all interactions is mandatory to correctly describe extended systems.

V. CONCLUSIONS

Large molecules or supramolecular aggregates with 500-1000 atoms are often difficult to characterize structurally and first-principles quantum chemical methods can be of great help in many experimental studies. A new simplified density functional scheme is proposed here for this purpose. The present work adheres to our general philosophy to simplify the methods as far as possible but keeping the correct physics without introducing too much empiricism. In this spirit, we coupled an existing hybrid density functional in modified form to well-established atom pair-wise corrections for long-range (London) dispersion effects and the more short-ranged BSSE. The method is dubbed PBEh-3c to indicate the origin and basically the three introduced modifications (corrections) to a standard method using a common basis set (modified def2-SV(P), not included in the abbreviation). The approach continues previous successful work using minimal basis set Hartree-Fock (termed HF-3c). By a systematic cancellation of errors between the density functional and the applied small Gaussian orbital basis set, accurate bond lengths and overall molecular structures are obtained. Most striking is the roughly "MP2-quality" (or slightly better) obtained for the non-covalent complexes in the S22/S66 sets and equilibrium structures (B_e values) for medium-sized organic molecules in the ROT34 set. Because we insisted on a large portion of non-local Fock-exchange (42%), self-interaction related errors are alleviated and artificial charge transfer in large (charged) systems can be avoided. As shown in a few examples, the Hartree-Fock character of the method is still small enough to provide reasonable structures of transition metal complexes (e.g., the difficult carbon-iron bond distance in ferrocene is in error by less than 1 pm).

We demonstrated that the method can be applied to organic and ice crystals without any modification and that the results are close to those of TPSS-D3 evaluated in a huge PAW basis set. The accuracy of PBEh-3c for both the structures and the lattice energy of the medium polar systems in the X23 set is close to the reference accuracy, i.e., the MAD to the unit cell volume reference and to the lattice energy reference is 2.7% and 1.3 kcal/mol, respectively. Slightly lower but still very reasonable accuracy is obtained for ten ice polymorphs with

varying density. Thus, PBEh-3c may also be viable method for simulations of water under liquid conditions or at interfaces.

Although primarily designed for computing structures (and possibly vibrational frequencies which are not investigated here), the method performs also reasonably well for thermochemistry, barrier heights, and general NCIs. This has been extensively tested for the GMTKN30 database and additionally for the L7 and S30L large molecule NCI benchmarks. Clearly, due to the use of the small basis set in PBEh-3c, the high accuracy obtained with, e.g., dispersion-corrected (double)hybrid functionals and quadruple-zeta basis sets cannot be reached. However, the energetic description is of similar quality (or slightly better) to that of the more empirical and highly parametrized M06-2X/"double-zeta" approach. It is expected to be good enough to provide undistorted potential energy surfaces also far away from equilibrium situations in order to allow reactive bio-molecular simulations. In fact, the largest errors in the GMTKN30 sets occur for small molecule reactions which are not in the main focus of the method. More important, however, is the very good description of NCIs which can compete with the best available DFT methods. Notably, NCIs with PBEh-3c are generally better than at the MP2/"large basis set" level. Compared to minimal basis set methods (e.g., HF-3c), particularly the description of charged systems is improved.

As a by-product of our detailed study of molecular structures, we found that the standard PBE0-D3/def2-TZVP method performs very well and robustly in general. All calculated structures are exceptionally close to the corresponding reference values. While Furche and co-workers promote RPA@PBE geometries as alternative to MP2 based structures,[140] we have shown that PBE0-D3/def2-TZVP can reliably be used in general and in particular also for transition metals. It yields molecular structures (e.g., ROT34) significantly better than MP2 and non-covalent intermolecular distances (e.g., for S66×8) with similar accuracy. The main advantage of PBEh-3c over the strongly related PBE0-D3 standard approach is three-fold: (a) alleviated SIE by increased Fock exchange admixture (42% vs. 25%), (b) large computational savings (factor of 10-30) if PBE0-D3 is appropriately applied with a large basis set of at least triple-zeta quality, and (c) consistent treatment of dispersion and BSSE which is not the case if standard PBE0-D3 is applied with the same DZ basis as used in PBEh-3c.

Because the proposed PBEh-3c method does not require any further corrections for basis set superposition error and yields consistent energies and structures (atomic forces), it can be applied straightforwardly for various purposes (optimization, screening, and molecular dynamics) and is proposed as general tool in quantum chemistry. It covers all basic physical effects in molecules, their aggregates, and crystals and may fail badly only in strongly correlated systems, if the self-interaction (over-delocalization) error is extreme, or the basis set demands are unusually high. Many of our conclusions are likely transferable to the liquid state which has not been tested so far. PBEh-3c is suggested as a replacement of the still widely used B3LYP/"double-zeta" (e.g., 6-31G*) model chemistry which is by far outperformed in almost all tested cases. B3LYP/6-31G* does not contain all necessary physics to describe large systems correctly as discussed in detail recently.[50] Regarding computational speed, PBEh-3c is about one order of magnitude faster than a common "hybrid functional/triple-zeta" combination and about two orders of magnitude faster than MP2/"near complete basis set" treatments. It can be routinely applied using common desktop computers for optimizations of systems with several hundreds of atoms. It is also a viable method for the computation of molecular crystals and their polymorphs as shown convincingly for the X23 and ICE10 test sets. The polymorph ranking problem in the context of crystal structure prediction will be studied in more detail in our laboratory. The PBEh-3c method will replace TPSS-D3/def2-TZVP from now on as the default structure optimization level in our group.

ACKNOWLEDGMENTS

This work was supported by the DFG in the framework of the "Gottfried-Wilhelm-Leibniz" prize. The authors thank Stephan Ehrlich for cross-checking protein ligand interactions.

[1]R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, Oxford, 1989).
[2]W. Kohn, Rev. Mod. Phys. **71**, 1253 (1998).
[3]Y. Zhao and D. G. Truhlar, Acc. Chem. Res. **41**, 157 (2008).
[4]R. Paverati and D. G. Truhlar, Philos. Trans. R. Soc., A **372**, 20120476 (2014).
[5]L. Goerigk and S. Grimme, J. Chem. Theory Comput. **7**, 291 (2011).
[6]N. Mardirossian, J. A. Parkhill, and M. Head-Gordon, Phys. Chem. Chem. Phys. **13**, 19325 (2011).
[7]W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory* (Wiley-VCH, New York, 2001).
[8]C. Riplinger, B. Sandhoefer, A. Hansen, and F. Neese, J. Chem. Phys. **139**, 134101 (2013).
[9]S. A. Maurer, D. S. Lambrecht, J. Kussmann, and C. Ochsenfeld, J. Chem. Phys. **138**, 014101 (2013).
[10]M. Schütz, O. Masur, and D. Usvyat, J. Chem. Phys. **140**, 244107 (2014).
[11]L. Goerigk and J. R. Reimers, J. Chem. Theory Comput. **9**, 3240 (2013).
[12]J. G. Brandenburg, M. Hochheim, T. Bredow, and S. Grimme, J. Phys. Chem. Lett. **5**, 4275 (2014).
[13]C. Puzzarini, M. Heckert, and J. Gauss, J. Chem. Phys. **128**, 194108 (2008).
[14]M. Heckert, M. Kallay, D. P. Tew, W. Klopper, and J. Gauss, J. Chem. Phys. **125**, 044108 (2006).
[15]S. Grimme and M. Steinmetz, Phys. Chem. Chem. Phys. **15**, 16031 (2013).
[16]S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).
[17]L. A. Burns, A. Vazquez-Mayagoitia, B. G. Sumpter, and C. D. Sherrill, J. Chem. Phys. **134**, 084107 (2011).
[18]J. Klimes and A. Michaelides, J. Chem. Phys. **137**, 120901 (2012).
[19]S. Grimme, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **1**, 211 (2011).
[20]S. Grimme, Angew. Chem., Int. Ed. **45**, 4460 (2006).
[21]N. C. Handy, C. W. Murray, and R. D. Amos, J. Phys. Chem. **97**, 4392 (1993).
[22]B. G. Johnson, P. M. W. Gill, and J. A. Pople, J. Chem. Phys. **98**, 5612 (1993).
[23]A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).
[24]F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, and J. H. van Lenthe, Chem. Rev. **94**, 1873 (1994).
[25]S. F. Boys and F. Bernardi, Mol. Phys. **19**, 553 (1970).
[26]A. Hamza, A. Vibok, G. J. Halasz, and I. Mayer, J. Mol. Struct.: THEOCHEM **501**, 427 (2000).
[27]T. Helgaker, J. Gauss, P. Jørgensen, and J. Olsen, J. Chem. Phys. **106**, 6430 (1997).
[28]R. Sure and S. Grimme, J. Comput. Chem. **34**, 1672 (2013).
[29]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996); Erratum, **78**, 1396 (1997).
[30]C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999).

[31]H. Kruse and S. Grimme, J. Chem. Phys. **136**, 154101 (2012).
[32]S. Grimme, S. Ehrlich, and L. Goerigk, J. Comput. Chem. **32**, 1456 (2011).
[33]J. A. Pople, in *Modern Theoretical Chemistry* (Plenum, New York, 1976), Vol. 4.
[34]E. R. Davidson and D. Feller, Chem. Rev. **86**, 681 (1986).
[35]W. Kołos, Theor. Chim. Acta **51**, 219 (1979).
[36]E. N. Brothers and G. E. Scuseria, J. Chem. Theory Comput. **2**, 1045 (2006).
[37]R. D. Adamson, P. M. W. Gill, and J. A. Pople, Chem. Phys. Lett. **284**, 6 (1998).
[38]Y. Zhao, B. J. Lynch, and D. G. Truhlar, J. Phys. Chem. A **108**, 4786 (2004).
[39]I.-C. Lin, M. D. Coutinho-Neto, C. Felsenheimer, O. A. von Lilienfeld, I. Tavernelli, and U. Rothlisberger, Phys. Rev. B **75**, 205131 (2007).
[40]G. A. DiLabio and M. Koleini, J. Chem. Phys. **140**, 18A542 (2014).
[41]M. Goldey and M. Head-Gordon, J. Phys. Chem. Lett. **3**, 3592 (2012).
[42]B. Brauer, M. K. Kesharwani, and J. M. L. Martin, J. Chem. Theory Comput. **10**, 3791 (2014).
[43]P. Mori-Sanchez, A. J. Cohen, and W. Yang, J. Chem. Phys. **125**, 201102 (2006).
[44]F. Weigend, F. Furche, and R. Ahlrichs, J. Chem. Phys. **119**, 12753 (2003).
[45]F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).
[46]N. Mardirossian and M. Head-Gordon, J. Chem. Theory Comput. **9**, 4453 (2013).
[47]N. Mardirossian and M. Head-Gordon, Phys. Chem. Chem. Phys. **16**, 9904 (2014).
[48]A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[49]P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, J. Phys. Chem. **98**, 11623 (1994).
[50]H. Kruse, L. Goerigk, and S. Grimme, J. Org. Chem. **77**, 10824 (2012).
[51]L. Goerigk, H. Kruse, and S. Grimme, ChemPhysChem **12**, 3421 (2011).
[52]A. Hansen, C. Bannwarth, S. Grimme, P. Petrović, C. Werlé, and J.-P. Djukic, ChemistryOpen **3**, 177 (2014).
[53]J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, Phys. Rev. Lett. **91**, 146401 (2003).
[54]E. R. Johnson and A. D. Becke, J. Chem. Phys. **123**, 024101 (2005).
[55]E. R. Johnson and A. D. Becke, J. Chem. Phys. **124**, 174104 (2006).
[56]P. Jurecka, J. Sponer, J. Cerny, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).
[57]J. Řezáč, K. E. Riley, and P. Hobza, J. Chem. Theory Comput. **7**, 2427 (2011).
[58]R. Sedlak, T. Janowski, M. Pitoňák, J. Řezáč, P. Pulay, and P. Hobza, J. Chem. Theory Comput. **9**, 3364 (2013).
[59]T. Risthaus and S. Grimme, J. Chem. Theory Comput. **9**, 1580 (2013).
[60]R. A. DiStasio, V. V. Gobre, and A. Tkatchenko, J. Phys.: Condens. Matter **26**, 213202 (2014).
[61]J. F. Dobson, Int. J. Quantum Chem. **114**, 1157 (2014).
[62]M. R. Kennedy, A. R. McDonald, A. E. DePrince, M. S. Marshall, R. Podeszwa, and C. D. Sherrill, J. Chem. Phys. **140**, 121104 (2014).
[63]J. G. Brandenburg, M. Alessio, B. Civalleri, M. F. Peintinger, T. Bredow, and S. Grimme, J. Phys. Chem. A **117**, 9282 (2013).
[64]J.-D. Chai and M. Head-Gordon, Phys. Chem. Chem. Phys. **10**, 6615 (2000).
[65]A. D. Becke, J. Chem. Phys. **107**, 8554 (1997).
[66]A. D. Becke, J. Chem. Phys. **104**, 1040 (1996).
[67]C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[68]H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao, J. Chem. Phys. **115**, 3540 (2001).
[69]Y. Zhang and W. Yang, Phys. Rev. Lett. **80**, 890 (1998).
[70]J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke, Phys. Rev. Lett. **100**, 136406 (2008).
[71]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **80**, 891 (1998).
[72]G. I. Csonka, A. Ruzsinszky, J. P. Perdew, and S. Grimme, J. Chem. Theory Comput. **4**, 888 (2008).
[73]G. I. Csonka, J. P. Perdew, and A. Ruzsinszky, J. Chem. Theory Comput. **6**, 3688 (2010).
[74]S. Grimme, W. Hujo, and B. Kirchner, Phys. Chem. Chem. Phys. **14**, 4875 (2012).
[75]A. D. Boese and J. M. L. Martin, J. Chem. Phys. **121**, 3405 (2004).
[76]A. Schäfer, H. Horn, and R. Ahlrichs, J. Chem. Phys. **97**, 2571 (1992).
[77]K. A. Peterson, D. Figgen, E. Goll, H. Stoll, and M. Dolg, J. Chem. Phys. **119**, 11113 (2003).
[78]W. J. Hehre, R. Ditchfield, and J. A. Pople, J. Chem. Phys. **56**, 2257 (1972).
[79]See supplementary material at http://dx.doi.org/10.1063/1.4927476 for def2-mSVP basis set definition, explicit bond distances of the small molecules and the LB12 set, X23 and ICE10 lattice energies and unit cell volumes, and geometries of all structures at PBEh-3c level.
[80]See http://www.thch.uni-bonn.de/ for freely available codes of the D3 and gCP programs.
[81]R. Ahlrichs, M. Bär, M. Häser, H. Horn, and C. Kölmel, Chem. Phys. Lett. **162**, 165 (1989).
[82]TURBOMOLE V6.6 2014, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.
[83]F. Neese, ORCA - an *ab initio*, density functional and semiempirical program package, version 3.0 (current development version), Max Planck Institute for Chemical Energy Conversion, Germany, 2015.
[84]F. Neese, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 73 (2012).
[85]D. Rappoport and F. Furche, J. Chem. Phys. **133**, 134105 (2010).
[86]E. J. Baerends, D. E. Ellis, and P. Ros, Chem. Phys. **2**, 41 (1973).
[87]B. I. Dunlap, W. D. Connolly, and J. R. Sabin, J. Chem. Phys. **71**, 3396 (1979).
[88]K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, Chem. Phys. Lett. **240**, 283 (1995).
[89]F. Weigend, Phys. Chem. Chem. Phys. **8**, 1057 (2006).
[90]K. Eichkorn, F. Weigend, O. Treutler, and R. Ahlrichs, Theor. Chem. Acc. **97**, 119 (1997).
[91]R. Dovesi, R. Orlando, A. Erba, C. M. Zicovich-Wilson, B. Civalleri, S. Casassa, L. Maschio, M. Ferrabone, M. De La Pierre, P. D'Arco, Y. Noël, M. Causà, M. Rérat, and B. Kirtman, Int. J. Quantum Chem. **114**, 1287 (2014).
[92]P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[93]G. Kresse and J. Joubert, Phys. Rev. B **59**, 1758 (1999).
[94]G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993).
[95]G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996).
[96]*Entropy and Heat Capacity of Organic Compounds*, NIST Chemistry WebBook, NIST Standard Reference Database Number 69, edited by P. Linstrom and W. Mallard (National Institute of Standards and Technology, Gaithersburg, MD, 2015).
[97]P. Pyykkö and M. Atsumi, Chem. - Eur. J. **15**, 186 (2009).
[98]W. Hujo and S. Grimme, J. Chem. Theory Comput. **9**, 308 (2013).
[99]T. Schwabe, S. Grimme, and J. P. Djukic, J. Am. Chem. Soc. **131**, 14156 (2009).
[100]J. Moellmann and S. Grimme, Phys. Chem. Chem. Phys. **12**, 8500 (2010).
[101]J. Moellmann and S. Grimme, Organometallics **32**, 3784 (2013).
[102]A. A. Fokin, L. V. Chernish, P. A. Gunchenko, E. Y. Tikhonchuk, H. Hausmann, M. Serafin, J. E. P. Dahl, R. M. K. Carlson, and P. R. Schreiner, J. Am. Chem. Soc. **134**, 13641 (2012).
[103]O. Ekkert, G. Kehr, R. Fröhlich, and G. Erker, J. Am. Chem. Soc. **133**, 4610 (2011).
[104]N. Mitzel, U. Losehand, A. Wu, D. Cremer, and D. Rankin, J. Am. Chem. Soc. **122**, 4471 (2000).
[105]Q. Shen and R. L. Hilderbrandt, J. Mol. Struct. **64**, 257 (1980).
[106]T. Cameron, R. Deeth, I. Dionne, H. Du, H. Jenkins, I. Krossing, J. Passmore, and H. Roobottom, Inorg. Chem. **39**, 5614 (2000).
[107]S. Grimme and J. P. Djukic, Inorg. Chem. **49**, 2911 (2010).
[108]R. A. Pascal, C. G. Winans, and D. Van Engen, J. Am. Chem. Soc. **111**, 3007 (1989).
[109]M. Davydova and Y. Struchkov, J. Struct. Chem. **6**, 98 (1965).
[110]G. Mandel and J. Donohue, Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. **28**, 1313 (1972).
[111]J. Zong, J. T. Mague, and R. A. Pascal, J. Am. Chem. Soc. **135**, 13235 (2013).
[112]D. W. Stephan and G. Erker, Angew. Chem., Int. Ed. **49**, 46 (2010).
[113]T. Risthaus, M. Steinmetz, and S. Grimme, J. Comput. Chem. **35**, 1509 (2014).
[114]H. Valdes, K. Pluháčková, M. Pitonák, J. Řezáč, and P. Hobza, Phys. Chem. Chem. Phys. **10**, 2747 (2008).
[115]J. M. Martin, J. El-Yazal, and J.-P. François, Mol. Phys. **86**, 1437 (1995).
[116]K. P. Huber and G. Herzberg, in *Constants of Diatomic Molecules*, Molecular Spectra and Molecular Structure Vol. 4 (Van Nostrand, Princeton, 1979).
[117]M. Bühl and H. Kabrede, J. Chem. Theory Comput. **2**, 1282 (2006).
[118]E. A. Coutsias, C. Seok, and K. A. Dillu, J. Comput. Chem. **25**, 1849 (2004).
[119]J. Witte, M. Goldey, J. B. Neaton, and M. Head-Gordon, J. Chem. Theory Comput. **11**, 1481 (2015).
[120]B. Civalleri, C. M. Zicovich-Wilson, L. Valenzano, and P. Ugliengo, CrystEngComm **10**, 405 (2008).
[121]M. Ferrero, B. Civalleri, M. Rerat, R. Orlando, and R. Dovesi, J. Chem. Phys. **132**, 214704 (2009).

[122]J. G. Brandenburg, S. Grimme, P. G. Jones, G. Markopoulos, H. Hopf, M. K. Cyranski, and D. Kuck, Chem. - Eur. J. **19**, 9930 (2013).
[123]S. L. Price, Chem. Soc. Rev. **43**, 2098 (2014).
[124]M. A. Neumann, F. J. J. Leusen, and J. Kendrick, Angew. Chem., Int. Ed. **47**, 2427 (2008).
[125]C. C. Pantelides, C. S. Adjiman, and A. V. Kazantsev, Top. Curr. Chem. **345**, 25 (2014).
[126]A. Otero-de-la-Roza and E. R. Johnson, J. Chem. Phys. **137**, 054103 (2012).
[127]A. M. Reilly and A. Tkatchenko, J. Chem. Phys. **139**, 024705 (2013).
[128]J. G. Brandenburg and S. Grimme, Top. Curr. Chem. **345**, 1 (2014).
[129]J. Moellmann and S. Grimme, J. Phys. Chem. C **118**, 7615 (2014).
[130]D. J. Carter and A. L. Rohl, J. Chem. Theory Comput. **10**, 3423 (2014).
[131]A. Ambrosetti, A. M. Reilly, R. A. DiStasio, and A. Tkatchenko, J. Chem. Phys. **140**, 18A508 (2014).
[132]J. G. Brandenburg and S. Grimme, J. Phys. Chem. Lett. **5**, 1785 (2014).
[133]J. G. Brandenburg, T. Maas, and S. Grimme, J. Chem. Phys. **142**, 124104 (2015).
[134]J. S. Chickos, Netsu Sokutei **30**, 116 (2003).
[135]V. S. Bryantsev, M. S. Diallo, A. C. T. van Duin, and W. A. Goddard, J. Chem. Theory Comput. **5**, 1016 (2009).
[136]B. Santra, J. Klimeš, A. Tkatchenko, D. Alfè, B. Slater, A. Michaelides, R. Car, and M. Scheffler, J. Chem. Phys. **139**, 154702 (2013).
[137]M. Macher, J. Klimeš, C. Franchini, and G. Kresse, J. Chem. Phys. **140**, 084502 (2014).
[138]R. Sure and S. Grimme, "Comprehensive benchmark of association (free) energies of realistic host–guest complexes," J. Chem. Theory Comput. (published online 2015).
[139]DLPNO-CCSD(T) with tight cutoff values and a newly developed approximate CBS extrapolation scheme denoted CBS* (unpublished).
[140]A. M. Burow, J. E. Bates, F. Furche, and H. Eshuis, J. Chem. Theory Comput. **10**, 180 (2014).