ChemPubSoc Europe

DOI: 10.1002/open.201500192

Open Access ChemistryOPEN Reviews

Small Atomic Orbital Basis Set First-Principles Quantum Chemical Methods for Large Molecular and Periodic Systems: A Critical Analysis of Error Sources

Rebecca Sure, Jan Gerit Brandenburg, and Stefan Grimme*[a]

In quantum chemical computations the combination of Hartree–Fock or a density functional theory (DFT) approximation with relatively small atomic orbital basis sets of double-zeta quality is still widely used, for example, in the popular B3LYP/6-31G* approach. In this Review, we critically analyze the two main sources of error in such computations, that is, the basis set superposition error on the one hand and the missing London dispersion interactions on the other. We review various strategies to correct those errors and present exemplary calculations on mainly noncovalently bound systems of widely varying size. Energies and geometries of small dimers, large supramolecular complexes, and molecular crystals are covered. We conclude that it is not justified to rely on fortunate error compensation, as the main inconsistencies can be cured by modern correction schemes which clearly outperform the plain mean-field methods.

1 Introduction

Kohn–Sham density functional theory (KS-DFT, or simply DFT in the following)[1,2] has evolved to be today's most widely used electronic structure method and has emerged as the theory of choice for application to various problems in the chemical and physical sciences. Due to its good cost-accuracy ratio this especially holds for large molecular systems and solids. The number of collaborative experimental and theoretical studies has grown tremendously in the last decade. The usage of complementary theoretical and experimental information can generate valuable new insights, and it is nowadays possible to explain and describe various phenomena in a detailed mechanistic way based on routine quantum chemical calculations.

DFT is considered as the natural theory for extended systems but its current, partially semiempirical character requires extensive benchmarking on theoretical or experimental reference values. Over the past years, such benchmark studies have been carried out with diligence, mainly focusing on energetic properties[3,4,5] and more recently also regarding structures of small to medium-sized molecules.[4,6,7,8] However, the number of proposed density functionals is already too huge to be covered comprehensively. Consequently, the task to select an appropriate and efficient level of theory for a specific problem is highly nontrivial. Thus, it comes as no surprise that nonexperts often choose methods purely because of their popularity, and those are not necessarily the best options for their application. This eventually results in a waste of computational as well as human resources.

One prominent example is the combination of the B3LYP functional[9,10,11,12] with the 6-31G* double-ζ one-particle atomic orbital (AO) basis set[13] in particular, or similar functionals with a small double-ζ (DZ) basis set in general. Over the last decade, computational power has increased immensely, and using a well-converged basis set (BS) is feasible in many cases. A search with SciFinder[14] for the exemplary B3LYP functional reveals that the ratio of journal articles using it in combination with a DZ basis, and those using it with a triple-ζ (TZ) basis, is roughly the same compared with the previous decade. In the years 1995 to 2005, this ratio was about 3.5:1, and it only dropped slightly to about 3:1 during the last decade (Figure 1).

In 2005 Ahlrichs et al. published the efficient def2-SVP (DZ) and def2-TZVP (TZ) BSs which were specifically designed for SCF calculations. However, the 6-31G* (DZ type) and 6-311G*[15] (TZ type) published by Pople et al. in 1972 and 1980, respectively, are still widely used in DFT calculations.

Compared with the citations of DZ and TZ basis sets, the number of articles employing B3LYP in combination with quadruple-ζ (QZ) type expansions is tiny. Utilizing QZ basis sets in HF or DFT calculations leads to results which are chemically very close to the complete basis set (CBS) limit, and this is our general recommendation if this level is affordable. Because of the faster BS convergence compared to correlated post-Hartree–Fock methods, normally no further BS extrapolation scheme is needed. However, these calculations are routinely possible on standard workstations only for medium-sized systems with about 100 atoms or less.

If the system size increases, or one has to perform very many calculations, and one is rather limited in the computational resources, as most mainly experimentally working

[a] R. Sure, Dr. J. G. Brandenburg, Prof. S. Grimme
Mulliken Center for Theoretical Chemistry
Institut für Physikalische und Theoretische Chemie, Universität Bonn
Beringstr. 4, 53115 Bonn (Germany)
E-mail: grimme@thch.uni-bonn.de

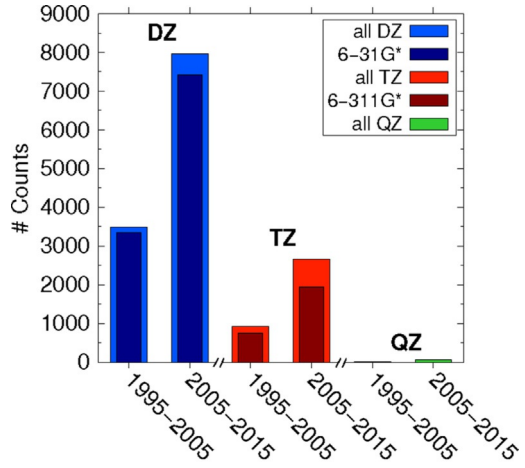
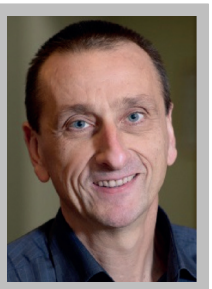


Figure 1. SciFinder[14] hits (dated May 2015) for journal articles containing the B3LYP functional in combination with DZ (6-31G*, def2-SVP, cc-pVDZ, aug-cc-pVDZ), TZ (6-311G*, def2-TZVP, cc-pVTZ, aug-cc-pVTZ), and QZ (def2-QZVP, cc-pVQZ, aug-cc-pVQZ) basis sets from the periods 1995–2005 and 2005–2015.

groups are, a DZ basis is sometimes the only choice. Even with modern computational equipment, a sufficiently fast, and at the same time, reasonably accurate and interaction-consistent electronic structure method is mandatory for the screening of a large conformational space, for instance in the fast growing field of organic crystal structure prediction.[16,17,18] Therefore, in this short Review article, we want to emphasize the problems that arise from using a small DZ or related BSs, give an overview of methods to circumvent these problems, and discuss some exemplary calculations to provide a survey on the accuracy of the selected methods. This work extends our previous activities in the field which were focused specifically on B3LYP/6-31G* thermochemistry.[19] For related papers concerning noncovalent interactions see refs. [20,21,22,23,24,25,26,27].



Stefan Grimme studied chemistry and finished his PhD in Physical Chemistry at the University of Braunschweig in 1991 on a topic in laser spectroscopy. He did his habilitation in theoretical chemistry in the group of Sigrid Peyerimhoff at the University of Bonn. In 2000, he got the C4 chair for theoretical organic chemistry at the University of Münster. In 2011, he accepted an offer as the head of the newly founded Mulliken Center for Theoretical Chemistry at the University of Bonn. He has published more than 380 research articles in various areas of quantum chemistry. He is the recipient of the 2013 Schroedinger medal of the World Organization of Theoretically Oriented Chemists (WATOC) and the 2015 Gottfried–Wilhelm–Leibniz price of the German Research Foundation (DFG). His main research interests are the development and application of quantum chemical methods for large molecules, density functional theory, noncovalent interactions and their impact in chemistry.

2 Problems of Double-ζ Basis Sets

There are two major shortcomings of small BS Hartree–Fock (HF) or DFT calculations. The first one is the BS error. This error can be split further into the basis set superposition error (BSSE) and the basis set incompleteness error (BSIE).

Almost all quantum chemical simulations rely on systematic error compensations between the initial (reactant) and final state (product) calculation. The BSSE is caused by the fact that with a small BS, the monomers and the complex in a reaction are not treated on equal footing which destroys the error compensation. Typically, this is discussed in the context of noncovalently bound complexes, but the same phenomenon also appears for covalent-bond-forming chemical reactions, as well as in intramolecular transformations. In a dimer complex *AB*, the BS is larger than the individual ones of the monomers *A* and *B* because the unoccupied orbitals from *A* can be used by *B* and vice versa. This variational 'borrowing' of basis functions leads to an artificial energy lowering of the complex.

The most common approach to circumvent an intermolecular BSSE is the counterpoise (CP) correction scheme proposed by Boys and Bernardi (BB-CP).[28] The BB-CP counterpoise correction ΔE^{CP} for a dimer complex *AB* is defined as

$$\Delta E^{CP} = E(A)_a - E(A)_{ab} + E(B)_b - E(B)_{ab} \quad (1)$$

where *a* and *b* are the BSs belonging to the monomers *A* and *B* in their frozen *AB* complex geometries. This approach is also termed molecular CP correction, as only two fragments (the former monomers) are taken into account. Although the BB-CP approach is not free of criticism,[29,30,31] it is widely used and found to be a robust approximation for the self-consistent field (SCF) methods HF and DFT when applied to molecular aggregates.

The BSSE depends on the number of virtual functions that are supplied by the additional fragment in the complex and on their respective overlap. Because the HF/DFT total energies converge exponentially with respect to the BS size, the initial increase of BSSE with BS size eventually decreases as the CBS limit is approached. The electron density decays exponentially with the distance and the corresponding exponent is determined by the ionization potential of the fragment.[32,33] Because the inclusion of Fock exchange in a hybrid functional increases the ionization potential, this leads to a more compact density, a smaller density overlap of neighboring atoms, and a smaller BSSE. This can be qualitatively described as

$$E_{BSSE} \propto N_{bf} \times exp(-N_{bf}) \times exp\left(-\sqrt{2I}\,r\right), \quad (2)$$

with the number of virtual basis functions N_{bf}, ionization potential *I*, and electron-molecule distance *r*. We have adjusted this function with variable prefactors to the Boys–Bernardi CP energy of the S66[34] noncovalent dimers (see below) for functionals with varying amount of Fock exchange (PBE: 0%, B3LYP: 20%, PBEh-3c: 42%, HF: 100%) and increasing BS size

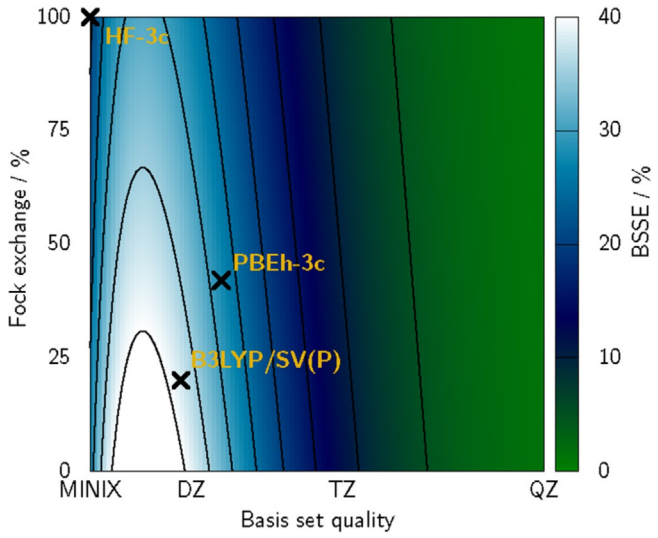


Figure 2. Contour plot of the relative BSSE as a function of the basis set size (minimal, DZ, TZ, QZ) and the amount of Fock-exchange. An interpolating function [Equation (2)] was fitted based on the S66 Boys-Bernardi counterpoise energies calculated for the functionals PBE, B3LYP, PBEh-3c, and HF.

(MINIX, def2-SV(P), def2-TZVP, def2-QZVP). The corresponding contour plot is shown in Figure 2.

The BSSE is most pronounced for medium-sized BSs of double-ζ quality and can be more than 40% of the binding energy. In a minimal BS, the neighboring fragment has only few (even zero for rare gas atoms) virtual orbitals, and the extension of the variational freedom is minor (small BSSE). In a CBS, the virtual space is huge, but the energy gain is zero because it is already converged in the single fragment basis (no BSSE). For medium sized BSs, the increase in the number of virtual functions and the corresponding lowering of the energy can be substantial. In general, BSSE leads to overestimated binding energies and underestimated interatomic distances.

Similar to the formation of a complex out of monomers, one part of a molecule, such as a functional group, can borrow basis function from another nearby part. This leads to the concept of intramolecular BSSE (IBSSE).[35,36,37] A uniform and clear definition of the IBSSE is missing, but its influence on energetics and structures of molecules has been recognized.

The BSIE is an inherent problem of any finite BS expansion. It leads to insufficient descriptions of physical effects such as Pauli repulsion, electrostatics and polarization, and thus, often to a systematic lengthening of bonds.[38] In practice, BSSE and BSIE are not strictly distinguishable, but we will focus on the effects due to BSSE in the following sections.

Though one should try to minimize the BSSE (corresponding to the 'green areas' in Figure 2), a small BSSE is not a sufficient criterion for a good basis set. A minimal BS, for instance, has a relatively small BSSE, but cannot describe certain physical effects like polarization well. Furthermore, additional basis functions do not automatically lead to a more complete basis. They need to have the proper shape, which is a nontrivial requirement, and basis set optimizations have been carried out for decades. We typically find the Ahlrichs sets optimal for molecular SCF-type calculations, and they have been only slightly adjusted and optimized for composite methods like PBEh-3c.[39] Similarly, the amount of HF exchange should not be increased too much to lower the BSSE because this would eliminate the account of important (static) electron correlation effects. The correct electron density can be best reproduced with a medium amount of HF exchange (about 20% to 50%), but other options (GGA or plain HF) can have advantages, too.

The second major shortcoming of common HF and (semi-local) DFT approximations is the inherent lack of a correct description of the London dispersion energy. For large interatomic distances >4.5 Å, the interaction between atoms or comparably nonpolar molecules is dominated by long-range correlation effects, called London dispersion. This type of interaction has a $-C_6/R^6$ distance dependence and is not included in any semi-local exchange-correlation functional. Modern density functionals exist,[40,41] which include correlation effects in the medium-distance regime (2.5 to 4.5 Å) to a strongly varying degree, but they do not provide the correct asymptotic behavior. The inclusion of London dispersion interactions in the mean field HF and DFT framework can be achieved by mainly three strategies. The first one is the construction of a nonlocal density functional correlation kernel (DFT-NL) often referred to as van der Waals density functional vdW-DF or its simplified variant VV10. In vdW-DF, the dynamical charge density response function is approximated by local dipole models which leads to tractable integral formulations. The VV10 method has especially been shown to yield reasonably good geometries and reasonably accurate binding energies.[42,43,44] While these special nonlocal functionals can, in principle, also be evaluated in small basis sets, this combination is rarely applied, and in the present Review we focus on inherently more efficient methods. The above-mentioned local response can be partitioned to atomic contributions leading to semiclassical London dispersion corrections. In the modern variants, the computed leading order dispersion coefficients are used to compute higher order contributions as done in the D3[45] (see below), exchange-dipole (XDM),[46,47] and many-body dispersion (MBD) models.[48] In contrast to the first two approaches, modified effective one-electron potentials do not describe the correct physical origin of London dispersion forces. However, they can be trained to mimic these interaction to some degree as shown by the dispersion correcting potentials (DCP).[49,50]

The reason why small BS DFT (or HF) calculations like B3LYP/6-31G* can perform surprisingly well is immediately recognized when looking at the two largest error sources and their (partial) compensation. The first one is the BSSE which leads to too strongly bound complexes, while the second flaw is the missing London dispersion energy resulting in too weak interactions. The prerequisite for a favorable error compensation is that dispersion and BSSE are of similar magnitude in a sufficiently large distance regime. However, this does not hold in general due to the fundamentally different functional dependence of BSSE and dispersion with respect to the distance separation, (exponential vs. R^{-6}) which is highlighted in Figure 3. We have calculated the dispersion and BSSE contribution for the S66×8 set (66 molecular dimers at eight different center of

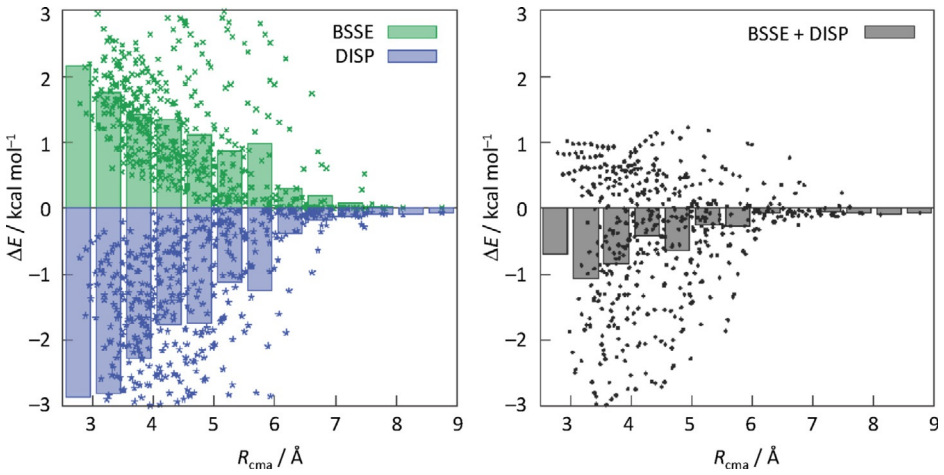


Figure 3. Distance behavior of the dispersion energy and BSSE as calculated with second-order DFT-SAPT and the Boys–Bernardi counterpoise scheme, respectively, for the PBE0/SV(P) method on the S66×8 molecular dimers. The individual energies with integrated contributions in 0.5 Å bins (bars, left) and the possible error compensation if both contributions are neglected (bars, right) are shown. The crosses refer to the individual values for the 528 complex geometries. Note that while the BSSE is a negative quantity which has the same sign as the stabilizing dispersion energy, the plotted CP correction is positive (repulsive).

mass distances) with DFT-SAPT (symmetry-adapted perturbation theory)[51,52] and the Boys–Bernardi method, respectively, at the PBE0/SV(P) level. While the two contributions roughly cancel each other on average (bars, right plot), the individual values for the complexes have a significant scatter showing that either dispersion or BSSE can dominate. It is clear that systematically accurate results cannot be obtained if both contributions are not properly included.

If we assume that BSSE and London dispersion effects cancel precisely at the equilibrium distance of the stacked benzene dimer, this cannot hold for non-equilibrium distances (compare with Figure 4). Thus, for reliable results, one needs to correct for both dispersion and BSSE at the same time. In the following section we will present an overview over existing methods that were designed for that purpose and are used in combination with HF or DFT.

These corrections directly determine the energy and the geometry of a given molecule and thus, indirectly also influence other properties, like the charge population, the chemical shift, the dipole moment, or the gap between the highest occupied and the lowest unoccupied molecular orbital. However, if a small basis set is, for example, lacking the necessary amount of polarization and diffuse functions to correctly calculate the dipole moment, the BSSE correction cannot repair this.

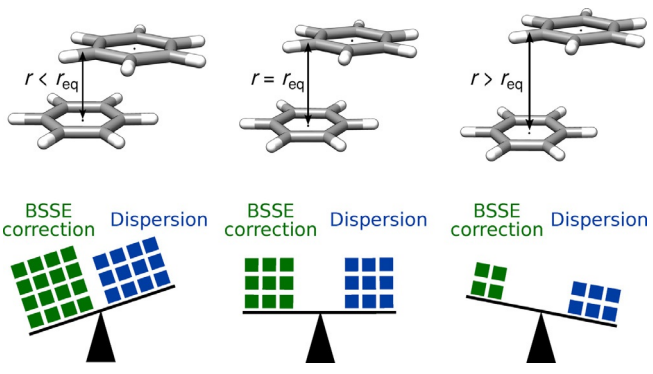


Figure 4. Sketch of the error compensation between BSSE correction and dispersion. In this example for the noncovalently bound benzene dimer, we assume that they are balanced at the intermolecular equilibrium distance. This is not the case for smaller or larger distances due to the different distance dependence of BSSE and dispersion. The BSSE decays exponentially whereas dispersion decreases slower with R^{-6}.

3 Methods Treating Dispersion and BSSE

One possibility to include London dispersion effects in DFT calculations is to correct the long-range interaction by atom-centered potentials, so called dispersion-correcting potentials (DCPs).[50,53,54] Though the correct physical terms leading to the London dispersion interaction (zero point energy of coupled frequency-dependent polarizabilities) cannot be described by DCPs, their mathematical form, together with parameter adjustment, can empirically capture attractive dispersion-like forces to a rather high degree. The DCPs designed by DiLabio et al. resemble traditional effective core potentials (ECPs) and are similar to the earlier plane-wave approach proposed for periodic DFT by Lilienfeld et al.[49,55] The general idea is to use a set of reference data and fit the interaction which is not covered by the density functional into additional atom-centered potentials. These potentials need a high degree of flexibility and should distinguish atoms in different hybridization states.

Typical potentials $U_l(r)$ are composed of atom-centered Gaussian-type functions and have the following form:

$$U_l(r) = r^{-2} \sum_{i=1}^{N_l} c_{li}\, r^{n_{li}}\, e^{-\xi_{li} r^2} \qquad (3)$$

where l is the angular momentum, N_l corresponds the number of Gaussian functions, n_{li} is the power of r (electron-nucleus distance), c_{li} is the coefficient of the Gaussian function, and ξ_{li} is its exponent.

For B3LYP-DCP[53] the n_{li} are fixed to the value of two, and the exponents and coefficients were optimized for a set of 16 noncovalently bonded dimers. Two sets of DCPs have been developed: one intended for use with a CP correction and one for use without. The latter also compensate for the BSSE to a significant extent. Thus, BSSE and dispersion are treated simultaneously but the correct R^{-6} asymptotic behavior of the London dispersion is not met (though this is in principle possible by a complete expansion with all orders in N_l).

The advantage of this approach is its simple implementation and that DCPs can be used with any computational chemistry program package that can handle ECPs. The corresponding exponents and coefficients for each element are provided in the input files, and no modifications of the program itself are necessary. Furthermore, with the use of DCPs or related approaches, dispersion effects are included on the electronic structure level, and the electron density can adjust to these effects. A disadvantage in particular for large systems is that the incorporation of DCPs into the self-consistent field (SCF) procedure often increases the number of cycles needed for convergence and, hence, the overall computational time. Another drawback is that a DCP has to be fitted for each element in combination with a specific density functional and a given BS. For each element, one has to gather enough reliable reference data, which can be difficult when aiming at an extension to heavier elements. Currently, for small DZ basis sets, which are the focus of this Review article, DCPs are available for the elements H, C, N, and O and the B3LYP functional.[53,54] They are suitable for the 6-31+G(d,p) basis set or larger, but the use of 6-31+G(2d,2p) is recommended. The first DCP for the carbon atom needed to be revised due to the too large exponents that hampered the correct description of C–C bond breaking or bond formation and with the revised DCP, noncovalent interactions and covalent chemical reactions are described with similar accuracy.[54] We will refer to and use these improved DCPs throughout this article.

In the original publication, the performance of B3LYP-DCP/6-31+G(2d,2p) for noncovalent interactions was tested on several benchmark sets, and we will give here only some examples. Its accuracy for the S66[34] test set of small noncovalently bound dimers is excellent. The mean absolute deviation (MAD) of the binding energies compared with the reference is only 0.19 kcal mol^{-1}. For comparison, B3LYP-D3(BJ) with the quadruple-ζ basis set def2-QZVP yields an MAD of 0.28 kcal mol^{-1}.[56] For the HSG[57] set of 21 dimers and trimers which are present in the complex of the inhibitor indinavir and HIV-II protease, the performance is also encouraging. The MAD compared to the revised reference values (HSG-A[58]) is 0.16 kcal mol^{-1}. Furthermore, B3LYP-DCP/6-31+G(2d,2p) was applied to the S12L[59] set of supramolecular complexes. One out of the 12 complexes contains a Cl atom, and another one involves Fe. As no DCPs exist for Cl and Fe, these atoms were left uncorrected. The final MAD for the S12L set is 2.6 kcal mol^{-1}. This result is similar to those obtained with PBE-D3 or PBE-NL in combination with def2-QZVP (2.1 and 2.3 kcal mol^{-1}, respectively).[60] Overall, these examples show, that B3LYP-DCP/6-31+G(2d,2p) as a method on the double-ζ level can provide results of quadruple-ζ quality.

Its good performance was confirmed by Goerigk who compared B3LYP-DCP, B3LYP-NL, and B3LYP-D3(BJ) in combination with the 6-31+G(2d,2p) basis set for noncovalent complexes, relative energies of conformers, basic properties, and reaction energies.[61] An overall comparison revealed B3LYP-NL as the most robust and accurate approach, closely followed by B3LYP-D3. However, for these two methods, the influence of BSSE effects on the binding energies of noncovalently bound complexes can be larger than it is the case for B3LYP-DCP. Furthermore, it was verified that the revised DCP for carbon actually does improve the overall performance, though the change for electron affinities and ionization potentials is negligible.

Recently, the DCP scheme was coupled to the atom-pairwise D3 dispersion correction (see below) for the BLYP functional and the 6-31+G(2d,2p) basis set.[62] In this BLYP-D3-DCP approach, the exponents of the DCP tend to be larger than those for the ones developed previously. Thus, they mostly have an impact on the electron density close to the nuclei and mainly influence the covalently bonded parts. This is reflected in the large improvement for barrier heights compared with BLYP-D3 but only small enhancements for noncovalent interaction energies. This result also indicates that typical GGA problems like the self-interaction-error (SIE) can be corrected with DCP (see also refs. [63,64]).

A different approach is the combination of a dispersion correction and a CP correction, which are developed independently from each other, but which are simultaneously employed in a calculation. For the treatment of London dispersion, we use our efficient semiclassical D3(BJ) correction[45,65] that can simply be added on top of a converged standard DFT or HF calculation. For reviews and overviews of other state-of-the-art dispersion corrections see refs. [23,24,25]. Within the D3(BJ) scheme, the energy contribution is calculated as a sum over all atom pairs *AB*

$$E_{disp}^{D3} = -\frac{1}{2}\sum_{A\neq B}\sum_{n=6,8} s_n \frac{C_n^{AB}}{R_{AB}^n + f(R_0^{AB})^n} \tag{4}$$

where, C_n^{AB} denotes the averaged coordination-number dependent (isotropic) n^{th} order dispersion coefficient for each atom pair *AB*. The order *n* equals 6 and 8, introducing an R_{AB}^{-6} long-range and an R_{AB}^{-8} medium-range term. The s_n are the global scaling factors. For common density functionals, s_6 is usually set to unity to ensure the correct asymptotic behavior, whereas s_8 is optimized for each functional. $f(R_0^{AB})$ is the damping function as introduced by Becke and Johnson[68,66]

$$f(R_0^{AB}) = a_1 R_0^{AB} + a_2 \tag{5}$$

with the fitting parameters a_1 and a_2, and the cut-off radii $R_0^{AB} = \sqrt{(C_8^{AB}/C_6^{AB})}$ or simplicity we will refer to D3(BJ) (which is the current default for the method) as D3 in the following. An

Axilrod–Teller–Muto (ATM)-type three-body (dipole-dipole-dipole) term is also available in the D3 code including its analytical derivatives.[67,68] The importance of many-body dispersion interactions has been recently analyzed by various groups,[69,70,71] but is not in the focus of this Review.

The D3 dispersion correction can, in principle, be combined with any BSSE correction, for example, with the standard BB-CP procedure. In this scheme, however, the computational cost quickly increases for larger complexes because full BS calculations for the fragments have to be conducted. If each atom is considered as an individual fragment, one can define an atomic counterpoise correction (ACP)[72] as done by Jensen. The ACP(x) correction $\Delta E^{ACP(x)}$ is expressed as a sum over all atoms A

$$\Delta E^{ACP(x)} = \sum_A E(A)_a - E(A)_{as} \tag{6}$$

where a denotes the regular basis set and as is a subset of a which always includes the regular basis function on A. For the intramolecular case, this subset further includes all basis functions from atoms x bonds apart, and for the intermolecular case, it contains all basis function of the other monomer. When all basis functions of the whole system are included in the subset, the ACP(1) correction equals the CP^{aa} correction published earlier by Galano and Alvarez–Idaboy.[73] These BSSE corrections have a highly reduced computational cost and the advantage to treat inter- and intra-molecular effects conceptually on the same level. Unfortunately, these approaches lack the availability of nuclear gradients. Therefore, we recently developed a geometrical counterpoise correction (gCP)[74] that solely depends on the molecular geometry. It provides a fast, conceptually simple, but physically reasonable energy and gradient correction for the BSSE in large molecules and condensed phase systems.

Within the gCP scheme, the difference in atomic energy E_A^{miss} between a large, nearly complete BS and the target basis (here DZ) is calculated (and tabulated) for each atom at the HF or B3LYP level and used as a measure for the BS incompleteness. The E_A^{miss} are then multiplied with a decay function depending on the interatomic distance R_{AB} and summed up over all atom pairs AB

$$E_{BSSE}^{gCP} = \sigma \sum_A \sum_{A \neq B} E_A^{miss} \frac{exp\left(-\alpha (R_{AB})^{\beta}\right)}{\sqrt{S_{AB} N_B^{virt}}} \tag{7}$$

where α, β, and σ are functional and BS-specific fitting parameters. As the density has an exponential tail, the decay function is exponential. Due to the strong dependence of the BSSE on the charge density overlap in SCF methods, this function is normalized by the square-root of the Slater-overlap S_{AB} times the number of virtual orbitals N_B^{virt} on atom B. The overlap integrals S_{AB} are evaluated over single s-type orbitals centered on each atom using optimized Slater exponents and weighted by the last fitting parameter η. The fit was performed for HF and B3LYP together with the target basis set on the S66×8 test set.[34] Standard BB-CP corrected interaction energies for the respective method were employed as reference values. The accuracy gained by a refit for different density functionals is negligible and thus, the use of the B3LYP parameters is recommended for common GGA or hybrid functionals.

One advantage of the D3-gCP combination is its availability for almost all elements in the periodic table, and the existence of analytical nuclear gradients. Furthermore, the scaling behavior with system size is low, and the computational prefactor is small. This results in very fast computations even for thousands of atoms. Another benefit is that the corrections can simply be added on top of any DFT or HF calculation without need for a specific implementation into a program package. A drawback is the semiempirical character of both corrections and thus, the need for a parameter fit for every functional in case of D3 and each functional/BS combination in case of gCP. However, as mentioned before, the gCP dependence on the functional was found to be negligible and, hence, only adjustments for each basis set and for HF or DFT have to be made. In addition, the corrections do not depend on the density and thus, the electronic structure is not directly affected, though, an indirect effect due to the altered geometry is present.

Note, that in general, the gCP scheme can be combined with any dispersion correction. One example is a recent publication by Yoshida et al., who used gCP for HF/6-31G(d) together with their own dispersion correction to describe the HIV-1 protease and its potent inhibitor KNI-10033.[75] The good performance of DFT-D3-gCP/DZ and HF-D3-gCP/DZ for noncovalent interactions was already noted in the original gCP publication. The gCP correction is able to provide a reasonable estimate for the intermolecular BSSE with an error of 10–30%. For the S22 benchmark set[76] for example, PW6B95-D3-gCP/def2-SVP yields an MAD of 0.84 kcal mol^{-1} for interaction energies. In case of B3LYP/6-31G*, the MAD can be reduced from 2.67 to 0.88 kcal mol^{-1} upon application of both the D3 and gCP correction. Geometry optimizations of the S22 complexes showed that B3LYP-D3-gCP/6-31G* as well as HF-D3-gCP/SV reproduce the reference structures well. In case of 9-helicene the nonbonded C-C distances can be accurately computed with HF-D3-gCP/SV within a few pm.[74] Somewhat unexpectedly, of all various method/basis set combinations tested, HF-D3-gCP/MINIS performs particularly well for noncovalent interactions.

In a recent publication, the shortcomings of the B3LYP/6-31G* model chemistry, as explained in the previous section, were analyzed and it was shown that D3-gCP can account for the major deficiencies and that B3LYP-D3-gCP/6-31G* yields reasonably accurate thermochemical results.[19] Benchmark calculations on the general main group thermochemistry, kinetics, and noncovalent interaction *meta*-database GMTKN30[5] showed a statistical improvement when both corrections are used. The weighted MAD decreased from 8.8 (B3LYP/6-31G*) to 6.9 kcal mol^{-1} (B3LYP-D3-gCP/6-31G*). It was statistically confirmed that the partial error compensation of missing dispersion and BSSE in plain B3LYP/6-31G* is unsystematic and depends on the chemical nature of the system at hand. The improvement gained with the D3-gCP scheme is largest for systems that exhibit noncovalent interactions, but reaction energies and barrier heights are also improved.

Goerigk and Reimers used DFT-D3-gCP/DZ and HF-D3-gCP/DZ for geometry optimizations of several test sets which aim at describing important interactions in protein structures.[77] Various functionals as well as HF in combination with different DZ basis sets were employed for the P26 test set,[78] in order to investigate their performance for conformers of five tri-peptides containing aromatic side chains. For the 6-31G* basis without any correction as an example, structural RMSDs around 0.5 Å are observed. When only gCP is employed the RMSDs rise, and with solely the D3 correction the RMSDs drop significantly. When the combined D3-gCP scheme is used, the RMSDs decrease to values of about 0.15 Å, which are slightly higher than those with the D3 correction only. It seems that in this specific case without gCP, a fortunate error compensation occurs which, however, does not hold in general as discussed above.

Martinez et al. showed that uncorrected DFT or HF with DZ basis sets can yield good geometries for small proteins.[79] They compiled a set of 58 proteins with up to 35 residues (up to 600 atoms) and compared their results to experimental X-ray or nuclear magnetic resonance derived structures. The ab initio methods HF and ωPBEh are able to provide geometries of the same quality as highly parametrized force fields and are consistently better at reproducing experimental structures for proteins with disordered regions, judged by standard health metrics.

Reimers et al. optimized a portion of an ensemble of conformationally flexible lysosome structures by a divide-and-conquer approach and compared their results with X-ray crystallography data.[80] The functionals BP86 and B3LYP, as well as HF, were employed together with the 6-31G* basis set and in combination with the D3-gCP scheme. Regarding the all atom RMSD and the R-factor, the best and most consistent structures are obtained when both the D3 and the gCP correction are used. Compared to the uncorrected methods, employing only D3 gives similar results and only gCP yields worse values. This observations resemble the ones made for small peptides[77] and again show that one cannot rely on error compensation effects.

Extension of the gCP correction to periodic HF/DFT calculations enables the use of the D3-gCP scheme for molecular crystals.[81] The corrections were applied to PBE and B3LYP for the X23 molecular crystal test set[82] and decrease the MAD of the sublimation energies significantly by more than 70% and 80%, respectively, to small residual MADs of about 2 kcal mol^{-1} (corresponding to 13% of the average sublimation energy). Furthermore, variation of the interlayer distances for graphite yielded a potential energy surface that is very close to the converged basis set reference and agrees very well with experimental stacking distances.

Up to now, D3 and gCP have been fitted independently of each other, but applied at the same time in a calculation. The effect of fitting both corrections together was briefly tested for HF-D3-gCP and showed small improvements over the independent fit, mainly due to removing redundancies in the potentials.[74] We introduced two composite methods that also make use of these corrections, but which were fitted or adjusted in the presence of each other and thus are suggested as one composite approach with a fixed basis set. As we noticed the good performance of HF-D3-gCP/MINIS for noncovalent interactions during the development of the gCP correction, we proposed HF-3c, a minimal basis set Hartree–Fock method with three atom-pairwise corrections: D3, gCP, and an additional term, which corrects for short-range basis set (SRB) incompleteness effects.[83] The six parameters of the gCP and D3 correction terms were fitted together on the S66 test set and were kept constant in the subsequent fitting procedure of the third SRB term. This composite method corrects for both dispersion and BSSE and is suggested as an alternative to semiempirical methods or DFT, in particular when SIE is acute. HF-3c yields reasonable noncovalent interaction energies and good geometries of small organic molecules, as well as supramolecular complexes and small proteins.[83,84,77] As this Review focuses on DZ basis sets, we will not discuss this method further.

A related composite approach is our recently developed PBEh-3c method, a global hybrid functional with a DZ basis set, that is meant to fill the gap between existing semiempirical methods or HF-3c and large basis set DFT with respect to the cost-accuracy ratio.[39] The term '3c' indicates its relation to HF-3c, and the corrections are a slightly modified gCP, D3, and minor modifications to the def2-SV(P) BS (dubbed def2-mSVP) for boron to neon in order to ensure consistent bond lengths for all elements. PBEh-3c yields accurate geometries which were verified for small molecules as well as medium sized molecules, noncovalently bound complexes, and molecular crystals. The overall deviations from reference structures are tiny and practically of MP2/def2-TZVPP quality, while the geometries are obtained at a much lower computational cost (speed-up of a factor of 50–100). All other DFT/small BS methods tested yielded larger deviations. For the S22 set of noncovalent complexes PBEh-3c agrees well with the MP2 reference geometries, the mean deviation (MD) for intermolecular center-of-mass distances is only 3 pm. For molecular crystals, the PBEh-3c accuracy for geometries in the X23 and ICE10[85] sets approaches TPSS-D3/'large BS' results. The mean absolute deviations in the computed unit cell volume are 2.7% and 5.0%, respectively, for X23 and ICE10. Although PBEh-3c was mainly designed for the computation of structures, it yields reasonable results for thermochemistry, barrier heights, and general noncovalent interactions. Clearly, due to the small BS, the accuracy of dispersion-corrected hybrid DFT in a QZ basis set cannot be reached.

4 Comparison of Methods for Noncovalently Bound Systems

In the following section, we will compare the performance of the various discussed methods for some exemplary noncovalently bound systems. We chose HF, HF-D3-gCP, B3LYP, B3LYP-D3-gCP, B3LYP-DCP, M06-2X, and PBEh-3c. The def2-SV(P) basis set will be applied in all cases (modified for PBEh-3c), except for B3LYP-DCP where the 6-31+G(2d,2p) basis will be used. An

Table 1. Overview of the applied methods and their capability to treat BSSE and dispersion.[a]

Method	BSSE correction	Dispersion correction
HF	no	no
HF-D3-gCP	yes	yes
B3LYP	no	no
B3LYP-D3-gCP	yes	yes
B3LYP-DCP	yes	(yes)[b]
M06-2X	no	(yes)[c]
PBEh-3c	yes	yes

[a] B3LYP-DCP is used with the 6-31+G(2d,2p) basis set, all other methods with the def2-SV(P) basis set (modified in case of PBEh-3c). [b] BSSE and dispersion are treated together in one ECP leading to the wrong asymptotic behavior for the dispersion interaction. [c] The dispersion interaction has the wrong asymptotic behavior.

overview of the capability of these method to treat dispersion and BSSE is provided in Table 1.

4.1 Noncovalent interaction energies

The accuracy for noncovalent interaction energies of the aforementioned methods was tested on several benchmark sets. We chose three sets for the interaction of small to medium sized systems (WATER27,[86,87] S22,[76,58] S66[34]), two sets for large and supramolecular systems (L7,[88] S30L[84]), and two test sets for molecular crystals (ICE10,[85] X23[89,82]). For each test set one exemplary system is depicted in Figure 5.

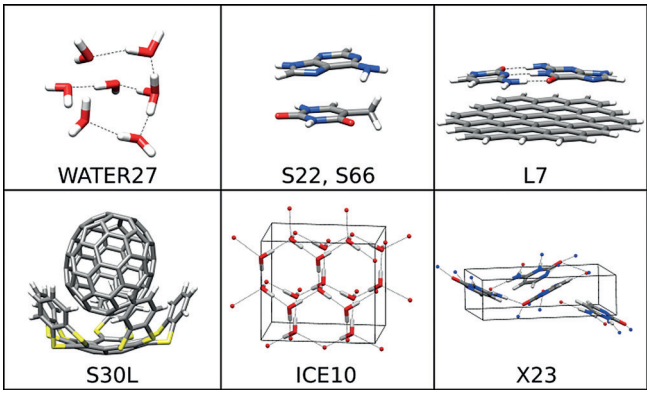


Figure 5. Example systems for each test set. Hydrogen bonds are indicated by dotted lines.

The WATER27 test set contains 27 neutral and charged water clusters with up to 20 water molecules. The S22 set consists of 22 noncovalently bound model complexes that show hydrogen bonding, dispersion interactions, and mixed electrostatic-dispersion binding motifs. The S66×8 test set is similar to S22 but with less emphasis on nucleobases. Furthermore, reference geometries and energies are provided at eight different distances of the monomers, which allows the extraction of the minimum of the intermolecular potential energy surface (PES) of a given method via an interpolation procedure. The reference energies for these three sets refer to the estimated CCSD(T)/CBS level of theory. For the S22 we use the revised values by Sherill et al.[58] For the $(H_2O)_{20}$ complex contained in the WATER27 set we use the reference values computed on the incremental CCSD(T)(F12*)|MP2-F12+ΔMP2 level by Friedrich.[87] The L7 test set comprises seven larger, mostly dispersion-stabilized complexes of organic molecules. We use the revised reference values on the estimated DLPNO-CCSD(T)/CBS* level of theory.[90] The S30L set is an extension of the S12L set,[59,60] which was the first test set for large host–guest complexes. It contains 30 realistic host–guest complexes with charges from −1 to +4 and up to 200 atoms, featuring various typical noncovalent binding motifs like hydrogen and halogen bonding, π–π stacking, nonpolar dispersion, CH–π, and cation-dipolar interactions. The reference association energies are back-corrected values from experimentally measured association free energies. ICE10 includes ten ice polymorphs and X23 compiles molecular crystals that show mainly van der Waals or hydrogen bonding or a mixture of these two interaction motifs. For these two sets the reference lattice energies were derived from experimental values which are further corrected for zero-point vibrational and thermal effects.

As the absolute interaction energies differ by almost three orders in magnitude, we give mean absolute relative deviations (MARDs in %) from the reference energies for all test sets and methods in Figure 6. Because of SCF convergence problems for some molecular crystals, the HF results for the periodic benchmarks were omitted. The values are color-coded as suggested by Martin[91] in order to provide an easy overview and the best two methods for each test set are highlighted.

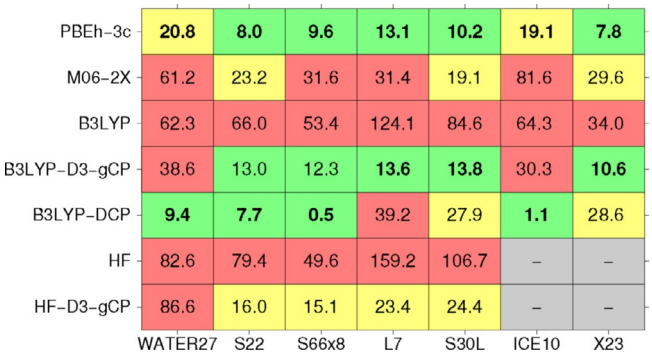


Figure 6. Mean absolute relative deviations (MARDs, in %) for different methods compared with the reference values for several test sets. MARDs below 15% are color-coded in green, those below 30% in yellow, and those higher than 30% in red. For each set, the two best performing methods are highlighted. PBEh-3c includes the ATM three-body dispersion term by default; for B3LYP-D3-gCP and HF-D3-gCP it was included for the large systems (L7 and S30L test sets). In case of B3LYP-DCP, two systems of the S30L were omitted due to missing functions of the 6-31+G(2d,2p) basis set for iodine, and for X23, eight systems had to be disregard due to SCF convergence problems.

As expected, the plain B3LYP functional or HF without any corrections cannot properly describe noncovalent interactions. Already for small systems contained in the WATER27, S22, or S66×8 sets huge MARDs of 50 to 80% are obtained. In many

cases, these methods yield unbound complex states. The same is observed for the supramolecular test sets and the MARDs for L7 and S30L are even larger (80 to 160%). The performance of B3LYP for the ice polymorphs is similar to the molecular WATER27 set. For both the MARD is about 60%. Surprisingly, the MARD for the X23 set of molecular crystals is with 34% much smaller than the corresponding values for the S22 and S66×8 sets (about 60%), and the performance is actually similar to M06-2X and B3LYP-DCP. For the mixed hydrogen-bonded crystals, the error compensation between missing dispersion and neglected BSSE in plain B3LYP is rather good around the corresponding equilibrium geometry. While this explains the slightly smaller error compared to the other test sets, this compensation does not hold for stronger hydrogen bonding (significant overbinding of the various ICE10 polymorphs due to dominant BSSE) nor for purely London-dispersion-bonded X23 systems (significant underbinding due to missing dispersion interaction).

When the D3 and gCP corrections are added, the plain HF and B3LYP results can be improved tremendously. In case of HF, no improvement is observed for the WATER27 set, but for all others the MARD for HF-D3-gCP drops to 15 to 25% which is very reasonable for such a simple method. For B3LYP-D3-gCP the enhancement for WATER27 and ICE10 is much smaller than for the other sets, but still, the MARD is reduced from 60% to about 35%. Very good results are obtained for the S22, S66×8, L7, S30L, and X23 test sets which have MARDs of 10 to 14%. Compared with B3LYP-D3 with the large def2-QZVP basis set, the MARD for B3LYP-D3-gCP/DZ on the S22 set is doubled (6.2%[5] vs. 13%). If a TZ basis like def2-TZVP is employed, the MARD for the S22 is 11% which indicates remaining BSSE. Adding the gCP correction for the TZ basis set decreases the MARD to 7% which is similar to the value obtained with the QZ basis set. For the large supramolecular complexes, B3LYP-D3-gCP/DZ yields equally good values or even better results than B3LYP-D3/QZ. The MARD for the S30L set is 13.8% for B3LYP-D3-gCP/DZ and very similar for B3LYP-D3/QZ (13.2%[84]). For the L7 set, B3LYP-D3-gCP/DZ yields an MARD of 13.6%, which is less than half of the value for B3LYP-D3/QZ (32.5%[88,90]).

If DCPs are used for B3LYP instead of the D3-gCP correction, the behavior is very different. First we note the extraordinary good performance for the water-containing systems. The MARD of B3LYP-DCP is 9.4% for WATER27 and 1.1% for ICE10. Even with large basis set dispersion-corrected DFT calculations, it is difficult to reach this accuracy. This can partially be attributed to the basis set (6-31+G(2d,2p)) which contains two sets of additional polarization functions, as well as a diffuse set of *sp*-functions on nonhydrogen atoms, which is known to be important for these systems.[5,86,87] Therefore, the number of basis functions per atom is more comparable to a TZ basis and much larger than in def2-SV(P). The B3LYP-DCP results for S22 and S66×8 are also very good and the MARDs of 7.7 and 0.5%, respectively, are the lowest ones reported here. The MARD for S22 is very close to the already mentioned B3LYP-D3/QZ result (6.2%[5]). For large systems, however, the performance of B3LYP-DCP deteriorates significantly. The MARDs for the S30L and L7 sets are 39.2 and 27.9%, respectively. Although these values are about three times smaller than those for plain B3LYP, they are still about three times larger than for B3LYP-D3-gCP. For the X23 set there is no improvement compared to plain B3LYP. Both MARDs are about 35%, which is again three times as large compared to B3LYP-D3-gCP.

Finally, we discuss the recently published composite method PBEh-3c.[39] Its overall performance for all test sets is very good, and consistent accuracy for small as well as large complexes is evident. The MARDs for the S22, S66×8, L7, S30L, and X23 sets are found to be in the range of 8 to 13%. The performance for WATER27 and ICE10 is worse; the MARDs are about 20%. This indicates that the applied corrections cannot repair the higher basis set requirements in condensed hydrogen-bonded systems compared with only medium polar dimers. But nevertheless, the similar MARDs for WATER27 and ICE10 as well as S22/S66×8, L7/S30L, and X23 show that PBEh-3c treats the noncovalent interactions in small, large and periodic systems with the same accuracy. PBEh-3c is always one of the two best performing methods on any test set. The others are either B3LYP-DCP (WATER27, S22, S66×8, ICE10) or B3LYP-D3-gCP (L7, S30L, X23).

As this article mainly focuses on large systems, a closer look to the supramolecular complexes of the S30L set is appropriate. The association energies ΔE range from -17.4 kcal mol^{-1} for the halogen-bonded complex **15** up to -135.5 kcal mol^{-1} for the doubly positive charged complex **24**. Figure 7 shows a comparison of the ΔE values for PBEh-3c, M06-2X, B3LYP-D3-gCP, and B3LYP-DCP with the reference values.

As one can easily see, B3LYP-DCP and M06-2X exhibit large systematic overbinding as indicated by MDs of -9.6 and -6.4 kcal mol^{-1}, respectively. The largest errors for B3LYP-DCP are observed for the charged systems **23** to **30** and range from -11 to -27 kcal mol^{-1}. Somewhat surprisingly, the errors are also large (>-10 kcal mol^{-1}) for most of the hydrogen-bonded systems (**17**, **19** to **21**). As seen before, B3LYP-DCP performs exceptionally well for the WATER27 and ICE10 sets, the hydrogen-bonded dimers in S22/S66×8, but seems to fail for hydrogen bonds in these supramolecular complexes. Furthermore, the errors are large (-8 to -14 kcal mol^{-1}) for some of the π–π-stacked systems (**5**, **7** to **10**). The complexes **25** and **26** which also exhibit π–π-stacking as major interaction show a similar error. Apparently, for the small systems (on which the DCPs are fitted), the description of dispersion and the compensation for BSSE are reasonable, and accurate results can be obtained. For these large supramolecular complexes, the balancing of dispersion effects and BSSE is different, which is difficult to describe by a correction potential lacking the correct physics. As explained above, missing dispersion results in too weakly and BSSE in too strongly bound complexes. B3LYP-DCP overestimates the binding energy for all host–guest complexes and thus, the remaining BSSE seems to be the major error source. Whether the diffuse functions in the 6-31+G(2d,2p) basis play an additional negative role in the larger systems due to a more long-ranged BSSE is currently not clear.

The largest errors for plain M06-2X (-10 to -19 kcal mol^{-1}) are obtained for the complexes **5**, **9**, **11**, **12**, **17**, **29**, and **30**.

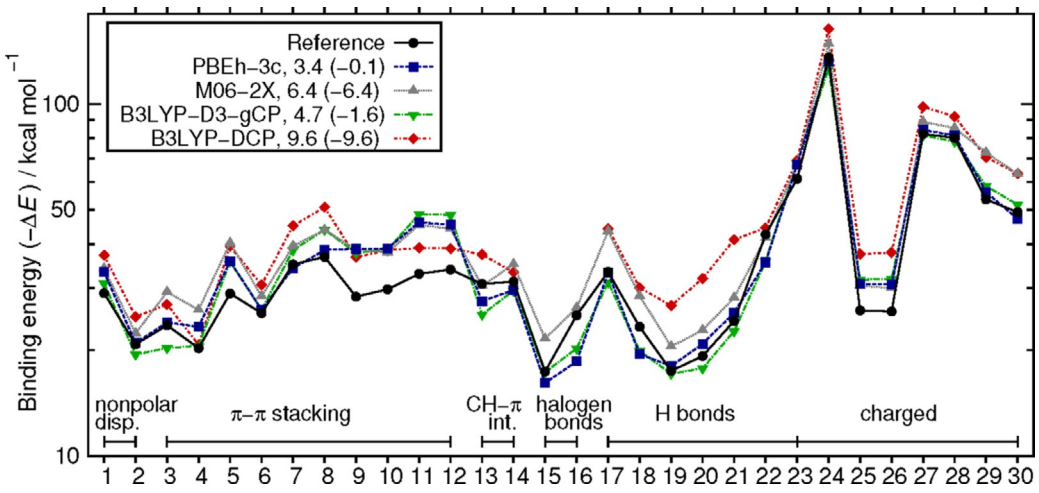


Figure 7. Comparison of the PBEh3c, M06-2X, B3LYP-D3-gCP, and B3LYP-DCP binding energies with reference values for the S30L test set. The complexes are sorted according to the most prominent type of interaction. In case of B3LYP-DCP, the complexes **15** and **16** were omitted due to missing functions of the 6-31 + G(2d,2p) basis set for iodine. The MADs and MDs in parentheses are provided for each method in kcal mol^{-1}.

Thus, M06-2X seems to have less trouble to accurately describe hydrogen-bonded complexes than to reproduce the reference values for the π–π-stacked systems. As the MARD for S30L is similar to that of S22 and even better than for S66×8 (Figure 6), the incorrect asymptotic treatment of the London dispersion seems not to be a major error source. Much more problematic is the unaccounted BSSE, and therefore, binding energies are overestimated. When a TZ basis is used, the MAD drops to 2.5 kcal mol^{-1} and the MD is just 1.4 kcal mol^{-1}, indicating underbinding due to the missing long-range dispersion contribution.[84]

When dispersion and BSSE are both separately accounted for, as it is in B3LYP-D3-gCP and PBEh-3c with two different potentials, the errors decrease substantially. For B3LYP-D3-gCP the largest errors of −12 to −21 kcal mol^{-1} are observed for complexes **9** to **12**. For PBEh-3c the complexes **11** to **13**, **22**, and **24** show the largest errors of 6 to 9 kcal mol^{-1}. B3LYP-D3-gCP and PBEh-3c reach MDs of −1.6 to −0.1 kcal mol^{-1}, respectively, indicating small to almost no systematic overbinding. Compared with M06-2X and B3LYP-DCP, the MAD values for B3LYP-D3-gCP and PBEh-3c are with 4.7 and 3.4 kcal mol^{-1}, respectively, much lower. For comparison, the previous best results for S30L were obtained with PW6B95-D3/def2-QZVP, which yields an MAD of 2.4 and an MD of −0.1 kcal mol^{-1}.[84] B3LYP-D3/def2-QZVP is one of the worst performers at the large BS level and has an MAD of 4.1 and an MD −2.7 kcal mol^{-1}.[84] Thus, B3LYP-D3-gCP is able to provide close to QZ quality results but at a small fraction of computational cost, and PBEh-3c almost approaches the accuracy of PW6B95-D3/QZ. These examples show clearly how important it is to properly and consistently treat both dispersion and BSSE in large systems.

In Figure 8, we summarize the different contributions to the binding energy of the S66 dimers (minimum extracted from S66×8 potentials) for the HF and B3LYP methods evaluated in a def2-SV(P) basis set with and without correction schemes. We show the statistics of the deviations to CCSD(T) references as normal error distributions. The behavior of HF and B3LYP mean-field methods is very similar, which is typical for purely noncovalent interactions. Without any correction, the error spread is large (broad distribution) with a slight systematic underbinding. When only the gCP correction is applied, the error

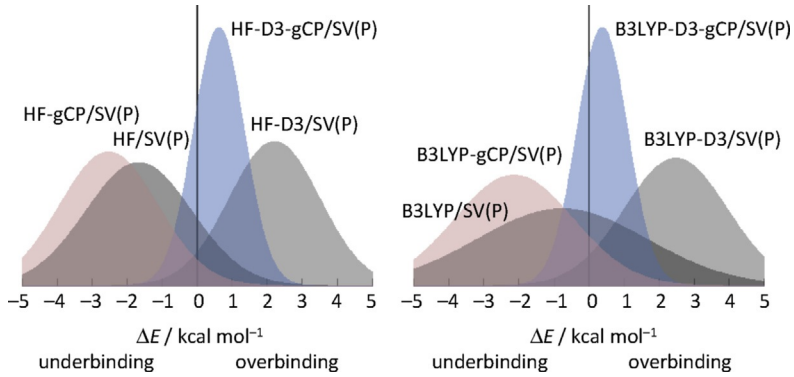


Figure 8. Error statistics of S66 equilibrium binding energies for corrected and uncorrected HF and B3LYP in a SV(P) basis set converted into normal error distributions.

spread decreases, but the underbinding is increased. The sole application of the D3 correction leads similarly to a smaller error spread and a systematic overbinding. Only the combination of both schemes leads to an excellent agreement with the reference data with MAD of 0.8 kcal mol^{-1} and 0.6 kcal mol^{-1} for HF-D3-gCP and B3LYP-D3-gCP, respectively.

4.2 Structures of noncovalently bound systems

In the following, the accuracy of the methods for optimized structures of noncovalent complexes is evaluated. As examples for small systems we chose the S22 and S66×8 sets that were already employed for the interaction energies as well as the P26[78] set, which contains different conformers of four peptides. For S22 and P26, the reference geometries were calculated on the MP2/TZ level of theory, and the root mean square deviation (RMSD) of the heavy atom positions, as well as the deviation of the intramolecular center-of-mass distance of the monomers, are used as performance measures. For the S66×8, the PES is used to determine the optimal intramolecular center-of-mass distance of the monomers, and therefore, the CCSD(T)/CBS level of theory is the reference as conducted similarly before.[39,92] For the supramolecular systems, we face the problem that there are no reference geometries available. The L7 and S30L systems were optimized on the TPSS-D3/TZ level of theory, which is certainly a good choice but not accurate enough to serve as a reference. In order to show exemplary the influence of London dispersion and BSSE in large supramolecular complexes, system **5** from S30 L and the **phe** complex from L7 were reoptimized on the TPSS-D3/def2-QZVP(-g/f) level, and these structures were used for comparison.

For the molecular crystal test sets ICE10 and X23, the reference geometries refer to experimental X-ray data which are isotropically corrected for zero-point vibrational and thermal effects.[85,39] Here, we use the deviation of the unit cell volume as measure to evaluate the accuracy of the tested methods.

Figure 9 depicts the geometries of two exemplary systems, the π–π stacked and hydrogen-bonded cytosine–uracil base pair, as obtained with the DFT methods in comparison with the reference structures. The results for HF and HF-D3-gCP are similar to B3LYP and B3LYP-D3-gCP and, therefore, these geometries are not shown.

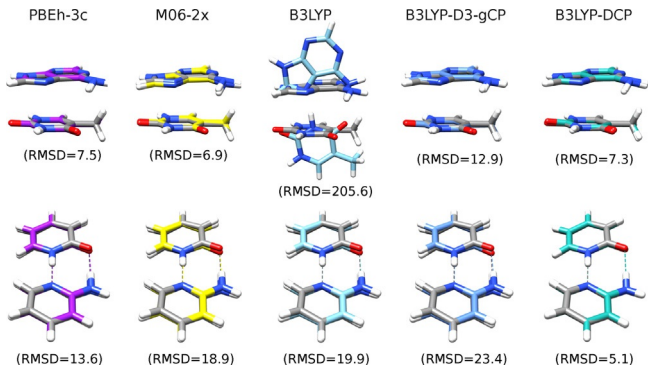


Figure 9. Comparison of the PBEh-3c, M06-2X, B3LYP, B3LYP-D3, B3LYP-D3-gCP, and B3LYP-DCP geometries (colored) with the reference MP2/TZ structure (in gray) for the π–π stacked cytosine-uracil base pair (top) and the hydrogen-bonded cytosine-uracil base pair (bottom) from the S22 test set. Hydrogen bonds are indicated by dotted lines. The RMSDs for the heavy atom positions are given in pm.

One can see immediately, that plain B3LYP cannot even qualitatively correctly reproduce the π–π stacked structure (Figure 9, top). As the London dispersion is completely missing, the π–π stacked dimer is not a minimum on the PES, and the optimization leads to the hydrogen-bonded structure. In contrast, all other four methods which include dispersion effects can describe the π–π stacking. B3LYP-DCP and PBEh-3c yield the most, and B3LYP-D3-gCP the least accurate structure. The missing dispersion terms in plain B3LYP are less problematic for the hydrogen-bonded dimer because hydrogen bonds are mainly caused by electrostatic and induction interactions which most density functionals cover rather accurately. Thus, for the hydrogen-bonded dimer, plain B3LYP yields a geometry which is as good as with M06-2X and even slightly better than the B3LYP-D3-gCP structure. Again, B3LYP-DCP and PBEh-3c have the smallest RMSD compared with the reference.

Table 2 presents the statistical data of the S22, S66×8, and P26 test sets for small systems. The two best performing methods for each set are highlighted, and the data are converted to normal error distributions in Figure 10.

Table 2. Statistical data for the deviations of the intramolecular center-of-mass distances R_{CMA}[a] and the mean RMSDs.[b]

	S22			S66×8			S22	P26
	MD	MAD	MAX	MD	MAD	MAX	$\overline{\text{RMSD}}$	$\overline{\text{RMSD}}$
PBEh-3c	7.6	12.7	98.4	3.2	**5.7**	16.7	9.9	16.1
M06-2X	−7.7	**8.4**	−44.7	−9.5	9.6	−24.6	**6.5**	**8.4**
B3LYP	78.7	82.6	389.9	40.2	43.0	362.5	55.1	48.8
B3LYP-D3-gCP	4.6	11.2	43.8	10.1	10.2	28.0	8.2	17.5
B3LYP-DCP	−4.4	**5.7**	33.8	1.3	20.5	59.6	**3.9**	**9.4**
HF	81.8	82.2	263.2	54.2	54.2	362.5	53.2	49.6
HF-D3-gCP	8.6	15.6	135.4	1.7	**2.8**	f 18.4	11.7	12.7

[a] From the reference values for the S22 and S66×8 test set. [b] $\overline{\text{RMSD}}$ for the heavy atom positions in case of S22 and P26 compared with the reference. An MD > 0 denotes too large intermolecular distances. All values are given in pm. The two best performing methods are highlighted.

In general, plain B3LYP and HF cannot correctly reproduce the structures. Especially, when π–π stacking or nonpolar dispersion interactions are involved, these methods yield a practically unbound geometry or a different conformation, like the hydrogen bonded dimer in the example shown above. Thus, the mean (absolute) deviations for the intermolecular center-of-mass distances R_{CMA} for the S22 and S66×8 sets, as well as the heavy atom positions in the S22 and P26 sets, are unacceptably large. When the D3-gCP corrections are added, the MDs and MADs for the R_{CMA} of about 80 pm for S22 and about 40–50 pm for S66×8 drop significantly. B3LYP-D3-gCP yields an MD of only 5 pm and an MAD of 11 pm in case of S22 and the same values for both measures of 10 pm for the S66×8 set. HF-D3-gCP gives an MD of 9 pm and and MAD of 16 pm for the S22 set. For S66×8, HF-D3-gCP in fact is one of the two

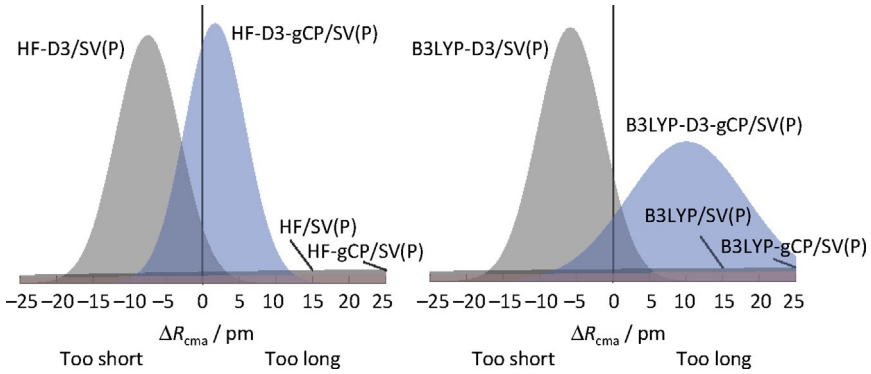


Figure 10. Error statistics of S66 equilibrium center-of-mass distances for corrected and uncorrected HF and B3LYP in a SV(P) basis set converted into normal error distributions.

best methods and yields an MD of only 2 pm and and MAD of only 3 pm. The RMSD for the heavy atom positions are similar for B3LYP-D3-gCP and HF-D3-gCP, about 10 pm for the S22 and about 15 pm for the P26.

B3LYP-DCP performs somewhat better than B3LYP-D3-gCP for S22 and gives the best results for this set with an MAD of −4 pm, an MAD 6 pm, and an RMSD value for the heavy atom positions of only 4 pm. In case of the S66×8 set, the MD of 1 pm is even lower, and the best value obtained but the MAD of 21 pm is much larger than those for all other dispersion-corrected methods. M06-2X yields one of the two best values for the heavy-atom RMSDs of 7 and 8 pm for the S22 and P26 set, respectively. It is the only method which consistently gives too small intermolecular center-of-mass distances R_{CMA} for both the S22 and S66×8 set. The MDs are −8 and −10 pm, respectively. The MADs are very similar, indicating that this error is systematic. PBEh-3c yields results comparable to HF-D3-gCP and slightly worse than M06-2x, with an MD of 8 pm, an MAD of 13 pm and an RMSD value of 10 pm for the S22 set. In case of the S66×8, the performance is better. The MD is 3 pm and the MAD is with 6 pm the second best value obtained.

Note that the S66 set consists of the most reliable reference data, while the S22 and P26 systems are only optimized at the MP2 level. A comparison of the S66 equilibrium CMA distances calculated at MP2 level reveals an MAD of 3.3 pm compared with the coupled cluster reference. Therefore, MAD values of a few pm on the S22 and P26 sets do not indicate significant deviations and are within the MP2 error. Even more important in this context are systematic errors in the reference structures. For instance some of the largest PBEh-3c outliers for S22 occur for the π-stacked benzene dimer, for which MP2 is known to overbind significantly.[93] Presumably, in this case the MP2 reference is in fact off (too short distance) as indicated by a distance underestimation by 3.4% compared to the CCSD(T) reference for the benezene dimer in the S66×8 set.

The influence of the dispersion and counterpoise correction schemes for the S66 equilibrium distances is summarized in Figure 10. In analogy to the binding energy analysis in the previous paragraph, we show HF and B3LYP deviations with and without correction converted into normal error distributions. These distributions mainly confirm the analysis given above. While the uncorrected methods yield rather bad structures, the most consistent methods are the dispersion and counterpoise corrected ones, though B3LYP-D3-gCP yields slightly too large intermolecular distances.

Among the dispersion-corrected methods B3LYP-DCP seems to be the best performer for all three test sets but we could not identify a method which is clearly superior to others. It is important to note, that due to the larger 6-31+G(2d,2p) basis set and the dispersion correcting potentials themselves, the geometry optimizations with B3LYP-DCP are an order of magnitude slower than with all other methods employed.

In order to show the influence and the interplay of dispersion and BSSE for supermolecular systems, we optimized complex **5** of the S30L and the **phe** complex of the L7 set with plain B3LYP, B3LYP-D3, B3LYP-gCP, and B3LYP-D3-gCP. The overlays of these geometries with the reference structure are presented in Figure 11.

As already observed for the small complexes, the plain B3LYP functional gives too large distances for the π–π stacked systems due to the dominant effect of missing dispersion. The RMSD of the heavy atom positions is 59 pm for **5** and 94 pm

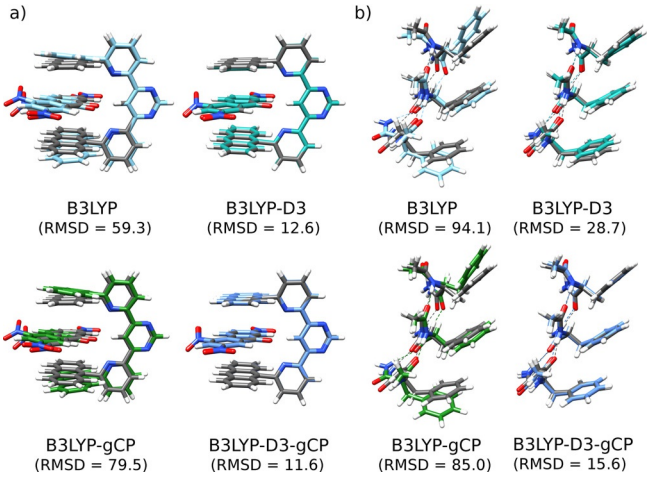


Figure 11. Comparison of the B3LYP, B3LYP-D3, B3LYP-gCP, and B3LYP-D3-gCP geometries (colored) with the reference TPSS-D3/def2-QZVP(-gf) structure (in gray) for a) complex **5** of the S30L set and b) the **phe** complex of the L7 set. The RMSDs for the heavy atoms are given in pm.

for **phe**. If the D3 scheme is employed, the distances are slightly too small due to the BSSE. The RMSDs drop to 13 and 29 pm, respectively. When we only correct for the BSSE by gCP, the distances are again far too large and the RMSD values are similar to those of plain B3LYP. Only if dispersion and BSSE corrections are employed together (B3LYP-D3-gCP level), accurate geometries are obtained. Visually, the agreement with the reference structure is very good, and the heavy-atom RMSDs are only 12 pm for **5** and 16 pm for **phe**. These two examples clearly show, that not only for interaction energies, but also for geometries of large systems, it is important to treat London dispersion and BSSE on the same footing.

Finally, we investigate the structures of the previously introduced X23 and ICE10 solid state benchmark sets. As noted before, we use the experimental crystal densities (or crystal volumes) from X-ray measurements. These mass densities have been back-corrected for zero-point and thermal effects, which are important as they can alter the mass density by 1 to 5% with a typically decreased density (enlarged unit cell volume). In Table 3, we give the statistical deviations from the reference unit cell volumes for the methods PBEh-3c, M06-2X, B3LYP, B3LYP-D3-gCP, and B3LYP-DCP all with the same basis sets as in the molecular calculations.

Table 3. Statistics (MD, MAD, SD, MAX)[a] for the relative deviations of the cell volume for the X23 and ICE10 sets.

	X23				ICE10			
	MD	MAD	SD	MAX	MD	MAD	SD	MAX
PBEh-3c	1.8	**2.7**	3.2	10.2	2.5	**5.0**	7.7	16.6
M06-2X[b]	−12.5	12.5	4.4	23.3	−14.9	14.9	2.2	17.3
B3LYP	22.1	22.1	15.5	57.3	−5.7	5.7	1.4	8.3
B3LYP-D3-gCP	5.5	7.6	6.4	14.7	7.6	8.3	5.5	15.8
B3LYP-DCP[b]	−3.5	**3.6**	2.1	7.5	−4.4	**4.4**	1.9	7.0

All values are given in %. The two best performing methods with smallest MAD are highlighted. [a] An MD > 0 denotes cell volume that is too large. [b] Only 70% of the systems could be converged.

The general picture that emerged from the molecular complexes is confirmed for the crystals. However, because of the larger long-range contributions to the interaction, the differences between the tested methods are more pronounced. Again, M06-2X and B3LYP-DCP are numerically problematic and suffer from SCF convergence problems. As already seen for the molecular dimers, M06-2X suffers from BSSE, which leads to systematically too small unit cells by 13% and 15% for the X23 and ICE10 set, respectively. For plain B3LYP, inconsistent behavior for the two test sets is found. For the more dispersion dominated X23 systems, the unit cells are substantially too large by more than 20%, though for some systems, the error compensation leads to better results than B3LYP is inherently capable. For instance, the geometries of the oxalic acid polymorphs are very reasonable with only 4% deviation from the reference density. The ice polymorphs are more strongly dominated by electrostatic and induction effects with only small dispersion contribution. Here, the BSSE is even larger compared with the missing dispersion leading to too small unit cells. Applying both correction schemes (D3 and gCP) results in a more consistent performance. At this level, for both test sets, a reasonable MAD of about 8% is obtained. A combined optimization of the gCP and D3 parameters would probably lead to even better geometries. B3LYP-DCP is based on a larger basis set with rather diffuse functions, which explains some of the convergence problems. However, this also minimizes the BSSE, and the results are good with MADs for the X23 and ICE10 reference unit cell slightly below and slightly above 4%, respectively. Again, especially the ice polymorphs are described to a high accuracy consistent with the excellent lattice energies. On the X23 set, B3LYP-DCP is only outperformed by the new PBEh-3c composite method. The geometries are competitive to more expensive calculations based on converged PAW basis sets with typical unit cell errors of about 3%.[39,94]

As prototypical example for a London-dispersion-dominated crystal, we investigate the benzene crystal in more detail. It has various energetically-close-lying polymorphs,[95,96] and it was used extensively to test and judge electronic structure methods (including wavefunction expansions,[97,98,99,100,101] dispersion corrected DFT,[89,82,102,103,94,104] and semiempirical MO methods[105,106,107]). We show a potential energy surface (PES) scan of the benzene crystal in Figure 12. Each structure corresponds to a constrained volume optimization at the TPSS-D3 level in a converged PAW[108,109] basis set, and we additionally highlight the equilibrium point. The reference point refers to the back-corrected experimental unit cell volume combined with a highly accurate CCSD(T) computed lattice energy.[113]

Because of SCF convergence problems, M06-2X results are not included. Concerning the other methods, substantial differences in the computational speed are observed. With identical numerical setups, the relative timing for one single-point energy calculation of PBEh-3c, B3LYP/SV(P), and B3LYP-DCP/6-31+G* are 1.0:1.2:8.2 with PBEh-3c being the fastest and B3LYP-DCP/6-31+G* the slowest method. The higher computational cost is mainly due to the larger and more diffuse basis set, which leads (especially in periodic boundaries) to a substantially higher number of computed integrals.

The benzene crystal nicely reflects the basic properties of the described methods regarding the treatment of dispersion dominated systems. Plain B3LYP just shows a shallow BSSE-related minimum. This correctly disappears when the gCP correction is applied. Only in combination with both correction schemes (B3LYP-D3-gCP), a very reasonable PES is obtained with nearly perfect lattice energy (−13.16 (B3LYP-D3-gCP) vs. −13.22 kcal mol^{-1} (reference)), but slightly too large unit cell (by 2.6% too low mass density). The B3LYP-DCP approach shows a clear minimum which is somewhat too low (lattice energy of −15.22 kcal mol^{-1}) and too small unit cell (by 4.5% too large mass density). The potential of PBEh-3c agrees excellently with the reference, and both the unit cell volume and the lattice energy are within 1.2% and 0.8 kcal mol^{-1}, respectively.

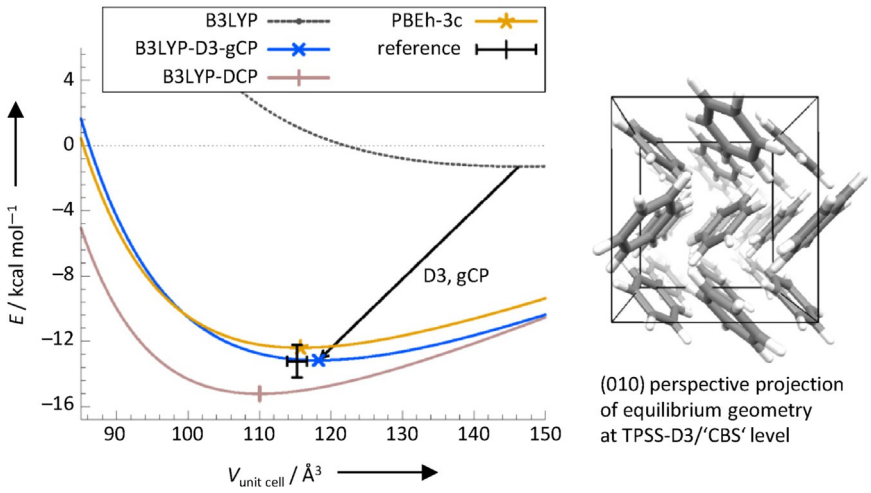


Figure 12. Lattice energy of the benzene crystal along a PES based on contained volume optimizations (TPSS-D3/'CBS' level). The experimental geometry is back-corrected for zero-point and thermal effects as described in refs. [85,39], and the reference lattice energy corresponds to a CCSD(T) estimate.[110]

5 Conclusion

In this short topical Review, we have critically analyzed widely used quantum chemical HF and DFT computations employing relatively small single-particle basis sets of double-zeta quality. As indicated by the tremendous number of publications which are based on this or similar theoretical levels, these methodologies are practically very relevant. We highlighted the two main error sources in standard applications, namely the BSSE and the missing London dispersion interaction. Different strategies to treat and correct the errors were reviewed and tested on mainly noncovalently bound systems with varying size. We analyzed both energetic and geometric properties. Due to the efficiency of the methods, their main applications are large supramolecular or periodic systems, which were also the focus of our analysis.

As main result of our investigations, it is nowadays not justified to rely on fortuitous error compensation as for example, in the popular B3LYP/6-31G* approach. Without additional computational overhead, the main error sources can be treated with semiclassical potentials, and the composite method B3LYP-D3-gCP/DZ outperforms the plain functional clearly. Further improved results are obtained, when the BSSE and London dispersion are directly included in the method design as recently done in the PBEh-3c functional with good to excellent results on all tested geometries and reference energies. Only for some systems with particular high requirements on the basis set (e.g. very strong hydrogen bonds or anions), the performance is slightly worse compared with, for example, a 'hybrid'/QZ level.

Using ECP-type potentials to simultaneously cure the functional and basis set errors works very well for small complexes similar to those used in the training sets for the method. However, with the tested B3LYP-DCP scheme, the computational costs are closer to that of a triple-zeta basis set. More importantly, the quality of the results for larger systems deteriorates and the performance is unsatisfying. We attribute this inconsistency to a wrong distance behavior of the correction potential as the finite ECP expansion cannot recover the correct R^{-6} limit of the dispersion interaction.

In summary, it is indeed possible to effectively use quantum chemical methods with small basis set expansions when all arising errors are treated properly. The good results for both, energetic and geometric properties of large and periodic systems is encouraging, and we expect this to translate into globally accurate potential energy surfaces, which is important for thermodynamic properties and ab initio molecular dynamics.

In this context we would like to mention the problem of solvation effects that was not discussed in this Review. Dispersion effects are omnipresent and also occur for any molecule when it is solvated as in most chemical applications. Molecular dispersion (or BSSE) effects are then partly quenched, that is, intramolecular contributions are replaced by intermolecular ones with the solvent. An accurate account of these effects requires sophisticated solvation models with the same high accuracy as the quantum chemical treatments which is difficult to obtain at present. Whenever comparisons of computed molecular to experimental liquid phase data are made, we recommend to include consistent continuum solvation models like COSMO-RS[111,112] or DCOSMO-RS[111] (for geometry optimizations). Only such treatments eventually will lead to the 'right answer for the right reason'. In any case, due to the broad area of possible applications of for example, the new PBEh-3c composite scheme in describing host–guest binding enthalpies, lattice enthalpies of organic crystals, and structures of larger biologically relevant molecules, the future for quantum chemical modeling of these systems seems bright.

6 Computational Details

For the single-point energy calculations on the benchmark sets S22,[76,93] S66×8,[34] WATER27,[86] L7,[88] S30L,[84] ICE10,[85] and X23[89,82] the geometries were taken as provided in the corresponding references. The computations for the molecular systems were carried out with either the current development version of ORCA

3.0[113,114] in case of B3LYP[9, 10, 11, 12]-DCP[53, 54]/6-31+G(2d2p)[13] or TURBOMOLE 7.0[115,116] for all other methods (HF, B3LYP, M06-2X,[40] PBE0,[117] and PBEh-3c[39]) in combination with the double-ζ basis set def2-SV(P)[118] (modified in case of PBEh-3c for boron to neon, for details see ref. [39]). The resolution-of-identity (RI) approximation for the Coulomb integrals[119] was applied in all cases except B3LYP-DCP using matching default auxiliary basis sets.[120] For the integration of the exchange-correlation contribution the numerical quadrature grids *m4* (*m5* in case of M06-2X)[121] and grid 6 were employed in TURBOMOLE and ORCA, respectively. The energy convergence criteria were set to 10^{-7} E_h in all cases.

The periodic calculations were conducted with a developer version of the CRYSTAL14 program.[122] It is the ideal choice for cost-effective DFT calculations in small basis sets as it can exploit full point and space group symmetry. The Brillouin zone is sampled with a Γ-centered *k*-mesh with grid density of approximately 0.025 $Å^{-1}$ (for details see refs. [85,82]). Standard integral thresholds and large DFT integration grids were used.

The calculation of the D3(BJ)[45,65] dispersion correction and the gCP BSSE correction[74] were carried out with our own programs dftd3 and gcp, respectively. These programs are freely available from our website.[123]

For the structure optimizations of the benchmark sets S22 and P26,[78] the geometries as provided in the corresponding references were taken as start coordinates. Again, for B3LYP-DCP/6-31+G(2d,2p), ORCA 3.0 was employed and TURBOMOLE 6.6 was used in for HF, B3LYP, M06-2X, PBE0, and PBEh-3c in combination with the def2-SV(P) basis set (modified in case of PBEh-3c). The D3(BJ) and gCP corrections to the gradients were again calculated with our own programs dftd3 and gcp. The convergence criteria were set to 10^{-7} E_h for energies and 10^{-5} E_h/Bohr for gradients.

Acknowledgements

This work was supported by the German Research Foundation (DFG) in the framework of the Gottfried Wilhelm Leibniz Prize.

Keywords: basis set superposition error · density functional theory (DFT) · double-zeta AO basis set · London dispersion energy · molecular complexes · molecular crystals

[1] P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *136*, B864.
[2] W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133.
[3] Y. Zhao, D. G. Truhlar, *Acc. Chem. Res.* **2008**, *41*, 157.
[4] R. Peverati, D. G. Truhlar, *Philos. Trans. R. Soc. London Ser. A* **2014**, *372*, 20120476.
[5] L. Goerigk, S. Grimme, *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670.
[6] M. Bühl, C. Reimann, D. A. Pantazis, T. Bredow, F. Neese, *J. Chem. Theory Comput.* **2008**, *4*, 1449.
[7] S. Grimme, M. Steinmetz, *Phys. Chem. Chem. Phys.* **2013**, *15*, 16031.
[8] T. Risthaus, M. Steinmetz, S. Grimme, *J. Comput. Chem.* **2014**, *35*, 1509.
[9] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.
[10] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.
[11] S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*, 1200.
[12] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623.
[13] W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257.
[14] Scifinder 2015, Chemical Abstracts Service: Columbus, OH, https://scifinder.cas.org (accessed June 2, **2015**).
[15] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, *72*, 650.
[16] S. L. Price, *Chem. Soc. Rev.* **2014**, *43*, 2098.
[17] M. A. Neumann, F. J. J. Leusen, J. Kendrick, *Angew. Chem. Int. Ed.* **2008**, *47*, 2427; *Angew. Chem.* **2008**, *120*, 2461.
[18] C. C. Pantelides, C. S. Adjiman, A. V. Kazantsev, *Top Curr. Chem.* **2014**, *345*, 25.
[19] H. Kruse, L. Goerigk, S. Grimme, *J. Org. Chem.* **2012**, *77*, 10824.
[20] S. Grimme, J. Antony, T. Schwabe, C. Mück-Lichtenfeld, *Org. Biomol. Chem.* **2007**, *5*, 741.
[21] E. R. Johnson, I. D. Mackie, G. A. DiLabio, *J. Phys. Org. Chem.* **2009**, *22*, 1127.
[22] K. Müller-Dethlefs, P. Hobza, *Chem. Rev.* **2000**, *100*, 143.
[23] L. A. Burns, A. Vázquez-Mayagoitia, B. G. Sumpter, C. D. Sherrill, *J. Chem. Phys.* **2011**, *134*, 084107.
[24] S. Grimme, *WIREs Comput. Mol. Sci.* **2011**, *1*, 211.
[25] J. Klimeš, A. Michaelides, *J. Chem. Phys.* **2012**, *137*, 120901.
[26] Y. Ikabata, H. Nakai, *Int. J. Quantum Chem.* **2015**, *115*, 309.
[27] J. P. Wagner, P. R. Schreiner, *Angew. Chem. Int. Ed.* **2015**, *54*, 12274–12296; *Angew. Chem.* **2015**, *127*, 12446–12471 .
[28] S. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553.
[29] L. M. Mentel, E. J. Baerends, *J. Chem. Theory Comput.* **2014**, *10*, 252.
[30] F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, J. H. van Lenthe, *Chem. Rev.* **1994**, *94*, 1873.
[31] M. Gutowski, G. ChaŃasiński, *J. Chem. Phys.* **1993**, *98*, 5540.
[32] J. Katriel, E. R. Davidson, *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 4403.
[33] C.-O. Almbladh, U. von Barth, *Phys. Rev. B* **1985**, *31*, 3231.
[34] J. Rezáč, K. E. Riley, P. Hobza, *J. Chem. Theory Comput.* **2011**, *7*, 2427.
[35] L. F. Holroyd, T. van Mourik, *Chem. Phys. Lett.* **2007**, *442*, 42.
[36] H. Valdés, V. Klusák, M. Pitonák, O. Exner, I. Starý, P. Hobza, L. Rulísek, *J. Comput. Chem.* **2008**, *29*, 861.
[37] T. van Mourik, P. G. Karamertzanis, S. L. Price, *J. Phys. Chem. A* **2006**, *110*, 8.
[38] T. Helgaker, J. Gauss, P. Jorgensen, J. Olsen, *J. Chem. Phys.* **1997**, *106*, 6430.
[39] S. Grimme, J. G. Brandenburg, C. Bannwarth, A. Hansen, *J. Chem. Phys.* **2015**, *143*, 054107.
[40] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215.
[41] Y. Zhao, D. G. Truhlar, *J. Chem. Theory Comput.* **2008**, *4*, 1849.
[42] O. A. Vydrov, T. Van Voorhis, *J. Chem. Phys.* **2010**, *133*, 244103.
[43] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, B. I. Lundqvist, *Phys. Rev. Lett.* **2004**, *92*, 246401.
[44] J. Klimeš, D. R. Bowler, A. Michaelides, *J. Phys. Condens. Matter* **2010**, *22*, 022201.
[45] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
[46] A. D. Becke, E. R. Johnson, *J. Chem. Phys.* **2005**, *123*, 154101.
[47] A. D. Becke, E. R. Johnson, *J. Chem. Phys.* **2007**, *127*, 154108.
[48] A. Tkatchenko, R. A. DiStasio, R. Car, M. Scheffler, *Phys. Rev. Lett.* **2012**, *108*, 236402.
[49] O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger, D. Sebastiani, *Phys. Rev. Lett.* **2004**, *93*, 153004.
[50] I. D. Mackie, G. A. Dilabio, *Phys. Chem. Chem. Phys.* **2010**, *12*, 6092.
[51] A. Heßelmann, G. Jansen, *Chem. Phys. Lett.* **2002**, *357*, 464.
[52] A. Heßelmann, G. Jansen, *Chem. Phys. Lett.* **2002**, *362*, 319.
[53] E. Torres, G. A. DiLabio, *J. Phys. Chem. Lett.* **2012**, *3*, 1738.
[54] G. A. DiLabio, M. Koleini, E. Torres, *Theor. Chem. Acc.* **2013**, *132*, 1389.
[55] O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger, D. Sebastiani, *Phys. Rev. B* **2005**, *71*, 195119.
[56] L. Goerigk, H. Kruse, S. Grimme, *ChemPhysChem* **2011**, *12*, 3421.
[57] J. C. Faver, M. L. Benson, X. He, B. P. Roberts, B. Wang, M. S. Marshall, M. R. Kennedy, C. D. Sherrill, K. M. Merz, Jr., *J. Chem. Theory Comput.* **2011**, *7*, 790.
[58] M. S. Marshall, L. A. Burns, C. D. Sherrill, *J. Chem. Phys.* **2011**, *135*, 194102.
[59] S. Grimme, *Chem. Eur. J.* **2012**, *18*, 9955.
[60] T. Risthaus, S. Grimme, *J. Chem. Theory Comput.* **2013**, *9*, 1580.
[61] L. Goerigk, *J. Chem. Theory Comput.* **2014**, *10*, 968.
[62] J. A. van Santen, G. A. DiLabio, *J. Phys. Chem. A* **2015**, *119*, 6703.
[63] J. P. Perdew, A. Zunger, *Phys. Rev. B* **1981**, *23*, 5048.
[64] B. Baumeier, P. Krüger, J. Pollmann, *Phys. Rev. B* **2006**, *73*, 195205.
[65] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456.
[66] E. R. Johnson, A. D. Becke, *J. Chem. Phys.* **2005**, *123*, 024101.
[67] B. M. Axilrod, E. Teller, *J. Chem. Phys.* **1943**, *11*, 299.

[68] Y. Muto, *Proc. Phys.-Math. Soc. Jpn.* **1943**, *17*, 629.
[69] R. A. DiStasio, V. V. Gobre, A. Tkatchenko, *J. Phys. Condens. Matter* **2014**, *26*, 213202.
[70] J. F. Dobson, *Int. J. Quantum Chem.* **2014**, *114*, 1157.
[71] M. R. Kennedy, A. Ringer McDonald, A. E. DePrince, M. S. Marshall, R. Podeszwa, C. D. Sherrill, *J. Chem. Phys.* **2014**, *140*, 121104.
[72] F. Jensen, *J. Chem. Theory Comput.* **2010**, *6*, 100.
[73] A. Galano, J. R. Alvarez-Idaboy, *J. Comput. Chem.* **2006**, *27*, 1203.
[74] H. Kruse, S. Grimme, *J. Chem. Phys.* **2012**, *136*, 154101.
[75] T. Yoshida, T. Hayashi, A. Mashima, H. Chuman, *Bioorg. Med. Chem. Lett.* **2015**, *25*, 4179.
[76] P. Jurečka, J. Šponer, J. Černý, P. Hobza, *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.
[77] L. Goerigk, J. R. Reimers, *J. Chem. Theory Comput.* **2013**, *9*, 3240.
[78] H. Valdes, K. Pluháčková, M. Pitonák, J. Rezác, P. Hobza, *Phys. Chem. Chem. Phys.* **2008**, *10*, 2747.
[79] H. J. Kulik, N. Luehr, I. S. Ufimtsev, T. J. Martinez, *J. Phys. Chem. B* **2012**, *116*, 12501.
[80] L. Goerigk, C. A. Collyer, J. R. Reimers, *J. Phys. Chem. B* **2014**, *118*, 14612.
[81] J. G. Brandenburg, M. Alessio, B. Civalleri, M. F. Peintinger, T. Bredow, S. Grimme, *J. Phys. Chem. A* **2013**, *117*, 9282.
[82] A. M. Reilly, A. Tkatchenko, *J. Chem. Phys.* **2013**, *139*, 024705.
[83] R. Sure, S. Grimme, *J. Comput. Chem.* **2013**, *34*, 1672.
[84] R. Sure, S. Grimme, *J. Chem. Theory Comput.* **2015**, *11*, 3785.
[85] J. G. Brandenburg, T. Maas, S. Grimme, *J. Chem. Phys.* **2015**, *142*, 124104.
[86] V. S. Bryantsev, M. S. Diallo, A. C. T. van Duin, W. A. Goddard, *J. Chem. Theory Comput.* **2009**, *5*, 1016.
[87] T. Anacker, J. Friedrich, *J. Comput. Chem.* **2014**, *35*, 634.
[88] R. Sedlak, T. Janowski, M. Pitoňák, J. Řezáč, P. Pulay, P. Hobza, *J. Chem. Theory Comput.* **2013**, *9*, 3364.
[89] A. Otero-de-la-Roza, E. R. Johnson, *J. Chem. Phys.* **2012**, *137*, 054103.
[90] H. Kruse, A. Mladek, K. Gkionis, A. Hansen, S. Grimme, J. Sponer, *J. Chem. Theory Comput.* **2015**, *11*, 4972.
[91] U. R. Fogueri, S. Kozuch, A. Karton, J. M. L. Martin, *J. Phys. Chem. A* **2013**, *117*, 2269.
[92] J. Witte, M. Goldey, J. B. Neaton, M. Head-Gordon, *J. Chem. Theory Comput.* **2015**, *11*, 1481.
[93] T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall, C. D. Sherrill, *J. Chem. Phys.* **2010**, *132*, 144104.
[94] J. G. Brandenburg, S. Grimme, *Top. Curr. Chem.* **2014**, 345, 1.
[95] S. Block, C. E. Weir, G. J. Piermarini, *Science* **1970**, *169*, 586.
[96] B. P. van Eijck, A. L. Spek, W. T. M. Mooij, J. Kroon, *Acta Crystallogr. Sect. B* **1998**, *54*, 291.
[97] W. B. Schweizer, J. D. Dunitz, *J. Chem. Theory Comput.* **2006**, *2*, 288.
[98] A. Ringer, C. Sherrill, *Chem. Eur. J.* **2008**, *14*, 2542.
[99] O. Bludský, M. Rubeš, P. Soldán, *Phys. Rev. B* **2008**, *77*, 092103.
[100] S. Wen, G. J. O. Beran, *J. Chem. Theory Comput.* **2011**, *7*, 3733.
[101] R. Podeszwa, B. M. Rice, K. Szalewicz, *Phys. Rev. Lett.* **2008**, *101*, 115503.
[102] D. J. Carter, A. L. Rohl, *J. Chem. Theory Comput.* **2014**, *10*, 3423.
[103] G. J. O. Beran, K. Nanda, *J. Phys. Chem. Lett.* **2010**, *1*, 3480.
[104] J. Moellmann, S. Grimme, *J. Phys. Chem. C* **2014**, *118*, 7615.
[105] G. Raabe, *Z. Naturforsch.* **2004**, *59a*, 609.
[106] J. G. Brandenburg, S. Grimme, *J. Phys. Chem. Lett.* **2014**, *5*, 1785.
[107] J. G. Brandenburg, M. Hochheim, T. Bredow, S. Grimme, *J. Phys. Chem. Lett.* **2014**, *5*, 4275.
[108] P. E. Blöchl, *Phys. Rev. B* **1994**, *50*, 17953.
[109] G. Kresse, J. Joubert, *Phys. Rev. B* **1999**, *59*, 1758.
[110] J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schutz, G. K.-L. Chan, *Science* **2014**, *345*, 640.
[111] A. Klamt, *J. Phys. Chem.* **1995**, *99*, 2224.
[112] A. Klamt, *WIREs Comput. Mol. Sci.* **2011**, *1*, 699.
[113] F. Neese, *"ORCA: An ab initio, density functional and semiempirical program package"*, Version 3.0 (Current Development Version), Max Planck Institute for Bioinorganic Chemistry, Mülheim an der Ruhr (Germany), **2014**.
[114] F. Neese, *WIREs Comput. Mol. Sci.* **2012**, *2*, 73.
[115] R Ahlrichs, F. Furche, C. Hättig, W. Klopper, M. Sierka, F. Weigend, TURBOMOLE 7.0, Universität Karlsruhe, Karlsruhe (Germany), **2015**; See also: http://www.turbomole.com.
[116] F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, F. Weigend, *WIREs Comput. Mol. Sci.* **2014**, *4*, 91.
[117] C. Adamo, V. Barone, *J. Chem. Phys.* **1999**, *110*, 6158.
[118] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
[119] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs, *Chem. Phys. Lett.* **1995**, *242*, 652.
[120] F. Weigend, *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057.
[121] O. Treutler, R. Ahlrichs, *J. Chem. Phys.* **1995**, *102*, 346.
[122] 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, B. Kirtman, *Int. J. Quantum Chem.* **2014**, *114*, 1287.
[123] See http://www.thch.uni-bonn.de/tc/?section=downloads&lang=english.

Received: October 5, 2015
Published online on November 25, 2015

