Ψ_k **Scientific Highlight Of The Month**

No. 144 **January 2019**

Simplified DFT methods for consistent structures and energies of large systems

Eike Caldeweyher[1] and Jan Gerit Brandenburg[2,3,*]

[1] *Mulliken Center for Theoretical Chemistry, Institute for Physical and Theoretical Chemistry, University of Bonn, Beringstr. 4, 53115 Bonn, Germany*

[2] *Thomas Young Center, University College London, Gower Street, London WC1E 6BT, United Kingdom*

[3] *Institute for Physical Chemistry, University of Göttingen, Tammannstr. 6, 37077 Göttingen, Germany*

This Psi-k Scientific Highlight is a reformatted preprint version of *E. Caldeweyher and J. G. Brandenburg 2018 J. Phys.: Condens. Matter 30 213001* and is reproduced with permission by the Institute of Physics.

Abstract

Kohn-Sham density functional theory (DFT) is routinely used for the fast electronic structure computation of large systems and will most likely continue to be the method of choice for the generation of reliable geometries in the foreseeable future. Here, we present a hierarchy of simplified DFT methods designed for consistent structures and non-covalent interactions of large systems with particular focus on molecular crystals. The covered methods are a minimal basis set Hartree-Fock (HF-3c), a small basis set screened exchange hybrid functional (HSE-3c), and a generalized gradient approximated functional evaluated in a medium-sized basis set (B97-3c), all augmented with semi-classical correction potentials. We give an overview on the methods design, a comprehensive evaluation on established benchmark sets for geometries and lattice energies of molecular crystals, and highlight some realistic applications on large organic crystals with several hundreds of atoms in the primitive unit cell.

*gerit.brandenburg@thch.uni-bonn.de

1 Introduction

Computational materials science is a dynamic area of modern scientific research. Approaches that are based on the fundamental laws of quantum mechanics are now integral to almost any materials design initiative in academia and industry [1, 2]. A typical materials discovery workflow (see sketch in Fig. 1) starts with a set of requirements on the properties of a material, which could be a combination of stability, and mechanical, or electronic features. However, the possible materials space is huge and the task of theoretical models is in reducing this space to a smaller number of promising candidates. Our ability to do these predictions are regularly

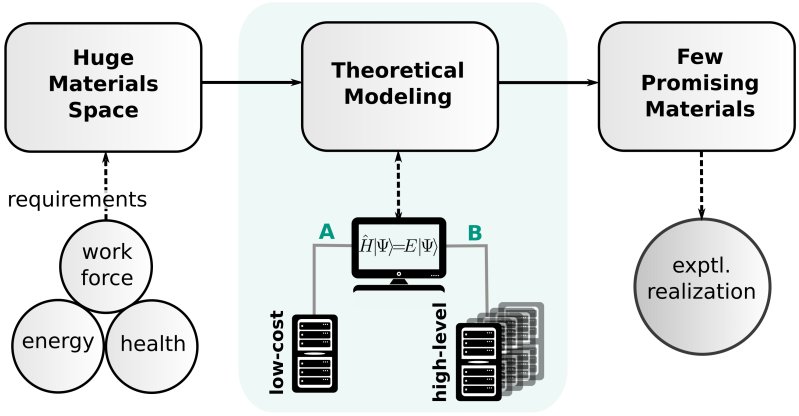

Figure 1: Sketch of a possible materials discovery work-flow. The compromise between high accuracy and affordable computational cost in solving the Schrödinger equation is shown.

tested by blind competitions like the statistical assessment of the modeling of proteins and ligands (SAMPL [3, 4]) and crystal structure prediction of organic compounds (CSP [5–7]). While the static Schrödinger equation is a formally exact description, its direct solution is prohibitive for many-particle systems. Thus, approximations to either the Hamiltonian or the many-particle wavefunction ansatz are needed. Depending on the required accuracy and the available computational resources, different electronic structure methods are needed ranging from fast "low-cost" to more involved high-level methods. Exciting progress has been made in the field of high-level methods (variant B in Fig. 1). On the one hand, embedding techniques [8–10] and local approaches of coupled cluster theories [11–13] made the gold standard of quantum chemistry applicable to molecular systems with a few hundred atoms [14] and molecular crystals of small molecules [15, 16]. On the other hand, new algorithmic developments in the field of quantum Monte-Carlo lead to substantial computational savings [17], enabling the calculation of lattice energies of a range of molecular crystals within reasonable computational effort [18]. Another approach, which is less accurate but more affordable is the random phase approximation (RPA) [19–23] that has recently shown good results, in particular if singles corrections are introduced [24–26]. While RPA is not considered a benchmark type method, it has the advantage of lower computational cost. Specifically combined with efficient implementations of analytical nuclear gradients, this enabled the geometry optimization of large complexes and even molecular dynamics applications [27, 28]. When numerical thresholds are pushed to the limit, the different benchmark type methods have been shown to agree within sub-chemical accuracy, i.e. within

1 kcal/mol or 43 meV [29].

However, aiming at increasing system size and large-scale screening of many compounds, more efficient methods that typically start from an effective one-particle theory are needed (variant A in Fig. 1). Even with growing computational facilities, the routine application of electronic structure methods to large systems requires efficient methodologies. Here, Kohn-Sham density functional theory is often the method of choice due to its excellent accuracy to computational cost ratio [30–32]. Recent developments of semi-local density functional approximations combine exact constraints with various degrees of parametrization in order to improve the short- to medium range electron correlation [33–35]. Another substantial development in the last decade is in the incorporation of London dispersion interactions (also known as the attractive part of van der Waals forces [36]) in the DFT framework [37–39]. This enabled the large scale application of electronic structure methods for large non-covalently bound systems important for host-guest complexes, protein-ligand binding energies, molecular crystals, porous cages, and nanostructured materials. A comprehensive benchmark for main-group thermochemistry, kinetics, and non-covalent interactions (GMTKN55) demonstrated that the improved density functionals perform excellently for these energetic properties [40]. Indeed, the dispersion corrected SCAN-D3 functional [41] is the best performing non-hybrid method in most test sets. But considering the improvement of high-level methods enabling affordable single-point energies, the major strength of the low-cost methods will be in providing reliable geometries, perform conformational samplings, and running molecular dynamics. With this in mind, we and other groups have presented a couple of low-cost methods in the past years employing similar strategies. One central idea, which has a long tradition in the quantum-chemistry community, is to use the error compensation between a small atom-centered single-particle basis set and an incomplete treatment of electron correlation [42–46]. We have been working on a hierarchy of simplified DFT methods that were specifically designed for consistent structures and interactions of large systems like molecular crystals. The covered methods are a minimal basis set Hartree-Fock (HF-3c [47]), a small basis set global hybrid functional (PBEh-3c [48]) and its screened exchange variant (HSE-3c [49]), and a generalized gradient approximated functional evaluated in a medium-sized basis set (B97-3c [50]), all augmented with semi-classical correction potentials. As other groups have followed with similar strategies [51–54], the methods are implemented in several program packages, and are well received by the chemical physics community, it seems appropriate to give an overview in this Topical Review. Most methods discussed are based on a DFT description in specialized single-particle expansions. We give a brief (educational) background of the theoretical framework in section 2.1, summarize the treatment of London dispersion interactions with semi-classical potentials in section 2.2, and explain the construction of the "3c" hierarchy of simplified DFT methods. In section 3, we compare the described methods based on established molecular crystal benchmark sets and highlight a few selected realistic applications to large organic crystals with several hundreds of atoms in the primitive unit cell. A future perspective is given in section 4.

2 Methodologies

2.1 Theoretical framework

Density functional theory (DFT) is based on the inspiring work of Hohenberg and Kohn [55]. In the following the electronic energy is written as a functional of the electron density $n(r)$

$$E[n(r)] = T[n(r)] + V_{ne}[n(r)] + V_{ee}[n(r)] = \int n(r)\, v(r)\, dr + F[n(r)] \tag{1}$$

where $T[n(r)]$ is the kinetic energy, $V_{ne}[n(r)]$ is the electron-nucleus, $V_{ee}[n(r)]$ is the electron-electron interaction energy, and $F[n(r)]$ is a universal functional of $n(r)$. The first Hohenberg-Kohn theorem states that any approximate density $\tilde{n}(r)$ determines its own external potential $\tilde{v}(r)$ and hence its own wave function $\tilde{\Psi}$. The application of the standard variational principle gives, therefore,

$$\left\langle \tilde{\Psi} | \hat{H} | \tilde{\Psi} \right\rangle = \int \tilde{n}(r)\, \tilde{v}(r)\, dr + F[\tilde{n}(r)] = E[\tilde{n}(r)] \geq E[n(r)]. \tag{2}$$

Thus, the exact electron density minimizes the exact energy expression and is accessible through a direct minimization of the electronic energy with respect to density variations. In Kohn-Sham (KS) theory the total energy expression is given as a sum of the single-particle kinetic energy contributions expressed in terms of orbitals,

$$T_s[n(r)] \equiv T_s[\phi_i] = -\frac{1}{2} \sum_i \int dr\, \phi_i^*(r) \nabla^2 \phi_i(r) \tag{3}$$

the Hartree (Coulomb) repulsion of the density with itself,

$$J[n(r)] = \frac{1}{2} \int dr \int dr'\, \frac{n(r)\, n(r')}{|\, r\, -\, r'\,|} \tag{4}$$

and the so-called exchange-correlation (XC) functional $E_{XC}[n(r)]$

$$F[n(r)] = T_s[n(r)] + J[n(r)] + E_{XC}[n(r)]. \tag{5}$$

The XC functional is defined as

$$E_{XC}[n(r)] = T[n(r)] - T_s[n(r)] + V_{ee}[n(r)] - J[n(r)]. \tag{6}$$

The KS equations describe non-interacting particles ($T = T_s$ and $V_{ee} = 0$) moving in an external potential $v_{eff}(r)$, defined by

$$v_{eff}(r) = v(r) + \frac{\delta\, J[n(r)]}{\delta\, n(r)} + \frac{\delta\, V_{ee}[n(r)]}{\delta\, n(r)}. \tag{7}$$

The Hamiltonian for a system with non-interacting electrons moving in an external potenital $v_{eff}(r)$ is simply given by

$$\hat{H} = -\sum_i \frac{1}{2} \nabla_i^2 + \sum_i v_{eff}(r_i). \tag{8}$$

This operator is separable and can, therefore, be expressed *via* an anti-symmetrized product of single-particle eigenfunctions ϕ_i (KS orbitals), the solutions of a coupled set of non-linear single-particle equations

$$\left(-\frac{1}{2}\, \nabla^2 + v_{eff}(r) \right)\, \phi_i(r) = \varepsilon_i\, \phi_i(r). \tag{9}$$

DFT is exact provided that the XC energy is exact. In practice, the XC energy has to be approximated, which is regularly classified according to the Jacob's ladder hierarchy [56]. The simplest approximation assumes a (locally) uniform electron density and is thus dubbed local spin density approximation (LDA). In the frame of this approximation, the expression for the XC energy is given by

$$E_{XC}[n(r)] = \int n(r)\varepsilon_{XC}^{LDA}[n(r)]\,dr. \tag{10}$$

Here, ε_{XC}^{LDA} is the exchange-correlation energy density per particle of an electron gas with uniform spin densities [57]. LDA is still used in the solid state community with recent extensions to finite temperature free energies [58]. While extended metallic systems can be described reasonably well by LDA, typical molecular systems require inclusion of the density gradient as in the generalized gradient approximation (GGA).

$$E_{XC}^{GGA} = \int \varepsilon_{XC}^{LDA}[n(r)]\, g\left(n(r), \nabla\, n(r)\right)\, dr \tag{11}$$

Different expressions are possible for the so-called enhancement factor g. The most prominent GGAs are the Perdew-Burke-Enzerhof (PBE) exchange and correlation functionals [59] and the Becke exchange (B88 [60]) combined with the Lee-Yang-Parr (LYP) correlation functional [61]. A natural extension to GGAs is to use higher-order derivatives of the electron density or other semilocally-available information, leading to the meta-GGA class. A typically employed variable is the KS kinetic energy density $\tau = \frac{1}{2}\sum_i |\nabla\phi_i|^2$. Popular meta-GGAs are the Tao-Perdew-Staroverov-Scuseria (TPSS) functional [62] and the Minnesota functionals M06L [63], M11L [64], and MN12L [65] by Truhlar and coworkers. A recently introduced empirical meta-GGA with a smoothness constraint and a VV10 long-range dispersion correction (long-range dispersion effects see below), B97M-V, was presented by Mardirossian and Head-Gordon [66]. Constraint-satisfaction based meta-GGA functionals have gained more attention [67–70] with a most recent development being the strongly constrained and appropriately normed (SCAN) functional [33]. Driven from the fact that LDA and GGAs suffer from self-interaction errors (integer discontinuity [71]) techniques have been developed for constructing functionals which combine a fractional amount a_X of non-local (one-determinantal) Hartree-Fock exchange with local XC functionals. Such hybrid functionals have reduced self-interaction error which drastically improved, e.g., the description of band-gaps of periodic materials, thermochemistry, and kinetics of chemical reactions. These hybrid DFAs were originally introduced by Becke and are motivated by the adiabatic connection [72]. The hybrid energy expression is then given as

$$E_{XC}^{hybrid} = (1 - a_X)\, E_X^{GGA} + a_X E_X^{HF} + E_C^{GGA}. \tag{12}$$

In principle, any semi-local XC component can be combined with Fock exchange, popular ones are PBE0 [73] and B3LYP [74, 75]. Heyd and co-workers decomposed the Coulomb operator into short- (SR) and long-range (LR) contributions of the form

$$\frac{1}{r} = \frac{\mathrm{erf}\,(\omega\, r)}{r} + \frac{\mathrm{erfc}\,(\omega\, r)}{r} \tag{13}$$

where $\mathrm{erf}(x)$ is the normal error function, $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$, and ω is an adjustable range separation parameter [76]. Those range-separated hybrids (rsh) have the energy expression

$$\begin{aligned} E_{XC}^{rsh} = \quad & (1 - a_X)\, E_X^{GGA,SR}(\omega) + a_X E_X^{HF,SR}(\omega) \\ & + E_X^{GGA,LR} + E_C^{GGA} \end{aligned} \tag{14}$$

In molecular calculations especially long-range corrected functionals are widely used in the calculation of excited states by means of time-dependent DFT because the orbital energies obtained with them are much more amenable for such calculations [77–81]. In the procedure of solving the Kohn-Sham equations in systems with translational invariance, several simplifications can be applied due to symmetry. The effective potential, which enters the Kohn-Sham Hamiltonian, is periodic with respect to the translation vector T

$$v_{eff}(r) = v_{eff}(r+T) \tag{15}$$

where T is a Bravais lattice vector of the solid. This periodic boundary condition (PBC) leads to Bloch's theorem which states that the periodicity of the bulk material constrains the one-electron wave function to obey

$$\phi_{i,k}(r+T) = \exp(i\,k\,T)\exp(i\,k\,r)\,u_{i,k}(r) = \exp(i\,k\,T)\,\phi_{i,k}(r) \tag{16}$$

where k is the vector of reciprocal space and $\phi_{i,k}(r)$ as product of a lattice-periodic Bloch function $u_{i,k}(r)$ and a single-particle basis set. Translational symmetry now ensures that one has to consider only k-vectors which lie inside the first Brillouin zone [82] when solving the Kohn-Sham equations under PBCs. The integration in the first Brillouin zone is then replaced by an integration over a k-point mesh (special point theorem [83]). Different expansions for the single-particle basis sets are possible. One possibility is to solve the Kohn-Sham equations directly by using a grid where functions are represented by their value over a set of points in real space [84]. Furthermore, numerically tabulated atom-centered orbitals (NAOs) are well known in the literature allowing the creation of optimized element-dependent basis sets that are compact as well as accurate in production calculations with respect to total energy convergence [85]. Conceptually somehow different are the so-called Daubechies wavelets which have the characteristic that they form an orthogonal and smooth basis set, localized both in real and Fourier spaces [86]. Another possibility in generating orbitals relies on an expansion in Slater type orbitals (STOs) where integrals may be calculated numerically [87] or in a mixed scheme with analytical and recursion/expansion-based evaluations [88]. Most quantum chemical codes use Gaussian type orbitals (GTOs) to solve integrals analytically [89–91] while others use a combinations of GTOs with projector augmented-wave (PAW) methods [92, 93]. Especially the PAW ansatz is often used in material science [94, 95] and will thus be discussed in more detail. In the PAW method the all-electron (AE) wave function Ψ_N is transformed into the so-called pseudo Hilbert space (PS) by means of a linear transformation

$$|\Psi_N\rangle = |\tilde{\Psi}_N\rangle + \sum_i \left(|\phi_i\rangle - |\tilde{\phi}_i\rangle\right)\langle\tilde{p}_i|\tilde{\Psi}_N\rangle. \tag{17}$$

Here, the PS wave functions $\tilde{\Psi}_N$ are the variational quantities and i is a shorthand for the atomic site. The AE partial waves ϕ_i are obtained for a reference atom whereas the PS partial waves $\tilde{\phi}_i$ are equivalent to the AE partial waves outside a defined radius [96]. The projector functions $\tilde{p}_i$ are dual to the partial waves

$$\langle\tilde{p}_i|\tilde{\phi}_j\rangle = \delta_{ij}\,. \tag{18}$$

By incorporating equation 17, one is able to show that in the PAW method the AE charge density is given by

$$n(r) = \tilde{n}(r) + n^1(r) - \tilde{n}^1(r), \tag{19}$$

where $\tilde{n}(r)$ is the soft pseudo-charge-density calculated directly from the pseudo-wave-functions on a plane-wave grid

$$\tilde{n}(r) = \sum_N f_N \langle \tilde{\Psi}_N | r \rangle \langle r | \tilde{\Psi}_N \rangle. \tag{20}$$

Here, f_N denotes the occupation number. The onsite charge densities $n^1(r)$ and $\tilde{n}^1(r)$ are treated on a radial support grid [97–101]. Both charge densities are defined as

$$n^1(r) = \sum_{(i,j)} \rho_{ij} \langle \phi_i | r \rangle \langle r | \phi_j \rangle \tag{21}$$

and

$$\tilde{n}^1(r) = \sum_{(i,j)} \rho_{ij} \langle \tilde{\phi}_i | r \rangle \langle r | \tilde{\phi}_j \rangle \tag{22}$$

where ρ_{ij} are the occupancies of each augmentation channel (i, j) which are calculated from the pseudo-wave-functions applying the projector functions

$$\rho_{ij} = \sum_N f_N \langle \tilde{\Psi}_N | \tilde{p}_i \rangle \langle \tilde{p}_j | \tilde{\Psi}_N \rangle. \tag{23}$$

It is common to expand only those plane-waves-functions which exhibit small kinetic energies. Hence, the plane-wave basis set can be truncated to include only plane-wave-functions within a particular cutoff energy. The truncation of the basis set at a finite cutoff energy will lead to errors in the computation of the energy and its derivatives. To minimize this error in a systematic way, it is recommended to increase the value of the cutoff energy until the calculated total energy converges within a required tolerance. For states with metallic characteristic and small fluctuations in the charge density, very few planes-waves are sufficient. In contrast, systems with rather localized electron density as in molecular crystals, large plane-wave basis sets are needed which make the DFT calculation rather costly [102]. In particular, the convergence of unit cells in free optimizations requires a very accurate stress tensor due to the shallow potential energy surface.

In these situations, atom centered functions like GTOs can significantly reduce the number of basis functions. This approach is based on the expansion of orbitals in Bloch sums

$$\phi_{i,k}(r) = \sum_\mu C_{\mu i}(k) \phi_{\mu,k}(r) \quad \mu = 1, \ldots, M \tag{24}$$

where μ labels all atomic orbitals (AOs) in the reference primitive cell and the Bloch functions are created from atom-centered basis functions χ_μ

$$\phi_{\mu,k}(r) = \sum_T \exp(i\, k\, T)\ \chi_\mu (r - A_\mu - T)\,. \tag{25}$$

Gaussian-type orbitals benefit from a fast convergence behavior for describing the core electrons due to strong localization. However, in contrast to plane-waves small AO basis sets strongly suffer from basis set incompleteness errors, especially the basis set superposition error (BSSE). Already semi-diffuse AOs can exhibit near linear dependencies in periodic calculations which directly excludes the reduction of the BSSE by a systematic increase of the AO basis. To overcome this problem, a general geometrical procedure has been developed in 2012 which

corrects for the BSSE in a semi-empirical way [103]. This geometrical counterpoise correction (gCP) has been further extended to be applicable to periodic systems in 2013 [104]

$$E_{BSSE}^{gCP} = \frac{\sigma}{2} \sum_{A} \sum_{A \neq B} V_A^{gCP} \left(|| \, r_A - r_B + T \, || \right) \tag{26}$$

with

$$V_A^{gCP} \left(R_{AB} \right) = \, E_{miss}^{A} \, \frac{\exp\left(-\alpha \, R_{AB}^{\beta} \right)}{\sqrt{S_{AB} \, N_B^{virt}}}. \tag{27}$$

The difference in atomic energy between a large (nearly complete) basis set and the target basis set for each free atom, E_{miss}^{A}, is used as a measure to generate the repulsive potential V_A^{gCP}. Here, 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 set. 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. The gCP parameter were obtained by least-squares minimization against counterpoise correction data obtained by the scheme of Boys and Bernadi (BB-CP [105]). Several basis sets have been parametrized and it could be shown that the gCP correction removes most of the BSSE and is able to closely reproduce BB-CP corrected curves for, e.g., lattice energies of molecular crystals [104].

2.2 London dispersion interaction

Mean field electronic structure methods like Hartree-Fock (HF) or semi-local DFT are widely used in computational chemistry and physics. These methods do not describe long-range electronic correlation effects, and hence they cannot account for so-called London dispersion effects. Nonetheless, such interactions are mandatory for describing the chemistry or physics of large or condensed-phase systems in an accurate and asymptotically correct way. The exact expression for treating long-range correlation effects is accessible by introducing the adiabatic connection fluctuation dissipation theorem [106] which uses response functions to express the effect of an external perturbation (Coulomb interaction scaled by λ) acting on the electron density.

$$E_c = -\frac{1}{2} \int_0^1 d\lambda \int dr dr' \, \frac{1}{|\, r \, - \, r' \,|} \int_0^\infty d\omega \left[\chi_\lambda(r, r', i\omega) \, - \, \chi_0(r, r', i\omega) \right] \tag{28}$$

The induced density and the external potential are related through a position- and frequency-dependent charge density susceptibility

$$\chi_0(r, r', i\omega) = -4 \sum_i \sum_a \, \frac{\omega_{ai}}{\omega_{ai}^2 + \omega^2} \, \phi_i(r) \phi_a(r) \phi_a(r') \phi_i(r'). \tag{29}$$

Unfortunately, most systems can not be treated exactly due to increasing computational costs with increasing system size. In order to circumvent this obstacle, semi-classical methods were developed to enable the computation of London dispersion interactions [107] even for large system sizes. The computationally efficient semi-classical D3 scheme requires only the geometry as input to calculate the inter- and intra-molecular dispersion energy based on pre-calculated time-dependent density functional theory (TD-DFT) data. Within the D3 method the Coulomb

operator is expanded into multipoles where a coarse-grain partitioning to atomic polarizabilities at imaginary frequency enables the calculation of interatomic dipole-dipole dispersion coefficients [108]. The exact Casimir-Polder equation gives access to such pairwise dispersion coefficients

$$C_6^{AB} = \int_0^\infty d\omega\, \alpha^A(i\omega)\alpha^B(i\omega). \tag{30}$$

Aside from the pure geometrical D3 dispersion model, electron density dependent approaches like the Tkatchenko-Scheffler [109] (TS) model with its many-body analogon [110] (MBD@scsTS), the exchange dipole moment [111–113] (XDM) model of Becke and Johnson, or the local-response dispersion [114, 115] (LRD) model by Sato and Nakai exist. Somewhat different approaches that also rely on a local description of the response function are the non-local density functional based dispersion corrections for which the well established family of van-der-Waals density functionals [116–120] (vdw-DFs) and VV10 [121, 122] are appreciable examples. In 2010 the DFT-D approach dubbed D3 [123] has been proposed. In distinction to earlier proposed approaches, the molecular environment is explicitly taken into account by the empirical concept of fractional coordination numbers (CNs) which are constructed by a pairwise sum. Through the concept of fractional coordination numbers, different hybridization conditions are represented for each element which is in agreement with chemical intuition. In this way the chemical environment is taken into account only *via* the molecular geometry. System specific dispersion coefficients are then derived, by taking the electrical dipole polarizability $\alpha(i\omega)$ for differently coordinated symmetric hydrides (A_mH_n and B_kH_l) into account. For all elements with $Z \leq 94$, several of such reference hydrides have been treated non-empirically by TD-DFT calculations with a variant of the PBE0 hybrid functional with a modified amount of non-local exchange ($a_X = 37.5\%$). A modified Casimir-Polder equation [124] has been used to calculate atom pairwise dispersion coefficients $C_{6,ref}^{AB}(CN^A, CN^B)$ for atoms A and B in these reference systems

$$\begin{aligned} C_{6,ref}^{AB}(CN^A, CN^B) = \; & \tfrac{3}{\pi} \int_0^\infty \tfrac{1}{m} \left[\alpha^{A_mH_n}(i\omega) - \tfrac{n}{2}\,\alpha^{H_2}(i\omega)\right] \\ & \times \tfrac{1}{k} \left[\alpha^{B_kH_l}(i\omega) - \tfrac{l}{2}\,\alpha^{H_2}(i\omega)\right] d\omega. \end{aligned} \tag{31}$$

Here, A_mH_n and B_kH_l are the reference systems for A and B with the corresponding coordination number CN^A and CN^B. These reference systems describe different bonding situations and are distinguished by their CN. In D3 every possible element and hybridization combination is precalculated and stored. During a calculation of a target system, the atom pairwise $C_{6,ref}^{AB}(CN^A, CN^B)$ dispersion coefficients are interpolated *via* a Gaussian weighting function to generate system specific $C_6^{AB}(CN^A, CN^B)$ dispersion coefficients as follows

$$C_6^{AB}(CN^A, CN^B) = \frac{\sum_i^{N_A} \sum_j^{N_B} C_{6,ref}^{AB}(CN^A, CN^B)\, L_{ij}}{\sum_i^{N_A} \sum_j^{N_B} L_{ij}} \tag{32}$$

with

$$L_{ij} = \exp\left[-4\left((CN^A - CN_i^A)^2 + (CN^B - CN_j^B)^2\right)\right]. \tag{33}$$

This strategy ensures an efficient calculation of accurate system specific dispersion coefficients simply by geometrical means (cf. Fig. 2). For clarity, the lengthy notation $C_6^{AB}(CN^A, CN^B)$ is

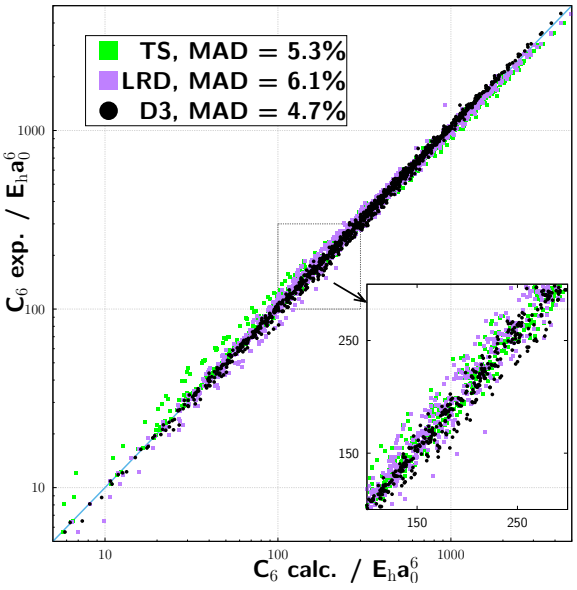

Figure 2: Comparison of experimentally derived molecular dispersion coef- ficients (from dipole oscillator strength distributions) with calculated ones (1225 cases, compilation by A. Tkatchenko). Note the logarithmic scale and variation of the coefficients over three orders of magnitude. The test set contains small to medium-sized, inorganic and organic molecules (H_2–C_8H_{18}).

abbreviated in favor of C_6^{AB}. The atom pairwise dipole-quadrupole interaction $E_{disp}^{(8)}$ is included with the respective C_8^{AB} dispersion coefficient, which is generated in D3 approximately from recursion relations. Higher order multi-polar terms are neglected in D3 but effectively absorbed in the s_8 scaling factor into the $1/R^8$ term. The atom pairwise two-body dispersion energy is than calculated as

$$E_{disp}^{D3} = -\frac{1}{2} \sum_{AB} \sum_{n=6,8} s_n \frac{C_n^{AB}}{R^n} f_{damp}^{(n)}(R_0^{AB}). \tag{34}$$

Different approaches are available for choosing $f_{damp}^{(n)}(R^{AB})$ where R^{AB} is an inter-atomic distance for atom pair AB. The rational Becke-Johnson damping function has become the default in combination with D3:

$$f_{damp,BJ}^{(n)}(R_0^{AB}) = \frac{R_{AB}^n}{R_{AB}^n + \left(a_1 R_0^{AB} + a_2\right)^n} \tag{35}$$

The scheme incorporates functional-specific parameter a_1 and a_2 as well as the radii $R_0^{AB} = \sqrt{C_8^{AB}/C_6^{AB}}$. The leading non additive dispersion contribution dubbed Axilrod-Teller-Muto [125, 126] (ATM) term is defined as derived from third-order perturbation theory as

$$E^{ABC} = \frac{C_9^{ABC}\,(3\cos\theta_a\,\cos\theta_b\,\cos\theta_c + 1)}{(R_{AB}\,R_{BC}\,R_{CA})^3}. \tag{36}$$

Here, θ_a, θ_b, and θ_c are the internal angles of the triangle formed by R_{AB}, R_{BC}, and R_{CA} while C_9^{ABC} is the triple-dipole constant defined by

$$C_9^{ABC} = \frac{3}{\pi} \int_0^\infty d\omega\, \alpha^A(i\omega)\, \alpha^B(i\omega)\, \alpha^C(i\omega). \tag{37}$$

Due to the fact that the total three-body contributions are rather small (∼5-10%) their coefficients are approximated by geometric means of dipole-dipole dispersion coefficients as

$$C_9^{ABC} \approx -\sqrt{C_6^{AB}\, C_6^{BC}\, C_6^{CA}}. \tag{38}$$

This approximation was tested for different element combinations so that a general deviation of about 10-20% to the exact expression could be determined [123]. The energy expression for the three-body contribution is then given as

$$E_{disp}^{(9)} = \sum_{ABC}^{\text{triples}} E^{ABC}\, f_{damp,(3)}(\overline{R}_{ABC}), \tag{39}$$

where the sum is over all atom triples ABC in the system applied with a damping scheme proposed by Chai and Head-Gordon [127]

$$f_{damp,(3)}(\overline{R}_{ABC}) = \frac{1}{1 + 6\left(\overline{R}_{ABC}/\left(4/3\,\overline{R}_0^{ABC}\right)\right)^{-16}}. \tag{40}$$

In equation 40 the geometrically averaged inter-atomic distances $\overline{R}_{ABC}$ as well as cutoff radii $\overline{R}_0^{ABC}$ are used. In densely packed systems the correction is in general repulsive. Attractive contributions are only found for more linear arrangements. As demonstrated by a database of 1225 inter-molecular C_6^{AB} coefficients (see Fig. 2), the semi-classical D3 schemes yields highly accurate results competitive to approaches that directly use electronic structure informations to scale tabulated atomic references. The accuracy can be further increased by incorporating a partitioned atomic charge information as recently shown [128].

2.3 Hierarchy of cost-efficient density functionals

Nowadays increasingly large systems can be computed routinely with good accuracy by dispersion corrected DFT together with relatively large basis sets (triple-zeta quality or better in the GTO expansion). Still, despite of the good cost-accuracy ratio of DFT for large systems, larger model systems and smaller time scales for the computation might still make the conventional DFT approaches unfeasible. To overcome this problem while keeping the good DFT accuracy, several composite schemes were developed during the past five years. They are based on established mean-field methods like HF or semi-local DFT and share a treatment within modified basis sets to gain substantial speed ups. As noted in the previous section, several basis set options are in principle available. We aimed for a seamless description of 0D to 3D systems and do not specifically target bulk solids with very delocalized electron density. Thus, the use of GTOs is most efficient and major parts of the effort has been devoted in revising established GTO basis sets of minimal (MINIX), double-zeta (def2-mSVP), and triple-zeta (mTZVP) quality with focus on organic and organo-metallic systems. Each basis set is the foundation for one of the 3c-methods and the general energy expression is:

$$E_{tot}^{3c} = E_{tot}^{HF,DFT/basis} + E_{disp}^{D3} + E_{BSSE}^{gCP} + E^{SRB} \tag{41}$$

The occurring basis set superposition errors are mostly removed by incorporating the gCP correction (see equations 26,27) or by explicit parametrization of the underlying model of theory

and remaining short-range basis set errors on bond length are corrected by an atom-pairwise potential, dubbed short-range basis set (SRB) correction. Missing long-range correlation effects are accounted for by applying the semi-classical D3 dispersion correction (see section 2.2). A

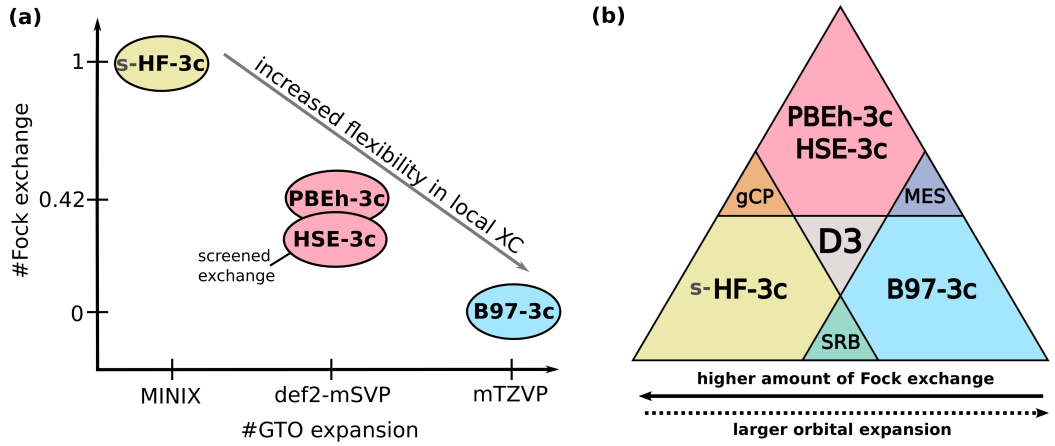

Figure 3: Sketch of the 3c composite methods according to (a) their basis set size and amount of Fock exchange and to (b) their semi-classical correction potential. Here, MES is short for modified electronic structure, i.e. an adapted treatment of the semi-local XC, and gCP and SRB are the short-range basis set correction and geometrical counterpoise correction, respectively. Adapted with permission from Ref [50]. Copyright 2018 AIP.

Table 1: Comparison of the hierarchy of "3c" low-cost electronic structure methods.

	HF-3c	PBEh-3c[(a)]	B97-3c
AO basis set	MINIX	def2-mSVP	mTZVP
#parameters in F_x [(b)]	0	2	4
#parameters in F_c [(c)]	0	1	6
Fock exchange [%]	100	42	0
D3 dispersion	yes	yes	yes
SRB correction	yes	no	yes
gCP correction	yes	yes	no

[(a)] HSE-3c identical apart from long-range screening of Fock exchange (range-separation $\omega = 0.11$). [(b)] Exchange enhancement compared to LDA. [(c)] Correlation enhancement compared to LDA, B97-3c has separate parametrization of same-spin and opposite spin.

sketch of the methods ingredients is given in Fig. 3 and listed in Table 1. With increasing basis set expansion, the amount of Fock exchange can be reduced and still yield (on average) correct geometries. For the smaller basis sets (MINIX and def2-mSVP), we exploit that elongated main group bond lengths due to the small GTO basis can be systematically compensated by higher amounts of Fock exchange. Following Occam's razor, we choose for each 3c-method the least empirical (less flexible) XC form that still yields competitive results. In the minimal basis set HF method (HF-3c), short-range electron correlation is completely neglected, which reduces the area of applicability, but has several advantages as listed below. In a slightly larger basis set (as

in PBEh-3c, and HSE-3c), the semi-local description could be improved, but still uses one of the simplest exchange- and correlation enhancement factors of the PBE [59] form. Various flexible forms have been tested but they were not able to yield a systematical improvement. Finally, the triple-zeta basis set (as in B97-3c) is large enough to profit from a more involved XC treatment and we choose Becke's B97 form [129]. While HF-3c needs all three semi-classical correction potentials, the SRB correction is not needed for PBEh-3c/HSE-3c and the (quite empirical) gCP correction can be dropped for B97-3c, i.e. the empiricism in these corrections decreases in the order HF-3c, PBEh-3c/HSE-3c, and B97-3c. In the following sections the theoretical main ingredients of all composite scheme are shortly summarized.

2.3.1 HF-3c

In 2013, the first three-fold corrected composite scheme based on HF theory has been published by Sure and Grimme. Its main area of application is the computation of structures, vibrational frequencies, and non-covalent interaction energies in huge molecular systems. Using HF has the following advantages: First, in contrast to DFT, HF does not suffer from self interaction errors and extended charged systems are unproblematic even when treated without an implicit solvation model (*in vacuo*). Second, due to the simple analytical expressions of the HF equations gradients and Hessians are easily accessible without numerical noise in geometry optimizations or frequency calculations. Third, HF is inherently able to treat the important hydrogen bonding motif so that there is no need for atom-type dependent H-bond corrections which are often applied to semi-empirical approaches [130, 131]. The proposed HF method can be applied without further parametrization to almost any element of the periodic table and includes important effects like Pauli-exchange intrinsically. By following the standard definition of electron correlation, it is clear that HF is not able to describe Coulomb correlation effects. The suggested method is hence not generally applicable but yields simple properties as equilibrium structures, vibrational frequencies or non-covalent interaction energies, when changes in the electronic structure during a chemical process remain small. Processes which require to account especially for dynamic correlation effects (reaction energies for, e.g., bond cleavage) are not covered here. The total energy in HF-3c is

$$E_{tot}^{HF-3c} = E_{tot}^{HF/MINIX} + E_{disp}^{D3(BJ)} + E_{BSSE}^{gCP} + E^{SRB} \tag{42}$$

In the following, all corrections are shortly described: First, the D3(BJ) dispersion correction scheme is incorporated into the HF-3c method for describing non-covalent interactions in a reasonable manner. Second, the basis set is of minimal quality for typical ("organic") elements H, C, N, O and mostly of split-valence (SV) or polarized split-valence (SVP) quality for the other elements (cf. table 2). The basis set superposition error (BSSE) is significant for a small basis set and is mostly corrected by a gCP correction (cf. equation 26). Importantly, also intra-molecular BSSE is treated this way, which is difficult to correct efficiently otherwise. According to the well-established practice, basis set effects are separated into BSSE and the basis set incompleteness error (BSIE). Following this sense, the gCP scheme accounts for the atom-pairwise part of the BSSE and the third short-ranged basis correction (SRB) corresponds to the BSIE. SRB corrects for systematically overestimated bond lengths for electronegative elements (e.g., N, O, F) and

Table 2: Composition ot the MINIX basis set.

Element	**Basis**
H-He, B-Ne	MINIS
Li-Be	MINIS+1(p)
Na-Mg	MINIS+1(p)
Al-Ar	MINIS+1(d)
K-Zn	SV
Ga-Kr	SVP
Rb-Xe	def2-SV(P) with ECP

is given as a pairwise sum in the following

$$E^{SRB} = -\frac{s}{2} \sum_{A \neq B} (Z_A \, Z_B)^{3/2} \exp\left(-\gamma (R_0^{AB})^{3/4} \, R_{AB}\right). \tag{43}$$

In equation 43, R_0^{AB} are the default Becke-Johnson cutoff radii as determined *ab-initio* for the D3 dispersion correction scheme and Z_A, Z_B are the nuclear charges of atom A, B. The correction is applied for all atom types with $Z \leq 18$. The empirical fitting parameters $s = 0.03$ and $\gamma = 0.7$ were determined to minimize HF-3c forces for given B3LYP-D3(BJ)/def2-TZVPP equilibrium structures for 107 small organic molecules. $E_{disp}^{D3(BJ)}$ and E_{BSSE}^{gCP} have been included during the fitting process of E^{SRB}, which was carried out to minimize the HF-3c RMS gradient for reference geometries. Here, all D3(BJ) and gCP parameters were held constant at their initially optimized values within this procedure. All empirically derived parameters of the HF-3c composite method are given in table 3. Molecular crystals represent an important benchmark to study the accuracy

Table 3: Empirical parameters of the HF-3c method.

	Parameter			
Contribution			s	γ
E^{SRB}			0.03	0.7
		a_1	a_2	s_8
$E_{disp}^{D3(BJ)}$		0.4171	2.9149	0.8777
	α	β	σ	η
E_{BSSE}^{gCP}	1.1549	1.1763	0.1290	1.1526

of the theoretical method in order to describe intermolecular interactions in a reasonable way. Computing cohesive energies as well as optimized cell volumes is nowadays affordable for medium to large molecular crystals, which is, furthermore, comparable to a broad pool of experimental data. For this purpose different benchmark sets have been proposed to assess the performance of different methods for describing molecular crystals. One well established example is the X23 benchmark set which consists of 23 different molecular crystals ranging from pure dispersion driven to H-bonded systems. The original HF-3c methods performs rather well (comparable to LMP2 methods [132]) but shows a tendency of overbinding molecular crystals, especially for

weakly bound systems. To cure this drawback a refinement of the HF-3c has been attempted by tuning the dipole-quadrupole dispersion term in the energy expression (see Ref. [132]). This method termed as scaled HF-3c (sHF-3c) has a by 0.7 scaled s_8 factor which describes the $-C_8^{AB}/R_{AB}^8$ contribution. While the performance of cohesive energy prediction slightly worsens, optimized unit cell volumes are in much better agreement with experiment.

2.3.2 PBEh3c

In 2015, a DFT-based composite electronic structure approach has been proposed for the efficient calculation of structures and interaction energies in large chemical systems. It is based on the well-known and numerically robust PBE generalized gradient approximation in a modified global hybrid functional with 42% of non-local Fock-exchange dubbed PBEh-3c. The starting point for calculating the electronic energy is a standard hybrid density functional treatment with a small basis set. Generally, the Ahlrich's-type split valence basis set def2-SV(P) [133] (together with its ECP for heavier elements) is employed. For the elements B-Ne, the similar double-zeta sets are used which contain one extra uncontracted s-function leading to slightly shortened (improved) bond lengths. Exponents for d-polarization functions are resembling those from the 6-31G* basis set [134] to values of $\alpha(d) = 0.8$ for carbon to fluorine whereas standard values are applied to boron $\alpha(d) = 0.5$ and neon $\alpha(d) = 1.8$. A weakness of the def2-SV(P) basis set is the systematic overestimation of bond distances involving hydrogen which is caused by missing p-polarization functions on hydrogen. This effect could be compensated by scaling all hydrogen s-function exponents by a value of 1.2^2 which accounts for an increased nuclear charge in typical molecular environments. This factor has been manually adjusted to match X-H bond lengths as well as atomization energies of the methane, ammonia, and water molecule. For all other elements, the standard def2-SV(P) basis set is used throughout without any changes and the entire new basis set is termed def2-mSVP (where "m" stands for modified). Note that def2-SV(P) and def2-mSVP involve practically the same computational costs. Two terms are added to correct the DFT energy E_{tot}^{PBEh} in order to include long-range correlation effects and to correct for the occurring BSSE.

$$E_{tot}^{PBEh-3c} = E_{tot}^{PBEh} + E_{disp}^{D3^{\mathrm{ATM}}(BJ)} + E_{BSSE}^{damped,gCP} \tag{44}$$

In equation 44, PBEh denotes the density functional used (PBEh in the double-zeta basis set). For the present method, all D3(BJ) and gCP parameters have been re-parametrized using interaction energies of the S22 set [135]. Due to linear dependencies between D3 and gCP parameters during the fitting procedure, the s_8 was set to zero for this composite scheme, i.e. a purely dipole-dipole interaction. Furthermore, three-body interactions (dipole-dipole-dipole interactions n=9) are included *via* the ATM term. The second correction term denotes a slight modification of equation 26. In contrast to the earlier proposed HF-3c method, the PBEh-3c composite scheme puts more emphasis on accurate thermochemistry. Therefore, the effect of the applied gCP correction, given in equation 26, in relation to this property has been further investigated. It turned out that in few cases the 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 effect could be canceled *via* the additional E_{BSIE}^{SRB} term which is omitted here. Hence, the basic form of the gCP correction has been maintained but damped to be excluded

Table 4: Empirical parameters of the PBEh-3c method.

	Parameter			
Contribution	κ	μ	β_{PBE}	a_X
E_{XC}^{PBEh}	1.0245	0.123457	0.03	0.42
	a_1	a_2	s_8	a_3
$E_{disp}^{(2)(3)}$	0.4860	4.5	0^a	1^a
	α	β	σ	η
$E_{BSSE}^{damped,gCP}$	0.27649	1.9560	1^a	1.32492
			k_{dmp}^1	k_{dmp}^2
			4	6

[a] Constrained value.

in the short-range regime. Furthermore, the global hybrid density functional part is based on PBE GGA exchange and correlation according to

$$E_{XC}^{PBEh} = (1 - a_X)\, E_X^{PBEh} + a_X\, E_X^{HF} + E_C^{PBE}, \tag{45}$$

where the parameter $a_X = 0.42$ describes the amount of non-local Fock exchange. In order to have more flexibility in the electronic part, the enhancement factor g_X of the PBE exchange, given by

$$g_X^{PBE} = 1 + \frac{\mu\, s^2}{1 + \frac{\mu\, s^2}{\kappa}}, \tag{46}$$

has been modified to yield better atomization energies and certain difficult electron correlation effects. The parameter $\mu = 10/81$ is equal to the one of PBEsol [136] whereas $\kappa = 1.0245$ is the average between the values of PBEsol and revPBE [137]. The PBE correlation functional E_C^{PBE} contains a parameter $\beta_{PBE} = 0.03$ which controls the admixture of GGA correction to the LDA correlation energy. This parameter was adjusted *via* matching atomization and total energies of a few small molecules and reaction energies from the GMTKN30 [138] benchmark set. All parameters are given in table 4.

2.3.3 HSE-3c

In 2016 a screened non-local exchange variant of the global hybrid PBEh-3c based on the Henderson-Janesko-Scuseria [139, 140] exchange hole model (cf. equation 47) has been published [49]. The relatively large amount of Fock exchange within the PBEh-3c composite method hampers SCF convergence, which is especially pronounced in condensed systems with metallic characteristic, i.e. with orbital energy gaps smaller than 3.5 eV. Unfortunately, this excluded important classes of solids or other condensed phase systems from a numerically stable description (e.g., the naphthalene crystal falls into this category). For instance, numerical problems have been detected by describing organic semiconductors, extended protein structures, and pharmaceutically relevant organic polymorph structures. The problem has been solved by combining PBEh-3c with the Heyd-Scuseria-Ernzerhof [141, 142] (HSE06) exchange functional. The method dubbed HSE-3c has the following energy expression

$$E_{tot}^{HSE-3c} = E_{tot}^{HSE-3c} + E_{disp}^{D3^{\mathrm{ATM}}(BJ)} + E_{BSSE}^{gCP}. \tag{47}$$

Here, E_{tot}^{HSE-3c} includes among others (e.g., the kinetic energy) the exchange-correlation (XC) functional which is given as follows

$$\begin{aligned} E_{XC}^{HSE-3c} &= (1-a_x)\, E_X^{PBEh-3c,SR}(\omega) + a_x E_X^{HF,SR}(\omega) \\ &\quad + E_X^{PBEh-3c,LR}(\omega) + E_C^{PBEh-3c}. \end{aligned} \tag{48}$$

The XC functional is evaluated in an unmodified small Gaussian AO basis set of double-zeta quality termed def2-mSVP, as described in section 2.3.2. The short-range (SR) Fock admixture parameter a_x, the modified PBE exchange, and the modified PBE correlation are fixed to the PBEh-3c values. The HJS exchange hole model $H(s_\sigma(r))$ has been parametrized to match the enhancement factor of the PBEh-3c exchange functional (cf. equation 46)

$$H(s_\sigma(r)) = \frac{\sum_{k=2}^{7} a_k\, s_\sigma^k(r)}{\sum_{k=1}^{9} b_k\, s_\sigma^k(r)} \tag{49}$$

with parameters $\{a_i, b_i\}$ and the reduced spin density gradient $s_\sigma(r)$

$$s_\sigma(r) = \frac{\nabla n_\sigma(r)}{n_\sigma^{4/3}(r)}. \tag{50}$$

The error function separation is controlled *via* the standard parameter $\omega = 0.11$ (fixed at the HSE06 value) which switches the HSE-3c potential to a pure GGA at long-range. For treating long-range London dispersion interactions the semi-classical D3(BJ) dispersion energy correction is applied, respectively. Furthermore, three-body interactions (dipole-dipole-dipole interactions n=9) are included. Due to the small atom-centered single particle basis set, the significantly occurring BSSE is effectively removed by the gCP correction. All empirically derived parameter are given in Ref [49].

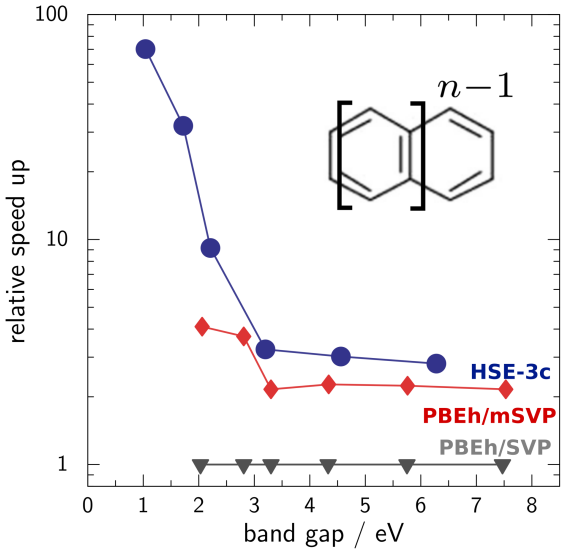

Figure 4: Relative speed of HSE-3c compared to its global hybrid variant for a series of acene crystals (benzene to hexacene). The smaller the band gap, the larger the speed up due to the screening of exchange interactions with identical technical settings (CRYSTAL17 program).

We test the applicability of HSE-3c for small gap systems on a series of acene crystals (see Fig. 4). This is an ideal test case as the band gap decreases with increased π system. We see indeed a substantial improvement compared to the global hybrid PBEh-3c up to two orders of magnitude for hexacene. Additionally, a comparison with Ahlrich's SVP basis set shows that the def2-mSVP basis is indeed faster while yielding variationally lower total energies for the tested molecular crystals.

2.3.4 B97-3c

In 2018 a revised version of the well-established B97-D DFA has been proposed. It is a pure GGA variant of Becke's power-series ansatz from 1997 and is explicitly parametrized by including long-range electron correlation effect in terms of the D3 dispersion correction. Different from other composite schemes, the orbitals are expanded in a modified triple-zeta Gaussian AO basis set dubbed mTZVP, which is available for all elements up to Rn. The remaining BSSE is adsorbed in the modified B97 parameterization and thus, an additional treatment by a gCP correction is not needed. It could be shown that B97-3c is excellent for molecular and condensed phase geometries, competing with the performance of most hybrid functionals evaluated in larger basis set expansion [50]. B97-3c can be applied to hundreds of atoms on a single processor and due to lacking non-local exchange, it is suggested as a robust computational tool in particular for systems with partial multi-reference character. The B97 functional is based on a remapping of the reduced spin gradient variable $s_\sigma(n_\sigma(r))$, where $n_\sigma(r)$ denotes the electron density with $\sigma = \alpha/\beta$ spin. The XC functional is given as

$$E_{XC}^{B97} = E_X + E_{C\alpha\beta} + \sum_\sigma E_{C\sigma\sigma}. \tag{51}$$

The correction factors are expanded in a power series in the remapped variable $u(s^2)$. Becke, as well as Head-Gordon, [143, 144] showed that three-terms in the series are a good compromise between flexibility and "robustness" for a GGA functional. The B97-3c energy expression is calculated as

$$E_{tot}^{B97-3c} = E_{tot}^{mB97} + E_{disp}^{D3^{\mathrm{ATM}}(BJ)} + E^{SRB}. \tag{52}$$

The first term contains the re-fitted energy term of the B97 functional applied with the modified mTZVP basis set. Furthermore, London dispersion effects are treated for pair- and triple interactions as described before in the D3(BJ) version. The last correction term E^{SRB} denotes a slight modification of the short-range basis set correction developed for HF-3c (cf. equation 43).

$$E^{SRB} = -\frac{q_{scal}}{2} \sum_{A,B} \sqrt{Z_A Z_B} \exp\left(-r_{scal} \frac{R_{AB}}{R_0^{AB}}\right) \tag{53}$$

Here, q_{scal} and r_{scal} are fitting parameter, R_{AB} is the inter-nuclear distance, $Z_{A/B}$ are the nuclear charges and R_0^{AB} standard D3 damping radii. The SRB correction mostly corrects for artificially elongated covalent bond lengths that occur for most GGA density functionals.

2.3.5 Molecular dipole moments

As the in section 2.3 presented methods employ rather compact orbital basis set expansions, it is challenging to describe the charge density very accurately, which might lead to errors in

the electrostatic, induction, and exchange interactions. For polar molecules, the electrostatic

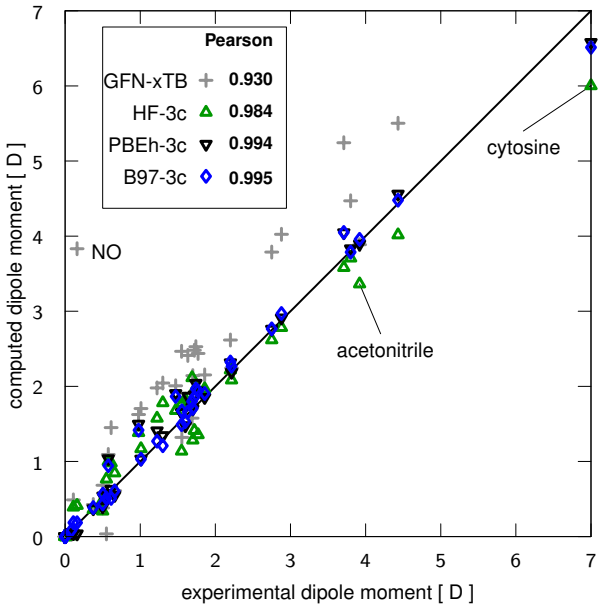

Figure 5: Molecular dipole benchmark: We compare computed dipole moments from three of our "3c" methods (see section 2.3) as well as a modern tight-binding Hamiltonian GFN-xTB [145] with experimental measurements [146].

contribution is particularly important and we thus test the leading order dipole for a set of 43 molecules [146]. Correlation between computed and calculated dipole moments is shown in Fig. 5. The correlation increases in the order HF-3c, PBEh-3c, and B97-3c, which demonstrates the relation to the larger basis set expansion. Results of PBEh-3c and B97-3c are quite encouraging with Peason correlations of 0.994 and 0.995, and mean absolute deviations (MADs) of 0.11 and 0.09 Debye, respectively. This is competitive to results of other density functionals evaluated in much larger basis set expansions, e.g. PBE, PBE0, and B3LYP/aug-cc-pvTZ all yield an MAD of 0.09 Debye [146]. HF-3c uses a minimal basis set and is thus in between traditional semi-empirical methods and standard DFT. While the dipole moments are not particularly good with MAD of 0.21 Debye, it is similar to Hartree-Fock in the popular 6-31G* basis and clearly superior to the tight-binding Hamiltonian. Additionally, some errors in the electrostatic interaction of HF-3c compensate with errors from the Pauli exchange repulsion as analyzed in Ref. [103]. As we do not use any further approximation for Coulomb electrostatics (as done in traditional semiempirical methods), we expect theses results to be transferable to larger systems. Indeed, intermolecular interaction energies of large complexes and molecular crystals turn out to be quite accurate (see section 3).

2.3.6 Implementations and computational cost

Due to the compact basis sets with low angular momentum functions, the presented "3c"-methods can be ideally evaluated on modern CPU and GPU based architectures. All methods have been implemented in a developer version of CRYSTAL17 [147]. CRYSTAL17 is an ideal program for large scale screening of many crystal structures as it can employ all point and space group symmetries [148]. Furthermore, it scales well on high-performance computational facilities

with up to 30 000 cores and electronic structure calculations on 14 000 atoms in the primitive unit cell have been presented recently [149]. A simple scaling example is given in Fig. 6 where different supercells of the one dimensional adenine-thymine helix are calculated at the sHF-3c level of theory. We first see the efficient treatment of the full periodic structure by only one base

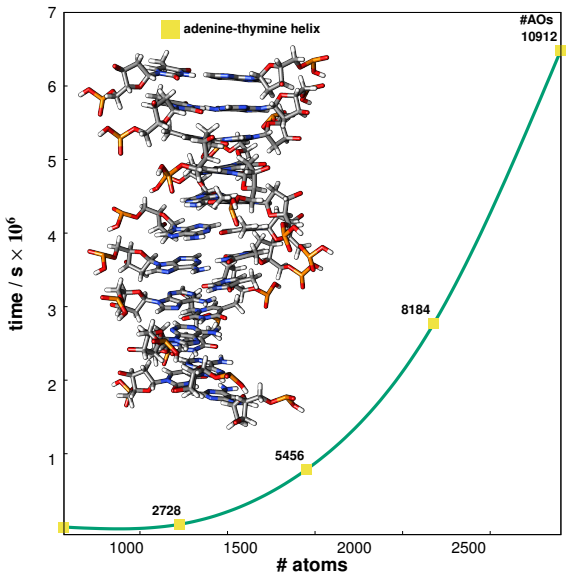

Figure 6: Scaling behavior of the adenine-thymine (poly(dA)-poly(dT)) helix (one dimensional periodicity) calculated at the sHF-3c level of theory using the CRYSTAL17 software suite. Each cross represents an augmented supercell starting with the irreducible primitive cell (66 atoms). All single-point calculations have been executed at the same hardware (one node with 28 CPUs).

pair in the irreducible unit (66 atoms). The eleven symmetry operators to generate the primitive cell reduce the computational costs significantly by a factor of 15. While not all crystals have such a high symmetry group, the most common space group for molecular crystals ($P2_1/c$) has four symmetry operators that yield up to a 64 fold speed up for a cubic scaling method. We can furthermore see that the electronic structure solution of a helix with 2904 atoms in the primitive cell can be computed with less than 2000 CPU hours. This enables routine geometry optimizations and even computation of phonon spectra for large complex systems on standard computational facilities. The cost effective calculations by the sHF-3c method is intrinsically due to their construction in restricted orbital expansions. All other numerical and implementation enhancements like linear scaling DFT [150] can in principle be employed additionally. Apart from the CRYSTAL17 package, additional implementations of most of the described low-cost methods are available in the (mostly) molecular codes Orca [151], Turbomole [152], Psi4 [153], CP2K [154–156], Fermions [157], and TerraChem [158, 159]. While CRYSTAL17 uses continuous fast multipole methods to approximate long-range Coulomb and exchange integrals [160, 161], all the other codes additionally employ density fitting techniques [162–164] with recent iterative low-memory extensions to periodic systems [165–167]. This can make a substantial difference for larger basis set calculations, especially for pure GGA functionals like B97-3c. In practice, this leads to the relative speed for larger molecular complexes of 1:10:50 for HF-3c, B97-3c, and PBEh-3c, respectively (see Ref. [50]). The GPU implementation of HF in TerraChem can lead to a substantial boost in computational speed, making the full electronic structure description

(HF-3c and PBEh-3c level) of protein ligand binding affinities feasible [168].

3 Applications

All composite methods presented in the last sections are intended to cover a broad spectrum of applications in the field of organic, inorganic and organometallic chemistry. Due to the cost advantageous nature of each method, they are particularly suitable for the application of structural properties and non-covalent interactions occurring, e.g., in molecular crystals. Specifically for predicting crystal structures of organic molecules as highlighted in the introduction, London dispersion corrected DFT has been shown to be promising. Neumann and coworkers have been successfully applying combinations of molecule specific force fields with DFT methods including the D3 dispersion correction in the CSP blind tests organized by the Cambridge structural database [6, 169–171]. We have extended this analysis to a broader range of DFT methods, where the most recent developments showed promising accuracies [172]. In the following the performances of the introduced low-cost methods are compared to high-level reference data in relation to several benchmark studies. Furthermore, two selected examples are given to highlight the benefits of the low-cost methods to solve quantum chemical problems.

3.1 Molecular crystal benchmark sets

3.1.1 Lattice energies compared to high-level references

Molecular crystals are an important class of systems with a huge impact on material science [173] as well as on pharmaceutical industry [174]. Unfortunately, high-level methods are limited in their application to a smaller range of systems. Therefore, efficient DFT methods are crucial for routine applications. The central quantity to assess the stability of a crystal is given by its lattice energy E_{latt} which is by definition the energy per molecule gained upon adopting the crystalline form with respect to the gas state

$$E_{latt} = \frac{1}{Z} E_{crys} - E_{gas}. \tag{54}$$

Here, E_{crys} is the energy of the crystal with Z molecules within the primitive cell and E_{gas} is the energy of the isolated molecule in its lowest energy conformation. Recently, reference energies based on high-level diffusion Monte-Carlo (DMC) have been calculated for a set of eight molecular crystals [18] which contain different binding motifs like strong hydrogen bonds as well as unsaturated π-systems with dominant London dispersion interactions. Table 5 shows the results of the introduced composite methods in comparison to affordable many-body methods like the random phase approximation (RPA@PBE) and second order perturbation (MP2). The mean absolute (unsigned) deviations (MADs) of sHF-3c, HSE-3c, and B97-3c are comparable with each other (1.1, 1.3, and 1.4 $\mathrm{kcal\,mol^{-1}}$) and outperform even the RPA@PBE method which yields an MAD of 1.9 $\mathrm{kcal\,mol^{-1}}$ [18]. While RPA can be signifiantly improved by singles corrections, the calculations get substantially more costly and are currently not applicable to larger molecular crystals [26]. The commonly used MP2 perturbation theory is able to describe simple hydrogen bonded systems, but fails for delocalized systems with small orbital gaps. In

Table 5: **DMC8:** Comparison of lattice energies for 8 molecular crystals (subsets of ICE10 [175] and X23 [176, 177]) compared to high-level DMC references [18]. Energies are given in kcal/mol, RPA results from [26] and MP2 from [132, 178]

			deviation $E^{\text{calc}} - E^{\text{ref}}$				
	system	Ref.	RPA	MP2	sHF-3c	HSE-3c	B97-3c
1	ice Ih	-14.2	1.7	0.1	-1.8	-2.8	-2.2
2	ice II	-14.1	1.8	0.2	-1.3	-2.2	-1.6
3	ice VIII	-13.7	1.9	0.2	0.7	-0.8	-0.7
4	carbon dioxide	-6.7	1.0	-0.2	-0.0	-0.1	0.9
5	ammonia	-8.9	1.3	-0.5	-0.1	-3.7	-2.7
6	benzene	-12.7	1.9	-2.0	1.4	-0.2	-1.9
7	naphtalene	-18.8	2.5	-3.0	1.4	0.2	-3.3
8	anthracene	-25.2	3.1	-5.1	2.2	0.6	-4.4
mean deviation			1.9	-1.3	0.3	-1.1	-0.4
mean absolute deviation			1.9	1.4	1.1	1.3	1.4

this regard, the presented low-cost methods are reliable even though this set of systems is quite challenging for effective mean field methods.

3.1.2 Lattice energies and geometries from experimental measurements

Another 33 molecular crystals are taken from the ICE10 [175] and the X23 [176, 177] benchmark sets to process a statistically larger amount of data points. Here, semi-experimental references are derived from measured sublimation enthalpies (ΔH_{sub}) by removing zero-point (E_{ZPVE}) and thermal effects with measurement temperatures T and adiabatic heat capacity differences ΔC_p

$$E_{latt} = -\Delta H_{sub}(T) + \Delta E_{ZPVE} + \int_0^T dT' \Delta C_p(T'). \tag{55}$$

The uncertainty of these references have a range of 1-1.5 kcal mol^{-1} due to uncertainties in the sublimation measurement [180]. The zero-point and thermal effects on the unit cell volumes have been estimated for both benchmark sets with an uncertainty of about 1-2%. Individual results of the presented composite schemes as well as for the B3LYP hybrid functional applied with a def2-SV(P) basis set are given in Fig. 7. The strong hydrogen bonds in the ICE10 set are particularly challenging for small basis set expansions which is ideally compensated by the scaled minimal basis set HF-3c method. HSE-3c is performing a bit worse, probably due to underestimated BSSE, but it performs better than related hybrid functionals in a comparable basis set expansion (cf. B3LYP/def2-SV(P) results in Fig. 7). B97-3c is performing in the upper range of GGA functionals applied with a converged basis set like BLYP-D3 [175] with an overall MAD of 1.2 kcal mol^{-1} for calculated lattice energies. On average 0.7% mass densities are 0.7% too small.

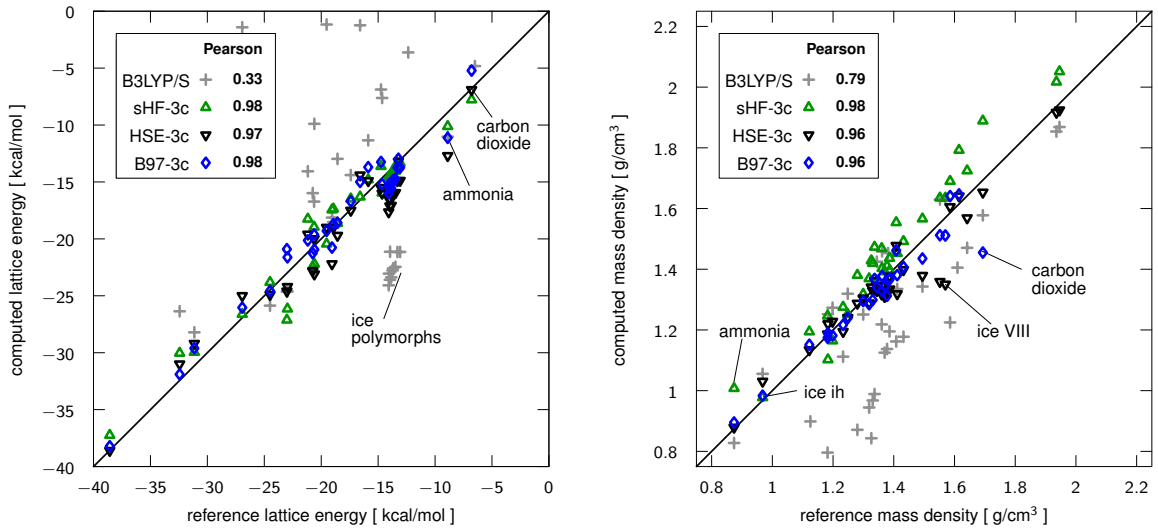

Figure 7: Lattice energies and mass densities of 33 molecular crystals from the ICE10 [175] and X23 [176, 179] benchmark set. In addition to the three composite methods, we show results from B3LYP in a small basis set taken from Ref. [48] and give the Pearson linear correlation value. Adapted with permission from Ref [50]. Copyright 2018 AIP.

3.2 Selected examples

3.2.1 Shortest H···H contact

Sterically overcrowded molecular structures originate from short H···H contacts that may correspond to potential energy minima or conformational transition states [181]. Especially the latter involve severe H···H repulsions leading to substantial interconversion barriers. The accompanying unusual and therefore interesting characteristics of many overcrowded molecules depend decisively on the type of their non-covalent H···H interactions at short distances. The development of theoretical tools for the computer-assisted treatment of such overcrowded molecules, especially of methods based on quantum mechanics, requires reliable experimental reference data of structures with short H···H distances on which the theoretical models can be tested and calibrated. Key features are the non-covalent interactions dominated by London dispersion forces. The aim here is to use a low-cost approach which is able to describe such structural motifs occurring, e.g., in molecular crystals within a reasonable amount of computational costs. By taking a closer look at experimental X-ray data it is well known that hydrogen positions obtained from such measurements suffer from large systematic errors due to missing scattering abilities of hydrogen atoms whose electron densities usually do not coincide with the proton locations, respectively [182, 183]. This effect is already visible when examining the electron density distribution of normal C-H bonds. In this case, the maximum electron density is shifted by 0.1 Å to the carbon atom, so that C-H bond lengths determined by X-ray experiments are generally underestimated. Other experimental reference methods such as neutron diffraction at low temperatures are completely free of such disadvantages, since the ability to bend neutrons does not differentiate much by going from hydrogen to heavier atoms. Such low-temperature neutron diffraction data (NRD) are therefore ideally suited as a reference to test theoretical models. Recently, low-temperature single-crystal neutron diffraction measurements of the crys-

talline tri(3,5-tert-butylphenyl)methane, abbreviated as TPM in the following, (cf. Fig. 8) at 20.0 K have been reported showing an intermolecular H$\cdots$H distances of 1.566(5) Å, which is the shortest non-covalent H$\cdots$H contact to date [184]. To rationalize this unusual structural

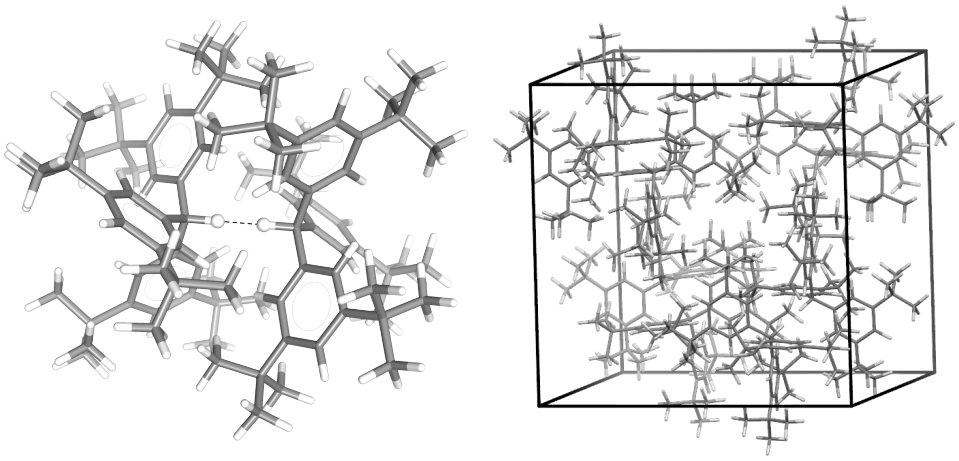

Figure 8: Representation of TPM as molecular dimer (left) and the primitive cubic cell with $P2_13$ symmetry (right). The mentioned hydrogen-hydrogen bond is highlighted. The mentioned short hydrogen bonds are stressed out on the left side.

motif, electronic structure simulations of the experimentally derived starting structure at the HSE-3c level [49] were performed. The starting structure exhibits a cubic space group of $P2_13$ symmetry with eight molecular dimers in the primitive cell (856 atoms). All atomic positions and cell lengths were relaxed until convergence has been reached. The good agreement with the experimentally derived solid-state structure is reflected in particularly in the excellent description of the intermolecular H$\cdots$H distance. The calculated solid-state structure generally matches the extrapolated experimental bond lengths at 0.0 Kelvin extremely well. The volume of the primitive cell is reduced by 2% which is in the typical range for organic crystals [185] due to missing zero-point vibrational and thermal effects. The finite temperature volume of the primitive cell has been determined by exploiting the quasi-harmonic approximation to include zero-point and thermal vibrational effects which gives an absolute deviation of 0.1% with respect to the experimentally measured volume. Packing effects have been estimated by recomputing the TPM dimer in the gas phase at the same level of theory, results are summarized in table 6. This strongly indicated that the short H$\cdots$H contact originated from a London dispersion driven crystal packing effect. To strengthen this hypothesis, an energy decomposition analysis [186] (EDA) at the B3LYP-D3^{ATM}(BJ)/def2-TZVPP level of theory is used to separate the various energy contributions to the overall stability. Through this EDA approach, the hydrogen bond interaction can be decomposed into electrostatic (ES), exchange-repulsion (EXR), orbital-relaxation (OR), local correlation (loc-Corr), and dispersion interactions (DISP, covered through the D3 correction). Figure 9 shows the EDA of the optimized dimer geometry and a full dissociation curve with fixed monomer geometries. London dispersion interactions are the dominant binding motif leading to the overall stability of the TPM dimer. To put the efficiency of the presented low-cost electronic structure methods into perspective, we compare the timings for a single self-consistent field solution of the TPM crystal in its $Pa\overline{3}$ modification.

Table 6: Structural data of TPM determined by neutron diffraction (abbreviated as NRD) and by the composite method HSE-3c [184].

	NRD			HSE-3c	
T[K]	20.0	100.0	200.0	0.0(s)	0.0(g)
R-value	0.031	0.054	0.096	–	–
Volume[Å^3]	7833.1(4)	7901.4(4)	8040.4(4)	7676.9	–
H$\cdots$H [Å]	1.566(5)	1.577(6)	1.594(9)	1.555	1.634

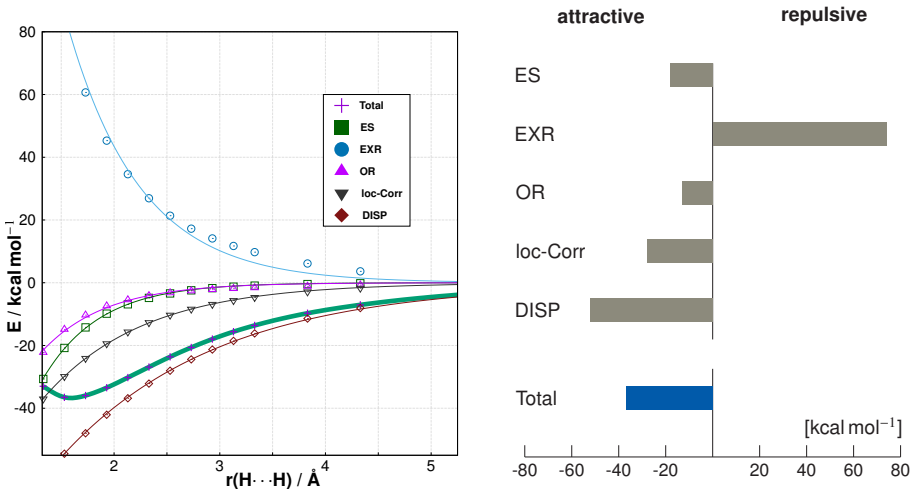

Figure 9: Energy decomposition analysis at the B3LYP-D3$^{\mathrm{ATM}}$(BJ)/def2-TZVPP level of theory with respect to the H$\cdots$H distance of the TPM dimer. *Left:* The total interaction energy (bold, green curve) shows a minimum around 1.59 Å (highlighted as dotted line). *Right:* The individual components of the energy decomposition at the minimum geometry is shown.

Here, a semi-empirical DFTB3-D3 [187–189] method is compared with the three low-cost methods sHF-3c, PBEh-3c, and HSE-3c. Additionally, a projector augmented-wave method (PBE-D3/PAW,1000eV) is used as cross-check computed with the VASP program [94, 95, 97, 98]. Wall times computed on a single node with 16 CPUs (Intel(R) Xeon(R) E5-2660 v4) are given in Table 7. The explicit timing should be interpreted carefully as different program packages are used and only CRYSTAL17 can fully employ all symmetry operators. In terms of computational times, the presented composite schemes are in between the traditional semi-empirical methods and the full DFT treatment in converged basis sets while keeping the good DFT-D accuracy (see section 3.1).

3.2.2 Shortest P=O$\cdots$I bond

The description of halogen bonding is commonly based on the concept of the so-called σ-hole [190–192]. When the electrostatic potential, produced by the nuclei and electrons of a molecule at any point r, is calculated on a molecular surface one obtains the surface electrostatic

Table 7: Timings of different methods for one single-point calculation of the TPM crystal in $Pa\overline{3}$ symmetry. Absolute timings are given in hours (h), minutes (m), and seconds (s), relative timings are given relative to the "cheapest" method.

Method	[h:m:s]	rel. time
DFTB3-D3	00:00:22	1
sHF-3c	00:02:54	8
HSE-3c	00:21:50	60
PBEh-3c	00:29:35	81
PBE-D3/PAW,1000eV	21:56:33	3599

potential (SEP). The SEP of a molecule is an important physical characteristic for understanding non-covalent interactions. Significant progress in describing halogen (XB) bonding theoretically has been achieved during the recent years and different levels of computational methods have been employed with the main focus of finding the best accuracy-cost ratio [193]. Nonetheless, examples are reported for which the σ-hole concepts fails to explain the experimental finding [194–196]. Beneath well established methods like NMR spectroscopy and UV/Vis absorption, and emission titration, X-ray crystallography is frequently cited as unambiguous proof of XB interactions present in the solid-state [197, 198]. Those interactions are crucial in understanding the often underestimated influence of crystal packing effects in the solid-state [199–201]. Recently, Schaub et al. have proposed an XB complex which represents one of the rare cases where oxygen does not act as a bifurcated halogen bond acceptor [202]. Here, a linear coordination of the halogen bond donor 1,4-diiodotetrafluorobenzene and the phosphoryl (P=O) group of the triarylphosphine oxide (cf. Fig. 10 for the dimer structure cutout) is found. Most importantly,

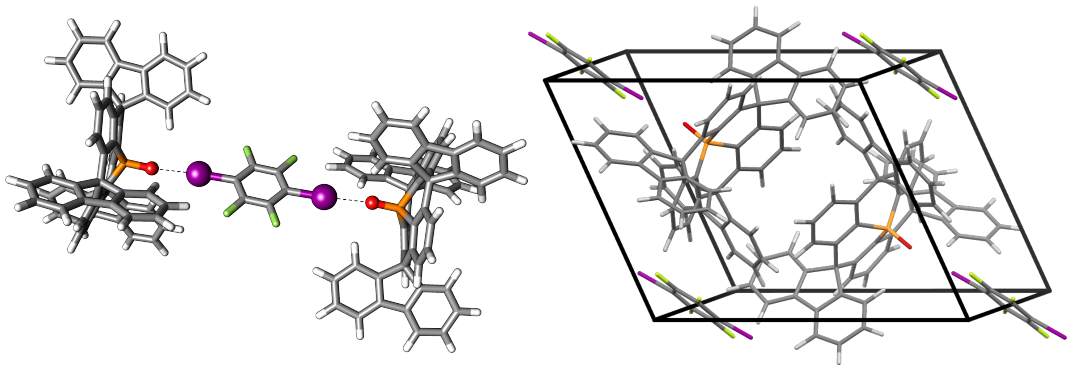

Figure 10: Depiction of the dimer structure cutout from the experimentally derived crystal structure (left) as well as the triclinic primitive cell with space group P-1 (right). The mentioned short halogen bonds are stressed out on the left side.

the O$\cdots$I distance of 2.683(5) Å is significantly shorter than the sum of the van der Waals radii for iodine (1.98 Å) and oxygen (1.52 Å), with an overall shortening of 23.3 %. This complex features the shortest documented halogen bond compromising a P=O$\cdots$I motif and hence might imply that the electron-rich oxygen atom at the central phosphoryl unit acts as an exceptionally strong halogen-bond acceptor. However, this finding is in contrast to titration experiments,

which pointed at the absence of halogen bonding in solution. Quantum chemical calculations of the experimentally measured X-ray structure with HSE-3c confirmed the structural motif of the XB complex. Here, the short experimental O···I bond of 2.683(5) Å could be reproduced to a value of 2.625 Å. Furthermore, the computed P=O···I bond angle differed by less than one degree from the measured one, which intensified a virtually linear arrangement. In order to

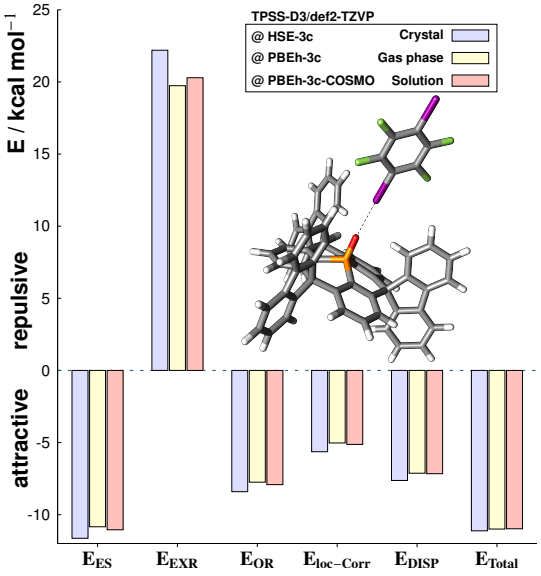

Figure 11: EDA results for the halogen bond interaction of the dimer shown in Fig. 10. Structures were extracted from the HSE-3c optimized crystal structure, PBEh-3c optimized gas phase structure, and PBEh-3c-COSMO optimized structure in solution (structures are given in the ESI). The analysis has been executed at the TPSS-D3/def2-TZVP level of theory.

provide a closer look at the halogen bonding situation an EDA has been conducted for the dimer structure (cf. left part of Fig. 10) at the PBEh-3c optimized structure in the gas phase, PBEh-3c-COSMO optimized structure in solution, and HSE-3c optimized crystal structure. Here, only one triarylphosphine oxide and the halogen bond donor are analyzed. The analysis has been executed at the TPSS-D3/def2-TZVP level of theory (cf. Fig. 11). The total interaction energy does not differ significantly for the three setups used (gas phase, solution, and crystal). With more than one third of the overall interaction energy, the major contribution to the halogen-bonding interaction is provided by the electrostatic interaction, whereas the other interactions roughly contribute equally. Since the halogen bond motif in the crystal is not much larger than in the free dimer, this cannot be considered a driving force in the crystallization process. A closer look to the environment of the halogen bond donor and acceptor within the crystal structure shows that the driving force is by far the stacking interaction of electron-deficient 1,4-diiodotetrafluorobenzene with fluorenyl flanks of the triarylphosphine oxides as has been shown recently [202]. This study showed how the low-cost electronic structure methods could be used to qualitatively and quantitatively explain the supposingly contradicting experimental findings by an effective electronic structure description of gas, solution, and solid state of the XB complex.

4 Future perspective

DFT has been the 'work-horse' in electronic structure computation in both the solid state physics and theoretical chemistry communities. In the past years, many-body electronic structure theories emerged that will partially replace DFT for computing electronic properties on given atomic configurations. However, we still see DFT as an irreplaceable tool for (a) the routine calculation of structures and properties of systems with medium size of about 10^2 atoms and (b) the electronic structure description of increasingly large systems with well above 10^3 atoms. In order to keep DFT as a state-of-the-art methodology, both functional as well as numerical developments will be important. In this topical review, we presented a set of low-cost methods and mainly focused on the electronic structure part by combining compact orbital basis sets with semi-classical correction potentials. Substantial speed-ups of one to three orders of magnitude can be achieved while keeping the good DFT-D accuracy. A major focus was on accurate intra- and intermolecular geometries as this will in our view be a most important area for DFT applications. This hierarchy of methods is well suited for the every-day calculation on systems of modest to large size, but more importantly it will be able to make use of new developments in employing modern hardware, density fitting, or linear-scaling techniques as long as they can be mapped to a similar orbital expansion. We expect a significant impact of efficient DFT methodologies on the crystal structure prediction algorithms and on large scale material screenings in general. In a complementary multilevel treatment, the generated DFT structures can be further used to compute the desired property by any, possible more accurate, electronic structure method.

Acknowledgements
EC thanks his PhD supervisor Prof. Dr. Stefan Grimme for financial support who is himself appointed by the DFG in the framework of the "Gottfried-Wilhelm-Leibniz" prize. EC thanks Christoph Bannwarth and Fabian Bohle for proof reading the manuscript. JGB acknowledges support by the Alexander von Humboldt foundation within the Feodor-Lynen program.

References

References

[1] S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy. The high-throughput highway to computational materials design. *Nat. Mat.*, 12(3):191–201, March 2013.

[2] N. Marzari. The frontiers and the challenges. *Nat. Mat.*, 15(4):381–382, April 2016.

[3] R. Sure, J. Antony, and S. Grimme. Blind prediction of binding affinities for charged supramolecular hostguest systems: Achievements and shortcomings of dft-d3. *J. Phys. Chem. B*, 118:3431–3440, 2014.

[4] J. Yin, N. M. Henriksen, D. R. Slochower, M. R. Shirts, M. W. Chiu, D. L. Mobley, and

M. K. Gilson. Overview of the sampl5 host–guest challenge: Are we doing better? *J. Comput. Aided Mol. Des.*, 31:1–19, Jan 2017.

[5] D. A. Bardwell, C. S. Adjiman, Y. A. Arnautova, E. Bartashevich, S. X. M. Boerrigter, D. E. Braun, A. J. Cruz-Cabeza, G. M. Day, R. G. Della Valle, G. R. Desiraju, B. P. van Eijck, J. C. Facelli, M. B. Ferraro, D. Grillo, M. Habgood, D. W. M. Hofmann, F. Hofmann, K. V. J. Jose, P. G. Karamertzanis, A. V. Kazantsev, J. Kendrick, L. N. Kuleshova, F. J. J. Leusen, A. V. Maleev, A. J. Misquitta, S. Mohamed, R. J. Needs, M. A. Neumann, D. Nikylov, A. M. Orendt, R. Pal, C. C. Pantelides, C. J. Pickard, L. S. Price, S. L. Price, H. A. Scheraga, J. van de Streek, T. S. Thakur, S. Tiwari, E. Venuti, and I. K. Zhitkov. Towards crystal structure prediction of complex organic compounds – a report on the fifth blind test. *Acta Cryst. B*, 67:535–551, November 2011.

[6] Reilly, A. M. *et al.* Report on the sixth blind test of organic crystal structure prediction methods. *Acta Cryst. B*, 72(4):439–459, 2016.

[7] S. L. Price and S. M. Reutzel-Edens. The potential of computed crystal energy landscapes to aid solid-form development. *Drug Discov. Today*, 21:912–923, 2016.

[8] P. J. Bygrave, N. L. Allan, and F. R. Manby. The embedded many-body expansion for energetics of molecular crystals. *J. Chem. Phys.*, 137(16):164102, 2012.

[9] O. Masur, M. Schütz, L. Maschio, and D. Usvyat. Fragment-based direct-local-ring-coupled-cluster doubles treatment embedded in the periodic hartreefock solution. *J. Chem. Theory Comput.*, 12(10):5145–5156, 2016.

[10] G. J. O. Beran, S. Wen, K. Nand, Y. Huang, and Y. Heit. Accurate and Robust Molecular Crystal Modeling Using Fragment-Based Electronic Structure Methods. *Top. Curr. Chem.*, 345:59–93, 2014.

[11] C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, and F. Neese. Sparse maps – a systematic infrastructure for reduced-scaling electronic structure methods. ii. linear scaling domain based pair natural orbital coupled cluster theory. *J. Chem. Phys.*, 144:024109, 2016.

[12] S. A. Maurer, D. S. Lambrecht, J. Kussmann, and C. Ochsenfeld. Efficient distance-including integral screening in linear-scaling møller-plesset perturbation theory. *J. Chem. Phys.*, 138(1):014101, 2013.

[13] M. Schütz, O. Masur, and D. Usvyat. Efficient and accurate treatment of weak pairs in local CCSD(T) calculations. II. Beyond the ring approximation. *J. Chem. Phys.*, 140(24):244107, 2014.

[14] C. Riplinger, B. Sandhoefer, A. Hansen, and F. Neese. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J. Chem. Phys.*, 139:134101, 2013.

[15] G. J. O. Beran. Modeling Polymorphic Molecular Crystals with Electronic Structure Theory. *Chem. Rev.*, 116:5567–5613, 2016.

[16] J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schütz, and G. K. L. Chan. Ab initio determination of the crystalline benzene lattice energy to sub-kilojoule/mole accuracy. *Science*, 345:640–643, 2014.

[17] A. Zen, S. Sorella, M. J. Gillan, A. Michaelides, and D. Alfè. Boosting the accuracy and speed of quantum Monte Carlo: size-consistency and time-step. . *Phys. Rev. B*, 93:241118(R), 2016.

[18] A. Zen, J. G. Brandenburg, J. Klimeš, A. Tkatchenko, D. Alfè, and A. Michaelides. Fast and accurate quantum monte carlo for molecular crystals. *Proc. Natl. Acad. Sci. U.S.A.*, 115:1724–1729, 2017.

[19] P. Nozières and D. Pines. Correlation Energy of a Free Electron Gas. *Phys. Rev.*, 111(2):442–454, July 1958.

[20] J. Harl and G. Kresse. Cohesive energy curves for noble gas solids calculated by adiabatic connection fluctuation-dissipation theory. *Phys. Rev. B*, 77(4):045136, January 2008.

[21] J. Harl and Georg Kresse. Accurate Bulk Properties from Approximate Many-Body Techniques. *Phys. Rev. Lett.*, 103(5):056401, July 2009.

[22] H. Eshuis, J. E. Bates, and F. Furche. Electron correlation methods based on the random phase approximation. *Theo. Chem. Acc.*, 131:1084, 2012.

[23] J. Wilhelm, P. Seewald, M. Del Ben, and J. Hutter. Large-scale cubic-scaling random phase approximation correlation energy calculations using a gaussian basis. *J. Chem. Theory Comput.*, 12:5851–5859, 2016.

[24] L. Schimka, J. Harl, A. Stroppa, A. Grüneis, M. Marsman, F. Mittendorfer, and G. Kresse. Accurate surface and adsorption energies from many-body perturbation theory. *Nat. Mater.*, 9(9):741–744, July 2010.

[25] X. Ren, A. Tkatchenko, P. Rinke, and M. Scheffler. Beyond the Random-Phase Approximation for the Electron Correlation Energy: The Importance of Single Excitations. *Phys. Rev. Lett.*, 106(15):153003, April 2011.

[26] J. Klimeš. Lattice energies of molecular solids from the random phase approximation with singles corrections. *J. Chem. Phys.*, 145(9):094506, September 2016.

[27] A. M. Burow, J. E. Bates, F. Furche, and H. Eshuis. Analytical first-order molecular properties and forces within the adiabatic connection random phase approximation. *J. Chem. Theory Comput.*, 10:180–194, 2014.

[28] B. Ramberger, T. Schäfer, and G. Kresse. Analytic interatomic forces in the random phase approximation. *Phys. Rev. Lett.*, 118:106403, 2017.

[29] Y. S. Al-Hamdani, M. Rossi, D. Alfè, T. Tsatsoulis, B. Ramberger, J. G. Brandenburg, A. Zen, G. Kresse, A. Grüneis, A Tkatchenko, and A. Michaelides. Properties of the water to boron nitride interaction: From zero to two dimensions with benchmark accuracy. *J. Chem. Phys.*, 147:044710, 2017.

[30] R. G. Parr and W. Yang. *Density-Functional Theory of Atoms and Molecules*. Oxford University Press, Oxford, 1989.

[31] W. Kohn. Electronic structure of matter — wave functions and density functionals. *Rev. Mod. Phys.*, 71:1253–1266, 1998.

[32] K. Burke. Perspective on density functional theory. *J. Chem. Phys.*, 136(15):150901, 2012.

[33] J. Sun, A. Ruzsinszky, and J. P. Perdew. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.*, 115:036402, 2015.

[34] N. Mardirossian and M. Head-Gordon. wb97x-v: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.*, 16:9904–9924, 2014.

[35] Y. Wang, X. Jin, H. S. Yu, D. G. Truhlar, and X. He. Revised m06-l functional for improved accuracy on chemical reaction barrier heights, noncovalent interactions, and solid-state physics. *Proc Natl Acad Sci*, 114(32):8487–8492, 2017.

[36] A. J. Stone. *The Theory of Intermolecular Forces*. Oxford University Press, Oxford, 1997.

[37] S. Grimme, A. Hansen, J. G. Brandenburg, and C. Bannwarth. Dispersion-corrected mean-field electronic structure methods. *Chem. Rev.*, 116:5105–5154, 2016.

[38] L. A. Burns, A. Vazquez-Mayagoitia, B. G. Sumpter, and C. D. Sherrill. A comparison of dispersion corrections (dft-d), exchange-hole dipole moment (xdm) theory, and specialized functionals. *J. Chem. Phys.*, 134:084107, 2011.

[39] J. Klimeš and A. Michaelides. Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory. *J. Chem. Phys.*, 137:120901, 2012.

[40] Lars Goerigk, Andreas Hansen, Christoph Alexander Bauer, Stephan Ehrlich, Asim Najibi, and Stefan Grimme. A look at the density functional theory zoo with the advanced gmtkn55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.*, 2017. DOI: 10.1039/C7CP04913G.

[41] J. G. Brandenburg, J. E. Bates, J. Sun, and J. P. Perdew. Benchmark tests of a strongly constrained semilocal functional with a long-range dispersion correction. *Phys. Rev. B*, 94:115144, 2016.

[42] J. A. Pople. *Modern Theoretical Chemistry*, chapter 1. Plenum, New York, vol. 4. edition, 1976.

[43] E. R. Davidson and D. Feller. Basis set selection for molecular calculations. *Chem. Rev.*, 86(4):681–696, 1986.

[44] W. Kołos. Possible improvements of the interaction energy calculated using minimal basis sets. *Theor. Chim. Acta*, 51(3):219–240, 1979.

[45] E. N. Brothers and G. E. Scuseria. Parametrization of atomic energies to improve small basis set density functional thermochemistry. *J. Chem. Theory Comp.*, 2:1045, 2006.

[46] R. D. Adamson, P. M. W. Gill, and J. A. Pople. Empirical density functionals. *Chem. Phys. Lett.*, 284:6–11, 1998.

[47] R. Sure and S. Grimme. Corrected small basis set hartree-fock method for large systems. *J. Comput. Chem.*, 34:1672–1685, 2013.

[48] S. Grimme, J. G. Brandenburg, C. Bannwarth, and A. Hansen. Consistent structures and interactions by density functional theory with small atomic orbital basis sets. *J. Chem. Phys.*, 143:054107, 2015.

[49] J. G. Brandenburg, E. Caldeweyher, and S. Grimme. Screened exchange hybrid density functional for accurate and efficient structures and interaction energies. *Phys. Chem. Chem. Phys.*, 18:15519–15523, 2016.

[50] J. G. Brandenburg, C. Bannwarth, A. Hansen, and S. Grimme. B97-3c: A revised low-cost variant of the b97-d density functional method. *J. Chem. Phys.*, 148:064104, 2018.

[51] J. Hostaš and J. Řezáč. Accurate dft-d3 calculations in a small basis set. *J. Chem. Theory Comput.*, 13:3575–3585, 2017.

[52] J. Witte, J. B. Neaton, and M. Head-Gordon. Effective empirical corrections for basis set superposition error in the def2-svpd basis: gcp and dft-c. *J. Chem. Phys.*, 146:234105, 2017.

[53] A. Otero-de-la Roza and Gino A. DiLabio. Transferable atom-centered potentials for the correction of basis set incompleteness errors in density-functional theory. *J. Chem. Theory Comput.*, 13:3505–3524, 2017.

[54] K. Miyamoto, T. F. Miller, and F. R. Manby. Fock-matrix corrections in density functional theory and use in embedded mean-field theory. *J. Chem. Theory Comput.*, 12:5811–5822, 2016.

[55] W. Kohn. Nobel lecture: Electronic structure of matterwave functions and density functionals. *Rev. Mod. Phys.*, 71:1253–1266, 1999.

[56] J. P. Perdew and K. Schmidt. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conf. Proc.*, 577(1):1–20, 2001.

[57] U. von Barth and L. Hedin. A local exchange-correlation potential for the spin polarized case. *J. Phys. C*, 5:1629, 1972.

[58] S. B. Trickey. Approximate electron removal energies in density-functional theory from post-hoc correction of local-spin-density eigenvalues. *Phys. Rev. Lett.*, 56:881–881, 1986.

[59] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, 1996. erratum *Phys. Rev. Lett.* **78**, 1396 (1997).

[60] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38:3098–3100, 1988.

[61] C. Lee, W. Yang, and R. G. Parr. *Phys. Rev. B*, 37:785–789, 1988.

[62] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria. Climbing the density functional ladder: Nonempirical meta generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.*, 91:146401, 2003.

[63] Y. Zhao and D. G. Truhlar. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.*, 125(19):194101, 2006.

[64] R. Peverati and D. G. Truhlar. M11-L: A Local Density Functional That Provides Improved Accuracy for Electronic Structure Calculations in Chemistry and Physics. *J. Phys. Chem. Lett.*, 3(1):117–124, 2012.

[65] R. Peverati and D. G. Truhlar. An improved and broadly accurate local approximation to the exchange-correlation density functional: The MN12-L functional for electronic structure calculations in chemistry and physics. *Phys. Chem. Chem. Phys.*, 14(38):13171–13174, 2012.

[66] N. Mardirossian and M. Head-Gordon. Mapping the genome of meta-generalized gradient approximation density functionals: The search for b97m-v. *J. Chem. Phys.*, 142(7):074111, 2015.

[67] J. M. del Campo, J. L. Gázquez, S. B. Trickey, and A. Vela. A new meta-GGA exchange functional based on an improved constraint-based GGA. *Chem. Phys. Lett.*, 543:179–183, 2012.

[68] J. Sun, B. Xiao, and A. Ruzsinszky. Communication: Effect of the orbital-overlap dependence in the meta generalized gradient approximation. *J. Chem. Phys.*, 137(5):051101, 2012.

[69] J. Sun, R. Haunschild, B. Xiao, I. W. Bulik, G. E. Scuseria, and J. P. Perdew. Semilocal and hybrid meta-generalized gradient approximations based on the understanding of the kinetic-energy-density dependence. *J. Chem. Phys.*, 138(4):044113, 2013.

[70] J. Sun, J. P. Perdew, and A. Ruzsinszky. Semilocal density functional obeying a strongly tightened bound for exchange. *Proc. Natl. Acad. Sci. U.S.A.*, 112:685, 2015.

[71] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz. Density-functional theory for fractional particle number: Derivative discontinuities of the energy. *Phys. Rev. Lett.*, 49:1691–1694, 1982.

[72] A. D. Becke. A new mixing of hartree–fock and local density-functional theories. *J. Chem. Phys.*, 98:1372–1377, 1993.

[73] C. Adamo and V. Barone. Toward reliable density functional methods without adjustable parameters: The pbe0 model. *J. Chem. Phys.*, 110:6158–6170, 1999.

[74] A. D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. *J. Chem. Phys.*, 98:5648–5652, 1993.

[75] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.*, 98:11623–11627, 1994.

[76] J. Heyd, G. E. Scuseria, and M. Ernzerhof. Hybrid functionals based on a screened coulomb potential. *J. Chem. Phys.*, 118(18):8207–8215, 2003.

[77] E. J. Baerends, O. V. Gritsenko, and R. Van Meer. The Kohn–Sham gap, the fundamental gap and the optical gap: the physical meaning of occupied and virtual Kohn–Sham orbital energies. *Phys. Chem. Chem. Phys.*, 15(39):16408–16425, 2013.

[78] T. Yanai, D. P. Tew, and N. C. Handy. A new hybrid exchange–correlation functional using the coulomb-attenuating method (cam-b3lyp). *Chem. Phys. Lett.*, 393(1):51–57, 2004.

[79] J.-D. Chai and M. Head-Gordon. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.*, 10(44):6615–6620, 2008.

[80] Y.-S. Lin, G.-D. Li, S.-P. Mao, and J.-D. Chai. Long-range corrected hybrid density functionals with improved dispersion corrections. *J. Chem. Theory Comput.*, 9(1):263–272, 2012.

[81] N. Mardirossian and M. Head-Gordon. ωB97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.*, 16(21):9904–9924, 2014.

[82] J. C. Slater. *The Self Consistent Field for Molecules and Solids.* McGraw-Hill, Singapore, 1974.

[83] D. J. Chadi and M. L. Cohen. Special points in the brillouin zone. *Phys. Rev. B*, 8:5747–5753, 1973.

[84] X. Andrade, D. Strubbe, U. De Giovannini, A. H. Larsen, M. J. T. Oliveira, J. Alberdi-Rodriguez, A. Varas, I. Theophilou, N. Helbig, M. J. Verstraete, L. Stella, F. Nogueira, A. Aspuru-Guzik, A. Castro, M. A. L. Marques, and A. Rubio. Real-space grids and the octopus code as tools for the development of new simulation approaches for electronic systems. *Phys. Chem. Chem. Phys.*, 17(47):31371–31396, 2015.

[85] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler. Ab initio molecular simulations with numeric atom-centered orbitals. *Comput. Phys. Commun.*, 180(11):2175–2196, 2009.

[86] X. Gonze, J. M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G. M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, Ph. Ghosez, J. Y. Raty, and D. C. Allan. First-principles computation of material properties: the ABINIT software project. *Comput. Mater. Sci.*, 25(3):478–492, 2002.

[87] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, and T. Ziegler. Chemistry with adf. *J. Comp. Chem.*, 22:931–967, 2001.

[88] F. J. Rico, R. Lpez, A. Aguado, I. Ema, and G. Ramrez. New program for molecular calculations with slater-type orbitals. *Int. J. Quantum Chem.*, 81(2):148–153, 2001.

[89] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Gaussian16 Revision A.03, 2016. Gaussian Inc. Wallingford CT.

[90] TURBOMOLE V7.2 2017, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from `http://www.turbomole.com`.

[91] Frank Neese. The ORCA program system. *Wiley Interdiscip. Rev.-Comput. Mol. Sci.*, 2(1):73–78, 2012.

[92] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele. CP2K: atomistic simulations of condensed matter systems. *Wiley Interdiscip. Rev.-Comput. Mol. Sci.*, 4(1):15–25, 2014.

[93] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen. Real-space grid implementation of the projector augmented wave method. *Phys. Rev. B*, 71:035109, 2005.

[94] P. E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50:17953, 1994.

[95] G. Kresse and J Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59:1758, 1999.

[96] J. Furthmüller, P. Käckell, F. Bechstedt, and G. Kresse. Extreme softening of vanderbilt pseudopotentials: General rules and case studies of first-row and d-electron elements. *Phys. Rev. B*, 61:4576–4587, 2000.

[97] G. Kresse and J. Hafner. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B*, 47:558, 1993.

[98] G. Kresse and J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *J. Comp. Mat. Sci.*, 6:15, 1996.

[99] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169–11186, 1996.

[100] G Kresse and J Hafner. Norm-conserving and ultrasoft pseudopotentials for first-row and transition elements. *J. Phys. Condens. Matter*, 6(40):8245, 1994.

[101] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59:1758–1775, 1999.

[102] J. G. Brandenburg, S. Grimme, P. G. Jones, G. Markopoulos, H. Hopf, M. K. Cyranski, and D. Kuck. Unidirectional molecular stacking of tribenzotriquinacenes in the solid state: A combined x-ray and theoretical study. *Chem. Eur. J.*, 19:9930–9938, 2013.

[103] H. Kruse and S. Grimme. A geometrical correction for the inter- and intra-molecular basis set superposition error in hartree-fock and density functional theory calculations for large systems. *J. Chem. Phys.*, 136(15):154101, 2012.

[104] J. G. Brandenburg, M. Alessio, B. Civalleri, M. Peintinger, T. Bredow, and S. Grimme. Geometrical correction for the inter- and intramolecular basis set superposition error in periodic density functional theory calculations. *J. Phys. Chem. A*, 117:9282–9292, 2013.

[105] S. F. Boys and F. Bernardi. The calculation of small molecular interactions by the differences of separate total energies. some procedures with reduced errors. *Mol. Phys.*, 19:553–566, 1970.

[106] A. Zangwill and Paul Soven. Density-functional approach to local-field effects in finite systems: Photoabsorption in the rare gases. *Phys. Rev. A*, 21:1561–1572, 1980.

[107] R. Eisenschitz and F. London. Über das Verhältnis der van der Waalsschen Kräfte zu den homöopolaren Bindungskräften. *Z. Phys.*, 60:491–527, 1930.

[108] H. B. G. Casimir and D. Polder. The influence of retardation on the london-van der waals forces. *Phys. Rev.*, 73:360–372, 1948.

[109] A. Tkatchenko and M. Scheffler. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.*, 102:073005, 2009.

[110] R. A. DiStasio, V. V. Gobre, and A. Tkatchenko. Many-body van der waals interactions in molecules and condense matter. *J. Phys.: Condens. Matter*, 26:213202, 2014.

[111] A. D. Becke and E. R. Johnson. Exchange-hole dipole moment and the dispersion interaction. *J. Chem. Phys.*, 122:154104, 2005.

[112] A. D. Becke and E. R. Johnson. A density-functional model of the dispersion interaction. *J. Chem. Phys.*, 123:154101, 2005.

[113] A. D. Becke and E. R. Johnson. Exchange-hole dipole moment and the dispersion interaction: High-order dispersion coefficients. *J. Chem. Phys.*, 124:014104, 2006.

[114] T. Sato and H. Nakai. Density functional method including weak interactions: Dispersion coefficients based on the local response approximation. *J. Chem. Phys.*, 131:224104, 2009.

[115] T. Sato and H. Nakai. Local Response Dispersion Method. II. Generalized Multicenter Interactions. *J. Chem. Phys.*, 133:194101, 2010.

[116] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist. Van der Waals Density Functional for General Geometries. *Phys. Rev. Lett.*, 92:246401, 2004.

[117] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist. Erratum: Van der Waals Density Functional for General Geometries [Phys. Rev. Lett. **92**, 246401 (2004)]. *Phys. Rev. Lett.*, 95:109902, 2005.

[118] K. Berland and P. Hyldgaard. Exchange functional that tests the robustness of the plasmon description of the van der Waals density functional. *Phys. Rev. B*, 89:035412, 2014.

[119] D. C. Langreth, B. I. Lundqvist, S. D. Chakarova-Käck, V. R. Cooper, M. Dion, P. Hyldgaard, A. Kelkkanen, J. Kleis, L. Kong, and S. Li. A density functional for sparse matter. *J. Phys.: Condens. Matter*, 21:084203, 2009.

[120] K. Berland, V. R. Cooper, K. Lee, E. Schröder, T. Thonhauser, P. Hyldgaard, and B. I. Lundqvist. van der Waals forces in density functional theory: a review of the vdW-DF method. *Rep. Prog. Phys.*, 78:066501, 2015.

[121] O. A. Vydrov and T. Van Voorhis. Improving the accuracy of the nonlocal van der Waals density functional with minimal empiricism. *J. Chem. Phys.*, 130:104105, 2009.

[122] O. A Vydrov and T. Van Voorhis. Nonlocal van der Waals Density Functional Made Simple. *Phys. Rev. Lett.*, 103:063004, 2009.

[123] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *J. Chem. Phys.*, 132:154104, 2010.

[124] H. B. G. Casimir and D. Polder. The Influence of Retardation on the London-van der Waals Forces. *Phys. Rev.*, 73:360–372, 1948.

[125] B. M. Axilrod and E. Teller. Interaction of the van der Waals Type Between Three Atoms. *J. Chem. Phys.*, 11:299–300, 1943.

[126] Y. Muto. Force between nonpolar molecules. *Proc. Phys. Soc. Jpn.*, 17:629, 1943.

[127] J.-D. Chai and M. Head-Gordon. Long-range corrected double-hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.*, 10:6615–6620, 2000.

[128] E. Caldeweyher, C. Bannwarth, and S. Grimme. Extension of the d3 dispersion coefficient model. *J. Chem. Phys.*, 147(3):034112, 2017.

[129] A. D. Becke. Density functional thermochemistry. v. systematic optimization of exchange-correlation functionals. *J. Chem. Phys.*, 107:8554–8560, 1997.

[130] M. Korth, M. Pitoňák, J. Řezáč, and P. Hobza. A transferable h-bonding correction for semiempirical quantum-chemical methods. *J. Chem. Theory Comput.*, 6:344–352, 2010.

[131] V. M. Miriyala and J. Řezáč. Description of non-covalent interactions in scc-dftb methods. *J. Comput. Chem.*, 38(10):688–697, 2017.

[132] M. Cutini, B. Civalleri, M. Corno, R. Orlando, J. G. Brandenburg, L. Maschio, and P. Ugliengoa. Assessment of different quantum mechanical methods for the prediction of structure and cohesive energy of molecular crystals. *J. Chem. Theory Comput.*, 12:3340–3352, 2016.

[133] F. Weigend and R. Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, 7:3297–3305, 2005.

[134] W. J. Hehre, R. Ditchfield, and J. A. Pople. Selfconsistent molecular orbital methods. xii. further extensions of gaussiantype basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.*, 56:2257–2261, 1972.

[135] P. Jurečka, J. Šponer, J. Cerny, and P. Hobza. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.*, 8:1985–1993, 2006.

[136] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke. Restoring the density-gradient expansion for exchange in solids and surfaces. *Phys. Rev. Lett.*, 100:136406, 2008.

[137] Y. Zhang and W. Yang. Comment on "generalized gradient approximation made simple". *Phys. Rev. Lett.*, 80:890–890, 1998.

[138] L. Goerigk and S. Grimme. Efficient and accurate double-hybrid-meta-gga density functionals-evaluation with the extended gmtkn30 database for general main group thermochemistry, kinetics, and noncovalent interactions. *J. Chem. Theory Comput.*, 7:291–309, 2011.

[139] T. M. Henderson, B. G. Janesko, and G. E. Scuseria. Generalized gradient approximation model exchange holes for range-separated hybrids. *J. Chem. Phys.*, 128(19):194105, 2008.

[140] T. M. Henderson, B. G. Janesko, and G. E. Scuseria. Range Separation and Local Hybridization in Density Functional Theory. *J. Phys. Chem. A*, 112(49):12530–12542, 2008.

[141] J. Heyd, G. E. Scuseria, and M. Ernzerhof. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.*, 118(18):8207–8215, 2003.

[142] B. G. Janesko, T. M. Henderson, and G. E. Scuseria. Screened hybrid density functionals for solid-state chemistry and physics. *Phys. Chem. Chem. Phys.*, 11(3):443–454, 2009.

[143] A. D. Becke. Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals. *J. Chem. Phys.*, 107(20):8554–8560, 1997.

[144] H. L. Schmider and A. D. Becke. Optimized density functionals from the extended g2 test set. *J. Chem. Phys.*, 108(23):9624–9631, 1998.

[145] Stefan Grimme, Christoph Bannwarth, and Philip Shushkov. A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (z = 1-86). *J. Chem. Theory Comput.*, 13(5):1989–2009, 2017.

[146] A. Leif Hickey and Christopher N. Rowley. Benchmarking quantum chemical methods for the calculation of molecular dipole moments and polarizabilities. *J. Phys. Chem. A*, 118(20):3678–3687, 2014.

[147] 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. CRYSTAL14: A Program for the Ab Initio Investigation of Crystalline Solids. *Int. J. Quantum Chem.*, 114:1287–1317, 2014.

[148] R. Orlando, M. De La Pierre, C. M. Zicovich-Wilson, A. Erba, and R. Dovesi. On the full exploitation of symmetry in periodic (as well as molecular) self-consistent-field ab initio calculations. *J. Chem. Phys.*, 141:104108, 2014.

[149] A. Erba, J. Baima, I. Bush, R. Orlando, and R. Dovesi. Large-scale condensed matter dft simulations: Performance and capabilities of the crystal code. *J. Chem. Theory Comput.*, 13:5019–5027, 2017.

[150] D. R. Bowler and T. Miyazaki. $\mathcal{O}(N)$ methods in electronic structure calculations. *Rep. Prog. Phys.*, 75(3):036503, 2012.

[151] F. Neese. The orca program system. *WIREs Comput. Mol. Sci.*, 2:73–78, 2012.

[152] F. Furche, R. Ahlrichs, C. Httig, W. Klopper, M. Sierka, and F. Weigend. Turbomole. *WIREs Comput Mol Sci*, 4:91–100, 2014.

[153] R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Yu. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill. Psi4 1.1: An open-source electronic structure program emphasizing automation, advanced libraries, and interoperability. *J. Chem. Theory. Comput.*, 13:3185–3197, 2017.

[154] The CP2K developers group, 2015. CP2K is freely available from http://www.cp2k.org/.

[155] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele. cp2k: atomistic simulations of condensed matter systems. *WIREs Comput Mol Sci*, 4:15–25, 2014.

[156] The implementation of B97-3c was done by Dr. Eva Perlt from the University of Bonn. A slightly different basis set is employed, details will be published elsewhere.

[157] Jörg Kussmann and Christian Ochsenfeld. Pre-selective screening for matrix elements in linear-scaling exact exchange calculations. *J. Chem. Phys.*, 138(13):134114, 2013.

[158] Ivan S. Ufimtsev and Todd J. Martinez. Quantum chemistry on graphical processing units. 3. analytical energy gradients, geometry optimization, and first principles molecular dynamics. *J. Chem. Theory Comput.*, 5:2619–2628, 2009.

[159] Nathan Luehr, Ivan S. Ufimtsev, and Todd J. Martinez. Dynamic Precision for Electron Repulsion Integral Evaluation on Graphical Processing Units (GPUs). *J. Chem. Theory Comput.*, 7:949–954, 2011.

[160] R. Dovesi, C. Pisani, C. Roetti, and V. R. Saunders. Treatment of coulomb interactions in hartree-fock calculations of periodic systems. *Phys. Rev. B*, 28:5781, 1983.

[161] V. R. Saunders, C. Freyria-Fava, R. Dovesi, L. Salasco, and C. Roetti. On the electrostatic potential in crystalline systems where the charge density is expanded in gaussian functions. *Mol. Phys.*, 77:629, 1992.

[162] K. Eichkorn, O. Treutler, H. Oehm, M. Haeser, and R. Ahlrichs. Auxiliary basis sets to approximate coulomb potential. *Chem. Phys. Lett.*, 240:283, 1995.

[163] M. Sierka, A. Hogekamp, and R. Ahlrichs. Fast evaluation of the coulomb potential for electron densities using multipole accelerated resolution of identity approximation. *J. Chem. Phys.*, 118:9136, 2003.

[164] F. Neese. An improvement of the resolution of the identity approximation for the calculation of the coulomb matrix. *J. Comp. Chem.*, 24:1740–1747, 2003.

[165] R. Lazarski, A. M. Burow, and M. Sierka. Density functional theory for molecular and periodic systems using density fitting and continuous fast multipole methods. *J. Chem. Theory Comput.*, 11:3029–3041, 2015.

[166] R. Lazarski, A. M. Burow, L. Grajciar, and M. Sierka. Density functional theory for molecular and periodic systems using density fitting and continuous fast multipole methods: Analytical gradients. *J. Comput. Chem.*, 57:2518–2526, 2016.

[167] L. Grajciar. Low-memory iterative density fitting. *J. Comput. Chem.*, 36:1521–1535, 2015.

[168] S. Ehrlich, A. H. Göller, and S. Grimme. Towards full quantum-mechanics-based protein-ligand binding affinities. *ChemPhysChem*, 18:898–905, 2017.

[169] M. A. Neumann, J. van de Streek, F. P. A. Fabbiani, P. Hidber, and O. Grassmann. Combined crystal structure prediction and high-pressure crystallization in rational pharmaceutical polymorph screening. *Nature Communications*, 6:1463–1473, 2015.

[170] Marcus A. Neumann. Tailor-made force fields for crystal-structure prediction. *J. Phys. Chem. B*, 112:9810–9829, 2008.

[171] Marcus A. Neumann and Marc-Antoine Perrin. Energy ranking of molecular crystals using density functional theory calculations and an empirical van der waals correction. *J.Phys. Chem. B*, 109:15531–15541, 2005.

[172] J. G. Brandenburg and S. Grimme. Organic crystal polymorphism: A benchmark for dispersion corrected mean field electronic structure methods. *Acta Cryst. B*, 72:502–513, 2016.

[173] Angeles Pulido, Linjiang Chen, Tomasz Kaczorowski, Daniel Holden, Marc A Little, Samantha Y Chong, Benjamin J Slater, David P McMahon, Baltasar Bonillo, Chloe J Stackhouse, Andrew Stephenson, Christopher M Kane, Rob Clowes, Tom Hasell, Andrew I Cooper, and Graeme M Day. Functional materials discovery using energy–structure–function maps. *Nature*, 543(7647):657–664, 2017.

[174] A. J. Cruz-Cabeza, S. M. Reutzel-Edens, and J. Bernstein. Facts and fictions about polymorphism. *Chem. Soc. Rev.*, 44:8619–8635, 2015.

[175] J. G. Brandenburg, T. Maas, and S. Grimme. Benchmarking dft and semiempirical methods on structures and lattice energies for ten ice polymorphs. *J. Chem. Phys.*, 142:124104, 2015.

[176] A. Otero-de-la-Roza and E. R. Johnson. A benchmark for non-covalent interactions in solids. *J. Chem. Phys.*, 137:054103, 2012.

[177] A. M. Reilly and A. Tkatchenko. Understanding the role of vibrations, exact exchange, and many-body van der waals interactions in the cohesive properties of molecular crystals. *J. Chem. Phys.*, 139:024705, 2013.

[178] S. Wen and G. J. O. Beran. Accurate Molecular Crystal Lattice Energies from a Fragment QM/MM Approach with On-the-Fly Ab Initio Force Field Parametrization. *J. Chem. Theory Comput.*, 7(11):3733–3742, 2011.

[179] A. Ambrosetti, A. M. Reilly, R. A. DiStasio, and A. Tkatchenko. Long-range correlation energy calculated from coupled atomic response functions. *J. Chem. Phys.*, 140:18A508, 2014.

[180] J. S. Chickos. Enthalpies of Sublimation after a Century of Measurement: A View as Seen through the Eyes of a Collector. *Netsu Sokutei*, 30:116, 2003.

[181] O. Ermer, S. A. Mason, F. A. L. Anet, and S. S. Miura. Ultrashort nonbonded hydrogen... hydrogen distance in a half-cage pentacyclododecane. *J. Am. Chem. Soc.*, 107(8):2330–2334, 1985.

[182] N. Stojilovic. Why cant we see hydrogen in x-ray photoelectron spectroscopy? *J. Chem. Educ.*, 89(10):1331–1332, 2012.

[183] V. L. Deringer, V. Hoepfner, and R. Dronskowski. Accurate hydrogen positions in organic crystals: Assessing a quantum-chemical aide. *Cryst. Growth Des*, 12:1014–1021, 2012.

[184] S. Rösel, H. Quanz, C. Logemann, J. Becker, E. Mossou, L. Cañadillas-Delgado, E. Caldeweyher, S. Grimme, and P. R. Schreiner. London Dispersion Enables the Shortest Intermolecular Hydrocarbon H···H Contact. *J. Am. Chem. Soc.*, 139(22):7428–7431, 2017.

[185] C. C. Sun. Thermal expansion of organic crystals and precision of calculated crystal density: A survey of cambridge crystal database. *J. Pharm. Sci.*, 96(5):1043–1052, 2007.

[186] P. Su and H. Li. Energy decomposition analysis of covalent bonds and intermolecular interactions. *J. Chem. Phys.*, 131(1):014102, 2009.

[187] M Elstner, D Porezag, G Jungnickel, J Elsner, M Haugk, T Frauenheim, S Suhai, and G Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*, 58:7260–7268, 1998.

[188] B. Aradi, B. Hourahine, and Th. Frauenheim. Dftb+, a sparse matrix-based implementation of the dftb method. *J. Phys. Chem. A*, 111(26):5678–5684, 2007.

[189] J. G. Brandenburg and S. Grimme. Accurate modeling of organic molecular crystals by dispersion-corrected density functional tight binding (dftb). *J. Phys. Chem. Lett.*, 5:1785–1789, 2014.

[190] A. Priimagi, G. Cavallo, P. Metrangolo, and G. Resnati. The Halogen Bond in the Design of Functional Supramolecular Materials: Recent Advances. *Acc. Chem. Res.*, 46(11):2686–2695, 2013.

[191] A. Mukherjee, S. Tothadi, and G. R. Desiraju. Halogen Bonds in Crystal Engineering: Like Hydrogen Bonds yet Different. *Acc. Chem. Res.*, 47(8):2514–2524, 2014.

[192] E. Persch, O. Dumele, and F. Diederich. Molecular recognition in chemical and biological systems. *Angew. Chem. Int. Ed.*, 54(11):3290–3327, 2015.

[193] L. P. Wolters, P. Schyman, M. J. Pavan, W. L. Jorgensen, F. M. Bickelhaupt, and S. Kozuch. The many faces of halogen bonding: a review of theoretical models and methods. *Wiley Interdiscip. Rev.-Comput. Mol. Sci.*, 4(6):523–540, 2014.

[194] S. M. Huber, E. Jimenez-Izal, J. M. Ugalde, and I. Infante. Unexpected trends in halogen-bond based noncovalent adducts. *Chem. Commun.*, 48(62):7708–7710, 2012.

[195] S. V. Rosokha and M. K. Vinakos. Hybrid Network Formation via Halogen Bonding of the Neutral Bromo-Substituted Organic Molecules with Anionic MetalBromide Complexes. *Cryst. Growth Des*, 12(8):4149–4156, 2012.

[196] J. Řezáč. On the role of charge transfer in halogen bonding. *Phys. Chem. Chem. Phys.*, 19(1):791–803, 2017.

[197] C. B. Aakeröy, T. K. Wijethunga, J. Desper, and M. aković. Electrostatic Potential Differences and Halogen-Bond Selectivity. *Cryst. Growth Des*, 16(5):2662–2670, 2016.

[198] F. Topić and K. Rissanen. Systematic Construction of Ternary Cocrystals by Orthogonal and Robust Hydrogen and Halogen Bonds. *J. Am. Chem. Soc.*, 138(20):6610–6616, 2016.

[199] K. Gao and N. S. Goroff. Two New Iodine-Capped Carbon Rods. *J. Am. Chem. Soc.*, 122(38):9320–9321, 2000.

[200] M. T. Messina, P. Metrangolo, W. Panzeri, T. Pilati, and G. Resnati. Intermolecular recognition between hydrocarbon oxygen-donors and perfluorocarbon iodine-acceptors: the shortest OI non-covalent bond. *Tetrahedron*, 57(40):8543–8550, 2001.

[201] H. Y. Gao, Q. J. Shen, X. R. Zhao, X. Q. Yan, X. Pang, and W. J. Jin. Phosphorescent co-crystal assembled by 1{,}4-diiodotetrafluorobenzene with carbazole based on C-I[three dots{,} centered][small pi] halogen bonding. *J. Mater. Chem.*, 22(12):5336–5343, 2012.

[202] T. A. Schaub, R. Sure, F. Hampel, S. Grimme, and M. Kivala. Quantum chemical dissection of the shortest p=oi halogen bond: The decisive role of crystal packing effects. *Chem. Eur. J.*, 23(24):5687–5691, 2017.