A geometrical correction for the inter- and intra-molecular basis set superposition error in Hartree-Fock and density functional theory calculations for large systems

ARTICLE *in* THE JOURNAL OF CHEMICAL PHYSICS · APRIL 2012

Impact Factor: 2.95 · DOI: 10.1063/1.3700154 · Source: PubMed

2 AUTHORS:

Holger Kruse
Institute of Biophysics ASCR
17 PUBLICATIONS **654** CITATIONS
SEE PROFILE

Stefan Grimme
University of Bonn
475 PUBLICATIONS **31,053** CITATIONS
SEE PROFILE

A geometrical correction for the inter- and intra-molecular basis set superposition error in Hartree-Fock and density functional theory calculations for large systems

Holger Kruse and Stefan Grimme

Citation: J. Chem. Phys. **136**, 154101 (2012); doi: 10.1063/1.3700154
View online: http://dx.doi.org/10.1063/1.3700154
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v136/i15
Published by the AIP Publishing LLC.

A geometrical correction for the inter- and intra-molecular basis set superposition error in Hartree-Fock and density functional theory calculations for large systems

Holger Kruse[1] and Stefan Grimme[2,a)]
[1]*Theoretische Organische Chemie, Organisch-Chemisches Institut der Universität Münster, Corrensstr. 40, D-48149 Münster, Germany*
[2]*Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie der Universität Bonn, Beringstr. 4, D-53115 Bonn, Germany*

(Received 9 February 2012; accepted 15 March 2012; published online 16 April 2012)

A semi-empirical counterpoise-type correction for basis set superposition error (BSSE) in molecular systems is presented. An atom pair-wise potential corrects for the inter- and intra-molecular BSSE in supermolecular Hartree-Fock (HF) or density functional theory (DFT) calculations. This geometrical counterpoise (gCP) denoted scheme depends only on the molecular geometry, i.e., no input from the electronic wave-function is required and hence is applicable to molecules with ten thousands of atoms. The four necessary parameters have been determined by a fit to standard Boys and Bernadi counterpoise corrections for Hobza's S66×8 set of non-covalently bound complexes (528 data points). The method's target are small basis sets (e.g., minimal, split-valence, 6-31G*), but reliable results are also obtained for larger triple-ζ sets. The intermolecular BSSE is calculated by gCP within a typical error of 10%–30% that proves sufficient in many practical applications. The approach is suggested as a quantitative correction in production work and can also be routinely applied to estimate the magnitude of the BSSE beforehand. The applicability for biomolecules as the primary target is tested for the crambin protein, where gCP removes intramolecular BSSE effectively and yields conformational energies comparable to def2-TZVP basis results. Good mutual agreement is also found with Jensen's ACP(4) scheme, estimating the intramolecular BSSE in the phenylalanine-glycine-phenylalanine tripeptide, for which also a relaxed rotational energy profile is presented. A variety of minimal and double-ζ basis sets combined with gCP and the dispersion corrections DFT-D3 and DFT-NL are successfully benchmarked on the S22 and S66 sets of non-covalent interactions. Outstanding performance with a mean absolute deviation (MAD) of 0.51 kcal/mol (0.38 kcal/mol after D3-refit) is obtained at the gCP-corrected HF-D3/(minimal basis) level for the S66 benchmark. The gCP-corrected B3LYP-D3/6-31G* model chemistry yields MAD=0.68 kcal/mol, which represents a huge improvement over plain B3LYP/6-31G* (MAD=2.3 kcal/mol). Application of gCP-corrected B97-D3 and HF-D3 on a set of large protein-ligand complexes prove the robustness of the method. Analytical gCP gradients make optimizations of large systems feasible with small basis sets, as demonstrated for the inter-ring distances of 9-helicene and most of the complexes in Hobza's S22 test set. The method is implemented in a freely available FORTRAN program obtainable from the author's website. © *2012 American Institute of Physics*. [http://dx.doi.org/10.1063/1.3700154]

I. INTRODUCTION

Quantum chemically computed supermolecular interaction energies and molecular structures are subject to the basis set superposition error (BSSE) in an incomplete, atom-centered one-particle basis set.[1–4]

We like to split the general term *basis set error* (BSE), following Ref. 1, into the *basis set superposition error* (BSSE) and the *basis set incompleteness error* (BSIE). The BSSE arises due to an unbalanced basis set expansion of monomers and the ensuing multimer complex in the supermolecular approach of calculating interaction energies. In a dimer complex of the moieties A and B, the basis set of the complex is larger than the basis sets of A and B because unoccupied basis functions from A can be used by B to lower the energy (and vice versa). In the case of complexation (interaction) energies, this lowering of the energy leads to artificial- or over-binding of complexes. The famous counterpoise (CP) scheme by Boys and Bernadi[5] (BB) can be used to remedy this unbalanced description. For a dimer it reads

$$\Delta E_{CP} = [E(A)_a - E(A)_{ab}] + [E(B)_b - E(B)_{ab}]$$

where a, b are the basis functions of monomer A, B (in their frozen complex geometries). ΔE_{CP}, which is always positive in this formulation, is added to the binding energy (BE) of the complex yielding the corrected BE_{CP}

$$BE_{CP} = BE + \Delta E_{CP}.$$

The BB-CP correction covers only the case of intermolecular BSSE of non-covalently bound dimer complexes, but in fact does any close lying molecular assembly (another monomer,

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

alkyl side-chains or bulky substituents) "offer" its unoccupied (virtual) basis functions to another part of the molecule, reducing the energy of the whole assembly. This extends the idea of BSSE to the intramolecular case (IBSSE), although a uniform, clear definition is missing.

Experience shows that already at polarized triple-ζ basis quality ΔE_{CP} is reduced to about 10% of the interaction energy in self-consistent field (SCF) type treatments. However, *ab initio* studies on large biochemical systems using orbital-based quantum chemical methods are limited by non-linearly increasing computational costs if reasonable large triple-ζ basis sets (or above) are employed. Schemes to deal with this problem either on a technical (code parallelization[6,7]), algorithmical (linear scaling techniques,[8,9] divide-and-conquer approaches[10–12]) or theoretical (hybrid schemes such as QM/MM (Ref. 13)) level are still actively developed and not yet "black box".

Today's best cost-accuracy ratio is provided by density functional theory[14,15] (DFT). For large systems such as biomolecules,[16,17] nanostructured materials,[18,19] or supramolecular chemistry[20–22] weak interactions play a dominant role in their stability and reactivity. A fundamental problem of all current semi-local and conventional hybrid DFT functionals is their inability to provide the asymptotically correct $-C_6/R^6$ dependence of the London dispersion interaction energy on the inter-atomic/molecular distance R for these weak, non-covalent interaction.[23–25] Various approaches to tackle this problem were proposed in the literature; for a recent review see, e.g., Ref. 26. Herein, we will consider the latest dispersion-correction schemes DFT-D3 (Refs. 27 and 28) and DFT-NL.[29,30]

Very few limitations are set to the computational chemist having the latest computer codes and the newest supercomputer at hand. Within typical limited resources, however, one is forced to apply rather small double-ζ basis sets of various sizes (6-31G*, 3-21G) often out of necessity.[31,32] The hope is that the BSSE of relative energies is tolerable if only a qualitative understanding is sought. Validation studies are only viable on smaller systems, and as no simple BSSE propagation from smaller to larger systems can be expected, the errors for these (very) large systems remain uncertain. Even for smaller systems (20–80 atoms) it seems that many authors find the use of triple-ζ basis sets prohibitive for their computational studies using hybrid density functionals, which can be seen from the vast number of B3LYP/6-31G* applications.

The BSSE problem has many aspects that have been investigated in the past and we mention here some of which are relevant in the present context. Different formulations for multi-body complexes (clusters) are described in the literature,[33–38] they differ in the number of terms used to calculate the counterpoise correction. A central question is, if and how the so-called second order basis-set effects (affecting the individual pair-interactions in the cluster) should be accounted for. A critical examination of the different approaches can be found in Ref. 38. Developments of (supermolecular) methods that exclude BSSE by constructions as the chemical Hamiltonian approach (CHA) (Ref. 39) have not found widespread use, partially due to its involved theoretical foundation and technical difficulties.[40] Gill *et al.* propose to reduce the BSSE in the framework of dual-basis set schemes, using the monomer basis as primary basis and the complex basis as secondary basis in his Hartree-Fock (HF) or density functional perturbative correction (HFPC or DFPC) approach, and reports major computational savings.[41] Aiming at the biomolecular community, Merz *et al.* proposed a statistical, fragment-based model to estimate quickly intra- and inter-molecular BSSE of protein structures,[42] but the need of an internal classification of the fragment type makes it less universal.

The problem of IBSSE,[43–47] e.g., for conformational energies or isomerization reactions, can be approximated in the framework of the BB-CP correction – if one allows breaking of covalent bonds – by defining a suitable number of fragments.[48,49] The choice of the fragmentation scheme is crucial and introduces an unwelcoming arbitrariness, besides technical difficulties of charge and spin-state of the fragments. It can be broken down all the way towards individual atoms, leading to N_{atoms} additional computations. Valdés *et al.* demonstrated the application of this in their atom by atom scheme (CP^{aa}) called approach[50] for estimating the intramolecular BSSE in phenylalanine-glycine-phenylalaine (FGF) tripeptide. Jensen extended this idea in his atomic counterpoise correction (ACP(x)) (Ref. 51) by selecting only a specific subset of the whole complex-basis set (including basis functions from atoms x bonds apart), arguing that only not-directly bound atoms contribute to the BSSE. Jensen's ACP(1), i.e., including also directly neighboring atoms, equals Galanos CP^{aa} method. It was further argued by Jensen that the restriction to atoms reduces the computational costs of each calculation tremendously, because of efficient integral screening effects and a certain locality of the BSSE.[51] One nice feature is that inter- and intra-molecular BSSE are treated on the same conceptual level.

Many of the newly proposed BSSE correction schemes lack discussions of nuclear gradients. These are of utmost importance since a BSSE contaminated structure optimization is not expected to yield a good interaction energy at any higher level of theory.

Herein, we propose a correction scheme that can estimate the inter- and intra-molecular BSSE for HF/DFT calculations with various (small to medium sized) basis sets. The main aims of this project are: (1) To provide a fast, conceptually simple energy- and gradient-correction for the BSSE in computations of large molecules, where small basis sets often cannot be avoided because of limited computer resources. (2) To supply a tool for a quick BSSE estimation without the explicit need of expensive CP calculations.

The central idea is to estimate the CP correction solely based on the Cartesian coordinates of the molecule or complex, i.e., no input from its wave-function and no information about connectivity ("bonding") are required. A strong motivation of our scheme is certainly to keep the computational overhead as low as possible. We denote the scheme *geometrical counterpoise* (gCP) correction. The name implies that it is fitted against the conventional BB-CP correction and that the geometry of the molecules is the only information required. It is an atomic correction in the sense that the BSSE of a molecule or complex is evaluated by additive

atomic contributions. One main assumption here is that the BB-CP correction is "right". Although this is still theoretically under debate,[1,38,52–54] clearly it works very well in particular for small basis sets, which are our primary target. Obviously, in a complete basis no BSSE arises. Therefore, one needs a measure of basis set completeness (quality) that we define as the energy differences between a large (nearly complete) and a smaller (target) basis. We can use this measure to estimate atomic BSSE contributions in molecules, by accounting for the interatomic distance decay of the BSSE with a semi-empirical potential (four global parameters per basis set/method combination). The potential includes the number of unoccupied basis functions assigned to the atoms (virtual orbitals) and an estimate for the overlap of the valence orbitals. The parameters are fitted against the BB-CP corrections for the recent S66×8 benchmark set[55] of non-covalent interactions. We want to stress at this point already that the proposed correction is not meant as some general substitute for calculations that can easily be done with triple-ζ quality basis sets or higher. Large basis sets are the preferable way to go whenever this is technically possible and we are performing more and more routine calculations in our group at the quadruple-ζ level.

The outline of the paper is as follows: First, the theory of the method is presented. Second, some technical details and the fit procedure are discussed. Finally, the performance of the correction is demonstrated on a few illustrative examples including inter- and intramolecular BSSE, and for geometry optimizations.

II. THEORY

The central idea is to add in a semi-empirical fashion an energy correction ΔE_{gCP} to the energies of molecular systems in order to remove artificial overbinding effects from BSSE. As the focus lies on the contribution of individual atoms a natural outcome is its ability to yield also intramolecular BSSE corrections. The parameterization is constructed such, that it approximates the BB-CP correction ΔE_{CP} for dimers

$$\Delta E^{CP} \approx \Delta E_{gCP}, \tag{1}$$

where, e.g., for a complexation reaction $A + B \rightarrow C$ our correction is given by

$$\Delta E_{gCP} = E_{gCP}(C) - E_{gCP}(A) - E_{gCP}(B). \tag{2}$$

In practice, E_{gCP} can simply be added to the HF/DFT energy

$$E_{total} = E_{HF/DFT} + E_{gCP}. \tag{3}$$

In the following, the details of the gCP correction will be outlined. E_{gCP} contains four parameters specific for a given Hamiltonian (HF or KS-DFT) and basis set combination. The atomic contributions are globally scaled by the fit parameter σ

$$E_{gCP} = \sigma \cdot \sum_{a}^{atoms} E_a^{atom}. \tag{4}$$

The atomic contributions E_a^{atom} are obtained by multiplying the energy e_a^{miss}, which measures the incompleteness for the chosen target basis set for atom a with a decay function $f_{dec}(R_{ab})$ depending on the inter-atomic distance R_{ab}

$$E_a^{atom} = \sum_{b \neq a}^{atoms} e_a^{miss} \cdot f_{dec}(R_{ab}). \tag{5}$$

The e_a^{miss} terms are obtained from the atomic energy difference between a large (nearly complete) basis set and the target basis set. The quantity is pre-computed and supplied for a wide range of basis sets (see Sec. IV A) and computed as

$$e_a^{miss} = E_{target\ basis} - E_{large\ basis}\ |_{F=0.6a.u.}, \tag{6}$$

where the index $F = 0.6$ a.u. denotes an applied electric field of 0.6 a.u. Energy minimization of an atomic wave-function will generally not properly populate higher angular momentum (polarization) functions in the basis set. The ground state energies of a single hydrogen atom at the HF/SV and HF/SVP level of theory, e.g., are identical. To account for the population of (molecular) polarization functions that occur in molecules, we apply a weak electric field F of 0.6 a.u. in the first Cartesian quadrant ($x = y = z = 0.03464$) and perform restricted open-shell Hartree-Fock[56] (ROHF) calculations to obtain $E_{target\ basis}$ and $E_{large\ basis}$. The calculations are based on the Turbomole code[57–59] without any symmetry constraints and refer to state average solutions for 3d-transition metals. Because of SCF convergence problems in the corresponding ROKS calculations, we use the e_a^{miss} energies from ROHF calculations also for the DFT parameterization.

From test calculations and comparisons for H_2 and CH_4 molecules, a field strength of 0.6 a.u. is found to populate the p/d-orbitals in the atomic calculations reasonably well. Note, that the correct modeling of the atomic orbital populations in molecules is neither desired – nor necessary – here. The field strength, if kept within reasonable limits, has only a minor influence on the overall performance of the model.

The second term in Eq. (5), the decay function $f_{dec}(R_{ab})$, is given by

$$f_{dec}(R_{ab}) = \frac{\exp\left(-\alpha \cdot {R_{ab}}^{\beta}\right)}{\sqrt{S_{ab} \cdot N_b^{virt_b}}}, \tag{7}$$

where the interaction factor $\exp(-\alpha \cdot {R_{ab}}^{\beta})$ is normalized by the square-root of the Slater-overlap S_{ab} times the number of virtual orbitals on atom b. Many different kinds and combinations of functions were tested. Starting from a more flexible function (more fit parameters) some of them could be eliminated during the fitting procedure, and it turned out that a rational function of two exponentials (the overlap is considered as a complicated exponential) yields favorable performance. The square-root in the denominator results from one of the eliminated fit parameters. The parameters α and β are crucial and determine the performance most strongly. The number of virtual orbitals N_b^{virt} on atom b is straightforwardly obtained by subtracting the number of electron pairs (N_b^{el} being the total number of electrons of atom b) from the number of basis functions N_b^{bf} in the target basis set for b

$$N_b^{virt} = N_b^{bf} - \frac{N_b^{el}}{2}. \tag{8}$$

The overlap integral in Eq. (7) is evaluated over a single Slater-type orbital centered on each atom. Optimized Slater exponents ζ_{opt} are taken from an extended Hückel theory study by Herman.[60] The s and p valence exponents are averaged to get a single s-type orbital ζ_s. Thus, the valence overlap S_{ab} is calculated as

$$\zeta_s = \eta \cdot \frac{\zeta_{opt,s} + \zeta_{opt,p}}{2}, \tag{9}$$

$$S_{ab} = \langle \zeta_s | \zeta_s \rangle, \tag{10}$$

where η is the last of the four fit parameters.

III. IMPLEMENTATION DETAILS

The method is implemented into a freely available FORTRAN code.[61] Overlap integrals up to principle quantum number three, i.e., up to $S_{ab} = \langle 3s|3s\rangle$, are available. For all 4th-row and higher elements 3s-type valence orbitals are employed. The d-functions of transition metals are treated in the same manner as the s, p-valence shells, i.e., the s, p, and d exponents are reduced to a s-function by a simple arithmetic mean of the three exponents. To allow for elements that are not parameterized the program will internally substitute elements by their 4th-row homologous, e.g., Pb will be replaced with Ga, I with Br, Os with Fe, and so on. Unparametrized 4th-row elements will give zero contribution in the pair-potential. The error introduced by this fall-back algorithm should be small if only few atoms within a larger complex are replaced or neglected.

The computational complexity of energy and gradient evaluation formally scales with $(N_{atoms})^2$. The prefactor is very small, resulting in very fast computations (a few seconds) even for large molecules with 2000 atoms. Jensen's investigations[51] indicate a certain locality of the BSSE by showing that for HF with regular basis sets ACP contributions of atoms beyond 7 Å (for augmented basis sets 10 Å) can be neglected for an accuracy of about 0.1 kcal/mol. A huge speed up is gained by exploiting this locality with a proper cut-off distance. We make use of a value of 60 Bohrs (conservative upper bound), which results in practically no loss in accuracy but leads already to a large decrease in computation time.

For each atom-pair one overlap integral has to be evaluated. The analytical implementation is done via the auxiliary integrals A and B (below including sum representation) with $x_a = (\zeta_a + \zeta_b)/2$ and $x_b = (\zeta_a - \zeta_b)/2$

$$A_k(x_a) = \int_1^\infty \xi^k e^{-x_a \xi}\, \mathrm{d}\xi, \tag{11}$$

$$= e^{-x_a} \frac{k!}{(x_a)^{k!+1}} \sum_{v=0}^{k} \frac{(x_a)^v}{v!}, \tag{12}$$

$$B_k(x_b) = \int_{-1}^{+1} \chi^k e^{-x_b \chi}\, \mathrm{d}\chi, \tag{13}$$

$$= e^{x_b} \frac{k!}{(x_b)^{k!+1}} \sum_{v=0}^{k} \frac{(-x_b)^v}{v!} \tag{14}$$

$$-e^{-x_b} \frac{k!}{(x_b)^{k!+1}} \sum_{v=0}^{k} \frac{(x_b)^v}{v!}. \tag{15}$$

Starting from the overlap integral for quantum number $na = nb = 0$

$$\langle 0s|0s\rangle = \frac{1}{2} A_0 B_0,$$

all s-type overlap integrals over A and B integrals can be generated by applying successively the following rule (Lofthus algorithm, see Ref. 62): If na increases by 1 every $A_k B_l$ reformulates into $(A_{k+1}B_l + A_k B_{l+1})$. Similarly does the increase of nb by 1 lead to $(A_{k+1}B_l - A_k B_{l+1})$.

Analytical derivatives of each overlap integral-type are generated using the Maple algebra tool.[63] The resulting gCP gradient takes only a little more (below a factor of two) time than the gCP energy correction. A test molecule with 48 827 atoms takes about 76 s for the energy and 118 s for the gradient correction on an Intel Xeon E5430 (2.66 Ghz) workstation.

IV. COMPUTATIONAL DETAILS

Most calculations were either carried out with the TURBOMOLE suite of programs (a locally modified version of TURBOMOLE 5.9 and the recent version 6.3 (Refs. 57–59)) or with a development version of ORCA (Ref. 64). The GGA functionals BLYP,[65–67] B97-D,[68] revPBE,[69] the meta-GGA TPSS (Ref. 70) the hybrids PW6B95,[71] B3LYP,[72,73] and revPBE0 (Ref. 74) are used. For the M06-2X (Ref. 75) meta-hybrid functional GAUSSIAN09 (Ref. 76) was used.

The DFT-D3 corrections[27,28] (both damping-function variants) were applied with our group own program *dftd3*. Becke-Johnson (BJ) damping[28,77,78] is the default damping function used, i.e., DFT-D3 always corresponds to DFT-D3(BJ). In the case of M06-2X, the original D3-damping function (denoted zero-damping) is applied. The functionals' specific parameters were determined in Refs. 27, 28, 74, and 79. A variety of basis sets is employed: The Ahlrichs-type basis sets SV,[80] def2-SVP, def2-TZVP,[81,82] and def2-QZVP,[82,83] Hunzingas valence-scaled version of his minimal basis MINI (denoted MINIS) is used as provided by the EMSL basis set exchange website[84,85] and the Pople-style basis set 6-31G*.[86]

For the def2-SVP, def2-TZVP, and def2-QZVP basis sets, the TURBOMOLE calculations for (meta-)GGA and hybrid functionals uses the resolution of the identity (RI-J) approximation for the Coulomb part.[87] ORCA employs the Split-RI-J variant.[88] For def2-QZVP calculations with Turbomole and ORCA, the RI-K approximation to the exchange integrals[89] is additionally used. All auxiliary basis functions were taken from the Turbomole basis set library.[90,91] If not denoted otherwise the Turbomole grid *m5*[91] was used.[92–97] The calculations for the crambin protein employed the smaller *m3* grid.

The development version of ORCA was used for all calculations using the non-local (NL), density-dependent dispersion correction DFT-NL.[29,30] The keywords `grid4` and `vdwgrid3` specify the integration accuracy of the exchange-correlation and the NL-part, respectively. The coupled electron pair (CEPA, version 1) calculations make use of the local pair natural orbital (LPNO) approximation (denoted LPNO-CEPA/1) (Refs. 98 and 99) as implemented[100] in ORCA. A complete basis set (CBS) two-point extrapolation for LPNO-CEPA/1 was done using the Halkier extrapolation scheme[101,102] (separate extrapolation of SCF and correlation energy) using def2-TZVP and def2-QZVP (termed CBS(3,4)).

The symmetry adapted perturbation theory (SAPT) (Ref. 103) calculations are done with MOLPRO (Ref. 104) (default settings) employing the aug-cc-pVTZ (Ref. 105) basis set.

A. Fitting procedure

The recently introduced benchmark set S66×8 is used to generate CP data on which the gCP parameters are fitted. The set consists of 66 dimers at eight different distances. The dimers are combinations of 14 different monomers with one another: acetic acid, acetamide, benzene, cyclopentane, ethene, ethyne, neopentane, *n*-pentane, methylamine, methanol, *N*-methylacetamide, pyridine, uracil, and water (for a list of all 66 dimers, see Ref. 55). They have been classified according to dominating intermolecular interactions so that 23 complexes are formed by hydrogen bonds, 23 are dominated by dispersion interactions and 20 complexes are equally dominated by dispersion and electrostatic interactions. Prominent interactions such as π-π stacking (ten systems), aliphatic-aliphatic interactions (five systems), and π-aliphatic interactions (eight systems) display a wide range of interactions typical in large organic and biological systems.

For all 66 complexes counterpoise-corrected MP2/TZVP structures were computed. For each dimer "the closest intermolecular distance in the complex along an intermolecular axis" (Ref. 55) was identified and the monomer-monomer distances were varied (frozen monomers) along this axis so that a dissociation curve is generated (two distances below and five above equilibrium distance), where the longest distance is twice the equilibrium distance. The first five structures of these systems were taken and by interpolating MP2/CBS+ΔCCSD(T)/aDZ single-point energies with a fourth-order polynomial the energetically optimal intermolecular distance was determined. These "minima" constitute the S66 benchmark set used in Sec. V C for benchmarking noncovalent interactions.

For each of the 528 dimers in the set the standard BB-CP correction is calculated for a target method (e.g., HF/MINIS), and the four parameters in the gCP scheme are fitted in a least-squares sense (i.e., minimization of the root-mean-square deviation, RMSD). The weight of the errors for the shortest distance in the set is reduced to 0.5. This focuses the correction slightly to equilibrium geometries (important for good structures) and longer distances (important for large biomolecules).

TABLE I. Parameters for the gCP correction and fit quality (RMS deviation in kcal/mol) of the fit against Boys and Bernadi CP values for the S66×8 dimer geometries. The arithmetic mean $\overline{CP}$ of the BB-CP correction is also given (in kcal/mol).

	σ	η	α	β	$\overline{CP}$	RMSD
HF/MINIS	0.1290	1.1526	1.1549	1.1763	1.02	0.30
HF/SV	0.1724	1.2804	0.8568	1.2342	1.16	0.32
HF/SVP	0.2054	1.3157	0.8136	1.2572	1.10	0.41
HF/6-31G(d)	0.2048	1.5652	0.9447	1.2100	1.02	0.40
HF/def2-TZVP	0.3127	1.9914	1.0216	1.2833	0.20	0.12
B3LYP/MINIS	0.2059	0.9722	1.1961	1.1456	1.10	0.34
B3LYP/SV	0.4048	1.1626	0.8652	1.2375	1.64	0.56
B3LYP/SVP	0.2990	1.2605	0.6438	1.3694	1.64	0.65
B3LYP/6-31G(d)	0.3405	1.6127	0.8589	1.2830	1.43	0.48
B3LYP/def2-TZVP	0.2905	2.2495	0.8120	1.4412	0.32	0.20

In principle, parameters have to be fitted for every combination of Hamiltonian and basis set (model chemistry). The efforts to account for the broad range of common density functionals (plus Hartree-Fock) combined with a variety of basis sets are huge and beyond the scope of this study. We decided to supply a few practically useful combinations as listed in Table I. We suggest to use the B3LYP fit parameters also for any other density functional. As will be shown later, the accuracy gained by a re-fit for different density functionals is negligible because different functionals provide rather similar orbitals and the mechanism for the BSSE is the same in all SCF methods.

As can be seen in the last column in Table I, the representation of the BB-CP correction by the gCP approach is surprisingly good keeping in mind that no electronic information has been used and only four parameters had to be determined. With most method/basis combinations RMSDs of 0.1 to 0.4 kcal/mol are obtained. This corresponds to a typical relative accuracy of 10%–30%, which is sufficient for many purposes as will be demonstrated in detail below. For the larger def2-TZVP basis, the RMSD is rather small but the relative accuracy deteriorates although we still consider this as practically useful for a quick estimate of the magnitude of the BSSE. Not unexpectedly, the best fits (small RMSD and best relative accuracy) are obtained with the compact minimal basis set. A further discussion of the accuracy of the gCP method is provided in Sec. V A. Except for η the behavior of the four gCP parameters is unsystematic. The global scaling factor σ varies mostly between 0.1 and 0.5. For α values are found between 0.8 and 1.2 except for B3LYP/SVP that has a small α of 0.6. The parameter β ranges between 1.1 and 1.5. A systematic increase from smaller to larger basis sets is found for η, which scales the Slater exponent (see Eq. (9)). This makes the Slater functions more compact (less diffuse) and reduces the overlap S_{ab} as a result.

V. RESULTS AND DISCUSSION

A. BSSE in HF and DFT calculations for the S66×8 set

During the course of the project a few thousand CP corrections have been calculated (data points for the fit). As a side

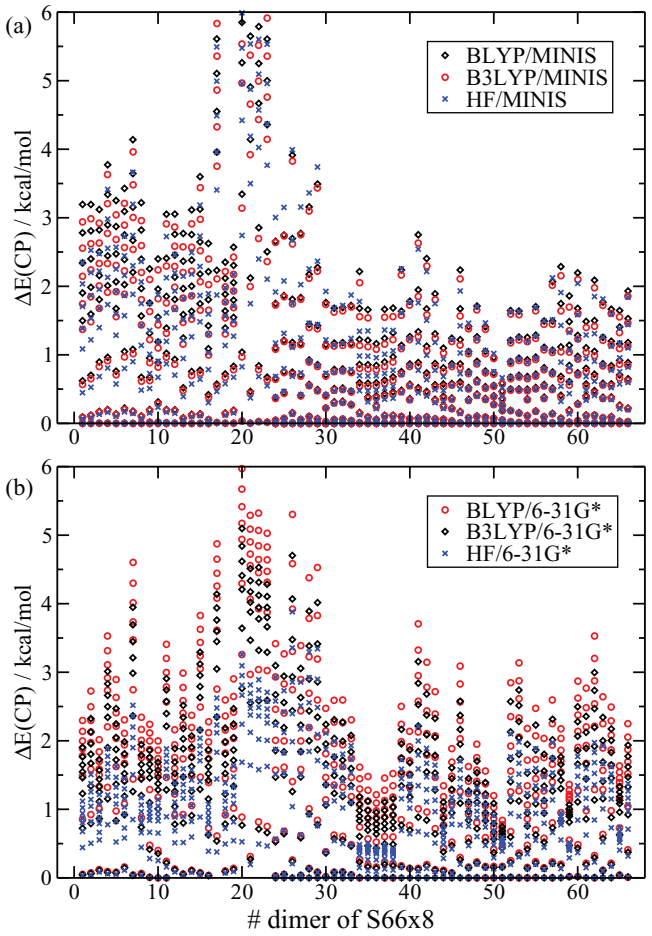

FIG. 1. Computed Boys and Bernadi CP correction for the S66×8 benchmark set with HF, B3LYP, and BLYP with the basis sets (a) MINIS and (b) 6-31G*. The set consists of 66 systems (abscissa) with eight varying distances. The CP correction for each distance (ordinate) results in 528 data points.

effect, this allows to analyze the behavior of the BSSE for a statistically very meaningful number of data. For the two basis sets MINIS and 6-31G* CP corrections for HF, B3LYP, and BLYP are plotted in Figures 1(a) and 1(b). Shown are for each of the 66 systems all eight CP values for the different inter-fragment distances (eight values on the ordinate for one complex on the abscissa).

For the MINIS basis set the data points for HF, B3LYP, and BLYP are hardly distinguishable, giving in all three cases a similar CP correction. Largest deviations (although still very small in magnitude) are seen for the hydrogen-bonded systems (entries 1–23). The largest CP correction (at the shortest distance) is obtained in most cases for BLYP. HF shows a tendency to yield smaller CP corrections than DFT. The data points for 6-31G* are more clearly separated. The HF CP correction is the smallest, while BLYP gives again the largest values. Comparing MINIS with 6-31G*, the former yields only for the hydrogen-bonded systems a larger BSSE, for the dispersion and mixed-systems 6-31G* tends to larger BSSE values. Rather small CP corrections are observed for system 35–38 for both basis sets. These four systems (cyclopentane, neopentane, and pentane in different combinations) are less prone to BSSE due to large inter-fragment distances. As

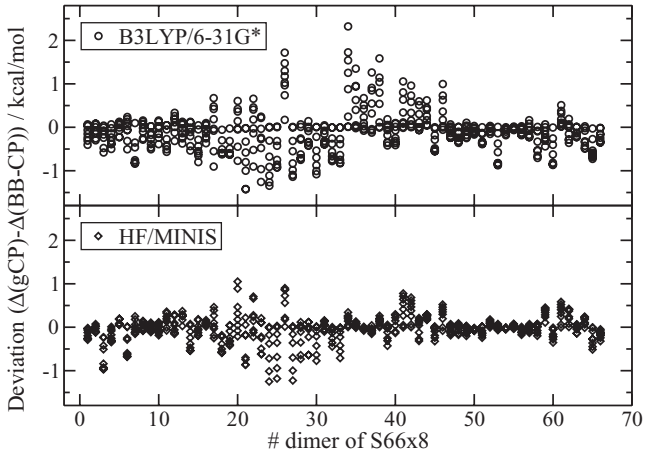

FIG. 2. Deviations between the Boys and Bernadi CP correction (ΔBB-CP) and the gCP correction (ΔgCP) for the S66×8 dimers in kcal/mol for B3LYP/6-31G* and HF/MINIS. The set consists of 66 systems (abscissa) with eight varying distances. The deviation (ΔgCP − ΔBB-CP) for each distance (ordinate) results in 528 data points. Positive values indicate an overestimation.

Figure 1 demonstrates, the order of magnitude of the CP correction is HF < B3LYP < BLYP, which indicates the amount of Fock-exchange as the major influencing factor.

These results for two different basis sets support the notion that HF is slightly less affected by BSSE than DFT, and that the DFT BSSE is reasonably invariant to the functional type. This is one motivation to globally adjust the gCP parameters only to HF and DFT, instead to each DFT functional specifically. This also suggests to estimate the true CP correction for hybrid-functionals with a GGA functional to save computation time with little loss of accuracy (hybrid functionals are a factor of 3–5 more costly with efficient density-fitting (RI) schemes).

The differences between the BB-CP and our gCP correction are given in Figure 2 for two representative examples (as above eight values for each complex). The results for B3LYP/6-31G* indicate good performance for the hydrogen-bonding dominated and the mixed-type systems. Some outliers are found for the dispersion dominated systems. The three largest deviations are observed for systems 34 (pentane···pentane), 26 (uracil···uracil) and 38 (cyclopentane···cyclopentane), all showing overestimation of the BSSE by gCP (positive values). HF/MINIS yields a much smaller error range with fewer outliers, though the systems 20 to 30 seem to be problematic at both levels (underestimation of the BSSE). For the largest dimer distances (twice the equilibrium value), the HF/MINIS BB-CP correction is essentially zero, which is well reproduced by gCP. This asymptotic decay of the BSSE is a very important property for large systems and should be accurately reproduced by any approximate scheme.

B. Adjustment of the London dispersion correction DFT-D3

Although the gCP correction is proposed as an independent, stand-alone procedure, it is clear that HF/DFT in general requires dispersion corrections even in the basis set limit.

We will therefore apply our DFT-D3 scheme[27,28] (in the default Becke-Johnson damping variant[28,77,78]) and also comment on the use of the modified Vydrov and van Voorhis (VV10) (Ref. 30) non-local (NL), density dependent dispersion correction.[29]

In the original publication[27] the dispersion correction has been developed with AO basis sets close to the complete basis set (CBS) limit and remaining, tiny BSSE effects have been absorbed in the DFT-D3 short-range damping functions. Hence it is in general recommended to apply the gCP correction together with the default DFT-D3 parameterization. Nevertheless, we studied in quite detail if any re-fit of the DFT-D3 short-range part has a positive effect. In the case of HF-gCP-D3/MINIS, we found a large gain in performance when the DFT-D3 parameters are re-determined in the presence of the gCP correction. The resulting level of theory (termed HF-gCP-D3(fit)/MINIS) performs extremely well providing a MAD for the S66 set of only 0.38 kcal/mol which is of about the same accuracy as obtained by most DFT-D3/large basis methods and also much better than, e.g., MP2/CBS (see below). In a similar fashion the short-range damping parameter b (see Ref. 29 for details) in the HF-gCP-NL/MINIS method was re-fitted, but without the same success. The MAD for the S66 drops from 0.68 to 0.62 kcal/mol.

C. Interaction energies for S66 and S22 sets

The validation of the gCP scheme is shown for the two benchmark sets S22 and S66. The S22 set is the de-facto test set for non-covalent interactions.[106] According to Hobza and co-workers, some interaction motifs are underrepresented in S22.[55] As a consequence, they recently published the S66 set as an extension and revision of S22.[55]

Reference values for the interaction energies were taken from the original work and are based on the estimated CCSD(T)/CBS level of theory (MP2/CBS values based on aTZ and aQZ basis set extrapolation were combined with the difference of MP2 and CCSD(T)/aDZ correlation energies (ΔCCSD(T))). Very recently, revised reference values for S66 and another extension were published; this extension was dubbed S66a8 and describes the same dimers at different inter-monomeric angles.[107] While the original S66 focuses on benchmarking wave-function based methods, DFT in the framework of DFT-D3 was extensively tested on both the S66 and S66×8 in our group.[108] The performance of the very recently proposed DFT-NL method on the S66 can be found in Ref. 29. Note, that the S22 set is not included in the fitting set, while S66 is partly included through the S66×8 set that has the same systems but at slightly different equilibrium distances.

Results for various methods are shown in Table II. Well performing methods have MAD values less than about 0.5–0.6 kcal/mol for the S66 interaction energies. The decrease of the MAD after adding the gCP correction is striking, in particular for the smaller basis sets. The already high accuracy of methods such as BLYP-D3, B3LYP-D3, and PW6B95-D3 with def2-TZVP can be further improved to reach exceptionally small MAD values of 0.28–0.33 kcal/mol.

TABLE II. Mean absolute deviations (MADs) for S66 benchmark set in kcal/mol. The column "+ gCP" indicates that the gCP correction is added to the DFT-D3 results. The methods are ordered according to increasing basis set quality.

Method	w.o. corr.[a]	DFT-D3	+ gCP
Parametrized methods			
HF/MINIS	2.95	1.86	0.51
HF/MINIS[b]	2.95	1.54	0.38
HF/SV	2.96	2.72	1.32
HF/def2-SVP	2.83	2.06	0.83
HF/6-31G*	2.72	2.08	0.89
HF/def2-TZVP	3.97	0.88	0.67
B3LYP/MINIS	3.38	2.26	1.07
B3LYP/SV	3.08	2.92	1.15
B3LYP/def2-SVP	2.61	2.33	0.68
B3LYP/6-31G*	2.30	2.20	0.68
B3LYP/def2-TZVP	2.96	0.57	0.33
Methods that employ the "dft" parameters for gCP			
BLYP/MINIS	3.88	2.54	1.25
BLYP/MINIS[c]	3.88	2.54	1.22
B97-D/MINIS	3.94	1.90	1.00
TPSS/MINIS	3.43	1.79	1.25
M06-2X/MINIS[d]	1.48	1.72	0.73
PW6B95/MINIS	2.22	1.57	0.95
BLYP/SV	3.35	3.08	1.19
BLYP/SV[c]	3.35	3.08	1.07
TPSS/6-31G*	2.22	1.76	0.66
BLYP/def2-SVP	2.91	2.53	0.72
TPSS/def2-SVP	2.45	1.92	0.80
M06-2X/def2-SVP	1.60	1.41	1.15
PW6B95/def2-SVP	1.38	1.85	0.66
PW6B95/def2-SVP[c]	1.38	1.85	0.60
BLYP/def2-TZVP	3.71	0.46	0.32
PW6B95/def2-TZVP	1.36	0.29	0.28

[a]Result without any dispersion or gCP correction.
[b]DFT-D3 parameters fitted to the respective gCP corrected level of theory.
[c]gCP parameters fitted for the respective level of theory.
[d]DFT-D3 with zero-damping.

The MAD decreases compared to the "pure" results in most cases by adding the DFT-D3 dispersion correction. A look at the mean deviation (MDs, see supplementary material[109]) shows that the "pure" results still suffer from underbinding, which is typically found for DFT/HF for non-covalent interactions since they are incapable of describing London dispersion interactions. However, the artificial binding from BSSE becomes obvious for the smaller basis sets after applying DFT-D3. A typical example is B3LYP/6-31G* that gives a MD of −1.29 kcal/mol (underbinding) without any correction and changes to 2.20 kcal/mol (overbinding) at the DFT-D3 level. Only after correction with gCP, a very reasonable MD of 0.32 kcal/mol is obtained. The relatively large basis set def2-TZVP, which reduces the BSSE typically below 10% of the interaction energy, already leads to good results without gCP correction.

From the first block of data in Table II, which depicts the performance of all parameterized methods, it becomes clear that the minimal and polarized double-ζ basis sets (MINIS, def2-SVP, 6-31G*) work better in conjunction with gCP than the seemingly too unbalanced SV basis set. Remarkable is the

excellent performance of HF-gCP-D3/MINIS with a MAD of 0.51 kcal/mol, which can be further improved by adjusting the DFT-D3 parameters for this model chemistry (see Sec. V B). The adjusted HF-gCP-D3(fit)/MINIS method yields the smallest MAD of 0.38 kcal/mol for basis sets below triple-ζ quality in this work and beats even MP2/CBS results[55] (MAD = 0.45 kcal/mol). This very good performance clearly relies on strong, but systematic error compensation that is analyzed further below.

The second data block in Table II applies the gCP correction with the "dft" parameters or a special fit to the respective level of theory. We focus again mainly on the gCP-D3 corrected results. The use of MINIS with a GGA or meta-GGA cannot be recommended as the MADs are all above 1 kcal/mol. The gCP re-fit for BLYP/MINIS shows no significant improvement (MAD 1.25 versus 1.22 kcal/mol). The inclusion of Fock-exchange (with contributions of 20% (B3LYP) or 28% for PW6B95) improves the results (1.07 and 0.95 kcal/mol). M06-2X-gCP-D3 (54% Fock-exchange) gives a reasonable MAD of 0.73 kcal/mol. The gCP re-fit is also done for BLYP/SV, but again the improvement is small. GGAs and meta-GGAs perform much better with the more balanced def2-SVP basis and the MADs drop to 0.72 (BLYP) and 0.80 (TPSS). Even better results are obtained for corrected PW6B95/def-SVP and B3LYP/def2-SVP yielding good values of 0.66 and 0.68 kcal/mol. The gCP re-fit for PW6B95/def2-SVP further improves the MAD to 0.60 kcal/mol, which is the second best MAD next to HF/MINIS's for the "small" basis sets (below triple-ζ quality). The def2-TZVP results show that gCP works in principle also for large basis sets, but that the performance gain is small (e.g., the MAD for HF/def2-TZVP decreases from 0.88 to only 0.67 kcal/mol). We again note here that def2-TZVP is not the main target for gCP, but the consistent improvement of the MAD even for BLYP/def2-TZVP and PW6B95/def2-TZVP proves the generality of the gCP scheme. One important point is that the MAD in all tested cases is not increased by gCP correction and a statistically positive, or at least neutral, effect is obtained.

It can be argued that DFT-NL is less suited than DFT-D3 for small basis set calculations, because the dispersion contribution arises through integration of the density, and this density is also subject to basis set errors. From an empirical point of view, comparing Table II and Table III, DFT-NL yields very similar results to DFT-D3, proving that DFT-NL is capable to deliver good dispersion energies also with small basis sets. Studies on the large GMTKN30 benchmark set have shown that DFT-NL and DFT-D3 perform statistically at the same, high level, but individual functional/basis combination will be slightly in favor of one or the other dispersion correction. We observe the same apparent error compensation for HF/MINIS as in the DFT-D3 example, such that HF-gCP-NL/MINIS yields also a very good MAD of only 0.64 kcal/mol.

S22 results are reported in Table IV. The updated reference values from Sherill's group are used.[110] HF-gCP-D3/MINIS and its re-fitted variant provide again the best performance with a MAD of 0.64 and 0.55 kcal/mol, respectively. Also PW6B95-gCP-D3 shows again high accuracy (MAD=0.66 kcal/mol) after re-fit of gCP. Note that the re-fit is not done on the S22, but still on the S66×8 dimers. The overall picture of performance is quite similar to the S66 set. There is, however, the tendency that the S22 MADs are slightly higher than for the S66 set. The gCP-corrected model chemistries that reach a MAD below 1 kcal/mol are, besides the already mentioned HF/MINIS, B3LYP, and PBE0 combined with 6-31G*, as well as PW6B95/def2-SVP. Obviously, the list of tested functionals and basis sets is far from exhaustive and other, well-performing hybrid functionals are sure

TABLE III. Mean absolute deviations (MADs) for DFT-NL for the S66 benchmark set in kcal/mol. The column "+ gCP" indicates that the gCP correction is added to the DFT-NL results. The DFT-NL calculations have been performed non-self-consistently, except those indicated by (SC).

Method	DFT-NL	+ gCP
HF/MINIS	1.09	0.64
HF/SV	2.06	0.96
HF/def2-TZVP	0.27	0.27
B3LYP/MINIS	2.01	1.31
B3LYP/def2-SVP	2.05	0.98
BLYP/MINIS	2.28	1.47
PW6B95/SV	2.48	0.97
PW6B95/def2-SVP	1.85	0.70
revPBE0/SV	2.34	1.09
revPBE0/SV (SC)	2.37	1.08
revPBE/SV	2.56	1.06
revPBE/def2-SVP	1.95	0.77
revPBE/def2-SVP (SC)	1.99	0.75
TPSS/MINIS	1.89	1.42
TPSS/def2-SVP	1.95	0.91

TABLE IV. Mean absolute deviations (MADs) for S22 benchmark set in kcal/mol. The column "+ gCP" indicates that the gCP correction is added to the DFT-D3 or DFT-NL results. The DFT-NL calculations have been performed non-self-consistently.

Method	w.o. corr.[a]	DFT-D3	+ gCP
HF/MINIS	3.50	2.19	0.64
HF/MINIS[b]	3.50	1.80	0.55
B3LYP/MINIS	4.33	3.22	1.70
PW6B95/MINIS	2.99	2.32	1.24
PW6B95/MINIS[c]	2.99	2.32	1.27
B3LYP/6-31G*/CP	3.79	0.66	...
B3LYP/6-31G*	2.67	2.55	0.88
PBE0/6-31G*	2.13	2.17	0.77
B3LYP/def2-SVP	3.41	3.23	1.05
B3LYP-NL/def2-SVP	2.94	...	1.37
PW6B95/def2-SVP	1.44	2.01	0.84
PW6B95/def2-SVP[c]	1.44	2.01	0.66
BLYP/SV	3.69	3.36	1.10
BLYP/SV[c]	3.69	3.36	1.02
TPSS/SV	3.14	2.77	0.99
TPSS/SV[d]	3.14	2.77	1.06
TPSS-NL/SV	2.87	...	1.17

[a]Result without dispersion or gCP correction.
[b]DFT-D3 parameters fitted to the respective gCP corrected level of theory.
[c]gCP parameter fitted for the respective level of theory.
[d]gCP parameters fitted for the GGA functional BLYP/SV.

to be found. The re-fitting of gCP parameters for BLYP and TPSS shows similar to the S66 only minor improvements. In the case of TPSS/SV the standard parameters suit even better than the for BLYP re-optimized gCP parameters.

We also investigated in detail the performance of the very common B3LYP/6-31G* combination for S22 and additionally calculated its BB-CP corrections (B3LYP/6-31G*/CP). It can be seen that the artificial BSSE binding contributes about 1 kcal/mol to the MAD reduction. CP-corrected B3LYP/6-31G* gives 3.79 kcal/mol, while the uncorrected MAD is 2.67. This means that the widely used B3LYP/6-31G* method benefits from BSSE in many intramolecular situations, while its inherent accuracy for non-covalent interactions is catastrophic. B3LYP-D3/6-31G*/CP gives a slightly better MAD than the gCP corrected B3LYP-D3/6-31G* (0.66 versus 0.88 kcal/mol). Adding only DFT-D3 to B3LYP/6-31G* is insufficient, as it gives a MAD of 2.55, such that a larger basis set or a BSSE correction is necessary for reliable results.

The amount of Fock-exchange plays a major role for the BSSE of very small basis sets such as MINIS. The MADs for GGA functionals with basis sets smaller def2-SVP are usually higher than 1 kcal/mol (e.g., BLYP-gCP-D3/MINIS gives 1.25 kcal/mol). Hybrid functional seem to perform better (PW6B95-gCP-D3/MINIS MAD of 0.95 kcal/mol), but not uniformly (B3LYP-gCP-D3 for MINIS gives MAD of 1.07 kcal/mol). Obviously only a very small selection of the plethora of functionals is tested here, but we believe that hybrid functionals or HF are better suited in combination with very small basis sets such as MINIS and SV. The Fock-exchange is deemed necessary for proper error compensation between induction, electrostatics and Pauli-repulsion as discussed below.

The BSSE is reduced by gCP with a typical error of 10%-30% (see Sec. IV A), which already improves the MADs for the S66 and S22 set in a robust manner, i.e., addition of gCP never statistically worsens the results among all test methods. This validates small basis set calculations of non-covalent interaction with the combination of gCP and DFT-D3.

D. Energy decomposition analysis of minimal basis set HF interactions

The very good performance of HF-D3 for non-covalent interactions was already noted in Ref. 28. That this (at least for the here investigated systems) also holds for the gCP-corrected HF/MINIS method is surprising and demands further analysis, as briefly given here. Pople already observed that the geometries of small molecules with HF/STO-3G turn out to be excellent as well, better than HF is inherently capable of yielding.[111,112] Similar observations for CP-corrected minimal basis set HF interaction energies were reported by Kołos,[113] who already enhanced HF with a correction for the London dispersion energy.

To gain insight into the apparent error compensation an energy decomposition analysis (EDA) as proposed by Kitaura and Morokuma[114] and implemented in GAMESS-US (Ref. 115) is done for HF with increasing basis set sizes. The very large def2-QZVP basis set already includes semi-diffuse functions and gives results near the basis set limit for HF and DFT. Investigations of BSSE effects within the EDA (Ref. 116) demonstrated significant BSSE in the Pauli repulsion (less repulsion due to BSSE) and in the charge-transfer term (more binding due to BSSE). The decomposed interaction energy for the benzene-water system from S66 is presented in Table V. It is classified in the original work as of "mixed" interaction type, meaning that both, dispersion and electrostatics, play a dominant role.

TABLE V. Energy decomposition analysis of the interaction energy of the benzene-water complex (system 54) of the S66 benchmark set in kcal/mol. The "def2" label is omitted for clarity. The individual terms are electrostatic interaction ES, Pauli, or exchange repulsion EX, induction energy IND, charge-transfer energy CT, and the high-order-coupling term MIX. The gCP correction and the Boys and Bernadi counterpoise correction BB-CP are given. The BB-CP-correction only affects the EX,CT and MIX terms.

Term	MINIS	MINIS/CP	SVP	TZVP	QZVP
ES	− 2.14	− 2.14	− 3.18	− 3.37	− 3.16
EX	1.86	2.21	2.70	3.38	3.45
IND	− 0.09	− 0.09	− 0.39	− 0.77	− 1.05
CT	− 0.93	− 0.78	− 1.23	− 1.05	− 0.93
MIX	0.05	0.27	− 0.07	0.37	0.82
INT	− 1.24	− 0.53	− 2.18	− 1.45	− 0.88
gCP	0.69	...	0.88	0.15	...
BB-CP	0.72	...	0.99	0.48	0.12
INT+CP	− 0.52	...	− 1.19	− 0.97	− 0.76

The electrostatic energy (ES) is already almost exact at def2-SVP, but is too small in the MINIS basis. Exchange repulsion (EX) is only half as strong in MINIS as in def2-QZVP. This can be understood from the high compactness of the minimal basis that reduces the overlap between the fragments. The absolute induction energy (IND) slowly increases from MINIS (below 0.1 kcal/mol) to 1 kcal/mol with def2-QZVP. The very small value for MINIS stems from the inflexibility of the minimal basis set. The charge-transfer term (CT) is least affected by the basis set size. The high-order-coupling term (MIX), which takes into account the mutual influence of each interaction term, is small (below 0.1 kcal/mol) up to def2-TZVP where it reaches 0.37 kcal/mol. The CP-corrected EDA raises the repulsion stemming from EX and introduces a significant amount of term-coupling (MIX = 0.27 kcal/mol). The CT is insignificantly affected. The major effects comparing MINIS/CP with def2-QZVP can be found in the lowered ES and IND terms (about 2 kcal/mol), while at the same time the EX value is reduced (1.2 kcal/mol). The behavior of the high-order-coupling term MIX is not easily understood and it complicates the interpretation of the error compensation. It is together with charge-transfer the smaller effect, and both seem to partly cancel each other as well. The investigation of only a single complex limits the generality of our conclusions. However, the benzene-water interaction is a representative interaction type not only in the S66 benchmark set, but of biomolecular systems in general. HF's physical correctness of Fermi-correlation and the error compensation with minimal basis sets turns it into an intriguing method if the coulomb-correlation is dominated by dispersion interactions, which in turn can be accurately evaluated by DFT-D3.

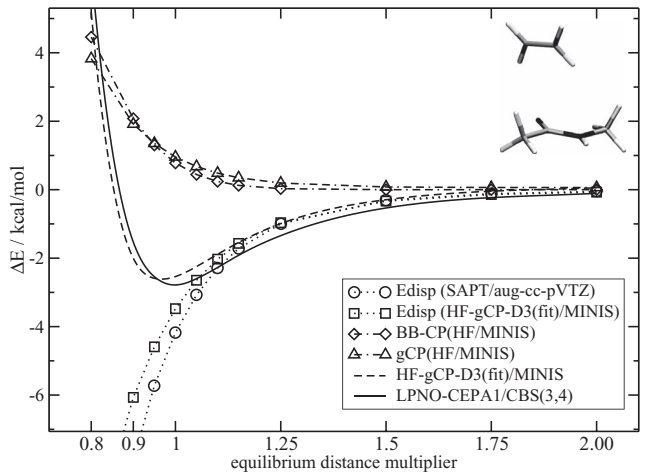

FIG. 3. Potential energy curve of of a peptide-ethene complex (taken from the S66 set, depicted in the top right corner). The dimers are displaced along the vectors connecting their center-of-mass points by the multipliers 0.8, 0.9, 0.95, 1.00, 1.05, 1.10, 1.25, 1.50, 1.75, 2.00. The dispersion energy E_{disp} from SAPT and DFT-D3 is plotted along with the Boys and Bernadi counterpoise (BB-CP) and gCP correction.

Figure 3 presents the potential energy curve for the peptide-ethene complex in the S66 benchmark set and compares the dispersion energy obtained through SAPT treatment with that from the D3 method, and the BB-CP correction scheme with gCP. The monomers are displaced along the vector that connects their center-of-mass points. Similarly to the S66×8 set, multipliers (0.8, 0.9, 0.95, 1.00, 1.05, 1.10, 1.25, 1.50, 1.75, 2.00) are used to generate displaced structures from the equilibrium geometry, which is taken from the S66 set. Reference interaction energies[117] are obtained by a two-point extrapolation to the CBS limit using def2-TZVP and def2-QZVP (CBS(3,4)) for LPNO-CEPA/1. The dispersion energy E_{disp} is calculated by SAPT/aug-cc-pVTZ and by DFT-D3 using the parameters optimized for HF-gCP-D3(fit)/MINIS. Both curves are in good agreement, especially in the asymptotical region, where the dispersion energy clearly dominates the interaction energy. The BB-CP correction almost vanishes already at a multiplier of 1.25 (below 0.1 kcal/mol). The gCP corrections follows closely the BB-CP curve, but is a little bit more long-ranged (gCP vanishes at multiplier 1.5 instead of 1.25). The close agreement between DFT-D3 and SAPT, and the gCP and BB-CP scheme, is encouraging and supports our idea to built a total correction on physically sound components.

E. Multiple fragments

The performance of the atomic (pair-wise additive) gCP scheme for multi-fragment situations is a justified question as the fitting procedure includes only dimeric complexes. Four systems of the WATER27[118] test set of water clusters are used as test cases. The BSSE is calculated for multiple fragments using the simplest, the so-called site-site function counterpoise (ssf-CP) approach,[37] thus neglecting second-order basis set effects. In the ssf-CP approach, the BSSE is calculated for n-fragments by summing up the difference between the fragment in its own basis set and the fragment in the complex basis set. For a trimer it reads

$$\begin{aligned}\Delta E_{CP}(ABC) = {} & E(A)_a - E(A)_{abc} \\ & + E(B)_c - E(B)_{abc} \\ & + E(C)_c - E(C)_{abc}\end{aligned}$$

where A, B, C denote the fragments and a,b,c the corresponding basis sets.

TABLE VI. Selected systems from the WATER27 set as a test case for a cluster-type CP correction. ssf-CP and gCP values in kcal/mol are compared for the HF/MINIS and DFT(=B3LYP)/def2-SVP level of theory.

	HF/MINIS		DFT/def2-SVP	
# Fragments	ssf-CP	gCP	ssf-CP	gCP
2	1.96	1.59	2.46	2.02
3	6.22	5.50	7.74	6.97
4	10.10	8.84	12.49	10.96
5	12.57	11.25	15.70	13.69

As can be seen from Table VI, the estimates of the BSSE by the ssf-CP and gCP approaches are in good mutual agreement. The gCP correction is constantly smaller by only 10%–20%, which demonstrates its usefulness also for clusters of multiple fragments for which other methods become prohibitively costly for large systems.

F. Intramolecular BSSE and biomolecules

The treatment of large biomolecules is one major aspect of the gCP scheme. Their typical system sizes make compromises at the one-particle basis set inescapable. The limitation to small double-ζ basis sets is error-prone, not only regarding energetics, but also regarding structures. An extensive study of biomolecular systems is beyond the scope of this work, instead prove-of-principle examples are presented. One ideal test system is the crambin protein, which finds prominent use in many computational and experimental studies due to its small size.[119–121] We apply the GGA functional BLYP combined with the basis sets MINIS, SV, and def2-TZVP to elucidate intramolecular BSSE in a folded and extended conformer of crambin.

BLYP-D3/def2-QZVP performs exceptionally well for the PCONF test set[122,123] of peptide conformers (MAD = 0.5 kcal/mol) and is expected to yield reliable results at the triple-ζ level for our model conformational problem. Additionally to the standard B3LYP parameters, optimized gCP parameters for BLYP have been tested, but again without any significant impact.

After adding hydrogen-atoms to the crystal structure (1CRN.pdb) using OPENBABEL,[124,125] the generated structure is manually corrected to give a sensible starting point for a PM6-DH+ optimization[126–128] using MOPAC,[129] which results in the folded conformer (see Figure 4). A 5000 steps, gas-phase molecular dynamics simulation using Amber/GAFF[130] build into the *ambmd* program (supplied through the MOLDEN[131] package) at 298 K leads to a strongly

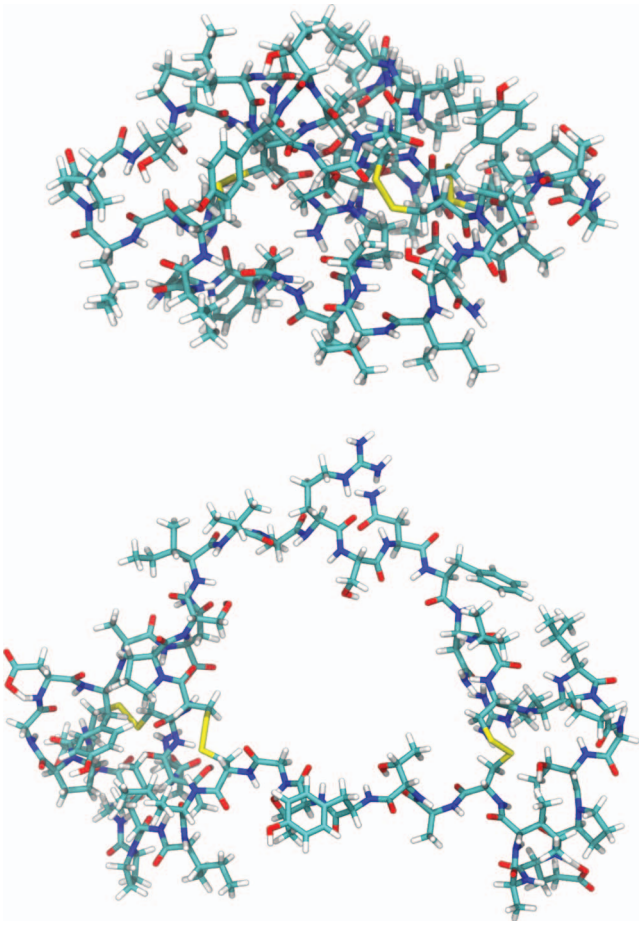

FIG. 4. View of the folded (top) and extended (bottom) PM6-DH+ optimized (gas-phase) conformers of crambin.

unfolded crambin. A sub-sequential PM6-DH+ optimization regained a somewhat tighter packaging, and yields the final, extended conformer. The extended conformer is thus purely artificial and serves only test purposes. The Cartesian coordinates can be found in the supplementary material.

In Table VII, the relative energies between both conformers are computed at different levels of theory. Note that already for the def2-TZVP basis set (12 063 basis functions) Turbomole needed to be slightly modified in order to run the calculations (mostly raising some hard coded memory allocation limits).

The gCP uncorrected results for ΔE with SV and def2-TZVP basis sets deviate strongly from each other (by 85 kcal/mol), which indicates a huge IBSSE (overstabilization of the folded conformer). The MINIS and def2-TZVP values agree much better, which is rather unexpected and can be attributed to error compensation. As mentioned in Sec. V C, GGAs with the MINIS basis set should be used with care. The gCP correction for the SV basis is largest (about twice the MINIS correction) while it diminishes for def2-TZVP (−17.4 kcal/mol). Results for SV and def2-TZVP differ by only 3 kcal/mol which can be considered as excellent in our opinion.

The HF-gCP-D3 results are equally promising. gCP corrections for MINIS amount to 60 kcal/mol, while for SV, a seemingly too small correction of about 80 kcal/mol leads to a too high ΔE. The excellent performance of HF-gCP-D3/MINIS for the S66 set, on the other hand, is also reflected in the very good agreement with BLYP-gCP-D3/def2-TZVP (11 kcal/mol difference corresponding to only 5% for ΔE). Although this has to be verified for more systems, HF-gCP-D3(fit)/MINIS seems to be a very promising (non-DFT) candidate for related biochemical problems.

TABLE VII. Relative energies (kcal/mol) of two crambin conformers (extended and folded) calculated with BLYP-D3 and various basis sets. Structures are optimized at the PM6-DH+ level of theory.

Method	ΔE[a]	ΔgCP (≈ IBSSE)
BLYP-D3/MINIS	223.1	...
BLYP-D3/SV	323.5	...
BLYP-D3/def2-TZVP	238.1	...
BLYP-gCP-D3/MINIS	161.4	− 61.7
BLYP-gCP-D3/SV	223.4	− 100.1
BLYP-gCP-D3/def2-TZVP	220.7	− 17.4
HF-gCP-D3/MINIS	244.5	− 59.7
HF-gCP-D3(fit)/MINIS	231.4	− 59.7
HF-gCP-D3/SV	290.5	− 78.5

[a]Folded → extended.

One well-known example in the literature for major IBSSE effects is the phenylalanine-glycine-phenylalanine (FGF) tripeptide,[43,50,51] and as such it became a common test case. Table VIII shows a comparison of the ACP and the CP^{aa} correction for the FGF tripeptide with the gCP scheme. The values are taken from Ref. 51 where the haug-cc-pVDZ basis set (heavy-augmented, i.e., no diffuse functions on hydrogen) has been applied. This basis is not parameterized and so instead we compare against def2-SVP and def2-TZVP results.

Our comparison suffers from the differences in the basis set, but the BSSE for def2-SVP should be slightly higher, or at least of comparable magnitude, than haug-cc-pVDZ. For the HF/def2-SVP parameterization gCP yields an IBSSE of 1.72 while ACP(2) gives 0.76 kcal/mol and ACP(4) 1.56 kcal/mol. ACP(1) or its equivalent CP^{aa} underestimates the IBSSE for this system (0.38 kcal/mol as discussed by Jensen[51]). Similar underestimation can be ascribed to ACP(2) that gives only slightly larger values and good mutual agreement is seen between ACP(4) and gCP. One serious drawback of IBSSE calculations is the absence of a well-established reference method, thus it is difficult to estimate "correct" values. However, the performance of gCP fits well into the other approaches, having the same sign and comparable magnitudes.

TABLE VIII. Intramolecular BSSE calculated for the FGF-tripeptide in kcal/mol. Comparison with Jensen's ACP approach.

Method	Δ
gCP(HF/def2-SVP)	1.72
gCP(HF/def2-TZVP)	0.27
ACP(1) or CP^{aa}[a]	0.38
ACP(2)*	0.76
ACP(4)*	1.56

[a]Evaluated for HF/haug-cc-pVDZ (no diffuse functions on hydrogen), values taken from Ref. 51.

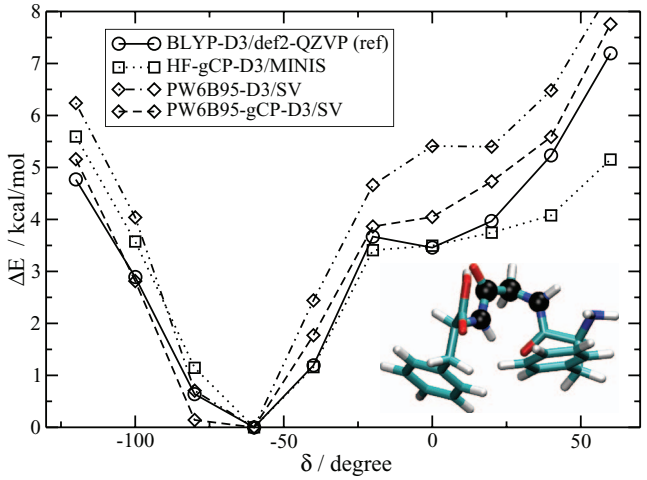

FIG. 5. Relaxed rotational energy profile of the FGF tripeptide in kcal/mol. The dihedral angle δ between the atoms indicated in black is varied in steps of 20° and held fixed in subsequent optimizations.

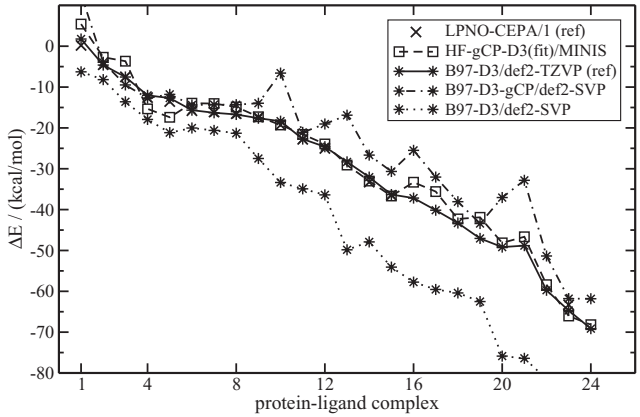

FIG. 6. B97-D3 and HF-D3 interaction energies for 24 protein-ligand complexes with and without gCP correction. The LPNO-CEPA/1 and B97-D3/def2-TZVP reference values are taken from Ref. 132.

An additional benefit of gCP is the practically vanishing computational cost.

A relaxed rotational energy profile is calculated for the FGF-tripetide for fixed dihedral angles as depicted in Figure 5. The influence of the IBSSE on peptide conformations is discussed, e.g., by van Mourik *et al.*[45,46] for the Tyr-Gly structure, where it was found that some conformations are obscured by IBSSE. For each level of theory, the central NCCN dihedral angle (black atoms in Figure 5) is varied by 20°. The remaining part of the tripeptide is optimized, while the dihedral angle is held fixed. The rotation profile for BLYP-D3/def2-QZVP, which serves as the reference, shows two minima: the lowest at −60° and a very flat minimum at 0°. The best agreement is reached with HF-gCP-D3/MINIS that gives an almost quantitatively correct transition barrier between both conformers and indicates the second minimum with a rather flat progression around 0°. PW6B95-D3/SV shows a too high barrier at −20° and continues its strong increase to 0°, not showing the flat region of BLYP-D3/def2-QZVP or HF-gCP-D3/MINIS. Applying PW6B95-gCP-D3/SV yields the quasi-correct (0.2 kcal/mol deviation) barrier and regains some of the flatness around 0°, although the basis set errors (intramolecular BSSE or BSIE) are still significant in this region. HF-gCP-D3/MINIS and PW6B95-gCP-D3/SV qualitatively reproduce the rest of the energy profile. The high sensitivity of peptide structure towards the level of theory is well-known[45,46,122,123] and basis set effects, as well as an accurate description of dispersion interactions are crucial to consider for high accuracy. We suggest gCP-corrected HF/DFT-D3 as an efficient tool in screening a large number of peptide conformations and pre-selecting conformations that can be subsequently subjected to high-level calculations.

Recently, Antony *et al.*[132] presented a set of protein-ligand complexes, where the drug molecule together with parts of the binding pocket of the protein within certain radii was cut out of X-ray structures. Reference energies for the smaller complexes could be obtained by LPNO-CEPA/1. Good agreement was found with B97-D3/def2-TZVP calculations that were proposed as reference values for the larger complexes. The performance of B97-D3/def2-SVP calculations with and without gCP correction is demonstrated in Figure 6, where both kinds of reference values are included. The overbinding of gCP-uncorrected B97-D3/def2-SVP is clearly visible for all interaction strengths. Relative errors without gCP can be large, e.g., for system 2, where the error is about 80% or about 40% for complex 18. The gCP correction (B97-D3-gCP/def2-SVP) reduces the errors to 13% and 12%, respectively. The HF-gCP-D3(fit)/MINIS results are also plotted in Figure 6. The agreement with the reference is better than for B97-D3-gCP/def2-SVP for this set which is very encouraging because this benchmark is very close to typical target applications in biochemistry. However, gCP cannot correct for inherent limitations by the basis set and a few outliers can be found; most notable system 1, where many flourine atoms and a flourine-hydrogen bond are present.

G. S22 optimizations with small basis sets

To systematically investigate the applicability of the gCP correction for geometry optimizations, the structures of the well-known S22 benchmark set[106] were optimized with HF-D3/MINIS, HF-D3/SV, and B3LYP-D3/6-31G*, applying a standard BB-CP optimization scheme and our gCP correction. The resulting structures are compared with the original S22 reference geometries that are based on MP2 calculations. We focus on the differences in the intermolecular center-of-mass distances as a measure of deviation from the reference structure, as shown in Figure 7. Only dispersion corrected results are presented.

As can be seen in Figure 7(a), HF-D3/SV results in too short distances (BSSE induced overbinding). Compared to the reference data, CP-optimization yields too short distances (underbinding), uncovering shortcomings of the model chemistry that can be in large parts addressed to a significant BSIE. The gCP correction improves the results compared to raw HF-D3/SV in all cases and (due to favorable error compensation) overall provides the best results compared to the reference data. This favorable error cancellation is not transferable to other methods as the results for B3LYP-D3/6-31G*

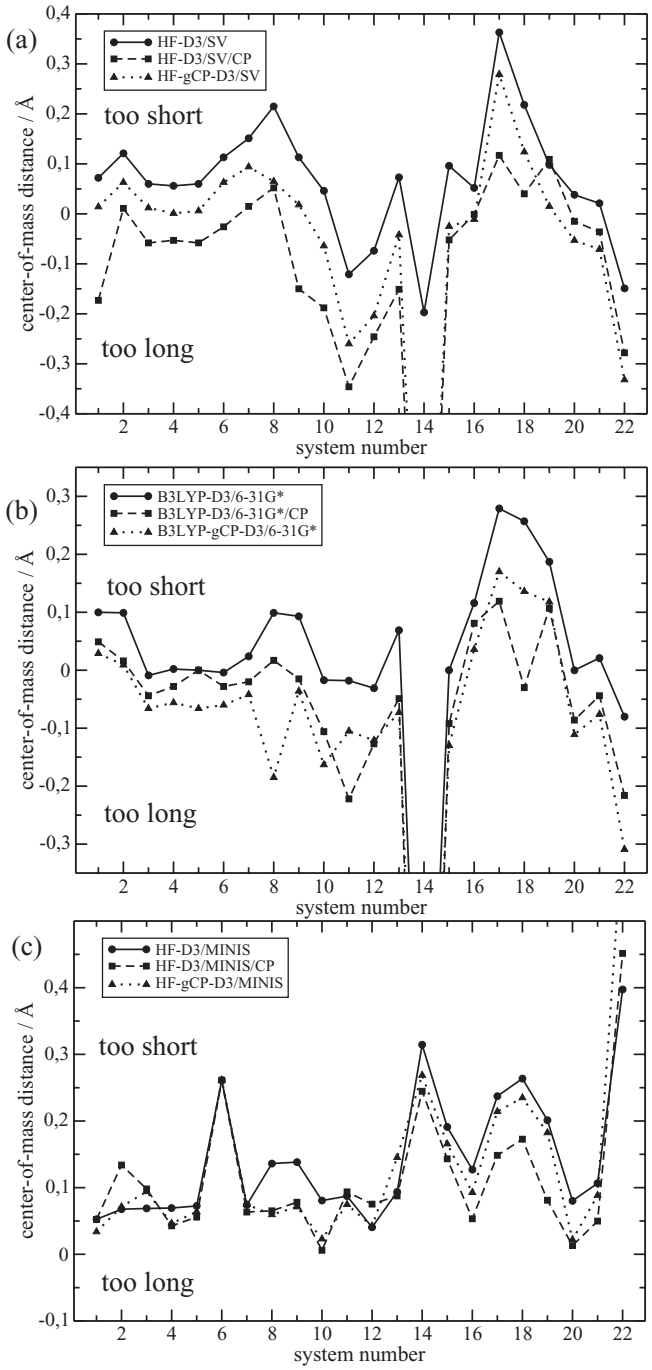

FIG. 7. Comparison of BB-CP- and gCP-optimized structures with uncorrected ones for the S22 benchmark set. Deviation from the center-of-mass distances of the MP2 reference structures is plotted in Å. (a) HF-D3/SV, (b) B3LYP-D3/6-31G*, and (c) HF-D3/MINIS.

show (Figure 7(b)), where the gCP optimization produces slightly longer distances than the CP-optimization. However, the correction emulates the CP-optimization fairly well, having larger deviations only for complex number eight (methane dimer) and number 18 (benzene···NH_3). The overall agreement between CP-optimized and gCP-optimized geometries is better than in the HF-D3/SV case. Large deviations can be seen in both plots for system 14 (indole···benzene complex) that undergoes a conformational change from a parallel to a T-shaped arrangement. HF-D3/SV does not show this peak, but this is an artifact from BSSE since the CP-optimizations yield the other conformer. The gCP scheme is able to reproduce this correctly.

Figure 7(c) shows the results for HF/MINIS. Notable is system six (2-pyridoxine 2-aminopyridine complex) that is not affected by BSSE at all. For most systems reproduces gCP the CP-optimized structures very well. Some minor deviations can be seen for the mixed-type complexes. Complex 22 (phenol dimer) gives a large deviation from the reference at this model chemistry, which is, however, specifically related to this complex and not to the gCP procedure.

With the exception of the sensitive systems 14 and 22, the general structural features of the S22 complexes are preserved with CP-optimized DFT-D3 in comparison to the MP2 reference data. The computationally much faster gCP method reproduces the CP-optimized structure well and seems a reliable alternative for larger systems. The good performance of HF-gCP-D3/MINIS in most cases is also notable and suggests this method for many problems in structural bio-chemistry as an alternative to semiempirical methods.

H. 9-Helicene

The helical structure of helicene polycyclic aromatic hydrocarbons is rather sensitive to the account of dispersion interactions in the quantum chemical treatment and intramolecular BSSE.[43] Hobza *et al.* showed that DFT-D is able to correct for missing dispersion interactions in DFT for this system.[43] Geometry optimizations with gCP are tested for HF-D3/SV on the 9-helicene structure. Artificial IBSSE stabilization for this system is revealed in a compressed helical geometry with too short inter-ring distances. The values for the two in Figure 8 shown, representative distances are compared to those of a HF-D3/def2-QZVP reference structure that can be considered almost IBSSE free.

The inter-ring distances for HF-D3/SV are significantly too short by about 6 and 10 pm, respectively.

The HF-gCP-D3/SV geometry optimization, on the other hand, yields an excellent agreement for the two non-bonded C–C distances, which are too short by only 1 and 3 pm. In passing it is noted that it takes a few days with the def2-QZVP

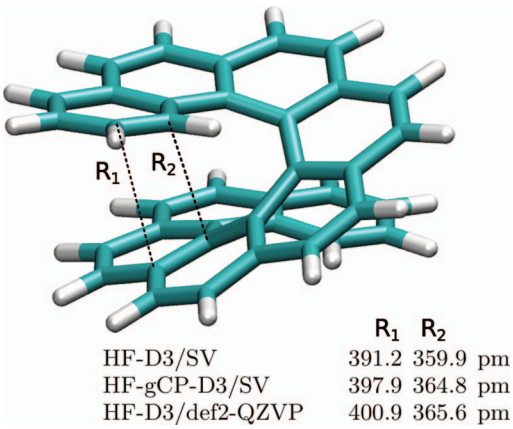

FIG. 8. HF-D3/def2-QZVP optimized structure of 9-helicene. Two intramolecular distances of raw and gCP-corrected HF-D3/SV structures are compared with the shown reference structure.

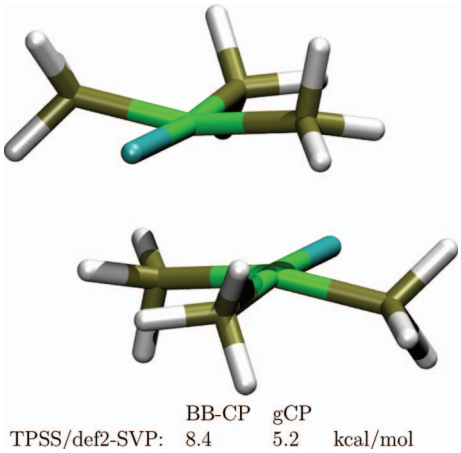

FIG. 9. TPSS-D3/def2-QZVP structure of the $(RhCl(PH_3))_2$ dimer.

basis to optimize 9-helicene, while the calculation with the small SV basis finished over night (same starting structure and a comparable number of optimization cycles on a typical workstation).

I. A transition metal dimer

The focus of this work clearly lies on organic and biological molecules. However, active centers in the latter often host transition metals (TMs). Currently, the gCP parameterization for TMs is available only for the basis sets def2-SVP and def2-TZVP and only for 3d metals. The earlier mentioned fall-back algorithm substitutes the parameters for 4d and 5d metals internally to those of the 3d series. The proof-of-principle for this approach is shown on a model system dimer of the square-planar Wilkinson catalyst,[133] where the PPh_3 ligands are replaced by PH_3. Normally one expects that the gross of the BSSE between two TM-complexes stems from the bulky, stabilizing ligands and the error of our fall-back algorithm is minor. The small PH_3 ligands makes the BSSE contributions from the (two) metal centers far more important and providing a worst case scenario with an upper bound error estimation in a system with larger ligands.

The dimer is optimized at the TPSS-D3/def2-QZVP level of theory (depicted in Figure 9). In this example, the program substitutes the Rh by Co atoms.

The evaluation of the BSSE for the interaction energy with gCP using the "dft/svp" parameters (fitted for B3LYP/def2-SVP) yields a somewhat too small but reasonable value of 5.2 kcal/mol in comparison to the "true" BB-CP result of 8.4 kcal/mol.

VI. CONCLUSIONS

A semi-empirical approach to correct for the BSSE in HF and DFT calculations is presented that requires only the geometric information of the molecule and no wave-function input. The approach dubbed gCP can be applied to supermolecular interaction energies in the spirit of Boys and Bernadi's counterpoise correction (BB-CP), but additionally is able to correct for intramolecular BSSE. The correction is constructed from overlap integrals over Slater functions and employs computed measures for the quality (incompleteness) of the target basis set. A few central points of the gCP scheme are:

1. Only four adjustable parameters are necessary to produce accurate fits (RMSD mostly between 0.1 and 0.4 kcal/mol) against BB-CP corrections of the S66×8 dimers. Adjustments for different DFT functionals are not necessary and we discriminate only between HF and DFT.
2. Low-order scaling behavior with system size and a small computational pre-factor for both, energy and gradient evaluations results in very fast computations even for thousands of atoms (gradient for 48 000 atoms below 2 min computation time on a typical workstation).
3. The gCP correction provides reasonable estimates for the intermolecular BSSE in the benchmark sets S66 and S22 with typical errors of 10%-30% for the BSSE. Common dispersion-corrections (DFT-D3 and DFT-NL) work well in combination with gCP for total interaction energies. For the polarized double-ζ basis sets 6-31G* and def2-SVP, HF and all considered functional types (GGAs, meta-GGAs, and hybrid functionals) perform quite well.
4. The analytical gradient has been applied successfully for optimizations of the non-covalent systems in the S22 set and for the 9-helicene structure where agreement for non-bonded C–C distances with the (almost IBSSE free) reference is obtained within a few pm.
5. For conformers of the small protein crambin and the tripeptide FGF, the significant intramolecular BSSE can be efficiently removed so that, e.g., in the case of crambin BLYP-D3/def2-TZVP, HF-gCP-D3/MINIS, and BLYP-gCP-D3/SV are in agreement within 10%-15% of the folding energy while deviations without gCP reach 50%. A relaxed rotational energy profile for the FGF-tripeptide proves sensible towards intramolecular BSSE, which could partly be removed for HF-D3/MINIS and PW6B95-D3/SV by gCP to yield results close to BLYP-D3/def2-QZVP quality.
6. Calculated interaction energies with B97-D3/def2-SVP and HF-D3/MINIS for a set of 24 large protein-ligand complexes are significantly improved by removing intermolecular BSSE with gCP. The structures includes fluorine, phosphorus, and chlorine atoms, which are not included within the gCP training set, demonstrating wide applicability of the scheme.
7. For a few water clusters, gCP shows good agreement with site-site function counterpoise computation of the BSSE, thus demonstrating its applicability also for multimer-BSSE.
8. The program *gcp* is supplied[61] as open-source to make the correction easily accessible to the community. It provides a user friendly fall-back algorithm to compute systems with unparameterized elements ($Z > 36$).
9. An important finding of our detailed investigations of various method/basis set combinations is that the minimal basis set MINIS together with HF-D3 and the gCP

correction yields exceptionally high accuracy for typical non-covalent interactions due to a relatively systematic error compensation. Hence we recommend this so-called HF-gCP-D3/MINIS method (or slightly enhanced variants) for studies on large bio-molecular systems as an alternative to DFT because it does not suffer from the self-interaction error of common density functionals. A comparison with the dispersion energy from SAPT calculations and Boys and Bernadi counterpoise corrections for a peptide-ethene potential energy curve shows that D3 and gCP (in HF-gCP-D3/MINIS) provide a physically sound description of the interaction energy components.

The diverse set of examples investigated here already provide a good initial basis for the validation of the gCP approach and future studies and applications likely will further solidify it. One yet unattended topic, e.g., is the impact of the correction for force constant (vibrational frequency) calculations and the derived statistical thermodynamic corrections.

ACKNOWLEDGMENTS

This work was supported by the DFG in the framework of the International Research Training Group 1444 "Generation of Supramolecular Functional Cavities – Container Molecules, Macrocycles, and Related Compounds" with a scholarship to H.K.

[1] F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, and J. H. van Lenthe, Chem. Rev. **94**, 1873 (1994).
[2] B. Liu and A. D. McLean, J. Chem. Phys. **59**, 4557 (1973).
[3] N. R. Kestner, J. Chem. Phys. **48**, 252 (1968).
[4] H. B. Jansen and P. Ros, Chem. Phys. Lett. **3**, 140 (1969).
[5] S. Boys and F. Bernardi, Mol. Phys. **19**, 553 (1970).
[6] I. S. Ufimtsev and T. J. Martinez, J. Chem. Theory Comput. **5**, 1004 (2009).
[7] K. A. Wilkinson, P. Sherwood, M. F. Guest, and K. J. Naidoo, J. Comput. Chem. **32**, 2313 (2011).
[8] C. Ochsenfeld, J. Kussmann, and D. S. Lambrecht, in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and T. R. Cundari (Wiley, Hoboken, NJ, USA, 2007), Vol. 23, pp. 1–82.
[9] C. Ochsenfeld, C. A. White, and M. Head-Gordon, J. Chem. Phys. **109**, 1663 (1998).
[10] G. D. Fletcher, D. G. Fedorov, S. R. Pruitt, T. L. Windus, and M. S. Gordon, J. Chem. Theory Comput. **8**, 75 (2012).
[11] N. J. Mayhall and K. Raghavachari, J. Chem. Theory Comput. **7**, 1336 (2011).
[12] R. Alizadegan, K. J. Hsia, and T. J. Martinez, J. Chem. Phys. **132**, 034101 (2010).
[13] H. M. Senn and W. Thiel, Angew. Chem., Int. Ed. **48**, 1198 (2009).
[14] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, Oxford, 1989).
[15] W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory* (Wiley VCH, New York, 2001).
[16] S. Grimme, J. Antony, T. Schwabe, and C. Mück-Lichtenfeld, Org. Biomol. Chem. **5**, 741 (2007).
[17] J. Šponer and P. Hobza, in *Encyclopedia of Computational Chemistry*, edited by P. von Ragué Schleyer (Wiley, New York, 2001).
[18] H. Kruse and S. Grimme, J. Phys. Chem. C **113**, 17006 (2009).
[19] N. Hansen, T. Kerber, J. Sauer, A. T. Bell, and F. J. Keil, J. Am. Chem. Soc. **132**, 11525 (2010).
[20] J.-M. Lehn, *Supramolecular Chemistry. Concepts and Perspectives* (VCH, Weinheim, 1995).
[21] N. Kannan and S. Vishveshwara, Protein Eng. **13**, 753 (2000).
[22] M. P. Waller, H. Kruse, C. Mück-Lichtenfeld, and S. Grimme, Chem. Soc. Rev. **41**, 3119 (2012).
[23] K. Müller-Dethlefs and P. Hobza, Chem. Rev. **100**, 143 (2000).
[24] A. J. Stone, *The Theory of Intermolecular Forces* (Oxford University Press, Oxford, 1997).
[25] I. G. Kaplan, *Intermolecular Interactions* (Wiley, Chichester, 2006).
[26] S. Grimme, WIREs Comput. Mol. Sci. **1**, 211 (2011).
[27] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).
[28] S. Grimme, S. Ehrlich, and L. Goerigk, J. Comput. Chem. **32**, 1456 (2011).
[29] W. Hujo and S. Grimme, J. Chem. Theory Comput. **7**, 3866 (2011).
[30] O. A. Vydrov and T. Van Voorhis, J. Chem. Phys. **133**, 244103 (2010).
[31] P. Canfield, M. G. Dahlbom, N. S. Hush, and J. R. Reimers, J. Chem. Phys. **124**, 024301 (2006).
[32] I. S. Ufimtsev and T. J. Martinez, J. Chem. Theory Comput. **5**, 2619 (2009).
[33] D. Hankis, J. W. Moskowitz, and F. H. Stillinger, J. Chem. Phys. **53**, 4544 (1970).
[34] P. Valiron and I. Mayer, Chem. Phys. Lett. **275**, 46 (1997).
[35] L. Turi and J. J. Dannenberg, J. Phys. Chem. **97**, 2488 (1993).
[36] P. Salvador and M. M. Szcześniak, J. Chem. Phys. **118**, 537 (2003).
[37] B. H. Wells and S. Wilson, Chem. Phys. Lett. **101**, 429 (1983).
[38] M. Gutowski and G. Chałasiński, J. Chem. Phys. **98**, 5540 (1993).
[39] I. Mayer, Int. J. Quantum Chem. **23**, 341 (1983).
[40] I. Mayer, Int. J. Quantum Chem. **70**, 41 (1998).
[41] J. Deng, A. T. B. Gilbert, and P. M. W. Gill, J. Chem. Phys. **135**, 081105 (2011).
[42] J. C. Faver, Z. Zheng, and J. K. M. Merz, J. Chem. Phys. **135**, 144110 (2011).
[43] H. Valdés, V. Klusák, M. Pitoňák, O. Exner, I. Starý, P. Hobza, and L. Rulíšek, J. Comput. Chem. **29**, 861 (2008).
[44] D. Moran, A. C. Simmonett, F. E. Leach, W. D. Allen, P. v. R. Schleyer, and H. F. Schaefer, J. Am. Chem. Soc. **128**, 9342 (2006).
[45] T. van Mourik, P. G. Karamertzanis, and S. L. Price, J. Phys. Chem. A **110**, 8 (2006).
[46] L. F. Holroyd and T. van Mourik, Chem. Phys. Lett. **442**, 42 (2007).
[47] R. M. Balabin, J. Chem. Phys. **129**, 164101 (2008).
[48] D. Asturiol, M. Duran, and P. Salvador, J. Chem. Phys. **128**, 144108 (2008).
[49] R. M. Balabin, J. Chem. Phys. **132**, 231101 (2010).
[50] A. Galano and J. R. Alvarez-Idaboy, J. Comput. Chem. **27**, 1203 (2006).
[51] F. Jensen, J. Chem. Theory Comput. **6**, 100 (2010).
[52] D. B. Cook, J. A. Sordo, and T. L. Sordo, Int. J. Quantum Chem. **48**, 375 (1993).
[53] R. Wieczorek, L. Haskamp, and J. J. Dannenberg, J. Phys. Chem. A **108**, 6713 (2004).
[54] I. Mayer and L. Turi, J. Mol. Struct.: THEOCHEM **227**, 43 (1991), and references therein.
[55] J. Řezáč, K. E. Riley, and P. Hobza, J. Chem. Theory Comput. **7**, 2427 (2011).
[56] C. C. J. Roothaan, Rev. Mod. Phys. **32**, 179 (1960).
[57] R. Ahlrichs, M. K. Armbruster, M. Bär, H.-P. Baron, R. Bauernschmitt, N. Crawford, P. Deglmann, M. Ehrig, K. Eichkorn, S. Elliott, F. Furche, F. Haase, M. Häser, C. Hättig, A. Hellweg, H. Horn, C. Huber, U. Huniar, M. Kattannek, C. Kölmel, M. Kollwitz, K. May, P. Nava, C. Ochsenfeld, H. Öhm, H. Patzelt, D. Rappoport, O. Rubner, A. Schäfer, U. Schneider, M. Sierka, O. Treutler, B. Unterreiner, M. von Arnim, F. Weigend, P. Weis, and H. Weiss, TURBOMOLE (version. 5.9 and 6.3), Universität Karlsruhe 2008 & 2011; see also http://www.turbomole.com.
[58] R. Ahlrichs, M. Bär, M. Häser, H. Horn, and C. Kölmel, Chem. Phys. Lett. **162**, 165 (1989).
[59] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, Chem. Phys. Lett. **240**, 283 (1995).
[60] A. Herman, Modell. Simul. Mater. Sci. Eng. **12**, 21 (2004).
[61] See website by the group of Prof. Stefan Grimme for a FORTRAN program implementing the gCP method: http://www.thch.uni-bonn.de/tc.
[62] A. Lofthus, Mol. Phys. **5**, 105 (1962).
[63] MAPLE 12, Maplesoft, http://www.maplesoft.com/, Waterloo ON, Canada.
[64] F. Neese, ORCA - an *ab initio*, density functional and semiempirical program package, Max Planck Institute for Bioinorganic Chemistry, D-45470 Muelheim/Ruhr, Germany, 2011.
[65] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[66] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[67] B. Miehlich, A. Savin, H. Stoll, and H. Preuss, Chem. Phys. Lett. **157**, 200 (1989).

[68] S. Grimme, J. Comput. Chem. **27**, 1787 (2006).
[69] Y. Zhang and W. Yang, Phys. Rev. Lett. **80**, 890 (1998).
[70] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, Phys. Rev. Lett. **91**, 146401 (2003).
[71] Y. Zhao and D. G. Truhlar, J. Phys. Chem. A **109**, 5656 (2005).
[72] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[73] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, J. Phys. Chem. **98**, 11623 (1994).
[74] L. Goerigk and S. Grimme, Phys. Chem. Chem. Phys. **13**, 6670 (2011).
[75] Y. Zhao and D. G. Truhlar, Theor. Chem. Acc. **120**, 215 (2008).
[76] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, GAUSSIAN 09, Revision A.02, Gaussian, Inc., Wallingford, CT, 2009.
[77] A. D. Becke and E. R. Johnson, J. Chem. Phys. **122**, 154101 (2005).
[78] E. R. Johnson and A. D. Becke, J. Chem. Phys. **123**, 024101 (2005).
[79] L. Goerigk and S. Grimme, J. Chem. Theory Comput. **7**, 291 (2011).
[80] A. Schäfer, H. Horn, and R. Ahlrichs, J. Chem. Phys. **97**, 2571 (1992).
[81] A. Schäfer, C. Huber, and R. Ahlrichs, J. Chem. Phys. **100**, 5829 (1994).
[82] F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).
[83] F. Weigend, F. Furche, and R. Ahlrichs, J. Chem. Phys. **119**, 12753 (2003).
[84] K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, and T. L. Windus, J. Chem. Inf. Model. **47**, 1045 (2007).
[85] D. Feller, J. Comput. Chem. **17**, 1571 (1996).
[86] W. J. Hehre, R. Ditchfeld, and J. A. Pople, J. Chem. Phys. **56**, 2257 (1972).
[87] R. Ahlrichs, Phys. Chem. Chem. Phys. **6**, 5119 (2004).
[88] F. Neese, J. Comput. Chem. **24**, 1740 (2003).
[89] F. Weigend, Phys. Chem. Chem. Phys. **4**, 4285 (2002).
[90] F. Weigend, M. Häser, H. Patzelt, and R. Ahlrichs, Chem. Phys. Lett. **294**, 143 (1998).
[91] K. Eichkorn, F. Weigend, O. Treutler, and R. Ahlrichs, Theor. Chem. Acc. **97**, 119 (1997).
[92] J. M. L. Martin, C. W. Bauschlicher, Jr., and A. Ricca, Comput. Phys. Commun. **133**, 189 (2001).
[93] B. N. Papas and H. F. Schaefer III, J. Mol. Struct.: THEOCHEM **768**, 175 (2006).
[94] S. Dressler and W. Thiel, Chem. Phys. Lett. **273**, 71 (1997).
[95] V. Termath and J. Sauer, Chem. Phys. Lett. **255**, 187 (1996).
[96] E. R. Johnson, R. A. Wolkow, and G. A. DiLabio, Chem. Phys. Lett. **394**, 334 (2004).
[97] J. Gräfenstein, D. Izotov, and D. Cremer, J. Chem. Phys. **127**, 164113 (2007).
[98] F. Neese, A. Hansen, and D. G. Liakos, J. Chem. Phys. **131**, 064103 (2009).
[99] F. Neese, F. Wennmohs, and A. Hansen, J. Chem. Phys. **130**, 114108 (2009).
[100] D. G. Liakos, A. Hansen, and F. Neese, J. Chem. Theory Comput. **7**, 76 (2011).
[101] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, Chem. Phys. Lett. **286**, 243 (1998).
[102] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, and J. Olsen, Chem. Phys. Lett. **302**, 437 (1999).
[103] B. Jeziorski and K. Szalewicz, in *Encyclopedia of Computational Chemisty*, edited by P. von Rague-Schleyer (Wiley, New York, 1998), Vol. 2, p. 1376.
[104] H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, *et al.*, MOLPRO, version 2010.1, a package of *ab initio* programs, 2010, see http://www.molpro.net.
[105] R. A. Kendall, T. H. Dunning, and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
[106] P. Jurečka, J. Sponer, J. Cerny, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).
[107] J. Řezáč, K. E. Riley, and P. Hobza, J. Chem. Theory Comput. **7**, 3466 (2011).
[108] L. Goerigk, H. Kruse, and S. Grimme, Chem. Phys. Chem. **12**, 3421 (2011).
[109] See supplementary material at http://dx.doi.org/10.1063/1.3700154 for cartesian coordinates of the crambin conformers and tables with mean deviations for the S66 and S22 sets.
[110] T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall, and C. D. Sherrill, J. Chem. Phys. **132**, 144104 (2010).
[111] E. R. Davidson and D. Feller, Chem. Rev. **86**, 681 (1986).
[112] J. A. Pople, in *Modern Theoretical Chemistry*, edited by H. F. Schaefer III (Plenum, New York, 1976), Vol. 4.
[113] W. Kołos, Theor. Chem. Acc. **51**, 219 (1979).
[114] K. Kitaura and K. Morokuma, Int. J. Quantum Chem. **10**, 325 (1976).
[115] W. Chen and M. S. Gordon, J. Phys. Chem. **100**, 14316 (1996).
[116] R. Cammi, R. Bonaccorsi, and J. Tomasi, Theor. Chem. Acc. **68**, 271 (1985).
[117] F. Neese, A. Hansen, F. Wennmohs, and S. Grimme, Acc. Chem. Res. **42**, 641 (2009).
[118] V. S. Bryantsev, M. S. Diallo, A. C.T. van Duin, and W. A. Goddard III, J. Chem. Theory Comput. **5**, 1016 (2009).
[119] H. Umeda, Y. Inadomi, T. Watanabe, T. Yagi, T. Ishimoto, T. Ikegami, H. Tadano, T. Sakurai, and U. Nagashima, J. Comput. Chem. **31**, 2381 (2010).
[120] M. Wojciechowski and M. Cieplak, BioSystems **94**, 248 (2008).
[121] G. Stracquadanio and G. Nicosia, Comput. Chem. Eng. **35**, 464 (2011).
[122] L. Goerigk and S. Grimme, J. Chem. Theory Comput. **6**, 107 (2010).
[123] D. Řeha, H. Valdes, Vondrášek, P. Hobza, A. Abu-Riziq, B. Crews, and M. S. de Vries, Chem.-Eur. J. **11**, 6803 (2005).
[124] The OPENBABEL Package, version 2.3.1, see http://openbabel.org.
[125] N. O'Boyle, M. Banck, C. James, C. Morley, T. Vandermeersch, and G. Hutchison, J. Cheminf. **3**, 33 (2011).
[126] J. J. P. Stewart, J. Mol. Model. **13**, 1173 (2007).
[127] M. Korth, M. Pitoňák, J. Řezáč, and P. Hobza, J. Chem. Theory Comput. **6**, 344 (2010).
[128] M. Korth, J. Chem. Theory Comput. **6**, 3808 (2010).
[129] J. J. P. Stewart, MOPAC2009, Stewart Computational Chemistry, Version 11.052W, available at http://openmopac.net.
[130] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, J. Comput. Chem. **25**, 1157 (2004).
[131] G. Schaftenaar and J. H. Noordik, J. Comput.-Aided Mol. Des. **14**, 123 (2000).
[132] J. Antony, S. Grimme, D. G. Liakos, and F. Neese, J. Phys. Chem. A **115**, 11210 (2011).
[133] J. A. Osborn, F. H. Jardine, J. F. Young, and G. Wilkinson, J. Chem. Soc. A 1711 (1966).