Systematic Evaluation of Counterpoise Correction in Density Functional Theory

Montgomery Gray, Paige E. Bowling, and John M. Herbert*
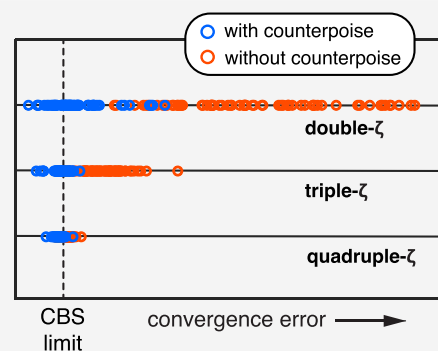
Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: A widespread belief persists that the Boys−Bernardi function counterpoise (CP) procedure "overcorrects" supramolecular interaction energies for the effects of basis-set superposition error. To the extent that this is true for correlated wave function methods, it is usually an artifact of low-quality basis sets. The question has not been considered systematically in the context of density functional theory, however, where basis-set convergence is generally less problematic. We present a systematic assessment of the CP procedure for a representative set of functionals and basis sets, considering both benchmark data sets of small dimers and larger supramolecular complexes. The latter include layered composite polymers with ∼150 atoms and ligand−protein models with ∼300 atoms. Provided that CP correction is used, we find that intermolecular interaction energies of nearly complete-basis quality can be obtained using only double-ζ basis sets. This is less expensive as compared to triple-ζ basis sets without CP correction. CP-corrected interaction energies are less sensitive to the presence of diffuse basis functions as compared to uncorrected energies, which is important because diffuse functions are expensive and often numerically problematic for large systems. Our results upend the conventional wisdom that CP "overcorrects" for basis-set incompleteness. In small basis sets, CP correction is mandatory in order to demonstrate that the results do not rest on error cancellation.

1. INTRODUCTION

1.1. Density Functional Theory for Noncovalent Interactions.

The description of noncovalent interactions within density functional theory (DFT) has improved dramatically over the past two decades, relative to a starting point where generalized gradient approximations (GGAs) predict interaction potentials that are unbound for dispersion-dominated complexes such as rare-gas dimers or even $(C_6H_6)_2$.[1] The first functionals to deal with this problem in a generally effective way were those with empirical atom−atom dispersion potentials added *ad hoc*.[1,2] These dispersion-corrected or "DFT+D" methods remain tremendously important in contemporary DFT,[1−3] and modern versions of the dispersion corrections resemble proper atomic C_6 coefficients in the separated-atom limit.[4,5] (Double counting of correlation effects at van der Waals contact distances does remain an issue, however.[1,6,7]) Early versions of the DFT+D approach[8] were quickly followed by semi-empirical meta-GGA functionals that significantly outperformed GGAs for van der Waals complexes,[9−12] although not always for the right reasons.[13−15] Rather, because the semilocal meta-GGA framework has a much longer range in real space, due to its dependence on the density Laplacian and the kinetic energy density, meta-GGAs afford nonvanishing interactions in the region of nonbonded close contacts (unlike typical GGAs), and noncovalent interactions may therefore emerge upon sufficient parameterization. Separately, the development of

nonlocal correlation functionals,[16−22] which have the right physics to describe dispersion from first principles, has paved the way for a new generation of empirical functionals[23−25] based on the B97 model.[26] At present, these B97-based functionals are among the best performers for noncovalent interactions, with accuracy approaching 1−2 kcal/mol for standard benchmark data sets.[3] Other methods attempt to extract *in situ* atomic C_6 (or C_8, C_{10}, ...) coefficients from a DFT density for a first-principles approach to modeling dispersion. Examples include the Becke−Johnson exchange-hole dipole moment method[27−32] and the Tkatchenko−Scheffler atoms-in-molecules polarizability approach.[33−35] Finally, there is double-hybrid (DH-)DFT,[36,37] where a fraction of the second-order Møller−Plesset (MP2) correlation energy is introduced. These functionals can also describe dispersion from first principles, although it is worth mentioning that MP2 itself is not a benchmark-quality method for dispersion interactions.[38−41]
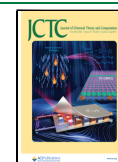
In DFT, noncovalent interaction energies ΔE_{int} are typically computed via the supramolecular approach:

$$\Delta E_{int} = E_{AB} - E_A - E_B \tag{1}$$

When atom-centered Gaussian basis sets are used, this approach suffers from the well-known deficiency of basis-set superposition error (BSSE).[42−46] This reflects the fact that in the dimer calculation (E_{AB}), either monomer may use basis functions centered on the other to improve the quality of its wave function, but that possibility is precluded in the monomer calculations (E_A and E_B). This leads to a significant overestimation of ΔE_{int} that disappears slowly as the monomer-centered basis sets approach completeness.[47−50] The Boys−Bernardi "function counterpoise" (CP) procedure is a means to sidestep this problem, which dates to the early days of quantum chemistry.[51] The CP procedure recognizes that eq 1 is an unbalanced approximation unless all three energies are computed in the same (dimer) basis set.

Despite its easy-to-understand provenance, use of the CP procedure has been controversial[44] and remains so in some more recent work.[52−54] The chronological evolution of this controversy is detailed in Section 1.2, but the gist is a perception that the CP procedure "overcorrects" ΔE_{int}.[42−44] When small basis sets are used, $|\Delta E_{int}|$ is much too large as compared to the complete basis set (CBS) limit computed at the same level of theory, but $|\Delta E_{int}|$ may be too small upon CP correction. This observation is not an indictment of the CP procedure *per se* because BSSE is interwoven with basis-set incompleteness error (BSIE). Due to partial compensation of these effects,[55] it has been suggested that fixing one of these issues (namely BSSE, by means of the CP procedure) without fixing the other (BSIE, by use of larger basis sets) results in an approximation that is also unbalanced.[53] The use of a "half-CP" procedure,[56] which averages the CP-corrected and uncorrected values of ΔE_{int}, can be seen as a remedy for this imbalance. For correlated wave function methods, the half-CP average converges to the CBS limit faster than either the corrected or uncorrected interaction energy on its own.[49,50,55]

For DFT, there is an additional consideration in that the performance for noncovalent interactions varies greatly from one functional to another, and finite-basis error may offset inherent functional error in some cases.[57] To the extent possible, one should therefore attempt to uncouple these errors by examining the performance of DFT in the CBS limit. Although the CP procedure has been carefully benchmarked for correlated wave function models,[50,55] its behavior in DFT calculations has not been examined so systematically. A recent comparison with explicitly correlated wave function results recommended CP correction for converging DFT to the CBS limit,[56] although the selection of basis sets and functionals was somewhat limited. Another assessment of dispersion-corrected GGA functionals using Dunning basis sets concluded that CP-corrected aug-cc-pVDZ results were close to quadruple-ζ results but that uncorrected aug-cc-pVDZ results were not.[58] The systematic examination contained herein considers a wider variety of functionals and basis sets that are more typically used for DFT calculations.

With the exception of double-hybrid functionals, it can be expected that BSSE is smaller for DFT than it is for correlated wave function methods due to the more rapid basis-set convergence of DFT that originates in the absence of electron coalescence cusps. Dependence on virtual orbitals also increases the BSSE associated with correlated wave function

methods.[52] The much slower convergence of MP2 as compared to traditional semilocal or hybrid DFT underlies a recent recommendation to use the half-CP procedure with double-hybrid functionals versus "full CP" for other functionals.[56] In the early days of molecular DFT calculations, small Pople-style basis sets were quickly judged to be adequate,[59] and this recommendation seems to have been ported to noncovalent problems without careful calibration. However, the conventional wisdom that double-ζ basis sets are adequate for DFT does not always hold for modern meta-GGAs and B97-based functionals, which converge more slowly than their predecessors with respect to both basis set[23,24,60] and quadrature grid.[24,61,62] In view of this, a thorough assessment of CP correction in DFT seems timely.

1.2. History of CP Correction. We consider the historical record in an effort to understand the emergence of a conventional wisdom that the Boys−Bernardi procedure "overcorrects" for BSSE. Evidence debunking that viewpoint had appeared already in reviews dating to the late 1980s and early 1990s;[42−44] nevertheless, this sentiment persists in more recent literature.[52−54,63]

Very early literature on the efficacy of CP correction is muddled by the use of low-quality basis sets, for which both BSSE and BSIE are sizable. It is difficult to take seriously the early criticisms of CP correction based on minimal-basis calculations,[64] where BSSE can be so large as to convert repulsive interactions into attractive ones.[65] There was also an early suggestion[66] that the Boys−Bernardi procedure erroneously allows the occupied orbitals of one monomer to occupy regions of space that should be taken up by occupied orbitals of the partner monomer in the supramolecular complex, thus causing the CP procedure to underestimate the true interaction energy. This view was later shown to be false, however.[67−69] In fact, spatial restriction of the occupied orbitals upon formation of the A···B complex lies at the heart of the Pauli exclusion principle![44]

The erratic nature of BSSE in small basis sets[70,71] partly motivated the development of intermolecular perturbation theory in the 1970s[72] because this approach is inherently free of BSSE.[73] Gaussian basis sets had become more systematized by the 1980s, and a 1988 review[74] observes that it is "generally accepted" that the CP correction brings Hartree−Fock interaction energies closer to the CBS limit, except in the case of minimal basis sets. However, the effectiveness of the CP procedure was questioned in a 1985 study by Schwenke and Truhlar,[75] who performed calculations on $(HF)_2$ using 34 different Pople-style basis sets and found no systematic improvement when CP was used. That said, there was little variation among the basis sets tested, and a subsequent study of alanine dimer using many of the same basis sets reached the opposite conclusion, namely, that CP correction typically improves the results.[76] Other work from the same time period suggested that the CP procedure actually insulates the results from otherwise erratic changes in ΔE_{int} upon relatively small changes in the basis set.[77]

This understanding was summarized in a 1994 review of the "overcorrection debate".[44] The Boys−Bernardi procedure is not only compatible with the Pauli principle,[67] but furthermore, CP-corrected supramolecular calculations afford results of comparable accuracy to intermolecular perturbation theories that are inherently free of BSSE.[43,67−69] (In fact, supramolecular perturbation theory calculations may afford nonsensical results for energy components if these are not

corrected for BSSE.[43]) Similarly, comparison of CP-corrected MP2 to the corresponding "chemical Hamiltonian" version of that method,[78] which is inherently free of BSSE, results in excellent agreement for basis sets of moderate quality.[79,80] By late 1990s (but arguably much earlier than that),[42−44] it was therefore evident that CP correction generally improves the quality of supramolecular interaction energies. Where exceptions can be found in the older literature, they are almost always attributable to the use of low-quality basis sets. Nevertheless, reviews by van Lenthe *et al.* in 1987[42] and by the same authors again in 1994[44] bemoan the persistent and widespread belief that the Boys−Bernardi procedure "overcorrects" as well as the notion that the CP correction is merely an "estimate" of the BSSE rather than a well-defined procedure that recognizes the origin of the problem (namely, imbalance in the supramolecular formula of eq 1) and eliminates it, essentially by definition.

To a significant extent, this older discussion was hindered by inability to reach the CBS limit for systems with more than a few atoms or to obtain reliable *ab initio* benchmarks. Those problems have been overcome, and correlated wave function benchmarks clearly indicate that CP-corrected and uncorrected energies converge to the CBS limit from opposite directions.[50] We regard this as an incompleteness problem rather than an "overcorrection" because the CP-corrected results extrapolate more smoothly to the CBS limit, resulting in smaller error bars.[50] (This behavior was originally predicted by Dunning.[81]) Half-CP results converge even faster,[50,55] and other molecular properties also converge more smoothly with CP correction than without.[47] Reported cases where uncorrected results converge faster to the CBS limit,[52] or where the basis-set error (relative to the CBS limit) is larger *with* CP correction,[54] typically involve systems where the difference between the CP-corrected and uncorrected results is quite small, whereas the systematic study in ref 50 considered a wide range of cases. As such, we conclude that there is nothing in wave function theory to suggest that CP correction is ill-advised.

1.3. Overview of the Present Work. Whereas the CP correction for correlated wave function methods has been considered systematically,[50,55] as has that of double-hybrid functionals,[56] the latter constitute a relatively small niche in the overall pantheon of DFT methods and the CP correction has not been considered systematically for other classes of functionals. In addition, we want to understand how basis sets of double- and triple-ζ quality behave, as these are the largest basis sets that are typically employed in DFT calculations. We will therefore examine the effects of CP correction for interaction energies computed using a set of density functionals that perform well for noncovalent interactions, using basis sets as small as 6-31G*. These tests span a range of system sizes because BSSE is size-extensive and large supramolecular complexes exhibit dramatically larger CP corrections as compared to small dimers. Finally, we will consider whether CP correction is more or less economical than simply enlarging the basis set, as a strategy to obtain converged DFT/CBS values of ΔE_{int}.

2. METHODS

Basis sets tested here include Dunning's correlation-consistent sequence, aug-cc-pVXZ (with X = D, T, and Q);[82,83] Karlsruhe "def2" basis sets through def2-QZVPD;[84,85] and Pople basis sets 6-31G*, 6-31+G*, 6-311G*, and 6-311+G*.[86] Pople basis sets use a common orbital exponent for s and p functions in a

given shell, which makes them much more efficient than alternatives with a comparable number of functions,[87,88] provided that the electronic structure program takes advantage of this simplification. As such, Pople basis sets continue to see widespread use in DFT calculations, especially in large systems. A double-ζ basis set called DZVP,[89] which is not in widespread use, is also tested for large supramolecular complexes because it is reported to afford accurate interaction energies for small dimers, without the need for CP correction.[90] (In the basis set exchange,[91] this basis set is called DGauss-DZVP.)

We selected a small set of functionals that perform well for noncovalent interactions.[3] These include BLYP+D3(BJ), a GGA functional that employs the D3 empirical dispersion correction[4] with a Becke−Johnson (BJ) damping function;[6] PBE0+D4, a hybrid GGA using the relatively new D4 dispersion correction;[5] ωB97X-V,[23] a hybrid GGA with nonlocal VV10 correlation;[21] and finally the hybrid meta-GGAs ωB97M-V[24] and M06-2X.[10] Extensive benchmarking suggests that ωB97X-V, ωB97M-V, and BLYP+D3(BJ) are among the best all-around options for noncovalent interactions,[3] although M06-2X remains widely used in that capacity and PBE0+D4 performs very well for large supramolecular complexes.[39] We also consider the dispersion-corrected M06-2X+D3(0) functional, where "D3(0)" indicates the original damping function developed for D3,[2,4] and finally the meta-GGA functional M06-L.[92] The latter performs less well for noncovalent interactions but exhibits exceptionally poor convergence properties with respect to basis set,[60] making for an interesting test. A few double-hybrid functionals are also tested: PBE-QIDH,[93] B2GP-PLYP,[94] ωB97X-2(LP),[95] and ωB97M(2).[24] For these functionals, the MP2 correlation energy is evaluated within the resolution-of-identity approximation using orbitals obtained from the underlying hybrid functional.

The overall performance of these methods is illustrated in Table 1 using several noncovalent data sets. Importantly, the

Table 1. Statistical Accuracy of DFT for Noncovalent Interactions

functional	RMSD (kcal/mol)[a]			
	NCED[b]	NCEC[c]	NCD[d]	S66[e]
BLYP+D3(BJ)	0.34	2.18	2.82	0.28
M06-L	0.55	2.20	1.87	1.16
M06-2X	0.43	2.52	0.99	0.34
ωB97X-V	0.24	0.64	1.23	0.20
ωB97M-V	0.18	0.48	1.13	0.23
PBE0+D4				0.50
ωB97X-2(LP)				0.38
ωB97M(2)				0.35

[a]Root-mean-square deviations (RMSDs) with respect to CCSD(T)/ CBS benchmarks. [b]Noncovalent "easy" dimers (NCED), from ref 3. [c]Noncovalent "easy" clusters (NCEC), from ref 3. [d]Noncovalent "difficult" dimers (NCD), from ref 3. [e]Evaluated as part of the present work.

DFT calculations used to compute these error statistics were performed at or near the CBS limit and are compared to benchmark values of ΔE_{int} obtained at the level of coupled-cluster theory with single, double, and perturbative triple excitations [CCSD(T)], also extrapolated to the CBS limit. As such, these error statistics probe the inherent accuracy of DFT itself, free from BSSE or BSIE. The data in Table 1 are

Table 2. Basis-Set Convergence Errors for S66 Interaction Energies

| | mean absolute error (kcal/mol), vs DFT/CBS[a] | | | | | | | | | | | |
| | M06-2X | | M06-L | | BLYP+D3(BJ) | | PBE0+D4 | | ωB97X-V | | ωB97M-V | |
basis set	no CP	with CP	no CP	with CP	no CP	with CP	no CP	with CP	no CP	with CP	no CP	with CP
aug-cc-pVDZ	0.75	0.14	0.79	0.43	0.71	0.16	0.66	0.06	0.65	0.07	0.79	0.08
aug-cc-pVTZ	0.14	0.07	0.85	0.49	0.11	0.07	0.10	0.05	0.12	0.04	0.17	0.02
aug-cc-pVQZ	0.05	0.05	0.20	0.20	0.04	0.04	0.04	0.04	0.03	0.03	0.02	0.02
def2-SVP	1.54	0.43	1.48	0.60	2.54	0.26	1.85	0.28	1.82	0.36	2.04	0.37
def2-SVPD	1.38	0.14	1.24	0.57	1.52	0.13	1.37	0.09	1.36	0.10	1.56	0.18
def2-TZVP	0.26	0.19	0.82	0.58	0.48	0.15	0.35	0.13	0.34	0.12	0.42	0.10
def2-TZVPD	0.08	0.14	0.75	0.54	0.09	0.13	0.09	0.09	0.08	0.08	0.12	0.05
def2-QZVP	0.07	0.11	0.59	0.51	0.12	0.06	0.08	0.04	0.07	0.04	0.10	0.03
def2-QZVPD	0.05	0.09	0.59	0.52	0.02	0.05	0.02	0.04	0.01	0.04	0.02	0.03
6-31G*	1.36	0.52	1.42	0.75	2.27	0.39	1.62	0.39	1.62	0.42	1.84	0.44
6-31+G*	0.63	0.28	0.96	0.68	0.54	0.29	0.62	0.24	0.56	0.24	0.63	0.25
6-311G*	1.64	0.39	1.76	0.78	1.86	0.41	1.63	0.34	1.67	0.32	1.95	0.37
6-311+G*	1.07	0.45	1.35	0.82	0.69	0.37	0.93	0.32	0.88	0.33	1.02	0.43
6-311++G**	0.75	0.28	1.08	0.70	0.43	0.23	0.96	0.32	0.61	0.18	0.75	0.29

[a]DFT/CBS is defined as the average of CP-corrected and uncorrected aug-cc-pVQZ interaction energies.

intended to demonstrate what level of accuracy is presently feasible with available exchange−correlation functionals. In the rest of this work, we will mostly be concerned with convergence to the CBS limit for a particular functional and therefore subsequent error statistics will be defined with respect to that limit (DFT/CBS), rather than comparing to CCSD(T)/CBS results.

Often in DFT calculations, quadruple-ζ energies are taken to be equivalent to the CBS limit.[38,58,60] Except for double-hybrid functionals, we will define the DFT/CBS limit using the average of CP-corrected and uncorrected values of ΔE_{int} computed using a quadruple-ζ basis set, namely, aug-cc-pVQZ for small dimers and def2-ma-QZVP[88] for larger systems. For dimers in the S66 data set,[96] CP correction at the aug-cc-pVQZ level typically amounts to <0.1 kcal/mol, indicating convergence to the DFT/CBS limit. For larger systems, the magnitude of the CP correction in the def2-ma-QZVP basis set affords an estimate of the uncertainty in the DFT/CBS limit. For double-hybrid functionals, we estimate the CBS limit using the SCF/aug-cc-pVQZ hybrid DFT energy plus a two-point "T/Q" extrapolation of the MP2 correlation energy, where the latter is computed using aug-cc-pVTZ and aug-cc-pVQZ.[97]

All calculations were performed using Q-Chem v. 5.4.[98] Integral screening and shell-pair drop tolerances were both set to $\tau_{ints} = 10^{-12}$ a.u., and the self-consistent field (SCF) convergence threshold was set to $\tau_{SCF} = 10^{-8}$ Ha, except for the protein−ligand systems in Section 3.3.2, for which we use $\tau_{ints} = 10^{-10}$ a.u. and $\tau_{SCF} = 10^{-6}$ Ha. The SG-2 quadrature grid is used for all B97-based functionals and SG-3 for the Minnesota functionals.[62] For other functionals, we use SG-1.[99]

3. RESULTS AND DISCUSSION

3.1. S66 Data Set. We will use the S66 data set[96] to obtain a systematic understanding of the effect of CP correction on ΔE_{int}. The largest of the S66 dimers contains 34 atoms (pentane dimer), and this small size allows us to use quadruple-ζ calculations even for computationally demanding functionals and thus to definitively establish the CBS limit.

3.1.1. Convergence Errors for Conventional DFT. A statistical summary of basis-set convergence errors for S66 is

provided in Table 2, where the errors are defined with respect to the DFT/CBS limit for each functional. More detailed statistics can be found in Tables S1−S7, including a breakdown into subsets consisting of hydrogen-bonded dimers, dispersion-bound dimers, and dimers of mixed-influence interactions. Data for M06-2X+D3(0) are omitted from Table 2 but can be found in Table S2.

It proves useful to examine Table 2 by class of basis set. For the Dunning basis sets, CP correction has a significant effect only at the double-ζ level. For aug-cc-pVDZ, the mean absolute error (MAE) with respect to the DFT/CBS limit is reduced from 0.7 to 0.2 kcal/mol (averaging across functionals) when the CP correction is included. (These statistics exclude the double-hybrid functionals, whose basis-set convergence is quite different and which are considered later.) With the exception of the M06-L functional, this difference of ≈0.5 kcal/mol between CP and non-CP interaction energies is larger than the difference between DFT/CBS and CCSD(T)/CBS interaction energies for the same data set (Table 1). When the basis set is extended to aug-cc-pVTZ, the CP correction is nearly negligible (≲0.1 kcal/mol), except in the case of M06-L, and for aug-cc-pVQZ it is negligible in all cases. Notably, the CP-corrected aug-cc-pVDZ convergence errors are generally close to the *uncorrected* aug-cc-pVTZ values, such that results of triple-ζ quality can be obtained at double-ζ cost by applying CP correction.

Most practical DFT calculations do not employ correlation-consistent basis sets, so we next turn to the Karlsruhe basis sets that were designed specifically for SCF calculations. We tested versions with[85] and without[84] augmentation by diffuse functions, finding that diffuse functions afford systematically smaller convergence errors both before and after CP correction. Otherwise, the behavior of the Karlsruhe basis sets is similar to that of the Dunning basis sets, albeit with somewhat larger convergence errors, especially at the double-ζ level. Similar to the aug-cc-pVDZ case, however, def2-SVPD with CP correction is sufficient to reduce the convergence error so that it is smaller than the intrinsic accuracy of the functional itself. In the absence of the CP correction, this cannot be said of the double-ζ basis sets. A corollary is that once again, triple-ζ basis sets are not required to reach the

Table 3. Basis-Set Convergence Errors for S66 Interaction Energies

| | mean absolute error (kcal/mol) vs DFT/CBS | | | | | | | | |
| | MP2 | | PBE-QIDH | | B2GP-PLYP | | ωB97X-2(LP) | | ωB97M(2) | |
basis set	no CP	with CP	no CP	with CP	no CP	with CP	no CP	with CP	no CP	with CP
aug-cc-pVDZ	1.70	0.80	1.31	0.18	1.29	0.24	1.54	0.28	1.34	0.18
aug-cc-pVTZ	0.78	0.32	0.96	0.10	1.01	0.11	1.31	0.14	0.99	0.11
aug-cc-pVQZ	0.28	0.16	0.36	0.48	0.36	0.10	0.48	0.13	0.39	0.11
def2-SVP	1.02	2.27	1.52	0.73	1.70	0.87	1.46	1.27	1.65	0.86
def2-SVPD	3.31	0.99	2.35	0.18	2.41	0.26	2.72	0.35	2.43	0.18
def2-TZVP	0.31	0.99	0.42	0.29	0.48	0.31	0.48	0.41	0.44	0.32
def2-TZVPD	0.79	0.63	0.57	0.17	0.61	0.18	0.79	0.22	0.56	0.16
def2-QZVP	0.13	0.34	0.31	0.10	0.34	0.10	0.41	0.13	0.33	0.09
def2-QZVPD	0.27	0.28	0.39	0.09	0.42	0.09	0.57	0.12	0.41	0.09
6-31G*	0.89	2.34	1.21	0.85	1.38	0.93	1.09	1.32	1.34	0.93
6-31+G*	0.77	2.00	0.92	0.61	0.84	0.73	0.85	0.85	0.86	0.66
6-311G*	0.84	2.12	1.36	0.69	1.41	0.82	1.22	1.07	1.42	0.77
6-311+G*	1.07	1.78	1.20	0.55	1.06	0.66	1.13	0.81	1.08	0.57
6-311++G**	0.48	0.39	0.99	0.39	0.86	0.51	0.96	0.66	0.88	0.43

intrinsic accuracy limit, as CP-corrected def2-SVPD convergence errors are within ~0.1 kcal/mol of def2-QZVPD values.

Pople basis sets remain workhorses in DFT, and it is therefore interesting to note that in order to reduce convergence errors below 1.0 kcal/mol, CP correction is required in most (although not all) cases. With the exception of the slowly convergent M06-L functional, convergence errors for all CP-corrected functionals are ≲0.5 kcal/mol. These errors are systematically smaller when diffuse functions are added, yet CP correction is still required to reduce the convergence error to something comparable to the intrinsic accuracy of the functional in the DFT/CBS limit.

So far, this discussion has been confined to error statistics with respect to the DFT/CBS limit, but it is important not to lose sight of accuracy with respect to benchmark interaction energies. Table S8 lists error statistics with respect to CCSD(T)/CBS benchmarks[96] for the functionals and basis sets in Table 2. With the exception of the notably less accurate M06-L functional, the errors fall within <0.6 kcal/mol of the benchmarks when CP correction is employed but exceed 2 kcal/mol in some double-ζ basis sets when the CP correction is omitted. With CP correction and diffuse functions (i.e., excluding the basis sets 6-31G* and def2-SVP), the double-ζ errors are ≲0.4 kcal/mol. CP correction is thus crucial not only for achieving the CBS limit but also for improving the accuracy of small-basis interaction energy calculations.

3.1.2. Convergence Errors for Double-Hybrid Functionals. Convergence error statistics with respect to the CBS limit are summarized for double-hybrid functionals in Table 3. Inclusion of MP2 correlation means that we expect the convergence behavior of these functionals to resemble that of correlated wave function methods, which is generally much slower than DFT, and the convergence behavior of MP2 itself is also examined in Table 3. We observe that MP2 convergence errors in Pople basis sets and in def2-SVP are larger with CP correction than without, which can simply be ascribed to the fact that these basis sets are not appropriate for a correlated wave function method. A more incisive explanation is gleaned by examining the raw data for S66 interaction energies computed at the MP2 level (Table S9), which reveals that MP2 severely underestimates ΔE_{int} for the dispersion-bound subset of S66 when these basis sets are employed. Since the

absence of CP correction leads to a larger interaction energy, there is a partial error cancellation that is upset by CP correction. This behavior contrasts with the systematically smaller convergence errors obtained when CP correction is applied to GGA and hybrid functionals, and it sets up a competition between that behavior and the behavior of MP2 correlation in the case of double-hybrid functionals, when very small basis sets are employed. This leads to slightly less systematic convergence behavior than what was observed for GGA and hybrid functionals; CP correction generally reduces the convergence errors, but by amounts that are somewhat functional-dependent.

That said, if we exclude def2-SVP and the Pople basis sets, then convergence to the DH-DFT/CBS limit is reasonable and is facilitated by CP correction, especially in double-ζ basis sets such as def2-SVPD. A graphical example is presented in Figure 1 for Karlsruhe basis sets; see Figures S1−S3 for additional functionals and basis sets. Using CP correction in conjunction with def2-SVPD results in a mean convergence error of 0.2 kcal/mol, comparable to the CP-corrected def2-TZVPD error. While diffuse functions prove to be important (e.g., the def2-TZVP convergence error is a bit larger at 0.3 kcal/mol, even with CP correction), this analysis demonstrates that
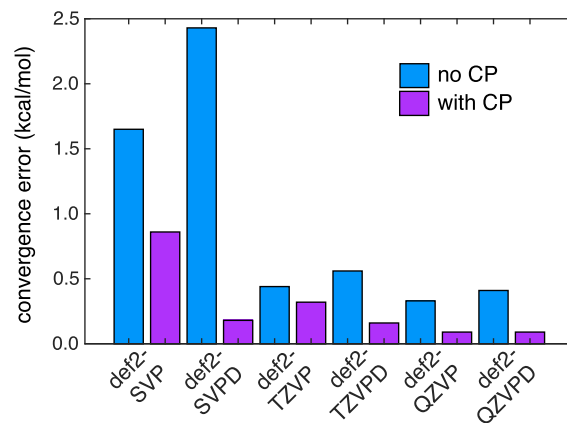

Figure 1. Mean absolute convergence errors (with respect to the CBS limit) for S66 interaction energies computed using the double-hybrid functional ωB97M(2) in various Karlsruhe basis sets.

correlation-consistent basis sets are not required to reach the DH-DFT/CBS limit, which can facilitate significant cost savings. Convergence errors remain larger in Pople basis sets (see Figure S3) and are only somewhat mitigated (in most cases) by CP correction. These basis sets are therefore not recommended for double-hybrid functionals.

3.1.3. Error Distributions. Figure 2 plots how the basis-set convergence errors are distributed with respect to zero for one
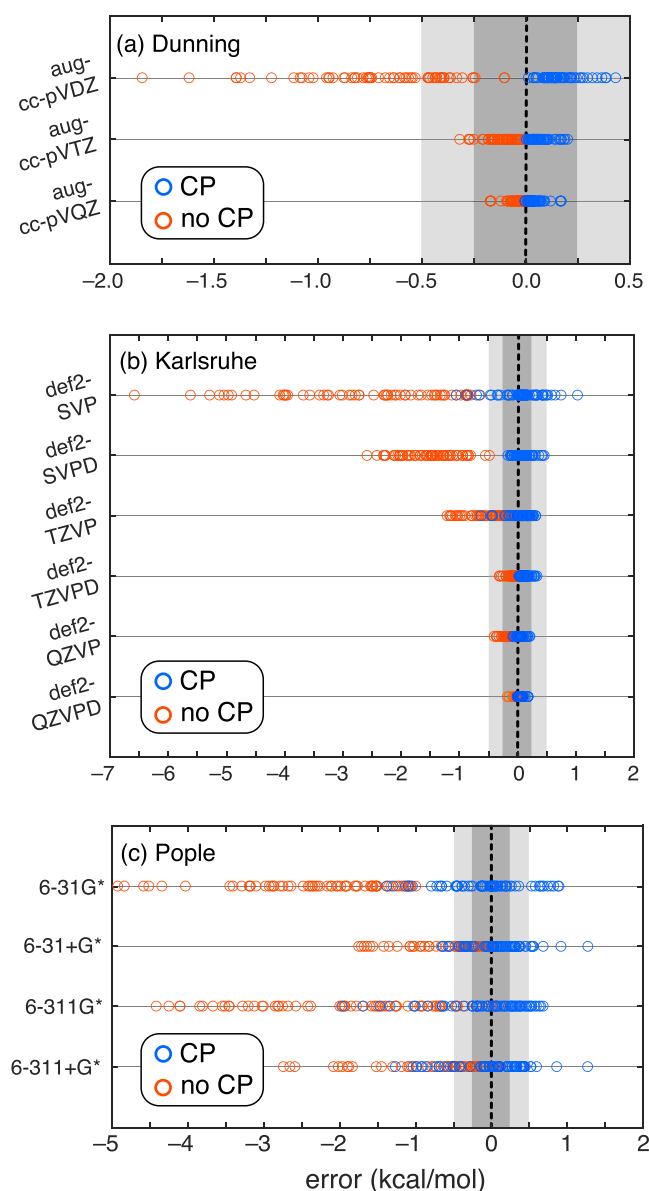

Figure 2. Convergence errors (with respect to the DFT/CBS limit) for the BLYP+D3(BJ) functional using (a) Dunning, (b) Karlsruhe, and (c) Pople basis sets. Gray rectangles delineate errors of ± 0.25 and ± 0.50 kcal/mol.

particular functional using a variety of basis sets and comparing both CP-corrected and uncorrected results. (For other functionals, including double hybrids, see Figures S4 and S5.) In larger basis sets, the error distributions narrow considerably and def2-TZVPD is the smallest basis set for which all the data points lie within ± 0.5 kcal/mol of the CBS limit, which is true in that case even without CP correction. For aug-cc-pVDZ and def2-SVPD, the CP-corrected errors are

much more tightly grouped together as compared to the uncorrected results, with a mean signed error that tilts toward underbinding but by a much smaller amount as compared to the overbinding that is observed in the absence of CP correction. In the aug-cc-pVXZ data, one can perhaps find justification for the statement that "CP overcorrects for BSSE", yet it is equally true that the uncorrected results systematically underbind the S66 complexes, and as a result the half-CP average is more reliable than either the CP-corrected or uncorrected values of ΔE_{int}. That said, CP correction reduces the MAE with respect to the CBS limit and also significantly reduces the spread of the convergence errors. Quantitative data for each functional and basis set, as measured by root-mean-square deviations from the DFT/CBS limit, can be found in Tables S15 and S16.

These trends are more muddled in the case of Pople basis sets, where CP-DFT does not consistently under- or overbind the complexes, and the distribution of CP-DFT convergence errors is centered close to zero. For example, in the case of the BLYP+D3(BJ) functional, the mean signed errors range from −0.07 to +0.10 kcal/mol for the four Pople basis sets shown in Figure 2c; see Table S4 for the quantitative data. If CP correction is neglected, the dimers are generally overbound, with the mean signed errors ranging up to −2.3 kcal/mol. Karlsruhe basis sets (Figure 2b) represent something of an intermediate case, in that CP-corrected results are not always underbound but convergence to the CBS limit is systematic.

Other functionals behave similarly to BLYP+D3(BJ), except for the problematic M06-L (Figure S4), for which CP-corrected convergence errors show a significant spread on both sides of zero and the distributions of CP-corrected and uncorrected values are difficult to distinguish. As such, CP correction may either improve or degrade the results but does so more or less at random, indicative of an overall failure of systematic convergence to the CBS limit.[60] For the double-hybrid functionals (Figure S5), Dunning and Karlsruhe basis sets behave similarly to what is seen for BLYP+D3(BJ) in Figure 2b,c, but Pople basis sets clearly underbind the complexes when CP correction is applied, whereas the results without CP correction are overbound.

In general, it can be said that the CP-corrected convergence errors span a smaller range as compared to convergence errors when CP correction is omitted. This discrepancy is especially prevalent in double-ζ basis sets and in all Pople basis sets. These observations parallel what has been observed for wave function methods,[50] although the magnitude of the CP correction is different, reflecting the different convergence behavior of DFT. The same trends hold for double-hybrid functionals if Pople basis sets (for which CP correction has only a modest effect) are excluded.

3.1.4. Timing Data. We have observed that CP-corrected double-ζ calculations often afford interaction energies on par with uncorrected triple-ζ results, so it is interesting to consider which of these methods is less expensive. We assess this using pentane dimer as it is the largest of the S66 complexes, and Table S17 presents timing data using each of the functionals and basis sets whose convergence statistics are given in Table 2. In any given basis set, CP correction adds about a factor of 2.5× to the cost of the calculation, which reflects the cost of performing two monomer calculations in the dimer rather than the monomer basis set. This additional overhead pales in comparison to the increased cost of extending the basis set from double- to triple-ζ, however. All three double-ζ

calculations in the dimer basis set can typically be performed in 25−50% of the time required for a single triple-ζ calculation of the dimer. If a triple-ζ level of convergence is desired, then CP correction is about 4× less expensive than a single quadruple-ζ calculation on the dimer.

An illustration of the trade-off between accuracy and cost is presented in Figure 3 using convergence errors for the whole of
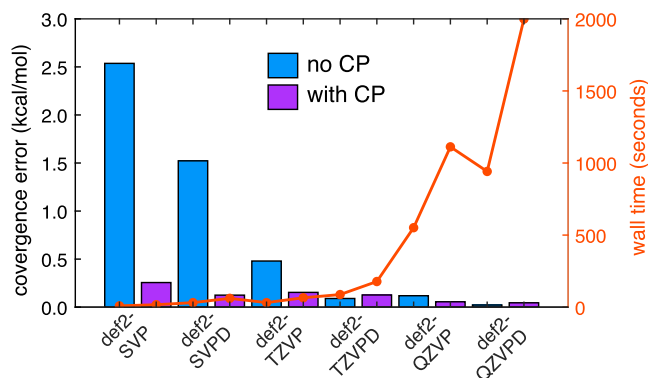

Figure 3. Mean absolute convergence errors for BLYP+D3(BJ) calculations of the S66 complexes (bar graph, axis at left) vs wall time required for pentane dimer (orange dots, axis at right). Calculations were performed on a single compute node with 14 processors.

S66, with BLYP+D3(BJ) and Karlsruhe basis sets. (For other functionals and basis sets, see Figures S6−S8.) Timing data are shown for pentane dimer. In the absence of CP correction, double-ζ basis sets afford unacceptably large convergence errors (1.5−2.5 kcal/mol on average), which the CP correction reduces below 0.25 kcal/mol, smaller than the intrinsic accuracy of the functional for this data set. All three calculations required for the CP-corrected double-ζ result can be obtained in about 75% of the compute time required for a single triple-ζ calculation, yet basis-set convergence errors and accuracy with respect to CCSD(T)/CBS are comparable to those obtained using def2-TZVPD. This comparison demonstrates that a CP-corrected double-ζ calculation is a cost-effective strategy to obtain results of triple-ζ quality.

3.2. (Coronene)$_2$ at Displaced Geometries. The S66 calculations reported above were performed exclusively at equilibrium geometries. Since BSSE arises from overlapping monomer-centered basis functions, it should be sensitive to

monomer separation, so we next examine results for coronene dimer, $(C_{24}H_{12})_2$, as a function of intermolecular separation. We placed the dimer in an eclipsed cofacial ("sandwich") orientation at intermolecular separations ranging from $R = 3$ to 10 Å. Potential energy curves, both with and without CP correction, are shown in Figure 4 for the ωB97M-V functional using Karlsruhe basis sets. Results for other functionals and basis sets can be found in Figures S9 and S10 but are quite similar to the data in Figure 4.

In the absence of CP correction (Figure 4a), the minimum-energy separation of the dimer changes as a function of basis set, increasing from 3.50 Å in double-ζ basis sets to 3.75 Å in larger basis sets. The potential curve is also much shallower in the larger basis sets, consistent with too-large values of ΔE_{int} due to BSSE, and the effect is larger than it was for the smaller S66 dimers. The def2-SVP and def2-SVPD basis sets overestimate ΔE_{int} by 7.0 and 5.0 kcal/mol, respectively, as compared to the def2-QZVPD value. On the other hand, triple-ζ results are essentially converged to the CBS limit.

Upon CP correction (Figure 4b), all Karlsruhe basis sets afford very similar potential energy curves. There is no longer any meaningful distinction between def2-SVP and def2-SVPD results, both of which overestimate ΔE_{int} by about 2.0 kcal/mol relative to triple- and quadruple-ζ results that are indistinguishable from one another. Moreover, the shapes of the CP-DFT/double-ζ potentials are the same as the converged results, with the same minimum-energy separation of $R = 3.75$ Å. In the absence of CP correction, we conclude that double-ζ basis sets cannot be assumed to afford correct geometries.

A new CCSD(T)/CBS interaction energy has recently been reported for coronene dimer,[100] and in Table 4 we use that benchmark to assess errors in ωB97M-V interaction energies using Karlsruhe basis sets. Although this functional affords sub-kcal/mol errors for the S66 dimers, that level of accuracy is not realized here, where the errors remain 2−3 kcal/mol upon CP correction. However, in the absence of CP correction, the relative errors (as a percentage of the total interaction energy) are actually larger for the S66 data set, where they range from 3% (def2-QZVPD) to 38% (def2-SVP). For CP-corrected calculations, errors in the S66 data set range from 3% (def2-QZVPD) to 9% (def2-SVP). Combined with the (coronene)$_2$ results, this suggests an intrinsic accuracy of ≈10% for ωB97M-V, but that intrinsic limit cannot be reached in double-ζ basis sets unless CP correction is used.
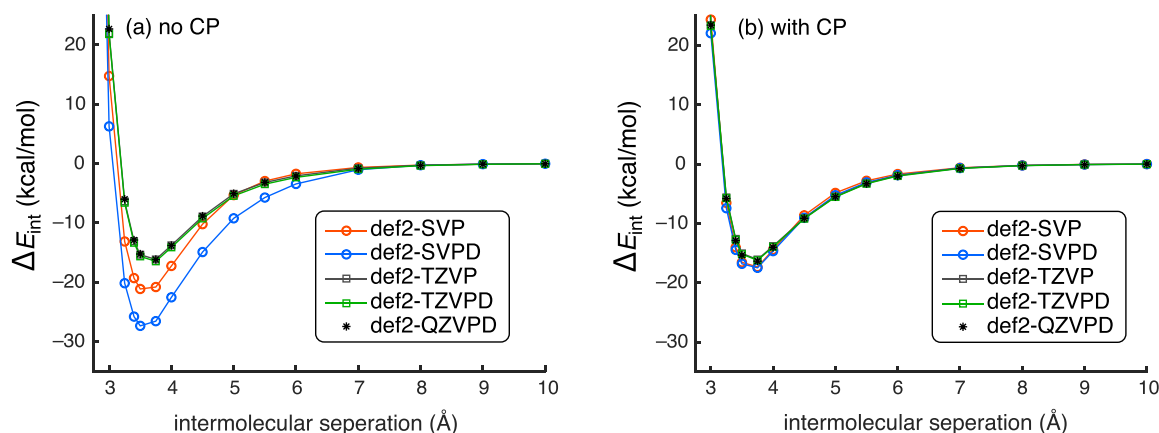

Figure 4. Potential energy profiles for the eclipsed-cofacial configuration of (coronene)$_2$, computed using ωB97M-V in various basis sets (a) without CP correction and (b) with CP correction.

Table 4. Accuracy of ωB97M-V Interaction Energies for Coronene Dimer

| | error vs CCSD(T)/CBS[a] | | | |
| | absolute (kcal/mol) | | relative (%) | |
basis set	no CP	CP	no CP	CP
def2-SVP	7.7	2.9	37	14
def2-SVPD	14.8	3.3	70	16
def2-TZVP	2.5	1.9	12	9
def2-TZVPD	2.8	2.0	13	10
def2-QZVP	2.5	2.2	12	11
def2-QZVPD	2.5	2.2	12	11

[a]Benchmark from ref 100.

A peculiarity in the coronene dimer results is that the def2-SVPD error is significantly larger (at 15 kcal/mol) than the def2-SVP error (8 kcal/mol). While CP correction reduces both errors to 3 kcal/mol, the larger error when diffuse functions are added defies the conventional wisdom that a good description of the tails of the density is important to the description of noncovalent interactions. In fact, larger errors for def2-SVPD than for def2-SVP are a feature of the dispersion-bound subset of S66 and also for every functional that we have examined (see Tables S1–S7), although the overall S66 convergence errors are reduced in the presence of diffuse functions. (This is primarily due to a reduction in convergence errors for the hydrogen-bonded subset of S66, and we have observed that diffuse functions are particularly important for the description of hydrogen bonds.[88]) For the dispersion-bound complexes in S66, results with def2-SVP lead to overbinding for all functionals examined, and since CP correction generally reduces interaction energies, these results benefit from some error cancellation that disappears when diffuse functions are added. The effect is less overbinding with def2-SVPD. However, this type of error cancellation feels fragile and from our point of view, CP-corrected results obtained using def2-SVPD provide a more reliable representation of the true performance of double-ζ basis sets.

3.3. Large Complexes. For the S66 dimes, CP-DFT in small basis sets can be used to obtain interaction energies that are within ∼0.2 kcal/mol of the DFT/CBS limit for the same functional, a convergence error that is smaller than the intrinsic accuracy of the best contemporary functionals for these same complexes. For coronene dimer, however, the CP correction is larger in double-ζ basis sets. In this section, we examine much larger molecular systems in order to understand how the magnitude of BSSE grows with system size. Systems examined include a set of layered complexes of oligothiophenes and graphene nanoribbons,[101,102] such as the example shown in Figure 5a. These range in size from 134 to 174 atoms and have ΔE_{int} in the range of 60−100 kcal/mol, depending on the functional and basis set. We also examine four protein−ligand systems ranging from 261 to 323 atoms, whose DFT/CBS interaction energies range from 15 to 100 kcal/mol; two of these are shown in Figure 5b,c. As in the calculations reported above, our primary goal is not to evaluate ΔE_{int} against some benchmark value but rather to examine convergence to the DFT/CBS limit and to assess the efficacy of CP correction in obtaining that limit.

3.3.1. Graphene−Oligothiophene Complexes. We examine several layered trimers of polythiophene (nPT, meaning n thiophene units) with graphene nanoribbons: (5PT)(C$_{38}$H$_{22}$)-
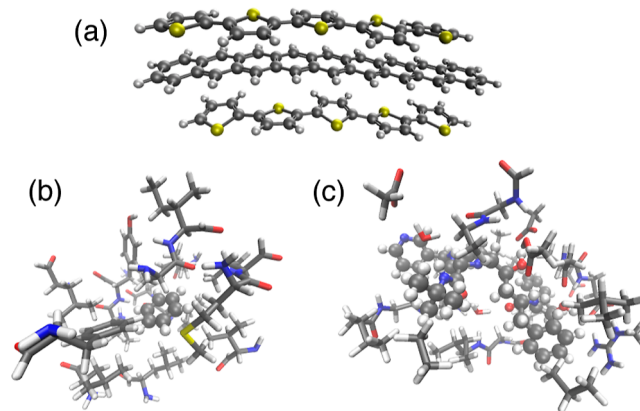

Figure 5. Examples of large molecular assemblies considered in this work: (a) (5PT)(C$_{38}$H$_{22}$)(5PT), (b) 181L:benzene, and (c) 1HSG:indinavir. In (b,c), the ligand is shown in a ball-and-stick representation and the protein binding site is shown using a tubular representation.

(5PT), (5PT)(C$_{59}$H$_{20}$)(5PT), and (7PT)(C$_{46}$H$_{26}$)(7PT). Geometries are taken from ref 102 and are provided in the Supporting Information. As representative functionals, we consider M06-2X and ωB97M-V using a subset of the basis sets examined above, mostly representative of practical choices for large systems but including the high-quality def2-TZVPD basis set. For the S66 dimers, the latter basis affords absolute convergence errors smaller than 0.15 kcal/mol for the two functionals considered here, and CP-DFT/def2-TZVPD affords interaction energies comparable to DFT/def2-QZVPD for the S66 dimers.

Interaction energies for these layered materials are listed in Table 5. The trends are similar for M06-2X and ωB97M-V, so our discussion will focus on the latter. When that functional is used with a double-ζ basis set, the difference between CP-corrected and uncorrected interaction energies is large, ranging from 12 to 48 kcal/mol. The DZVP basis set, which was previously found to afford accurate interaction energies in small-molecule data sets such as S66,[90] here affords CP corrections that are comparable to those obtained using 6-31+G* and certainly not small.

It should be noted that double-ζ basis sets are not recommended for use with ωB97M-V and similar, combinatorially optimized B97-based functionals.[23,24] The def2-SVP basis is specifically suggested to be "incompatible" with ωB97M-V,[24] and def2-TZVPPD is suggested as the smallest recommended basis set. These recommendations are based in part on thermochemical data and may be overly conservative for noncovalent interactions; thus, it is relevant to consider smaller double-ζ basis sets here. The M06-2X functional, which is routinely used with small Pople basis sets, actually converges to the DFT/CBS limit in a manner that is very similar to ωB97M-V, as measured by the difference between CP-corrected and uncorrected values of ΔE_{int}.

Despite the recommendation against double-ζ basis sets, with CP correction the ωB97M-V interaction energies are remarkably close to def2-TZVPD results, within 3 kcal/mol in all cases except 6-31G*. In fact, M06-2X exhibits slightly larger deviations between CP-corrected double- and triple-ζ results. It is perhaps surprising to note that diffuse functions have only a small effect when CP correction is used: $\lesssim$4 kcal/mol difference between def2-SVP and def2-SVPD results, and <0.5

Table 5. Interaction Energies (in kcal/mol) for Graphene−Oligothiophene Complexes

method	(5PT)(C$_{59}$H$_{20}$)(5PT)			(5PT)(C$_{38}$H$_{22}$)(5PT)			(7PT)(C$_{46}$H$_{26}$)(7PT)		
	no CP[a]	CP[b]	diff.[c]	no CP[a]	CP[b]	diff.[c]	no CP[a]	CP[b]	diff.[c]
ωB97M-V/DZVP	−105.37	−90.68	14.69	−78.24	−66.53	11.71	−104.60	−89.11	15.48
ωB97M-V/6-31G*	−100.81	−75.75	25.06	−76.91	−54.37	22.54	−102.31	−73.25	29.07
ωB97M-V/6-31+G*	−103.06	−88.95	14.11	−75.73	−65.26	10.48	−101.50	−87.45	14.05
ωB97M-V/def2-SVP	−110.15	−88.50	21.65	−82.03	−63.45	18.58	−110.21	−85.59	24.61
ωB97M-V/def2-SVPD	−133.97	−90.61	43.36	−102.99	−67.00	35.99	−137.17	−89.85	47.32
ωB97M-V/def2-TZVP	−89.70	−86.57	3.13	−65.38	−62.81	2.57	−87.69	−84.49	3.20
ωB97M-V/def2-TZVPD	−89.82	−86.67	3.15	−65.78	−63.12	2.66	−88.23	−84.89	3.34
M06-2X/DZVP	−88.00	−68.36	19.64	−70.69	−55.20	15.49	−94.51	−73.76	20.75
M06-2X/6-31G*	−71.20	−51.41	19.80	−58.91	−41.12	17.79	−78.49	−55.24	23.25
M06-2X/6-31+G*	−80.10	−66.22	13.88	−63.52	−53.40	10.12	−85.21	−71.26	13.95
M06-2X/def2-SVP	−84.89	−65.21	19.68	−67.15	−50.62	16.54	−90.46	−68.15	22.31
M06-2X/def2-SVPD	−109.26	−66.60	42.66	−89.59	−53.94	35.65	−119.53	−72.21	47.33
M06-2X/def2-TZVP	−65.53	−61.60	3.92	−52.91	−49.65	3.26	−71.02	−66.55	4.48
M06-2X/def2-TZVPD	−66.91	−62.41	4.51	−53.79	−50.14	3.65	−72.20	−67.05	5.15

[a]ΔE_{int} without CP correction. [b]ΔE_{int} with CP correction. [c]Difference between CP-corrected and uncorrected values of ΔE_{int}.

kcal/mol difference between def2-TZVP and def2-TZVPD. Presumably, the use of a supramolecular basis set facilitates a good description of the tails of the density, whereas in the absence of CP correction diffuse functions are needed to describe these tails in the monomer calculations. This observation is good news for calculations on systems of this size, where diffuse functions add significantly to the cost and also incur problems with linear dependencies. That said, when double-ζ basis sets are employed, CP correction is essential in order to mitigate the effect of the diffuse functions; without it, def2-SVP and def2-SVPD interaction energies differ by >20 kcal/mol in some cases. For triple-ζ basis sets, the effect of diffuse functions is $\lesssim 1$ kcal/mol even without CP correction.

Regarding the results with the smallest basis sets, such as 6-31G* (for which the CP-corrected and uncorrected values of ΔE_{int} differ by as much as 30 kcal/mol), we note that the ratio $|\Delta E_{int}(\text{no CP})/\Delta E_{int}(\text{CP})| \approx 1.4$. This is only slightly smaller than the ratio of ≈ 1.5 that is observed for the S66 dimers, using the same functional and basis set. For ωB97M-V/def2-SVP, this ratio is ≈ 1.3 for the graphene−oligothiophene systems. The similarity of these ratios suggests that neither the behavior of the functional nor that of the CP correction is fundamentally changed by moving to these larger systems. Instead, the larger convergence errors observed for the layered complexes reflects the size extensivity of BSSE, which is often overlooked.

When ωB97M-V is used with a larger def2-TZVPD basis set, the magnitude of the CP correction is reduced below 4 kcal/mol or about 4% of $|\Delta E_{int}|$ for these complexes. The ratio $|\Delta E_{int}(\text{no CP})/\Delta E_{int}(\text{CP})|$ is reduced to ≈ 1.04, which is nearly the same as the ratio of ≈ 1.03 that is obtained for S66 calculations in this basis set.

3.3.2. Protein−Ligand Complexes. Finally, we consider four protein−ligand complexes: 181L:benzene (Figure 5b),[103] 1LI2:phenol,[104] 1HSG:indinavir (Figure 5c),[105] and 1O44:(C$_{34}$H$_{35}$N$_3$O$_9$).[106] In the latter complex, the ligand is a malonic acid derivative known as RU85052, and we henceforth refer to this system as 1O44:RU85052. The indinavir and RU85052 complexes involve much larger ligands as compared to benzene and phenol, and the computed values of ΔE_{int} are correspondingly larger for these two systems.

The 1HSG:indinavir complex consists of a small ligand bound to HIV-2 protease and has become a model system for testing quantum-chemical methods for noncovalent interactions.[39,107−110] This structure was taken from ref 108. For the other three complexes, crystal structures were obtained from the protein data bank,[111] which were then protonated using the H++ web server.[112] The resulting structures were relaxed using the GFN2-xTB semi-empirical quantum chemistry method,[113] using a generalized Born model for water.[114] Relaxed structures were trimmed using a 5 Å cutoff (with respect to the ligand) for the 181L and 1LI2 complexes and a 2.5 Å cutoff in the case of 1O44, where the ligand is much larger. The coordinates that we used in the quantum chemistry calculations can be found in the Supporting Information.

We will again test M06-2X and ωB97M-V as representative functionals, using a variety of basis sets, and the corresponding interaction energies are listed in Table 6. The basis sets are standard except for minimally augmented ("ma") and heavy-augmented ("ha") versions of the Karlsruhe basis sets, which were introduced in ref 88. Due to the size of these complexes, we use the average of CP-corrected and uncorrected DFT/def2-ma-QZVP interaction energies to estimate the DFT/CBS limit for the ligand−protein complexes. By examining the entries for def2-ma-QZVP in Table 6, one can see that the two values that are averaged lie within 0.8 kcal/mol of one another when ωB97M-V is used and within 1.2 kcal/mol for M06-2X. This provides some measure of the uncertainty in establishing the DFT/CBS limit.

Broadly speaking, these systems exhibit the same trends observed above, namely, that CP corrections are sizable in double-ζ basis sets but modest in triple-ζ bases. For the more weakly interacting complexes where the ligand is benzene or phenol, CP correction moves ΔE_{int} about 2 kcal/mol closer to the CBS limit for M06-2X/6-31+G* and about 3 kcal/mol closer for ωB97M-V/6-31+G*. For ωB97M-V/def2-SVP, the CP correction reduces the convergence error by 1−2 kcal/mol, although for M06-2X/def2-SVP, the CP correction actually increases the convergence error by 1−3 kcal/mol. CP-ωB97M-V/6-31+G* results for 181L:benzene and 1LI2:phenol are essentially converged to the CBS limit, and the CP-M06-2X/6-31+G* results are only about 1 kcal/mol from that limit. In these cases, the CP correction improves the agreement with the DFT/CBS result by 3−4 kcal/mol or 20−25% of the total interaction energy. For the more strongly bound complexes

Table 6. Interaction Energies (in kcal/mol) for Ligand−Protein Complexes[a]

method	181L:benzene		1LI2:phenol		1HSG:indinavir		1O44:RU85052	
	no CP[b]	CP[c]	no CP[b]	CP[c]	no CP[b]	CP[c]	no CP[b]	CP[c]
ωB97M-V/DZVP	−22.4	−19.6	−25.0	−21.4	−146.9	−129.3	−119.9	−108.0
	(−1.9)	(+0.9)	(−3.2)	(+0.4)	(−16.7)	(+0.9)	(−12.0)	(−0.1)
ωB97M-V/6-31+G*	−23.3	−20.7	−25.5	−22.0	−143.2	−129.3	−120.2	−108.0
	(−2.8)	(−0.2)	(−3.7)	(−0.2)	(−13.1)	(+0.9)	(−12.2)	(+0.0)
ωB97M-V/def2-SVP	−23.9	−17.9	−27.0	−19.0	−175.9	−125.6	−136.9	−106.8
	(−3.4)	(+2.6)	(−5.2)	(+2.8)	(−40.8)	(+4.6)	(−28.9)	(+1.1)
ωB97M-V/def2-ma-SVP	−25.5	−18.7	−27.9	−20.1	−151.5	−127.6	−126.5	−108.8
	(−5.0)	(+1.8)	(−6.1)	(+1.7)	(−21.3)	(+2.6)	(−18.5)	(−0.8)
ωB97M-V/def2-ha-SVP	−28.9	−20.1	−30.0	−21.3	−157.9	−130.1	−128.2	−108.7
	(−8.4)	(+0.4)	(−8.2)	(+0.5)	(−27.7)	(+0.0)	(−20.2)	(−0.7)
ωB97M-V/def2-SVPD	−32.0	−20.5	−33.0	−21.6	−167.5	−131.0	−132.7	−109.1
	(−11.4)	(+0.1)	(−11.2)	(+0.2)	(−37.3)	(−0.9)	(−24.7)	(−1.1)
ωB97M-V/def2-TZVP	−21.6	−20.5	−23.5	−21.7	−138.9	−130.6	−112.6	−107.8
	(−1.1)	(+0.0)	(−1.7)	(+0.1)	(−8.7)	(−0.4)	(−4.7)	(+0.2)
ωB97M-V/def2-ma-TZVP	−21.7	−20.5	−23.3	−21.8	−134.4	−130.1	−110.8	−107.8
	(−1.1)	(+0.0)	(−1.5)	(+0.0)	(−4.3)	(+0.1)	(−2.8)	(+0.2)
ωB97M-V/def2-ha-TZVP	−21.3	−20.5	−22.7	−21.7	−132.8	−131.9	−109.3	−107.7
	(−0.8)	(+0.0)	(−0.9)	(+0.1)	(−2.6)	(−1.7)	(−1.3)	(+0.3)
ωB97M-V/def2-TZVPD	−21.3	−20.5	−22.6	−21.7	−132.8	−130.2	−109.3	−107.7
	(−0.8)	(+0.0)	(−0.9)	(+0.1)	(−2.6)	(−0.1)	(−1.3)	(+0.3)
ωB97M-V/def2-QZVP	−20.7	−20.4	−22.1	−21.6	−131.8	−129.9	−108.6	−108.3
	(−0.2)	(+0.2)	(−0.3)	(+0.2)	(−1.6)	(+0.2)	(−0.6)	(−0.3)
ωB97M-V/def2-ma-QZVP	−20.7	−20.4	−22.0	−21.5	−130.6	−129.8	−108.1	−107.8
	(−0.2)	(+0.2)	(−0.2)	(+0.2)	(−0.4)	(+0.4)	(−0.2)	(+0.2)
M06-2X/DZVP	−16.7	−14.5	−18.9	−16.0	−123.7	−108.6	−104.5	−93.0
	(−1.5)	(+0.7)	(−3.0)	(−0.2)	(−16.4)	(−1.2)	(−12.7)	(−1.3)
M06-2X/6-31+G*	−18.3	−16.4	−19.8	−16.9	−122.6	−108.6	−105.1	−92.2
	(−3.1)	(−1.2)	(−4.0)	(−1.1)	(−15.2)	(−1.3)	(−13.3)	(−0.5)
M06-2X/def2-SVP	−16.4	−11.2	−18.7	−12.0	−142.1	−99.6	−114.3	−89.2
	(−1.1)	(+4.0)	(−2.8)	(+3.8)	(−34.7)	(+7.7)	(−22.5)	(+2.6)
M06-2X/def2-ma-SVP	−18.7	−12.4	−20.6	−13.4	−125.5	−102.6	−109.0	−91.6
	(−3.5)	(+2.8)	(−4.7)	(+2.4)	(−18.1)	(+4.7)	(−17.2)	(+0.2)
M06-2X/def2-ha-SVP	−23.3	−14.5	−23.6	−15.1	−134.3	−106.3	−111.5	−91.8
	(−8.1)	(+0.7)	(−7.7)	(+0.8)	(−26.9)	(+1.1)	(−19.8)	(−0.1)
M06-2X/def2-SVPD	−26.3	−14.6	−26.9	−15.4	−144.8	−107.3	−116.5	−92.3
	(−11.1)	(+0.6)	(−11.0)	(+0.4)	(−37.5)	(+0.0)	(−24.7)	(−0.5)
M06-2X/def2-TZVP	−15.2	−14.8	−16.4	−15.6	−113.1	−106.5	−95.1	−90.4
	(−0.0)	(+0.4)	(−0.5)	(+0.2)	(−5.7)	(+0.9)	(−3.4)	(+1.3)
M06-2X/def2-ma-TZVP	−15.3	−14.8	−16.4	−15.7	−110.7	−106.3	−94.3	−90.5
	(−0.1)	(+0.4)	(−0.6)	(+0.2)	(−3.3)	(+1.1)	(−2.6)	(+1.2)
M06-2X/def2-ha-TZVP	−14.9	−14.8	−15.7	−15.6	−109.1	−106.4	−92.7	−90.3
	(+0.3)	(+0.4)	(+0.1)	(+0.2)	(−1.8)	(+1.0)	(−1.0)	(+1.5)
M06-2X/def2-TZVPD	−15.0	−14.8	−15.7	−15.7	−109.0	−106.5	−92.5	−90.3
	(+0.2)	(+0.4)	(+0.1)	(+0.2)	(−1.7)	(+0.9)	(−0.8)	(+1.5)
M06-2X/def2-QZVP	−15.1	−15.2	−16.0	−15.6	−108.4	−106.7	−92.5	−91.0
	(+0.1)	(+0.0)	(−0.2)	(+0.2)	(−1.1)	(+0.6)	(−0.8)	(+0.7)
M06-2X/def2-ma-QZVP	−15.2	−15.3	−15.7	−16.0	−107.9	−106.8	−92.3	−91.2
	(+0.1)	(−0.1)	(+0.1)	(−0.1)	(−0.5)	(+0.5)	(−0.6)	(+0.6)

[a]Errors relative to the DFT/CBS limit are given in parentheses. [b]ΔE_{int} without CP correction. [c]ΔE_{int} with CP correction.

(1HSG:indinavir and 1O44:RU85052), CP-corrected 6-31+G* results are even more impressive, affording results that are only about 1 kcal/mol from the DFT/CBS limit. Nevertheless, CP correction is essential and exceeds 10 kcal/mol. This is true even for the DZVP basis set that afforded good accuracy (without CP correction) for small-molecule calculations reported in ref 90, including the S66 data set. Those results likely benefit from fortuitous error cancelation

made possible by the relatively small magnitude of the CP corrections for the S66 dimers.

Notably, the addition of diffuse functions to def2-SVP significantly *increases* the convergence errors for both functionals. This behavior was also observed for coronene dimer (Section 3.2), where it was explained by analogy to the basis-set convergence properties of the dispersion-bound subset of the S66 complexes. Consistent with those examples, here we find that CP correction significantly insulates the results
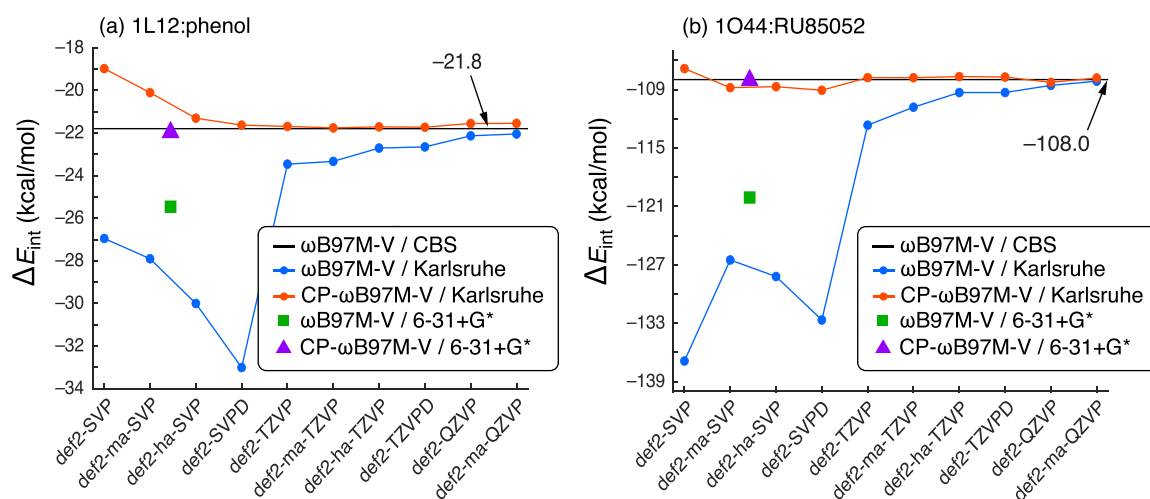
Figure 6. Interaction energies for (a) 1LI2:phenol and (b) 1O44:RU85052 computed using ωB97M-V in various basis sets. Connected points illustrate convergence of ΔE_{int} in Karlsruhe basis sets, both with and without CP correction. For comparison, values of ΔE_{int} in the 6-31+G* basis set are also shown. The ωB97M-V/CBS limit, obtained by averaging def2-ma-QZVP interaction energies with and without CP correction, is indicated by the horizontal line, and its numerical value is also indicated.

against this effect. With CP correction, the def2-ha-SVP basis set (in which hydrogen atoms do not have diffuse functions) performs very similar to fully augmented def2-SVPD, although the results in the minimally augmented def2-ma-SVP basis set are noticeably different. This suggests def2-ha-SVP (with CP correction) as a low-cost choice for large systems, in which convergence errors have been reduced to $\lesssim 1$ kcal/mol with respect to the DFT/CBS limit. This basis is defined for most of the periodic table, whereas 6-31+G* is only defined for elements up to argon.

A visual perspective on convergence to the ωB97M-V/CBS limit for these systems is presented in Figure 6 for two of the ligand−protein complexes. (The corresponding plots for the other two complexes can be found in Figure S11 and for M06-2X in Figure S12.) In small basis sets, the CP-corrected value of ΔE_{int} is much closer to the CBS limit as compared to the uncorrected interaction energy, which is systematically over-bound. CP correction reduces the interaction energy in a systematic way as the CBS limit is approached, and although its effect is diminished in larger basis sets, the CP-corrected interaction energies hew more closely to the DFT/CBS limit than do the uncorrected values, with the former being close to convergence already for def2-SVPD. This parallels what we have seen for other systems and suggests that double-ζ interaction energies are completely unreliable for large systems unless CP correction is employed. Unlike the results in smaller systems, however, the triple-ζ results cannot be considered to be converged in these examples (even in the presence of diffuse functions) unless CP correction is applied, although the fully augmented def2-TZVPD basis set comes close. Timing data for the 1HSG:indinavir complex (Figure S13) demonstrate that CP-ωB97M-V/def2-SVPD is 6× less expensive than the def2-ma-QZVP basis set that is otherwise needed to reach the CBS limit.

Finally, in an effort to gauge the absolute accuracy of these ligand−protein interaction energies, we take advantage of a recent CCSD(T)/CBS benchmark for the 1HSG:indinavir system.[110] Table 7 shows how ωB97M-V performs against this benchmark, using various basis sets. CP-corrected values of ΔE_{int} lie within 10 kcal/mol of the benchmark (representing 8% error), even for basis sets as small as def2-SVP. This is

Table 7. Errors in ΔE_{int} for 1HSG:Indinavir Using ωB97M-V

	error vs CCSD(T)/CBS[a]			
	absolute (kcal/mol)		relative (%)	
basis set	no CP[b]	CP[c]	no CP[b]	CP[c]
def2-SVP	54.4	4.1	44.8	3.4
def2-ma-SVP	30.0	6.1	24.7	5.0
def2-ha-SVP	36.4	8.6	29.9	7.1
def2-SVPD	46.0	9.5	37.9	7.9
def2-TZVP	17.4	9.1	14.3	7.5
def2-ma-TZVP	12.9	8.6	10.6	7.1
def2-ha-TZVP	11.3	10.4	9.3	8.6
def2-TZVPD	11.3	8.4	9.3	7.2
def2-QZVP	10.3	8.4	8.5	6.9
def2-ma-QZVP	9.1	9.3	7.5	6.8

[a]Versus the benchmark value $\Delta E_{int} = -121.50$ kcal/mol from ref 110. [b]Error without CP correction. [c]Error with CP correction.

certainly a much larger absolute error as compared to those observed for S66 dimers, but the percentage errors are typically 7−8% when CP correction is used. (The ωB97M-V/def2-SVP value is slightly more accurate.) Errors increase substantially in the absence of CP correction, up to 54 kcal/mol (45% error) in the smallest basis set tested.

4. CONCLUSIONS

Far from being an "overcorrection", the Boys−Bernardi CP procedure is an essential aspect of noncovalent quantum chemistry that significantly reduces the finite-basis error in intermolecular interaction energies and thus showcases the true behavior of the functional in question, free of error cancellation. CP-corrected interaction energies computed using double-ζ basis sets are comparable in quality to uncorrected triple-ζ results and therefore close to the DFT/CBS limit, even for combinatorially optimized functionals such as ωB97X-V and ωB97M-V that are considered to be more demanding in their basis-set dependence.[23,24] Perhaps surprisingly, these conclusions also hold for double-hybrid functionals that contain MP2 correlation and therefore exhibit

slower basis-set convergence (and also increased cost). Even for lower-cost GGA and hybrid functionals, the use of CP correction with double-ζ basis sets results in computational savings, as this approach is 3−4× faster than computing CP-corrected triple-ζ interaction energies and somewhat faster (1.2−1.6×) than computing triple-ζ interaction energies with no CP correction.

For double-ζ basis sets, the use of a supramolecular basis (*i.e.*, CP correction) significantly reduces the importance of diffuse basis functions. For example, we find that CP-corrected def2-SVP and def2-SVPD values of ΔE_{int} are similar except in the case of double-hybrid functionals, whereas in the absence of CP correction the diffuse functions significantly alter ΔE_{int}. For the coronene dimer, we find that double-ζ basis sets predict an incorrect intermolecular separation unless CP correction is applied, whereas the CP-corrected double-ζ geometry agrees with the results obtained using quadruple-ζ basis sets. These are important observations, given that diffuse functions are expensive and numerically problematic for large systems.

In systems with hundreds of atoms, including layered nanocomposite materials and protein−ligand complexes, the magnitude of the CP correction is much larger than it is in small benchmark dimers, which reflects the size extensivity of BSSE. In the absence of CP correction, double-ζ results for systems of this size may be 30−40 kcal/mol (or more) from the DFT/CBS limit and should be considered unreliable, which is notable given the prevalence of such calculations in the literature. On the other hand, CP-corrected interaction energies converge toward the CBS limit in a systematic manner, and they do so much more rapidly than the uncorrected results. CP correction again proves to be cost-effective despite the added expense of computing the monomer energies in a supramolecular basis set. We recommend CP-DFT/def2-ha-SVP as a robust and relatively low-cost approach to compute ΔE_{int} in large systems.

We find no realistic use cases where CP correction is disadvantageous. (CP correction can increase the convergence error when double-hybrid functionals are used with small Pople basis sets or with def2-SVP, but this reflects the inadequacy of those basis sets for MP2 correlation rather than a limitation of the CP procedure.) CP correction increases the cost of the calculation by a modest factor, typically ≈2.5×, even for large protein−ligand complexes with hundreds of atoms, and is therefore feasible whenever the supramolecular calculation itself is feasible. CP correction can be used with double-ζ basis sets to obtain reliable results in large systems where triple-ζ calculations are intractable. We suggest that its use should be considered mandatory for supramolecular calculations of ΔE_{int} in large systems, especially when only double-ζ calculations are computationally feasible, in order to ensure that the results do not benefit from error cancellation that may not be robust. Even with triple-ζ basis sets, CP correction can compensate for the absence of diffuse functions and furthermore provides a means to estimate how far the finite-basis value of ΔE_{int} may be from the DFT/CBS limit.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c00883.

    Additional data (PDF)

    Coordinates for the large systems (XYZ)

■ AUTHOR INFORMATION

Corresponding Author

    John M. Herbert − *Department of Chemistry and Biochemistry and Biophysics Graduate Program, The Ohio State University, Columbus, Ohio 43210, United States;* ⓞ orcid.org/0000-0002-1663-2278; Email: herbert@chemistry.ohio-state.edu

Authors

    Montgomery Gray − *Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States*

    Paige E. Bowling − *Department of Chemistry and Biochemistry and Biophysics Graduate Program, The Ohio State University, Columbus, Ohio 43210, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.2c00883

Notes

The authors declare the following competing financial interest(s): J.M.H. serves on the board of directors of Q-Chem Inc.

■ REFERENCES

(1) Grimme, S. Density functional theory with London dispersion corrections. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 211−228.

(2) Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-corrected mean-field electronic structure methods. *Chem. Rev.* **2016**, *116*, 5105−5154.

(3) Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: An overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017**, *115*, 2315−2372.

(4) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate *ab initio* parameterization of density functional dispersion correction (DFT-D) for the 94 elements H−Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(5) Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A generally applicable atomic-charge dependent London dispersion correction. *J. Chem. Phys.* **2019**, *150*, 154122.

(6) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.

(7) Price, A. J. A.; Bryenton, K. R.; Johnson, E. R. Requirements for an accurate dispersion-corrected density functional. *J. Chem. Phys.* **2021**, *154*, 230902.

(8) Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* **2004**, *25*, 1463−1473.

(9) Zhao, Y.; Truhlar, D. G. Design of density functionals that are broadly accurate for thermochemistry, thermochemical kinetics, and nonbonded interactions. *J. Phys. Chem. A* **2005**, *109*, 5656−5667.

(10) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, non-

covalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215−241; Erratum: *Theor. Chem. Acc.* **2008**, *119*, 525.

(11) Zhao, Y.; Truhlar, D. G. Density functionals with broad applicability in chemistry. *Acc. Chem. Res.* **2008**, *41*, 157−167.

(12) Mardirossian, N.; Head-Gordon, M. How accurate are the Minnesota density functionals for noncovalent interactions, isomerization energies, thermochemistry, and barrier heights involving molecules composed of main-group elements? *J. Chem. Theory Comput.* **2016**, *12*, 4303−4325.

(13) Goerigk, L. Testing London-dispersion effects with the latest Minnesota density functionals: Problems and possible solutions. *J. Phys. Chem. Lett.* **2015**, *6*, 3891−3896.

(14) Shahbaz, M.; Szalewicz, K. Do semilocal density-functional approximations recover dispersion energies at small intermonomer separations? *Phys. Rev. Lett.* **2018**, *121*, 113402.

(15) Shahbaz, M.; Szalewicz, K. Evaluation of methods for obtaining dispersion energies used in density functional calculations of intermolecular interactions. *Theor. Chem. Acc.* **2019**, *138*, 25.

(16) Langreth, D. C.; Dion, M.; Rydberg, H.; Schröder, E.; Hyldgaard, P.; Lundqvist, B. I. Van der Waals density functional theory with applications. *Int. J. Quantum Chem.* **2005**, *101*, 599−610.

(17) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. Van der Waals density functional for general geometries. *Phys. Rev. Lett.* **2004**, *92*, 246401.

(18) Langreth, D. C.; et al. A density functional for sparse matter. *J. Phys.: Condens. Matter* **2009**, *21*, 084203.

(19) Lee, K.; Murray, É. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. Higher-accuracy van der Waals density functional. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *82*, 081101.

(20) Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals density functional theory made simple. *Phys. Rev. Lett.* **2009**, *103*, 063004.

(21) Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals density functional: The simpler the better. *J. Chem. Phys.* **2010**, *133*, 244103.

(22) Vydrov, O. A.; Voorhis, T. Nonlocal van der Waals density functionals based on local response models. In *Fundamentals of Time-Dependent Density Functional Theory*; Marques, M. A. L., Maitra, N. T., Nogueira, F. M. S., Gross, E. K. U., Rubio, A., Eds.; Springer-Verlag: Berlin, 2012; Vol. *837*, Chapter 23, pp 443−456.

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

(24) Mardirossian, N.; Head-Gordon, M. ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016**, *144*, 214110.

(25) Mardirossian, N.; Head-Gordon, M. Mapping the genome of meta-generalized gradient approximation density functionals: The search for B97M-V. *J. Chem. Phys.* **2015**, *142*, 074111.

(26) Becke, A. D. Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals. *J. Chem. Phys.* **1997**, *107*, 8554−8560.

(27) Becke, A. D.; Johnson, E. R. Exchange-hole dipole moment and the dispersion interaction. *J. Chem. Phys.* **2005**, *122*, 154104.

(28) Becke, A. D.; Johnson, E. R. A density-functional model of the dispersion interaction. *J. Chem. Phys.* **2005**, *123*, 154101.

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

(30) Becke, A. D.; Johnson, E. R. Exchange-hole dipole moment and the dispersion interaction revisited. *J. Chem. Phys.* **2007**, *127*, 154108.

(31) Otero-de-la-Roza, A.; Johnson, E. R. Van der Waals interactions in solids using the exchange-hole dipole moment model. *J. Chem. Phys.* **2012**, *136*, 174109.

(32) Otero-de-la-Roza, A.; Johnson, E. R. Many-body dispersion interactions from the exchange-hole dipole moment model. *J. Chem. Phys.* **2013**, *138*, 054103.

(33) Tkatchenko, A.; Scheffler, M. Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. *Phys. Rev. Lett.* **2009**, *102*, 073005.

(34) Ambrosetti, A.; Reilly, A. M.; DiStasio, R. A., Jr.; Tkatchenko, A. Long-range correlation energy calculated from coupled atomic response functions. *J. Chem. Phys.* **2014**, *140*, 18A508.

(35) Hermann, J.; DiStasio, R. A., Jr.; Tkatchenko, A. First-principles models for van der Waals interactions in molecules and materials: Concepts, theory, and applications. *Chem. Rev.* **2017**, *117*, 4714−4758.

(36) Goerigk, L.; Grimme, S. Double-hybrid density functionals. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 576−600.

(37) Martin, J. M. L.; Santra, G. Empirical double-hybrid density functional theory: A 'third way' in between WFT and DFT. *Isr. J. Chem.* **2020**, *60*, 787−804.

(38) Lao, K. U.; Herbert, J. M. Atomic orbital implementation of extended symmetry-adapted perturbation theory (XSAPT) and benchmark calculations for large supramolecular complexes. *J. Chem. Theory Comput.* **2018**, *14*, 2955−2978.

(39) Carter-Fenk, K.; Lao, K. U.; Herbert, J. M. Predicting and understanding non-covalent interactions using novel forms of symmetry-adapted perturbation theory. *Acc. Chem. Res.* **2021**, *54*, 3679−3690.

(40) Nguyen, B.; Chen, G. P.; Agee, M. M.; Burow, A. M.; Tang, M.; Furche, F. Divergence of many-body perturbation theory for noncovalent interactions of large molecules. *J. Chem. Theory Comput.* **2020**, *16*, 2258−2273.

(41) Nguyen, B. D.; Hernandez, D. J.; Flores, E. V.; Furche, F. Dispersion size-consistency. *Electron. Struct.* **2022**, *4*, 014003.

(42) van Lenthe, J. H.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B. Weakly bonded systems. *Adv. Chem. Phys.* **1987**, *69*, 521−566.

(43) Chalasinski, G.; Gutowski, M. Weak interactions between small systems. Models for studying the nature of intermolecular forces and challenging problems for ab initio calculations. *Chem. Rev.* **1988**, *88*, 943−962.

(44) van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. State of the art in counterpoise theory. *Chem. Rev.* **1994**, *94*, 1873−1885.

(45) Tschumper, G. S. Reliable electronic structure computations for weak noncovalent interactions in clusters. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Cundari, T. R., Eds.; Wiley-VCH: Hoboken, 2009; Vol. *26*, Chapter 2, pp 39−90.

(46) Riley, K. E.; Pitoňák, M.; Jurečka, P.; Hobza, P. Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories. *Chem. Rev.* **2010**, *110*, 5023−5063.

(47) van Mourik, T.; Wilson, A. K.; Peterson, K. A.; Woon, D. E.; Dunnning, T. H., Jr. The effect of basis set superposition error (BSSE) on the convergence of molecular properties calculated with the correlation consistent basis sets. *Adv. Quantum Chem.* **1998**, *31*, 105−135.

(48) Varandas, A. J. C. Can extrapolation to the basis set limit be an alternative to the counterpoise correction? A study on the helium dimer. *Theor. Chem. Acc.* **2008**, *119*, 511−521.

(49) Richard, R. M.; Lao, K. U.; Herbert, J. M. Achieving the CCSD(T) basis-set limit in sizable molecular clusters: Counterpoise corrections for the many-body expansion. *J. Phys. Chem. Lett.* **2013**, *4*, 2674−2680.

(50) Burns, L. A.; Marshall, M. S.; Sherrill, C. D. Comparing counterpoise-corrected, uncorrected, and averaged binding energies for benchmarking noncovalent interactions. *J. Chem. Theory Comput.* **2014**, *10*, 49−57.

(51) Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separated total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553−566.

(52) Alvarez-Idaboy, J. R.; Galano, A. Counterpoise corrected interaction energies are not systematically better than uncorrected

ones: Comparison with CCSD(T) CBS extrapolated values. *Theor. Chem. Acc.* **2010**, *126*, 75−85.

(53) Sheng, X.-W.; Mentel, L.; Gritsenko, O. V.; Baerends, E. J. Counterpoise correction is not useful for short and van der waals distances but may be useful at long range. *J. Comput. Chem.* **2011**, *32*, 2896−2901.

(54) Mentel, Ł. M.; Baerends, E. J. Can the counterpoise correction for basis set superposition effect be justified? *J. Chem. Theory Comput.* **2014**, *10*, 252−267.

(55) Brauer, B.; Kesharwani, M. K.; Martin, J. M. L. Some observations on counterpoise corrections for explicitly correlated calculations on noncovalent interactions. *J. Chem. Theory Comput.* **2014**, *10*, 3791−3799.

(56) Brauer, B.; Kesharwani, M. K.; Kozuch, S.; Martin, J. M. L. The S66x8 benchmark for noncovalent interactions revisited: Explicitly correlated *ab initio* methods and density functional theory. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20905−20925.

(57) Jensen, F. Method calibration or data fitting? *J. Chem. Theory Comput.* **2018**, *14*, 4651−4661.

(58) Smith, D. G. A.; Burns, L. A.; Patkowski, K.; Sherrill, C. D. Revised damping parameters for the D3 dispersion correction to density functional theory. *J. Phys. Chem. Lett.* **2016**, *7*, 2197−2203.

(59) Boese, A. D.; Martin, J. M. L.; Handy, N. C. The role of the basis set: Assessing density functional theory. *J. Chem. Phys.* **2003**, *119*, 3005−3014.

(60) Mardirossian, N.; Head-Gordon, M. Characterizing and understanding the remarkably slow basis set convergence of several Minnesota density functionals for intermolecular interaction energies. *J. Chem. Theory Comput.* **2013**, *9*, 4453−4461.

(61) Wheeler, S. E.; Houk, K. N. Integration grid errors for meta-GGA-predicted reaction energies: Origin of grid errors for the M06 suite of functionals. *J. Chem. Theory Comput.* **2010**, *6*, 395−404.

(62) Dasgupta, S.; Herbert, J. M. Standard grids for high-precision integration of modern density functionals: SG-2 and SG-3. *J. Comput. Chem.* **2017**, *38*, 869−882.

(63) Galano, A.; Alvarez−Idaboy, J. R. A new approach to counterpoise correction to BSSE. *J. Comput. Chem.* **2006**, *27*, 1203−1210.

(64) Johansson, A.; Kollman, P.; Rothenberg, S. An application of the functional Boys-Bernardi counterpoise method to molecular potential energy surfaces. *Theor. Chem. Acc.* **1973**, *29*, 167−172.

(65) Kochanski, E. Intermolecular interactions: Use of small basis sets in ab initio calculations. *J. Mol. Struct.: THEOCHEM* **1984**, *107*, 49−57.

(66) Jönsson, B.; Nelander, B. Infrared spectrum and structure of an ammonia−carbon dioxide complex. A comparison between ab initio MO calculations and matrix isolation results. *Chem. Phys.* **1977**, *25*, 263−269.

(67) Gutowski, M.; van Duijneveldt, F. B.; Chałasiński, G.; Piela, L. Proper correction for the basis set superposition error in SCF calculations of intermolecular interactions. *Mol. Phys.* **1987**, *61*, 233−247.

(68) Gutowski, M.; Van Lenthe, J. H.; Verbeek, J.; Van Duijneveldt, F. B.; Chałasinski, G. The basis set superposition error in correlated electronic structure calculations. *Chem. Phys. Lett.* **1986**, *124*, 370−375.

(69) Gutowski, M.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H.; van Duijneveldt, F. B. Accuracy of the Boys and Bernardi function counterpoise method. *J. Chem. Phys.* **1993**, *98*, 4728−4737.

(70) Urban, M.; Hobza, P. Weak intermolecular interaction. IV. An application of the function counterpoise method to the calculation of the interaction energy and its components. *Theor. Chem. Acc.* **1975**, *36*, 215−220.

(71) Jeziorski, B.; van Hemert, M. Variation−perturbation treatment of the hydrogen bond between water molecules. *Mol. Phys.* **1976**, *31*, 713−729.

(72) Jeziorski, B.; Szalewicz, K.; Chalasinski, G. Symmetry forcing and convergence properties of perturbation expansions for molecular interactions energies. *Int. J. Quantum Chem.* **1978**, *14*, 271−287.

(73) Szalewicz, K.; Patkowski, K.; Jeziorski, B.Intermolecular interactions via perturbation theory: From diatoms to biomolecules. In *Intermolecular Forces and Clusters II*; Wales, D. J., Ed.; Springer-Verlag: Berlin, 2005; Vol. *116*.

(74) Hobza, P.; Zahradnik, R. Intermolecuar interactions between medium-sized systems. Nonempirical and empirical calculations of interaction energies. Successes and failures. *Chem. Rev.* **1988**, *88*, 871−897.

(75) Schwenke, D. W.; Truhlar, D. G. Systematic study of basis set superposition errors in the calculated interaction energy of two HF molecules. *J. Chem. Phys.* **1985**, *82*, 2418−2426; Erratum: *J. Chem. Phys.* **1986**, *84*, 4113; *J. Chem. Phys.* **1987**, *86*, 3760.

(76) Sordo, J. A.; Sordo, T. L.; Fernández, G. M.; Gomperts, R.; Chin, S.; Clementi, E. A systematic study on the basis set superposition error in the calculation of interaction energies of systems of biological interest. *J. Chem. Phys.* **1989**, *90*, 6361−6370.

(77) Szczęśniak, M.; Scheiner, S. Correction of the basis set superposition error in SCF and MP2 interaction energies. The water dimer. *J. Chem. Phys.* **1986**, *84*, 6328−6335.

(78) Mayer, I. The "chemical Hamiltonian approach" and the SCF method. *J. Mol. Struct.: THEOCHEM* **1988**, *165*, 255−272.

(79) Mayer, I. The chemical Hamiltonian approach for treating the BSSE problem of intermolecular interactions. *Int. J. Quantum Chem.* **1998**, *70*, 41−63.

(80) Mayer, I.; Valiron, P. Second order Møller−Plesset perturbation theory without basis set superposition error. *J. Chem. Phys.* **1998**, *109*, 3360−3373.

(81) Dunning, T. H., Jr. A road map for the calculation of molecular binding energies. *J. Phys. Chem. A* **2000**, *104*, 9062−9080.

(82) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796−6806.

(83) Woon, D. E.; Dunning, T. H., Jr. Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties. *J. Chem. Phys.* **1994**, *100*, 2975−2988.

(84) Weigend, F.; Ahlrichs, R. 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.* **2005**, *7*, 3297−3305.

(85) Rappoport, D.; Furche, F. Property-optimized Gaussian basis sets for molecular response calculations. *J. Chem. Phys.* **2010**, *133*, 134105.

(86) Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.* **1984**, *80*, 3265−3269.

(87) Herbert, J. M.; Paul, S. K. Interaction energy analysis of monovalent inorganic anions in bulk water versus air/water interface. *Molecules* **2021**, *26*, 6719.

(88) Gray, M.; Herbert, J. M. Comprehensive basis-set testing of extended symmetry-adapted perturbation theory and assessment of mixed-basis combinations to reduce cost. *J. Chem. Theory Comput.* **2022**, *18*, 2308−2330.

(89) Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. Optimization of Gaussian-type basis sets for local spin density functional calculations. Part I. Boron through neon, optimization technique and validation. *Can. J. Chem.* **1992**, *70*, 560−571.

(90) Hostaš, J.; Řezáč, J. Accurate DFT-D3 calculations in a small basis set. *J. Chem. Theory Comput.* **2017**, *13*, 3575−3585.

(91) Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibson, T. D.; Windus, T. L. New basis set exchange: An open, up-to-date resource for the molecular sciences community. *J. Chem. Inf. Model.* **2019**, *59*, 4814−4820.

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

(93) Brémond, E.; Sancho-García, J. C.; Pérez-Jiménez, A. J.; Adamo, C. Communication: Double-hybrid functionals from adiabatic-connection: The QIDH model. *J. Chem. Phys.* **2014**, *141*, 031101.

(94) Karton, A.; Tarnopolsky, A.; Lamère, J.-F.; Schatz, G. C.; Martin, J. M. L. Highly accurate first-principles benchmark data sets for the parameterization and validation of density functional and other approximate methods. Derivation of a robust, generally applicable, double-hybrid functional for thermochemistry and thermochemical kinetics. *J. Phys. Chem. A* **2008**, *112*, 12868−12886.

(95) Chai, J.-D.; Head-Gordon, M. Long-range corrected double-hybrid density functionals. *J. Chem. Phys.* **2009**, *131*, 174105.

(96) Řezác, J.; Riley, K. E.; Hobza, P. S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427−2438; Erratum: *J. Chem. Theory Comput.* **2014**, *10*, 1359−1360.

(97) Helgaker, T.; Jørgensen, P.; Olsen, J.*Molecular Electronic-Structure Theory*; Wiley: New York, 2000.

(98) Epifanovsky, E.; et al. Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *J. Chem. Phys.* **2021**, *155*, 084801.

(99) Gill, P. M. W.; Johnson, B. G.; Pople, J. A. A standard grid for density-functional calculations. *Chem. Phys. Lett.* **1993**, *209*, 506−512.

(100) Ballesteros, F.; Dunivan, S.; Lao, K. U. Coupled cluster benchmarks of large noncovalent complexes: The L7 dataset as well as DNA-ellipticine and buckycatcher-fullerene. *J. Chem. Phys.* **2021**, *154*, 154104.

(101) Ashraf, A.; Carter-Fenk, K.; Herbert, J. M.; Farooqi, B. A.; Farooq, U.; Ayub, K. Interaction of graphene quantum dots with oligothiophene: A comprehensive theoretical study. *J. Phys. Chem. C* **2019**, *123*, 29556−29570.

(102) Ashraf, A.; Herbert, J. M.; Muhammad, S.; Farooqi, B. A.; Farooq, U.; Salman, M.; Ayub, K. Theoretical approach to evaluate the gas-sensing performance of graphene nanoribbon/oligothiophene composites. *ACS Omega* **2022**, *7*, 2260−2274.

(103) Morton, A.; Matthews, B. W. Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: Linkage of dynamics and structural plasticity. *Biochemistry* **1995**, *34*, 8576−8588.

(104) Wei, B. Q.; Baase, W. A.; Weaver, L. H.; Matthews, B. W.; Shoichet, B. K. A model binding site for testing scoring functions in molecular docking. *J. Mol. Biol.* **2002**, *322*, 339−355.

(105) Chen, Z.; Li, Y.; Chen, E.; Hall, D. L.; Darke, P. L.; Culberson, C.; Shafer, J. A.; Kuo, L. C. Crystal structure at 1.9-Å resolution of human immunodeficiency virus (HIV) II protease complexed with L-735,524, an orally bioavailable inhibitor of the HIV proteases. *J. Biol. Chem.* **1994**, *269*, 26344−26348.

(106) Lange, G.; Lesuisse, D.; Deprez, P.; Schoot, B.; Loenze, P.; Bénard, D.; Marquette, J.; Broto, P.; Sarubbi, E.; Mandine, E. Requirements for specific binding of low affinity inhibitor fragments to the SH2 domain of ^{pp60}Src are identical to those for high affinity binding of full length inhibitors. *J. Med. Chem.* **2003**, *46*, 5184−5195.

(107) Ucisik, M. N.; Dashti, D. S.; Faver, J. C.; Merz, K. M., Jr. Pairwise additivity of energy components in protein-ligand binding: The HIV II protease-indinavir case. *J. Chem. Phys.* **2011**, *135*, 085101.

(108) Lao, K. U.; Herbert, J. M. Accurate and efficient quantum chemistry calculations of noncovalent interactions in many-body systems: The XSAPT family of methods. *J. Phys. Chem. A* **2015**, *119*, 235−252.

(109) Liu, K.-Y.; Carter-Fenk, K.; Herbert, J. M. Self-consistent charge embedding at very low cost, with application to symmetry-adapted perturbation theory. *J. Chem. Phys.* **2019**, *151*, 031102.

(110) Villot, C.; Ballesteros, F.; Wang, D.; Lao, K. U. Coupled cluster benchmarking of large noncovalent complexes in L7 and S12L as well as the C60 dimer, DNA−ellipticine, and HIV−indinavir. *J. Phys. Chem. A* **2022**, *126*, 4326−4341.

(111) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(112) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: Automating p*K* prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulation. *Nucleic Acids Res.* **2012**, *40*, W537−W541.

(113) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—an accurate and broadly parameterized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652−1671.

(114) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.

(115) Ohio Supercomputer Center. http://osc.edu/ark:/19495/f5s1ph73 (accessed 2022-10-05).