

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Development and Assessment of Electronic Structure Approaches for Non-Covalent Interactions

Permalink

<https://escholarship.org/uc/item/7mc3k8tm>

Author

Witte, Jonathon Kendall

Publication Date

2017

Peer reviewed|Thesis/dissertation

**Development and Assessment of Electronic Structure Approaches for
Non-Covalent Interactions**

by

JONATHON KENDALL WITTE

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Martin P. Head-Gordon, Co-chair

Professor Jeffrey B. Neaton, Co-chair

Professor Eric W. Neuscamman

Professor Mark D. Asta

Spring 2017

**Development and Assessment of Electronic Structure Approaches for
Non-Covalent Interactions**

Copyright 2017
by
JONATHON KENDALL WITTE

Abstract

Development and Assessment of Electronic Structure Approaches for Non-Covalent Interactions

by

JONATHON KENDALL WITTE

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Martin P. Head-Gordon, Co-chair

Professor Jeffrey B. Neaton, Co-chair

This thesis is primarily concerned with the development and assessment of electronic structure approaches for intermolecular interactions. Various aspects of existing approaches – most notably the choices of method, grid, and basis set – are examined with respect to performance in novel ways, and new semi-empirical corrections intended to rectify deficiencies in standard methods are introduced.

Contents

| | |
|---|------------|
| Contents | i |
| List of Figures | iii |
| List of Tables | x |
| 1 Introduction | 1 |
| 1.1 Quantum Mechanics | 1 |
| 1.2 The Born-Oppenheimer Approximation | 2 |
| 1.3 Hartree-Fock Theory | 3 |
| 1.4 Wavefunction-Based Approaches | 5 |
| 1.5 Density Functional Theory | 6 |
| 1.6 Outline | 8 |
| 2 Case Study: The CO₂-Benzene Complex | 11 |
| 2.1 Introduction | 11 |
| 2.2 Computational Methods | 12 |
| 2.3 Results and Discussion | 14 |
| 2.4 Conclusion | 21 |
| 3 A New Paradigm for Method Assessment: Geometries | 23 |
| 3.1 Introduction | 23 |
| 3.2 Computational Methods | 25 |
| 3.3 Results and Discussion | 30 |
| 3.4 Conclusion | 39 |
| 4 Basis Sets for Intermolecular Interactions | 43 |
| 4.1 Introduction | 43 |
| 4.2 Computational Methods | 45 |
| 4.3 Results and Discussion | 46 |
| 4.4 Discussion and Conclusion | 62 |
| 5 Empirical Dispersion: DFT-D3 | 65 |

| | | |
|----------|--|------------|
| 5.1 | Introduction | 65 |
| 5.2 | Theory | 66 |
| 5.3 | Computational Details | 70 |
| 5.4 | Results and Discussion | 71 |
| 5.5 | Discussion and Conclusion | 78 |
| 6 | Modeling Basis Set Superposition Error | 80 |
| 6.1 | Introduction | 80 |
| 6.2 | Theory and Methods | 81 |
| 6.3 | Results and Discussion | 87 |
| 6.4 | Discussion and Conclusions | 93 |
| A | Geometry Study | 95 |
| A.1 | Damping Function | 95 |
| A.2 | MP2 Basis | 96 |
| A.3 | Intramolecular Metrics | 97 |
| B | Basis Set Study | 98 |
| B.1 | Numbering Convention: Key for S22 | 98 |
| B.2 | Justification of CP pc-4 as CBS limit | 99 |
| B.3 | Semilocal Exchange-Correlation Grid | 100 |
| B.4 | Nonlocal Correlation Grid | 101 |
| B.5 | SPW92 RMSE versus CBS and S22B | 102 |
| B.6 | B3LYP RMSE versus CBS and S22B | 103 |
| B.7 | B97M-V RMSE versus CBS and S22B | 104 |
| B.8 | Nonlocal Correlation Basis Set Insensitivity | 105 |
| B.9 | Basis Set Convergence Figures | 106 |
| C | DFT-D3 Study | 109 |
| C.1 | Additional DFT-D3(op) Parameterizations | 109 |
| C.2 | Rare Gas Dimers (RG10) | 110 |
| D | DFT-C Study | 112 |
| D.1 | Additional Figures and Tables | 112 |
| D.2 | gCP | 114 |
| D.3 | DFT-C | 116 |
| | References | 131 |

List of Figures

| | | |
|-----|--|----|
| 2.1 | Calculated equilibrium structure of the benzene-CO ₂ system, determined at the DB-MP2 level of theory with the dual-basis pairing for Dunning’s augmented correlation-consistent polarized valence triple-zeta basis. The complex possesses C _s symmetry. | 14 |
| 2.2 | Binding energy curves for the benzene-CO ₂ complex obtained with various wavefunction methods. The region denoted ΔCCSD(T)/CBS is bounded by CP- and non-CP-corrected ΔCCSD(T) curves, and represents our best guess at the true binding curve. None of the other methods employ the CP correction. The inset is a close-up of the region around the equilibrium distance (note different scale), and the curves connecting the points are merely smooth fits to the data. The equilibrium intermolecular separations for each method (listed to the right of the legend as <i>r_e</i>) were determined by cubic spline interpolation of the data. The reported <i>r_e</i> for ΔCCSD(T)/CBS is the average of the ΔCCSD(T)/CBS noCP and CP equilibrium intermolecular separations. For ease of analysis, the legend has been ordered (with the exception of the benchmark ΔCCSD(T)/CBS region) to mirror the ordering of the binding curves near the equilibrium separation. | 15 |
| 2.3 | Binding energy curves for the benzene-CO ₂ complex obtained with various standard, hybrid, and long-range corrected hybrid GGA functionals in the aug-cc-pVTZ basis with a (75,302) integration grid. For further details, see Figure 2.2. | 17 |
| 2.4 | Binding energy curves for the benzene-CO ₂ complex obtained with various standard, hybrid, and range-separated hybrid meta-GGA functionals in the aug-cc-pVTZ basis with a (75,302) integration grid. For further details, see Figure 2.2. | 18 |
| 2.5 | Binding energy curves for the benzene-CO ₂ complex obtained with various DFT-D2 methods in the aug-cc-pVTZ basis with a (75,302) integration grid. For further details, see Figure 2.2. | 19 |
| 2.6 | Binding energy curves for the benzene-CO ₂ complex obtained with various methods employing environmentally-dependent C ₆ corrections in the aug-cc-pVTZ basis with a (75,302) integration grid. For further details, see Figure 2.2. | 20 |
| 2.7 | Binding energy curves for the benzene-CO ₂ complex obtained with various non-local methods in the aug-cc-pVTZ basis with a (75,302) integration grid. For further details, see Figure 2.2. | 21 |

| | | |
|-----|---|----|
| 3.1 | Structures of the systems in the M12 dataset. A light blue sphere corresponds to the center of mass of a particular molecule. The type of interaction is indicated by the color of the text: hydrogen-bonded systems are blue, dispersion-bound systems are green, and systems with mixed interactions are red. Green double-headed arrows indicate the relevant intermolecular axis for each system. In brackets, abbreviations for the complexes are introduced. | 27 |
| 3.2 | Structures of the systems in the A21x12 dataset. The type of interaction is indicated by the color of the text: hydrogen-bonded systems are blue, dispersion-bound systems are green, and systems with mixed interactions are red. Green double-headed arrows connect the centers of masses of the two molecules in each system, and hence indicate the relevant intermolecular axis for each system. . . | 29 |
| 3.3 | Structure of the coronene dimer. Light blue spheres correspond to the centers of masses of the coronene monomers. A green double-headed arrow indicates the relevant intermolecular axis. | 30 |
| 3.4 | Average signed and unsigned errors in closest-contact distances in geometries of complexes from the A21 Dataset. The methods are listed in order of ascending unsigned error. All calculations were performed in the aug-cc-pVTZ basis with tight convergence criteria. Calculations involving density functionals utilized a (99,590) grid. | 32 |
| 3.5 | Selection of binding energy curves for the cyclopentane dimer (CyCy in the M12 dataset). Equilibrium separations associated with each method are denoted by vertical dashed lines, and were determined via interpolation with a cubic spline. All calculations were performed in the aug-cc-pVTZ basis. Calculations involving density functionals utilized a (99,590) grid. | 34 |
| 3.6 | Average signed (ASE) and unsigned (AUE) errors (in units of Å) in interpolated equilibrium intermolecular separations in geometries of complexes in various subsets of the M12 dataset. Maximum error for each method (MAX) is given in last column (in units of Å). Within each subset, methods are sorted in order of ascending AUE. Signed values are colored as such: positive errors are blue, negative errors are red, and the tint of the color correlates with the magnitude of the error. Equilibria correspond to the interpolated (cubic spline) minima of the binding energy curves. All calculations were performed in the aug-cc-pVTZ basis. Calculations involving density functionals utilized a (99,590) grid. | 35 |
| 3.7 | Errors in interpolated equilibrium intermolecular separations (in units of Å) in geometries of complexes of the M12 dataset. The last column, the average unsigned error (AUE) across all systems (in units of Å), represents the metric by which the methods are sorted. Signed values are colored as such: positive errors are blue, negative errors are red, and the tint of the color correlates with the magnitude of the error. Equilibria correspond to the interpolated (cubic spline) minima of the binding energy curves. All calculations were performed in the aug-cc-pVTZ basis. Calculations involving density functionals utilized a (99,590) grid. For the abbreviations, see Figure 3.1. | 37 |

| | | |
|-----|--|----|
| 3.8 | Errors in interpolated equilibrium intermolecular separations (in units of Å) and binding energies (in units of kcal mol ⁻¹) for the parallel-displaced coronene dimer. The methods are listed in order of ascending error in intermolecular separation. Equilibria correspond to the interpolated (cubic spline) minima of the binding energy curves. All calculations were performed in the aug-cc-pVDZ basis. Calculations involving density functionals utilized a (99,590) grid. With the exception of attenuated MP2, all methods incorporate a correction for BSSE[91]. Errors in separation are relative to $R = 3.458$ Å reported by Janowski et al.[137], corresponding to QCISD(T)/h-aug-cc-pVDZ, and errors in binding energy are relative to $E_{\text{bind}} = -23.45$ kcal mol ⁻¹ , which corresponds to QCISD(T)/CBS as reported by Mardirossian and Head-Gordon[19]. | 39 |
| 4.1 | Structures of the systems in the S22 dataset. The systems are classified by interaction type as per the original work[181]. | 45 |
| 4.2 | Uncorrected (noCP) and counterpoise-corrected (CP) root mean square errors (RMSE) in SPW92 binding energies across the S22 set of molecules. Errors are expressed relative to SPW92/CBS. Blue corresponds to noCP, gold to CP. The bars are a visual representation of the actual RMSEs, which are tabulated for each basis set on the left side of the figure. Basis sets are listed in order of increasing number of basis functions. Note the difference between noCP (top) and CP (bottom) axes. | 48 |
| 4.3 | Uncorrected (noCP) and counterpoise-corrected (CP) root mean square errors (RMSE) in B3LYP binding energies across the S22 set of molecules. For further details, see Figure 4.2. | 50 |
| 4.4 | Uncorrected (noCP) and counterpoise-corrected (CP) root mean square errors (RMSE) in B97M-V binding energies across the S22 set of molecules. For further details, see Figure 4.2. | 51 |
| 4.5 | Convergence of uncorrected (noCP) SPW92 normalized mean binding energies across the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets. Each binding energy was normalized to the corresponding SPW92/CBS value before averaging. The number of basis functions for each basis set was determined by averaging the number of basis functions for each system within each basis set across all systems in S22. | 52 |
| 4.6 | Convergence of counterpoise-corrected (CP) SPW92 normalized mean binding energies across the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets. For further details, see Figure 4.5. | 52 |
| 4.7 | Convergence of uncorrected (noCP) and counterpoise-corrected (CP) B97M-V normalized mean binding energies across the S22 set of molecules for the Truhlar, Dunning, Jensen, and Karlsruhe sequences of basis sets. Each binding energy was normalized to the corresponding B97M-V/CBS value before averaging. | 53 |

| | | |
|------|--|----|
| 4.8 | Convergence of uncorrected (noCP) SPW92 normalized mean binding energies across subsets of the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets. Within each subset, each binding energy was normalized to the CBS limit (counterpoise-corrected pc-4) before averaging. The three subsets – hydrogen-bonded, dispersion-bound, and mixed interactions – are the same as those in Figure 4.1. The number of basis functions for each basis set was determined by averaging the number of basis functions for each system within each basis set across all systems in the relevant subset of S22. | 55 |
| 4.9 | Convergence of counterpoise-corrected (CP) SPW92 normalized mean binding energies across subsets of the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets. For further details, see Figure 4.8. | 56 |
| 4.10 | Relationship between basis set superposition error (BSSE) and number of interacting atoms for several method/basis set combinations. The number of interacting atoms is defined as the number of unique atoms in each system for which the distance to another atom on a different fragment is less than 1.1 times the sum of the van der Waals radii of the atoms. | 58 |
| 4.11 | Mean basis set superposition error (BSSE) per interaction across the S22 set of molecules for SPW92, B3LYP, and B97M-V in a variety of basis sets. The number of interactions per system were determined as in Figure 4.10. Each column is color-coded with a gradient from dark red (highest BSSE) to white (median BSSE) to dark blue (lowest BSSE). | 59 |
| 4.12 | Relationship between remaining basis set incompleteness error (rBSIE) and number of interacting atoms for several method/basis set combinations. The number of interactions per system were determined as in Figure 4.10. | 60 |
| 4.13 | Root mean square errors (RMSE) in uncorrected (noCP) and counterpoise-corrected (CP) binding energies for each method within the pc- n sequence of basis sets. Errors here correspond to errors relative to CCSD(T)/CBS results[186]. The number of basis functions is determined as in Figure 4.5. | 61 |
| 5.1 | Visualization of the effects of varying parameters α_1 , α_2 , and β from eq. (5.7). The leftmost plots are generated by varying α_1 with fixed α_2 and β ; the center plots are generated by varying α_2 with fixed α_1 and β ; and the rightmost plots are generated by varying β with fixed α_1 and α_2 . When not being varied, the parameters are fixed to $\alpha_1 = 0.5$, $\alpha_2 = 5$, and $\beta = 14$, with $r_{0,ij} = 2 \text{ \AA}$. The top plot in each section shows the damping function for the sixth-order term, $f_{\text{damp},6}^{\text{D3(op)}}(r_{ij})$ from eq. (5.7). The bottom plot in each section shows the sixth-order contribution to the two-body dispersion energy, $E^{(2)}$, which has been normalized to span the range $[0,1]$ | 69 |

| | | |
|-----|--|----|
| 5.2 | Root-mean-square errors across various datasets for all combinations of functionals and dispersion corrections examined. Abbreviations of the dataset names are used; the full names are, in order, NCEDTest, HSG, S22, NCECTest, Shields38, 3B-69-TRIM, IETest, Pentane14, YMPJ519, S66x8 BL, and S66x8 BE. The Product column corresponds to a simple product across the aggregate dataset RMSEs (times 100). All RMSEs are in units of kcal/mol, with the exception of the equilibrium binding lengths (BL), which are in units of angstroms. Each column within every functional block is color-coded for ease of reading, with darker cells corresponding to lower RMSEs; note, however, that the color-gradient is not uniform, but rather skewed to emphasize significant differences among the top fits for each functional. Additionally, the lowest RMSE achieved by any functional-dispersion combination on each aggregate dataset is indicated in bold. | 74 |
| 5.3 | Root-mean-square errors (RMSEs) across various datatypes (see Table 5.3) for the top-performing D3-corrected hybrid GGA functional examined in this work, as well as three other state-of-the-art hybrid GGA functionals. The left plot – (a) – contains datasets pertaining to standard non-covalent interactions and isomerization energies, whereas the right plot – (b) – encompasses other data, such as thermochemistry, which is beyond the realm of a D3 correction. Within each plot, the methods are ordered from best performance across the constituent datasets at the top, to worst performance at the bottom. Tables of RMSEs are provided below the bar graphs to facilitate quantitative comparison. | 77 |
| 5.4 | Root-mean-square errors (RMSEs) across various datatypes (see Table 5.3) for the top-performing D3-corrected pure meta-GGA functional examined in this work, as well as three other state-of-the-art pure meta-GGA functionals. For further details, refer to the caption of Figure 5.3. | 77 |
| 5.5 | Root-mean-square errors (RMSEs) across various datatypes (see Table 5.3) for the top-performing D3-corrected hybrid meta-GGA functional examined in this work, as well as three other state-of-the-art pure hybrid meta-GGA functionals. For further details, refer to the caption of Figure 5.3. | 78 |
| 6.1 | Visualization of how adding a third atom C impacts the contribution of basis functions centered at B to the BSSE on atom A , as per $h_{AB^*}(\{A, B, C\})$. When atom C is sufficiently far away from A and B (lighter areas), the model reduces to a pairwise approach. In this example, C and B are assumed to have the same number of virtual orbitals, and A and B are located at (-1.5 a.u., -1.5 a.u.) and (1.5 a.u., 1.5 a.u.), respectively. | 84 |
| 6.2 | Dependence of actual and predicted neon atom SPW92/def2-SVPD BSSEs on distance to argon ghost functions. | 86 |

| | | |
|-----|---|-----|
| 6.3 | Root-mean-square errors of B97M-V with (CP) and without (noCP) the Boys-Bernardi correction for BSSE in two small basis sets relative to B97M-V in the def2-QZVPPD basis, near the basis set limit. SVP and SVPD correspond to def2-SVP and def2-SVPD, respectively. Methods in the chart are ordered from lowest overall RMSE at the top, to highest overall RMSE at the bottom. A table of values is provided below the chart to facilitate quantitative comparison. . . . | 88 |
| 6.4 | Normalized root-mean-square errors (NRMSEs) of gCP and DFT-C predicted BSSEs versus Boys and Bernardi BSSEs at the LSDA level of DFT in the def2-SVPD basis. The datatypes NCED, NCD, and NCEC are defined in Table 6.1. The normalized root-mean-square error is obtained by dividing the RMSE by the mean reference value in the dataset, as described in the text. Direct use of LSDA/def2-SVPD without any correction would result in 100% NRMSE. . . . | 89 |
| 6.5 | Normalized root-mean-square errors (NRMSEs) of gCP and DFT-C predicted BSSEs versus Boys and Bernardi BSSEs for three GGA density functionals in the def2-SVPD basis. For further details, see Figure 6.4. | 90 |
| 6.6 | Normalized root-mean-square errors (NRMSEs) of gCP and DFT-C predicted BSSEs versus Boys and Bernardi BSSEs for three meta-GGA density functionals in the def2-SVPD basis. For further details, see Figure 6.4. | 91 |
| 6.7 | Root-mean-square errors of B97M-V versus high-level reference values at five levels of theory: uncorrected in the def2-SVPD basis (noCP); counterpoise-corrected in def2-SVPD (CP); with the geometrical counterpoise correction in def2-SVPD (gCP); with the correction introduced in this work in the def2-SVPD basis (DFT-C); and near the complete-basis set limit (CBS), in def2-QZVPPD. Methods in the chart are ordered from lowest overall RMSE at the top, to highest overall RMSE at the bottom. A table of values is provided below the chart to facilitate quantitative comparison. | 92 |
| 6.8 | Root-mean-square errors in kcal/mol of several pure meta-GGA density functionals relative to high-level reference values. B97M-V-C corresponds to B97M-V with the DFT-C correction. Results for the additional density functionals are taken from a previous study[44]. SVPD corresponds to def2-SVPD, and QZVPPD corresponds to def2-QZVPPD. Each datatype category is color-coded, with the darkest color corresponding to the lowest RMSE within that category. | 93 |
| B.1 | Convergence of uncorrected (noCP) SPW92 normalized mean binding energies across the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets. | 106 |
| B.2 | Convergence of counterpoise-corrected (CP) SPW92 normalized mean binding energies across the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets. | 106 |
| B.3 | Convergence of uncorrected (noCP) B3LYP normalized mean binding energies across the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets. | 107 |

| | | |
|-----|--|-----|
| B.4 | Convergence of counterpoise-corrected (CP) B3LYP normalized mean binding energies across the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets. | 107 |
| B.5 | Convergence of uncorrected (noCP) B97M-V normalized mean binding energies across the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets. | 108 |
| B.6 | Convergence of counterpoise-corrected (CP) B97M-V normalized mean binding energies across the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets. | 108 |
| D.1 | Root-mean-square errors of B97M-V/def2-SVPD versus B97M-V/def2-QZVPPD with no correction (noCP), the standard counterpoise correction (CP), the geometrical counterpoise correction (gCP), and the treatment introduced here (DFT-C). All RMSEs are in units of kcal/mol. Methods are ordered from lowest overall RMSE at the top, to highest overall RMSE at the bottom. | 112 |
| D.2 | Root-mean-square errors of B97M-V/def2-SVPD versus “exact” reference values with no correction (noCP), the standard counterpoise correction (CP), the geometrical counterpoise correction (gCP), and the treatment introduced here (DFT-C). All RMSEs are in units of kcal/mol. Note for SW49 and most of IE, the standard counterpoise correction is not possible, and so for these datasets the noCP and CP methods are identical. | 113 |

List of Tables

| | | |
|-----|---|----|
| 3.1 | Justification of methods. We have considered the quality of the reference data with regard to basis set size on both small (top left) and moderately-sized (bottom left) systems, the impact of the specific form of function used for interpolation (top right), and the difference between interpolation and constrained optimization (bottom right). UAD, SAD, and MAX denote respectively the unsigned, signed, and maximum average difference in interpolated equilibrium separation across the indicated dataset, in units of Å. r and BE denote the equilibrium distance (Å) and binding energy (kcal mol ⁻¹), respectively. | 28 |
| 3.2 | Average overall root-mean-square deviations (RMSD) and weighted root-mean-square deviations (wRMSD) in geometries of complexes in various subsets of the A21 Dataset. Weights were determined from the relative binding energies of each complex. Within each subset, the methods are listed in order of ascending RMSD or wRMSD. All calculations were performed in the aug-cc-pVTZ basis with tight convergence criteria. Calculations involving density functionals utilized a (99,590) grid. | 31 |
| 3.3 | Pearson’s correlation coefficient, R , for percent errors in interpolated equilibrium energy and intermolecular separation for the M12 dataset. Methods are divided into two sets – wavefunction-based (WFT) and density functional theory (DFT) – and listed in order of descending correlation coefficient within each set. For details about the calculations, see Figure 3.7. | 40 |
| 4.1 | Complete basis set (CBS) binding energies for each system in S22 at various levels of theory. For the density functional approximations, counterpoise-corrected pc-4 constitutes the CBS limit. Benchmark CCSD(T)/CBS results from Marshall et al.[186] are provided for comparison. | 47 |
| 4.2 | Uncorrected (noCP) and counterpoise-corrected (CP) root mean square errors (RMSE) in B97M-V binding energies across S22 for variants of Dunning-style basis sets and comparably-sized Karlsruhe alternatives. | 54 |
| 5.1 | Summary of empirical dispersion corrections considered in this study. | 69 |

| | | |
|-----|--|-----|
| 5.2 | Summary of density functionals. The names in the first column are standard, with two exceptions: B97 corresponds to Grimme’s pure functional B97-D[30], which has had the dispersion tail stripped away; and B97h corresponds to the original hybrid functional B97, as parameterized by Becke[17]. The third column details the existing parameterized DFT-D-style dispersion corrections we consider in this study, and the fourth column lists the references for the method. | 70 |
| 5.3 | Summary of datasets that comprise the training and test sets. For more details, see Ref. 24. | 72 |
| 5.4 | Optimized values of -D3(op) fit parameters for each density functional. | 72 |
| 6.1 | Summary of datatypes. For more details, see Ref. 24. | 87 |
| A.1 | Impact of damping function. All results pertain to A21 set, aug-cc-pVTZ basis, (99,590) grid, tight convergence criteria (details in article). Average RMS and signed errors in the closest point of contact for each optimized structure are shown. The method marked CHG combines -D2 C_6 coefficients with a -D3 style damping function. The method marked Fermi combines -D3 C_6 coefficients with the -D2 damping function (a Fermi-type damping function). | 95 |
| A.2 | Effect of finiteness of basis on MP2 A21 calculations. Since the reference structures are Δ CCSD(T)/CBS, it is possible that our MP2 results suffer from considerable BSSE. We thus show here average contact signed errors across the A21 dataset in the aug-cc-pVQZ basis (in which we expect our results should be well-converged with respect to basis size). | 96 |
| A.3 | Mean bond length (MBL) and bond angle (MBA) RMS errors across the A21 set. Methods are sorted in order of ascending MBL. | 97 |
| B.1 | Numbering convention of S22 used to abbreviate systems in the remainder of this Appendix to make tables more compact. | 98 |
| B.2 | Differences in SPW92 binding energies calculated in the aug-pc-4 basis set with (CP) and without (noCP) counterpoise correction, as compared to CP pc-4 values. Two systems could not be converged in the aug-pc-4 basis set (issues with linear dependencies); these are listed as "n/a." | 99 |
| B.3 | Impact of semilocal exchange-correlation grid. All results are for SPW92 using the cc-pVDZ basis set. Each column (SG-1 and the following columns) refers to a different Lebedev grid. The "incremental change in BSSE" refers to the change in BSSE upon switching to that grid from the next smallest grid, e.g. the numbers under (99,590) correspond to the change in BSSE accompanying the change from the (75,302) to the (99,590) grid. | 100 |

| | | |
|-----|--|-----|
| B.4 | Impact of nonlocal correlation grid. All results are for B97M-V and the cc-pVDZ basis set with a (99,590) semilocal grid. Both uncorrected (noCP) and counterpoise-corrected (CP) binding energies (BEs) are provided for each grid (SG-1 and (99,590)) to demonstrate that the fact that BSSE is converged is not a result of some fortuitous error cancellation. | 101 |
| B.5 | Root mean square errors (RMSE) of uncorrected (noCP) and counterpoise-corrected (CP) SPW92 versus two different types of references: complete-basis SPW92 (CBS) and complete-basis CCSD(T) (S22B[186]). | 102 |
| B.6 | Root mean square errors (RMSE) of uncorrected (noCP) and counterpoise-corrected (CP) B3LYP versus two different types of references: complete-basis B3LYP (CBS) and complete-basis CCSD(T) (S22B[186]). | 103 |
| B.7 | Root mean square errors (RMSE) of uncorrected (noCP) and counterpoise-corrected (CP) B97M-V versus two different types of references: complete-basis B97M-V (CBS) and complete-basis CCSD(T) (S22B[186]). | 104 |
| B.8 | Nonlocal (VV10) contributions to B97M-V binding energies across S22 with a small basis (uncorrected def2-SVPD) and the complete basis (counterpoise-corrected pc4). This does not arise solely from fortuitous error cancellation between the dimers and their constituent monomers; absolute energies are converged to nearly this same level of precision. | 105 |
| C.1 | Optimized values of DFT-D3(op) fit parameters for four additional density functionals. | 109 |
| C.2 | Root mean square errors (RMSEs) for meta-GGAs across the RG10 set of rare gas dimers. BL and BE refer to interpolated equilibrium binding length and energy, respectively, and the rightmost column (RG10) is an RMSE across all 569 datapoints. RMSEs are in kcal/mol unless otherwise noted. | 110 |
| C.3 | Root mean square errors (RMSEs) for GGAs across the RG10 set of rare gas dimers. BL and BE refer to interpolated equilibrium binding length and energy, respectively, and the rightmost column (RG10) is an RMSE across all 569 datapoints. RMSEs are in kcal/mol unless otherwise noted. | 111 |
| D.1 | RMSEs (kcal/mol) for gCP and gCPmod predicted BSSEs in B3LYP/def2-SVPD versus Boys-Bernardi BSSEs. | 115 |

Acknowledgments

First and foremost, I would like to thank my advisers, Martin Head-Gordon and Jeffrey Neaton, for all their support throughout the past five years, and for their unwavering enthusiasm. I could not have asked for better mentors, and I will definitely miss our (approximately) weekly meetings.

I would also like to thank Leslie Silvers, Meg Holm, and Jessie Woodcock for all their help navigating the administrative hurdles.

I would like to thank all the undergraduates, graduate students, and postdocs with whom I have had temporal overlap in the Head-Gordon and Neaton groups, particularly Narbe Mardirossian, the only person I know who spends more time at Yosemite than I do, and without whom none of my methodological projects would have been possible.

I would like to thank all the other incredible people I have met in the Berkeley chemistry and physics departments.

Lastly, I would like to thank Julie Beckmann, Kendall Witte, Ryan Beckmann, and Amy Patnode, for all their love and support.

Chapter 1

Introduction

The driving force for the work described within this thesis is the desire to better understand a novel class of materials: nanoporous materials, particularly metal-organic frameworks (MOFs)[1]. MOFs are three-dimensional, nanoporous solids built from metal-containing nodes connected by organic ligands; a cross-section of a MOF looks something like a honeycomb, only at the molecular – rather than macro – level. As a result of this molecular honeycomb structure, MOFs possess an incredibly high ratio of surface area to volume. This, in turn, makes them particularly well suited to a variety of applications ranging from catalysis[2] to gas storage[3]. Moreover, as a result of their modularity, MOFs are highly tunable, and different structures can give rise to significantly different functionality.

There is thus a sort of combinatorial problem with MOFs; there are far too many possible combinations of metal nodes and linkers to feasibly synthesize and characterize. Ideally, we would like to be able to accurately model these materials so synthetic efforts can be directed towards only the most promising candidates. Unfortunately, this is no easy endeavor. The relevant functional properties of MOFs arise from molecular-scale interactions. Such is the domain of quantum mechanics.

1.1 Quantum Mechanics

In the early twentieth century, quantum mechanics arose out of an effort to understand various observed phenomena – most notably black-body radiation and the photoelectric effect – that were not accounted for by conventional physical models. One of the central equations in quantum mechanics is the time-independent Schrödinger equation, a wave equation for particles,

$$\hat{H} |\Psi\rangle = E |\Psi\rangle, \quad (1.1)$$

where \hat{H} is the Hamiltonian operator, E is the energy of the system, and $|\Psi\rangle$ is the wavefunction (WF), or probability amplitude, from which all observable properties may be derived.

Due to the correspondence principle, equation 1.1 and its time-dependent analogue may be used to describe any physical system.

The molecular Hamiltonian may be expressed, in atomic units, as a sum of five terms,

$$\begin{aligned} \hat{H} = & - \sum_i^n \frac{1}{2} \nabla_i^2 - \sum_A^M \frac{1}{2m_A} \nabla_A^2 - \sum_i^n \sum_A^M \frac{Z_A}{|\vec{r}_i - \vec{R}_A|} \\ & + \sum_i^n \sum_{j>i}^n \frac{1}{|\vec{r}_i - \vec{r}_j|} + \sum_A^M \sum_{B>A}^M \frac{Z_A Z_B}{|\vec{R}_A - \vec{R}_B|}, \end{aligned} \quad (1.2)$$

where i and \vec{r}_i correspond to the indices and coordinates of the n electrons, and A , \vec{r}_A , Z_A , and m_A represent the indices, coordinates, atomic numbers, and masses of the M nuclei. The first term in equation 1.2, \hat{T}_e , is the kinetic energy of the electrons; the second term, \hat{T}_n , is the kinetic energy of the nuclei; the third term, \hat{V}_{en} , corresponds to the potential energy of attraction between the electrons and nuclei; the fourth term, \hat{V}_{ee} , is the potential energy of repulsion between the electrons; and the fifth term, \hat{V}_{nn} , is the potential energy of repulsion between the nuclei. Both the Hamiltonian and the wavefunction are functions of the n electronic and M nuclear coordinates; accordingly, solving the time-independent Schrödinger equation requires solving a $3(n + M)$ -dimensional partial differential equation. This is untenable for most systems of interest; hence, we must typically resort to a sequence of approximations.

1.2 The Born-Oppenheimer Approximation

The first approximation commonly employed to simplify the Schrödinger equation is the Born-Oppenheimer approximation, wherein the nuclear and electronic degrees of freedom are separated. Since electrons are nearly 2000 times less massive than the nuclei they orbit, their motion is “instantaneous” relative to that of the nuclei: the electrons can be roughly viewed as moving in a field of fixed nuclei. This approximation yields the electronic Hamiltonian, the central operator in electronic structure theory,

$$\begin{aligned} \hat{H}_e = & \hat{T}_e + \hat{V}_{en} + \hat{V}_{ee} \\ = & - \sum_i^n \frac{1}{2} \nabla_i^2 - \sum_i^n \sum_A^M \frac{Z_A}{|\vec{r}_i - \vec{R}_A|} + \sum_i^n \sum_{j>i}^n \frac{1}{|\vec{r}_i - \vec{r}_j|}. \end{aligned} \quad (1.3)$$

Since \hat{H}_e treats the nuclear coordinates as parameters, rather than explicit variables, the electronic Schrödinger equation is an eigenvalue problem in $3n$ dimensions,

$$\hat{H}_e |\Psi_e\rangle = E_e |\Psi_e\rangle, \quad (1.4)$$

where $|\Psi_e\rangle$ and E_e are the electronic wavefunction and energy, respectively. Unfortunately, this equation is still intractable for all but the most simple systems; accordingly, additional approximations are needed.

1.3 Hartree-Fock Theory

Perhaps the simplest, most naive, approximation for the n -electron wavefunction is a product of n single-particle wavefunctions dependent on general coordinates (σ) encompassing both spin and position. This simple product of single-particle wavefunctions, or spin orbitals $|\chi_i(\sigma_i)\rangle$, constitutes a Hartree product wavefunction,

$$\begin{aligned} |\Psi_e(\sigma_1, \sigma_2, \dots, \sigma_n)\rangle &\approx |\Phi_H(\sigma_1, \sigma_2, \dots, \sigma_n)\rangle \\ &= \prod_i^n |\chi_i(\sigma_i)\rangle. \end{aligned} \quad (1.5)$$

A Hartree product, however, is not antisymmetric with respect to change of electrons, and thus is not a valid fermionic wavefunction. The simplest fermionic ansatz – and hence simplest valid approximation to the electronic wavefunction – is not merely a simple product of spin orbitals, but rather a determinant, commonly referred to as a Slater determinant:

$$\begin{aligned} |\Psi_e(\sigma_1, \sigma_2, \dots, \sigma_n)\rangle &\approx |\Phi_S(\sigma_1, \sigma_2, \dots, \sigma_n)\rangle \\ &= \frac{1}{\sqrt{n!}} \begin{vmatrix} |\chi_1(\sigma_1)\rangle & |\chi_2(\sigma_1)\rangle & \dots & |\chi_n(\sigma_1)\rangle \\ |\chi_1(\sigma_2)\rangle & |\chi_2(\sigma_2)\rangle & \dots & |\chi_n(\sigma_2)\rangle \\ \vdots & \vdots & \ddots & \vdots \\ |\chi_1(\sigma_n)\rangle & |\chi_2(\sigma_n)\rangle & \dots & |\chi_n(\sigma_n)\rangle \end{vmatrix}. \end{aligned} \quad (1.6)$$

Combining this single Slater determinant approximation to the wavefunction with the time-independent electronic Schrödinger equation yields the Hartree-Fock approach (HF), in which electron-electron interactions are treated in a mean-field manner: each electron interacts with the field generated by the other $n - 1$ electrons. The HF energy is given by the expectation value of the electronic Hamiltonian in the state $|\Phi_S\rangle$, i.e.

$$\begin{aligned} E_{\text{HF}} &= \langle \Phi_S | \hat{H} | \Phi_S \rangle \\ &= \sum_i^n \langle \chi_i | \hat{h} | \chi_i \rangle + \frac{1}{2} \sum_i^n \sum_j^n \langle \chi_i \chi_j | \chi_j \chi_i \rangle, \end{aligned} \quad (1.7)$$

where the one-electron core Hamiltonian is given by

$$\hat{h}(\vec{r}_i) = -\frac{1}{2}\nabla_i^2 - \sum_A^M \frac{Z_A}{|\vec{R}_A - \vec{r}_i|}. \quad (1.8)$$

This is a variational method; the best spin orbitals are those that minimize the energy, subject to the constraint that the various spin-orbitals remain orthonormal. These best spin orbitals satisfy the Hartree-Fock equations,

$$\hat{f} |\chi_i\rangle = \epsilon |\chi_i\rangle, \quad (1.9)$$

with the Fock operator, \hat{f} , being given by

$$\hat{f}(\vec{r}_1) = \hat{h}(\vec{r}_1) + \sum_i^{n/2} \int \left[\phi_i^*(\vec{r}_2) \frac{1}{|\vec{r}_2 - \vec{r}_1|} (2 - \hat{P}_{12}) \phi_i(\vec{r}_2) \right] d\vec{r}_2. \quad (1.10)$$

Note that spin-dependence has been integrated out to yield spatial orbitals ϕ_i , which are functions of position (\vec{r}). Note also that \hat{P}_{12} is a permutation operator.

In order to solve the Hartree-Fock equations, we expand the spatial orbitals in a basis of K atomic orbitals (AOs) – which are in turn constructed from different sequences of real-space functions, the subject of Chapter 4 – $|\omega_\mu\rangle$:

$$|\phi_i\rangle = \sum_{\mu}^K c_{\mu i} |\omega_\mu\rangle. \quad (1.11)$$

We then left-project with $\langle\omega_\nu|$ to convert the integro-differential equation – equation 1.9 – into a generalized eigenvalue equation,

$$\sum_{\mu}^K F_{\nu\mu} c_{\mu i} = \sum_{\mu}^K S_{\nu\mu} c_{\mu i} \epsilon_i, \quad (1.12)$$

or, in matrix form,

$$\mathbf{FC} = \mathbf{SC}\epsilon. \quad (1.13)$$

Although the overlap matrix \mathbf{S} is straightforward to compute, the elements of Fock matrix involve functions of all the electrons, and hence this equation – the Roothaan-Hall equation – must be solved iteratively, using what is known as the self-consistent field (SCF) procedure.

The Hartree-Fock method recovers more than 99% of the electronic energy. To the uninitiated, that sounds like HF should be more than sufficient; unfortunately, this is not the case. That remaining 1% – termed the “correlation energy” – is where all the magic happens; it describes how electronic motion is correlated, how electrons “talk” to each other, and it is vital for understanding most of the interesting questions in the field of chemical physics. Most modern endeavors in electronic structure are centered around coming up with good, cheap approximations to the correlation energy. These approximations can be broadly grouped into two distinct classes: wavefunction-based and density-based.

1.4 Wavefunction-Based Approaches

The first class of correlation approximations are based upon wavefunctions. For most standard methods, the approximate wavefunction is obtained by incorporating excitations from the Hartree-Fock reference.

Møller-Plesset Perturbation Theory

The simplest wavefunction-based approach to correlation is Møller-Plesset perturbation theory, which can be derived by performing Rayleigh-Schrödinger many-body perturbation theory[4] on the HF wavefunction. The assumption implicit in this approach is that the interaction between the electrons – the perturbation to the non-interacting HF reference – is small. The most widely used variant of perturbation theory in the electronic structure community is second-order Møller-Plesset perturbation theory (MP2)[5], which is the leading correction in the perturbation expansion. The MP2 correction to the electronic energy is given by

$$E^{(2)} = -\frac{1}{4} \sum_{ij}^{\text{occ}} \sum_{ab}^{\text{virt}} \frac{|\langle ij || ab \rangle|^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j}, \quad (1.14)$$

where i, j and a, b are canonical occupied and virtual orbitals, respectively, and ϵ_i is the Fock operator – see equations 1.9 and 1.10 – corresponding to orbital i . Unlike the HF energy, the MP2 energy is nonvariational, and the method struggles when the mean-field solution is a poor reference, such as in systems with small HOMO-LUMO gaps.

Coupled-Cluster Theory

A more sophisticated family of wavefunction-based methods for recovering correlation energy is the coupled-cluster (CC) approach[6, 7], wherein a wavefunction is constructed by operating on the HF reference – the Slater determinant of spin orbitals that minimize the HF energy – with an exponential excitation operator:

$$|\Phi_{\text{CC}}\rangle = e^{\hat{T}} |\Phi_{\text{HF}}\rangle. \quad (1.15)$$

The cluster operator, \hat{T} , is given by

$$\begin{aligned} \hat{T} &= \sum_{\kappa} \hat{T}_{\kappa} \\ &= \sum_{\kappa} \left[\frac{1}{(\kappa!)^2} \sum_{i,b} t_{i_1, i_2, \dots, i_{\kappa}}^{b_1, b_2, \dots, b_{\kappa}} \hat{a}_{b_{\kappa}}^{\dagger} \dots \hat{a}_{b_1}^{\dagger} \hat{a}_{i_1} \dots \hat{a}_{i_{\kappa}} \right], \end{aligned} \quad (1.16)$$

where t are excitation amplitudes, and \hat{a}^\dagger and \hat{a} are the standard creation and annihilation operators from second quantization. In practice, the cluster operator is truncated at a certain level of excitation (e.g. CCSD corresponds to single and double excitations, $\kappa = 1, 2$), and the excitation amplitudes (and hence also the energy) are obtained by left-projecting with the series of ground- and excited-state determinants and solving the system of equations thus formed.

Coupled-cluster is one of the most accurate electronic structure approaches, and it has a clear hierarchy: incorporating higher-order excitations leads to more accurate (i.e. closer to the exact electronic energy) results[8]. Unfortunately, this accuracy comes at significant cost: full inclusion of up to triple and quadruple excitations, CCSDT and CCSDTQ, is prohibitively expensive for most systems of interest. Thus, in practice, triple and quadruple excitations are often treated perturbatively[9]; this constitutes the the CCSD(T) and CCSDT(Q) approaches. CCSD(T) – coupled-cluster with full single and double excitations, and perturbative triple excitations – is colloquially known as the "gold-standard" in quantum chemistry.

1.5 Density Functional Theory

The second class of electronic structure approximations capable of handling correlation is density functional theory (DFT).

Hohenberg-Kohn DFT

The foundations of DFT are the Hohenberg-Kohn (HK) theorems[10].

HK Theorem 1 *There is a one-to-one mapping between the electron density and the external potential.*

HK Theorem 2 *Density functional theory is variational: the exact ground-state density minimizes the total energy.*

This idea is incredibly powerful, in principle. In WF-based methods, such as HF, MP2, and CC, the wavefunction – a function that lives in a vast Hilbert space of dimension $3n$ – is the central entity from which the energy is derived. In DFT, on the other hand, everything comes from the electron density $\rho(\vec{r})$, a function that lives in Euclidean 3-space. In practice, however, DFT is not straightforward. The electronic Hamiltonian contains three terms: electronic kinetic energy \hat{T}_e , electron-nuclear potential energy \hat{V}_{en} , and electron-electron potential energy \hat{V}_{ee} . The first Hohenberg-Kohn theorem tells us each of these terms can be written as a functional of the electron density. The electron-nuclear attraction energy can be straightforwardly expressed as

$$V_{en}[\rho(\vec{r})] = \int d\vec{r} \hat{v}(\vec{r})\rho(\vec{r}), \quad (1.17)$$

but the exact forms of the kinetic and electron-electron potential energies are not known. The contribution of the kinetic energy to the total electronic energy is massive, but the only system for which we have a good description of the kinetic energy is the uniform electron gas, for which the Thomas-Fermi model is exact. Thus, this sort of orbital-free formalism of DFT has failed to gain much traction within the electronic structure community, as its utility is limited to systems with nearly uniform electron densities.

Kohn-Sham DFT

Fortunately, there is an alternative formalism for DFT: Kohn-Sham DFT, or KS-DFT[11]. KS-DFT recasts the problem through the introduction of a Slater determinant of spin orbitals (Kohn-Sham orbitals) that corresponds to a fictitious, non-interacting system of electrons, yet yields the same ground-state density as the exact electronic wave function. Like orbital-free DFT, KS-DFT is an exact theory. Its main advantages lie in the fact that the problem has been transformed from a constrained search over densities to a constrained search over orbitals, such that the same SCF machinery employed by HF can be used; and the largest unknown in DFT, the electronic kinetic energy, can be much more accurately reproduced. The kinetic energy for the fictitious, non-interacting system can be straightforwardly expressed as

$$T_s[\rho(r)] = \sum_i^n \left\langle \phi_i \left| -\frac{1}{2} \nabla^2 \right| \phi_i \right\rangle, \quad (1.18)$$

where $|\phi_i\rangle$, the Kohn-Sham orbitals, are related to the electron density via

$$\rho(r) = \sum_i |\phi_i(\vec{r})|^2. \quad (1.19)$$

This non-interacting kinetic energy T_s is different from the exact kinetic energy, T_e ; however, this difference is small, far smaller than the difference between the Thomas-Fermi kinetic energy and the exact kinetic energy for typical systems. Thus, one of the principle shortcomings of DFT has been overcome.

At this point, only two entities within the KS-DFT energy expression remain unknown: the non-classical electron-electron interaction energy, and the difference between the exact and non-interacting kinetic energies. These two terms, each relatively small in magnitude, constitute the exchange-correlation term in DFT, $E_{xc}[\rho(\vec{r})]$. Methodological developments within the DFT community are primarily centered around newer (and hopefully better) approximations to the exchange-correlation functional. Local spin density approximations[12, 13] consider E_{xc} to be a functional of only the density; generalized gradient approximations[14–17] (GGAs) incorporate the gradient of the density as well; meta-GGA functionals[18–21] include either the Laplacian or (more commonly) the kinetic energy density; and hybrid functionals[22–26] employ some amount of exact exchange. Such is the basis for the pseudo-hierarchy provided by the Jacob’s ladder of density functional approximations[27].

Unfortunately, conventional density functionals are inherently (semi-)local; they lack the long-range components necessary for correctly describing, e.g., van der Waals interactions. As such, an additional “stairway” of dispersion corrections to DFT has been constructed over the years[28]. The most simple models for dispersion rely on the addition of pairwise atomic corrections with the correct r^{-6} asymptote, with either pre-tabulated[29–31] or self-consistently-computed[32–36] isotropic dispersion coefficients. In recent years, attempts have even been made to develop density functionals with explicit nonlocal correlation[37–40].

1.6 Outline

This work is primarily concerned with the assessment and development of electronic structure approaches in the context of intermolecular interactions, the type of interaction that drives a variety of chemical processes, including gas sorption in nanoporous materials. Chapters 2-4 are generally concerned with the assessment of existing electronic structure approximations, while chapters 5 and 6 involve the development of novel methods.

Chapter 2

Adsorption of gas molecules in metal-organic frameworks is governed by many factors, the most dominant of which are the interaction of the gas with open metal sites, and the interaction of the gas with the ligands. In this chapter, we examine the latter class of interaction in the context of CO₂ binding to benzene. We begin by clarifying the geometry of the CO₂-benzene complex. We then generate a benchmark binding curve using a coupled-cluster approach with single, double, and perturbative triple excitations [CCSD(T)] at the complete basis set (CBS) limit. Against this Δ CCSD(T)/CBS standard, we evaluate a plethora of electronic structure approximations: Hartree-Fock, second-order Møller-Plesset perturbation theory (MP2) with the resolution-of-the-identity approximation, attenuated MP2, and a number of density functionals with and without different empirical and nonempirical van der Waals corrections. We find that finite-basis MP2 significantly overbinds the complex. On the other hand, even the simplest empirical correction to standard density functionals is sufficient to bring the binding energies to well within 1 kJ/mol of the benchmark, corresponding to an error of less than 10%; PBE-D in particular performs well. Methods that explicitly include nonlocal correlation kernels, such as VV10, vdW-DF2, and ω B97X-V, perform with similar accuracy for this system, as do ω B97X and M06-L. This work[41] has been published in *The Journal of Chemical Physics*.

Chapter 3

Electronic structure approaches for calculating intermolecular interactions have traditionally been benchmarked almost exclusively on the basis of energy-centric metrics. In this chapter, we explore the idea of utilizing a metric related to geometry. On a diverse series of

non-covalently-interacting systems of different sizes, from the water dimer to the coronene dimer, we evaluate a variety of electronic structure approximations with respect to their abilities to reproduce coupled-cluster-level geometries. Specifically, we examine Hartree-Fock, second-order Møller-Plesset perturbation theory (MP2), attenuated MP2, scaled MP2, and a number of density functionals, many of which include empirical or nonempirical van der Waals dispersion corrections. We find a number of trends that transcend system size and interaction type. For instance, functionals incorporating VV10 nonlocal correlation tend to yield highly accurate geometries; ω B97X-V and B97M-V in particular stand out. We establish that intermolecular distance, as measured by, e.g., the center-of-mass separation of two molecules, is the geometric parameter that deviates most profoundly among the various methods. This property of the equilibrium intermolecular separation, coupled with its accessibility via a small series of well-defined single-point calculations, makes it an ideal metric for the development and evaluation of electronic structure methods. This work[42] has been published in *Journal of Chemical Theory and Computation*.

Chapter 4

With the aim of systematically characterizing the convergence of common families of basis sets such that general recommendations for basis sets can be made, we have tested a wide variety of basis sets against complete-basis binding energies across the S22 set of intermolecular interactions – noncovalent interactions of small and medium-sized molecules consisting of first- and second-row atoms – with three distinct density functional approximations: SPW92, a form of local-density approximation; B3LYP, a global hybrid generalized gradient approximation; and B97M-V, a meta-generalized gradient approximation with non-local correlation. We have found that it is remarkably difficult to reach the basis set limit; for the methods and systems examined, the most complete basis is Jensen’s pc-4. The Dunning correlation-consistent sequence of basis sets converges slowly relative to the Jensen sequence. The Karlsruhe basis sets are quite cost effective, particularly when a correction for basis set superposition error (BSSE) is applied: counterpoise-corrected def2-SVPD binding energies are better than corresponding energies computed in comparably sized Dunning and Jensen bases, and on par with uncorrected results in basis sets 3-4 times larger. These trends are exhibited regardless of the level of density functional approximation employed. A sense of the magnitude of the intrinsic incompleteness error of each basis set not only provides a foundation for guiding basis set choice in future studies, but also facilitates quantitative comparison of existing studies on similar types of systems. This work[43] has been published in *The Journal of Chemical Physics*.

Chapter 5

With the aim of improving the utility of the DFT-D3 empirical dispersion correction, in this chapter we generalize the DFT-D3 damping function by optimizing an additional parameter, an exponent, which controls the rate at which the dispersion tail is activated. This

method – DFT-D3(op), shorthand for “optimized power,” where power refers to the newly introduced exponent – is then parameterized for use with ten popular density functional approximations across a small set of non-covalent interactions and isomerization energies; the resulting methods are then evaluated across a large independent test set of 2475 non-covalent binding energies and isomerization energies. We find that the DFT-D3(op) tail represents a substantial improvement over existing damping functions, as it affords significant reductions in errors associated with non-covalent interaction energies and geometries. The revPBE0-D3(op) and MS2-D3(op) methods in particular stand out, and our extensive testing indicates they are competitive with other modern density functionals. This work[44] has been accepted by *Journal of Chemical Theory and Computation*.

Chapter 6

With the aim of mitigating the basis set error in density functional theory (DFT) calculations employing local basis sets, in this chapter we develop two empirical corrections for basis set superposition error (BSSE) in the def2-SVPD basis, a basis which – when stripped of BSSE – is capable of providing near-complete-basis DFT results for non-covalent interactions. Specifically, we adapt the existing pairwise geometrical counterpoise approach (gCP) to the def2-SVPD basis, and we develop a beyond-pairwise approach, DFT-C, which we parameterize across a small set of intermolecular interactions. Both gCP and DFT-C are evaluated against the traditional Boys-Bernardi counterpoise correction across a set of 3402 non-covalent binding energies and isomerization energies. We find that the DFT-C method represents a significant improvement over gCP, particularly for non-covalently-interacting molecular clusters. Moreover, DFT-C is transferable among density functionals and can be combined with existing functionals – such as B97M-V – to recover large-basis results at a fraction of the cost. This work[45] has been submitted to *The Journal of Chemical Physics*.

Chapter 2

Case Study: The CO₂-Benzene Complex

2.1 Introduction

Metal-organic frameworks (MOFs) constitute a promising class of materials for CO₂ separation[46]. MOFs are three-dimensional nanoporous solids consisting of metal centers connected by organic ligands; there exist a tremendous number of synthetically accessible combinations of ligands and metals[47]. Extensive studies have been conducted to determine the impact of changing the metal center on CO₂ adsorption[48–52], but there is a relative paucity of work emphasizing the contribution of the ligand[53]. Although open metal sites are commonly regarded as the most thermodynamically favorable binding locations for gas molecules, the ligands themselves also contribute significantly to adsorption, and are hence worthy of further investigation[53, 54].

A handful of studies have involved the investigation of the binding of CO₂ to various forms of organic linkers, predominantly functionalized variants of benzene[55–58], but to date only one study has considered the interaction at nonequilibrium distances[59]. These nonequilibrium distances are particularly relevant in the context of CO₂ uptake in MOFs. Additionally, no studies have systematically evaluated the ability of the various existing electronic structure approximations to describe this gas-ligand interaction. The existing literature has established density functional theory (DFT) as the primary workhorse for such systems, a consequence of its ability to balance computational cost [typically $O(N^3)$] with accuracy[10, 11]. However, the famed Jacob’s ladder of DFT does not offer a clear way to systematically improve results[27]. Moreover, the local nature of DFT within standard approximations renders it incapable of describing long-range London dispersion, an integral component of noncovalent interactions such as gas-ligand binding[60]. To account for this, a number of dispersion corrections to standard density functional theories have been developed: the DFT-D, DFT-D2, and DFT-D3 approaches of Grimme[29, 30, 61], the exchange-hole dipole moment (XDM6 and XDM10) model of Becke and Johnson[32–35, 62–

64], and methods incorporating fully nonlocal correlation kernels[37–40]. For an overview of these methods (and many more), the reader is referred to a recent review by Klimeš and Michaelides[28]. This additional “stairway” of dispersion corrections further complicates the DFT picture. Although efforts are underway to develop generally-applicable functionals, it is often not clear *a priori* what the best functional for a particular application is.

In stark contrast, the hierarchy of wavefunction-based (WF) methods offers a clear path to systematically obtaining more accurate energies, but at much steeper costs: even the simplest method of incorporating electronic correlation in WF theory, second-order Møller-Plesset perturbation theory (MP2), scales as $O(N^5)$, and methods that incorporate higher-order correlation effects, such as coupled-cluster theory with single, double, and perturbative triple excitations [CCSD(T)], suffer from even worse scaling. Even MP2 can be prohibitively expensive for calculations involving extended systems. The resolution-of-the-identity (RI) approximation partly addresses this cost by reducing the prefactor of the calculation, but the underlying scaling is left unchanged[65–68]. Additionally, MP2 performs quite poorly for certain classes of systems (notably those involving aromatic interactions) and converges quite slowly with respect to the size of the basis set[69, 70]. Attenuation of the Coulomb operator (attenuated MP2) attempts to address some of these problems (as well as that of basis set superposition error, or BSSE), with the added benefit of introducing sparsity that could be exploited to potentially reduce scaling[71, 72].

In this work, we evaluate a multitude of electronic structure approximations for the CO₂-benzene complex, a well-defined model system that is also a simple model of a gas binding to a nonfunctionalized organic linker. We begin by elucidating the equilibrium geometry of the complex, since there are conflicting reports in the literature[57, 58, 73]. We then characterize the CO₂-benzene potential energy surface using a wide variety of methods, assessing their strengths and weaknesses. We find MP2 significantly overbinds the complex, particularly in a finite basis, and hence is unsuitable as a benchmark in dispersion-bound systems. We also find standard semilocal density functionals are – with some notable exceptions – generally inadequate for this particular system, although this deficiency is rectifiable via the addition of a simple C_6 dispersion correction or an explicitly nonlocal correlation kernel.

2.2 Computational Methods

Throughout this study, all MP2 calculations employed the RI approximation, but the RI prefix has been omitted. For all geometry optimizations, we utilized the dual-basis approximation to HF in conjunction with MP2 (DB-MP2)[74–76], with basis set pairings from Steele et al.[77]. We feel this approximation is justified based on the high degree of accuracy it has demonstrated in previous studies[77, 78], as well as in this particular study.

A series of geometry optimizations were performed for the various possible symmetries (C_{2v} , C_s , and C_1) of the flat-on conformations of the benzene-CO₂ system at the DB-MP2 level with the dual-basis pairing for Dunning’s augmented correlation-consistent polarized valence double-zeta basis (aug-cc-pVDZ)[77, 79, 80]. The Hessians and geometries for these

calculations were then utilized as initial guesses in subsequent geometry optimizations at the DB-MP2 level with the dual-basis pairing for the larger aug-cc-pVTZ basis. Single point energy calculations were performed at the same level of theory for each structure in the aug-cc-pVQZ basis, then extrapolated to the basis set limit using the two-point (T → Q) formula of Helgaker[70]. The conformation exhibiting the lowest energy at the extrapolated CBS limit was determined to be the optimal structure and was utilized in subsequent benchmarking.

In order to evaluate various methods, a series of one-dimensional potential energy surfaces of the benzene-CO₂ system were generated. The intermolecular structural parameter for these binding curves was the projected distance between the carbon atom in CO₂ and the ring center of benzene onto the principal axis of benzene. A wide variety of electronic structure methods were considered in the aug-cc-pVTZ basis: MP2[5, 81], attenuated MP2[71, 72], CCSD(T)[9], B3LYP[15, 16, 22, 82], B97[17], B97-D[30], PBE[14], PBE0[83], TPSS[84], TPSSh[18], M06-L[20], M06-2X[25], M11[85], ωB97X[86], ωB97X-D[87], ωB97X-V[23], vdW-DF2 (rPW86+PW92)[14, 38, 88], VV10 (rPW86+PBE)[13, 40, 88], XDM6 (PW86+PBE)[14, 32, 33, 89], and Grimme -D2 and -D3 corrections to some of the standard GGA functionals[30, 31]. The MP2 correlation energy was extrapolated to the basis set limit (T → Q)[70], and a frozen-core (fc) CCSD(T) correction to the MP2/CBS correlation energy was calculated in the cc-pVTZ basis as[90]:

$$E_{\Delta\text{CCSD(T)}}^{\text{CBS}} = E_{\text{MP2}}^{\text{CBS}} + (E_{\text{CCSD(T) (fc)}} - E_{\text{MP2 (fc)}})^{\text{TZ}}. \quad (2.1)$$

This benchmark was calculated both with and without the counterpoise correction (CP) for basis set superposition error (BSSE)[91], and the region bounded by the two curves, denoted ΔCCSD(T)/CBS, represents the standard against which all other methods are evaluated. With the exception of this benchmark curve, all other results reported herein do not employ any form of counterpoise correction. Such corrections are expensive and the ideal method should not necessitate them; furthermore, in the aug-cc-pVTZ basis, these corrections are typically very small for density functionals, a consequence of the relatively fast convergence of most density functionals with respect to basis set size[92].

A fine Lebedev integration grid consisting of 75 radial points and 302 angular points was utilized in the computation of the exchange-correlation components of all of the density functionals. With the exception of CCSD(T) and MP2, all calculations were performed in the aug-cc-pVTZ basis. Moreover, for the RI calculations, the auxiliary basis sets of Weigend were used[93]. For all calculations, integral thresholds of 10⁻¹⁴ a.u. were used, and no symmetry was exploited, save the symmetry of the system itself during all geometry optimizations. All calculations were performed with a development version of Q-Chem 4.1[94]. Molecular structures were generated with Avogadro[95].

2.3 Results and Discussion

Geometry

In this study, the geometry of the benzene-CO₂ system was optimized at the DB-MP2 level in the dual-basis pairing for aug-cc-pVTZ. The optimized geometry of the benzene-CO₂ dimer is provided in Figure 2.1. The optimal dimer geometry, unlike that reported by Mackie and DiLabio[57], is C_s-symmetric, with the principal axis of CO₂ bisecting a carbon-carbon bond of benzene. The CO₂ molecule exhibits a slight deviation from linearity and is not strictly parallel to the benzene ring. These findings compare favorably with related studies by Besnard et al.[73]. and Chen et al.[58].

There exists a small (1 kJ/mol) thermodynamic driving force to break C_{2v} symmetry, but the potential energy surface appears to be largely invariant to rotation of the CO₂ about the principal axis of benzene, as evidenced by calculations involving the atom-centric C_s structure.

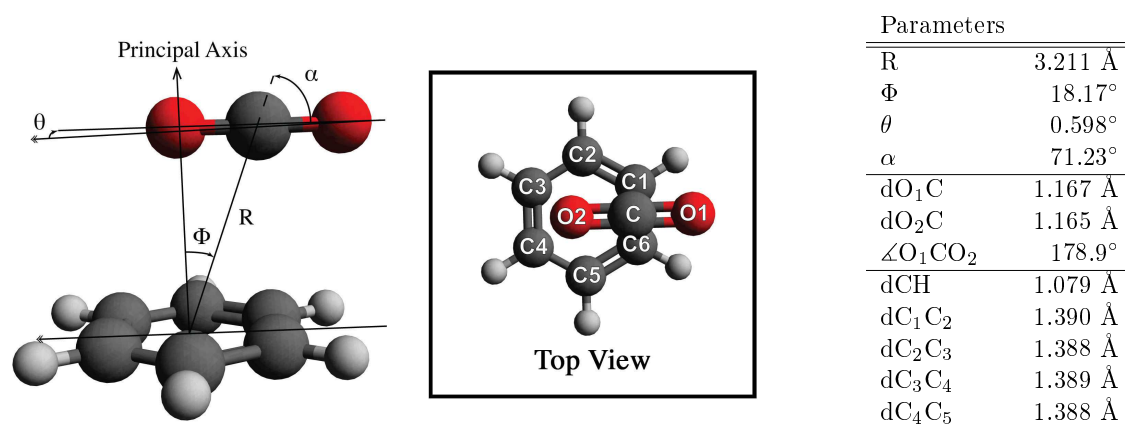


Figure 2.1: Calculated equilibrium structure of the benzene-CO₂ system, determined at the DB-MP2 level of theory with the dual-basis pairing for Dunning’s augmented correlation-consistent polarized valence triple-zeta basis. The complex possesses C_s symmetry.

Binding Curves

In order to assess the ability of various electronic structure methods to describe the interaction between benzene and CO₂, the optimized structure of CO₂ was rigidly translated along the principal axis of benzene. Twelve distinct intermolecular separations were thus sampled, where the intermolecular separation has been defined as the projection onto the principal axis of benzene of the distance between the carbon atom of CO₂ and the center of the ring. The resulting potential energy surfaces are portrayed in Figures 2.2-2.7.

We begin our analysis with the standard wavefunction-based methods, with results summarized in Figure 2.2. The region bounded by the counterpoise-corrected (CP) and noCP

$\Delta\text{CCSD(T)}/\text{CBS}$ curves represents our best guess of the true potential energy surface along this coordinate and the uncertainty associated therewith. The equilibrium intermolecular separation at this level of theory is found to be 3.25 ± 0.01 Å, corresponding to a binding energy on the order of 10.2 ± 0.4 kJ/mol. It is worth noting that, contrary to expectations, the CP curve actually lies below the noCP curve, which may be indicative of the limitations of extrapolation to the CBS limit.

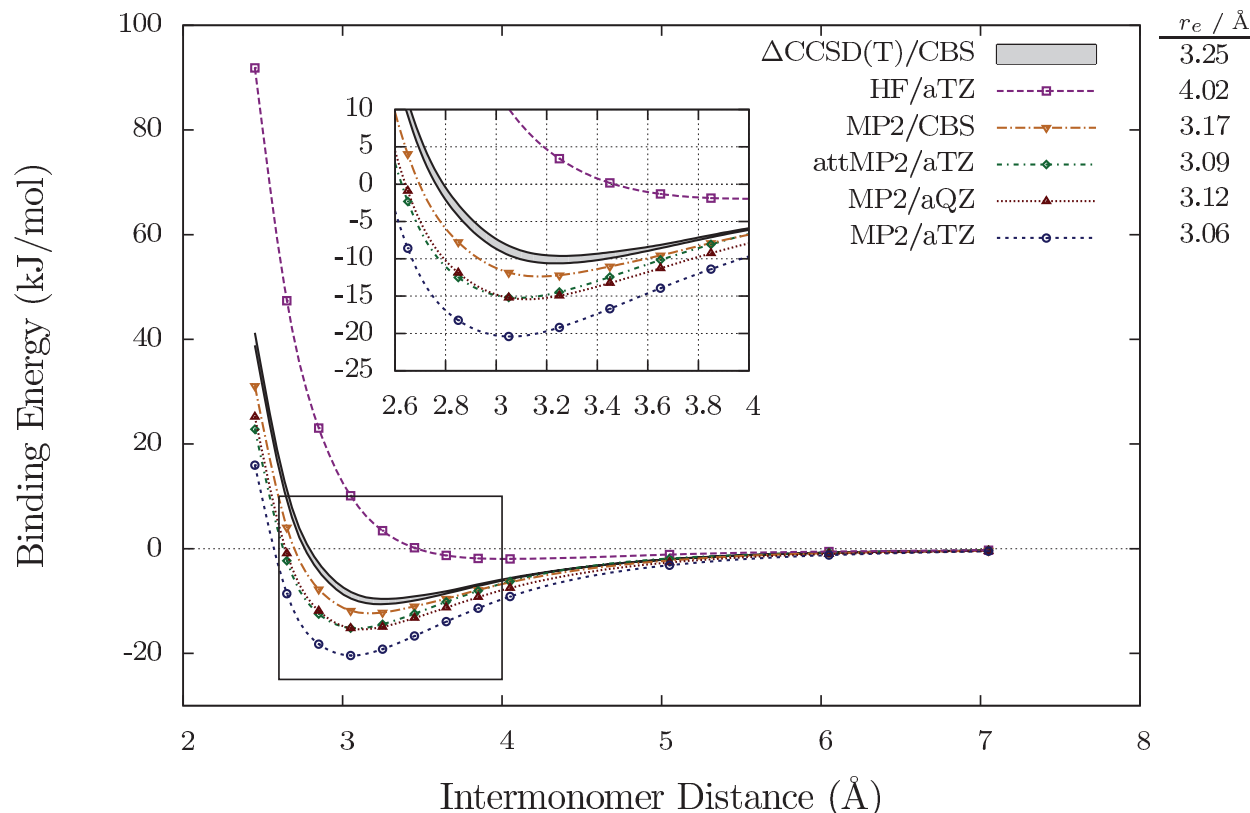


Figure 2.2: Binding energy curves for the benzene-CO₂ complex obtained with various wavefunction methods. The region denoted $\Delta\text{CCSD(T)}/\text{CBS}$ is bounded by CP- and non-CP-corrected $\Delta\text{CCSD(T)}$ curves, and represents our best guess at the true binding curve. None of the other methods employ the CP correction. The inset is a close-up of the region around the equilibrium distance (note different scale), and the curves connecting the points are merely smooth fits to the data. The equilibrium intermolecular separations for each method (listed to the right of the legend as r_e) were determined by cubic spline interpolation of the data. The reported r_e for $\Delta\text{CCSD(T)}/\text{CBS}$ is the average of the $\Delta\text{CCSD(T)}/\text{CBS}$ noCP and CP equilibrium intermolecular separations. For ease of analysis, the legend has been ordered (with the exception of the benchmark $\Delta\text{CCSD(T)}/\text{CBS}$ region) to mirror the ordering of the binding curves near the equilibrium separation.

Relative to the $\Delta\text{CCSD(T)}/\text{CBS}$ benchmark, Hartree-Fock theory (HF) provides a rather

poor description of the potential energy surface; this is unsurprising, given the lack of correlation beyond Pauli exclusion in HF. However, HF theory does yield a bound complex, with a binding energy curve whose asymptotic form scales with the inverse fifth power of intermolecular distance. This is consistent with expectations: both CO₂ and benzene are known to be characterized by significant quadrupole moments.

The addition of MP2 correlation significantly improves the description of the complex relative to HF. However, MP2 theory systematically predicts overbinding in this particular system, regardless of the choice of basis set. As the size of the basis set increases, so too does the quality of the description, but even at the extrapolated CBS (T → Q) limit, the minimum binding energy is in error by upwards of 15% relative to the benchmark $\Delta\text{CCSD(T)}/\text{CBS}$ value. Moreover, the equilibrium separation of the benzene and CO₂ molecules is significantly smaller (on the order of 0.2 Å) for finite-basis MP2 than for $\Delta\text{CCSD(T)}/\text{CBS}$, which is again indicative of overbinding. Fundamentally, this overbinding is attributable to the fact that the frequency-dependent polarizabilities (and hence C_6 coefficients, as per the Casimir-Polder relation) of MP2 are identical to those of uncoupled HF (UCHF) theory[96]. More specifically, the UCHF polarizabilities of benzene are likely the chief contributor to overbinding; the carbon dioxide dimer is well-described by MP2 theory, whereas the benzene dimer is known to be significantly overbound[69, 97]. It is clear from these results that MP2 is unsuitable as a benchmark for systems with significant dispersion interactions, even in the large aug-cc-pVQZ basis, and hence a more complete description of correlation is required.

Attenuation of the Coulomb operator in MP2 (attMP2 in Figure 2.2) addresses some of this overbinding. In fact, the attenuated MP2 method in the aug-cc-pVTZ basis performs similarly to standard MP2 in the much larger aug-cc-pVQZ basis, at least in the vicinity of the binding minimum. Unfortunately, this improvement comes at the expense of the description at larger intermolecular separations – the rate with which the attenuated MP2 interaction energy diminishes with separation is unphysical. Moreover, attenuated MP2 predicts a tighter equilibrium geometry of the complex than $\Delta\text{CCSD(T)}$. Despite these caveats, attenuated MP2 still represents a notable improvement over standard, finite-basis MP2 in this system.

In addition to the aforementioned wavefunction-based approaches, we have applied a number of density functionals of varying degrees of sophistication to this system. The simplest such functionals utilized here are of the GGA variety: PBE; the global hybrid GGAs PBE0, B97, and B3LYP; and the long-range corrected hybrid GGA ωB97X . As is evident from Figure 2.3, all of the GGA functionals employed, with the exception of ωB97X , produce qualitatively-incorrect binding energy curves that bear striking resemblance to the HF curve, and it is notable that all of these methods predict a bound complex. The strikingly-good performance of ωB97X could be, to some extent, a result of the incorporation of range-separated exchange, though it is more likely an artifact of training: unlike the other GGA functionals examined here, ωB97X was trained on noncovalent interactions.

Slightly more sophisticated than these GGA functionals are functionals of the meta-GGA variety. Here, we examined the empirical TPSS, its hybrid variant TPSSh, and a number of Minnesota functionals: the local M06-L, the global hybrid M06-2X, and the more recent

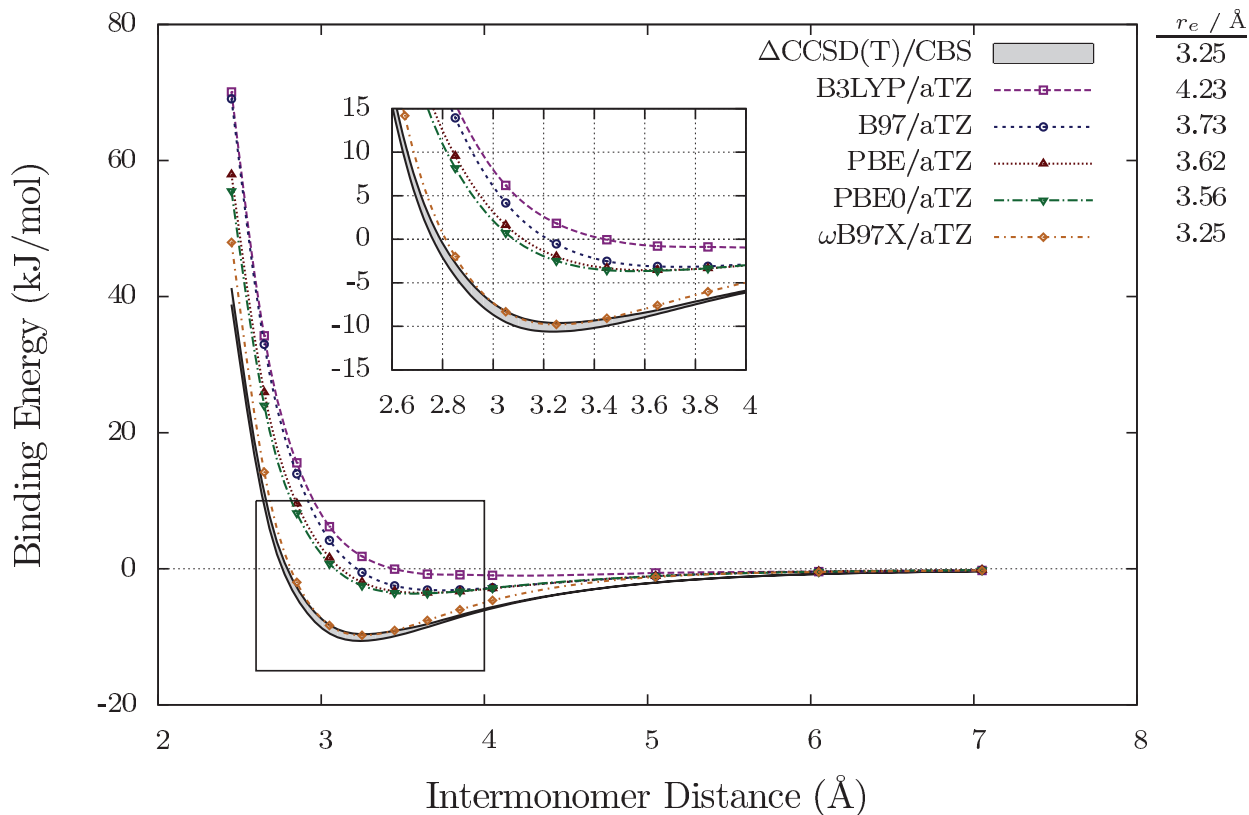


Figure 2.3: Binding energy curves for the benzene-CO₂ complex obtained with various standard, hybrid, and long-range corrected hybrid GGA functionals in the aug-cc-pVTZ basis with a (75,302) integration grid. For further details, see Figure 2.2.

range-separated hybrid M11. Binding curves obtained using these functionals are provided in Figure 2.4. We see that the empirical TPSS and TPSSH functionals suffer from similar issues as the GGA functionals: the kinetic energy density is not sufficiently nonlocal so as to capture dispersion interactions. The Minnesota functionals provide a much better description of the binding, though this is likely attributable to their high degrees of parametrization, as well as the fact that their training sets included noncovalent interactions. Unfortunately, this improved description is not without its flaws: the Minnesota functionals, particularly M06-2X and M11, are hampered by incorrect asymptotics and unphysical oscillations in binding energy with increasing separation[98, 99]. M06-L is clearly the top performer amongst the meta-GGA functionals for this particular system, even outperforming MP2 at the basis set limit in the vicinity of the equilibrium binding distance.

The next degree of sophistication of DFT explored here constitutes a departure from Perdew’s ladder – we’ll henceforth be traversing the stairway of Klimeš and Michaelides[28]. The first such step we consider involves Grimme’s empirical DFT-D2 correction to GGA functionals (henceforth referred to, for simplicity, as DFT-D, since the original iteration of

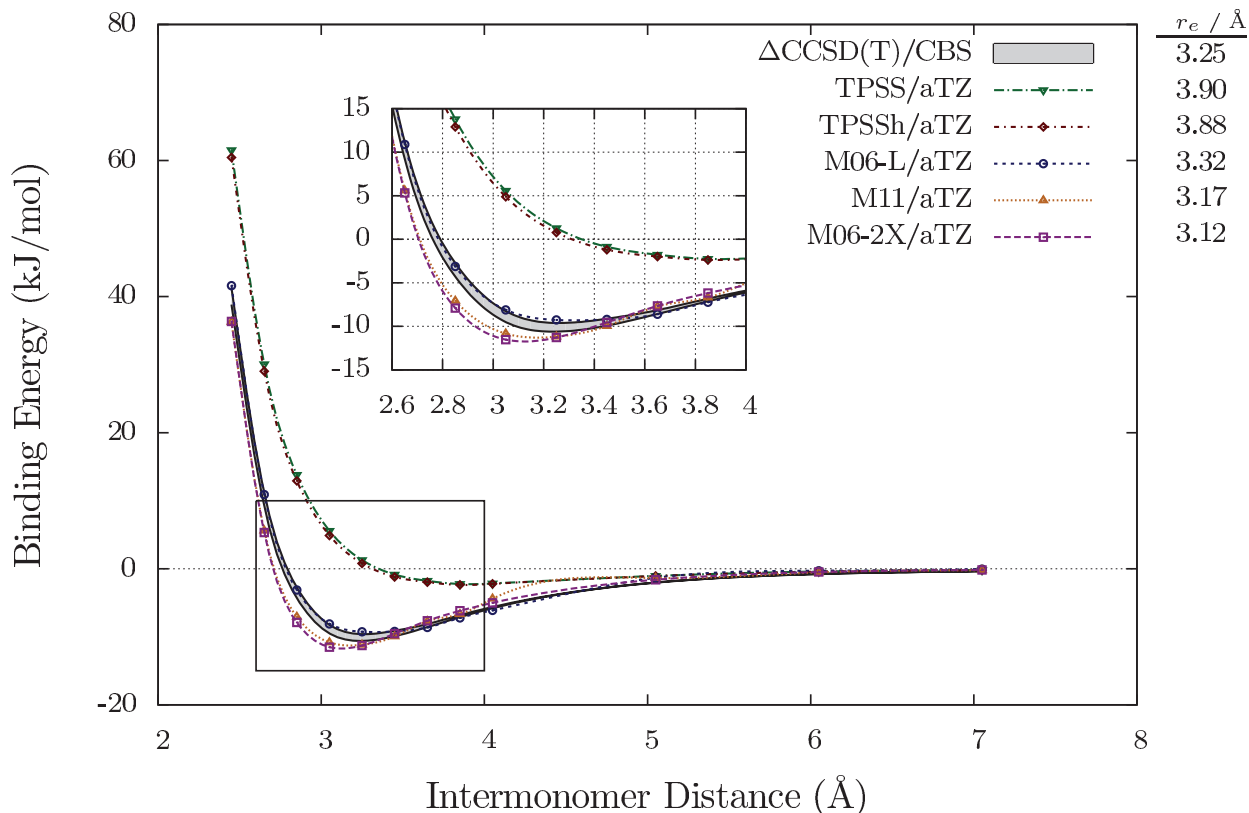


Figure 2.4: Binding energy curves for the benzene-CO₂ complex obtained with various standard, hybrid, and range-separated hybrid meta-GGA functionals in the aug-cc-pVTZ basis with a (75,302) integration grid. For further details, see Figure 2.2.

DFT-D is not in common use)[30]. We consider -D corrections to B3LYP, PBE, and PBE0, and we also examine Grimme’s unhybridized B97-D, as well as ω B97X-D. The binding energy curves associated with these methods are given in Figure 2.5. Comparing these results to those of Figure 2.3, we see that addition of even the simplest form of dispersion correction improves the description of binding in the benzene-CO₂ complex drastically. However, there is a caveat: the choice of damping function is quite important. PBE0-D exhibits significant overbinding near the equilibrium separation when the damping function associated with Grimme’s DFT-D is used; however, use of the damping function of Chai and Head-Gordon[87] (CHG) yields significantly better energetics near the equilibrium geometry, though at the cost of lethargic decay of the binding energy in the intermediately-stretched regime. See Figure 2.6 – the PBE0-D (CHG) and PBE0-D3 curves are very similar. Among the DFT-D methods, PBE-D exhibits the best performance, only deviating from the benchmark curves at very compressed geometries, where the overly repulsive PBE exchange contribution dominates.

The next step of complexity in describing dispersion interactions involves factoring in the local environment to the standard C_6 correction. This encompasses the DFT-D3 correction

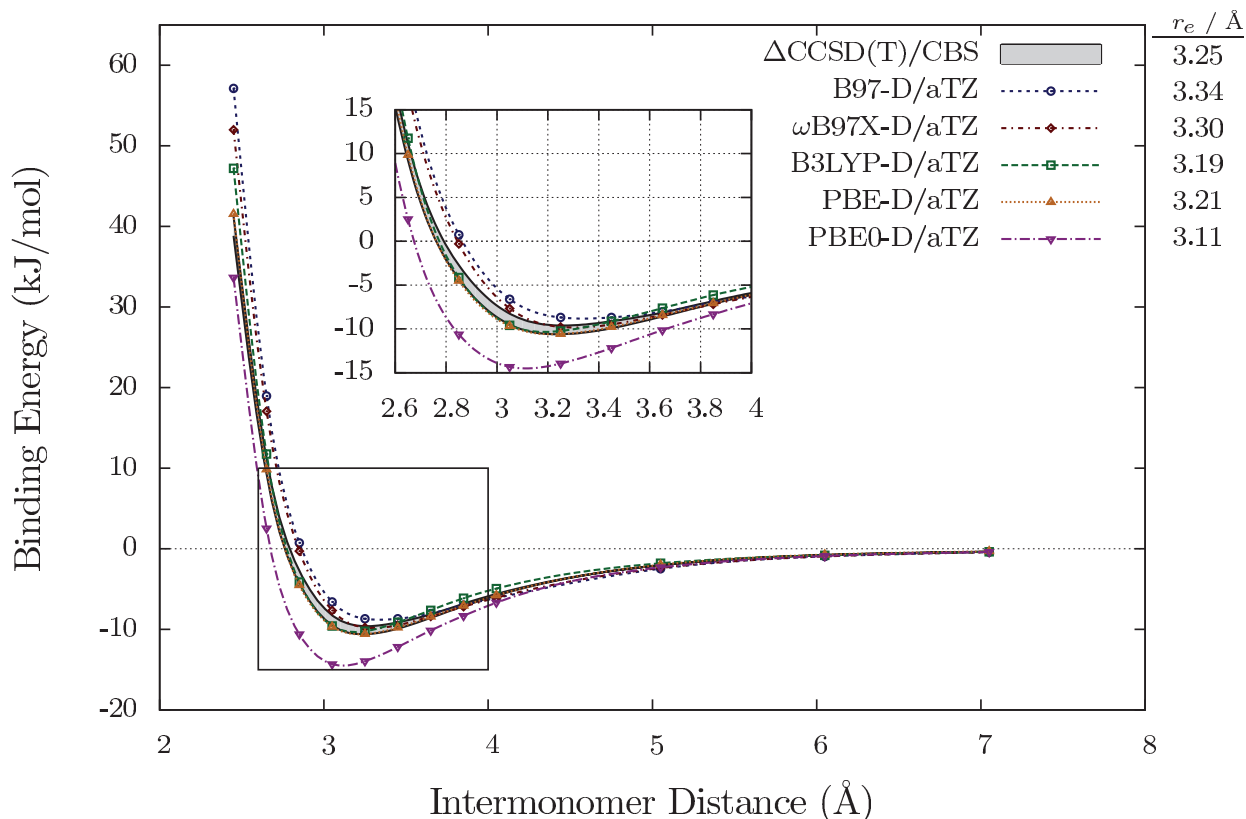


Figure 2.5: Binding energy curves for the benzene-CO₂ complex obtained with various DFT-D2 methods in the aug-cc-pVTZ basis with a (75,302) integration grid. For further details, see Figure 2.2.

of Grimme, as well as the XDM model of Becke and Johnson[32, 61]. Binding energy curves for PBE-D3, PBE0-D3, and XDM6 (with PW86 exchange and PBE correlation) are provided in Figure 2.6. Comparing the PBE-D3 and PBE0-D3 curves to the DFT-D curves of Figure 2.5, we see that although accounting for the local environment introduces another layer of complexity, this additional complexity does not necessarily correspond to an improvement of the description – the -D3 methods exhibit curves which are shifted significantly (around 0.1 Å) towards larger intermolecular separations, with broader wells. Moreover, comparison with the PBE0-D (CHG) results shows that perhaps the largest difference between the -D and -D3 methods is not actually this environmental dependence, but rather the choice of damping function: DFT-D3 utilizes the same form of damping function as -D (CHG).

The XDM approach, though also quite good here, comes with its own disclaimer: it is highly sensitive to the choice of exchange and correlation functionals. A number of other xc functionals were examined in conjunction with both XDM6 and XDM10 (using parameters recommended by various authors), but the performances of most are unsatisfactory – for instance, the PW86+PBE+XDM10 approach, using optimized parameters from Becke et

al.[63], shifts the binding minimum to the right by roughly 0.4 Å and only recovers 25% of the equilibrium binding energy).

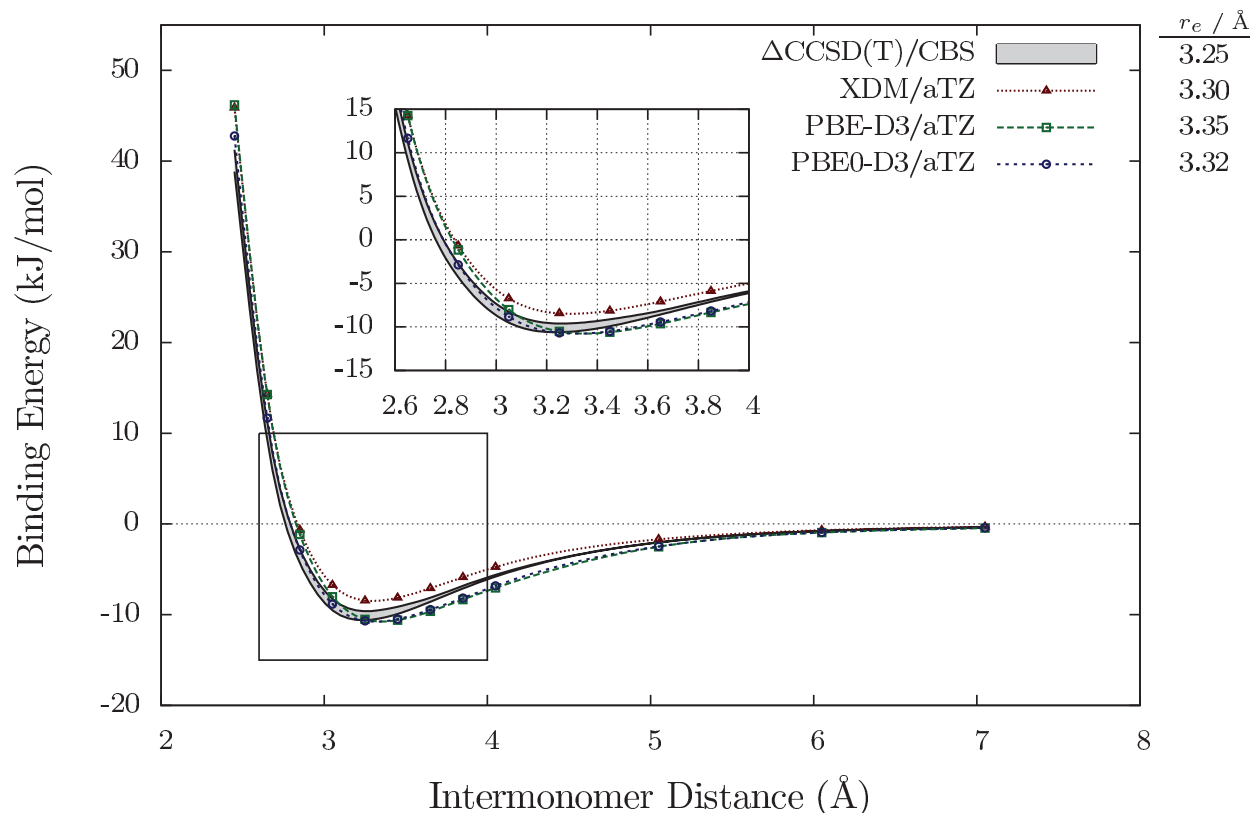


Figure 2.6: Binding energy curves for the benzene-CO₂ complex obtained with various methods employing environmentally-dependent C_6 corrections in the aug-cc-pVTZ basis with a (75,302) integration grid. For further details, see Figure 2.2.

The most sophisticated class of methods examined in this study explicitly incorporate nonlocal contributions to correlation: vdW-DF2, VV10, and $\omega\text{B97X-V}$. As illustrated in Figure 2.7, these methods describe the binding of CO₂ to benzene reasonably well through all intermolecular separations examined. All of the nonlocal methods overbind slightly at their respective equilibrium geometries, which is indicative of marginally inflated C_6 coefficients. The issues associated with exchange-matching in vdW-DF2 are manifest in fact that the entire vdW-DF2 curve appears to be shifted 0.1 Å to the right of the $\Delta\text{CCSD(T)}$ benchmark, but other than this, none of the methods here deviate by much more than 1 kJ/mol except at the most compressed geometry, where the overly repulsive exchange components begin to dominate.

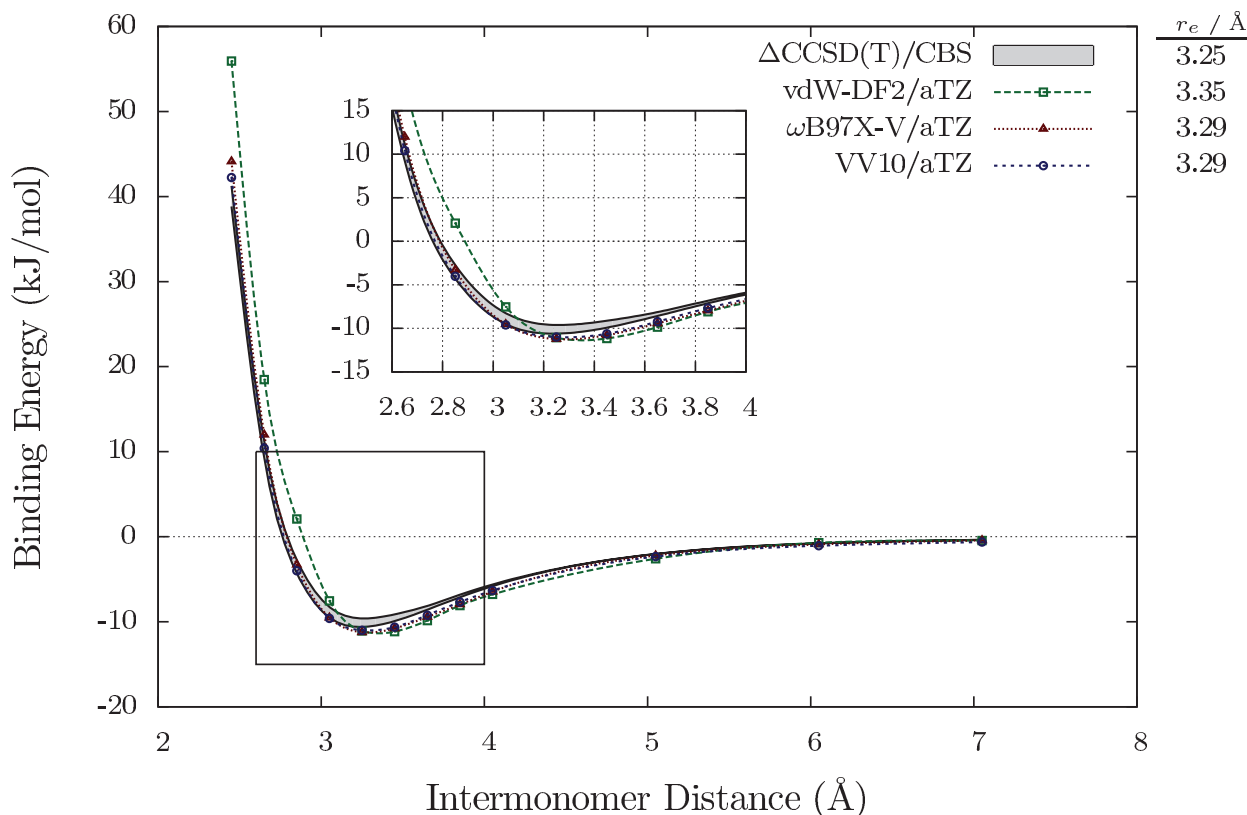


Figure 2.7: Binding energy curves for the benzene-CO₂ complex obtained with various non-local methods in the aug-cc-pVTZ basis with a (75,302) integration grid. For further details, see Figure 2.2.

2.4 Conclusion

In this work, we have clarified the geometry of the benzene-CO₂ complex and systematically generated a series of binding energy curves for the complex using a multitude of electronic structure approximations. More specifically, we have assessed the efficacy of HF, MP2, CCSD(T), and a number of DFT exchange-correlation functionals with and without different empirical and nonempirical van der Waals corrections with regard to the description of binding in the benzene-CO₂ complex.

We find that although HF predicts a bound complex – a consequence of a significant quadrupole-quadrupole electrostatic interaction – the energetics and intermonomer distance predicted at such a low level of theory leave much to be desired; the inclusion of electronic correlation is necessary to achieve any semblance of quantitative accuracy. Moreover, this description of electronic correlation must be sufficiently nonlocal, as exemplified by the failure of conventional semi-local density functionals (e.g. PBE, TPSS) to correctly reproduce the Δ CCSD(T) binding energy curve. Some semi-empirical functionals, namely those in

the Minnesota suite, perform quite well in the vicinity of the minimum, but they are hampered by incorrect asymptotics and unphysical oscillations. Standouts amongst the standard GGA and meta-GGA functionals are ω B97X and M06-L, which describe the binding quite well throughout a range of intermolecular separations; the exemplary performance of these functionals may be largely attributable to the inclusion of noncovalent interactions in their respective training sets.

The simplest wavefunction-based approach to correlation – MP2 theory – while qualitatively correct, yields substantial overbinding in finite basis sets, and overbinds noticeably even when extrapolated ($T \rightarrow Q$) to the CBS limit. Attenuation of the Coulomb operator addresses some of this overbinding and improves the performance of MP2 in the vicinity of the equilibrium separation, but at the cost of the long-range limit. In contrast, even the simplest density-functional-based method of treating dispersion, i.e. the addition of an empirical C_6 correction (DFT-D), is sufficient to bring standard GGA functionals within 1 kJ/mol of the Δ CCSD(T) benchmark, depending on the choice of damping function: PBE-D in particular performs quite well. Methods that explicitly address nonlocal correlation – VV10, vdW-DF2, and ω B97X-V – are also rather well-suited to this system, although vdW-DF2 predicts a characteristically too-large intermonomer distance, consistent with an overly-repulsive exchange component, and all of the nonlocal methods overbind slightly at their respective equilibrium separations.

The extent to which the impressive accuracies of the DFT-D, M06-L, ω B97X-V, and VV10 methods are transferable to larger, more diverse systems – namely a variety of MOFs – remains to be seen. The vdW-DF2 method has been validated on a wide variety of systems, and its performance has been remarkably consistent[54, 100–103]. To date, VV10 has only been benchmarked on noncovalent interactions in systems containing small molecules; although the results have been promising – particularly when range-separated exchange is employed – the method has yet to be validated for larger systems, such as MOFs[102, 104]. On the other hand, PBE-D has been demonstrated to be reasonably successful at predicting CO₂ adsorption energies in a few MOF variants[54, 103], though the failure of PBE-D3 to describe H₂ binding to Cu surfaces again underscores the potential sensitivity of the -D brand of methods to the choice of damping function and C_6 coefficients[101]. M06-L has recently been applied, with some success, to hydrocarbon adsorption in Fe-MOF74[105]. The ω B97X-V functional is quite new, and although it has excelled in a wide variety of systems, the range-separated exchange component hinders its adaptation into periodic codes[23]. This is not necessarily a problem, as calculations on cluster models of MOFs have the potential to be quite accurate[105, 106], but it is certainly a limitation of not only ω B97X-V, but also long-range corrected variants of VV10.

Chapter 3

A New Paradigm for Method Assessment: Geometries

3.1 Introduction

There exist a tremendous number of approaches to approximately solving the non-relativistic, time-independent Schrödinger equation for a many-electron system, and it is often not clear *a priori* what the optimal choice for a particular application is. The well-defined hierarchy of wavefunction-based (WF) methods offers a clear path to obtaining highly accurate energies, though at great expense. The simplest method for adding electronic correlation to the Hartree-Fock (HF) mean field ansatz[107, 108], second-order Møller-Plesset perturbation theory (MP2)[5], scales as $O(N^5)$, and more highly-correlated methods, such as coupled-cluster theory with single, double, and perturbative triple excitations – CCSD(T)[9] – exhibit even worse scaling ($O(N^7)$ in the case of CCSD(T)). The resolution-of-the-identity (RI) approximation[65–68] partly addresses this cost by reducing the computational prefactor, but the underlying scaling of the method to which it is applied is left unchanged. Additionally, correlated WF methods exhibit slow convergence with respect to basis set size, a consequence of their inclusion of excited-state determinants and the correspondingly large number of basis functions required to accurately describe the virtual space associated therewith[69, 70]. Though still in its infancy, attenuation of the Coulomb operator is one promising means of addressing not only these issues, but basis set superposition error (BSSE) as well[71, 72].

In stark contrast to the clear-cut hierarchy of WF methods, the well-known Jacob’s ladder[27] of density functional theory (DFT)[10, 11] offers no clear way to systematically improve results. Moreover, the inherently local description provided by DFT within standard approximations renders it incapable of recovering long-range dispersion[60]. In the absence of strong permanent electrostatic interactions, these second-order effects are crucial for the correct description of non-covalent interactions, which will be the focus of this chapter. In the past decade, a variety of means of accounting for long-range van der Waals dispersion

forces have been proposed, from simple pairwise C_6 corrections to the exchange-correlation energy[29, 30, 33, 36, 61] to the inclusion of fully nonlocal correlation kernels[37–40].

When evaluating the performance of any of these various electronic structure approaches, the recent literature has focused almost exclusively on the ability of the method to reproduce “exact” electronic energies. Whenever a promising new method is developed, a flurry of studies arise wherein this method is applied to a variety of systems of physical interest. The common thread in such studies is their focus on energies; for instance, in a non-covalently interacting system, the metric of choice is typically the binding energy, defined – at least for size-consistent methods – as the difference between the total energy of the system and the energies of its constituent molecules. Although there may be subtle differences in the precise definition of the binding energy among various studies (e.g., the fragments may or may not be allowed to relax) the fact remains that it is an objective metric: it provides a means of methodically comparing electronic structure approaches. Unfortunately, this sort of energy-centric approach to method evaluation is far from perfect. The ideal method would recover the entire “exact” potential energy surface, not just a single point; it would reproduce “exact” geometries as well as energies.

This energy-centric focus has been justified repeatedly over the years by studies in which intramolecular geometric parameters were considered. It has been well established that – in any reasonable basis – even HF yields accurate bond lengths and angles for many systems, and differences between various methods, as measured on the basis of these metrics, tend to be minimal[109–114]. Although these sorts of intramolecular metrics are sufficient for describing geometries of small single molecules, they are inadequate in the context of large molecules and systems composed of multiple molecules, i.e. systems involving significant non-covalent interactions. In such systems, quantitative comparison of geometries is difficult; distilling $3N - 6$ geometric degrees of freedom into a single objective metric is a distinctly non-trivial endeavor. Despite being difficult, this is an important avenue of research. Intermolecular interactions are orders of magnitude weaker than covalent bonds. They involve relatively shallow potential surfaces, and as such, some of these softer degrees of freedom may be useful for the evaluation of electronic structure approaches.

In the context of non-covalently-interacting systems, the systematic evaluation of electronic structure methods with regard to their description of geometries is a largely undeveloped idea. There have been a number of studies in which binding energy curves were generated and studied, though the principle metric of evaluation has in every case been the equilibrium binding energy, not the location of the minimum nor the shape of the curve[115–121]. There have also been a handful of studies in recent years in which hydrogen bond lengths predicted by various methods were compared[122–124]; additionally, there has been a study by Vydrov and Van Voorhis in which the performances of various van der Waals density functionals were evaluated on the basis of their ability to predict intermolecular separation in small CO_2 -containing complexes[102], and a study by Remya and Suresh in which a tremendous number of density functionals were screened on the basis of their abilities to minimize the overall root-mean-square deviation with regard to CCSD geometries of ten small complexes[125]. In the same time frame, there have been countless studies in which

electronic structure approaches have been evaluated solely on the basis of their abilities to reproduce single-point energies. Moreover, the conclusions that can be drawn from these few geometry-based studies are limited, in some cases by confinement to a single interaction motif, in others by the questionable quality of the reference structures, and in all cases by a focus on only small systems.

In this work, we evaluate a variety of electronic structure approximations with regard to their abilities to reproduce complete-basis CCSD(T)-level geometric parameters on a diverse set of systems for which high-quality reference data is readily available. We explore the impacts of interaction type and system size on the performances of the various methods. Moreover, we establish a procedure for obtaining geometric parameters for larger systems, for which multidimensional optimizations with CCSD(T) are prohibitively expensive. We find that although a number of deficiencies of various methods – such as the characteristic overbinding of MP2 – are simply amplified during the transition to larger systems, some are not.

3.2 Computational Methods

We have examined a wide variety of electronic structure methods with respect to their abilities to reproduce coupled-cluster geometries of non-covalently-interacting molecules. Methods examined include HF[107, 108], MP2[5, 81], attenuated MP2[71, 72], simple-scaled MP2 (sMP2, with same-spin and opposite-spin coefficients set to 0.60 for aug-cc-pVDZ, 0.75 for aug-cc-pVTZ)[71, 72, 126], spin-component scaled MP2 (SCS-MP2)[127], scaled opposite-spin MP2 (SOS-MP2)[128], B3LYP[15, 16, 22, 82], PBE[14], M06[25], M06-L[20], M06-2X[25], M11[85], ω B97X[86], ω B97X-D[87], ω B97X-V[23], B97M-V[19], vdW-DF2[13, 38, 88], VV10[14, 40, 88], LC-VV10[40], and Grimme -D2 and -D3 corrections to PBE and B3LYP[30, 31, 129]. For DFT-D3, two different damping functions were utilized: the original zero-damping scheme of Grimme[31], which we refer to simply as DFT-D3, and the damping scheme of Becke and Johnson[34, 61], which we denote as DFT-D3 (BJ).

Throughout this study, all MP2 calculations employ the RI approximation in conjunction with the auxiliary basis sets of Weigend et al.[93], but the RI prefix has been omitted. All calculations were performed with a development version of Q-Chem 4.2[94], with the exception of the calculations on the A21x12 dataset reported in Table 3.1, which were performed with PSI4[130]. Molecular structures were generated with Avogadro[95]. In this work, we have utilized three distinct datasets which can broadly be characterized by the sizes of their constituent systems; due to computational constraints, the manner in which these three classes of systems were treated differs, and will be detailed forthwith.

Small Systems

The first dataset, henceforth referred to as A21, is comprised of the first 21 systems in the A24 dataset[131]. These systems were optimized by Řezáč and coworkers[131] at the

Δ CCSD(T)/CBS level, with a two-point (aug-cc-pVTZ, aug-cc-pVQZ) Helgaker extrapolation of the counterpoise-corrected MP2 correlation energy and a counterpoise-corrected coupled-cluster correction in the aug-cc-pVDZ basis[70, 90, 91, 132]. The systems contained in this dataset are small enough that high-quality geometries are readily available, and so we have performed unconstrained relaxations using all of the methods detailed above, in order to compare the resultant structures to the benchmark-level structures of Řezáč and Hobza[131]. All geometry optimizations were initialized with the relevant Δ CCSD(T)/CBS structure and a Hartree-Fock Hessian generated in the 6-31+G* basis. Tight convergence criteria were employed: the DIIS error was converged to 10^{-8} , the maximum component of the gradient was converged to 1.5×10^{-4} , the maximum atomic displacement was converged to 6×10^{-5} , the energy change of successive optimization cycles was converged to 10^{-9} , and integral thresholds of 10^{-14} were used. All calculations were performed in the aug-cc-pVTZ basis[79, 80]. No symmetry was exploited, though the point group symmetry of each optimized structure matched in every instance the symmetry of the reference structure. Optimizations were performed in Cartesian coordinates. A fine Lebedev integration grid consisting of 99 radial points and 590 angular points was utilized in the computation of the semilocal exchange-correlation components of all of the density functionals; the coarser SG-1 grid was used for nonlocal correlation in the relevant methods, i.e. those involving vdW-DF2 or VV10 nonlocal correlation[133].

A variety of metrics were employed to compare the geometries associated with each method with the benchmark Δ CCSD(T)/CBS geometries. One such metric, the overall root-mean-square deviation (RMSD) is given by

$$RMSD = \sqrt{\frac{\sum_{i=j}^N d(x_i, y_i, z_i, x_j, y_j, z_j)^2}{N}} \quad (3.1)$$

where N is the number of atoms in the structure and $d(x_i, y_i, z_i, x_j, y_j, z_j)$ is the Euclidean distance between the points (x_i, y_i, z_i) and (x_j, y_j, z_j) . Specifically, we define the overall RMSD as the minimum such number, allowing for rigid transformations of the coordinate systems i and j associated with the reference structure and the optimized structure. The overall RMSD thus encompasses both inter- and intramolecular errors in geometry. We also utilized an intermolecular metric, namely the closest point of contact between the two molecules in each system, and various intramolecular metrics, specifically bond length and bond angle root-mean-square errors.

Medium Systems

The second dataset, henceforth referred to as M12, consists of a well-balanced subset of the S66x8 dataset of Řezáč et al.[116], as well as the CO₂-benzene complex[41]. Unfortunately, computational constraints have precluded explicit multi-dimensional optimizations of structures of this size at suitably high levels of theory (i.e. CCSD(T)); as a result, we have

utilized cubic interpolation of various single point energies corresponding to rigid displacement along a single intermolecular coordinate. The particular coordinates and displacements used are defined in the original works of Řezáč et al.[116] and Witte et al.[41], and are depicted in Figure 3.1.

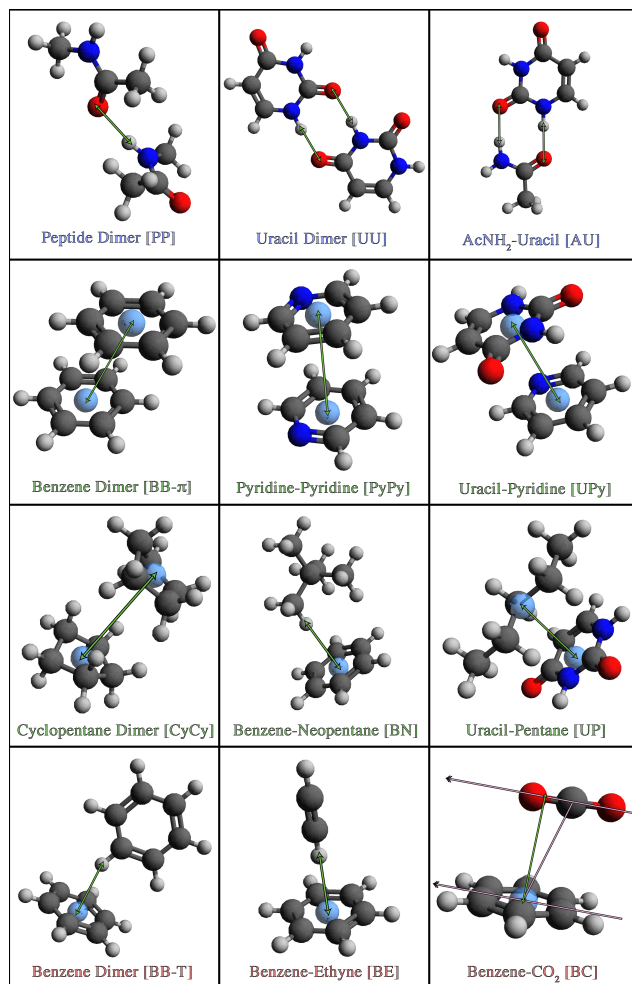


Figure 3.1: Structures of the systems in the M12 dataset. A light blue sphere corresponds to the center of mass of a particular molecule. The type of interaction is indicated by the color of the text: hydrogen-bonded systems are blue, dispersion-bound systems are green, and systems with mixed interactions are red. Green double-headed arrows indicate the relevant intermolecular axis for each system. In brackets, abbreviations for the complexes are introduced.

We have numerically examined the validity of this sort of approach. A representative summary of our findings is provided in Table 3.1. A variety of means of interpolating along the potential energy surface have been examined: a cubic spline, a quartic spline, and a

least-squares-optimized function consisting of a decaying exponential and a power series in r^{-1} , i.e. a simplification of the model of Tang and Toennies[134]. For the M12 set, the differences in interpolated equilibrium distances obtained with these methods are on the order of 0.001 Å.

| Benchmark Quality: Basis Size ^a | | | | Interpolation Type ^b | | | |
|--|-------|--------|--------------|--|-------|-------|--------|
| | UAD | SAD | MAX | | UAD | SAD | MAX |
| MP2: (aQ,a5) vs (aT,aQ) ^c | 0.001 | 0.000 | -0.002 | Quartic vs Cubic | 0.000 | 0.000 | 0.001 |
| Δ CC: aT vs aD ^d | 0.004 | -0.004 | -0.009 | LSQ vs Cubic | 0.001 | 0.000 | -0.004 |
| Benchmark Quality: CO ₂ -Benzene ^e | | | | Validity of Interpolation ^f | | | |
| | r | | BE | | r | | BE |
| MP2/CBS (aT,aQ) + Δ CC (aD) | 3.248 | -2.60 | Optimized | | 3.255 | -2.65 | |
| MP2/CBS (Q,5) + Δ CC (T) ^g | 3.251 | -2.66 | Interpolated | | 3.253 | -2.64 | |
| Difference | 0.003 | 0.06 | Difference | | 0.001 | -0.01 | |

^aResults pertain to the A21x12 set.

^bResults pertain to the M12 set.

^cExtrapolation to the CBS limit was performed according to the scheme of Helgaker et al.[70, 132]. aT, aQ, and a5 denote aug-cc-pVTZ, aug-cc-pVQZ, and aug-cc-pV5Z, respectively.

^dRI-CCSD(T) correction to MP2 correlation energy[90]. aD and aT denote aug-cc-pVDZ and aug-cc-pVTZ, respectively.

^eEquilibrium distances r and binding energies BE were interpolated with a cubic spline.

^fCase study on the CO₂-benzene complex using VV10.

^gT, Q, and 5 denote cc-pVTZ, cc-pVQZ, and cc-pV5Z, respectively.

Table 3.1: Justification of methods. We have considered the quality of the reference data with regard to basis set size on both small (top left) and moderately-sized (bottom left) systems, the impact of the specific form of function used for interpolation (top right), and the difference between interpolation and constrained optimization (bottom right). UAD, SAD, and MAX denote respectively the unsigned, signed, and maximum average difference in interpolated equilibrium separation across the indicated dataset, in units of Å. r and BE denote the equilibrium distance (Å) and binding energy (kcal mol⁻¹), respectively.

We have also investigated whether interpolation is a suitable substitute for explicitly performing a constrained optimization along the relevant axis. There appears to be no difference between the two approaches; the results of a case study of VV10 on the CO₂-benzene complex are provided in Table 3.1. Further proof is provided by a recent study on the parallel-displaced benzene dimer in which both the in-plane shift and interplane spacing were optimized at the CCSD(T)/aug-cc-pVTZ level of theory[135]. The optimized parameters correspond to a center-of-mass separation of 3.84 Å, which compares quite favorably with the value of 3.86 Å obtained by interpolating along the Δ CCSD(T)/CBS binding curve of

the M12 set.

As a final justification of our methodology, we have addressed the quality of our benchmarks. For the A21 and M12 sets, the reference data is $\Delta\text{CCSD(T)}/\text{CBS}$, with a two-point (aug-cc-pVTZ, aug-cc-pVQZ) extrapolation of the MP2 correlation energy and a correction for higher-order correlation effects in the aug-cc-pVDZ basis[70, 90, 132]. Specifically, we have investigated the impact of utilizing larger basis sets for each of these components on interpolated equilibrium distances in the A21x12 dataset – a set constructed in a manner analogous to the S66x8 set – wherein we rigidly have scaled the center-of-mass separation of each system in the A21 set by a factor of 0.9 to 2.0, in increments of 0.1. The A21 set, along with the relevant intermolecular axis utilized for the generation of the A21x12 set, is depicted in Figure 3.2. As is evident from Table 3.1, the reference data is indeed sufficiently high-quality: using larger bases changes the interpolated equilibrium intermolecular distances by less than 0.01 Å. As a case study of whether such a choice of basis sets is sufficient for larger systems, we have examined the CO₂-benzene complex at two different levels of theory, and found that our reference data is indeed sufficiently converged with respect to basis set size, as regards both equilibrium geometry and binding energy.



Figure 3.2: Structures of the systems in the A21x12 dataset. The type of interaction is indicated by the color of the text: hydrogen-bonded systems are blue, dispersion-bound systems are green, and systems with mixed interactions are red. Green double-headed arrows connect the centers of masses of the two molecules in each system, and hence indicate the relevant intermolecular axis for each system.

As in the case of the A21 set, our calculations on the M12 set utilize a (99,590) Lebedev integration grid for semilocal components of density functionals, with the SG-1 grid being used for nonlocal correlation. Furthermore, integral thresholds of 10^{-14} were used, the DIIS error was converged to 10^{-8} , and no symmetry was exploited. All calculations were performed in the aug-cc-pVTZ basis. For the MP2 calculations, the frozen core approximation was

employed[136]. MP2 results were corrected for basis set superposition error (BSSE) a la Boys and Bernardi[91].

Large System

In an attempt to probe the transferability of observed trends to even larger systems, we have treated the coronene dimer in an analogous manner to how the M12 set was treated. The reference geometry for the coronene dimer, determined at the QCISD(T)/h-aug-cc-pVDZ level by Janowski et al.[137], is depicted in Figure 3.3. This geometry corresponds to an interplane spacing of 3.458 Å, with an in-plane shift of 1.553 Å. We constructed a potential energy surface for each method with a series of seven of single-point energy calculations; in so doing, we sampled interplane separations from 3.008 Å to 3.908 Å, in increments of 0.158 Å, holding the in-plane shift and all intramolecular parameters constant. To aid convergence, densities determined at the LDA level were used as a starting point for all jobs, and the criterion for determining wavefunction convergence was lowered to 10^{-6} . All calculations were performed in the aug-cc-pVDZ basis, and all methods were corrected for BSSE in the usual manner[91]. A (99,590) Lebedev grid was used for the evaluation of semilocal components of density functionals, the SG-1 grid was used for the evaluation of nonlocal correlation, integral thresholds were set to 10^{-14} , and no symmetry was exploited.

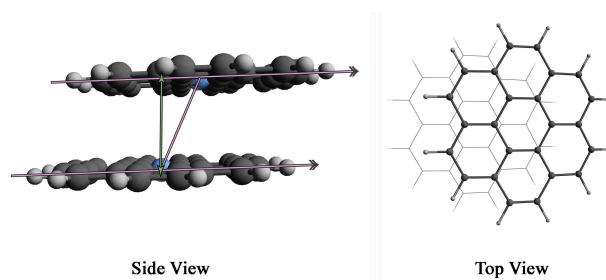


Figure 3.3: Structure of the coronene dimer. Light blue spheres correspond to the centers of masses of the coronene monomers. A green double-headed arrow indicates the relevant intermolecular axis.

3.3 Results and Discussion

Small Systems

A summary of results pertaining to the A21 dataset is provided in Table 3.2 and Figure 3.4. It is evident from Table 3.2 that of the methods examined, ω B97X-V is the top performer, treating the various classes of interactions equally well. Moreover, the similarity between the overall root-mean-square deviation (RMSD) and overall weighted root-mean-square deviation (wRMSD) indicates that ω B97X-V performs equally well for both weak

and strong interactions, as opposed to a method such as B3LYP, which performs disproportionately well on stronger interactions, i.e. hydrogen bonds. Within the A21 dataset, the various modifications of MP2 perform reasonably well for geometry optimizations, with attenuated MP2 (attMP2) slightly outperforming other versions of MP2. MP2 systematically underestimates intermolecular distances, as evidenced in Figure 3.4, which can be attributed somewhat to BSSE, but primarily to the overbinding endemic to the method. This overbinding and concomitant underestimation is alleviated somewhat by the attenuation of the Coulomb operator; some systems are actually underbound by attMP2, which suggests the addition of a long-range dispersion correction could be profitable. This has been explored by Huang et al., who found that attenuation of dispersion-corrected MP2 can indeed improve the description of intermolecular interactions, though at the possible cost of a poorer picture of intramolecular interactions[138]. Simple-scaling of the MP2 correlation energy, on the other hand, leads to systematic overestimation of intermolecular separation. This is an artifact of the fact that the scaling coefficient was optimized with respect to errors in interaction energies in the S66 dataset, and is hence too small for the purposes of this dataset. Scaling of individual components of the MP2 correlation energy (SCS-MP2 and SOS-MP2) is similarly underwhelming here.

| Hydrogen-Bonded | | Mixed Interactions | | Dispersion-Bound | | All | | | |
|-----------------|----------|--------------------|----------|------------------|----------|-----------------|----------|-----------------|-----------|
| Method | RMSD (Å) | Method | RMSD (Å) | Method | RMSD (Å) | Method | RMSD (Å) | Method | wRMSD (Å) |
| MP2 | 0.010 | ω B97X-V | 0.016 | attMP2 | 0.009 | ω B97X-V | 0.014 | ω B97X-V | 0.014 |
| ω B97X-V | 0.011 | LC-VV10 | 0.024 | MP2 | 0.012 | LC-VV10 | 0.028 | attMP2 | 0.022 |
| attMP2 | 0.012 | vdW-DF2 | 0.035 | ω B97X-V | 0.013 | attMP2 | 0.035 | MP2 | 0.022 |
| SCS-MP2 | 0.013 | ω B97X-D | 0.036 | B97M-V | 0.020 | vdW-DF2 | 0.036 | LC-VV10 | 0.027 |
| sMP2 | 0.013 | sMP2 | 0.039 | B3LYP-D3 | 0.021 | B97M-V | 0.037 | B97M-V | 0.027 |
| B3LYP | 0.014 | M11 | 0.050 | B3LYP-D3 (BJ) | 0.024 | ω B97X-D | 0.037 | sMP2 | 0.029 |
| B97M-V | 0.017 | SOS-MP2 | 0.053 | VV10 | 0.042 | MP2 | 0.038 | SCS-MP2 | 0.029 |
| M11 | 0.018 | B97M-V | 0.057 | vdW-DF2 | 0.042 | sMP2 | 0.042 | ω B97X-D | 0.031 |
| M06-2X | 0.019 | attMP2 | 0.062 | LC-VV10 | 0.043 | M11 | 0.045 | B3LYP-D3 | 0.032 |
| B3LYP-D3 | 0.020 | SCS-MP2 | 0.065 | SCS-MP2 | 0.048 | SCS-MP2 | 0.048 | B3LYP-D3 (BJ) | 0.034 |
| B3LYP-D3 (BJ) | 0.020 | PBE | 0.065 | ω B97X-D | 0.050 | B3LYP-D3 | 0.053 | vdW-DF2 | 0.037 |
| LC-VV10 | 0.021 | MP2 | 0.068 | M06-L | 0.054 | B3LYP-D3 (BJ) | 0.054 | M11 | 0.038 |
| SOS-MP2 | 0.023 | M06 | 0.071 | ω B97X | 0.057 | SOS-MP2 | 0.055 | ω B97X | 0.041 |
| ω B97X-D | 0.025 | B3LYP-D3 | 0.089 | M06 | 0.059 | M06 | 0.058 | SOS-MP2 | 0.042 |
| PBE | 0.027 | B3LYP-D3 (BJ) | 0.089 | M11 | 0.061 | VV10 | 0.067 | M06 | 0.045 |
| M06 | 0.029 | M06-L | 0.092 | PBE-D3 | 0.066 | M06-L | 0.067 | M06-2X | 0.049 |
| ω B97X | 0.031 | PBE-D3 (BJ) | 0.093 | PBE-D3 (BJ) | 0.067 | ω B97X | 0.069 | M06-L | 0.051 |
| PBE-D3 | 0.031 | ω B97X | 0.096 | B3LYP-D | 0.070 | PBE-D3 (BJ) | 0.072 | B3LYP-D | 0.052 |
| vdW-DF2 | 0.032 | VV10 | 0.097 | M06-2X | 0.070 | PBE-D3 | 0.074 | PBE-D3 | 0.056 |
| M06-L | 0.033 | PBE-D3 | 0.101 | sMP2 | 0.072 | M06-2X | 0.084 | VV10 | 0.060 |
| PBE-D3 (BJ) | 0.035 | PBE-D | 0.126 | PBE-D | 0.083 | B3LYP-D | 0.090 | PBE-D3 (BJ) | 0.061 |
| B3LYP-D | 0.035 | M06-2X | 0.126 | SOS-MP2 | 0.084 | PBE-D | 0.095 | PBE | 0.062 |
| VV10 | 0.038 | B3LYP-D | 0.130 | PBE | 0.205 | PBE | 0.096 | PBE-D | 0.074 |
| PBE-D | 0.047 | B3LYP | 0.152 | HF | 0.639 | HF | 0.330 | B3LYP | 0.138 |
| HF | 0.079 | HF | 0.270 | B3LYP | 1.128 | B3LYP | 0.398 | HF | 0.184 |

Table 3.2: Average overall root-mean-square deviations (RMSD) and weighted root-mean-square deviations (wRMSD) in geometries of complexes in various subsets of the A21 Dataset. Weights were determined from the relative binding energies of each complex. Within each subset, the methods are listed in order of ascending RMSD or wRMSD. All calculations were performed in the aug-cc-pVTZ basis with tight convergence criteria. Calculations involving density functionals utilized a (99,590) grid.

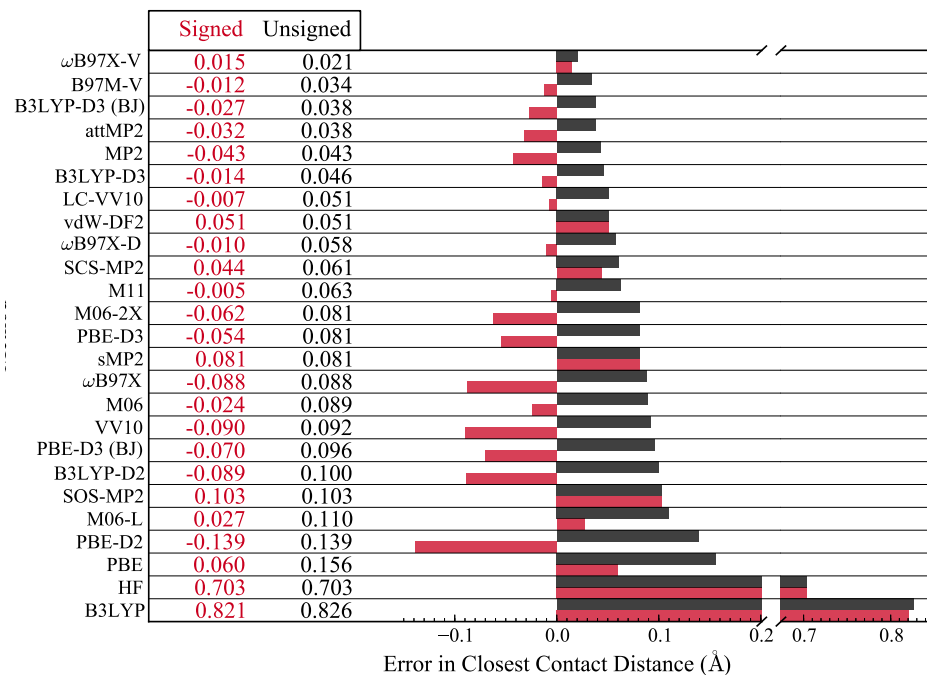


Figure 3.4: Average signed and unsigned errors in closest-contact distances in geometries of complexes from the A21 Dataset. The methods are listed in order of ascending unsigned error. All calculations were performed in the aug-cc-pVTZ basis with tight convergence criteria. Calculations involving density functionals utilized a (99,590) grid.

Most of the density functionals examined provide rather mediocre geometries for this dataset; the descriptions of mixed and dispersion-dominated interactions in particular leave much to be desired. There do appear to be trends in the performances of the various types of functionals, however. Functionals incorporating some amount of exact exchange tend to outperform those with approximate exchange kernels, particularly on hydrogen-bonded systems. Going a step further, long-range corrected hybrids appear to offer a superior description to global hybrids or functionals with no exact exchange. The best functionals incorporate range-separated exchange in conjunction with some sort of long-range dispersion correction. The nature of the dispersion tail is particularly important: the standard DFT-D2 treatment of Grimme yields complexes in which the closest-contact distance is vastly underestimated; in fact, PBE-D2 actually performs worse than PBE, a consequence of the fact that PBE exchange alone often overbinds in the context of non-covalent interactions[139]. The more repulsive DFT-D3 and DFT-D3 (BJ) variants are thus better suited for these smaller systems: B3LYP-D3 and PBE-D3 exhibit less severe underestimation of intermolecular separation than their -D2 counterparts, regardless of whether zero-damping or BJ-damping is employed, as evidenced in Figure 3.4. The differences between -D2 and -D3

can be attributed to both differences in their respective C_6 coefficients as well as differences in the damping functions of the two methods. Among the -D3 variants studied, both the zero-damping and BJ-damping schemes yield similar results. A variant employing Wu-Yang damping[140] – the function of choice in the original -D2 prescription – in conjunction with -D3 C_6 coefficients is given in Appendix A, and suggests the difference between -D2 and -D3 is in this case primarily due to the starkly different Fermi-type damping function.

The performances of the various nonlocal functionals are somewhat less intuitive. The standard exchange-matching issues of vdW-DF2 are manifest in the systematically too-large closest-contact distance, as illustrated in Figure 3.4. The behaviors of the various functionals with a VV10 tail, however, are less predictable: variants with range-separated exchange, namely ω B97X-V and LC-VV10, yield structures very similar to the Δ CCSD(T) benchmarks, whereas VV10 in its original iteration is somewhat lackluster in its description of the A21 set. This cannot be understood simply to be a result of a deficiency in any single component of the VV10 functional. As is apparent in Figure 3.4, VV10 underestimates intermolecular separations, whereas LC-VV10 on average overestimates them, despite the fact that the rPW86 exchange incorporated in VV10 is generally more repulsive than the (short-range) PBE exchange of LC-VV10. There is clearly a subtle interplay between the exchange components and the nonlocal tail at work. The fact that the parameters of the nonlocal tail were optimized on different datasets further complicates this issue.

One thing that bears mentioning here is that although we have distilled the comparison of geometries into two numbers – the overall RMSD and the error in closest-contact distance – these two metrics alone only tell part of the story. The overall RMSD encompasses all of the discrepancies between a structure associated with a particular method and the reference structure; what is lacking, however, is a breakdown of whence these disparities arise. The error in closest-contact distance, on the other hand, is a much more focused metric; it is primarily a measure of intermolecular separation, though it can be obfuscated by symmetry-preserving rotations, provided such transformations exist. A variety of other possible metrics exist. For instance, two simple intramolecular metrics might be bond length or bond angle RMS errors; these have been tabulated for the methods examined in the A21 dataset, and can be found in Appendix A. For the A21 set and the methods examined, these errors are an order of magnitude smaller than the overall RMSD; moreover, the distributions associated with these errors are much narrower than the distribution associated with either closest-contact error or overall RMSD, and hence the utility of such metrics is limited. Perhaps the most interesting story told by this supplementary data is the lack of a difference between the base functionals and those incorporating a -D2 or -D3 dispersion correction: addition of a simple empirical dispersion correction does not significantly impact intramolecular parameters in small molecules. This may not be the case, however, for large or extended molecules, particularly those involving significant non-covalent intramolecular interactions.

Medium Systems

It is evident from Table 3.2 and Figure 3.4 that the primary source of overall deviation in geometries in the chosen methods in the A21 set is the intermolecular separation. Thus, in systems where performing unconstrained optimizations are infeasible, it is possible to probe geometric differences solely on the basis of a single intermolecular coordinate. Moreover, we established earlier (see Table 3.1) that interpolation based on a series of intelligently-chosen points is, to a small degree of uncertainty (on the order of 0.01 Å), equivalent to explicitly performing a constrained optimization along the relevant coordinate. Examination of Figure 3.5 further supports this notion; with few exceptions, the methods investigated in this study yield well-behaved binding energy curves, and so it is no surprise that any sort of reasonable choice of interpolation scheme would identify the same minimum. Such is the basis for our treatment of the M12 set of medium-sized molecular systems.

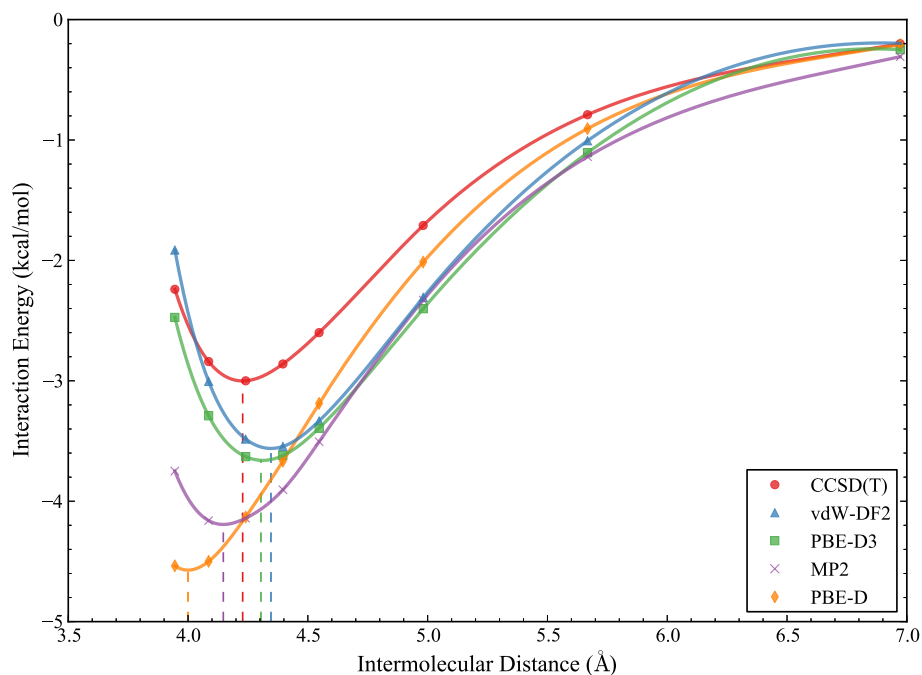


Figure 3.5: Selection of binding energy curves for the cyclopentane dimer (CyCy in the M12 dataset). Equilibrium separations associated with each method are denoted by vertical dashed lines, and were determined via interpolation with a cubic spline. All calculations were performed in the aug-cc-pVTZ basis. Calculations involving density functionals utilized a (99,590) grid.

A summary of results pertaining to the M12 dataset are provided in Figures 3.6 and 3.7. In light of the small uncertainty associated with the interpolation, any difference in intermolecular separation larger than 0.03 Å can be safely deemed significant. For certain cases

where the minimum associated with a particular method lies outside of the range explored, the second largest (or smallest, as appropriate) separation examined is reported, thereby providing a lower bound for the error. As a consequence of this and the general problem of interpolating a minimum from a flat surface, those errors associated with particularly under-binding methods, namely PBE, HF, and B3LYP on systems involving dispersive interactions, can only be interpreted in a qualitative sense.

| Hydrogen-Bonded | | | Dispersion (π - π) | | | Dispersion (Other) | | | Mixed | | | All | | | |
|-----------------|-------|------|------------------------------|-------|------|--------------------|-------|------|-----------------|-------|------|-----------------|-------|------|-------|
| Method | ASE | AUE | Method | ASE | AUE | Method | ASE | AUE | Method | ASE | AUE | Method | ASE | AUE | MAX |
| ω B97X | 0.00 | 0.00 | sMP2 | -0.01 | 0.02 | B3LYP-D3 (BJ) | 0.01 | 0.01 | B3LYP-D3 (BJ) | 0.01 | 0.01 | B3LYP-D3 (BJ) | 0.00 | 0.02 | 0.03 |
| M06-2X | 0.00 | 0.00 | SOS-MP2 | 0.02 | 0.02 | B97M-V | -0.01 | 0.01 | VV10 | 0.01 | 0.01 | B97M-V | 0.01 | 0.02 | 0.04 |
| MP2 (CP) | 0.00 | 0.00 | LC-VV10 | -0.02 | 0.02 | B3LYP-D3 | -0.02 | 0.02 | B97M-V | 0.02 | 0.02 | VV10 | 0.00 | 0.02 | -0.05 |
| ω B97X-V | 0.01 | 0.01 | ω B97X-D | 0.01 | 0.02 | VV10 | -0.01 | 0.02 | sMP2 | 0.00 | 0.02 | B3LYP-D3 | 0.01 | 0.03 | 0.05 |
| M06-L | -0.01 | 0.01 | B3LYP-D3 (BJ) | 0.00 | 0.02 | ω B97X | 0.01 | 0.02 | ω B97X | 0.01 | 0.02 | LC-VV10 | -0.03 | 0.03 | -0.06 |
| M11 | 0.01 | 0.01 | M06-L | -0.03 | 0.03 | LC-VV10 | -0.03 | 0.03 | B3LYP-D3 | 0.03 | 0.03 | sMP2 | 0.02 | 0.03 | 0.13 |
| B97M-V | 0.01 | 0.01 | B97M-V | 0.03 | 0.03 | MP2 (CP) | -0.03 | 0.03 | SOS-MP2 | 0.03 | 0.03 | ω B97X-D | -0.01 | 0.03 | -0.11 |
| M06 | 0.01 | 0.01 | VV10 | 0.03 | 0.03 | SCS-MP2 | 0.02 | 0.05 | ω B97X-D | 0.03 | 0.03 | ω B97X | 0.03 | 0.03 | 0.10 |
| B3LYP-D3 | -0.01 | 0.01 | M06 | 0.04 | 0.04 | ω B97X-V | 0.05 | 0.05 | SCS-MP2 | -0.04 | 0.04 | M06 | 0.01 | 0.03 | -0.10 |
| sMP2 | 0.01 | 0.01 | PBE-D2 | -0.04 | 0.04 | M06 | -0.05 | 0.05 | LC-VV10 | -0.04 | 0.04 | M06-L | -0.03 | 0.03 | -0.08 |
| ω B97X-D | -0.01 | 0.01 | B3LYP-D3 | 0.05 | 0.05 | attMP2 | -0.06 | 0.06 | M06-L | 0.00 | 0.04 | ω B97X-V | 0.04 | 0.04 | 0.08 |
| PBE-D3 | -0.01 | 0.01 | SCS-MP2 | -0.07 | 0.07 | PBE-D3 (BJ) | 0.06 | 0.06 | ω B97X-V | 0.04 | 0.04 | SCS-MP2 | -0.01 | 0.04 | 0.10 |
| VV10 | -0.01 | 0.01 | PBE-D3 (BJ) | 0.07 | 0.07 | sMP2 | 0.07 | 0.07 | M06 | 0.04 | 0.04 | SOS-MP2 | 0.04 | 0.04 | 0.16 |
| attMP2 | -0.02 | 0.02 | B3LYP-D2 | -0.07 | 0.07 | M06-L | -0.07 | 0.07 | PBE-D3 (BJ) | 0.05 | 0.05 | PBE-D3 (BJ) | 0.04 | 0.05 | 0.10 |
| B3LYP-D3 (BJ) | -0.02 | 0.02 | ω B97X-V | 0.07 | 0.07 | ω B97X-D | -0.07 | 0.07 | MP2 (CP) | -0.06 | 0.06 | MP2 (CP) | -0.06 | 0.06 | -0.17 |
| SCS-MP2 | 0.02 | 0.02 | M11 | -0.07 | 0.07 | PBE-D3 | 0.09 | 0.09 | PBE-D2 | -0.08 | 0.08 | M11 | -0.06 | 0.07 | -0.13 |
| PBE-D3 (BJ) | -0.02 | 0.02 | ω B97X | 0.09 | 0.09 | SOS-MP2 | 0.09 | 0.09 | M11 | -0.09 | 0.09 | attMP2 | -0.07 | 0.07 | -0.14 |
| PBE | 0.01 | 0.02 | M06-2X | -0.09 | 0.09 | M11 | -0.11 | 0.11 | PBE-D3 | 0.09 | 0.09 | M06-2X | -0.08 | 0.08 | -0.13 |
| LC-VV10 | -0.02 | 0.02 | attMP2 | -0.12 | 0.12 | M06-2X | -0.11 | 0.11 | M06-2X | -0.09 | 0.09 | PBE-D2 | -0.08 | 0.08 | -0.23 |
| MP2 | -0.03 | 0.03 | PBE-D3 | 0.12 | 0.12 | MP2 | -0.12 | 0.12 | attMP2 | -0.10 | 0.10 | PBE-D3 | 0.07 | 0.08 | 0.13 |
| B3LYP-D2 | -0.03 | 0.03 | MP2 (CP) | -0.14 | 0.14 | vdW-DF2 | 0.12 | 0.12 | B3LYP-D2 | -0.10 | 0.10 | B3LYP-D2 | -0.10 | 0.10 | -0.28 |
| PBE-D2 | -0.04 | 0.04 | vdW-DF2 | 0.14 | 0.14 | PBE-D2 | -0.16 | 0.16 | vdW-DF2 | 0.14 | 0.14 | vdW-DF2 | 0.12 | 0.12 | 0.16 |
| B3LYP | 0.04 | 0.04 | MP2 | -0.20 | 0.20 | B3LYP-D2 | -0.21 | 0.21 | MP2 | -0.14 | 0.14 | MP2 | -0.12 | 0.12 | -0.25 |
| SOS-MP2 | 0.04 | 0.04 | PBE | 0.63 | 0.63 | PBE | 0.51 | 0.51 | PBE | 0.34 | 0.34 | PBE | 0.38 | 0.38 | 0.86 |
| vdW-DF2 | 0.07 | 0.07 | HF | 1.30 | 1.30 | HF | 1.33 | 1.33 | B3LYP | 0.62 | 0.62 | HF | 0.89 | 0.89 | 1.78 |
| HF | 0.14 | 0.14 | B3LYP | 1.35 | 1.35 | B3LYP | 1.57 | 1.57 | HF | 0.77 | 0.77 | B3LYP | 0.89 | 0.89 | 1.78 |

Figure 3.6: Average signed (ASE) and unsigned (AUE) errors (in units of \AA) in interpolated equilibrium intermolecular separations in geometries of complexes in various subsets of the M12 dataset. Maximum error for each method (MAX) is given in last column (in units of \AA). Within each subset, methods are sorted in order of ascending AUE. Signed values are colored as such: positive errors are blue, negative errors are red, and the tint of the color correlates with the magnitude of the error. Equilibria correspond to the interpolated (cubic spline) minima of the binding energy curves. All calculations were performed in the aug-cc-pVTZ basis. Calculations involving density functionals utilized a (99,590) grid.

Due to the uncertainty associated with each interpolated equilibrium separation, it is difficult to draw the same sorts of conclusions for the M12 set as we did for the A21 set. Nevertheless, a few trends are apparent. Among the MP2 methods, standard MP2 is the worst performer, underestimating the intermolecular separation in every system and providing geometries worse than any of the standard density functional approaches (with the exception of PBE and B3LYP). The treatment of systems involving $\pi - \pi$ interactions is particularly bad – this is a manifestation of BSSE and the general unsuitability of uncoupled HF polarizabilities (and hence C_6 coefficients) for describing such systems[96]. In a larger basis, where BSSE is reduced, this underestimation of intermolecular distance is

somewhat less drastic, though still substantial, as evidenced by the counterpoise-corrected (CP) MP2 results in Figures 3.6 and 3.7. Attenuation of the Coulomb operator (attMP2) addresses the underestimation to a similar extent as CP-correction, though at a fraction of the cost. Simple-scaling of the MP2 correlation energy (sMP2) yields better agreement with the coupled-cluster benchmarks; this can be at least in part attributed to the optimization of the scaling coefficient on the S66 dataset, a set which contains eleven of the twelve systems in M12. However, the dangers of using a single scaling coefficient for a variety of systems are hinted at by the large error for the method on the cyclopentane dimer, as well as the lackluster performance of sMP2 with regard to the A21 set. Separate scaling of the different components of the MP2 correlation energy (SOS-MP2 and SCS-MP2) is generally inferior to simple-scaling.

Among the density functionals examined, most of the top performers incorporate some form of long-range dispersion correction. Specifically, those functionals with VV10 nonlocal correlation reproduce $\Delta\text{CCSD(T)}/\text{CBS}$ equilibrium separations well: B97M-V in particular is consistently quite accurate. Among the functionals lacking any form of long-range van der Waals correction, ωB97X , M06, and M06-L stand out. Their impressive performances can be at least somewhat attributed to the significant overlap between the systems examined here and the systems in their respective training sets. Interestingly enough, these methods actually perform better on these larger systems than on the small systems in the A21 set (cf. Figures 3.4 and 3.6). Thus, in going from small systems (A21) to medium systems (M12), we do not see an unambiguous amplification of deficiencies. The systematic underestimation of intermolecular separation associated with MP2 and its attenuated variant, the overestimation of vdW-DF2 and $\omega\text{B97X-V}$, and the overestimation of the DFT-D3 methods relative to their DFT-D2 counterparts are significantly magnified by the growth in system size, but the relative performances of certain other methods reverse completely. For instance, VV10 offers an excellent description of every system in the M12 set despite being one of the worst methods with respect to the A21 set.

There is one more interesting point to be made that is illustrated clearly by Figure 3.5: overbinding is not synonymous with underestimation of intermolecular separation, i.e. at least for some methods, horizontal and vertical motion of the binding energy curve associated with a given system are often decoupled. Throughout this article, we have made a point of maintaining this distinction by referring to methods as “overestimating intermolecular separations,” rather than by simply calling them “underbinding.” In general, we might expect that a method that yields a too-long intermolecular separation would be underbinding, but it is clear from Figure 3.5 that this is emphatically not the case. For instance, in the case of the cyclopentane dimer, PBE-D3 and vdW-DF2 both overestimate equilibrium intermolecular separations despite being overbinding with respect to energy. In a similar manner, M06 underbinds every system in M12 with respect to energy yet predicts a compressed geometry for half of the dispersion-bound systems. This highlights a serious deficiency in the standard approach of comparing methods solely on the basis of binding energies. This shortcoming is further amplified by the practice of comparing energies calculated with different methods on the same geometry: the binding energies of the cyclopentane dimer as predicted by PBE-

| Method | Hydrogen-Bonded | | | Dispersion (π - π) | | | Dispersion (Other) | | | Mixed | | | All |
|-----------------|-----------------|-------|-------|------------------------------|-------|-------|--------------------|-------|-------|-------|-------|-------|------|
| | PP | UU | AU | BB- π | PyPy | UPy | CyCy | BN | UP | BB-T | BE | BC | AUE |
| B3LYP-D3 (BJ) | -0.02 | -0.02 | -0.02 | -0.03 | -0.01 | 0.03 | -0.01 | 0.00 | 0.03 | 0.01 | 0.00 | 0.02 | 0.02 |
| B97M-V | 0.01 | 0.01 | 0.01 | 0.03 | 0.04 | 0.02 | -0.02 | 0.00 | -0.02 | 0.03 | 0.01 | 0.01 | 0.02 |
| VV10 | -0.02 | -0.01 | -0.01 | 0.03 | 0.04 | 0.04 | -0.05 | 0.00 | 0.01 | 0.02 | 0.00 | 0.01 | 0.02 |
| B3LYP-D3 | -0.01 | -0.01 | -0.01 | 0.04 | 0.04 | 0.05 | 0.00 | -0.03 | -0.02 | 0.01 | 0.02 | 0.04 | 0.03 |
| LC-VV10 | -0.02 | -0.03 | -0.03 | -0.01 | -0.02 | -0.03 | -0.01 | -0.04 | -0.03 | -0.03 | -0.06 | -0.02 | 0.03 |
| sMP2 | 0.02 | 0.01 | 0.01 | -0.02 | -0.01 | 0.01 | 0.13 | 0.03 | 0.05 | -0.01 | -0.02 | 0.04 | 0.03 |
| ω B97X-D | -0.02 | -0.01 | -0.01 | -0.02 | 0.00 | 0.04 | -0.11 | -0.06 | -0.03 | 0.00 | 0.02 | 0.07 | 0.03 |
| ω B97X | 0.00 | 0.00 | 0.00 | 0.10 | 0.09 | 0.07 | -0.02 | 0.03 | 0.02 | 0.04 | -0.01 | 0.02 | 0.03 |
| M06 | 0.00 | 0.01 | 0.01 | 0.03 | 0.05 | 0.04 | -0.10 | -0.02 | -0.03 | 0.04 | 0.05 | 0.04 | 0.03 |
| M06-L | 0.00 | -0.01 | 0.00 | -0.02 | -0.03 | -0.03 | -0.06 | -0.08 | -0.07 | 0.00 | -0.06 | 0.05 | 0.03 |
| ω B97X-V | 0.01 | 0.00 | 0.00 | 0.07 | 0.08 | 0.06 | 0.05 | 0.04 | 0.05 | 0.05 | 0.02 | 0.05 | 0.04 |
| SCS-MP2 | 0.02 | 0.02 | 0.02 | -0.09 | -0.08 | -0.03 | 0.10 | -0.04 | 0.01 | -0.06 | -0.05 | 0.00 | 0.04 |
| SOS-MP2 | 0.05 | 0.04 | 0.04 | 0.00 | 0.01 | 0.04 | 0.16 | 0.04 | 0.08 | 0.01 | 0.01 | 0.07 | 0.04 |
| PBE-D3 (BJ) | -0.01 | -0.03 | -0.03 | 0.05 | 0.06 | 0.09 | 0.05 | 0.04 | 0.10 | 0.05 | 0.02 | 0.09 | 0.05 |
| MP2 (CP) | 0.01 | 0.00 | 0.00 | -0.17 | -0.16 | -0.08 | 0.00 | -0.06 | -0.02 | -0.09 | -0.06 | -0.05 | 0.06 |
| M11 | 0.00 | 0.01 | 0.01 | -0.06 | -0.07 | -0.08 | -0.11 | -0.08 | -0.13 | -0.06 | -0.10 | -0.10 | 0.07 |
| attMP2 | -0.02 | -0.02 | -0.02 | -0.14 | -0.13 | -0.09 | -0.01 | -0.09 | -0.07 | -0.10 | -0.11 | -0.07 | 0.07 |
| M06-2X | -0.01 | 0.00 | -0.01 | -0.10 | -0.09 | -0.09 | -0.12 | -0.09 | -0.13 | -0.07 | -0.08 | -0.13 | 0.08 |
| PBE-D2 | -0.04 | -0.03 | -0.03 | -0.04 | -0.05 | -0.04 | -0.23 | -0.13 | -0.12 | -0.10 | -0.11 | -0.02 | 0.08 |
| PBE-D3 | 0.00 | -0.02 | -0.02 | 0.12 | 0.12 | 0.13 | 0.08 | 0.06 | 0.12 | 0.09 | 0.07 | 0.12 | 0.08 |
| B3LYP-D2 | -0.05 | -0.02 | -0.02 | -0.07 | -0.08 | -0.06 | -0.28 | -0.17 | -0.18 | -0.13 | -0.12 | -0.05 | 0.10 |
| vdW-DF2 | 0.07 | 0.07 | 0.07 | 0.14 | 0.13 | 0.16 | 0.12 | 0.12 | 0.13 | 0.16 | 0.14 | 0.11 | 0.12 |
| MP2 | -0.03 | -0.02 | -0.02 | -0.25 | -0.22 | -0.14 | -0.08 | -0.16 | -0.11 | -0.18 | -0.14 | -0.11 | 0.12 |
| PBE | 0.06 | -0.01 | -0.01 | 0.86 | 0.59 | 0.45 | 0.48 | 0.51 | 0.54 | 0.43 | 0.21 | 0.38 | 0.38 |
| HF | 0.22 | 0.10 | 0.09 | 1.78 | 1.37 | 0.75 | 1.44 | 1.26 | 1.30 | 1.00 | 0.51 | 0.81 | 0.89 |
| B3LYP | 0.09 | 0.01 | 0.01 | 1.78 | 1.67 | 0.62 | 1.44 | 1.58 | 1.68 | 0.85 | 0.37 | 0.64 | 0.89 |

Figure 3.7: Errors in interpolated equilibrium intermolecular separations (in units of \AA) in geometries of complexes of the M12 dataset. The last column, the average unsigned error (AUE) across all systems (in units of \AA), represents the metric by which the methods are sorted. Signed values are colored as such: positive errors are blue, negative errors are red, and the tint of the color correlates with the magnitude of the error. Equilibria correspond to the interpolated (cubic spline) minima of the binding energy curves. All calculations were performed in the aug-cc-pVTZ basis. Calculations involving density functionals utilized a (99,590) grid. For the abbreviations, see Figure 3.1.

D3 and MP2 at the $\Delta\text{CCSD(T)}$ minimum are indistinguishable, though at their respective minima they differ by nearly 0.2 kcal/mol.

Large System

Interpolated equilibrium interplane separations and binding energies predicted by each of the methods for the parallel-displaced coronene dimer are provided in Figure 3.8. Note that

the values of 3.91 Å and 0 kcal mol⁻¹ reported for PBE, B3LYP, and HF simply indicate that these methods were all repulsive at the maximum separation examined, 3.908 Å. Errors are expressed relative to current best guesses of 3.458 Å and 23.45 kcal mol⁻¹[19, 137]. It is worth mentioning, however, that these reference values are not nearly as ironclad as those used for the A21 and M12 sets: the interplane separation corresponds to QCISD(T)/h-aug-cc-pVDZ, i.e. cc-pVDZ on hydrogens and alternating carbons, and aug-cc-pVDZ on other carbons[137], and the binding energy is given by a counterpoise-corrected MP2/CBS (aTZ,aQZ Helgaker extrapolation)[70, 132] energy corrected for higher-order correlation in this same small basis[19]. Other notable reports for binding energy include -19.98 kcal mol⁻¹ and -24.36 kcal mol⁻¹, values which were obtained using different prescriptions for incorporating an MP2 energy correction and a different small basis for the Δ QCISD(T) correction[137, 141]. As a result of this general uncertainty in the “true” equilibrium interplane separation and binding energy, any discussion of our results for the coronene dimer can only be semi-quantitative.

It is apparent from Figure 3.8 that some of the general trends observed on the M12 set transfer reasonably well to the case of the coronene dimer: for instance, most methods involving VV10 nonlocal correlation perform well; DFT-D2 underestimates intermolecular separation relative to both variants of DFT-D3; some form of correction to MP2 is important, standard meta-GGA functionals perform surprisingly well; etc. This does not seem surprising, since the coronene dimer is often thought, to a first approximation, to be largely just a bigger version of the benzene dimer. This picture is very limited, however: comparison of our data for the coronene dimer with those for the parallel-displaced benzene-benzene dimer demonstrate that there exists only a very weak correlation between percent errors in geometries of the benzene dimer and those in the coronene dimer; this correlation is weaker still – if not entirely absent – when comparing errors in binding energies between the two systems. Thus, we advocate the use of care when extrapolating to larger systems. Moreover, as was previously illustrated in Figure 3.5 for the cyclopentane dimer, neither the signs nor relative magnitudes of the errors in geometry and energy for the coronene dimer are in any way correlated. This point is illustrated still further by Table 3.3, in which the Pearson’s correlation coefficient for percent errors in interpolated equilibrium energy and intermolecular separation across the M12 set is listed for each method.

It is clear that equilibrium binding energies and intermolecular separations are only weakly correlated for a number of methods; for some approaches, they are even somewhat anticorrelated, i.e. the methods overbind while overestimating intermolecular separation. It is also notable that wavefunction-based approaches, on average, seem to exhibit a stronger correlation between equilibrium binding energy and intermolecular separation than density functionals. This being said, this sort of orthogonality between energy and separation observed for a number of methods is not a flaw; rather, it is simply an interesting phenomenon that highlights yet again the fact that merely comparing energies is an insufficient means of assessing the performance of a given method for intermolecular interactions.

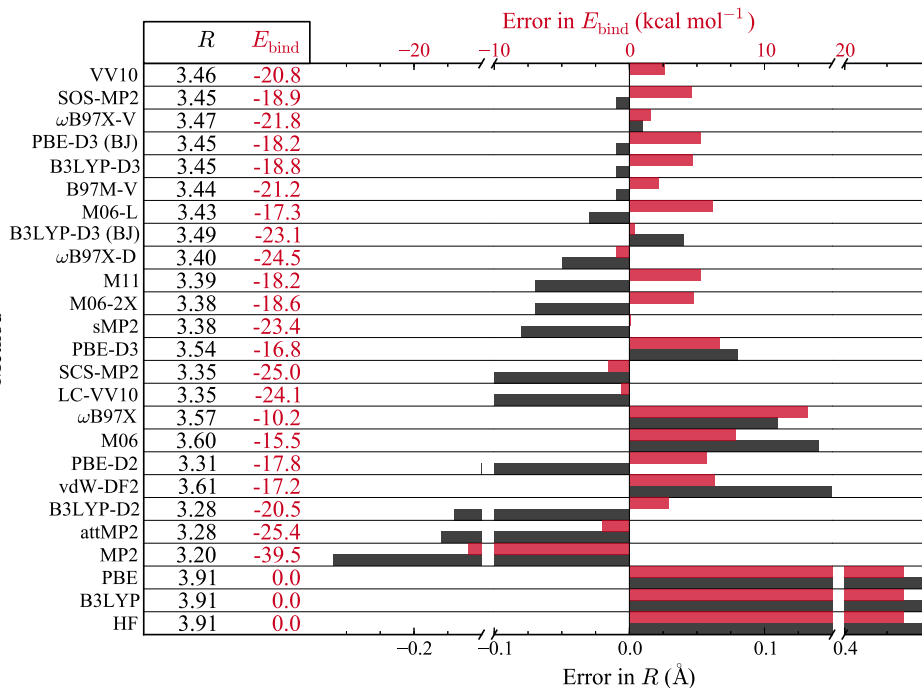


Figure 3.8: Errors in interpolated equilibrium intermolecular separations (in units of \AA) and binding energies (in units of kcal mol^{-1}) for the parallel-displaced coronene dimer. The methods are listed in order of ascending error in intermolecular separation. Equilibria correspond to the interpolated (cubic spline) minima of the binding energy curves. All calculations were performed in the aug-cc-pVDZ basis. Calculations involving density functionals utilized a (99,590) grid. With the exception of attenuated MP2, all methods incorporate a correction for BSSE[91]. Errors in separation are relative to $R = 3.458 \text{ \AA}$ reported by Janowski et al.[137], corresponding to QCISD(T)/h-aug-cc-pVDZ, and errors in binding energy are relative to $E_{\text{bind}} = -23.45 \text{ kcal mol}^{-1}$, which corresponds to QCISD(T)/CBS as reported by Mardirossian and Head-Gordon[19].

3.4 Conclusion

In this work, we have systematically assessed the abilities of a variety of electronic structure approximations to replicate coupled-cluster-level geometries of non-covalent complexes. Methods examined include HF, MP2, and several common DFT exchange-correlation functionals with and without various dispersion corrections. A variety of systems were studied: the A21 set of small (2-4 heavy atoms) systems, the M12 set of moderately-sized (8-14 heavy atoms) systems, and the parallel-displaced coronene dimer (48 heavy atoms). For the A21 set, $\Delta\text{CCSD(T)}/\text{CBS}$ geometries are readily available for comparison[131]. However, for the larger systems, multidimensional optimizations at such a level of theory are prohibitively

| WFT | | DFT | |
|----------|------|-----------------|-------|
| Method | R | Method | R |
| SOS-MP2 | 0.95 | PBE | 0.96 |
| MP2 (CP) | 0.94 | B3LYP | 0.86 |
| HF | 0.93 | PBE-D | 0.84 |
| SCS-MP2 | 0.93 | M06 | 0.84 |
| attMP2 | 0.88 | ω B97X | 0.83 |
| MP2 | 0.88 | ω B97X-D | 0.82 |
| sMP2 | 0.87 | LCVV10 | 0.74 |
| | | B97M-V | 0.71 |
| | | B3LYP-D3 | 0.63 |
| | | vdW-DF2 | 0.61 |
| | | PBE-D3 (BJ) | 0.40 |
| | | M06-L | 0.39 |
| | | PBE-D3 | 0.22 |
| | | M11 | 0.22 |
| | | M06-2X | 0.19 |
| | | VV10 | 0.17 |
| | | B3LYP-D3 (BJ) | 0.13 |
| | | B3LYP-D | -0.28 |
| | | ω B97X-V | -0.29 |

Table 3.3: Pearson’s correlation coefficient, R , for percent errors in interpolated equilibrium energy and intermolecular separation for the M12 dataset. Methods are divided into two sets – wavefunction-based (WFT) and density functional theory (DFT) – and listed in order of descending correlation coefficient within each set. For details about the calculations, see Figure 3.7.

expensive. Thus, we have established the validity of a protocol for utilizing binding energy curves along a single intermolecular coordinate to probe the performance of a given method with regard to geometries: interpolation with a cubic spline yields a minimum consistent with explicit constrained optimization, even with a relatively large distance between sampled geometries. Although the overall root-mean-square deviation is the most comprehensive metric for differentiating among methods, this sort of measure of error in intermolecular separation is a reasonable substitute.

We find that the relative performances of the various electronic structure methods for reproducing CCSD(T) geometries is dependent on not only the predominant interaction type, but also the size of the molecular system. Nevertheless, a number of general trends that transcend system size are evident. Those methods incorporating the VV10 brand of nonlocal correlation tend to yield quite accurate geometries; the recently-developed functionals ω B97X-V and B97M-V in particular are remarkably consistent, although the former tends to overestimate equilibrium separations, especially in the context of small systems. The vdW-DF2 method, on the other hand, systematically predicts too-large intermolecular separations, with the associated error dramatically increasing during the transition from small systems to moderately-sized systems. Conventional GGA functionals (e.g. PBE) yield wildly inaccurate geometries for systems in which dispersion interactions are domi-

nant. The addition of some form of correction for long-range dispersion generally improves their performances; furthermore, for the systems examined, DFT-D3 and DFT-D3 (BJ) are unambiguously superior to DFT-D2 with respect to geometries, though not necessarily with respect to equilibrium binding energies. Conventional semilocal functionals, such as ω B97X and the Minnesota functionals, yield decidedly mediocre geometries, particularly for dispersion-dominated interactions. Furthermore, some of these functionals – most notably M06 and M06-L – suffer from nonphysical grid-dependent oscillations, particularly for the coronene dimer[99]. Such oscillations introduce a large degree of uncertainty into the equilibrium intermolecular separations and binding energies; after all, an ill-behaved potential energy surface has an ill-defined minimum.

Among the wavefunction-based approaches examined, Hartree-Fock theory yields grossly inadequate geometries – even for small hydrogen-bonded complexes, on which it might be expected to perform reasonably well. A perturbative correction for electronic correlation vastly improves upon this HF picture: MP2 yields highly accurate geometries for small molecules. However, the shortcomings of the method are manifest in the deterioration of the quality of its predicted geometries of larger dispersion-dominated systems. In systems with upwards of 8 heavy atoms, MP2 yields geometries that are at best on par with standard density functionals, even when corrected for BSSE. Attenuation of the Coulomb operator achieves a similar effect to counterpoise correction; both approaches systematically underestimate intermolecular separation in dispersion-bound systems. It is likely that increasing the size of the parameter space – by, e.g., combining attenuation with scaling – would improve the description of geometries further, as it has been shown to do with energies[142].

The evaluation of electronic structure methods with regard to their description of geometries adds an important additional dimension to benchmarking binding energies. In the past, relative energies have served as the primary means of comparison of various methods. This procedure has several shortcomings. There is no information regarding the shape of the potential energy surface associated with each method. Moreover, it is not a particularly fair comparison: typically, the same geometry is used for all methods, such that the reported energy is identically the equilibrium energy for only one method. A self-consistent treatment with each method, wherein the structure is relaxed prior to the energy calculation, is a more balanced approach. Incorporating some form of geometric metric – e.g., some measure of intermolecular separation for non-covalent complexes – into the training and selection of new density functionals may lead to the development of more robust methods, and can be achieved with relative ease. This has been done, in a fashion, in the development of the HCTH, τ -HCTH, and BMK functionals[143–145]. These functionals were parameterized on experimental geometries of a number of small molecules by incorporating the computed gradient at these reference geometries into the penalty function for each method. What we are proposing here is extending this sort of idea to non-covalent interactions, which have a handful of highly-variable intermolecular degrees of freedom. The inclusion of a single additional metric is by no means an end-all solution, as information relating to the shape of the potential energy surface is still absent, but it is a step in the right direction, and it is something for which sufficiently high quality reference data already exists and further data

can be relatively straightforwardly generated.

Chapter 4

Basis Sets for Intermolecular Interactions

4.1 Introduction

The past two decades have seen an explosion of interest in Kohn-Sham density functional theory (DFT)[11], largely due to the potential of approximations within the formalism to strike a nice balance between computational expense and accuracy. Twenty years ago, state-of-the-art methods were generalized gradient approximations (GGA) with few, if any, nonempirical parameters[14–16, 22, 82]. Nowadays, density functionals abound: the most successful relics of the past (e.g. PBE, B3LYP) still live on, but the quest for the ultimate density functional continues[20, 23, 25, 85–87]; recent work by Mardirossian and Head-Gordon involved exploration of a space of over a billion meta-GGAs, a space orders of magnitude larger than the space of previously-existing density functionals (yet still a tiny fraction of the unexplored space of B97-esque functionals)[19]. Needless to say, there has been – and continues to be – a tremendous amount of effort dedicated to the development and testing of novel density functional approximations.

Although settling on a method is arguably the most important step one makes prior to running an electronic structure calculation, there remain other decisions that can significantly impact results, most notably grid – in the case of numerical calculations, as in DFT – and basis set. The issue of grid is relatively trivial to resolve: a standard semilocal DFT calculation is linear in the number of grid points, and so it is feasible to employ incredibly dense grids. The issue of basis set is a bit stickier, however, since it is the size of the basis that dominates the scaling. In extended and periodic systems, plane waves constitute the natural choice of basis function, though the delocalized nature of plane waves renders them ineffective at describing localized densities, e.g. core electrons. As a result, periodic calculations tend to employ some form of additional approximation to describe the effects of core electrons[146]. In calculations on molecular systems, local atomic orbital (AO) basis sets are arguably more physically relevant; common representations of AOs include Slater

orbitals[147] and Gaussian-type orbitals (GTOs)[148]. The latter are typically preferred; the Gaussian Product Theorem renders the necessary integrals more computationally tractable. The focus of this work will thus be GTO basis sets.

Even limiting oneself to existing GTO basis sets, the space of possibilities is enormous. There are a vast number of hierarchical basis sets that are in common use; for details pertaining to their construction, see Jensen’s recent review[149]. The fact that so many basis sets are regularly employed is a testament to the fact that there really is no unambiguously *best* basis set of a given size. Here, we focus primarily on three families of GTO basis sets: the Dunning correlation-consistent sequence[79, 80, 150], the Jensen polarization-consistent sequence[151–153], and the Karlsruhe property-optimized basis sets[154, 155]. The correlation-consistent basis sets have been designed to exploit the fact that the correlation energy converges as an inverse power series in the highest angular momentum of the basis[156, 157]; the result is systematic convergence with the cardinal number of the basis set. In the case of DFT, however, the convergence patterns of these basis sets lack the same theoretical underpinnings. Similarly, the convergence behaviors of the Jensen and Karlsruhe sequences, particularly in the context of intermolecular interactions – the domain of many interesting problems in modern chemistry – are not well-documented.

When local basis sets are utilized, two interrelated types of basis set incompleteness errors (BSIE) emerge: basis set superposition error (BSSE), which arises from the inconsistent treatment of a supersystem and its constituent fragments[91, 158], and what we will call the remaining basis set incompleteness error (rBSIE), the leftover incompleteness error once BSSE is removed that is due to the fact that the Schrödinger equation is being solved in just a fraction of the full Hilbert space. We will here briefly address the issue of BSSE, since unlike rBSIE it can be relatively cheaply eliminated. For a more detailed discussion of basis set errors, particularly in the context of small basis sets, the reader is referred to a recent review article by Sure et al.[159]. BSSE is often removed by performing fragment calculations in the full supersystem basis; this constitutes the counterpoise correction (CP) approach of Boys and Bernardi[91], though the downside of this approach is that there must exist some natural partitioning of the full supersystem into fragments. The validity of this and other BSSE correction schemes has long been a contentious issue[160–169], though a comprehensive review article by van Duijneveldt et al.[165] temporarily resolved the debate in favor of CP. Recent years have seen a resurgence of arguments against CP[166, 167], though it has been demonstrated that some of the data used to formulate the conclusions of Kalescky et al.[167] were impacted significantly by unrelated issues, namely mismatches between the choices of basis sets and the extent of correlation included[168]. Linear dependency issues in calculations involving ghost atoms may have affected the results, as well. Some authors have proposed a compromise – a half-CP approach, wherein uncorrected and counterpoise-corrected binding energies are averaged – on the basis of its stellar performance for certain methods[169]. Needless to say, the field has yet to reach a consensus on how BSSE should be addressed.

In this work, we have endeavored to fill in various gaps in the literature, to characterize the convergence patterns of several common hierarchical families of basis sets in the context

of noncovalent interactions as described by DFT. We have further distinguished between the two manifestations of basis set error, BSSE and rBSIE. Characterizing these errors in conventional basis sets enables us to make recommendations regarding which basis sets to use when studying noncovalent interactions with DFT.

4.2 Computational Methods

We have examined basis set errors of a wide variety of standard GTO basis sets in the context of density functional theory calculations on noncovalent interactions. Specifically, we have considered several Pople split-valence basis sets: 6-31G*, 6-31++G**, 6-311++G**, and 6-311++G(3df,3pd)[170–177]; the correlation-consistent basis sets of Dunning, cc-pVXZ, as well as their augmented variants aug-cc-pVXZ, with X=D,T,Q,5[79, 80, 150], the doubly-augmented versions d-aug-cc-pVDZ and d-aug-cc-pVTZ[79, 80, 178], the core-valence sets aug-cc-pCVDZ and aug-cc-pCVTZ[79, 80, 179], and Truhlar’s pruned jun-cc-pVXZ analogues[180]; the Karlsruhe sequence def2-SVP, def2-TZVP, and def2-QZVP[154], as well as the augmented variants def2-SVPD, def2-TZVPD, and def2-QZVPD[155]; and the Jensen polarization-consistent sequences pc-*n* and aug-pc-*n*[151–153]. We have utilized three density functional approximations: SPW92[10–13], a local-density approximation; B3LYP[15, 16, 22, 82], a global hybrid generalized gradient approximation; and B97M-V[19], a meta-generalized gradient approximation incorporating VV10 long-range correlation[40]. This set of approximations was chosen because it spans the space of complexity in common density functionals: SPW92 is one of the simplest Kohn-Sham density functionals; B3LYP is a bit more complex due to its incorporation of the gradient of the electron density, as well as a portion of exact exchange; and B97M-V is a state-of-the-art method with all the accompanying bells and whistles, most notably an explicit nonlocal correlation kernel. Calculations have been performed on the S22 set of molecules[181], depicted in Figure 4.1.

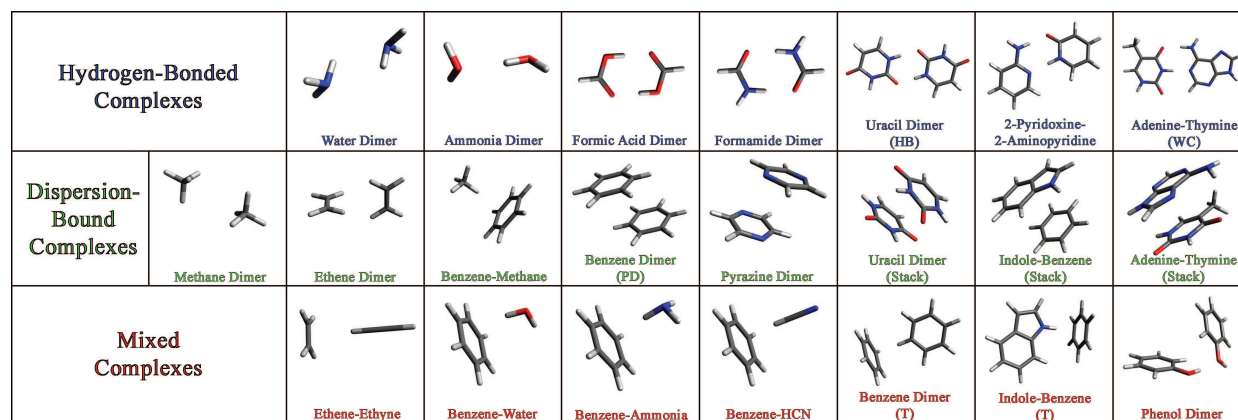


Figure 4.1: Structures of the systems in the S22 dataset. The systems are classified by interaction type as per the original work[181].

All calculations were performed with a development version of Q-Chem 4.3[182]. The DIIS error was converged to 10^{-8} , integral thresholds of 10^{-14} were used, and no symmetry was exploited. Molecular structures were generated with Avogadro[95]. For all systems, binding energies were determined both with and without the Boys and Bernardi correction for BSSE[91]. The occupied orbital resolution-of-the-identity approximation (occ-RI-K) was utilized to accelerate construction of the exact exchange matrix in B3LYP[183]. For the cc-pVXZ and def2- basis sets, optimized auxiliary basis sets from Weigend were used, though i functions were omitted[184, 185]. Auxiliary basis sets for the aug-cc-pVXZ basis sets were generated by adding an even-tempered diffuse function to each primitive set; the (aug-)cc-pVDZ auxiliary bases were generated by removing the highest angular momentum functions from the (aug-)cc-pVTZ auxiliary bases; and for the (aug-)pc- n and Pople basis sets, the corresponding Dunning auxiliary basis sets were used, e.g. cc-pVTZ-jkfit for pc-2.

Due to the constraints of double precision floating point numbers and linear-dependency issues in calculations on some systems with the larger basis sets, the precision to which we report all binding energies is 0.01 kcal/mol. For the smaller basis sets we could meaningfully reproduce binding energies to a much greater level of precision, but the same is not true for larger basis sets; this is particularly an issue in basis sets rife with diffuse functions, e.g. aug-pc-3. The desired level of precision dictates the grids necessary: a Lebedev integration grid consisting of 99 radial points and 590 angular points per atom was utilized to compute the semilocal exchange-correlation components of the energy, and the coarser SG-1 grid was used for nonlocal correlation in B97M-V[133]. This combination of grids yields binding energies that are converged to within 0.01 kcal/mol across the entire S22 set, as can be seen within Appendix B.

4.3 Results and Discussion

Although the principal objective of this study has been to elucidate the convergence patterns of various standard GTO basis sets in the context of density functional theory applied to intermolecular interactions, in the course of this work we have, at three different levels of density functional theory, established what we deem to be complete-basis (CBS) binding energies for every system in S22. These CBS binding energies correspond to counterpoise-corrected values in the pc-4 basis. We justify this choice of CBS limit – counterpoise-corrected pc-4 – in three ways: firstly, the pc-4 basis is the only basis examined for which the mean BSSE across the S22 set of molecules is not larger than the chosen level of precision, 0.01 kcal/mol; secondly, pc-4 absolute energies for any given system are lower than those computed with any other basis set in this study, and as such pc-4 is variationally the most complete basis examined; thirdly, counterpoise-corrected pc-4 SPW92 binding energies are converged to within 0.01 kcal/mol relative to those calculated in the much larger aug-pc-4 basis set, as is illustrated in Appendix B – in fact, even pc-4 absolute energies are converged to roughly this level of precision. Any reference to basis set limit SPW92, B3LYP, or B97M-V results henceforth corresponds to counterpoise-corrected pc-4, and all errors –

unless otherwise noted – are expressed relative to the basis set limit result for the relevant method. Since complete-basis results are costly to obtain and are of interest for e.g. anyone testing a novel basis set in one of these methods, we present them in Table 4.1. Reference CCSD(T)/CBS values generated by Marshall et al.[186] are also provided for comparison.

| System | SPW92 | B3LYP | B97M-V | CCSD(T) ^a |
|--------------------------------------|--------|--------|--------|----------------------|
| Ammonia dimer | -5.07 | -2.19 | -3.09 | -3.13 |
| Water dimer | -7.81 | -4.51 | -5.00 | -4.99 |
| Formic acid dimer | -26.98 | -17.45 | -18.69 | -18.75 |
| Formamide dimer | -21.92 | -14.07 | -15.62 | -16.06 |
| Uracil dimer h-bonded | -26.27 | -17.99 | -20.11 | -20.64 |
| 2-pyridoxine 2-aminopyridine complex | -22.89 | -13.83 | -16.39 | -16.93 |
| Adenine thymine Watson-Crick complex | -22.08 | -12.91 | -15.82 | -16.66 |
| Methane dimer | -0.83 | 0.39 | -0.43 | -0.53 |
| Ethene dimer | -2.47 | 0.48 | -1.31 | -1.47 |
| Benzene - Methane complex | -2.02 | 0.76 | -1.34 | -1.45 |
| Benzene dimer parallel displaced | -2.60 | 3.72 | -2.53 | -2.65 |
| Pyrazine dimer | -4.43 | 2.45 | -3.85 | -4.26 |
| Uracil dimer stack | -10.14 | -0.95 | -9.76 | -9.81 |
| Indole benzene complex stack | -4.36 | 4.64 | -4.35 | -4.52 |
| Adenine thymine complex stack | -11.95 | 1.29 | -11.75 | -11.73 |
| Ethene ethyne complex | -2.27 | -0.66 | -1.50 | -1.50 |
| Benzene water complex | -4.44 | -1.20 | -3.10 | -3.28 |
| Benzene ammonia complex | -3.04 | -0.11 | -2.13 | -2.31 |
| Benzene HCN complex | -5.82 | -1.97 | -4.21 | -4.54 |
| Benzene dimer T-shaped | -3.05 | 0.98 | -2.33 | -2.72 |
| Indole benzene T-shape complex | -6.27 | -0.55 | -5.02 | -5.63 |
| Phenol dimer | -9.01 | -2.99 | -6.57 | -7.10 |

^a CCSD(T) values taken from Marshall et al.[186].

Table 4.1: Complete basis set (CBS) binding energies for each system in S22 at various levels of theory. For the density functional approximations, counterpoise-corrected pc-4 constitutes the CBS limit. Benchmark CCSD(T)/CBS results from Marshall et al.[186] are provided for comparison.

Now that benchmarks for each method have been established, it is possible to assess the qualities of various standard local quantum chemistry basis sets in the context of noncovalent interactions as described by density functional theory. Note that the most meaningful point of comparison for each method-basis combination is the CBS limit for that method; by comparing to reference CCSD(T)/CBS results, we would be confounding method error with basis set error, whereas by comparing to CBS results within each method we are able to isolate basis set errors. The uncorrected (noCP) and counterpoise-corrected (CP) root mean square errors (RMSEs) versus the CBS limit for SPW92 are illustrated in Figure 4.2 for a variety of basis sets. Basis sets are listed in order of increasing size: the basis with the fewest functions (6-31G*) is at the top and that with the most (aug-cc-pV5Z) is at the bottom. Also noteworthy is the fact that there are two axes, one which corresponds to the uncorrected

RMSE (top, blue), and one which corresponds to the counterpoise-corrected RMSE (bottom, gold); the scales of these axes differ by roughly a factor of four in order to compensate for the fact that the CP RMSEs are significantly smaller, on average, than the noCP RMSEs.

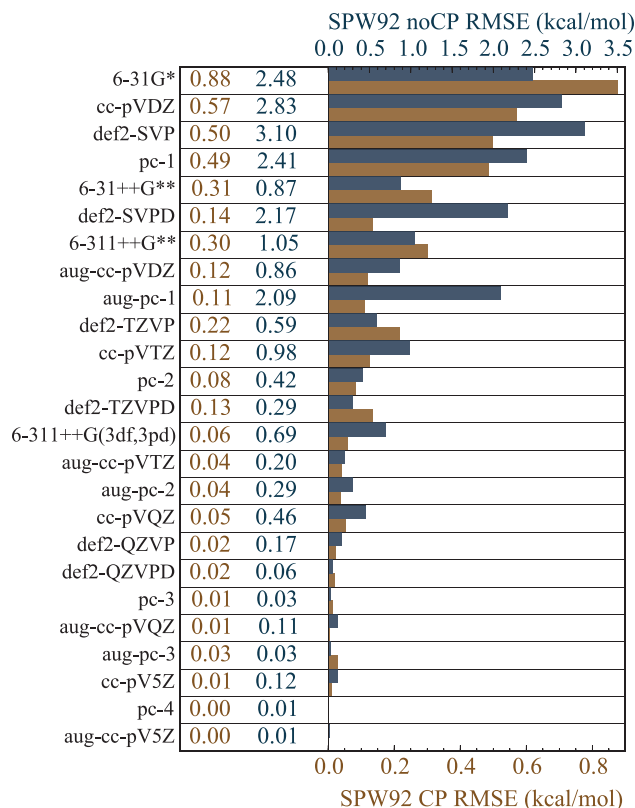


Figure 4.2: Uncorrected (noCP) and counterpoise-corrected (CP) root mean square errors (RMSE) in SPW92 binding energies across the S22 set of molecules. Errors are expressed relative to SPW92/CBS. Blue corresponds to noCP, gold to CP. The bars are a visual representation of the actual RMSEs, which are tabulated for each basis set on the left side of the figure. Basis sets are listed in order of increasing number of basis functions. Note the difference between noCP (top) and CP (bottom) axes.

From Figure 4.2, it is immediately evident that for any given basis set, the CP RMSE is significantly smaller than the noCP RMSE; that is, counterpoise-corrected binding energies across S22 are closer to the basis set limit than uncorrected ones. Since the counterpoise correction is designed to alleviate BSSE, the CP RMSE is a quantitative measure of the remaining basis set incompleteness error (rBSIE). The noCP RMSE, on the other hand, encompasses the entirety of the BSIE, i.e. both rBSIE and BSSE. Basis sets exhibiting a small CP RMSE can then be said to have a low intrinsic rBSIE: their spans form a good

approximation to the full Hilbert spaces of the systems in the S22 set. It is worth noting that when discussing basis set quality, rBSIE is arguably a more important metric than BSSE, since BSSE can be relatively cheaply removed – at least in the context of intermolecular interactions.

As the size of the basis increases, computed energies generally approach the basis set limit. This is trivially true within a given variational space, as exemplified by the difference between e.g. cc-pVDZ and aug-cc-pVDZ. The trend also weakly holds even when the spaces are different: a quadruple-zeta basis is usually closer to the basis set limit than a triple-zeta basis. Nevertheless, the different families of basis sets exhibit different rates of convergence to the basis set limits, such that it is possible for a basis set in one family to be smaller yet more complete than one in another family (compare def2-QZVPD to aug-cc-pVQZ). Certain basis sets are particularly cost-effective. This is represented within Figure 4.2 by smaller bars higher up, e.g. CP def2-SVPD. Other basis sets – most notably those in the Pople sequence, e.g. 6-31++G** – outperform their similarly-sized competitors when not corrected for BSSE, but dramatically underperform once a counterpoise correction is employed. Thus, different conclusions regarding relative qualities of basis sets may be drawn based on whether or not counterpoise correction is desired, which is evidenced by different trends between the gold and blue bars in Figure 4.2. Nevertheless, correcting for basis set superposition error is useful, particularly in the case of smaller basis sets. For instance, calculating a counterpoise-corrected binding energy for a dimer in the def2-SVPD basis entails less than 20% of the effort required to calculate the corresponding uncorrected energy in the aug-cc-pVTZ basis, despite the fact that the CP def2-SVPD calculation is – as evidenced by these data – more accurate.

Similar trends regarding the qualities of the various basis sets are observed for the other density functional approximations, B3LYP and B97M-V, as is illustrated by comparison with Figures 4.3 and 4.4, respectively. There are, of course, some exceptions: for instance, in the case of B97M-V, the noCP aug-pc- n results are disproportionately worse, which may simply be an artifact of its training. Nevertheless, a significant degree of transferability is expected for any density functional with a well-behaved inhomogeneity correction factor (ICF), regardless of whether exact exchange is incorporated or not. All bets are off, however, when considering functionals with strongly oscillatory ICFs; for more details, see Mardirossian and Head-Gordon[99].

One particularly interesting aspect of the similarities among functionals observed in this study is that the nonlocal VV10 correlation of B97M-V is no more sensitive to basis set than the semilocal exchange and correlation components. In fact, VV10 nonlocal correlation is vastly *less* sensitive to basis set size than even the local exchange and correlation of SPW92: there is effectively no difference between VV10 nonlocal contributions to binding energies in the def2-SVPD and pc-4 basis sets. Relevant data are provided in Appendix B. It has previously been established that nonlocal correlation energies are insensitive to grid[23, 187], but to our knowledge this is the first time basis set insensitivity has been reported. This is a conceptually interesting phenomenon which could be exploited to greatly reduce the cost of electronic structure calculations with VV10; such will be the focus of work to come.

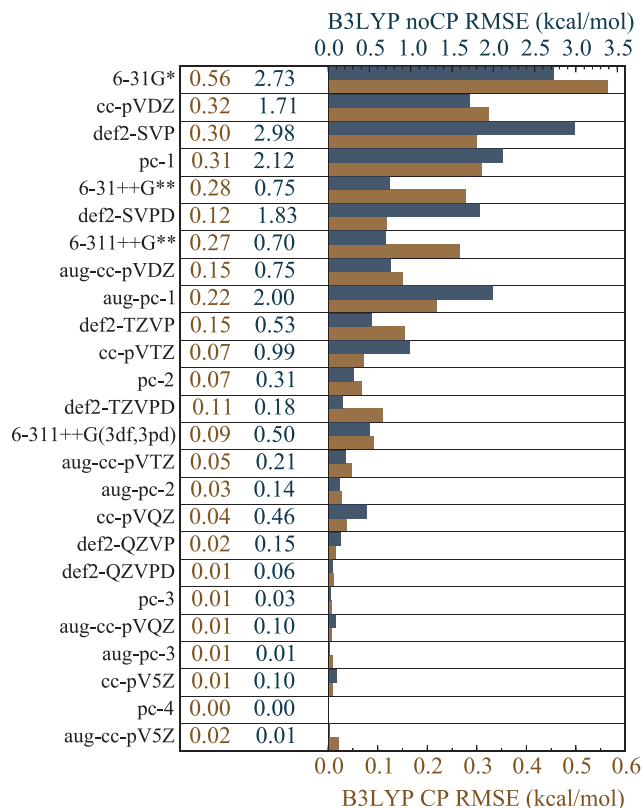


Figure 4.3: Uncorrected (noCP) and counterpoise-corrected (CP) root mean square errors (RMSE) in B3LYP binding energies across the S22 set of molecules. For further details, see Figure 4.2.

The convergence patterns of the Dunning, Jensen, and Karlsruhe basis sequences for uncorrected and counterpoise-corrected mean binding energies in S22 are summarized in Figures 4.5 and 4.6, respectively, at the local-density approximation level of DFT. For the Dunning and Jensen sequences, basis sets of double- through quintuple-zeta quality were employed (with the exception of aug-pc-4, since those calculations could not all be converged). For the Karlsruhe sequences, basis sets of double-zeta through quadruple-zeta quality were used. Similar convergence patterns are observed for B3LYP and B97M-V; relevant figures may be found in Appendix B. It is evident that the counterpoise-corrected results converge significantly more quickly than the uncorrected results with respect to the number of basis functions, regardless of the choice of basis sequence; note the difference in scales between Figure 4.5 and Figure 4.6. However, this accelerated convergence comes at the cost of a loss of systematicity in the error: whereas complexes are always overbound when uncorrected for BSSE (as is guaranteed by the variational borrowing of fragment functions in supersys-

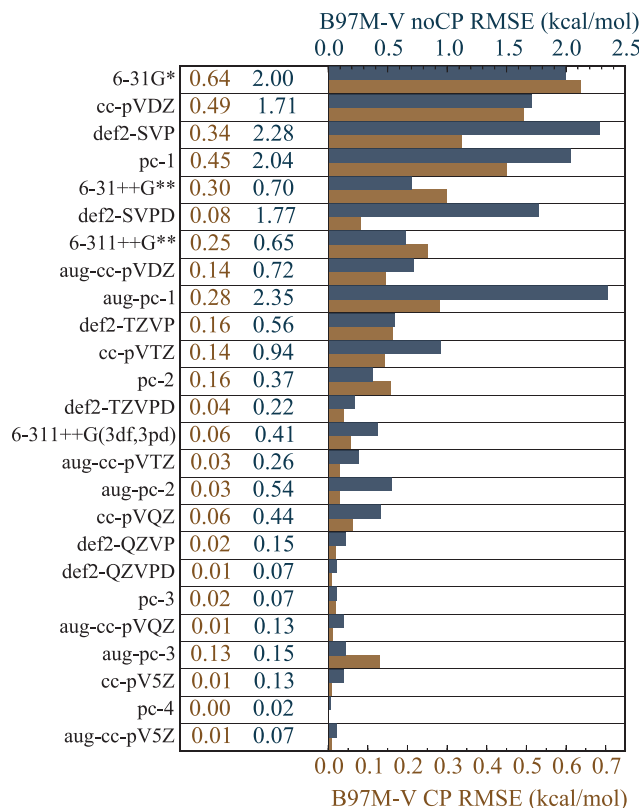


Figure 4.4: Uncorrected (noCP) and counterpoise-corrected (CP) root mean square errors (RMSE) in B97M-V binding energies across the S22 set of molecules. For further details, see Figure 4.2.

tem calculations), they are not always underbound following application of the counterpoise correction.

It is clear from Figure 4.5 that the Dunning sequences of basis sets cc-pVXZ and aug-cc-pVXZ converge remarkably slowly to the basis set limit for DFT. In fact, even binding energies at the uncorrected aug-cc-pV5Z level are not fully converged: in this basis set, the mean BSSE across S22 ranges from 0.01 to 0.06 kcal/mol, depending on the method. The pc-4 basis set, on the other hand, is essentially BSSE-free to this level of precision, despite being roughly 17% smaller. Even binding energies in the def2-QZVPD basis are converged to approximately the same level as those in aug-cc-pV5Z, despite the fact that def2-QZVPD is less than half the size of aug-cc-pV5Z. The Dunning sequences of basis sets are undeniably inefficient for DFT; that being said, they were designed with correlated wavefunction-based methods in mind, so this is not altogether unexpected. When considering counterpoise-corrected results, the picture is not nearly as bleak, though the Jensen sequences

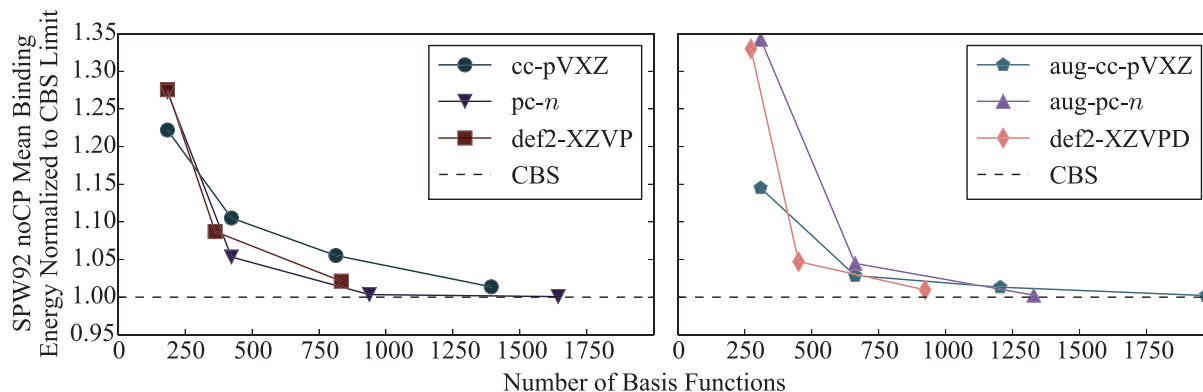


Figure 4.5: Convergence of uncorrected (noCP) SPW92 normalized mean binding energies across the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets. Each binding energy was normalized to the corresponding SPW92/CBS value before averaging. The number of basis functions for each basis set was determined by averaging the number of basis functions for each system within each basis set across all systems in S22.

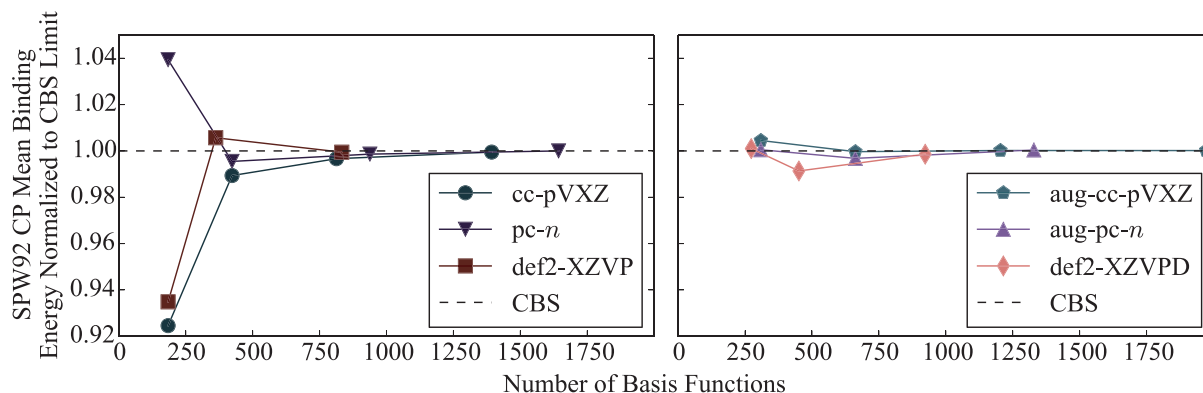


Figure 4.6: Convergence of counterpoise-corrected (CP) SPW92 normalized mean binding energies across the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets. For further details, see Figure 4.5.

of polarization-consistent basis sets still converge more quickly than the Dunning sequences, as is illustrated in Figure 4.6. On the basis of these results, it is difficult to justify the use of Dunning basis sets for density functional theory; for a given Dunning basis set, there exists a Jensen or Karlsruhe alternative that is simultaneously smaller and more accurate.

In their calendar basis set article, Papajak et al.[180] argue that the augmented Dunning basis sets contain more diffuse functions than are strictly necessary, and offer pruned versions at the double- through quadruple-zeta levels. We have thus examined one such sequence, the so-called jun-cc-pVXZ sequence of basis sets. The convergence pattern of this basis sequence with B97M-V across S22 is provided in Figure 4.7. At first glance, the jun-cc-pVXZ sequence seems superior to the aug-cc-pVXZ sequence, particularly in the absence of a correction for BSSE. Indeed, at the triple- and quadruple-zeta levels this is the case, but at the double-zeta level BSSE is simply being traded for rBSIE, as evidenced by the significantly increased CP RMSE in Table 4.2. Additionally, even though the jun-cc-pVTZ and jun-cc-pVQZ basis sets outperform their fully-augmented counterparts, they can still not really be recommended. They are not bad basis sets – by all measures examined here, they are better than the corresponding fully-augmented Dunning basis sets – but they are still less cost-effective than the Karlsruhe and Jensen alternatives. As a final note, as is evident in Table 4.2, other methods of altering the Dunning sequence, namely the addition of a new set of core functions (aug-cc-pCVXZ) or a set of diffuse functions (d-aug-cc-pVXZ), are also not useful for converging these binding energies. Again, these enhancements were optimized with other properties in mind, and cannot be expected to be effective in the present application.

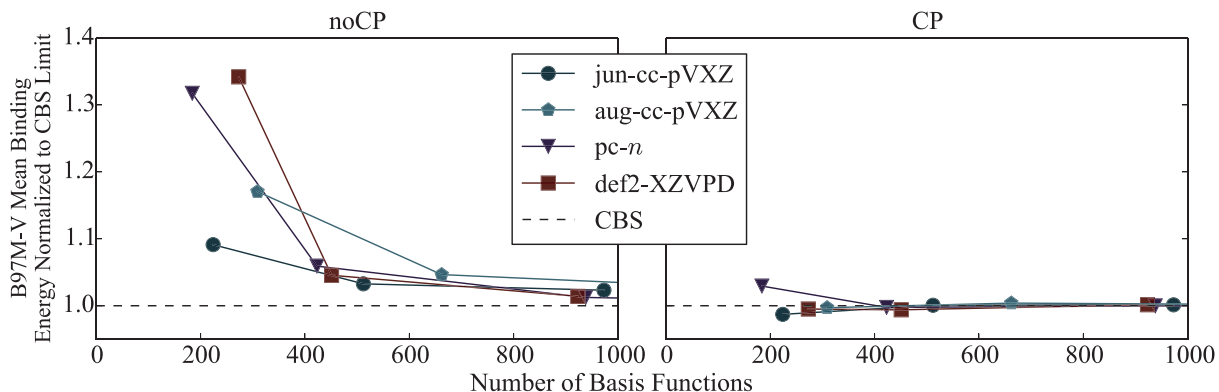


Figure 4.7: Convergence of uncorrected (noCP) and counterpoise-corrected (CP) B97M-V normalized mean binding energies across the S22 set of molecules for the Truhlar, Dunning, Jensen, and Karlsruhe sequences of basis sets. Each binding energy was normalized to the corresponding B97M-V/CBS value before averaging.

Figures 4.8 and 4.9 illustrate the convergence behavior of the Dunning, Karlsruhe, and Jensen sequences of basis sets across various subsets of S22 at the SPW92 level of theory. The subsets are delineated by interaction type: there is a category for hydrogen-bonded

| Basis | RMSE (kcal/mol) | |
|---------------|-----------------|------|
| | noCP | CP |
| aug-cc-pVDZ | 0.72 | 0.15 |
| d-aug-cc-pVDZ | 0.81 | 0.14 |
| aug-cc-pCVDZ | 0.75 | 0.13 |
| jun-cc-pVDZ | 0.57 | 0.39 |
| def2-SVPD | 1.77 | 0.08 |
| aug-cc-pVTZ | 0.26 | 0.03 |
| d-aug-cc-pVTZ | 0.32 | 0.03 |
| aug-cc-pCVTZ | 0.34 | 0.03 |
| jun-cc-pVTZ | 0.20 | 0.02 |
| def2-TZVPD | 0.22 | 0.04 |
| aug-cc-pVQZ | 0.13 | 0.01 |
| jun-cc-pVQZ | 0.14 | 0.01 |
| def2-QZVPD | 0.07 | 0.01 |

Table 4.2: Uncorrected (noCP) and counterpoise-corrected (CP) root mean square errors (RMSE) in B97M-V binding energies across S22 for variants of Dunning-style basis sets and comparably-sized Karlsruhe alternatives.

complexes, one for dispersion-bound complexes, and one for complexes bound by a combination of dispersion and permanent electrostatics; see Figure 4.1 a breakdown of which systems within S22 fall within each category. In these figures, the binding energies for each complex have been normalized to the SPW92 basis set limit – namely, counterpoise-corrected pc-4 – then these normalized binding energies have been averaged across each subset. Based on Figure 4.8, it is evident that in the absence of a correction for BSSE, the relatively slow convergence of the Dunning basis sequences in comparison to the Jensen and Karlsruhe sequences cannot be attributed to a failure on any particular interaction type; it is observed regardless of the type of dominant interaction. Through comparison with Figure 4.9, it becomes clear that this issue is a consequence of disproportionately high BSSE rather than rBSIE; after application of the counterpoise correction, the Dunning sequences converge at a rate similar to that of the Jensen and Karlsruhe sequences.

As is observed across the entirety of the S22 set (*cf.* Figures 4.5 and 4.6), comparison of Figures 4.8 and 4.9 demonstrates that the accelerated convergence counterpoise correction affords is observed regardless of interaction type, though again at a loss of systematicity. Thus, it stands to reason that BSSE is – for typical GTO basis sets in the context of typical noncovalent interactions – the predominant flavor of basis set error. One other particularly striking feature of Figure 4.9 is the relatively large deviation from the basis set limit exhibited by the unaugmented double-zeta basis sets. The performance across the hydrogen-bonded and mixed subsets of S22 is not strongly impacted by the decision to include diffuse functions, but for dispersion-bound complexes, particularly in the limit of smaller basis sets, the inclusion of diffuse functions is vital to eliminate rBSIE. This reinforces conventional wisdom: diffuse functions are necessary in order to accurately describe dispersion interactions, and, for certain classes of basis sets, other energetic properties[188]. In larger basis sets,

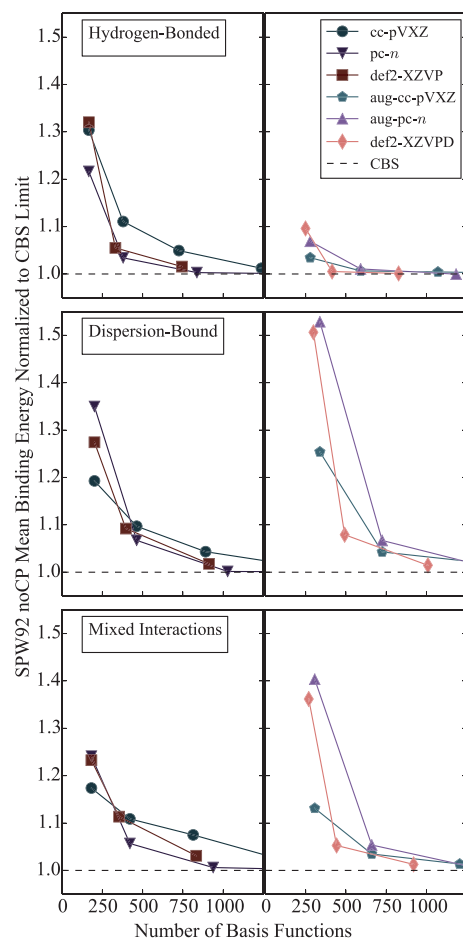


Figure 4.8: Convergence of uncorrected (noCP) SPW92 normalized mean binding energies across subsets of the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets. Within each subset, each binding energy was normalized to the CBS limit (counterpoise-corrected pc-4) before averaging. The three subsets – hydrogen-bonded, dispersion-bound, and mixed interactions – are the same as those in Figure 4.1. The number of basis functions for each basis set was determined by averaging the number of basis functions for each system within each basis set across all systems in the relevant subset of S22.

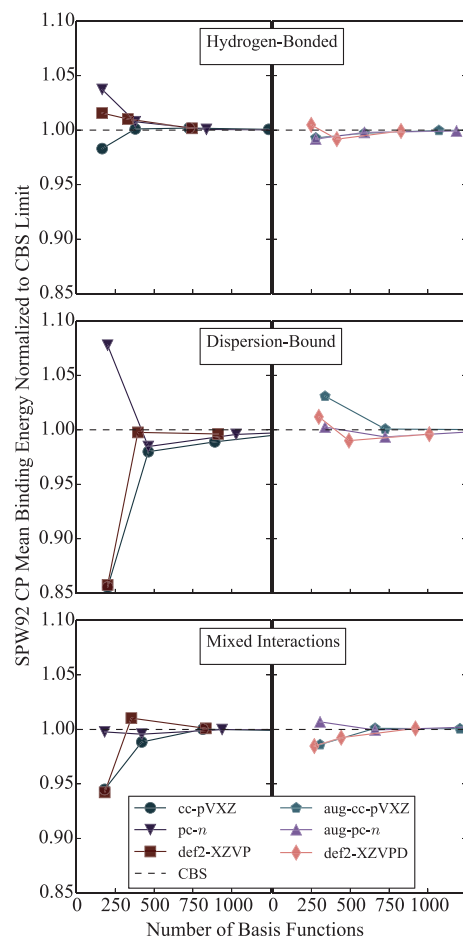


Figure 4.9: Convergence of counterpoise-corrected (CP) SPW92 normalized mean binding energies across subsets of the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets. For further details, see Figure 4.8.

additional diffuse functions may not be necessary depending on the system, since these basis sets tend to already contain basis functions with relatively small exponents, but in basis sets of double-zeta quality (def2-SVP, pc-1, cc-pVXZ), it is imperative to explicitly expand the basis sets to include such functions: def2-SVP is a terrible basis for describing dispersion.

One further interesting observation regarding basis set superposition error can be made on the basis of this work: BSSE is effectively extensive, growing with the number of significant interactions in the system. This is illustrated in Figure 4.10, where we have plotted BSSE versus the number of interacting atoms for three distinct method-basis pairings. We have defined the number of interacting atoms as simply the number of unique atoms within a given system for which the distance to another atom on a different molecule is less than 110% the sum of the van der Waals radii of the atoms. Even this incredibly simple and

naive approach yields a striking correlation between BSSE and system size, as measured by the number of interacting atoms. This property of extensivity justifies the development and use of geometric approaches to predicting BSSE, such as the interaction-specific approach of Merz[189, 190] and the more general approach of Grimme[191, 192].

Given this extensive nature of BSSE, it is possible to extract a meaningful measure of the BSSE associated with each method-basis pairing examined in this study, namely the mean BSSE per interacting atom across the S22 set. This is illustrated for a variety of basis sets with the SPW92, B3LYP, and B97M-V methods in Figure 4.11. Each column is color-coded from highest BSSE (dark red) to lowest BSSE (dark blue) for ease of reading; one thing that is immediately evident is that although the *absolute* BSSE within a given basis set is dependent on the density functional approximation employed, the *relative* BSSE is largely independent of method. Note that this extensivity does not extend to the remaining basis set incompleteness error: there is no such correlation between rBSIE and the number of interactions, as is illustrated in Figure 4.12.

A measure of mean BSSE per interaction effectively allows us to make back-of-the-envelope qualitative predictions of BSSE. For instance, for the equilibrium CO₂-benzene complex[41], we predict using this naive approach a BSSE of 1.7 kcal/mol for B3LYP in the def2-SVPD basis, which is not too far removed from the actual value of 1.1 kcal/mol. If we consider a larger system, such as the parallel-displaced coronene dimer with an interplane separation of 3.308 Å[42, 137], we predict a much larger BSSE: 9.9 kcal/mol for the SPW92 method in the def2-SVPD basis, which compares favorably with the actual value of 13.8 kcal/mol. This is not to suggest that the numbers in Figure 4.11 should be used for any quantitative purpose: they simply provide a rough indication of how much BSSE you can expect for a given system in a given basis set. In order to get a quantitative estimate of BSSE, it is necessary to be more clever in the counting of interactions, such as by incorporating an explicit distance dependence, as in the gCP approach of Kruse and Grimme[191]. Such a quantitative estimate has the potential to be incredibly useful, since for certain basis sets – e.g. the augmented Karlsruhe basis sets – BSSE constitutes the vast majority of basis set error. Thus, judicious choice of basis set could, in combination with some correction scheme for BSSE, yield effectively complete-basis results at a fraction of the effort.

Although this is intended to be primarily a study on basis sets, the availability of both DFT/CBS (Table 4.1) and CCSD(T)/CBS[186] data allows us to make one more meaningful analysis: namely, we can distinguish between apparent error and method error, as has been done previously for wavefunction-based methods[193–195]. The convergences – within the *pc-n* sequence of basis sets – of uncorrected and counterpoise-corrected binding energies for each method towards the CCSD(T)/CBS binding energies are visualized in Figure 4.13. Unlike RMSEs reported elsewhere in this study, the RMSEs in Figure 4.13 correspond to differences from “exact” binding energies. It is immediately evident that B97M-V is the most accurate of the methods examined, which is to be expected, given the nature of its construction and training. SPW92 and B3LYP are not intended to be used for the description of intermolecular interactions. SPW92 systematically overbinds across S22, and is subsequently worse in smaller basis sets without CP. B3LYP, on the other hand, overbinds

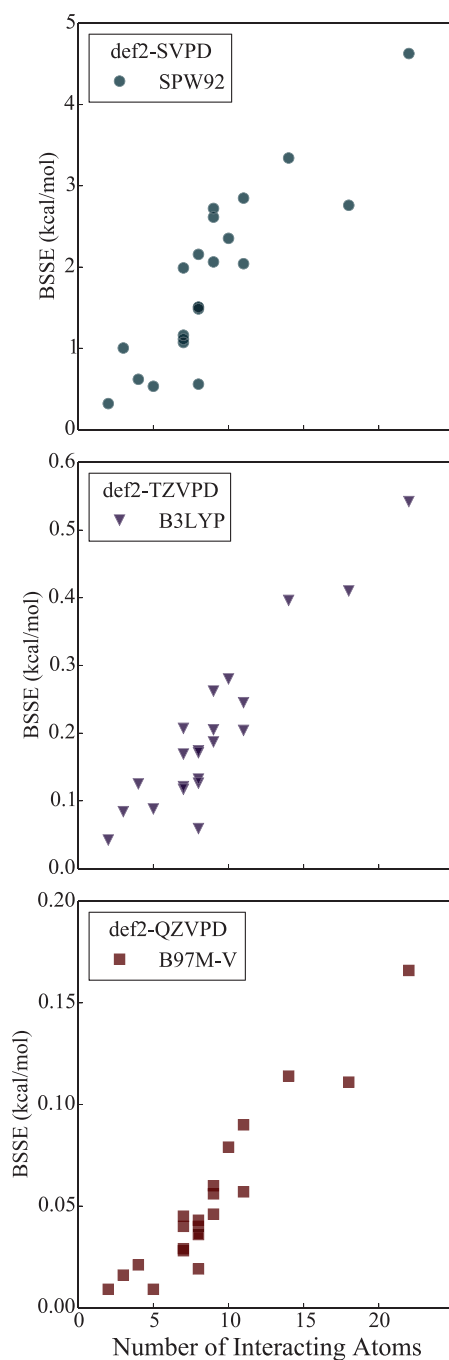


Figure 4.10: Relationship between basis set superposition error (BSSE) and number of interacting atoms for several method/basis set combinations. The number of interacting atoms is defined as the number of unique atoms in each system for which the distance to another atom on a different fragment is less than 1.1 times the sum of the van der Waals radii of the atoms.

| | Size ^a | Basis | Mean BSSE per Interaction (cal/mol) | | |
|-----------|-------------------|-------------------|--|-------|--------|
| | | | SPW92 | B3LYP | B97M-V |
| Pople | 204 | 6-31G* | 260.0 | 263.0 | 212.0 |
| | 300 | 6-31++G** | 76.9 | 76.2 | 67.6 |
| | 348 | 6-311++G** | 83.2 | 63.5 | 54.1 |
| | 684 | 6-311++G(3df,3pd) | 64.8 | 52.8 | 40.7 |
| Dunning | 228 | cc-pVDZ | 310.0 | 296.0 | 249.0 |
| | 528 | cc-pVTZ | 109.0 | 104.0 | 102.0 |
| | 1020 | cc-pVQZ | 50.3 | 47.6 | 47.9 |
| | 1752 | cc-pV5Z | 12.2 | 10.8 | 12.8 |
| | 384 | aug-cc-pVDZ | 89.4 | 92.6 | 78.7 |
| | 828 | aug-cc-pVTZ | 19.8 | 20.9 | 21.1 |
| | 1512 | aug-cc-pVQZ | 9.1 | 8.0 | 11.9 |
| | 2484 | aug-cc-pV5Z | 1.4 | 1.7 | 6.1 |
| Karlsruhe | 228 | def2-SVP | 319.0 | 291.0 | 238.0 |
| | 444 | def2-TZVP | 62.2 | 54.5 | 53.4 |
| | 1044 | def2-QZVP | 17.4 | 15.0 | 14.4 |
| | 336 | def2-SVPD | 206.0 | 185.0 | 158.0 |
| | 552 | def2-TZVPD | 32.1 | 22.1 | 22.3 |
| 1152 | def2-QZVPD | 5.9 | 4.5 | 5.4 | |
| Jensen | 228 | pc-1 | 206.0 | 194.0 | 182.0 |
| | 528 | pc-2 | 41.2 | 28.4 | 35.6 |
| | 1176 | pc-3 | 3.4 | 2.9 | 6.5 |
| | 2064 | pc-4 | 0.6 | 0.4 | 1.7 |
| | 384 | aug-pc-1 | 201.0 | 205.0 | 224.0 |
| | 828 | aug-pc-2 | 29.9 | 12.4 | 44.4 |
| 1668 | aug-pc-3 | 1.1 | 1.5 | 8.7 | |

^aSize here indicates the number of basis functions used for the parallel-displaced benzene dimer.

Figure 4.11: Mean basis set superposition error (BSSE) per interaction across the S22 set of molecules for SPW92, B3LYP, and B97M-V in a variety of basis sets. The number of interactions per system were determined as in Figure 4.10. Each column is color-coded with a gradient from dark red (highest BSSE) to white (median BSSE) to dark blue (lowest BSSE).

hydrogen-bonded complexes in small basis sets and underbinds them in large basis sets and when a correction for BSSE is applied, which accounts for the seemingly bizarre fact that B3LYP appears to be “better” in smaller basis sets. This highlights one of the dangers of judging the merits of a functional on the basis of its apparent error: oftentimes a low apparent error is simply a product of a fortuitous cancellation of basis set error and method error. Such cancellation is also manifest in the behavior of B97M-V, which apparently performs “best” in the pc-2 basis without counterpoise correction. The B97M-V noCP RMSE increases from pc-2 to pc-3, then again from pc-3 to pc-4. This behavior is in stark contrast to that of B3LYP and SPW92, the noCP RMSE errors of which change monotonically with increasing basis size within the same family of basis set, and is directly a result of its training; B97M-V was trained in the aug-cc-pVTZ basis set without counterpoise correction, and hence compensation for aug-cc-pVTZ basis set error was implicitly built in to the functional.

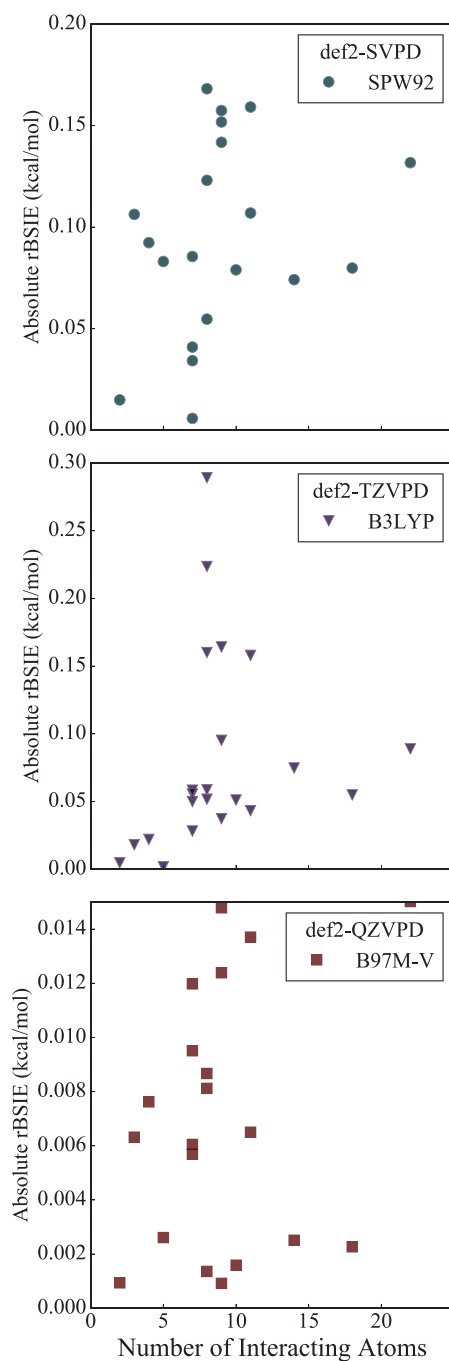


Figure 4.12: Relationship between remaining basis set incompleteness error (rBSIE) and number of interacting atoms for several method/basis set combinations. The number of interactions per system were determined as in Figure 4.10.

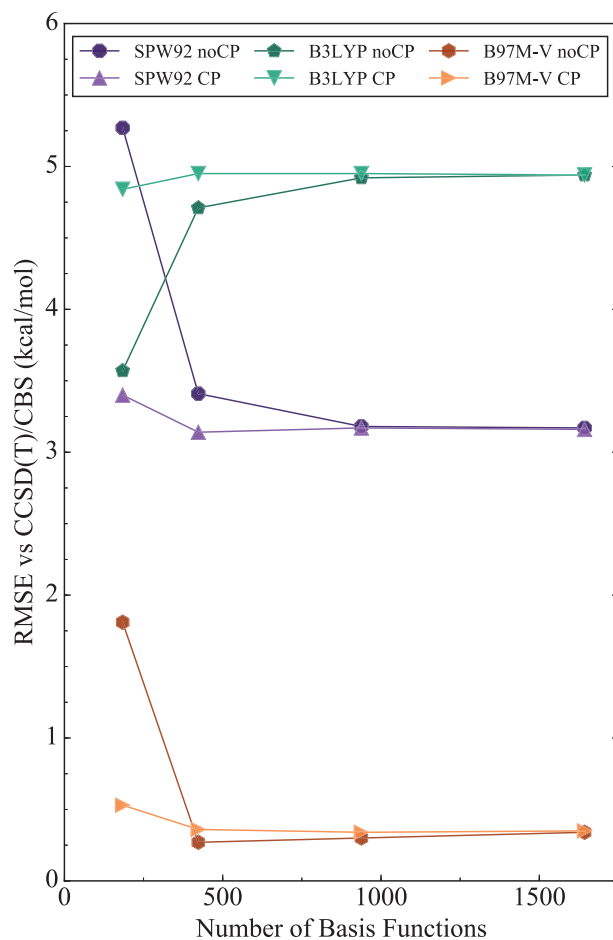


Figure 4.13: Root mean square errors (RMSE) in uncorrected (noCP) and counterpoise-corrected (CP) binding energies for each method within the $pc-n$ sequence of basis sets. Errors here correspond to errors relative to CCSD(T)/CBS results[186]. The number of basis functions is determined as in Figure 4.5.

4.4 Discussion and Conclusion

In this work, we have examined the efficacies of various popular families of basis sets with regards to their abilities to approach complete-basis binding energies across the S22 set of noncovalent interactions. More specifically, we have tested a number of Pople split-valence basis sets, the Dunning sequence of correlation-consistent basis sets, the Karlsruhe def2-basis sets, and the Jensen polarization-consistent sequence of basis sets with three distinctly different density functional approximations: SPW92, a form of local-density approximation, B3LYP, a global hybrid GGA, and B97M-V, a meta-GGA with nonlocal correlation.

Although ours is the first systematic study of the convergence patterns of these basis sets in the context of intermolecular interactions as described by DFT, there have been several other relevant studies comparing the Dunning and Jensen sequences of basis sets in slightly different contexts in the past decade. Shahbazian and Zahedi[196] have demonstrated that polarization-consistent basis sets outperform correlation-consistent basis sets for binding energies of diatomic molecules at the Hartree-Fock level of theory; we show here that this behavior applies to other self-consistent methods, namely DFT, and larger molecular systems. Kupka and Lim[197] have concluded that the pc- n sequence is similarly well-suited to the calculation of molecular and spectroscopic properties, namely geometries and vibrational frequencies, so our particular basis set recommendations may be relevant in the context of such properties. Elsohly and Tschumper[198] have previously shown that binding energies computed with Møller-Plesset perturbation theory to second order (MP2)[5] in the Dunning correlation-consistent sequence of basis sets converge more quickly than when the Jensen polarization-consistent sequence is employed. Correlation-consistent basis sets have their place; they reign supreme in the realm of correlated wavefunction-based calculations.

We have established that counterpoise correction accelerates convergence to the basis set limit for DFT in the context of noncovalent interactions – regardless of basis set sequence – though at the cost of a loss of systematicity of error. Previously, Eshuis and Furche[199] showed that counterpoise correction leads to faster convergence of random phase approximation (RPA) correlation energies across S22, though in the case of RPA, the corrected correlation errors have the added benefit of still being systematic. On the other hand, Elsohly and Tschumper[198] demonstrated that for five weakly bound clusters, counterpoise correction does not accelerate the convergence of MP2 correlation energies. We thus reiterate that our recommendation of counterpoise correction applies strictly to self-consistent methods, particularly well-behaved density functional approximations. We expect our conclusions to be transferable to Hartree-Fock (HF) theory[107, 108]; after all, HF and DFT have been previously shown to have similar basis set requirements[200], and here we have demonstrated that functionals with local and exact exchange within the Kohn-Sham formalism exhibit similar convergence patterns.

In this study, we have also established that it is remarkably difficult to truly reach the basis set limit; even the massive aug-cc-pV5Z basis is plagued by BSSE. In fact, in the context of S22 DFT binding energies, the only effectively BSSE-free basis set examined is pc-4. However, at 2064 basis functions for the benzene dimer, pc-4 is not a particularly pragmatic

basis for day-to-day use. A more economical alternative is the Karlsruhe def2-SVPD basis set. With only 336 basis functions for the benzene dimer, def2-SVPD is relatively affordable; moreover, when corrected for BSSE, it yields results that are comparable to those obtained with the analogous Dunning and Jensen basis sets, aug-cc-pVDZ and aug-pc-1, despite being significantly smaller. In fact, for the S22 set of systems, counterpoise-corrected def2-SVPD binding energies are on par with those obtained with significantly larger basis sets: on the order of 0.1 kcal/mol error across the S22 set of molecules. This is consistent with the findings of Mardirossian and Head-Gordon[19], who recommended counterpoise-corrected def2-SVPD as an alternative to aug-cc-pVTZ on the basis of its ability to reproduce reference coupled-cluster binding energies when combined with the B97M-V functional. We thus demonstrate here that this reproduction does not stem exclusively from fortuitous error cancellation, but rather emerges as a result of the strength of B97M-V and the small intrinsic rBSIE of def2-SVPD in the context of noncovalent interactions.

The complete-basis data presented in Table 4.1, Figure 4.13, and Appendix B shed light on an interesting aspect of functional development. The parameters of the functional B97M-V were optimized without counterpoise correction in the aug-cc-pVTZ basis set. As a result, B97M-V exhibits a smaller RMSE versus CCSD(T)/CBS across S22 without counterpoise correction in aug-cc-pVTZ than in any other basis, with or without a correction for BSSE: the noCP aug-cc-pVTZ RMSE of B97M-V is 0.23 kcal/mol, which is more than 30% smaller than the B97M-V/CBS RMSE of 0.35 kcal/mol. Similarly small RMSEs are exhibited for B97M-V in other basis sets of triple-zeta quality without counterpoise correction, such as noCP def2-TZVPD (0.25 kcal/mol). In larger basis sets, or when a correction for BSSE is applied, B97M-V systematically underbinds across the S22 set (with the exception of two systems: the uracil dimer stack and the ethene-ethyne complex, both of which are significantly overbound at the level of noCP aug-cc-pVTZ). Thus, in the case of methods trained in the presence of significant BSSE and rBSIE – such as B97M-V – it is not necessarily better to use a larger basis, since compensation for these basis set errors within the training basis has been implicitly built into the method. Consequently, we recommend future empirical methods be trained as close to the basis set limit as possible, such as in the def2-QZVPD basis set with counterpoise correction. When training in any finite basis, it is desirable to ensure that the method error is significantly larger than the basis set error; otherwise, the method will invariably rely on some cancellation of these errors and hence underperform when liberated from basis set error.

In addition to our previous small basis recommendation (CP def2-SVPD), we thus establish as our large basis of choice def2-QZVPD; when corrected for BSSE, this basis set is a practical alternative to pc-4. There is an argument to be made against training new density functionals in basis sets with significant rBSIE, since basis set error becomes confounded with method error, as is seen with B97M-V. Counterpoise-corrected def2-QZVPD thus constitutes an ideal level at which train new density functionals: it is sufficiently large to reproduce complete-basis results with errors an order of magnitude smaller than the intrinsic method errors, yet it is a small enough basis to be feasible. The principle downside of performing a counterpoise correction when training a new functional is the differential

treatment of noncovalent interactions and thermochemistry; this could potentially be remedied by correcting for BSSE in a manner that allows for the consideration of intramolecular BSSE, such as by utilizing an atomic[201, 202] or geometric[191] correction. Such will be the focus of work to come. An additional suitable course of future work would be to extend this study beyond second-row elements, to see whether the observed trends among and within the different sequences of basis sets hold for heavier atoms and transition metals.

Chapter 5

Empirical Dispersion: DFT-D3

5.1 Introduction

Kohn-Sham density functional theory (DFT)[11] is the most widely used formalism in electronic structure theory, a consequence of its relative simplicity and the nice balance it strikes between computational expense and accuracy. Nevertheless, DFT has its drawbacks. Although there is a degree of hierarchy within DFT, as exemplified by the proverbial Jacob’s ladder of DFT[203], there is no prescription for systematically improving results. Moreover, standard approximations within the formalism are inherently semi-local, and hence incapable of correctly describing long-range electron correlation, i.e. strong correlations and dispersion forces[60]. This latter deficiency is particularly troubling, since dispersion is integral to the correct description of non-covalent interactions. To address this issue, a “stairway” of dispersion corrections has been constructed over the years[28].

The simplest such corrections can be traced back to a Hartree–Fock+D approach[204, 205], which, channeling second-order Rayleigh–Schrödinger perturbation theory, adds in a pairwise atomic correction involving empirical isotropic dispersion coefficients with the correct r^{-6} asymptote. This scheme was later adapted to DFT[140], which introduced an additional complication: since DFT already describes local electron correlation, the added +D component needs to be damped at small separations in order to avoid double counting. Grimme systematized this approach, first with his introduction of the DFT-D method[29], then subsequently DFT-D2[30] and DFT-D3[31]. These methods are widely used for the same reason that DFT is so prolific within the electronic structure community: they are simple, incredibly efficient, and quite accurate for a variety of interesting systems.

In recent years, a number of additional approaches to dispersion have been developed. Von Lilienfeld et al.[206] proposed adding in dispersion-corrected atom-centered potentials (DCACPs) within the effective core-potential approximation; this approach was later adapted to atom-centered basis sets in what is now known as the DCP approach[207]. Several approaches for self-consistently calculating dispersion coefficients have been introduced, most notably the exchange-dipole moment (XDM) model[32–35], and the TS-vdW method[36].

The past decade has even seen the proliferation of methods that attempt to explicitly incorporate non-local correlation, such as the vdW-DF[37] and vdW-DF2[38] approaches, which are popular within the solid-state community, and the VV09[39] and VV10[40] methods. For a more thorough description of these various methods, the reader is referred to a recent review by Klimeš and Michaelides[28].

Within this study, we focus on the most computationally inexpensive brand of dispersion corrections, the aforementioned DFT-D2 and DFT-D3 approaches. Specifically, we explore the effect of including an additional degree of freedom within the -D3 damping function, thereby introducing a new, more general damping function. This new damping function is then optimized for several popular density functionals, and the resulting methods are compared to those obtained with existing -D3 damping functions. We find that this new iteration, which we term DFT-D3(op), shorthand for optimized power, substantially improves the description of non-covalent interactions – particularly those involving molecular clusters – and isomerization energies.

5.2 Theory

Within the DFT-D family of methods, the two-body component of the empirical dispersion energy is given by

$$E^{(2)} = - \sum_{i < j} \sum_{n=6,8,10,\dots} s_n \frac{C_{n,ij}}{r_{ij}^n} f_{\text{damp},n}(r_{ij}). \quad (5.1)$$

The first sum in eq. (5.1) runs over all unique pairs of atoms i and j ; $C_{n,ij}$ are isotropic n th-order dispersion coefficients for atom pair ij ; r_{ij} is the internuclear distance between atoms i and j ; s_n are global, density functional-dependent scaling parameters; and $f_{\text{damp},n}(r_{ij})$ are damping functions intended to address small- r_{ij} singularities, as well as double-counting of correlation effects. Early iterations of these dispersion models – namely the original DFT-D[29, 140], as well as DFT-D2[30] – truncated the sum at $n = 6$, employed chemically-insensitive, pre-tabulated dispersion coefficients $C_{6,ij}$ and van der Waals radii $r_{0,ij}$, and utilized Fermi-type damping functions of the form

$$f_{\text{damp},6}^{\text{D2}}(r_{ij}) = \left[1 + \exp \left(-\alpha \left(\frac{r_{ij}}{r_{0,ij}} - 1 \right) \right) \right]^{-1}, \quad (5.2)$$

with α generally fixed to 20.

In 2010, Grimme et al.[31] introduced the now widely used DFT-D3 scheme. Although many aspects of -D3 are similar to -D2, there are some key fundamental differences: a counting function is introduced to allow the $C_{6,ij}$ coefficients to be weakly environmentally-dependent; the sum in eq. (5.1) is extended to include the $n = 8$ term, and a CHG-style[87] damping function – given in eq. (5.3) – is used. For the damping function, Grimme et al.[31] chose to fix $s_{r,8} = 1$, $\alpha_6 = 14$, and $\alpha_8 = 16$, thereby optimizing only one nonlinear parameter

– $s_{r,6}$ – for each density functional. Moreover, for almost all density functionals, s_6 is fixed to unity, leaving only one linear parameter, s_8 . This version of DFT-D3 is now known as zero-damping, or DFT-D3(0).

$$f_{\text{damp},n}^{\text{D3(0)}}(r_{ij}) = \left[1 + 6 \left(\frac{r_{ij}}{s_{r,n}r_{0,ij}} \right)^{-\alpha_n} \right]^{-1} \quad (5.3)$$

One year later, Grimme et al.[61] combined the basic principles of DFT-D3 with the finite-damping scheme Johnson and Becke[34] had utilized in their XDM approach to dispersion; this is now the generally preferred style of -D3, termed DFT-D3(BJ). The damping function is of the form

$$f_{\text{damp},n}^{\text{D3(BJ)}}(r_{ij}) = \frac{r_{ij}^n}{r_{ij}^n + (\alpha_1 r_{0,ij} + \alpha_2)^n}, \quad (5.4)$$

where α_1 and α_2 are adjustable nonlinear parameters. At short internuclear distances r_{ij} , the dispersion energy $E^{(2)}$ in the zero-damping approach vanishes, since $f_{\text{damp},n}(r_{ij})$ decays more quickly than $C_{n,ij}r_{ij}^{-n}$. In the BJ-damping scheme, however, these two terms decay at the same rate, and hence $E^{(2)}$ asymptotes to a finite value. This is the key difference between the -D3(0) and -D3(BJ) approaches. Although van der Waals radii $r_{0,ij}$ in -D3(BJ) are given by $\sqrt{\frac{C_{8,ij}}{C_{6,ij}}}$ instead of their -D3(0) values, the tabulated coefficients $C_{n,ij}$ are the same, and – as with -D3(0) – s_6 is generally fixed to unity in -D3(BJ), leaving s_8 as the sole linear parameter.

This new version of -D3 with BJ-damping has become the preferred version of -D3 due to the fact that it consistently outperforms -D3(0)[61]. Recently, Schröder et al.[208] have attempted to simplify the model with their C-Six-Only (CSO) approach, wherein they introduce a sigmoidal interpolation function to approximate the eighth-order term. In so doing, they eliminate one linear parameter and one nonlinear parameter without significantly impacting performance across GMTKN30 or S66[208]. The damping function for this approach, -D3(CSO), is given by

$$f_{\text{damp},6}^{\text{D3(CSO)}}(r_{ij}) = \frac{r_{ij}^6}{r_{ij}^6 + (\alpha_3 r_{0,ij} + \alpha_4)^6} \left[1 + \frac{\alpha_1}{s_6 [1 + \exp(r_{ij} - \alpha_2 r_{0,ij})]} \right]. \quad (5.5)$$

Note the similarities between eqs. (5.4) and (5.5): α_3 and α_4 in the CSO scheme correspond to α_1 and α_2 in BJ-damping, respectively, and the bracketed term in eq. (5.5) is the interpolation function. For the density functionals they examined, Schröder et al.[208] found $\alpha_3 \approx 0$, $\alpha_4 \approx 6.25$, and $\alpha_2 \approx 2.5$, leaving α_1 as the sole functional-dependent parameter in -D3(CSO).

In addition to the dispersion corrections mentioned thus far, we consider in this work the modified version of -D3(BJ) recently proposed by Smith et al.[209]. The damping function in this approach, -D3M(BJ), is identical to that in eq. (5.4); the method constitutes a refitting of the BJ-damping parameters to a much broader set of data. Grimme et al.[61] originally

fit s_8 , α_1 , and α_2 on 130 data points; Smith et al.[209] utilize a training set of 1526 energies, with an emphasis on non-equilibrium – particularly compressed – geometries.

It is illustrative to rewrite the damping function for -D3(0) from eq. (5.3). By doing so, we obtain

$$f_{\text{damp},n}^{\text{D3}(0)}(r_{ij}) = \frac{r_{ij}^{\alpha_n}}{r_{ij}^{\alpha_n} + (\sqrt[n]{6}s_{r,n}r_{0,ij})^{\alpha_n}}. \quad (5.6)$$

The similarities between the functional forms of -D3(0) and -D3(BJ) are striking (cf. eqs. (5.4) and (5.6)). There is a direct correspondence between α_1 in BJ-damping and $s_{r,n}$. Whereas in -D3(BJ) α_1 is the same for the sixth- and eighth-order terms, in -D3(0) this is no longer the case, as $s_{r,6} \neq s_{r,8}$. The original zero-damping scheme thus resembles a slightly constrained version of BJ-damping, wherein α_2 (the constant added to the van der Waals radii) is zero, and the sixth-order power is 14 instead of 6.

In this work, we generalize the BJ-damping function, adding a parameter to control how quickly the damping occurs, i.e. the power. The damping function we employ is given by

$$f_{\text{damp},n}^{\text{D3}(\text{op})}(r_{ij}) = \frac{r_{ij}^{\beta_n}}{r_{ij}^{\beta_n} + (\alpha_1 r_{0,ij} + \alpha_2)^{\beta_n}} \quad (5.7)$$

This new scheme – optimized-power-damping, or -D3(op) – is mathematically similar to both BJ-damping and zero-damping. As is the case with both BJ- and zero-damping, we constrain $\beta_8 = \beta_6 + 2$. We utilize the same isotropic dispersion coefficients as both -D3(0) and -D3(BJ), the same van der Waals radii as -D3(BJ) – i.e. $r_{0,ij} = \frac{C_{8,ij}}{C_{6,ij}}$ – and optimize three nonlinear parameters ($\alpha_1, \alpha_2, \beta = \beta_6$) and one linear parameter (s_8). The effects of varying the nonlinear parameters in this model are visualized in Figure 5.1. The holdovers from BJ-damping – α_1 and α_2 – primarily control the distance r_{ij} at which the damping function switches off the dispersion correction: α_1 and α_2 are just multiplicative and additive terms, respectively, for the sum of van der Waals radii $r_{0,ij}$. The newly introduced β , on the other hand, controls the rate at which the dispersion correction is switched off; in the limit $\beta \rightarrow \infty$, $f_{\text{damp}}(r_{ij}) \rightarrow \theta(r_{ij})$, i.e. the damping function just becomes a step function. Most small changes in β correspond to subtle changes in the dispersion energy, $E^{(2)}$; the exception is the transition from $\beta = 6$ to $\beta = 6 + \epsilon$, which is a fundamentally different change than that from, say, $\beta = 10$ to $\beta = 10 + \epsilon$. For all $\beta > 6$, the contribution of atom pair ij to the dispersion energy at $r_{ij} = 0$ is zero; for $\beta = 6$, this contribution is nonzero. In practice, however, this difference is not so significant; due to Pauli repulsion, the limit $r_{ij} = 0$ is not particularly meaningful.

An overview of the six forms of dispersion corrections considered in this study – -D2, -D3(0), -D3(BJ), -D3M(BJ), -D3(CSO), and -D3(op) – can be found in Table 5.1. Note we have not considered the modified version of zero-damping, -D3M(0), as it was found to be inferior to -D3M(BJ) by the original authors[209].

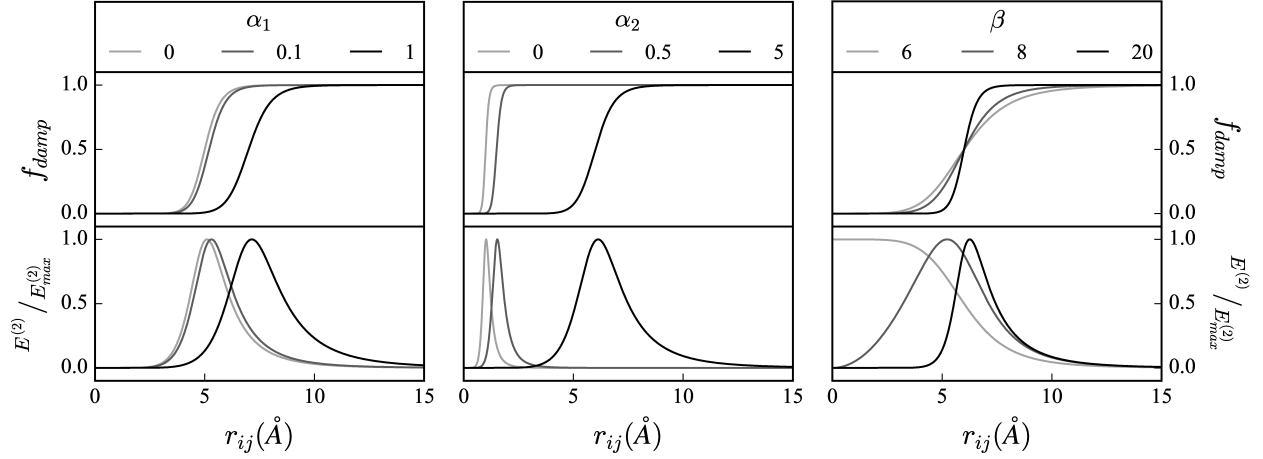


Figure 5.1: Visualization of the effects of varying parameters α_1 , α_2 , and β from eq. (5.7). The leftmost plots are generated by varying α_1 with fixed α_2 and β ; the center plots are generated by varying α_2 with fixed α_1 and β ; and the rightmost plots are generated by varying β with fixed α_1 and α_2 . When not being varied, the parameters are fixed to $\alpha_1 = 0.5$, $\alpha_2 = 5$, and $\beta = 14$, with $r_{0,ij} = 2$ Å. The top plot in each section shows the damping function for the sixth-order term, $f_{damp,6}^{D3(op)}(r_{ij})$ from eq. (5.7). The bottom plot in each section shows the sixth-order contribution to the two-body dispersion energy, $E^{(2)}$, which has been normalized to span the range $[0,1]$.

| Type | Fit Parameters | | $C_8?$ | Reference |
|----------|----------------|----------------------------------|--------|-----------|
| | Linear | Nonlinear | | |
| -D2 | s_6 | None | No | [30] |
| -D3(0) | s_8 | $s_{r,6}$ | Yes | [31] |
| -D3(BJ) | s_8 | α_1, α_2 | Yes | [61] |
| -D3M(BJ) | s_8 | α_1, α_2 | Yes | [209] |
| -D3(CSO) | None | α_1 | No | [208] |
| -D3(op) | s_6 or s_8 | $\alpha_1, \alpha_2,$ β | Maybe | This work |

Table 5.1: Summary of empirical dispersion corrections considered in this study.

5.3 Computational Details

We have optimized the DFT-D3(op) damping function given in eq. (5.7) for several density functional approximations and compared its performance to that of existing damping functions. Specifically, we have considered the ten density functionals outlined in Table 5.2.

| Functional | Class | Empirical Dispersion | Ref |
|------------|-----------------|-------------------------------------|------------------|
| BLYP | GGA | D2,D3(0),D3(BJ), D3M(BJ),D3(CSO) | [15, 16] |
| B3LYP | hybrid GGA | D2,D3(0),D3(BJ), D3M(BJ),D3(CSO) | [15, 16, 22, 82] |
| B97 | GGA | D2,D3(0),D3(BJ), D3M(BJ) | [30] |
| B97h | hybrid GGA | D2 | [17] |
| revPBE | GGA | D3(0),D3(BJ) | [14, 210] |
| revPBE0 | hybrid GGA | D3(0),D3(BJ) | [14, 83, 210] |
| TPSS | meta-GGA | D2,D3(0),D3(BJ), D3(CSO) | [84] |
| TPSSh | hybrid meta-GGA | D3(0),D3(BJ) | [18] |
| MS2 | meta-GGA | D3(0) | [211] |
| MS2h | hybrid meta-GGA | D3(0) | [211] |

Table 5.2: Summary of density functionals. The names in the first column are standard, with two exceptions: B97 corresponds to Grimme’s pure functional B97-D[30], which has had the dispersion tail stripped away; and B97h corresponds to the original hybrid functional B97, as parameterized by Becke[17]. The third column details the existing parameterized DFT-D-style dispersion corrections we consider in this study, and the fourth column lists the references for the method.

These ten representative density functionals were carefully chosen. There are five natural pairs of pure/hybrid functionals: BLYP/B3LYP, B97/B97h, revPBE/revPBE0, TPSS/TPSSh, and MS2/MS2h. The first three of these pairs are generalized gradient approximations (GGAs), and the last two are meta-GGAs. Moreover, each of these ten density functionals exhibits positive mean signed errors across every dataset of non-covalent interactions we considered; that is to say, they consistently underbind every type of system at which we have looked, and hence can all profit greatly from the addition of a dispersion correction. One popular functional we excluded from this study is PBE. Since PBE is known to overbind water clusters[212], we believe it is not a good candidate for a blanket dispersion correction. That being said, although PBE (and its complement, PBE0) are not found within the main study, parameterizations for both may be found in the Supporting Information.

All density functional calculations were performed in the def2-QZVPPD basis[154, 155], near the basis set limit for standard non-covalent interactions[43]. A fine Lebedev integration grid consisting of 99 radial shells – each with 590 angular points – was utilized in the computation of all semi-local components of exchange and correlation; non-local correlation in the VV10-containing functionals was calculated with the coarser SG-1 grid[133]. All calculations were performed within a development version of Q-Chem 4.4[182].

For each density functional, we performed an exhaustive determination of the optimal parameters for the DFT-D3(op) method. Specifically, we scanned α_1 from 0 to 1 in steps of 0.025, α_2 from 0 to 10 in increments of 0.25, and β from 6 to 18 in increments of 2; this resulted in 11767 possible forms of the -D3(op) tail for each density functional.

To identify the best of these many candidate fits, we utilized the comprehensive database assembled by Mardirossian and Head-Gordon[24]. This database contains 4419 data points which are spread out among 82 smaller datasets. These smaller constituent datasets are classified according to eight distinct datatypes: NCED (easy non-covalent interactions of dimers), NCEC (easy non-covalent interactions of clusters), NCD (difficult non-covalent interactions of dimers), IE (easy isomerization energies), ID (difficult isomerization energies), TCE (easy thermochemistry), TCD (difficult thermochemistry), and BH (barrier heights). “Difficult” interactions involve either strong correlation or self-interaction error, whereas “easy” interactions are not heavily characterized by either. In order to facilitate the testing of -D3(op) candidates, the datasets were divided into two categories. A training set was used to identify the best set of parameters, and a test set was used to assess the performance of the resulting method relative to existing dispersion corrections. A summary of the datasets can be found in Table 5.3.

Once all 11767 possible fits were generated for a given functional, we set s_6 to unity and performed a least-squares fit of s_8 to a small subset of NCEDTrain and IETrain, namely S66 and Butanediol65. The resulting methods were then sorted according to a simple product of root-mean-square errors (RMSEs) across eleven training datasets – S66, HB49, AlkBind12, H2O6Bind8, H2O20Bind4, HW6Cl, Butanediol65, Melatonin52, H2O16Rel5 and two geometric datasets: interpolated equilibrium binding lengths of BzDC215 and NBC10. To prevent the appearance of unphysical parameters (negative values of s_8), when s_8 optimized to a value less than 0.1, we set $s_8 = 0$ and performed a least-squares fit of s_6 instead. The only functional within this study for which this occurred is B97h.

5.4 Results and Discussion

In the course of this study, we have introduced a new damping function for use in the DFT-D3 empirical dispersion correction – DFT-D3(op) – which encompasses the space spanned by DFT-D3(BJ) and an unconstrained form of DFT-D3(0). We have optimized this new damping function for ten distinct density functionals across a small yet diverse training set. The resulting optimized fit parameters are listed in Table 5.4; parameterizations for additional functionals may be found in Appendix C.

This new variant of DFT-D3 was then evaluated in tandem with existing versions across a large independent test set. The results are given in Figure 5.2, within which we show for each method the root-mean-square errors (RMSEs) across the three energetic categories for which empirical dispersion methods are best suited – namely NCEDTest, NCECTest, and IETest – as well as one geometric category, S66x8 interpolated equilibrium binding lengths (and corresponding interpolated equilibrium binding energies). For each of NCEDTest, NCECTest,

| Set | Datatype | # | Constituent Datasets | References |
|-------|----------|------|--|--|
| Train | NCED | 127 | S66, HB49, AlkBind12 | [116, 213–217] |
| | NCEC | 18 | H2O6Bind8, H2O20Bind4, HW6C1 | [218–223] |
| | IE | 122 | Butanediol65, Melatonin52, H2O16Rel5 | [224–226] |
| | BL | 20 | Interpolated equilibrium binding lengths from BzDC215 and NBC10 | [119, 121, 186, 227, 228] |
| Test | NCED | 1617 | A24, DS14, HB15, HSG, NBC10, S22, X40, A21x12, BzDC215, HW30, NC15, S66x8, 3B-69-DIM, CO2Nitrogen16, Ionic43 | [42, 117, 119, 121, 131, 181, 186, 213, 227–236] |
| | NCEC | 225 | HW6F, FmH2O10, Shields38, SW49Bind345, SW49Bind6, WATER27, 3B-69-TRIM, CE20, H2O20Bind10 | [218–221, 234, 237–240] |
| | NCD | 91 | TA13, XB18, Bauza30, CT20, XB51 | [241–245] |
| | IE | 633 | AlkIsomer11, ACONF, CYCONF, Pentane14, SW49Rel345, SW49Rel6, H2O20Rel10, H2O20Rel4, YMPJ519 | [219–223, 238, 246–250] |
| | ID | 155 | EIE22, Styrene45, DIE60, ISOMERIZATION20, C20C24 | [239, 251–253] |
| | TCE | 947 | AlkAtom19, BDE99nonMR, G21EA, G21IP, TAE140nonMR, AlkIsod14, BH76RC, EA13, HAT707nonMR, IP13, NBPRC, SN13, BSR36, HNBrBDE18, WCPT6 | [188, 221, 246, 252, 254–261] |
| | TCD | 258 | BDE99MR, HAT707MR, TAE140MR, PlatonicHD6, PlatonicID6, PlatonicIG6, PlatonicTAE6 | [252, 262] |
| | BH | 206 | BHPERI26, CRBH20, DBH24, CR20, HTBH38, NHTBH38, PX13, WCPT27 | [221, 239, 240, 255, 256, 261, 263–267] |

Table 5.3: Summary of datasets that comprise the training and test sets. For more details, see Ref. 24.

| Functional | s_6 | s_8 | α_1 | α_2 | β |
|------------|---------|---------|------------|------------|---------|
| BLYP | 1.00000 | 1.31867 | 0.425 | 3.50 | 8 |
| B3LYP | 1.00000 | 0.78311 | 0.300 | 4.25 | 10 |
| B97 | 1.00000 | 1.46861 | 0.600 | 2.50 | 6 |
| B97h | 0.97388 | 0.00000 | 0.150 | 4.25 | 12 |
| revPBE | 1.00000 | 1.44765 | 0.600 | 2.50 | 6 |
| revPBE0 | 1.00000 | 1.25684 | 0.725 | 2.25 | 6 |
| TPSS | 1.00000 | 0.51581 | 0.575 | 3.00 | 14 |
| TPSSh | 1.00000 | 0.43185 | 0.575 | 3.00 | 14 |
| MS2 | 1.00000 | 0.90743 | 0.700 | 4.00 | 8 |
| MS2h | 1.00000 | 1.69464 | 0.650 | 4.75 | 6 |

Table 5.4: Optimized values of -D3(op) fit parameters for each density functional.

and IETest, both the aggregate results as well as results for two representative constituent datasets are provided.

From Figure 5.2, it is evident that the newly proposed -D3(op) dispersion correction represents an improvement over existing corrections for a diverse set of density functionals and systems. Even in cases where the power optimizes to $\beta = 6$, which corresponds to the DFT-D3(BJ) scheme, we see large improvements in some categories. This is the case, for instance, for revPBE. Relative to revPBE-D3(BJ), revPBE-D3(op) exhibits significantly lower RMSEs across all metrics related to non-covalent interactions; the reductions in errors on molecular clusters (NCEC, 63%) and geometries (BL, 45%) are particularly striking. This sort of improvement highlights the benefits a simple re-optimization incorporating the plethora of new high quality data that have been published in the past six years can bring; after all, revPBE-D3(op) is effectively just a reparameterization of revPBE-D3(BJ). Note that for this particular functional, there is no DFT-D3M(BJ) version with which to compare our reparameterization; Smith et al.[209] chose to parameterize PBE instead of revPBE.

Another manifestation of the importance of having a well-balanced training set is the performance of DFT-D3(CSO). In a recent study, it was found that DFT-D3(CSO) reproduces bond lengths and rotational constants quite well[268]. Here, we find that this satisfactory intramolecular performance does not transfer to intermolecular metrics, despite the fact that DFT-D3(CSO) is decent at reproducing accurate energies (particularly isomerization energies, at which it seems to excel). Regardless of density functional, -D3(CSO) exhibits significantly larger errors in equilibrium binding lengths than any of the other iterations of -D3; for B3LYP-D3(CSO), for instance, the RMSE across BL is nearly double the next-highest B3LYP-D3 BL RMSE. This is likely a consequence of the fact that DFT-D3(CSO) was trained exclusively on equilibrium systems, specifically S66. This is in stark contrast to other DFT-D3 variants, which were trained on geometries either implicitly through the inclusion of nonequilibrium systems – as was the case for -D3(0), -D3(BJ), and -D3M(BJ) – or by explicitly including geometries in the training metric, as is the case here for -D3(op).

Similarly, deviations among performances across single datasets highlight the need for diverse aggregate datasets. From the exemplary performance of BLYP-D3(CSO) on the Shields38 set, one might infer that this method should be particularly well-suited to water clusters. This is not the case, however; BLYP-D3(CSO) has the highest RMSE of all BLYP-D3 variants across the H2O20Bind10 set of large water clusters by nearly a factor of two. Oftentimes, outstanding performance of a particular method on a particular set of systems will transfer to similar systems, but that sort of transferability is not guaranteed. The larger and more diverse the independent test set, the more likely the results will be transferable.

This point is further driven home by the relative performances of the DFT-D3(BJ) and -D3M(BJ) methods. For instance, compare B3LYP-D3(BJ) and B3LYP-D3M(BJ) in Figure 5.2. The -D3M(BJ) method is simply a reparameterization of -D3(BJ) across a significantly larger training set. In the case of B3LYP-D3M(BJ), however, this reparameterization has resulted in significant loss of performance as measured by RMSEs across equilibrium non-covalent interactions. This is likely a consequence of the large emphasis placed on compressed geometries in the parameterization of -D3M(BJ) by both the composition of the datasets

| Functional | Non-Covalent: Dimers | | | Non-Covalent: Clusters | | | Isomerization | | | Equilibrium | | Overall |
|----------------|----------------------|-------------|-------------|------------------------|-------------|-------|---------------|-------------|-------------|--------------|-------------|-------------|
| | NCED* | HSG | S22 | NCEC* | Shields | 3B-69 | IE* | Pent14 | YMPJ | BL (Å) | BE | Product |
| BLYP | 3.46 | 3.39 | 6.00 | 14.20 | 7.82 | 8.39 | 1.58 | 0.55 | 1.51 | 0.225 | 3.80 | |
| -D2 | 0.54 | 0.48 | 0.25 | 2.13 | 3.23 | 0.73 | 1.17 | 0.83 | 1.18 | 0.095 | 0.58 | 12.85 |
| -D3(0) | 0.36 | 0.39 | 0.28 | 1.67 | 2.52 | 0.52 | 0.83 | 0.20 | 0.80 | 0.028 | 0.42 | 1.41 |
| -D3(BJ) | 0.33 | 0.40 | 0.32 | 2.07 | 1.59 | 0.71 | 0.71 | 0.48 | 0.70 | 0.027 | 0.23 | 1.32 |
| -D3M(BJ) | 0.33 | 0.29 | 0.34 | 2.57 | 3.81 | 0.69 | 0.72 | 0.45 | 0.71 | 0.038 | 0.31 | 2.33 |
| -D3(CSO) | 0.40 | 0.48 | 0.38 | 3.21 | 0.61 | 0.80 | 0.68 | 0.49 | 0.65 | 0.061 | 0.29 | 5.29 |
| -D3(op) | 0.31 | 0.40 | 0.20 | 1.77 | 2.24 | 0.75 | 0.73 | 0.41 | 0.74 | 0.029 | 0.21 | 1.14 |
| B3LYP | 2.82 | 2.68 | 4.89 | 8.67 | 4.54 | 6.61 | 1.34 | 0.47 | 1.28 | 0.215 | 3.12 | |
| -D2 | 0.58 | 0.45 | 0.61 | 5.19 | 5.14 | 0.80 | 0.89 | 0.54 | 0.88 | 0.100 | 0.81 | 26.63 |
| -D3(0) | 0.35 | 0.27 | 0.42 | 3.69 | 4.00 | 0.79 | 0.50 | 0.13 | 0.45 | 0.027 | 0.44 | 1.71 |
| -D3(BJ) | 0.30 | 0.21 | 0.43 | 2.67 | 3.21 | 0.74 | 0.51 | 0.20 | 0.49 | 0.022 | 0.38 | 0.89 |
| -D3M(BJ) | 0.34 | 0.19 | 0.52 | 3.32 | 3.79 | 0.80 | 0.50 | 0.19 | 0.49 | 0.023 | 0.44 | 1.32 |
| -D3(CSO) | 0.30 | 0.25 | 0.30 | 1.51 | 2.15 | 0.63 | 0.48 | 0.20 | 0.44 | 0.047 | 0.28 | 1.01 |
| -D3(op) | 0.26 | 0.30 | 0.25 | 1.28 | 1.88 | 0.56 | 0.55 | 0.20 | 0.54 | 0.027 | 0.24 | 0.49 |
| B97 | 3.66 | 3.70 | 6.49 | 21.42 | 15.11 | 9.30 | 1.87 | 0.64 | 1.77 | 0.239 | 3.86 | |
| -D2 | 0.57 | 0.72 | 0.62 | 5.63 | 3.66 | 1.42 | 1.28 | 0.66 | 1.33 | 0.068 | 0.53 | 27.50 |
| -D3(0) | 0.39 | 0.50 | 0.54 | 4.12 | 2.22 | 1.17 | 0.76 | 0.08 | 0.77 | 0.034 | 0.35 | 4.16 |
| -D3(BJ) | 0.46 | 0.55 | 0.49 | 4.75 | 2.75 | 1.11 | 0.82 | 0.28 | 0.82 | 0.047 | 0.38 | 8.40 |
| -D3M(BJ) | 0.40 | 0.40 | 0.38 | 2.75 | 0.94 | 1.17 | 0.61 | 0.19 | 0.60 | 0.032 | 0.30 | 2.14 |
| -D3(op) | 0.41 | 0.39 | 0.38 | 1.83 | 0.42 | 1.08 | 0.72 | 0.30 | 0.70 | 0.026 | 0.29 | 1.38 |
| B97h | 2.41 | 2.26 | 4.39 | 10.42 | 6.50 | 5.77 | 1.27 | 0.41 | 1.19 | 0.218 | 2.56 | |
| -D2 | 0.49 | 0.39 | 0.71 | 1.14 | 0.44 | 0.93 | 0.72 | 0.42 | 0.73 | 0.045 | 0.40 | 1.80 |
| -D3(op) | 0.29 | 0.33 | 0.44 | 1.29 | 0.30 | 0.55 | 0.56 | 0.10 | 0.59 | 0.034 | 0.26 | 0.71 |
| revPBE | 3.61 | 3.68 | 6.45 | 21.85 | 14.82 | 9.25 | 1.76 | 0.60 | 1.66 | 0.238 | 3.88 | |
| -D3(0) | 0.41 | 0.56 | 0.61 | 5.50 | 2.71 | 1.12 | 0.81 | 0.08 | 0.84 | 0.037 | 0.42 | 6.67 |
| -D3(BJ) | 0.48 | 0.60 | 0.62 | 6.56 | 3.51 | 1.25 | 0.81 | 0.39 | 0.81 | 0.056 | 0.43 | 14.08 |
| -D3(op) | 0.42 | 0.39 | 0.40 | 2.44 | 0.42 | 1.06 | 0.75 | 0.41 | 0.74 | 0.031 | 0.30 | 2.36 |
| revPBE0 | 3.08 | 3.14 | 5.45 | 16.52 | 12.07 | 7.78 | 1.51 | 0.58 | 1.41 | 0.229 | 3.38 | |
| -D3(0) | 0.32 | 0.34 | 0.36 | 2.14 | 1.56 | 0.53 | 0.59 | 0.26 | 0.62 | 0.032 | 0.40 | 1.32 |
| -D3(BJ) | 0.32 | 0.37 | 0.32 | 3.44 | 2.53 | 0.77 | 0.56 | 0.09 | 0.56 | 0.037 | 0.23 | 2.26 |
| -D3(op) | 0.33 | 0.26 | 0.39 | 0.83 | 0.33 | 0.80 | 0.56 | 0.09 | 0.55 | 0.024 | 0.29 | 0.37 |
| TPSS | 2.53 | 2.49 | 4.61 | 8.81 | 3.64 | 6.48 | 1.22 | 0.47 | 1.08 | 0.217 | 2.90 | |
| -D2 | 0.63 | 0.48 | 0.64 | 4.74 | 5.67 | 0.89 | 1.07 | 0.89 | 1.03 | 0.118 | 0.77 | 37.47 |
| -D3(0) | 0.34 | 0.25 | 0.44 | 2.75 | 3.90 | 0.74 | 0.67 | 0.38 | 0.53 | 0.059 | 0.35 | 3.73 |
| -D3(BJ) | 0.36 | 0.26 | 0.47 | 1.90 | 2.85 | 0.96 | 0.70 | 0.55 | 0.57 | 0.074 | 0.34 | 3.51 |
| -D3(CSO) | 0.38 | 0.25 | 0.43 | 1.85 | 2.69 | 0.90 | 0.69 | 0.56 | 0.59 | 0.075 | 0.33 | 3.62 |
| -D3(op) | 0.32 | 0.31 | 0.36 | 1.76 | 2.54 | 0.87 | 0.78 | 0.45 | 0.67 | 0.053 | 0.28 | 2.28 |
| TPSSh | 2.46 | 2.42 | 4.43 | 8.33 | 3.97 | 6.24 | 1.19 | 0.40 | 1.06 | 0.214 | 2.82 | |
| -D3(0) | 0.34 | 0.20 | 0.49 | 2.59 | 3.35 | 0.74 | 0.56 | 0.24 | 0.46 | 0.056 | 0.36 | 2.77 |
| -D3(BJ) | 0.35 | 0.23 | 0.44 | 1.49 | 2.13 | 0.86 | 0.63 | 0.42 | 0.53 | 0.072 | 0.33 | 2.37 |
| -D3(op) | 0.30 | 0.27 | 0.35 | 1.41 | 2.02 | 0.80 | 0.69 | 0.31 | 0.59 | 0.048 | 0.28 | 1.41 |
| MS2 | 1.24 | 1.28 | 2.36 | 4.83 | 2.77 | 3.48 | 0.68 | 0.29 | 0.67 | 0.126 | 1.70 | |
| -D3(0) | 0.38 | 0.38 | 0.41 | 1.38 | 1.24 | 0.62 | 0.49 | 0.09 | 0.41 | 0.034 | 0.43 | 0.86 |
| -D3(op) | 0.29 | 0.19 | 0.43 | 0.91 | 0.29 | 0.50 | 0.35 | 0.11 | 0.33 | 0.039 | 0.23 | 0.36 |
| MS2h | 1.27 | 1.31 | 2.36 | 4.49 | 2.84 | 3.50 | 0.68 | 0.34 | 0.67 | 0.130 | 1.72 | |
| -D3(0) | 0.34 | 0.33 | 0.34 | 1.68 | 1.14 | 0.59 | 0.43 | 0.12 | 0.34 | 0.033 | 0.39 | 0.80 |
| -D3(op) | 0.26 | 0.13 | 0.32 | 0.56 | 0.21 | 0.43 | 0.27 | 0.14 | 0.24 | 0.027 | 0.20 | 0.11 |

Figure 5.2: Root-mean-square errors across various datasets for all combinations of functionals and dispersion corrections examined. Abbreviations of the dataset names are used; the full names are, in order, NCEDTest, HSG, S22, NCECTest, Shields38, 3B-69-TRIM, IETest, Pentane14, YMPJ519, S66x8 BL, and S66x8 BE. The Product column corresponds to a simple product across the aggregate dataset RMSEs (times 100). All RMSEs are in units of kcal/mol, with the exception of the equilibrium binding lengths (BL), which are in units of angstroms. Each column within every functional block is color-coded for ease of reading, with darker cells corresponding to lower RMSEs; note, however, that the color-gradient is not uniform, but rather skewed to emphasize significant differences among the top fits for each functional. Additionally, the lowest RMSE achieved by any functional-dispersion combination on each aggregate dataset is indicated in bold.

and the unique error metric employed.

A consistent feature of functionals employing the DFT-D3(op) tail is enhanced performance on molecular clusters and geometries, as illustrated by low RMSEs across the NCECTest and S66x8 BL sets. For instance, B97-D3(op) performs similarly to B97-D3(BJ) and B97-D3(0) across NCEDTest and IETest, but with dramatic reduction in error across NCECTest and interpolated S66x8 binding lengths. Even in cases where performance on molecular dimers is significantly improved, e.g. B3LYP-D3(op) and MS2h-D3(op), we still see solid performances on clusters and geometries. Whereas with other dispersion tails, performance on molecular clusters is very dependent on the choice of base functional, the -D3(op) approach consistently captures cluster binding energies quite well. For instance, from Figure 5.2, it is clear that BLYP-D3(0) is the best BLYP-based approach examined on NCECTest, even beating out BLYP-D3(op) by 0.10 kcal/mol; on the other hand, B97-D3(0) has a NCECTest RMSE more than double that of B97-D3(op). This high degree of consistency in the new correction, -D3(op), is likely attributable to the incorporation of molecular clusters in the training set. It is worth noting that a three-body correction of the form suggested by Grimme et al.[31] was also considered in this study; its inclusion, however, did not impact results in any meaningful way. It is quite possible that for larger systems, a three-body term would become relevant.

The one category for which we consistently see the least improvement with the new DFT-D3(op) correction is IETest, i.e. isomerization energies. For certain functionals – most notably MS2 and MS2h – the -D3(op) tail actually improves performance on IETest dramatically. For others, such as TPSS and B3LYP, the new approach is slightly worse than the existing tails – though still significantly better than the base functionals – for isomerization energies. There is inevitably some degree of trade-off between performances across these various properties; of the 11767 combinations of parameters we examined for any given functional, there were certain fits which excelled at one particular category. Although in theory it would be possible to recommend specialized fits for each functional which are best suited for a particular property – B3LYP-D3(op,NCED), B3LYP-D3(op,NCEC), B3LYP-D3(op,IE), etc. – that sort of approach has limited utility. It is far simpler to recommend a parameterization that provides a balanced description of all relevant properties. Comparing B3LYP-D3(op) and B3LYP-D3(BJ), for instance, we see the former has 8% higher RMSE across IETest; however, this slight reduction in performance for isomerization energies is compensated by a 13% reduction in RMSE across NCEDTest and a 52% reduction in RMSE across NCECTest.

Based on the data in Figure 5.2, we can make recommendations regarding which fit should be used with each density functional. As a metric of optimality, we use a simple product of the RMSEs across each of the aggregate test sets in Figure 5.2, i.e. the Product column. It is worth noting that -D2 dramatically underperforms relative to all variants of -D3 for all functionals examined. That being said, even the -D2 corrections represent a major improvement over the base functional. For BLYP, B3LYP, TPSS, and TPSSh, the DFT-D3(op) correction is best, followed by -D3(BJ); for B97, the DFT-D3(op) correction is best, followed by -D3M(BJ); and for revPBE, revPBE0, MS2, and MS2h, DFT-D3(op) is

best, followed by -D3(0).

In a similar vein, we can use the data from Figure 5.2 to make recommendations for the “best” functional/tail combination within each rung of Jacob’s ladder. The top performer among pure GGAs is BLYP-D3(op); at the hybrid GGA level, revPBE0-D3(op); from the pure meta-GGAs, MS2-D3(op); and at the hybrid meta-GGA rung, MS2h-D3(op). Based on the data of Figure 5.2, these four functional/tail groupings provide the best, most balanced descriptions of non-covalent interactions and isomerization energies, the two datatypes which empirical dispersion corrections should be capable of improving.

Thus far, we have established two major points: first, when a functional consistently underpredicts intermolecular interactions, the addition of an empirical dispersion correction can greatly improve results; and second, among the dispersion corrections examined, for the systems and density functionals we considered, DFT-D3(op) is best, regardless of the choice of base functional.

To put these results in broader context, we can compare these “best-in-class” DFT-D3 functionals with state-of-the-art density functionals corresponding to the same rungs of Jacob’s ladder. In Figure 5.3, the RMSEs across all eight categories of Table 5.3 are plotted and tabulated for revPBE0-D3(op) and three other hybrid GGAs: ω B97X[86], ω B97X-D[87], and ω B97X-V[23]. The left side of Figure 5.3, (a), contains NCED, NCEC, and IE – the three categories a D3 correction is capable of improving – while the right side, (b), encompasses the remaining datatypes: difficult cases characterized by strong correlation or self-interaction error, as well as thermochemistry and barrier heights. Although revPBE0-D3(op) represents a substantial improvement over ω B97X as far as non-covalent interactions are concerned, the method is somewhat lacking when it comes to the other datatypes: its performance on standard thermochemical problems (TCE) is particularly lackluster. That being said, it represents a decent alternative to ω B97X-D, and even rivals ω B97X-V for non-covalent interactions.

The same sort of comparison is made at the pure meta-GGA level in Figure 5.4. Therein, the top performer from this study, MS2-D3(op), is compared to M06-L[20], TM[269], and B97M-rV[19, 24]. Here, it seems MS2-D3(op) is actually a viable alternative to standard meta-GGAs; the method generally outperforms both M06-L and TM across the 4419 systems examined. We see a 47% reduction in RMSE across NCED, a 60% reduction in error across NCEC, and a 51% reduction in RMSE across IE for MS2-D3(op) relative to M06-L. This strikes a stark contrast to the performance of the uncorrected MS2 functional, which, relative to M06-L, has a NCED RMSE of 1.27 kcal/mol, a NCEC RMSE of 4.99 kcal/mol, and an IE RMSE of 0.76 kcal/mol; without the DFT-D3(op) correction, MS2 is significantly worse than M06-L across all three categories. This is certainly a testament to the utility of dispersion corrections. That being said, the exemplary performance of MS2-D3(op) – particularly on non-covalent interactions and isomerization energies – is eclipsed by that of B97M-rV, which benefits from the inclusion of non-local van der Waals correlation.

As a final example, three popular hybrid meta-GGA functionals – M06-2X[25], MN15[26], and ω B97M-V[24] – are compared head-to-head with MS2h-D3(op) in Figure 5.5. Although the DFT-D3(op) dispersion correction allows MS2h to significantly outperform M06-2X and

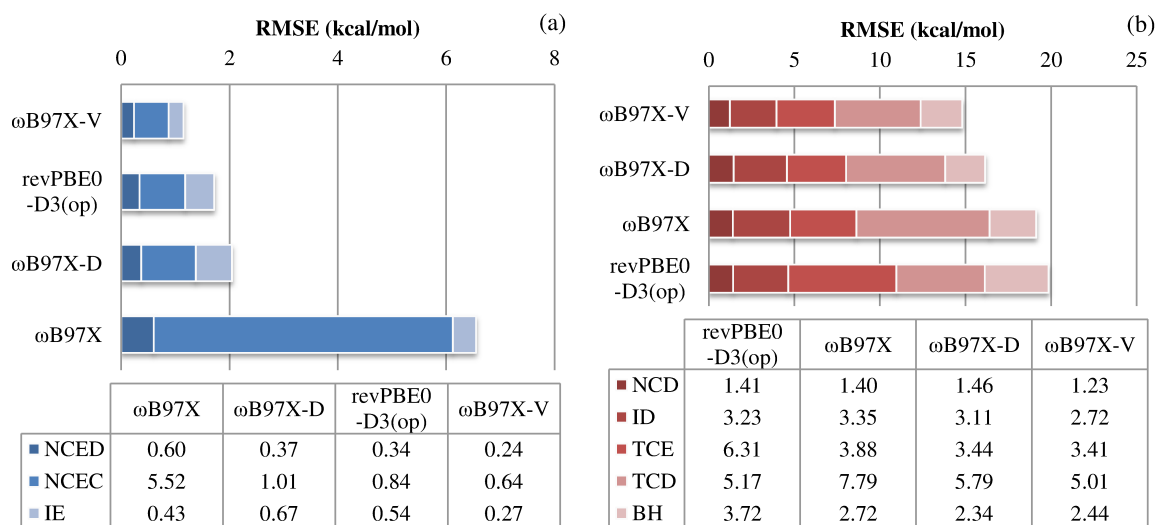


Figure 5.3: Root-mean-square errors (RMSEs) across various datatypes (see Table 5.3) for the top-performing D3-corrected hybrid GGA functional examined in this work, as well as three other state-of-the-art hybrid GGA functionals. The left plot – (a) – contains datasets pertaining to standard non-covalent interactions and isomerization energies, whereas the right plot – (b) – encompasses other data, such as thermochemistry, which is beyond the realm of a D3 correction. Within each plot, the methods are ordered from best performance across the constituent datasets at the top, to worst performance at the bottom. Tables of RMSEs are provided below the bar graphs to facilitate quantitative comparison.

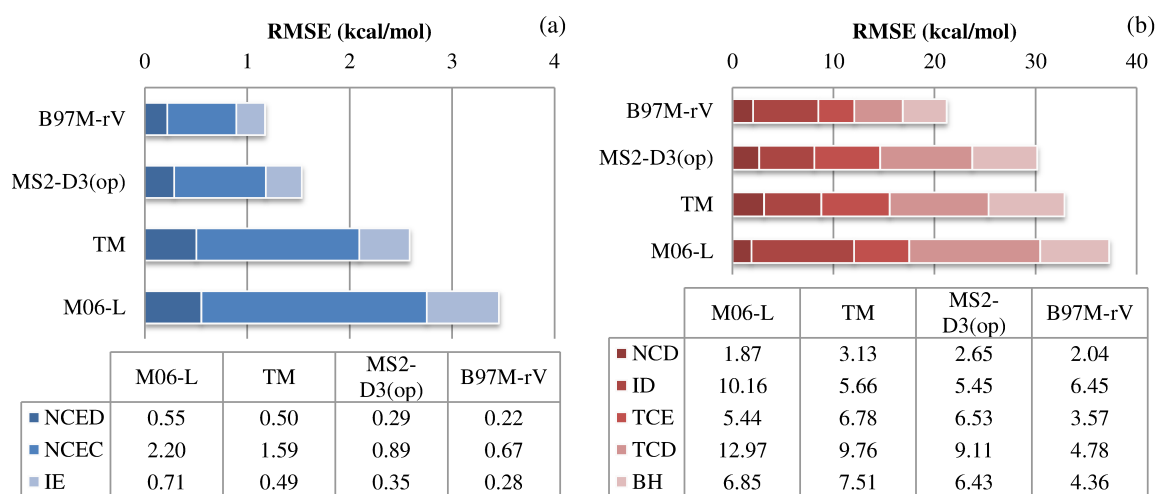


Figure 5.4: Root-mean-square errors (RMSEs) across various datatypes (see Table 5.3) for the top-performing D3-corrected pure meta-GGA functional examined in this work, as well as three other state-of-the-art pure meta-GGA functionals. For further details, refer to the caption of Figure 5.3.

MN15 for standard intermolecular binding energies and isomerization energies – and even come close to the stellar performance of ω B97M-V – it can do nothing to rectify the other deficiencies of the method, which manifest themselves in poor performance across the TCE and BH categories in particular. At the hybrid meta-GGA level, the top performer is unambiguously ω B97M-V.

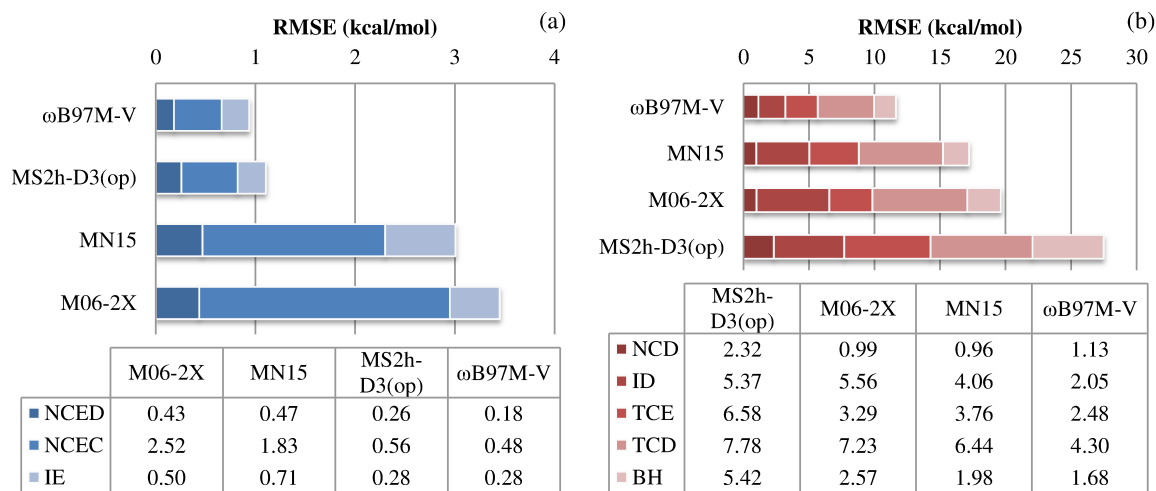


Figure 5.5: Root-mean-square errors (RMSEs) across various datatypes (see Table 5.3) for the top-performing D3-corrected hybrid meta-GGA functional examined in this work, as well as three other state-of-the-art pure hybrid meta-GGA functionals. For further details, refer to the caption of Figure 5.3.

5.5 Discussion and Conclusion

In this study, we have introduced a new damping function for the DFT-D3 brand of empirical dispersion corrections. This correction is effectively a generalization of DFT-D3(BJ) wherein the power is treated as an additional parameter, and is accordingly named DFT-D3(op), for “optimized-power”-damping. We have parameterized this method for ten distinct density functionals and compared its performance across an external test set to that of existing forms of DFT-D3.

This new approach, -D3(op), consistently yields substantial improvements in the descriptions of molecular clusters, regardless of the base functional with which it is paired. Moreover, it provides a well-balanced description of intermolecular binding energies, equilibrium geometries, and isomerization energies – the three broad classes of data that an empirical dispersion correction can reasonably be expected to improve. Unfortunately, the DFT-D3 correction is not intended to address deficiencies in other aspects – for instance, description of thermochemical properties – of the base functionals to which it is applied. From Figures

5.3, 5.4, and 5.5, it is apparent that there exists at each level of Jacob’s ladder beyond pure GGA at least one functional which will outperform the best comparable DFT-D3 method. This is not particularly surprising; after all, this isn’t a very fair comparison. Those better methods were built from the ground up; all their components were optimized together, self-consistently. If anything, it is remarkable just how well this sort of simple post-SCF dispersion correction performs.

The most impressive methods – namely ω B97X-V, B97M-rV, and ω B97M-V – explicitly include non-local correlation, which is a more robust, more powerful treatment of dispersion[104, 270]. Similarly, it might be the case that other self-consistent approaches, such as dispersion-corrected potentials[206] and TS-vdW,[271] will outperform the methods introduced here. Such comparisons are beyond the scope of this particular study; the aim here has been simply to introduce an updated, more accurate version of the most computationally economical form of dispersion correction, DFT-D3. In this particular endeavor, we have been successful: for all density functionals examined, DFT-D3(op) yields the best, most balanced performance across the available DFT-D3 dispersion tails.

Specifically, among the pure GGAs examined, BLYP-D3(op) is the stand-out candidate, offering an unparalleled account of non-covalent interactions. At the hybrid GGA level, revPBE0-D3(op) shines. As far as pure and hybrid meta-GGA functionals go, MS2-D3(op) and MS2h-D3(op) are both quite impressive. For these latter two functionals in particular, the new DFT-D3(op) tail represents a major improvement across all energetic categories over the existing -D3(0) versions. These DFT-D3(op) methods have a significant cost advantage over those incorporating VV10 non-local correlation. Moreover, they make relatively small sacrifices on performance across non-covalent interactions and isomerization energies, which makes them well suited for certain applications, e.g. calculations on large or condensed-phase systems, and molecular dynamics.

Chapter 6

Modeling Basis Set Superposition Error

6.1 Introduction

In an electronic structure calculation, two forms of basis set errors arise when local basis sets are employed: basis set superposition error (BSSE), which is a consequence of inconsistent treatment of a larger supersystem and its constituent subsystems[91, 158], and intrinsic basis set incompleteness error, the category to which we relegate all remaining basis set errors once BSSE has been removed. Intrinsic incompleteness error arises from the fact that the Schrödinger equation is being solved in just a fraction of the full Hilbert space, and no systematic means of removal – short of simply increasing the number of basis functions – has yet been discovered, though adaptive-basis approaches have shown some promise[272–277]. Basis set superposition error, on the other hand, has a long history within the electronic structure community[160–169]. In the case of distinct non-covalently interacting units, BSSE can be removed by performing fragment calculations within the basis of the full system, i.e. via the counterpoise correction (CP) first introduced by Boys and Bernardi[91].

The standard counterpoise correction has two principle shortcomings. First, it requires a partitioning of the full system into a number of fragments, $N_{\text{fragments}}$; for some systems, such as those with simple bimolecular interactions, this partitioning is straightforward, but for many interesting systems – such as those involving substantial intramolecular interactions – it is not. Second, although in principal a good approximation to counterpoise-corrected results may be obtained with minimal extra effort via standard energy decomposition analyses[278, 279], in practice the CP correction often ends up being quite computationally demanding: whereas an uncorrected binding energy requires only one calculation in the full supersystem basis, a counterpoise-corrected one requires $N_{\text{fragments}} + 1$ such calculations.

The issues of partitioning and the inability of the CP scheme to address intramolecular BSSE were first addressed by Galano and Alvarez-Idaboy with an atom-by-atom counterpoise correction[201]; Jensen later generalized this into the atomic counterpoise (ACP- n)

approach[202]. In the ACP- n scheme, BSSE is estimated as a sum of atomic BSSEs, where each atomic BSSE is calculated by considering basis functions up to n bonded atoms away. This approach has shown some promise in addressing intramolecular BSSE, though it suffers from the same partitioning problem as CP when ambiguous bonding patterns are involved – e.g. in transition states and hydrogen-bonded systems – and the computational complexity of the method is unchanged.

More recently, there have been attempts to develop empirical models for BSSE, as such approaches can potentially address both the partitioning and complexity issues. The first such model was proposed six years ago by Faver and Merz[189, 190], who constructed so-called “proximity functions” for molecular fragments from atomic pairs. Since the targets for this method are large biomolecules, the parameters are trained on a variety of proteinogenic systems. To date, this is the only empirical correction for BSSE developed for correlated wavefunction-based methods. The chief shortfall of the approach lies in its limited transferability: the parameters for modeling typical nonpolar, van der Waals-driven interactions are significantly different than those used for modeling hydrogen bonding.

Kruse and Grimme more recently introduced the so-called geometrical counterpoise (gCP) scheme[191], which was later adapted to the HF-3c, PBEh-3c, and HSE-3c methods[192, 280, 281]. The gCP scheme loosely resembles the proximity function approach of Faver and Merz, inasmuch as both methods are strictly pairwise atomic corrections. Unlike the proximity function-based correction, however, gCP has gained considerable traction within the electronic structure community, largely due to its low-cost, satisfactory transferability, and ease of use. The gCP approach is utilized in conjunction with very small basis sets – on the order of 6-31G* – and is capable of recovering most of the BSSE in typical systems.

Within this work, we adapt the gCP empirical correction for BSSE to the def2-SVPD basis. We focus exclusively on the def2-SVPD basis set[154, 155] due to its good balance of expense and performance; def2-SVPD has low intrinsic incompleteness error relative to other comparably-sized bases[43], and hence seems to us to be a particularly promising basis set for BSSE correction schemes. In addition, we develop an alternative beyond-pairwise empirical correction for BSSE within density functional theory: DFT-C. The many-body nature of the method accounts for the overcounting concomitant with any pairwise approach and allows DFT-C to treat both large and small systems in a consistent manner. Whereas gCP is developed for use with exceptionally small basis sets, with the aim of providing semi-quantitative results, we demonstrate DFT-C can recover near-basis-set-limit results at a fraction of the cost, particularly in the case of non-covalent interactions.

6.2 Theory and Methods

gCP

Here, we will briefly summarize the geometrical counterpoise (gCP) correction for BSSE; for further details, see the original study by Kruse and Grimme[191]. At the core of gCP lies

a function describing the decay of BSSE on atom A due to the presence of basis functions on atom B a distance r_{AB} away, which we denote $f_{AB^*}^{\text{gCP}}(r_{AB})$. This term is given by

$$f_{AB^*}^{\text{gCP}}(r_{AB}) = c_{AB} \exp\left(-\alpha r_{AB}^\beta\right), \quad (6.1)$$

and includes a multiplicative constant, c_{AB} as well as a universal decay parameter α and exponent β . The contributions of all atom-ghost pairs are summed up to yield the gCP correction for BSSE,

$$E_{\text{gCP}} = \sigma \sum_A c_A \sum_{B \neq A} f_{AB^*}^{\text{gCP}}(r_{AB}), \quad (6.2)$$

where c_A are atom-dependent parameters and σ is an overall scaling parameter. In practice, E_{gCP} is just added to the total electronic energy for a given system. The gCP approach is strictly pairwise additive with respect to nuclear centers.

Parameterization of gCP

Equations (6.1) and (6.2) contain several parameters: multiplicative constants c_{AB} , linear coefficients c_A , decay factors α and β , and an overall scaling factor σ . The pairwise multiplicative constants, c_{AB} , are calculated as

$$c_{AB} = \frac{1}{\sqrt{S_{AB} N_B^{\text{virt}}}}, \quad (6.3)$$

where N_B^{virt} is the number of virtual orbitals on atom B – given by $N_B^{\text{virt}} = N_B^{\text{basis functions}} - \frac{1}{2} N_B^{\text{electrons}}$ – and S_{AB} is a measure of the Slater overlap between atoms A and B . The overlap term is described in detail in the original study[191]; here, we will simply note that it involves an additional linear parameter, η .

The atomic linear coefficients, c_A , are calculated within the gCP approach as “missing energy” terms, i.e. c_A is calculated as the difference in restricted open-shell Hartree-Fock[282] energy between atom A in a target basis (here def2-SVPD) and a large basis, in the presence of an external electric field to populate higher angular momentum functions. We have utilized aug-pc-4 as the large basis[151–153].

The remaining parameters – three nonlinear (α , β , and η) and one linear (σ) – are obtained by minimizing the error in predicted gCP BSSE relative to Boys-Bernardi BSSE with the B3LYP[15, 16, 22, 82] density functional across the S66x8 dataset of intermolecular interactions[116, 283], within the def2-SVPD basis. As in the original work, the most compressed geometries are weighted for this optimization by a factor of 0.5 in order to emphasize equilibrium and long-distance structures.

The optimized set of parameters is provided in Appendix D; this set of parameters allows the existing gCP approach to be utilized with the def2-SVPD basis set for DFT. Briefly, we mention one particularly interesting aspect of the optimized parameters: the optimal value of η – the parameter controlling atomic overlap in the gCP model – in the def2-SVPD basis

is 0.00001, which suggests that for this particular basis set, the gCP expression can be simplified without degrading performance by simply removing the overlap term. We have verified that this is in fact true; we present in Appendix D a simpler formulation of gCP for def2-SVPD.

DFT-C

In addition to re-parameterizing the gCP method for use with the def2-SVPD basis, we also present a more complex, though physically-motivated, geometry-based empirical approximation for BSSE, which will henceforth be referred to as DFT-C. This model is in many ways similar to gCP[191]. At its core lies a term describing the decay of BSSE on atom A due to the presence of basis functions on atom B a distance r_{AB} away, which we denote $f_{AB^*}^{\text{DFT-C}}(r_{AB})$. This term is given by

$$f_{AB^*}^{\text{DFT-C}}(r_{AB}) = c_{AB} \exp(-\alpha_{AB} r_{AB}^2 + \beta_{AB} r_{AB}), \quad (6.4)$$

and includes a multiplicative constant, c_{AB} , a Gaussian decay parameter, α_{AB} , and an exponential decay parameter, β_{AB} . We expect the decay of BSSE to mirror that of the electron density; the exponential term accounts for the standard decay expression[284], and the Gaussian term reflects the nature of the basis functions employed. The DFT-C approach includes both an exponential and Gaussian term, with pair-dependent decay factors; these differences set it apart from the gCP core term given in eq. (6.1).

In DFT-C, we damp this atomic contribution to BSSE, much as the contribution of gCP is damped in PBEh-3c to potentially address short- r_{AB} issues that can arise in thermochemical problems[192]. We employ the same form of damping function as PBEh-3c[192],

$$d(r_{AB}) = \frac{1}{1 + k_{1,AB} (r_{AB}/r_{0,AB})^{-k_{2,AB}}}, \quad (6.5)$$

where $r_{0,AB}$ is the sum of the van der Waals radii of atoms A and B , and $k_{1,AB}$ and $k_{2,AB}$ are parameters that control the precise shape of the damping function. Whereas Grimme et al.[192] set $k_1 = 4$ and $k_2 = 6$ for all pairs of atoms A and B by inspection, we compute them systematically for each atom pair based on the sums of covalent and van der Waals radii such that $d(r_{\text{cov},AB}) = 0.05$ and $d(r_{0,AB}) = 0.95$. Doing so yields $k_{1,AB} = 19$ and $k_{2,AB} = 5.8889 [\log(r_{0,AB}/r_{\text{cov},AB})]^{-1}$. Moreover, we propose damping to a finite value, rather than zero, to more accurately reflect the actual short-range behavior of BSSE; after all, BSSE does not simply vanish in the covalent bonding distance regime. Thus, rather than simply multiplying the contribution from eq. (6.4) by the damping function in eq. (6.5), we define a damped contribution, $g_{AB^*}^{\text{DFT-C}}(r_{AB})$ as

$$g_{AB^*}^{\text{DFT-C}}(r_{AB}) = d(r_{AB}) f_{AB^*}^{\text{DFT-C}}(r_{AB}) + (1 - d(r_{AB})) f_{AB^*}^{\text{DFT-C}}(r_{\text{cov},AB}). \quad (6.6)$$

At long range, this term reduces to $f_{AB^*}^{\text{DFT-C}}(r_{AB})$, while at short range, it reduces to a pair-dependent constant, $f_{AB^*}^{\text{DFT-C}}(r_{\text{cov},AB})$.

Whereas the gCP correction is strictly pairwise, we incorporate into DFT-C a many-body component. We do so in the following physically-motivated though ad hoc way, by simply modifying each pairwise contribution by an additional term, $h_{AB^*}(\{A, B, \dots\})$, which is given by

$$h_{AB^*}(\{A, B, \dots\}) = \left[1 + \sum_{C \neq A, B} \frac{N_C^{\text{virt}}}{N_B^{\text{virt}}} \text{terfc}(r_{AC}, r_{AB}) \text{terfc}(r_{BC}, r_{AB}) \right]^{-1}, \quad (6.7)$$

where N_B^{virt} is the number of virtual orbitals on atom B – given by $N_B^{\text{virt}} = N_B^{\text{basis functions}} - \frac{1}{2}N_B^{\text{electrons}}$ as in gCP, with $N_B^{\text{electrons}}$ being the number of electrons on neutral atomic B and $N_B^{\text{basis functions}}$ corresponding to the number of basis functions centered at atom B – distances are in atomic units, and $\text{terfc}(x, y)$ is the attenuator defined by Dutoi and Head-Gordon[285],

$$\text{terfc}(x, y) = 1 - \frac{1}{2} [\text{erf}(x + y) + \text{erf}(x - y)]. \quad (6.8)$$

This additional correction, $h_{AB^*}(\{A, B, \dots\})$, addresses the nonzero overlap between the Hilbert space of atom B and the Hilbert spaces of all atoms $C \neq A, B$. As more and more atoms are added in the vicinity of atoms A and B , the contribution of the ghost functions centered at B to the atomic BSSE of A should decrease; eventually, once the space is saturated, adding additional atoms (i.e. ghost functions) does not change the BSSE of atom A . This phenomenon is not captured by a strictly pairwise approach. The many-body correction we employ is visualized for a planar 3-atom system in Figure 6.1.

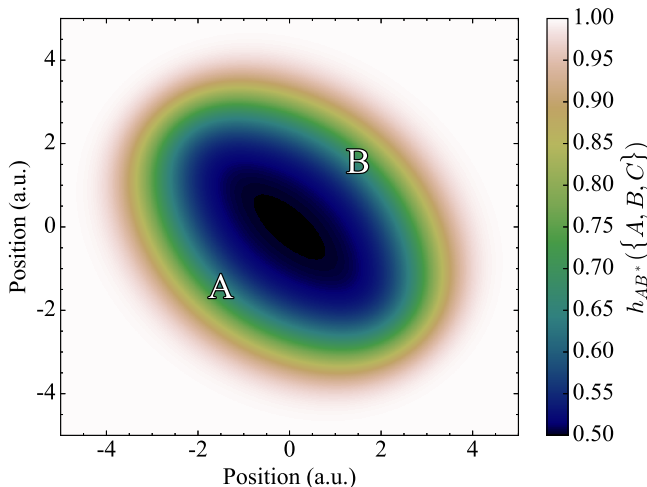


Figure 6.1: Visualization of how adding a third atom C impacts the contribution of basis functions centered at B to the BSSE on atom A , as per $h_{AB^*}(\{A, B, C\})$. When atom C is sufficiently far away from A and B (lighter areas), the model reduces to a pairwise approach. In this example, C and B are assumed to have the same number of virtual orbitals, and A and B are located at $(-1.5 \text{ a.u.}, -1.5 \text{ a.u.})$ and $(1.5 \text{ a.u.}, 1.5 \text{ a.u.})$, respectively.

The final form of the DFT-C correction for BSSE is given by

$$E_{\text{DFT-C}} = \sigma \sum_A c_A \sum_{B \neq A} g_{AB^*}^{\text{DFT-C}}(r_{AB}) h_{AB^*}(\{A, B, \dots\}), \quad (6.9)$$

where σ is an overall scaling coefficient, c_A is a linear coefficient that modifies the contributions of ghost functions on all atoms B to the BSSE on A , and the damped pairwise contribution, $g_{AB^*}^{\text{DFT-C}}(r_{AB})$, and many-body correction, $h_{AB^*}(\{A, B, \dots\})$, are defined in equations (6.6) and (6.7). With the exception of the many-body term, this expression for the DFT-C energy is mathematically similar to that for gCP – c.f. eqs. (6.9) and (6.2).

Parameterization of DFT-C

As can be seen from eq. (6.9), DFT-C has a large number of parameters. For each unique set of atom A and ghost functions centered at atom B , there are exponential and Gaussian decay parameters, α_{AB} and β_{AB} , and there is a multiplicative constant, c_{AB} . These parameters are obtained by generating BSSE curves for neutral atomic pairs AB^* using a form of local spin-density approximation (LSDA), SPW92[10–13], in the def2-SVPD basis. For each unique atom A and corresponding ghost atom B , we perform a least squares fit on a log BSSE curve generated over the range $[r_{\text{cov},AB}, 5r_{\text{cov},AB}]$ in units of $0.1 a_0$. To avoid overemphasizing the long-distance regime – where the atomic BSSE is nearly zero, and hence the logarithm of the BSSE is very large in magnitude – we weight each point by the inverse of the logarithm of the BSSE at each distance. The viability of this approach is demonstrated in Figure 6.2 for the neon component of neon-argon BSSE. The DFT-C method does a reasonable job of capturing BSSE throughout the entire distance regime, yielding an RMSE of 0.002 kcal/mol. Note the gCP RMSE for this system is an order of magnitude larger: the lack of a many-body term in gCP necessitates the systematic underprediction of pairwise atomic BSSEs. For pairs AB^* where the Gaussian decay parameter α_{AB} optimizes to a negative value, we set $\beta_{AB} = 0$ and re-optimize, so as to avoid divergence in the large- r_{AB} limit.

We have parameterized all 1296 combinations of the first 36 elements of the periodic table in this manner; the resulting c_{AB} , α_{AB} , and β_{AB} are tabulated in Appendix D. In the cases of manganese, iron, and cobalt, we have taken averages of the BSSEs for the two competing spin states. For elements heavier than krypton, we propose using the parameters from 4th-row analogues, as is done in gCP.

The linear coefficients c_A in eq. (6.9) are all unity, with the exception of those for hydrogen, carbon, nitrogen, and oxygen, which are fit via least-squares regression of DFT-C predicted BSSEs to actual BSSEs at the SPW92/def2-SVPD level across the S66 dataset of intermolecular interactions[116]. The overall scaling parameter, σ , is by definition unity for LSDA, and is allowed to vary for different density functionals. We have optimized σ for several generalized gradient approximations (GGAs) and meta-GGAs, again by minimizing the root-mean-square error (RMSE) across BSSEs in S66, using the pairwise parameters (c_{AB} , α_{AB} , and β_{AB}) and linear coefficients (c_A) obtained at the LSDA level. For GGA

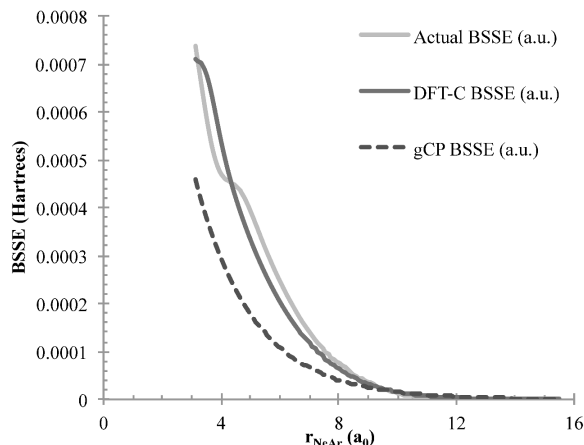


Figure 6.2: Dependence of actual and predicted neon atom SPW92/def2-SVPD BSSEs on distance to argon ghost functions.

functionals, the optimal value of σ is approximately 0.9, while for meta-GGA functionals, it is slightly lower, near 0.85. We thus propose using $\sigma = 1$ for LSDA, $\sigma = 0.9$ for GGAs, and $\sigma = 0.85$ for meta-GGAs.

Ultimately, almost all of the parameters associated with the DFT-C method are obtained from toy systems – neutral atom-ghost pairs – at the LSDA level. Four linear coefficients are trained on S66 BSSEs, also at the LSDA level, and for non-LSDA density functionals, we allow for one scaling parameter, which is trained on S66 BSSEs. An implementation of this method within the python programming language is provided in Appendix D. In practice, the DFT-C correction is applied in the same manner as gCP: the term from eq. 6.9 is simply added to the total electronic energy for a given system.

Datasets and Computational Details

To assess the performance of the gCP and DFT-C methods, we employ a subset of the comprehensive database assembled by Mardirossian and Head-Gordon[24]. The subset we utilize contains 3402 data points distributed over 48 distinct datasets. These smaller constituent datasets are classified according to five distinct datatypes: NCED (easy non-covalent interactions of dimers), NCEC (easy non-covalent interactions of clusters), NCD (difficult non-covalent interactions of dimers), IE (easy isomerization energies), and RG10 (binding curves of rare gas dimers). Unlike “easy” interactions, “difficult” interactions are characterized by strong correlation or self-interaction error. A summary of the datatypes may be found in Table 6.1.

In addition to the version of LSDA on which DFT-C is parameterized – SPW92[10–13] – we consider in this study three GGA and three meta-GGA functionals. At the GGA level, we examine a pure functional, PBE[14]; a global hybrid, B3LYP[15, 16, 22, 82] – the functional with which gCP is parameterized – and a range-separated hybrid, ω B97X-V[23]. At the

| Datatype | # | Constituent Datasets | References |
|----------|------|--|--|
| NCED | 1744 | S66, A24, DS14, HB15, HSG, NBC10, S22, X40, A21x12, BzDC215, HW30, NC15, S66x8, 3B-69-DIM, AlkBind12, CO2Nitrogen16, HB49, Ionic43 | [42, 116, 117, 119, 121, 131, 181, 186, 213–217, 227–236, 283] |
| NCEC | 243 | H2O6Bind8, HW6Cl, HW6F, FmH2O10, Shields38, SW49Bind345, SW49Bind6, WATER27, H2O20Bind4, 3B-69-TRIM, CE20, H2O20Bind10 | [218–223, 234, 237–240] |
| NCD | 91 | TA13, XB18, Bauza30, CT20, XB51 | [241–245] |
| IE | 755 | AlkIsomer11, Butanediol65, ACONF, CY-CONF, Pentane14, SW49Rel345, SW49Rel6, H2O16Rel5, H2O20Rel10, H2O20Rel4, Melatonin52, YMPJ519 | [219–226, 238, 246–250] |
| RG10 | 569 | RG10 | [286] |

Table 6.1: Summary of datatypes. For more details, see Ref. 24.

meta-GGA level, we test a pure functional, B97M-V[19]; a global hybrid, M06-2X[25]; and a range-separated hybrid, ω B97M-V[24].

All density functional calculations are performed in the def2-SVPD basis[154, 155]. A fine Lebedev integration grid of 99 radial shells – each with 590 angular points – is used to compute semi-local components of exchange and correlation, while non-local correlation in the VV10-containing functionals is calculated with the coarser SG-1 grid[133]. All calculations are performed within a development version of Q-Chem 4.4[182].

6.3 Results and Discussion

In this study, we have developed two geometry-based empirical corrections for BSSE in the def2-SVPD basis: gCP and DFT-C. This particular basis was chosen based on its low intrinsic basis set incompleteness error; BSSE-corrected results obtained within this basis are quite near the basis set limit. This is illustrated in Figure 6.3, wherein root-mean-square errors (RMSEs) for B97M-V with (CP) and without (noCP) counterpoise correction against B97M-V/def2-QZVPPD across the various non-covalent datatypes of Table 6.1 are shown. Within the def2-SVP basis, even when BSSE is removed (i.e. the CP SVP specification in Figure 6.3), the remaining basis set incompleteness error is quite large – significantly larger than method errors for typical density functionals. This indicates the def2-SVP basis is not suitable for a high-accuracy BSSE correction scheme; its utility would ultimately be contingent on significant cancellation of method and basis set errors. On the other hand, intrinsic incompleteness error in the def2-SVPD basis is quite small, and so a BSSE correction scheme developed in this basis can, in principle, allow for quantitative reproduction of large-basis results.

In addition to developing the DFT-C method, we have also parameterized the existing gCP scheme within the def2-SVPD basis for comparison. The first of these assessments is

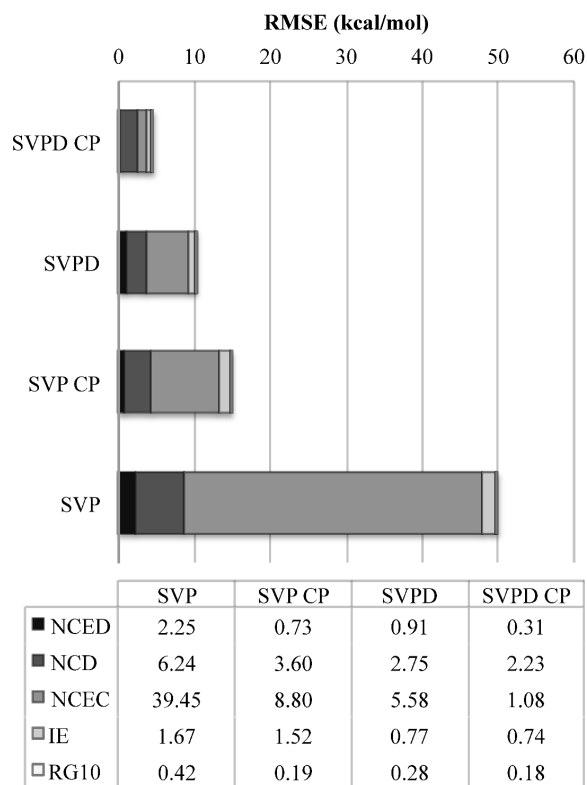


Figure 6.3: Root-mean-square errors of B97M-V with (CP) and without (noCP) the Boys-Bernardi correction for BSSE in two small basis sets relative to B97M-V in the def2-QZVPPD basis, near the basis set limit. SVP and SVPD correspond to def2-SVP and def2-SVPD, respectively. Methods in the chart are ordered from lowest overall RMSE at the top, to highest overall RMSE at the bottom. A table of values is provided below the chart to facilitate quantitative comparison.

shown in Figure 6.4, wherein we have plotted for the three non-covalent datatypes from Table 6.4 normalized root-mean-square errors (NRMSEs) for DFT-C and gCP predicted BSSEs at the LSDA level of DFT. The normalized RMSE is simply the RMSE divided by the mean of the reference data, and hence provides a measure of relative error. Its use facilitates comparison between e.g. NCED and NCEC, since the energy scales of those two datatypes differ by more than an order of magnitude.

Within Figure 6.4, it is evident that both gCP and DFT-C reproduce Boys-Bernardi BSSEs at the LSDA level reasonably well; either correction is a substantial improvement over no correction. The performance of DFT-C on molecular dimers is particularly promising, as is its consistency across the various datatypes: the lowest DFT-C NRMSE in SPW92 is 25%, for NCED, and the highest is 33% (NCD). On the other hand, the performance of gCP is quite variable; the method boasts an exceptionally low NRMSE of 19% across NCEC, but a significantly worse NRMSE of 56% for NCD. Neither correction can be considered a

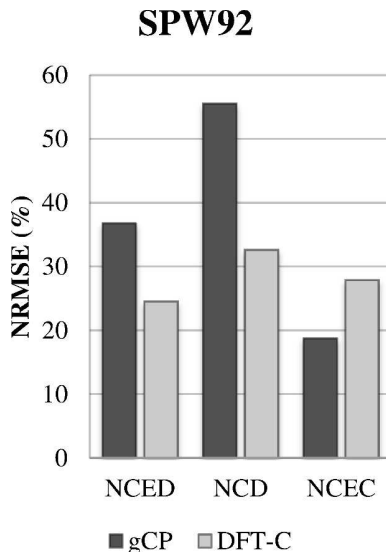


Figure 6.4: Normalized root-mean-square errors (NRMSEs) of gCP and DFT-C predicted BSSEs versus Boys and Bernardi BSSEs at the LSDA level of DFT in the def2-SVPD basis. The datatypes NCED, NCD, and NCEC are defined in Table 6.1. The normalized root-mean-square error is obtained by dividing the RMSE by the mean reference value in the dataset, as described in the text. Direct use of LSDA/def2-SVPD without any correction would result in 100% NRMSE.

quantitative replacement for the full counterpoise correction.

This same sort of comparison is made for three popular GGA functionals in Figure 6.5. Therein, NRMSEs for DFT-C and gCP BSSEs versus actual BSSEs obtained with a pure functional (PBE), a global hybrid (B3LYP), and a range-separated hybrid with non-local correlation (ω B97X-V) may be found. It is clear that for all three density functionals, both DFT-C and gCP are quite consistent with regards to their performances across the various datatypes. Moreover, comparing with Figure 6.4, this consistency extends across the LSDA-GGA gap for DFT-C, which bodes well for its transferability.

This same level of consistency is not seen for gCP, however: whereas gCP reproduces LSDA cluster BSSEs with unparalleled accuracy, the method is not nearly as good for clusters at the GGA level: the gCP NRMSE across NCEC in ω B97X-V is more than double that in SPW92. This is a consequence of the fact that gCP tends to overestimate BSSE in molecular clusters, and BSSEs obtained at the LSDA level are on average larger than those at the GGA level. The exceptional performance of gCP on SPW92 cluster BSSEs may thus be understood to be largely a consequence of the offsetting of these two phenomena.

It is also evident from Figure 6.5 that at the GGA level, DFT-C affords significant gains over gCP regardless of datatype or density functional. This is quite promising, as DFT-C is parameterized almost entirely at the LSDA level of theory, with only the overall scaling parameter changing from $\sigma = 1$ to $\sigma = 0.9$. On the other hand, gCP is parameterized at

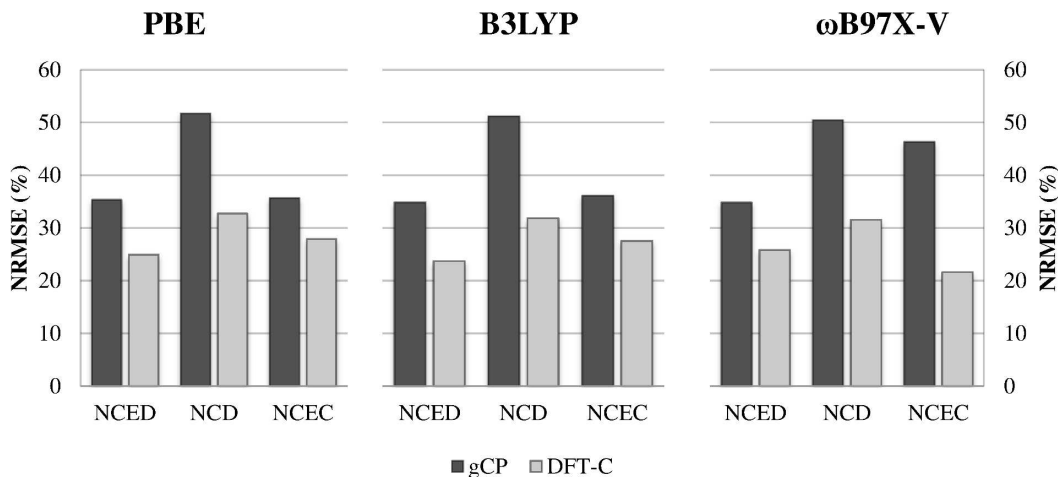


Figure 6.5: Normalized root-mean-square errors (NRMSEs) of gCP and DFT-C predicted BSSEs versus Boys and Bernardi BSSEs for three GGA density functionals in the def2-SVPD basis. For further details, see Figure 6.4.

the GGA level, specifically with B3LYP. It is still true that use of gCP is significantly better than no correction at all.

In Figure 6.6, we further assess the transferability of the gCP and DFT-C BSSE correction schemes across three distinct meta-GGAs: a pure meta-GGA B97M-V, a global hybrid M06-2X, and a range-separated hybrid ω B97M-V. Again, we see that across the three meta-GGA functionals, the relative performances of gCP and DFT-C are similar: for all three functionals, gCP exhibits a NRMSE of around 35% for NCED, 50% for NCD, and 60% for NCEC; the corresponding NRMSEs for DFT-C are 25%, 35%, and 20%. Similarly, we see the same sort of consistency for the DFT-C approach at the meta-GGA level as was seen at the GGA and LSDA levels (c.f. Figures 6.5 and 6.4). On the other hand, gCP is slightly worse at describing molecular clusters at the meta-GGA level than it was at the GGA level. Again, this can be traced back to the facts that gCP systematically over-predicts BSSE in molecular clusters, and meta-GGA BSSEs tend to be even lower than their GGA counterparts. This overcorrection by gCP can in turn be attributed to its strictly pairwise nature; due to the inclusion of a many-body correction, the DFT-C approach does not suffer from this overcounting issue.

Across the seven density functionals examined, the average NRMSE of the DFT-C approach across NCED is 30%, compared to the 42% of gCP; this corresponds to an improvement of more than 25%. For the NCD datatype, the gCP average NRMSE is 59%, compared to 38% – an improvement of 35%. Across the NCEC set of molecular clusters, we see a 46% improvement for DFT-C over gCP: a reduction in average NRMSE from 52% to 28%. It is clear that for a wide variety of systems, across a diverse set of density functionals, in the def2-SVPD basis, the DFT-C method is satisfactorily transferable and represents a significant improvement over gCP for the reproduction of Boys-Bernardi BSSEs. The re-

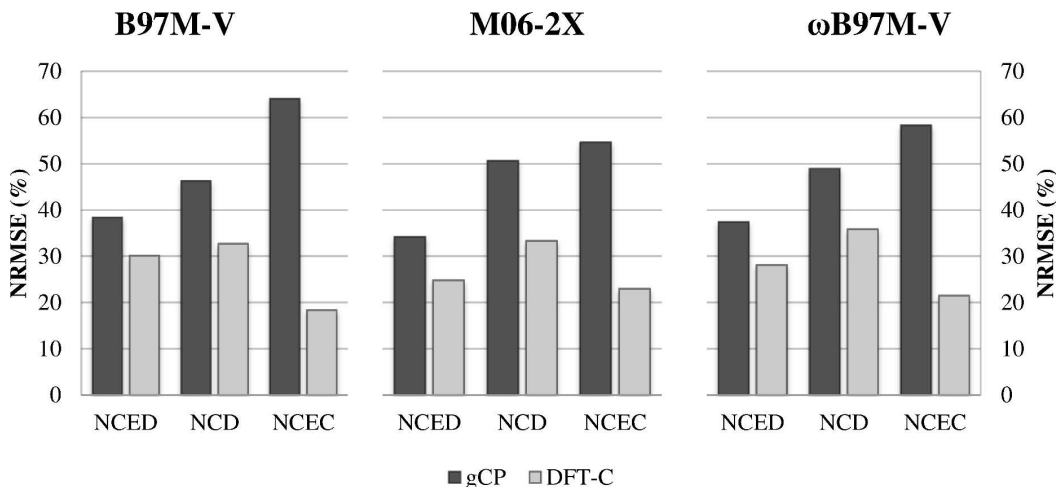


Figure 6.6: Normalized root-mean-square errors (NRMSEs) of gCP and DFT-C predicted BSSEs versus Boys and Bernardi BSSEs for three meta-GGA density functionals in the def2-SVPD basis. For further details, see Figure 6.4.

remaining DFT-C error of course represents the remaining gap to perfect reproduction of the Boys-Bernardi counterpoise correction.

Thus far, with the exception of the basis set comparison in Figure 6.3, all errors have been expressed relative to “exact” BSSEs. Although such metrics are relevant for this particular work, since the DFT-C and gCP methods are designed and trained to reproduce BSSEs, they are not of the same broad interest as, say, errors relative to high-level electronic structure methods. In Figure 6.7, we show root-mean-square errors (RMSEs) across the five datatypes from Table 6.1 for the B97M-V functional relative to high-level (generally CCSD(T)/CBS) results. The noCP and CP designations correspond to uncorrected and counterpoise-corrected B97M-V/def2-SVPD, respectively, and CBS corresponds to B97M-V/def2-QZVPPD – effectively B97M-V at the basis set limit. DFT-C and gCP refer to B97M-V/def2-SVPD with the corresponding approximation for BSSE included.

From Figure 6.7, it is immediately evident that any sort of BSSE correction is preferable to no correction. By correcting using the standard Boys-Bernardi approach, we are able to eliminate 90% of basis set error for NCED, 71% for NCD, 97% for NCEC, and even improve upon CBS results for RG10. Unfortunately, the standard counterpoise correction can not be applied for the vast majority of isomerization energies – it can only be applied for relative relative energies, such as relative binding energies – and so the CP and noCP results are almost identical for IE. On the other hand, both gCP and DFT-C offer solid improvements over noCP for every datatype examined, including isomerization energies, for which we are able to eliminate roughly 60% of basis set error. A full table of RMSEs across the individual datasets composing each aggregate datatype may be found in Appendix D.

For B97M-V/def2-SVPD, the DFT-C approach offers modest improvements over gCP

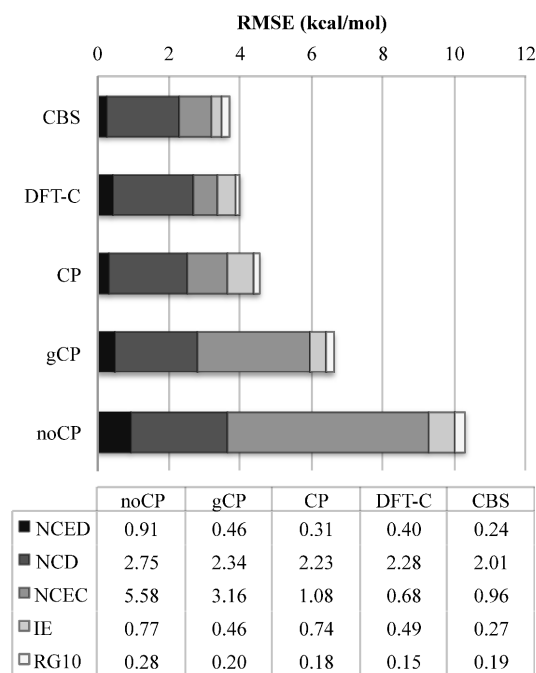


Figure 6.7: Root-mean-square errors of B97M-V versus high-level reference values at five levels of theory: uncorrected in the def2-SVPD basis (noCP); counterpoise-corrected in def2-SVPD (CP); with the geometrical counterpoise correction in def2-SVPD (gCP); with the correction introduced in this work in the def2-SVPD basis (DFT-C); and near the complete-basis set limit (CBS), in def2-QZVPPD. Methods in the chart are ordered from lowest overall RMSE at the top, to highest overall RMSE at the bottom. A table of values is provided below the chart to facilitate quantitative comparison.

for molecular dimers (NCED, NCD, and RG10), a significant improvement for molecular clusters (NCEC), and is slightly inferior for isomerization energies (IE). The DFT-C method outperforms the Boy-Bernardi counterpoise correction across the full dataset, with an overall RMSE of 0.56 kcal/mol compared to a CP RMSE of 0.63 RMSE; the large improvements it affords for NCEC and IE offset the small losses on NCED and NCD. As such, DFT-C is a viable alternative to the traditional counterpoise correction in the def2-SVPD basis set, yielding similar results to CP with effectively no increase in cost over noCP.

To further illustrate the power of the DFT-C BSSE-correction scheme, in Figure 6.8 we show RMSEs across the four aggregate datatypes for B97M-V with (B97M-V-C) and without (B97M-V) the DFT-C correction for BSSE in the def2-SVPD basis, as well as for four popular pure meta-GGA density functionals – B97M-V[19], MS2-D3(op)[44, 211], M06-L[20], and TM[269] – near the CBS limit, in the def2-QZVPPD basis. From Figure 6.8, it is clear that although B97M-V/def2-SVPD is not competitive with standard meta-GGAs at the basis set limit, B97M-V-C/def2-SVPD certainly is – despite requiring a small fraction

of the computational effort.

| Functional (Basis) | Datatype | | | |
|------------------------|----------|------|------|------|
| | NCED | NCD | NCEC | IE |
| B97M-V (SVPD) | 0.91 | 2.75 | 5.58 | 0.77 |
| B97M-V-C (SVPD) | 0.40 | 2.28 | 0.68 | 0.49 |
| B97M-V (QZVPPD) | 0.24 | 2.01 | 0.96 | 0.27 |
| M06-L (QZVPPD) | 0.55 | 1.87 | 2.20 | 0.71 |
| TM (QZVPPD) | 0.50 | 3.13 | 1.59 | 0.49 |
| MS2-D3(op) (QZVPPD) | 0.29 | 2.65 | 0.89 | 0.35 |

Figure 6.8: Root-mean-square errors in kcal/mol of several pure meta-GGA density functionals relative to high-level reference values. B97M-V-C corresponds to B97M-V with the DFT-C correction. Results for the additional density functionals are taken from a previous study[44]. SVPD corresponds to def2-SVPD, and QZVPPD corresponds to def2-QZVPPD. Each datatype category is color-coded, with the darkest color corresponding to the lowest RMSE within that category.

6.4 Discussion and Conclusions

In this study, we have introduced a physically-motivated empirical correction for basis set superposition error within the def2-SVPD basis set: DFT-C. This correction differs from the existing gCP approach – which we have also re-parameterized for use in the def2-SVPD basis – in two critical areas. First, whereas the linear coefficients within gCP include all manifestations of basis set incompleteness error, the DFT-C approach is constructed exclusively from basis set superposition errors. Second, although gCP is a strictly pairwise correction, in DFT-C each pairwise contribution is reduced by a many-body term to ameliorate the overcounting concomitant with the non-orthogonality of the Hilbert spaces of nearby atoms. We have evaluated both gCP and DFT-C across a diverse dataset containing 3402 non-covalent interactions and isomerization energies.

This new method, DFT-C, yields significantly more accurate BSSEs than gCP for a wide variety of interaction motifs. Moreover, the correction is transferable. DFT-C exhibits roughly the same relative performances across the various non-covalent datatypes regardless of the particular density functional with which it is paired: for non-covalently interacting dimers, DFT-C offers a modest improvement over gCP; in the case of molecular clusters

– particularly when a meta-GGA functional is employed – the improvement is more pronounced, which is likely attributable to the many-body nature of the method.

Whereas gCP has been developed as a general purpose tool that can be relatively easily adapted to any basis set, the DFT-C approach is much more complicated and specialized; tabulating the many pairwise coefficients and decay parameters is a nontrivial task. In this particular work, we have introduced a correction for def2-SVPD, a double-zeta basis set that has disproportionately low intrinsic basis set incompleteness error for how few basis functions it contains[43]. We are also exploring the possibility of extending this method to triple-zeta basis sets in order to truly push the basis set limit; such may be the focus of work to come.

Much as gCP is employed as a component of a small-basis functional in PBEh-3c, so too could DFT-C be adapted, with either some subset of the linear parameters c_A or simply the overall scaling parameter σ being allowed to vary. Even without modification, however, the method is immensely powerful; we have demonstrated it can be paired with an existing functional, B97M-V, to yield def2-SVPD results on par with def2-QZVPPD results for other state-of-the-art pure meta-GGA density functionals. DFT-C should prove immensely useful for recovering large-basis results for many energetic properties with small-basis effort – the correction scales with the number of atoms, not the number of basis functions, after all, and is essentially free on the scale of an electronic structure calculation – and it can be paired without modification with any density functional. This could allow us to obtain high-quality results for large systems which are currently out of the domain of quantitative electronic structure theory.

Appendix A

Geometry Study

A.1 Damping Function

| Method | Contact RMS Error (Å) | Contact Signed Error (Å) |
|----------------|-----------------------|--------------------------|
| PBE | 0.224 | 0.060 |
| PBE-D2 | 0.202 | -0.139 |
| PBE-D2 (CHG) | 0.165 | -0.054 |
| PBE-D3 | 0.155 | -0.082 |
| PBE-D3 (BJ) | 0.167 | -0.070 |
| PBE-D3 (Fermi) | 0.196 | -0.129 |

Table A.1: Impact of damping function. All results pertain to A21 set, aug-cc-pVTZ basis, (99,590) grid, tight convergence criteria (details in article). Average RMS and signed errors in the closest point of contact for each optimized structure are shown. The method marked CHG combines -D2 C_6 coefficients with a -D3 style damping function. The method marked Fermi combines -D3 C_6 coefficients with the -D2 damping function (a Fermi-type damping function).

A.2 MP2 Basis

| Basis | Contact Signed Error (\AA) |
|-------------|---------------------------------------|
| aug-cc-pVTZ | -0.043 |
| aug-cc-pVQZ | -0.030 |

Table A.2: Effect of finiteness of basis on MP2 A21 calculations. Since the reference structures are $\Delta\text{CCSD(T)}/\text{CBS}$, it is possible that our MP2 results suffer from considerable BSSE. We thus show here average contact signed errors across the A21 dataset in the aug-cc-pVQZ basis (in which we expect our results should be well-converged with respect to basis size).

A.3 Intramolecular Metrics

| Method | MBL (pm) | MBA (°) |
|-----------------|----------|---------|
| SOS-MP2 | 0.16 | 0.08 |
| SCS-MP2 | 0.17 | 0.08 |
| MP2 | 0.26 | 0.10 |
| attMP2 | 0.28 | 0.08 |
| B3LYP-D3 | 0.31 | 0.31 |
| B3LYP | 0.31 | 0.31 |
| B3LYP-D3 (BJ) | 0.32 | 0.30 |
| ω B97X-D | 0.33 | 0.26 |
| M06-2X | 0.34 | 0.24 |
| ω B97X | 0.34 | 0.28 |
| sMP2 | 0.34 | 0.14 |
| ω B97X-V | 0.35 | 0.21 |
| B3LYP-D2 | 0.35 | 0.28 |
| M06 | 0.41 | 0.33 |
| M06-L | 0.42 | 0.41 |
| M11 | 0.45 | 0.48 |
| vdW-DF2 | 0.50 | 0.23 |
| LC-VV10 | 0.51 | 0.41 |
| B97M-V | 0.56 | 0.30 |
| VV10 | 0.69 | 0.22 |
| PBE-D3 (BJ) | 0.92 | 0.26 |
| PBE | 0.92 | 0.27 |
| PBE-D3 | 0.93 | 0.27 |
| PBE-D2 | 1.00 | 0.22 |
| HF | 1.27 | 0.51 |

Table A.3: Mean bond length (MBL) and bond angle (MBA) RMS errors across the A21 set. Methods are sorted in order of ascending MBL.

Appendix B

Basis Set Study

B.1 Numbering Convention: Key for S22

| System Number | System Name |
|---------------|--------------------------------------|
| 1 | Ammonia dimer |
| 2 | Water dimer |
| 3 | Formic acid dimer |
| 4 | Formamide dimer |
| 5 | Uracil dimer h-bonded |
| 6 | 2-pyridoxine 2-aminopyridine complex |
| 7 | Adenine thymine Watson-Crick complex |
| 8 | Methane dimer |
| 9 | Ethene dimer |
| 10 | Benzene - Methane complex |
| 11 | Benzene dimer parallel displaced |
| 12 | Pyrazine dimer |
| 13 | Uracil dimer stack |
| 14 | Indole benzene complex stack |
| 15 | Adenine thymine complex stack |
| 16 | Ethene ethyne complex |
| 17 | Benzene water complex |
| 18 | Benzene ammonia complex |
| 19 | Benzene HCN complex |
| 20 | Benzene dimer T-shaped |
| 21 | Indole benzene T-shape complex |
| 22 | Phenol dimer |

Table B.1: Numbering convention of S22 used to abbreviate systems in the remainder of this Appendix to make tables more compact.

B.2 Justification of CP pc-4 as CBS limit

| System | BE difference from CP pc-4 (kcal/mol) | |
|--------|---------------------------------------|-------------|
| | noCP aug-pc-4 | CP aug-pc-4 |
| 1 | 0.000 | -0.001 |
| 2 | 0.002 | -0.001 |
| 3 | 0.002 | 0.000 |
| 4 | 0.001 | -0.001 |
| 5 | n/a | n/a |
| 6 | 0.003 | 0.000 |
| 7 | n/a | n/a |
| 8 | 0.001 | 0.000 |
| 9 | 0.003 | 0.001 |
| 10 | 0.002 | 0.001 |
| 11 | 0.003 | -0.001 |
| 12 | 0.005 | 0.002 |
| 13 | 0.005 | 0.004 |
| 14 | 0.005 | 0.003 |
| 15 | 0.008 | 0.001 |
| 16 | 0.001 | 0.000 |
| 17 | 0.003 | 0.001 |
| 18 | 0.004 | 0.001 |
| 19 | 0.003 | 0.000 |
| 20 | 0.002 | 0.001 |
| 21 | 0.003 | 0.001 |
| 22 | 0.009 | 0.001 |

Table B.2: Differences in SPW92 binding energies calculated in the aug-pc-4 basis set with (CP) and without (noCP) counterpoise correction, as compared to CP pc-4 values. Two systems could not be converged in the aug-pc-4 basis set (issues with linear dependencies); these are listed as "n/a."

B.3 Semilocal Exchange-Correlation Grid

| System | BSSE (kcal/mol) | Incremental Change in BSSE (kcal/mol) | | |
|--------|-----------------|---------------------------------------|----------|---------|
| | SG-1 | (75,302) | (99,590) | 250,974 |
| 1 | 3.18 | 0.01 | 0.00 | 0.00 |
| 2 | 3.85 | -0.01 | 0.00 | 0.00 |
| 3 | 6.40 | -0.02 | 0.00 | 0.00 |
| 4 | 5.91 | -0.05 | 0.00 | 0.00 |
| 5 | 4.72 | -0.15 | -0.01 | 0.00 |
| 6 | 4.57 | 0.03 | 0.03 | 0.00 |
| 7 | 5.43 | -0.02 | -0.02 | 0.00 |
| 8 | 0.19 | 0.00 | 0.00 | 0.00 |
| 9 | 0.74 | -0.04 | -0.01 | 0.00 |
| 10 | 0.44 | 0.01 | 0.00 | 0.00 |
| 11 | 1.26 | -0.13 | 0.06 | 0.00 |
| 12 | 1.53 | 0.12 | 0.00 | 0.00 |
| 13 | 3.62 | -0.21 | 0.03 | 0.00 |
| 14 | 1.55 | 0.26 | 0.00 | 0.00 |
| 15 | 4.30 | 0.18 | -0.02 | 0.00 |
| 16 | 0.39 | 0.00 | 0.00 | 0.00 |
| 17 | 1.30 | -0.04 | 0.02 | 0.00 |
| 18 | 0.76 | -0.07 | 0.02 | 0.00 |
| 19 | 0.59 | 0.01 | 0.01 | 0.00 |
| 20 | 0.67 | 0.06 | 0.01 | 0.00 |
| 21 | 1.06 | 0.06 | 0.02 | 0.00 |
| 22 | 3.19 | 0.24 | 0.02 | 0.00 |

Table B.3: Impact of semilocal exchange-correlation grid. All results are for SPW92 using the cc-pVDZ basis set. Each column (SG-1 and the following columns) refers to a different Lebedev grid. The "incremental change in BSSE" refers to the change in BSSE upon switching to that grid from the next smallest grid, e.g. the numbers under (99,590) correspond to the change in BSSE accompanying the change from the (75,302) to the (99,590) grid.

B.4 Nonlocal Correlation Grid

| System | noCP BE (kcal/mol) | | CP BE (kcal/mol) | | BSSE (kcal/mol) | |
|--------|--------------------|----------|------------------|----------|-----------------|----------|
| | SG-1 | (99,590) | SG-1 | (99,590) | SG-1 | (99,590) |
| 1 | -5.39 | -5.39 | -2.89 | -2.89 | 2.50 | 2.50 |
| 2 | -7.81 | -7.81 | -4.87 | -4.87 | 2.94 | 2.94 |
| 3 | -22.68 | -22.68 | -17.85 | -17.85 | 4.83 | 4.83 |
| 4 | -19.34 | -19.34 | -14.85 | -14.84 | 4.49 | 4.49 |
| 5 | -22.79 | -22.79 | -19.29 | -19.29 | 3.49 | 3.50 |
| 6 | -19.61 | -19.61 | -16.02 | -16.02 | 3.60 | 3.60 |
| 7 | -19.54 | -19.54 | -15.31 | -15.31 | 4.23 | 4.23 |
| 8 | -0.52 | -0.52 | -0.38 | -0.38 | 0.15 | 0.15 |
| 9 | -1.78 | -1.78 | -1.19 | -1.19 | 0.59 | 0.59 |
| 10 | -1.57 | -1.57 | -1.18 | -1.18 | 0.39 | 0.39 |
| 11 | -3.32 | -3.32 | -2.15 | -2.15 | 1.17 | 1.17 |
| 12 | -4.97 | -4.98 | -3.43 | -3.44 | 1.54 | 1.54 |
| 13 | -11.92 | -11.91 | -8.83 | -8.82 | 3.09 | 3.09 |
| 14 | -5.57 | -5.58 | -3.83 | -3.84 | 1.74 | 1.73 |
| 15 | -14.74 | -14.76 | -10.71 | -10.72 | 4.03 | 4.03 |
| 16 | -1.93 | -1.93 | -1.59 | -1.59 | 0.34 | 0.34 |
| 17 | -3.76 | -3.76 | -2.67 | -2.67 | 1.08 | 1.08 |
| 18 | -2.45 | -2.45 | -1.83 | -1.83 | 0.62 | 0.62 |
| 19 | -4.72 | -4.72 | -4.19 | -4.19 | 0.52 | 0.52 |
| 20 | -2.91 | -2.91 | -2.25 | -2.25 | 0.66 | 0.66 |
| 21 | -5.98 | -5.99 | -4.95 | -4.95 | 1.04 | 1.04 |
| 22 | -9.24 | -9.25 | -6.41 | -6.41 | 2.83 | 2.84 |

Table B.4: Impact of nonlocal correlation grid. All results are for B97M-V and the cc-pVDZ basis set with a (99,590) semilocal grid. Both uncorrected (noCP) and counterpoise-corrected (CP) binding energies (BEs) are provided for each grid (SG-1 and (99,590)) to demonstrate that the fact that BSSE is converged is not a result of some fortuitous error cancellation.

B.5 SPW92 RMSE versus CBS and S22B

| Basis | noCP RMSE (kcal/mol) | | CP RMSE (kcal/mol) | |
|-------------------|----------------------|---------|--------------------|---------|
| | vs CBS | vs S22B | vs CBS | vs S22B |
| 6-31G* | 2.48 | 5.54 | 0.88 | 3.36 |
| cc-pVDZ | 2.83 | 5.89 | 0.57 | 2.96 |
| def2-SVP | 3.10 | 6.12 | 0.50 | 3.29 |
| pc-1 | 2.41 | 5.27 | 0.49 | 3.40 |
| 6-31++G** | 0.87 | 3.60 | 0.31 | 3.09 |
| def2-SVPD | 2.17 | 4.76 | 0.14 | 3.26 |
| 6-311++G** | 1.05 | 3.70 | 0.30 | 3.10 |
| aug-cc-pVDZ | 0.86 | 3.69 | 0.12 | 3.09 |
| aug-pc-1 | 2.09 | 4.50 | 0.11 | 3.08 |
| def2-TZVP | 0.59 | 3.48 | 0.22 | 3.05 |
| cc-pVTZ | 0.98 | 4.00 | 0.12 | 3.13 |
| pc-2 | 0.42 | 3.41 | 0.08 | 3.14 |
| def2-TZVPD | 0.29 | 3.19 | 0.13 | 3.04 |
| 6-311++G(3df,3pd) | 0.69 | 3.49 | 0.06 | 3.11 |
| aug-cc-pVTZ | 0.20 | 3.26 | 0.04 | 3.13 |
| aug-pc-2 | 0.29 | 3.31 | 0.04 | 3.13 |
| cc-pVQZ | 0.46 | 3.51 | 0.05 | 3.17 |
| def2-QZVP | 0.17 | 3.27 | 0.02 | 3.16 |
| def2-QZVPD | 0.06 | 3.18 | 0.02 | 3.15 |
| pc-3 | 0.03 | 3.18 | 0.01 | 3.17 |
| aug-cc-pVQZ | 0.11 | 3.22 | 0.00 | 3.16 |
| aug-pc-3 | 0.03 | 3.16 | 0.03 | 3.15 |
| cc-pV5Z | 0.12 | 3.24 | 0.01 | 3.17 |
| pc-4 | 0.01 | 3.17 | 0.00 | 3.16 |
| aug-cc-pV5Z | 0.01 | 3.17 | 0.00 | 3.16 |

Table B.5: Root mean square errors (RMSE) of uncorrected (noCP) and counterpoise-corrected (CP) SPW92 versus two different types of references: complete-basis SPW92 (CBS) and complete-basis CCSD(T) (S22B[186]).

B.6 B3LYP RMSE versus CBS and S22B

| Basis | noCP RMSE (kcal/mol) | | CP RMSE (kcal/mol) | |
|-------------------|----------------------|---------|--------------------|---------|
| | vs CBS | vs S22B | vs CBS | vs S22B |
| 6-31G* | 2.72 | 3.59 | 0.56 | 5.21 |
| cc-pVDZ | 1.71 | 3.79 | 0.32 | 5.19 |
| def2-SVP | 2.98 | 3.68 | 0.30 | 5.09 |
| pc-1 | 2.11 | 3.57 | 0.31 | 4.84 |
| 6-31++G** | 0.75 | 4.34 | 0.28 | 5.04 |
| def2-SVPD | 1.84 | 3.24 | 0.12 | 5.02 |
| 6-311++G** | 0.70 | 4.37 | 0.26 | 5.00 |
| aug-cc-pVDZ | 0.75 | 4.23 | 0.15 | 5.05 |
| aug-pc-1 | 2.00 | 3.03 | 0.22 | 5.11 |
| def2-TZVP | 0.53 | 4.57 | 0.15 | 4.98 |
| cc-pVTZ | 0.99 | 4.28 | 0.07 | 4.97 |
| pc-2 | 0.31 | 4.71 | 0.07 | 4.95 |
| def2-TZVPD | 0.18 | 4.78 | 0.11 | 5.01 |
| 6-311++G(3df,3pd) | 0.50 | 4.47 | 0.09 | 5.02 |
| aug-cc-pVTZ | 0.21 | 4.74 | 0.05 | 4.95 |
| aug-pc-2 | 0.14 | 4.81 | 0.03 | 4.94 |
| cc-pVQZ | 0.46 | 4.61 | 0.04 | 4.94 |
| def2-QZVP | 0.15 | 4.83 | 0.02 | 4.95 |
| def2-QZVPD | 0.05 | 4.89 | 0.01 | 4.94 |
| pc-3 | 0.03 | 4.92 | 0.01 | 4.95 |
| aug-cc-pVQZ | 0.10 | 4.85 | 0.01 | 4.94 |
| aug-pc-3 | 0.02 | 4.93 | 0.01 | 4.95 |
| cc-pV5Z | 0.10 | 4.88 | 0.01 | 4.95 |
| pc-4 | 0.00 | 4.94 | 0.00 | 4.94 |
| aug-cc-pV5Z | 0.01 | 4.94 | 0.02 | 4.96 |

Table B.6: Root mean square errors (RMSE) of uncorrected (noCP) and counterpoise-corrected (CP) B3LYP versus two different types of references: complete-basis B3LYP (CBS) and complete-basis CCSD(T) (S22B[186]).

B.7 B97M-V RMSE versus CBS and S22B

| Basis | noCP RMSE (kcal/mol) | | CP RMSE (kcal/mol) | |
|-------------------|----------------------|---------|--------------------|---------|
| | vs CBS | vs S22B | vs CBS | vs S22B |
| 6-31G* | 2.00 | 1.74 | 0.64 | 0.75 |
| cc-pVDZ | 1.71 | 1.91 | 0.49 | 0.75 |
| def2-SVP | 2.28 | 2.05 | 0.34 | 0.53 |
| pc-1 | 2.04 | 1.81 | 0.45 | 0.53 |
| 6-31++G** | 0.70 | 0.56 | 0.30 | 0.50 |
| def2-SVPD | 1.77 | 1.57 | 0.08 | 0.34 |
| 6-311++G** | 0.65 | 0.51 | 0.25 | 0.41 |
| aug-cc-pVDZ | 0.72 | 0.55 | 0.15 | 0.44 |
| aug-pc-1 | 2.35 | 2.15 | 0.28 | 0.55 |
| def2-TZVP | 0.56 | 0.41 | 0.16 | 0.37 |
| cc-pVTZ | 0.94 | 0.75 | 0.14 | 0.40 |
| pc-2 | 0.37 | 0.27 | 0.16 | 0.36 |
| def2-TZVPD | 0.22 | 0.25 | 0.04 | 0.35 |
| 6-311++G(3df,3pd) | 0.41 | 0.35 | 0.06 | 0.38 |
| aug-cc-pVTZ | 0.26 | 0.23 | 0.03 | 0.33 |
| aug-pc-2 | 0.54 | 0.40 | 0.03 | 0.34 |
| cc-pVQZ | 0.44 | 0.34 | 0.06 | 0.35 |
| def2-QZVP | 0.15 | 0.28 | 0.02 | 0.35 |
| def2-QZVPD | 0.07 | 0.31 | 0.01 | 0.34 |
| pc-3 | 0.07 | 0.30 | 0.02 | 0.34 |
| aug-cc-pVQZ | 0.13 | 0.27 | 0.01 | 0.34 |
| aug-pc-3 | 0.15 | 0.38 | 0.13 | 0.43 |
| cc-pV5Z | 0.13 | 0.30 | 0.01 | 0.35 |
| pc-4 | 0.02 | 0.34 | 0.00 | 0.35 |
| aug-cc-pV5Z | 0.07 | 0.31 | 0.01 | 0.35 |

Table B.7: Root mean square errors (RMSE) of uncorrected (noCP) and counterpoise-corrected (CP) B97M-V versus two different types of references: complete-basis B97M-V (CBS) and complete-basis CCSD(T) (S22B[186]).

B.8 Nonlocal Correlation Basis Set Insensitivity

| System | Nonlocal BE (kcal/mol) | | | Difference |
|--------|------------------------|-------|--------|------------|
| | def2-SVPD | noCP | pc4 CP | |
| 1 | | -0.68 | -0.68 | 0.00 |
| 2 | | -0.62 | -0.61 | 0.00 |
| 3 | | -2.24 | -2.22 | 0.02 |
| 4 | | -2.10 | -2.09 | 0.01 |
| 5 | | -2.76 | -2.75 | 0.02 |
| 6 | | -3.03 | -3.01 | 0.02 |
| 7 | | -3.34 | -3.33 | 0.01 |
| 8 | | -0.52 | -0.53 | -0.01 |
| 9 | | -1.19 | -1.20 | -0.01 |
| 10 | | -1.45 | -1.45 | 0.00 |
| 11 | | -4.47 | -4.44 | 0.02 |
| 12 | | -4.61 | -4.59 | 0.02 |
| 13 | | -6.77 | -6.78 | -0.01 |
| 14 | | -6.49 | -6.47 | 0.02 |
| 15 | | -9.64 | -9.64 | 0.00 |
| 16 | | -0.61 | -0.60 | 0.00 |
| 17 | | -1.39 | -1.40 | -0.01 |
| 18 | | -1.42 | -1.43 | -0.02 |
| 19 | | -1.69 | -1.69 | 0.00 |
| 20 | | -2.49 | -2.50 | -0.01 |
| 21 | | -3.43 | -3.44 | -0.01 |
| 22 | | -3.03 | -3.03 | 0.00 |

Table B.8: Nonlocal (VV10) contributions to B97M-V binding energies across S22 with a small basis (uncorrected def2-SVPD) and the complete basis (counterpoise-corrected pc4). This does not arise solely from fortuitous error cancellation between the dimers and their constituent monomers; absolute energies are converged to nearly this same level of precision.

B.9 Basis Set Convergence Figures

Each figure in this section depicts normalized mean binding energies across the S22 set. Binding energies were normalized to the relevant [Method]/CBS value before averaging.

SPW92

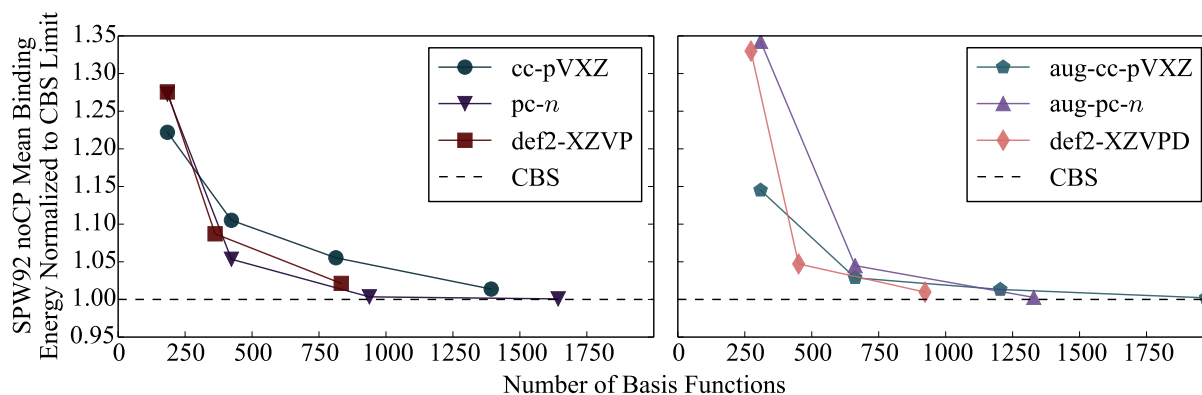


Figure B.1: Convergence of uncorrected (noCP) SPW92 normalized mean binding energies across the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets.

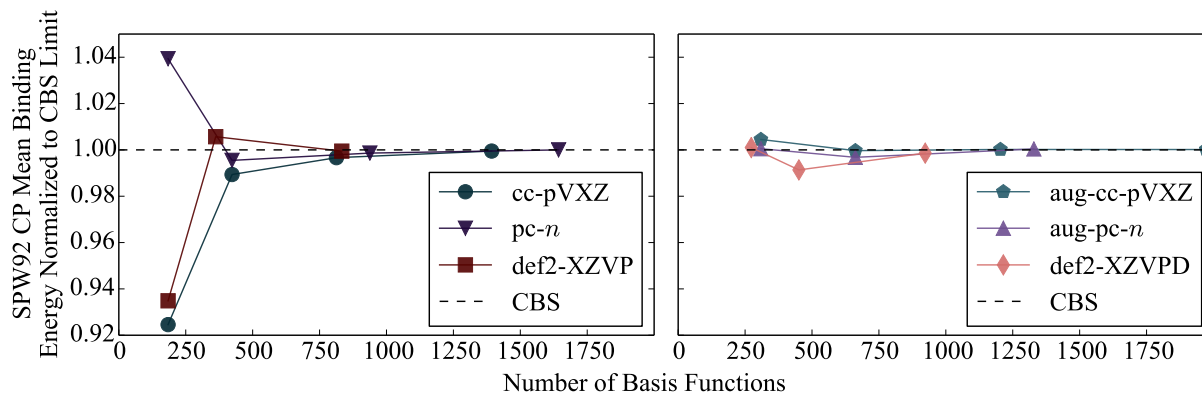


Figure B.2: Convergence of counterpoise-corrected (CP) SPW92 normalized mean binding energies across the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets.

B3LYP

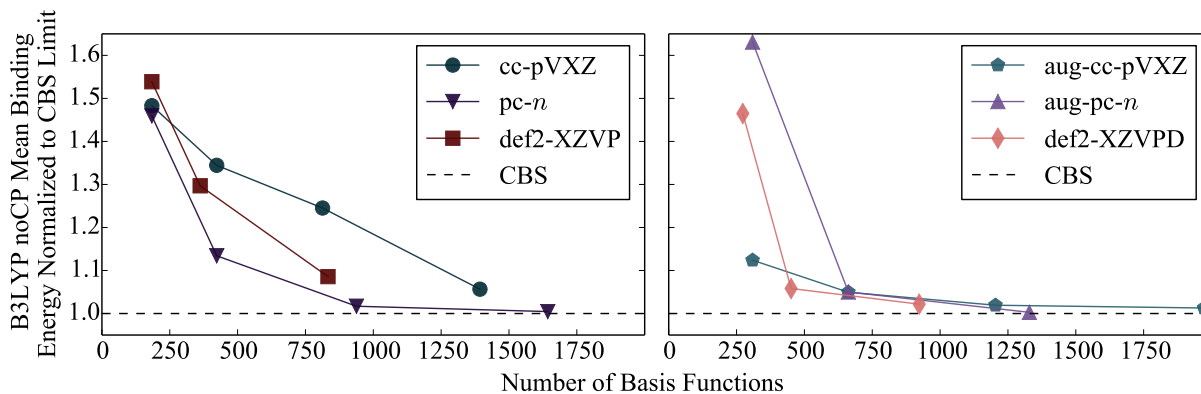


Figure B.3: Convergence of uncorrected (noCP) B3LYP normalized mean binding energies across the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets.

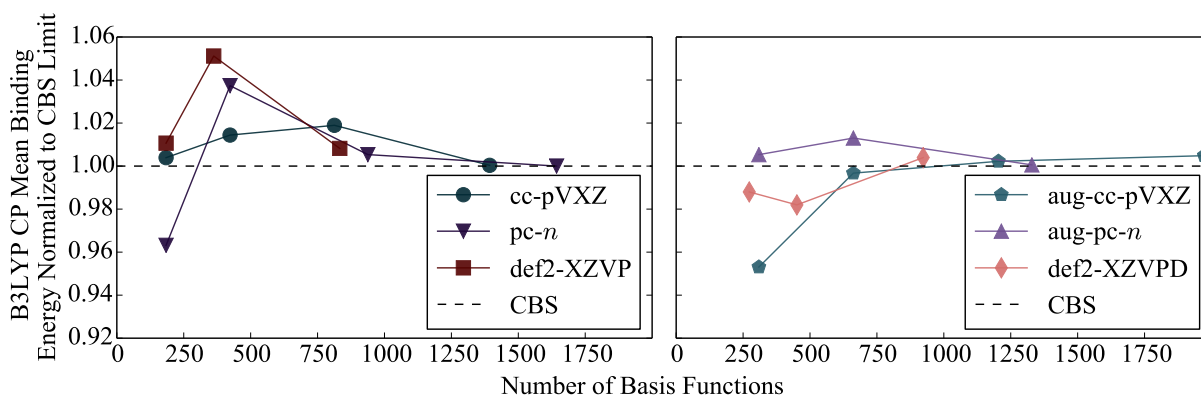


Figure B.4: Convergence of counterpoise-corrected (CP) B3LYP normalized mean binding energies across the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets.

B97M-V

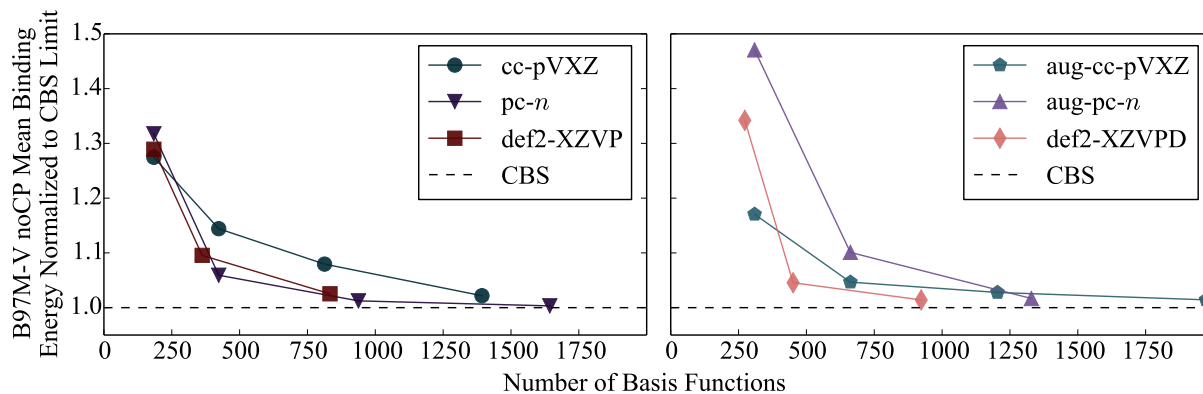


Figure B.5: Convergence of uncorrected (noCP) B97M-V normalized mean binding energies across the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets.

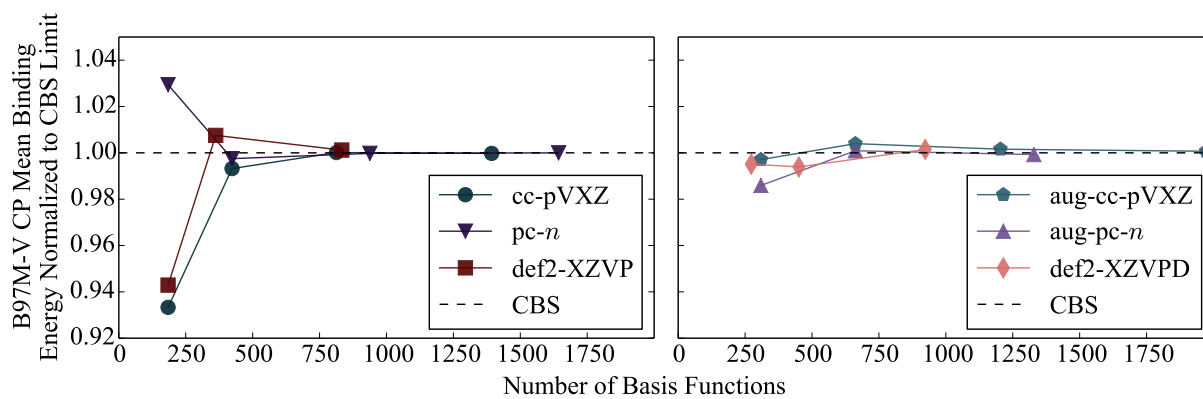


Figure B.6: Convergence of counterpoise-corrected (CP) B97M-V normalized mean binding energies across the S22 set of molecules for the Dunning, Jensen, and Karlsruhe sequences of basis sets.

Appendix C

DFT-D3 Study

C.1 Additional DFT-D3(op) Parameterizations

| Functional | s_6 | s_8 | α_1 | α_2 | β |
|------------|---------|---------|------------|------------|---------|
| PBE0 | 0.91826 | 0.00000 | 0.200 | 4.750 | 12 |
| PBE | 0.88290 | 0.00000 | 0.150 | 4.750 | 12 |
| revTPSS | 1.00000 | 0.27632 | 0.700 | 2.500 | 14 |
| revTPSSh | 1.00000 | 0.12467 | 0.575 | 3.000 | 16 |

Table C.1: Optimized values of DFT-D3(op) fit parameters for four additional density functionals.

C.2 Rare Gas Dimers (RG10)

| Functional | Correction | RG10 BL (\AA) | RG10 BE | RG10 |
|------------|------------|--------------------------|---------|------|
| MS2 | | 0.150 | 0.10 | 0.23 |
| | -D3(op) | 0.153 | 0.08 | 0.23 |
| | -D3(0) | 0.187 | 0.11 | 0.23 |
| MS2h | | 0.142 | 0.11 | 0.23 |
| | -D3(op) | 0.138 | 0.06 | 0.23 |
| | -D3(0) | 0.180 | 0.10 | 0.23 |
| TPSS | | 0.373 | 0.15 | 0.24 |
| | -D2 | 0.109 | 0.12 | 0.15 |
| | -D3(BJ) | 0.144 | 0.07 | 0.15 |
| | -D3(CSO) | 0.120 | 0.07 | 0.15 |
| | -D3(op) | 0.152 | 0.08 | 0.18 |
| | -D3(0) | 0.150 | 0.09 | 0.16 |
| | -D3(0) | 0.150 | 0.09 | 0.16 |
| TPSSh | | 0.382 | 0.16 | 0.24 |
| | -D3(BJ) | 0.147 | 0.06 | 0.15 |
| | -D3(op) | 0.153 | 0.06 | 0.18 |
| | -D3(0) | 0.145 | 0.07 | 0.15 |

Table C.2: Root mean square errors (RMSEs) for meta-GGAs across the RG10 set of rare gas dimers. BL and BE refer to interpolated equilibrium binding length and energy, respectively, and the rightmost column (RG10) is an RMSE across all 569 datapoints. RMSEs are in kcal/mol unless otherwise noted.

| Functional | Correction | RG10 BL (Å) | RG10 BE | RG10 |
|------------|------------|-------------|---------|------|
| B3LYP | | 3.677 | 0.21 | 0.28 |
| | -D2 | 0.140 | 0.06 | 0.16 |
| | -D3(BJ) | 0.062 | 0.04 | 0.15 |
| | -D3(CSO) | 0.514 | 0.07 | 0.15 |
| | -D3(op) | 0.021 | 0.06 | 0.13 |
| | -D3(0) | 0.072 | 0.05 | 0.16 |
| | -D3M(BJ) | 0.029 | 0.04 | 0.17 |
| B970 | | 0.673 | 0.18 | 0.51 |
| | -D2 | 0.167 | 0.07 | 0.36 |
| | -D3(BJ) | 0.220 | 0.05 | 0.26 |
| | -D3(op) | 0.234 | 0.03 | 0.24 |
| | -D3(0) | 0.198 | 0.03 | 0.27 |
| | -D3M(BJ) | 0.196 | 0.04 | 0.23 |
| B97h | | 0.250 | 0.14 | 0.21 |
| | -D2 | 0.118 | 0.10 | 0.15 |
| | -D3(op) | 0.100 | 0.07 | 0.13 |
| BLYP | | 3.697 | 0.21 | 0.39 |
| | -D2 | 0.839 | 0.10 | 0.22 |
| | -D3(BJ) | 1.006 | 0.09 | 0.14 |
| | -D3(CSO) | 1.100 | 0.11 | 0.14 |
| | -D3(op) | 1.069 | 0.12 | 0.14 |
| | -D3(0) | 0.588 | 0.08 | 0.20 |
| | -D3M(BJ) | 1.043 | 0.10 | 0.13 |
| revPBE | | 0.629 | 0.18 | 0.46 |
| | -D3(BJ) | 0.180 | 0.06 | 0.22 |
| | -D3(op) | 0.187 | 0.05 | 0.19 |
| | -D3(0) | 0.131 | 0.05 | 0.25 |
| revPBE0 | | 0.629 | 0.19 | 0.36 |
| | -D3(BJ) | 0.160 | 0.03 | 0.15 |
| | -D3(op) | 0.220 | 0.06 | 0.19 |
| | -D3(0) | 0.105 | 0.03 | 0.18 |

Table C.3: Root mean square errors (RMSEs) for GGAs across the RG10 set of rare gas dimers. BL and BE refer to interpolated equilibrium binding length and energy, respectively, and the rightmost column (RG10) is an RMSE across all 569 datapoints. RMSEs are in kcal/mol unless otherwise noted.

Appendix D

DFT-C Study

D.1 Additional Figures and Tables

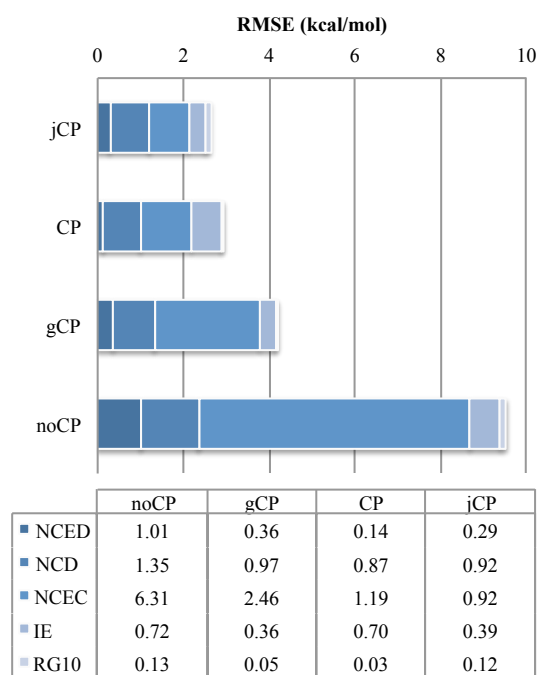


Figure D.1: Root-mean-square errors of B97M-V/def2-SVPD versus B97M-V/def2-QZVPPD with no correction (noCP), the standard counterpoise correction (CP), the geometrical counterpoise correction (gCP), and the treatment introduced here (DFT-C). All RMSEs are in units of kcal/mol. Methods are ordered from lowest overall RMSE at the top, to highest overall RMSE at the bottom.

| Dataset | noCP | CP | DFT-C | gCP |
|---------------|-------|------|-------|-------|
| S66 | 1.22 | 0.20 | 0.34 | 0.41 |
| A24 | 0.52 | 0.17 | 0.26 | 0.24 |
| DS14 | 0.86 | 0.16 | 0.27 | 0.29 |
| HB15 | 0.82 | 0.28 | 0.36 | 0.34 |
| HSG | 0.87 | 0.14 | 0.38 | 0.42 |
| NBC10 | 1.19 | 0.37 | 0.42 | 0.27 |
| S22 | 1.57 | 0.34 | 0.50 | 0.45 |
| X31 | 1.02 | 0.21 | 0.29 | 0.42 |
| A21x12 | 0.30 | 0.10 | 0.18 | 0.17 |
| BzDC215 | 0.70 | 0.28 | 0.28 | 0.26 |
| HW30 | 0.62 | 0.14 | 0.20 | 0.17 |
| NC15 | 0.43 | 0.10 | 0.23 | 0.25 |
| S66x8 | 0.95 | 0.25 | 0.36 | 0.42 |
| 3B-69-DIM | 1.01 | 0.26 | 0.32 | 0.38 |
| AlkBind12 | 1.34 | 0.20 | 0.73 | 0.60 |
| CO2Nitrogen16 | 0.94 | 0.21 | 0.46 | 0.31 |
| HB49 | 0.85 | 0.41 | 0.58 | 0.64 |
| Ionic43 | 1.27 | 1.11 | 1.41 | 1.90 |
| TA13 | 4.92 | 4.64 | 4.61 | 4.58 |
| XB18 | 2.09 | 1.20 | 1.41 | 1.51 |
| Bauza30 | 2.60 | 1.94 | 2.00 | 2.13 |
| CT20 | 0.92 | 0.32 | 0.51 | 0.53 |
| XB51 | 2.43 | 1.51 | 1.66 | 1.80 |
| H2O6Bind8 | 3.28 | 0.22 | 0.51 | 1.57 |
| HW6C1 | 2.64 | 0.44 | 0.84 | 1.42 |
| HW6F | 2.78 | 0.52 | 0.54 | 1.94 |
| FmH2O10 | 9.53 | 0.33 | 0.64 | 5.66 |
| Shields38 | 4.05 | 0.22 | 0.44 | 1.84 |
| SW49Bind345 | 1.17 | 1.17 | 0.35 | 1.39 |
| SW49Bind6 | 2.51 | 2.51 | 0.68 | 2.83 |
| Water27 | 3.27 | 1.08 | 0.96 | 1.61 |
| H2O20Bind4 | 18.11 | 0.49 | 1.83 | 8.75 |
| 3B-69-TRIM | 2.66 | 0.54 | 0.70 | 0.89 |
| CE20 | 2.84 | 0.55 | 0.64 | 1.37 |
| H2O20Bind10 | 18.72 | 2.63 | 0.58 | 11.23 |
| AlkIsomer11 | 0.46 | 0.46 | 0.37 | 0.36 |
| Butanediol65 | 0.37 | 0.37 | 0.18 | 0.12 |
| ACONF | 0.21 | 0.21 | 0.12 | 0.11 |
| CYCONF | 0.17 | 0.17 | 0.19 | 0.14 |
| Pentane14 | 0.07 | 0.07 | 0.15 | 0.11 |
| SW49Rel345 | 0.36 | 0.36 | 0.34 | 0.41 |
| SW49Rel6 | 0.38 | 0.38 | 0.42 | 0.70 |
| H2O16Rel5 | 1.44 | 1.44 | 1.02 | 0.36 |
| H2O20Rel10 | 0.82 | 0.13 | 0.64 | 0.39 |
| H2O20Rel4 | 2.59 | 0.49 | 1.88 | 0.62 |
| Melatonin52 | 0.73 | 0.73 | 0.49 | 0.70 |
| YMPJ519 | 0.83 | 0.83 | 0.51 | 0.47 |
| RG10 | 0.28 | 0.18 | 0.15 | 0.20 |

Figure D.2: Root-mean-square errors of B97M-V/def2-SVPD versus “exact” reference values with no correction (noCP), the standard counterpoise correction (CP), the geometrical counterpoise correction (gCP), and the treatment introduced here (DFT-C). All RMSEs are in units of kcal/mol. Note for SW49 and most of IE, the standard counterpoise correction is not possible, and so for these datasets the noCP and CP methods are identical.

D.2 gCP

Parameters

The following is the gcparr file for running gCP in the def2-SVPD basis. The first four numbers are σ , η , α , and β ; the following 36 lines are the number of basis functions followed by c_A , the missing energy terms for each atom.

```
# dft/svpd (comment line)
0.064821
0.000010825449487
0.576143886630077
0.946479745728949
8 0.002033712
8 0.006654316
12 0.245968569
12 0.093473481
20 0.193813847
20 0.059562241
20 0.068548897
23 0.095407956
23 0.12694673
23 0.163230942
18 0.370363706
21 0.274630257
24 0.343131329
24 0.23920287
24 0.147501274
27 0.154971698
27 0.160130459
27 0.180879081
27 0.336846485
27 0.404584507
34 0.383701913
34 0.336299149
34 0.18248286
34 0.274357288
34 0.108694613
34 0.277306996
34 0.262711007
34 0.317486445
34 0.314512981
34 0.300910789
38 0.501994778
38 0.423565755
38 0.355940629
41 0.333722951
41 0.3631844716
41 0.3815829321
```

gCPmod

As mentioned in the text, the low value of $\eta = 0.0000108$ suggests the overlap term is superfluous for gCP in the def2-SVPD basis. Moreover, since β is so close to unity, we also might be able to just assume exponential decay of BSSE. These approximations lead to a modified form of gCP, gCPmod:

$$E_{\text{gCPmod}} = \sigma \sum_A E_A^{\text{miss}} \sum_{B \neq A} \frac{1}{\sqrt{N_B^{\text{virt}}}} \exp(-\alpha r_{AB}) \quad (\text{D.1})$$

Here, the linear atomic coefficients c_A are the same as in gCP, namely E_A^{miss} ; the pairwise atomic coefficients now assume the atomic overlap $S_{AB} = 1$, and we have assumed $\beta = 1$. We are left with a much simpler model, with only one linear parameter (σ) and one nonlinear parameter (α), which we fit, as in gCP, to B3LYP BSSEs across S66x8.

The optimized parameters for gCPmod in the def2-SVPD basis are:

- $\sigma = 0.08032$
- $\alpha = 0.53848$

Root-mean-square deviations of gCP and gCPmod predicted BSSEs relative to Boys-Bernardi B3LYP/def2-SVPD BSSEs are given in Table D.2.

| Datatype | gCP | gCPmod |
|----------|------|--------|
| NCED | 0.32 | 0.32 |
| NCD | 0.45 | 0.48 |
| NCEC | 1.91 | 1.52 |
| RG10 | 0.14 | 0.14 |

Table D.1: RMSEs (kcal/mol) for gCP and gCPmod predicted BSSEs in B3LYP/def2-SVPD versus Boys-Bernardi BSSEs.

D.3 DFT-C

Python Implementation

The following python file, `dftc.py`, calculates the DFT-C empirical correction for BSSE within the def2-SVPD basis, given an `.xyz` file.

The program can be run for a meta-GGA functional, e.g., via `./dftc.py file.xyz -level meta-gga`. If the `-level` tag is omitted, the program defaults to GGA.

```
#!/usr/bin/env python

"""
*****
*-----DFT-C-----*
*****
  Copyright (C) 2017, Jonathon Witte, UC Berkeley

  This program is free software; you can redistribute it and/or modify
  it under the terms of the GNU General Public License as published by
  the Free Software Foundation; either version 1, or (at your option)
  any later version.

  This program is distributed in the hope that it will be useful,
  but WITHOUT ANY WARRANTY; without even the implied warranty of
  MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
  GNU General Public License for more details.
"""

import os, sys, numpy, math

# covalent radius
def covalent_radius(atom):
    covrad = [
        31,28, \
        127,96,84,76,71,66,57,58, \
        166,141,121,111,107,105,102,106,203,176,170,160,153,139,150,142,138,124, \
        132,122,122,120,119,120,120,116 \
    ]
    return covrad[atom-1]/100.0
# end of covalent_radius

# van der waals radius
def vdwradius(atom):
    vdwrad = [
        120,140, \
        182,153,192,170,155,152,147,154, \
        227,173,184,210,180,180,175,188,275,231,211,320,306,278,300,284,276,163, \
        140,139,187,211,185,190,185,202 \
    ]
    return vdwrad[atom-1]/100.0
# end of vdwradius

# read in xyz file
def read_xyz(filename):
    f = open(filename, 'r')
    lines = f.readlines()
    f.close()
    lines.pop(0) # skip number of atoms
    lines.pop(0) # skip comment line
```

```

coords = []
for line in lines:
    (at,x,y,z) = line.split()
    coords.append([at,[float(x),float(y),float(z)]])
return coords
# end of read_xyz

# convert atomic symbol to number
def atomic_number(atom):
    atoms = [
        "H", "He", "Li", "Be", "B", "C", "N", "O", "F", "Ne",
        "Na", "Mg", "Al", "Si", "P", "S", "Cl", "Ar", "K", "Ca",
        "Sc", "Ti", "V", "Cr", "Mn", "Fe", "Co", "Ni", "Cu", "Zn",
        "Ga", "Ge", "As", "Se", "Br", "Kr", "Rb", "Sr", "Y", "Zr",
        "Nb", "Mo", "Tc", "Ru", "Rh", "Pd", "Ag", "Cd", "In", "Sn",
        "Sb", "Te", "I", "Xe", "Cs", "Ba", "La", "Ce", "Pr", "Nd",
        "Pm", "Sm", "Eu", "Gd", "Tb", "Dy", "Ho", "Er", "Tm", "Yb",
        "Lu", "Hf", "Ta", "W", "Re", "Os", "Ir", "Pt", "Au", "Hg",
        "Tl", "Pb", "Bi", "Po", "At", "Rn", "Fr", "Ra", "Ac", "Th",
        "Pa", "U", "Np", "Pu",
    ]
    try:
        atno = int(atom)
    except ValueError:
        # given as a symbol
        atom = atom.upper()
        if len(atom) > 1:
            atom = atom[0]+atom[1].lower()
        try:
            atno = atoms.index(atom)+1
        except ValueError:
            print "DFT-C_only_supports_atoms_up_to_Pu"
            sys.exit(1)
# atno is now atomic number; remap to lower homologue if >36
if atno > 94:
    print "DFT-C_only_supports_atoms_up_to_Pu"
    sys.exit(1)
elif atno > 89:
    atno = 21
elif atno > 71:
    atno = atno - 2*18 - 14
elif atno > 57:
    atno = 21
elif atno > 54:
    atno = atno - 2*18
elif atno > 36:
    atno = atno - 18
return atno
# end atomic_number

# process the atom tags in the xyz file
def read_atoms(coords):
    for i,coord in enumerate(coords):
        atom = coord[0]
        # convert atom symbol to atomic number
        # also, if >36, remap to lower homologue
        coords[i][0] = atomic_number(atom)
    return coords
# end read_atoms

# number of basis functions
def num_basis(atom):
    # assumes def2-svpd basis

```

```

bas = [
    8,8, \
    12,12,20,20,20,23,23,23, \
    18,21,24,24,24,27,27,27,27,27,34,34,34,34,34,34,34, \
    34,34,38,38,38,41,41,41 \
]
return bas[atom-1]
# end of num_bas

# distance between two points c = (x,y,z)
def distance(c1, c2):
    (x1,y1,z1) = c1
    (x2,y2,z2) = c2
    return numpy.sqrt( \
        (float(x1)-float(x2))**2 + \
        (float(y1)-float(y2))**2 + \
        (float(z1)-float(z2))**2 \
    )
# end of distance

# terfc function
def terfc(x,y):
    return 1.0 - 0.5 * (math.erf(x+y) + math.erf(x-y))
# end terfc

# calculate dft-c empirical correction for bsse given a system
# system is a list of [atom,[x,y,z]]
def calculate_bsse(system):
    # conversion factors
    angs_to_bohr = 1.88973
    # parameters
    ca_list = set_linear_coeffs() # 1x36
    cab_list = set_coeffs() # 36x36
    alpha_list = set_alpha() # 36x36
    beta_list = set_beta() # 36x36
    # energy
    dftc_bsse = 0.0
    # loop over atom 1
    for a1 in system:
        a1_energy = 0.0
        # loop over atom 2
        for a2 in system:
            # no self-BSSE
            if a2 == a1:
                continue
            r12 = distance(a1[1],a2[1]) * angs_to_bohr
            rvdw = (vdw_radius(a1[0])+vdw_radius(a2[0])) * angs_to_bohr
            rcov = (covalent_radius(a1[0])+covalent_radius(a2[0])) * angs_to_bohr
            dampk1 = 19.0
            dampk2 = -5.8889/numpy.log10(rvdw/rcov)
            dampfn = 1.0 / (1.0 + dampk1 * (r12 / rcov) ** dampk2)
            # many-body term
            mbdy = 0.0
            # loop over atom 3
            for a3 in system:
                if a1 == a3 or a2 == a3:
                    continue
                num_virts_3 = num_basis(a3[0]) - a3[0]/2.0
                num_virts_2 = num_basis(a2[0]) - a2[0]/2.0
                r13 = distance(a1[1],a3[1]) * angs_to_bohr
                r23 = distance(a2[1],a3[1]) * angs_to_bohr
                mbdy = mbdy + (num_virts_3 / num_virts_2) * terfc(r13,r12) * terfc(r23,r12)
            # end loop over atom 3

```

```

    mbdy = 1.0 / (1.0 + mbdy)
    # array index for (a1, a2) element
    ind = 36*(a1[0]-1)+a2[0]-1
    ca = ca_list[a1[0]-1]
    cab = cab_list[ind]
    alp = alpha_list[ind]
    bet = beta_list[ind]
    al_energy += ca * cab * mbdy * \
                ((dampfn * numpy.exp(alp*r12**2.0 + bet*r12)) + \
                ((1.0-dampfn) * numpy.exp(alp*rcov**2.0 + bet*rcov)))
    dftc_bsse += al_energy
    return dftc_bsse
# end of calculate_bsse

# main function
def main():
    # ensure the program is being called correctly
    # i.e. via "python dftc.py file.xyz"
    usage = "usage: %s_filename_[-level_gga]" % os.path.basename(sys.argv[0])
    if len(sys.argv) < 2:
        print >> sys.stderr, usage
        sys.exit(1)
    filename = sys.argv[1]
    print "Running_DFT-C_on_%s" % filename
    if "-level" in sys.argv:
        level = sys.argv[sys.argv.index("-level")+1].upper()
        if level == "LDA" or level == "LSDA" or level == "L":
            sigma = 1.0
            print "LDA_level"
        elif level == "GGA" or level == "G":
            sigma = 0.9
            print "GGA_level"
        elif level == "METAGGA" or level == "META" or level == "M" \
              or level == "MGGA" or level == "META-GGA":
            sigma = 0.85
            print "meta-GGA_level"
        else:
            print "Unrecognized_level_of_DFT:_Defaulting_to_GGA"
            sigma = 0.9
    else:
        print "No_level_of_DFT_specified;_defaulting_to_GGA_(sigma=0.9)"
        print "For_LDA_or_meta-GGA,_use_'-level_LDA'_or_'-level_meta-GGA'"
        sigma = 0.9
    if not os.path.isfile(filename):
        print "File_not_found:%s" % filename
        sys.exit(1)
    # read in the xyz file
    coordinates = read_xyz(filename)
    # process the coordinates
    # change atom symbols to atomic numbers
    # also map elements >36 to lower homologues
    coordinates = read_atoms(coordinates)
    # calculate dft-c correction for bsse
    dftc_bsse = sigma * calculate_bsse(coordinates)
    print "DFT-C_energy_=_%.8f_Hartree,_%.4f_kcal/mol" \
          % (dftc_bsse, 627.5095*dftc_bsse)
# end main

# set linear atomic coefficients
def set_linear_coeffs():
    coef = [

```

```

0.938,1.000, \
1.000,1.000,1.000,0.601,0.339,0.483,1.000,1.000, \
1.000,1.000, \
1.000,1.000,1.000,1.000,1.000,1.000,1.000,1.000,1.000,1.000, \
1.000,1.000,1.000,1.000,1.000,1.000, \
1.000,1.000,1.000,1.000,1.000,1.000,1.000,1.000 \
]
return coef
# end of set_linear_coeffs()

# set pairwise multiplicative constants
def set_coeffs():
coef = [
4.7808E-05,8.9242E-05,4.2334E-04,5.6254E-04,6.5537E-04,6.0955E-04, \
3.8127E-04,4.9051E-04,3.1840E-04,1.1761E-04,3.8246E-04,4.8047E-04, \
4.1553E-04,3.9824E-04,3.2342E-04,4.4509E-04,4.2454E-04,3.7003E-04, \
3.6533E-04,3.1426E-04,4.1569E-04,4.3617E-04,4.1593E-04,3.6462E-04, \
4.1800E-04,4.0159E-04,3.8876E-04,3.8246E-04,3.6980E-04,3.6940E-04, \
4.0159E-04,3.6812E-04,3.2207E-04,3.7482E-04,3.6326E-04,3.5920E-04, \
1.5669E-03,1.0709E-04,2.4815E-03,2.6725E-03,1.5952E-03,1.6282E-03, \
2.1772E-03,1.6411E-03,1.5625E-03,1.5891E-03,2.7171E-03,2.1633E-03, \
1.2912E-03,1.2245E-03,1.2805E-03,1.0900E-03,1.0967E-03,1.1158E-03, \
4.6517E-03,1.7691E-03,1.3856E-03,1.2169E-03,1.3518E-03,1.5917E-03, \
1.1992E-03,1.3922E-03,1.4636E-03,1.5777E-03,1.5071E-03,1.5624E-03, \
1.2406E-03,1.2691E-03,1.2437E-03,1.0671E-03,1.0336E-03,1.0483E-03, \
6.4899E-04,6.0015E-04,2.2087E-04,2.4223E-04,2.6215E-03,7.9919E-04, \
7.2836E-04,8.2357E-04,9.6995E-04,7.9457E-04,2.4103E-04,2.7012E-04, \
5.9959E-04,9.9967E-04,2.2850E-03,4.5115E-04,4.7760E-04,4.3314E-04, \
5.2151E-04,3.2239E-04,2.5091E-04,2.3091E-04,2.2589E-04,2.5944E-04, \
2.1243E-04,2.3091E-04,2.4175E-04,2.6773E-04,2.4860E-04,2.6342E-04, \
8.3186E-04,6.3107E-04,1.2026E-03,3.9776E-04,4.1195E-04,4.2884E-04, \
6.2987E-04,7.4405E-04,7.0612E-04,7.0597E-04,5.0924E-03,1.1866E-03, \
1.2490E-03,2.3254E-03,8.0190E-04,7.5154E-04,9.5616E-04,1.2526E-03, \
3.2685E-03,5.1419E-03,4.9802E-03,1.1354E-03,9.9058E-04,1.5404E-03, \
5.5497E-04,4.0908E-04,1.3145E-03,1.3855E-03,1.4378E-03,1.6127E-03, \
1.4424E-03,1.4390E-03,1.3721E-03,1.4024E-03,1.5134E-03,9.4469E-04, \
2.8255E-03,4.6243E-03,5.0845E-03,1.3195E-03,1.0149E-03,1.0446E-03, \
4.9003E-04,5.7338E-04,6.3919E-04,6.8525E-04,9.7050E-04,8.8859E-04, \
8.6246E-04,1.1294E-03,2.0165E-03,2.0969E-03,8.7250E-04,6.4302E-04, \
7.9202E-04,7.0166E-04,6.3370E-04,3.7545E-04,3.9211E-04,3.9872E-04, \
1.7430E-03,5.4182E-04,5.2055E-04,3.4820E-04,3.4063E-04,3.6812E-04, \
2.8438E-04,3.0788E-04,3.1035E-04,3.4422E-04,3.2940E-04,3.4517E-04, \
7.4222E-04,7.0820E-04,6.2700E-04,3.2892E-04,3.1458E-04,3.4087E-04, \
2.0044E-03,6.3473E-04,2.4212E-03,2.8796E-03,2.5601E-03,1.0484E-03, \
7.3624E-04,2.4435E-03,2.2465E-03,1.3367E-03,3.1733E-03,2.5434E-03, \
1.4948E-03,1.1866E-03,7.4676E-04,1.4564E-03,9.1098E-04,6.5115E-04, \
6.6903E-03,1.9831E-03,1.8163E-03,1.2022E-03,1.1080E-03,1.3568E-03, \
8.8740E-04,8.5672E-04,8.5266E-04,9.7528E-04,8.9807E-04,8.2055E-04, \
1.3733E-03,1.0062E-03,7.6405E-04,1.4151E-03,9.5122E-04,7.7704E-04, \
3.4577E-03,8.0764E-04,6.6329E-03,5.2971E-03,3.3912E-03,3.2862E-03, \
1.1806E-03,2.8795E-03,2.8656E-03,2.9431E-03,9.8664E-03,6.5736E-03, \
5.2513E-03,4.9387E-03,2.6373E-03,3.9665E-03,3.9004E-03,3.3963E-03, \
3.8346E-03,7.5747E-03,6.8383E-03,4.0795E-03,3.5426E-03,4.3579E-03, \
2.6827E-03,2.5735E-03,2.5286E-03,3.1681E-03,2.7471E-03,2.6070E-03, \
5.3425E-03,5.2238E-03,3.4982E-03,4.8535E-03,4.6999E-03,4.2994E-03, \
5.8565E-04,6.4366E-04,6.5927E-04,1.1825E-03,9.4540E-04,3.3133E-03, \
4.0336E-03,1.4958E-03,1.4358E-03,1.7440E-03,6.7728E-04,7.3863E-04, \
1.0291E-03,1.0051E-03,2.2166E-03,7.0852E-04,9.3457E-04,1.0757E-03, \
8.9242E-04,6.6915E-04,1.2995E-03,1.5981E-03,1.5311E-03,1.3673E-03, \
1.7542E-03,1.8875E-03,2.0189E-03,1.5426E-03,1.9204E-03,1.8076E-03, \
1.1761E-03,9.4309E-04,1.7097E-03,7.1521E-04,9.1377E-04,1.0336E-03, \
9.0118E-04,8.2325E-04,8.7178E-04,1.2589E-03,7.3624E-04,1.8422E-03, \
4.0835E-03,2.4374E-03,2.1286E-03,1.1789E-03,1.2271E-03,1.4462E-03, \

```

1.6231E-03,9.4548E-04,1.2911E-03,5.6318E-04,7.6063E-04,1.1761E-03, \\
1.5813E-03,1.6417E-03,2.1571E-03,2.5914E-03,2.3593E-03,1.6688E-03, \\
2.5765E-03,2.7107E-03,2.4240E-03,1.4948E-03,1.9204E-03,1.7561E-03, \\
1.5346E-03,1.2446E-03,1.2163E-03,6.4254E-04,8.0142E-04,1.0593E-03, \\
4.2669E-04,5.6445E-04,1.0391E-03,1.5458E-03,1.6816E-03,9.5090E-04, \\
2.5348E-03,1.6895E-03,3.2047E-03,1.9219E-03,1.6354E-03,2.4430E-03, \\
2.0777E-03,1.9181E-03,1.3778E-03,7.6907E-04,8.9624E-04,1.5232E-03, \\
2.3036E-03,2.5791E-03,2.9810E-03,3.3098E-03,2.8398E-03,2.1737E-03, \\
3.0562E-03,2.9985E-03,2.7278E-03,1.7625E-03,2.2312E-03,2.0801E-03, \\
1.8896E-03,2.6855E-03,2.1208E-03,1.0403E-03,1.2995E-03,1.4185E-03, \\
7.2071E-04,5.8676E-04,8.1991E-04,3.8884E-04,1.1239E-03,1.1405E-03, \\
9.7552E-04,1.2444E-03,1.1911E-03,1.0527E-03,7.4102E-04,6.9800E-04, \\
1.2243E-03,1.0972E-03,5.4088E-03,9.4166E-04,9.9353E-04,1.0499E-03, \\
1.1072E-03,9.4548E-04,7.2278E-04,5.5346E-04,4.8406E-04,5.4692E-04, \\
3.8717E-04,3.6478E-04,3.5609E-04,3.8565E-04,3.8533E-04,3.6637E-04, \\
1.3087E-03,1.1499E-03,5.2319E-03,8.7967E-04,9.2126E-04,9.9329E-04, \\
1.0028E-03,4.0605E-04,1.9226E-03,1.2765E-03,1.6761E-03,9.6174E-04, \\
7.3098E-04,1.7350E-03,1.5743E-03,1.2822E-03,2.6334E-03,3.3633E-03, \\
1.8350E-03,1.7067E-03,1.1570E-03,1.1839E-03,1.3355E-03,1.5031E-03, \\
1.3805E-03,1.5740E-03,2.8612E-03,1.3562E-03,9.6453E-04,1.6513E-03, \\
4.0772E-04,3.0119E-04,2.4502E-04,3.6334E-04,3.7290E-04,2.5434E-04, \\
1.9506E-03,1.9002E-03,1.4973E-03,1.1589E-03,1.1946E-03,1.3048E-03, \\
2.9880E-04,1.2505E-03,8.0150E-04,9.4182E-04,9.8166E-04,9.6174E-04, \\
7.2573E-04,1.0507E-03,9.3752E-04,7.3720E-04,1.0338E-03,9.5855E-04, \\
1.4083E-03,6.0413E-04,4.7593E-04,8.6294E-04,5.8374E-04,4.2167E-04, \\
1.5617E-03,7.0437E-04,9.5871E-04,5.4979E-04,4.9840E-04,7.7130E-04, \\
3.8374E-04,3.7816E-04,3.7569E-04,4.8127E-04,4.4748E-04,3.8151E-04, \\
1.6996E-03,7.1170E-04,4.7282E-04,8.6437E-04,5.5051E-04,4.5816E-04, \\
2.3187E-04,4.7410E-04,2.6220E-03,2.3187E-03,9.9281E-04,6.6915E-04, \\
7.7831E-04,1.4473E-03,1.0083E-03,7.2859E-04,2.6055E-03,2.1872E-03, \\
2.0330E-03,7.1250E-04,4.1808E-04,1.1294E-03,6.9282E-04,4.8764E-04, \\
4.6378E-03,2.4951E-03,1.9137E-03,1.2022E-03,1.1223E-03,1.3112E-03, \\
9.3536E-04,9.1027E-04,8.8206E-04,1.0008E-03,9.1983E-04,7.8421E-04, \\
1.9900E-03,9.1505E-04,4.5569E-04,1.4381E-03,9.3250E-04,6.6342E-04, \\
7.9720E-04,2.3314E-04,4.2602E-03,3.7466E-03,2.6131E-03,5.5266E-04, \\
6.3895E-04,2.2769E-03,2.0431E-03,9.6955E-04,5.0238E-03,3.6454E-03, \\
2.9059E-03,1.5442E-03,5.4955E-04,2.4647E-03,1.7129E-03,8.7489E-04, \\
1.0678E-02,3.2401E-03,3.7262E-03,2.8956E-03,2.7215E-03,2.9064E-03, \\
2.3093E-03,2.2355E-03,2.1921E-03,2.4414E-03,2.2592E-03,2.0347E-03, \\
2.7537E-03,2.1000E-03,7.9146E-04,2.5357E-03,2.3219E-03,1.7044E-03, \\
1.0733E-03,2.3337E-03,1.0207E-03,1.8709E-03,6.1549E-03,7.4718E-03, \\
5.0143E-03,3.4094E-03,4.3161E-03,5.9234E-03,1.3984E-03,2.5529E-03, \\
4.2167E-03,6.3258E-03,1.2753E-03,3.3397E-03,3.7176E-03,4.9309E-03, \\
6.4979E-04,5.1202E-04,4.1344E-03,2.9139E-03,2.5637E-03,3.7408E-03, \\
2.1517E-03,2.4028E-03,2.1463E-03,2.5721E-03,2.6010E-03,2.1467E-03, \\
5.0557E-03,7.0504E-03,8.6924E-03,3.4192E-03,2.9023E-03,3.4674E-03, \\
1.0984E-03,4.2412E-03,9.3202E-04,2.3745E-03,4.7744E-03,9.1893E-03, \\
8.8297E-03,3.5547E-03,4.2277E-03,5.1833E-03,1.0677E-03,1.5681E-03, \\
1.8880E-03,5.1555E-03,1.2017E-03,2.9703E-03,2.8686E-03,3.2557E-03, \\
8.5058E-04,4.4971E-04,2.8087E-03,2.5071E-03,2.3020E-03,2.8258E-03, \\
2.2442E-03,2.4263E-03,2.5906E-03,3.1553E-03,2.7377E-03,3.2277E-03, \\
2.0044E-03,4.2492E-03,7.6302E-03,2.3445E-03,2.4642E-03,2.4484E-03, \\
6.9800E-04,7.1728E-04,9.2198E-04,2.1641E-03,2.2506E-03,5.5781E-03, \\
7.3204E-03,2.2406E-03,2.2848E-03,2.8111E-03,8.9839E-04,1.1832E-03, \\
1.1601E-03,2.1944E-03,4.9144E-03,2.0650E-03,2.5030E-03,2.8427E-03, \\
1.4753E-03,6.9354E-04,1.8088E-03,1.7337E-03,1.6637E-03,1.8932E-03, \\
1.7645E-03,1.9377E-03,2.0875E-03,2.5721E-03,2.1815E-03,2.6130E-03, \\
1.2155E-03,1.5847E-03,2.8746E-03,1.5375E-03,1.8013E-03,2.0782E-03, \\
4.8047E-04,1.1300E-03,3.4302E-04,3.7768E-04,7.7250E-04,6.4206E-04, \\
4.7067E-04,7.2365E-04,7.3114E-04,5.7656E-04,4.0238E-04,5.6413E-04, \\
5.6055E-04,5.8246E-04,5.2589E-04,6.1768E-04,6.0143E-04,5.5362E-04, \\
4.7967E-04,4.7951E-04,3.5203E-04,3.1888E-04,3.2629E-04,3.7163E-04, \\
3.1179E-04,3.2796E-04,3.3322E-04,3.4103E-04,3.2318E-04,3.3306E-04, \\

2.4669E-03,5.8198E-04,5.7218E-04,6.5401E-04,6.4039E-04,6.2310E-04, \\
4.5179E-04,3.8820E-04,3.9155E-04,5.4023E-04,4.3199E-03,7.3146E-04, \\
6.3107E-04,1.0325E-03,8.8740E-04,7.1712E-04,5.2390E-04,8.5337E-04, \\
1.2605E-03,9.4548E-04,4.6042E-03,1.0446E-03,1.0761E-03,1.0929E-03, \\
6.9162E-04,6.4206E-04,4.4940E-04,4.1784E-04,3.8366E-04,4.8031E-04, \\
3.5976E-04,3.7816E-04,3.9856E-04,4.7171E-04,4.1330E-04,4.8143E-04, \\
1.3554E-03,1.0938E-03,5.6225E-03,1.0288E-03,1.0411E-03,1.0556E-03, \\
1.3291E-03,9.0580E-04,6.8387E-03,2.5625E-03,7.6134E-04,9.9154E-04, \\
3.4814E-03,4.3298E-04,6.3385E-04,1.3711E-03,1.3105E-02,1.9482E-03, \\
1.1127E-03,7.6397E-04,8.8086E-04,2.7251E-04,2.6239E-04,3.2988E-04, \\
5.1091E-03,2.0589E-03,6.0706E-03,2.1808E-03,1.8609E-03,1.8020E-03, \\
1.2266E-03,1.0441E-03,8.9186E-04,7.9361E-04,9.1983E-04,6.2374E-04, \\
1.2908E-03,8.2039E-04,7.0923E-04,3.2892E-04,2.5466E-04,2.5290E-04, \\
1.5609E-03,1.6833E-03,2.1810E-02,5.8645E-03,3.3605E-03,2.4030E-03, \\
8.4090E-03,1.5624E-03,1.7601E-03,3.3676E-03,5.8147E-03,8.8230E-03, \\
3.4414E-03,2.6766E-03,2.6321E-03,1.3716E-03,1.2648E-03,1.4715E-03, \\
7.1059E-03,1.0769E-02,2.6230E-02,7.6128E-03,6.1744E-03,5.0415E-03, \\
3.6969E-03,3.1157E-03,2.6298E-03,2.4637E-03,2.4239E-03,1.9015E-03, \\
3.5211E-03,3.1027E-03,2.4737E-03,1.4534E-03,1.2246E-03,1.1179E-03, \\
2.5850E-02,4.4011E-02,4.2399E-02,1.5703E-02,3.6628E-02,4.3370E-02, \\
6.0193E-02,3.7000E-02,4.6699E-02,6.3677E-02,8.5776E-03,6.6515E-02, \\
2.3872E-02,4.7022E-02,5.8463E-02,1.8276E-02,2.6171E-02,4.4837E-02, \\
9.7608E-03,6.1733E-02,7.9580E-03,2.3689E-02,1.8197E-02,1.8504E-02, \\
1.1194E-02,1.0033E-02,9.0623E-03,9.3576E-03,1.0047E-02,8.2600E-03, \\
1.9301E-02,3.7794E-02,5.5091E-02,1.3887E-02,1.9346E-02,2.6713E-02, \\
6.9824E-03,1.5829E-03,2.1517E-02,1.1558E-02,1.9767E-02,2.0570E-02, \\
1.3484E-02,2.1616E-02,2.3058E-02,2.4267E-02,8.5230E-02,3.4627E-02, \\
1.3372E-02,2.4558E-02,3.0717E-02,1.0806E-02,1.5702E-02,2.6407E-02, \\
1.0899E-02,3.7175E-02,3.5532E-02,1.4119E-02,1.1639E-02,1.1899E-02, \\
9.2063E-03,8.6308E-03,8.3208E-03,8.6182E-03,9.1889E-03,8.7141E-03, \\
1.0746E-02,1.9872E-02,3.2145E-02,9.0032E-03,1.1381E-02,1.5519E-02, \\
5.3163E-02,3.9166E-03,4.2234E-03,2.3906E-02,7.7794E-02,1.5490E-01, \\
7.6911E-03,6.9586E-02,9.4869E-02,8.1723E-03,5.9043E-03,8.3282E-03, \\
4.1723E-02,1.0547E-01,1.9482E-01,4.1179E-02,5.6817E-02,9.5199E-02, \\
5.7672E-03,8.0163E-03,5.5987E-03,4.6806E-03,3.3518E-02,4.9994E-03, \\
1.5983E-02,1.5207E-02,1.2016E-02,1.4228E-02,1.5960E-02,1.1203E-02, \\
3.4472E-02,7.9202E-02,1.4951E-01,3.2138E-02,4.4453E-02,5.8670E-02, \\
1.9543E-02,1.2629E-02,1.4327E-02,4.9386E-02,1.7141E-01,2.5049E-02, \\
2.2334E-02,2.3275E-02,2.2061E-02,2.1456E-02,1.8814E-02,2.8588E-01, \\
6.8256E-02,4.1742E-01,2.8453E-02,8.1022E-02,2.2104E-01,2.8168E-02, \\
1.9965E-02,2.8085E-01,1.9837E-02,1.5334E-01,1.2727E-01,1.3480E-01, \\
7.4641E-02,6.9378E-02,7.2695E-02,7.3726E-02,8.1520E-02,8.2439E-02, \\
5.3465E-02,2.0838E-01,2.9366E-02,5.3992E-02,9.3444E-02,2.1391E-01, \\
2.0961E-02,1.5249E-02,1.4141E-02,6.5334E-02,2.5875E-02,2.6075E-02, \\
2.3328E-02,2.4937E-02,2.2604E-02,2.2948E-02,1.8579E-02,2.4162E-02, \\
1.3458E-01,2.6354E-02,3.0055E-02,1.6474E-01,4.5233E-01,2.7646E-02, \\
1.9524E-02,2.1776E-02,1.9508E-02,1.5475E-02,1.8203E-01,1.6615E-02, \\
1.0004E-01,1.2543E-02,8.1603E-02,1.2857E-01,9.8754E-02,1.0479E-01, \\
1.3611E-01,3.8290E-01,3.0675E-02,1.0938E-01,2.2374E-01,3.7973E-01, \\
1.8634E-01,2.2721E-02,1.6333E-02,8.1092E-02,5.4050E-01,3.6511E-02, \\
3.1814E-02,3.5607E-02,3.4930E-02,3.2792E-02,2.2832E-02,3.1015E-02, \\
1.9934E-01,3.7280E-02,3.8343E-02,3.6435E-01,3.6882E-02,3.6966E-02, \\
2.2561E-02,2.9901E-02,2.4189E-02,2.1497E-02,3.1427E-01,2.2185E-02, \\
1.8215E-01,1.7354E-01,1.6131E-01,1.5199E-01,1.6097E-01,1.4829E-01, \\
1.5972E-01,6.5145E-01,3.9727E-02,2.3074E-01,4.1908E-01,7.4941E-01, \\
2.1742E-01,1.7069E-02,1.2541E-01,5.1072E-02,2.6733E-01,2.9719E-02, \\
2.5637E-02,2.7705E-02,2.6920E-02,2.5925E-02,2.0253E-02,3.6806E-01, \\
1.1330E-01,5.5374E-01,3.3152E-02,1.5440E-01,3.6483E-01,3.1955E-02, \\
2.0498E-02,4.0305E-01,2.1683E-02,1.7440E-01,1.3009E-01,1.5556E-01, \\
8.8383E-02,8.3052E-02,7.9830E-02,7.6171E-02,8.0218E-02,8.2079E-02, \\
8.1519E-02,3.0637E-01,3.4848E-02,9.8867E-02,1.7831E-01,3.6433E-01, \\
1.5773E-01,1.6920E-02,8.3447E-03,6.8107E-02,2.0740E-02,2.3743E-02, \\
2.1822E-02,2.1916E-02,2.1958E-02,2.1430E-02,1.2593E-02,1.7340E-02, \\

```

1.7405E-01,2.2041E-02,2.4548E-02,2.1083E-02,2.1743E-02,2.2807E-02, \
1.2081E-02,1.6133E-02,1.3770E-02,1.2198E-02,1.0950E-02,1.3189E-02, \
9.6414E-03,9.6850E-03,1.1449E-01,1.0486E-02,1.0034E-02,9.2028E-02, \
1.5372E-01,2.1968E-02,2.4700E-02,2.2087E-01,2.1215E-02,2.1746E-02, \
6.6334E-04,7.0389E-04,5.3545E-04,6.0079E-04,7.9760E-04,9.0485E-04, \
9.6238E-04,1.1627E-03,1.8455E-03,1.5777E-03,6.8923E-04,6.0955E-04, \
8.1154E-04,4.2533E-04,4.6015E-04,5.9648E-04,4.3633E-04,3.0980E-04, \
1.2271E-03,4.1450E-04,5.0931E-04,3.5187E-04,3.2988E-04,4.5928E-04, \
2.8095E-04,2.8111E-04,2.8095E-04,3.5378E-04,3.1697E-04,2.9067E-04, \
9.8604E-04,5.1888E-04,4.2485E-04,5.1633E-04,4.0820E-04,3.0422E-04, \
2.4263E-04,1.0798E-03,1.4240E-03,1.4550E-03,6.2748E-04,7.4230E-04, \
1.0728E-03,8.8716E-04,1.0672E-03,9.2652E-04,1.6872E-03,1.4390E-03, \
1.2271E-03,5.0661E-04,4.8645E-04,6.4493E-04,5.6605E-04,4.3601E-04, \
3.1513E-03,1.2354E-03,1.0037E-03,6.6342E-04,6.3465E-04,7.2573E-04, \
5.5505E-04,5.8565E-04,5.4557E-04,6.0557E-04,5.9354E-04,5.5712E-04, \
1.2801E-03,5.9250E-04,4.2828E-04,7.4581E-04,5.5425E-04,4.6916E-04, \
2.8924E-04,5.1202E-04,2.8780E-03,2.9163E-03,1.0235E-03,6.8812E-04, \
1.2595E-03,1.0022E-03,9.4931E-04,7.6015E-04,3.2948E-03,2.8374E-03, \
1.7932E-03,8.2628E-04,6.3370E-04,1.0083E-03,7.6357E-04,6.2533E-04, \
6.8493E-03,2.3569E-03,1.9961E-03,1.3048E-03,1.2944E-03,1.4024E-03, \
1.0176E-03,1.0575E-03,1.0487E-03,1.0932E-03,1.0628E-03,9.7799E-04, \
1.8717E-03,9.3959E-04,6.0645E-04,1.3730E-03,9.4748E-04,7.4772E-04, \
7.5178E-04,6.6931E-04,1.5913E-03,4.4780E-04,1.0291E-03,8.7505E-04, \
8.4668E-04,9.4771E-04,1.0289E-03,8.8062E-04,5.5577E-04,7.3385E-04, \
8.0039E-04,9.2923E-04,8.5983E-04,3.5607E-03,4.2070E-03,7.7449E-04, \
4.2390E-04,1.4873E-03,8.0517E-04,7.3306E-04,3.5569E-03,7.9584E-04, \
2.3949E-03,2.5601E-03,2.0745E-03,2.1577E-03,2.6694E-03,1.7471E-03, \
8.7847E-04,9.3258E-04,8.9768E-04,3.7367E-03,3.5090E-03,4.0465E-03, \
8.0198E-04,5.3322E-04,1.1111E-03,1.8056E-03,9.1752E-04,7.9106E-04, \
6.2581E-04,3.6274E-03,9.4636E-04,9.0357E-04,1.8804E-03,2.8996E-03, \
2.4374E-03,7.8565E-04,7.7043E-04,2.6191E-03,2.5502E-03,3.1984E-03, \
1.3665E-03,9.3465E-04,3.9023E-03,2.9212E-03,2.5099E-03,3.2570E-03, \
2.2305E-03,2.3894E-03,2.2476E-03,2.4350E-03,2.5264E-03,2.0922E-03, \
2.7968E-03,4.4841E-03,7.6748E-04,2.4057E-03,2.2320E-03,2.3751E-03, \
5.6294E-04,3.2685E-04,8.7345E-04,1.9426E-03,3.9099E-03,8.1545E-04, \
5.4167E-04,1.7259E-03,2.2966E-03,5.5075E-04,1.2291E-03,1.7259E-03, \
1.4697E-03,3.5815E-03,6.7577E-04,2.3617E-03,2.5177E-03,2.8398E-03, \
1.1183E-03,6.1226E-04,2.6402E-03,2.4008E-03,2.2231E-03,2.4927E-03, \
2.2031E-03,2.3292E-03,2.3979E-03,2.6581E-03,2.4612E-03,2.5616E-03, \
1.5060E-03,2.7802E-03,6.7154E-04,1.8588E-03,2.0822E-03,2.2212E-03 \
]
return coef
# end of set_coeffs

# set pairwise Gaussian decay parameters
def set_alpha():
    alpha = [
        -6.1500E-02,-1.6700E-02,-2.5900E-02,-1.9300E-02,-1.6100E-02,-1.5500E-02, \
        -2.5900E-02,-2.6100E-02,-2.5000E-02,-3.8300E-02,-2.4000E-02,-2.2400E-02, \
        -2.9800E-02,-3.2500E-02,-3.9400E-02,-3.5400E-02,-3.9000E-02,-4.3600E-02, \
        -1.9000E-02,-3.0000E-02,-2.7800E-02,-3.0000E-02,-3.1700E-02,-3.5300E-02, \
        -3.3200E-02,-3.4800E-02,-3.6200E-02,-3.7900E-02,-3.7800E-02,-3.9700E-02, \
        -3.1400E-02,-3.5100E-02,-3.8100E-02,-3.6500E-02,-3.9500E-02,-4.1100E-02, \
        -1.0050E-01,-1.7300E-01,-2.3100E-02,-2.5200E-02,-6.5300E-02,-7.4300E-02, \
        -6.0100E-02,-8.1800E-02,-9.2600E-02,-9.6800E-02,-2.1300E-02,-3.1100E-02, \
        -5.9200E-02,-7.1100E-02,-7.6500E-02,-7.0300E-02,-7.9200E-02,-8.3700E-02, \
        -9.1000E-03,-3.5400E-02,-4.0700E-02,-4.8900E-02,-4.7200E-02,-4.3700E-02, \
        -5.2900E-02,-5.0800E-02,-5.0900E-02,-5.1000E-02,-5.0200E-02,-5.4700E-02, \
        -5.9400E-02,-6.5300E-02,-7.2500E-02,-6.2000E-02,-6.7900E-02,-7.3500E-02, \
        -2.8100E-02,-3.9300E-02,-9.3000E-03,-1.0700E-02,-8.0000E-04,-2.1700E-02, \
        -2.2100E-02,-2.2000E-02,-2.5300E-02,-2.5500E-02,-8.9000E-03,-1.0400E-02, \
        -1.0900E-02,-8.1000E-03,-9.0000E-04,-1.6400E-02,-1.8000E-02,-1.8800E-02, \
        -1.0400E-02,-1.0200E-02,-1.0000E-02,-1.0400E-02,-1.0400E-02,-1.1100E-02, \

```


-1.0600E-02, -1.1200E-02, -1.1600E-02, -1.2500E-02, -1.1700E-02, -1.3200E-02, \\
-8.4000E-03, -1.1900E-02, -7.1000E-03, -1.4600E-02, -1.6000E-02, -1.7200E-02, \\
-4.2700E-02, -2.2100E-02, -2.0500E-02, -1.3700E-02, -2.4000E-03, -1.9700E-02, \\
-1.5200E-02, -4.8000E-03, -3.4500E-02, -3.6900E-02, -1.4700E-02, -1.2700E-02, \\
-6.0000E-03, -3.3000E-03, -8.0000E-04, -1.6700E-02, -1.5000E-02, -9.5000E-03, \\
-1.4400E-02, -2.5400E-02, -1.7700E-02, -1.8000E-02, -1.7700E-02, -1.6900E-02, \\
-1.7400E-02, -1.7000E-02, -1.6800E-02, -1.5800E-02, -1.6700E-02, -1.7100E-02, \\
-7.2000E-03, -4.5000E-03, -2.9000E-03, -1.6600E-02, -1.7000E-02, -1.4900E-02, \\
-2.2800E-02, -4.3300E-02, -2.1600E-02, -1.8500E-02, -2.7200E-02, -3.4300E-02, \\
-3.8800E-02, -2.4000E-02, -7.5000E-03, -8.3000E-03, -1.5200E-02, -2.0800E-02, \\
-2.6800E-02, -3.2300E-02, -3.5500E-02, -3.8300E-02, -4.0400E-02, -4.0800E-02, \\
-5.7000E-03, -2.3100E-02, -2.6400E-02, -3.3300E-02, -3.4000E-02, -3.3300E-02, \\
-3.6900E-02, -3.6600E-02, -3.6900E-02, -3.6300E-02, -3.6600E-02, -3.7200E-02, \\
-2.7200E-02, -3.1600E-02, -3.4100E-02, -3.6100E-02, -3.8100E-02, -3.8400E-02, \\
-1.5500E-02, -4.1900E-02, -1.5500E-02, -1.3100E-02, -2.5100E-02, -4.9800E-02, \\
-6.4200E-02, -2.6400E-02, -2.9300E-02, -4.7800E-02, -1.1300E-02, -1.6800E-02, \\
-3.4000E-02, -4.4000E-02, -5.5200E-02, -3.5300E-02, -4.9100E-02, -5.9900E-02, \\
-2.4000E-03, -2.0800E-02, -2.3700E-02, -3.2400E-02, -3.3700E-02, -3.1200E-02, \\
-3.8200E-02, -3.9800E-02, -4.0600E-02, -4.0100E-02, -4.0000E-02, -4.4600E-02, \\
-3.5000E-02, -4.5600E-02, -5.3100E-02, -3.3800E-02, -4.3600E-02, -5.0500E-02, \\
-5.8200E-02, -7.7100E-02, -1.0800E-02, -1.7800E-02, -4.9900E-02, -4.8300E-02, \\
-7.7600E-02, -6.3800E-02, -6.7500E-02, -6.1200E-02, -7.6000E-03, -1.7700E-02, \\
-3.0100E-02, -3.5000E-02, -5.4600E-02, -3.6100E-02, -3.8300E-02, -4.5800E-02, \\
-2.0500E-02, -1.7900E-02, -1.5800E-02, -2.5400E-02, -2.7200E-02, -2.4300E-02, \\
-3.2400E-02, -3.3900E-02, -3.5000E-02, -3.2100E-02, -3.3800E-02, -3.7900E-02, \\
-2.8700E-02, -3.3200E-02, -4.6800E-02, -2.9200E-02, -3.1700E-02, -3.5900E-02, \\
-5.1300E-02, -7.9600E-02, -1.6100E-02, -1.1600E-02, -2.4800E-02, -8.6000E-03, \\
-9.4000E-03, -1.8800E-02, -1.7800E-02, -1.6300E-02, -1.5900E-02, -1.8000E-02, \\
-2.2000E-02, -2.1000E-02, -1.0100E-02, -3.3800E-02, -3.0700E-02, -2.6900E-02, \\
-1.1600E-02, -2.2000E-02, -1.4100E-02, -1.2100E-02, -1.1900E-02, -1.3400E-02, \\
-1.0400E-02, -9.6000E-03, -9.0000E-03, -1.4000E-02, -9.6000E-03, -1.2300E-02, \\
-2.0900E-02, -1.9800E-02, -1.1500E-02, -2.9300E-02, -2.6600E-02, -2.6200E-02, \\
-1.3500E-02, -8.2200E-02, -1.5100E-02, -1.2400E-02, -3.9500E-02, -2.4600E-02, \\
-1.3500E-02, -2.0300E-02, -1.5000E-02, -3.2100E-02, -1.1300E-02, -1.3400E-02, \\
-2.4400E-02, -2.7800E-02, -2.1100E-02, -4.3800E-02, -4.1400E-02, -3.3500E-02, \\
-7.8000E-03, -1.5900E-02, -9.5000E-03, -7.2000E-03, -7.3000E-03, -1.2000E-02, \\
-5.7000E-03, -5.1000E-03, -7.1000E-03, -1.6800E-02, -1.1100E-02, -1.5200E-02, \\
-2.5900E-02, -2.6100E-02, -2.0500E-02, -3.5900E-02, -3.4300E-02, -3.1100E-02, \\
-6.0300E-02, -7.1600E-02, -1.1800E-02, -6.4000E-03, -3.0300E-02, -3.8100E-02, \\
-2.2300E-02, -3.5600E-02, -1.4200E-02, -2.8600E-02, -7.6000E-03, -7.5000E-03, \\
-2.1600E-02, -2.2800E-02, -2.0800E-02, -3.8200E-02, -3.9500E-02, -3.0700E-02, \\
-3.1000E-03, -1.0500E-02, -5.3000E-03, -3.4000E-03, -4.0000E-03, -8.0000E-03, \\
-2.2000E-03, -2.3000E-03, -4.0000E-03, -1.2700E-02, -7.5000E-03, -1.1000E-02, \\
-2.3400E-02, -1.8200E-02, -1.7000E-02, -2.8800E-02, -2.6700E-02, -2.7500E-02, \\
-2.3900E-02, -2.8000E-02, -1.1600E-02, -9.6000E-03, -1.8900E-02, -2.1800E-02, \\
-2.2700E-02, -2.1800E-02, -2.3400E-02, -2.4200E-02, -1.0400E-02, -1.1100E-02, \\
-1.6700E-02, -1.8600E-02, -1.0000E-03, -1.5900E-02, -1.7600E-02, -1.9300E-02, \\
-1.0300E-02, -1.1400E-02, -1.1100E-02, -1.0900E-02, -1.0500E-02, -1.1000E-02, \\
-1.0200E-02, -1.0200E-02, -1.0400E-02, -1.0900E-02, -1.0600E-02, -1.1300E-02, \\
-1.6600E-02, -1.8000E-02, -2.0000E-03, -1.4200E-02, -1.5600E-02, -1.7100E-02, \\
-3.5700E-02, -3.3400E-02, -9.7000E-03, -3.4000E-03, -2.6800E-02, -2.7200E-02, \\
-2.6400E-02, -2.9300E-02, -3.0100E-02, -2.9600E-02, -5.0000E-03, -1.4000E-03, \\
-2.2900E-02, -2.6200E-02, -2.6500E-02, -2.1700E-02, -2.4000E-02, -2.6100E-02, \\
-8.4000E-03, -1.2200E-02, -8.5000E-03, -1.4100E-02, -1.6400E-02, -1.4200E-02, \\
-2.1800E-02, -2.3300E-02, -2.4100E-02, -2.0700E-02, -2.2000E-02, -2.2400E-02, \\
-2.2500E-02, -2.5500E-02, -2.6400E-02, -1.9600E-02, -2.1200E-02, -2.3200E-02, \\
-3.3400E-02, -1.7900E-02, -1.6900E-02, -1.4000E-02, -2.5700E-02, -3.1900E-02, \\
-3.6200E-02, -2.5600E-02, -2.7300E-02, -3.0600E-02, -1.3000E-02, -1.4500E-02, \\
-1.7200E-02, -3.1300E-02, -3.7400E-02, -2.4900E-02, -3.1500E-02, -3.6000E-02, \\
-7.9000E-03, -1.8000E-02, -1.7200E-02, -2.5300E-02, -2.6700E-02, -2.2200E-02, \\
-3.0600E-02, -3.1300E-02, -3.1800E-02, -2.9800E-02, -3.0300E-02, -3.3200E-02, \\
-1.4900E-02, -2.8500E-02, -3.5900E-02, -2.1900E-02, -2.9000E-02, -3.2500E-02, \\
-5.9200E-02, -4.4500E-02, -1.1100E-02, -1.3300E-02, -3.3900E-02, -4.4800E-02, \

```

-4.6900E-02, -3.2100E-02, -3.8900E-02, -4.4000E-02, -1.0900E-02, -1.4600E-02, \
-2.0900E-02, -3.8800E-02, -4.7400E-02, -3.0800E-02, -4.0800E-02, -4.7300E-02, \
-4.3000E-03, -1.3200E-02, -1.7400E-02, -2.5100E-02, -2.6400E-02, -2.4400E-02, \
-3.0000E-02, -3.0900E-02, -3.2100E-02, -3.1100E-02, -3.1400E-02, -3.5600E-02, \
-2.1000E-02, -3.4900E-02, -4.5600E-02, -2.4600E-02, -3.2700E-02, -3.9600E-02, \
-5.6800E-02, -7.0000E-02, -1.4500E-02, -1.9300E-02, -3.1100E-02, -6.0600E-02, \
-6.1900E-02, -3.7800E-02, -4.0200E-02, -5.7000E-02, -1.1900E-02, -1.9300E-02, \
-2.8900E-02, -4.0500E-02, -5.8900E-02, -3.1800E-02, -4.0500E-02, -5.5200E-02, \
-3.5000E-03, -2.1100E-02, -1.8600E-02, -2.3700E-02, -2.4700E-02, -2.4100E-02, \
-2.8000E-02, -2.9200E-02, -3.0100E-02, -2.9200E-02, -2.9600E-02, -3.3600E-02, \
-2.9900E-02, -3.5000E-02, -5.2600E-02, -3.0500E-02, -3.2600E-02, -3.9400E-02, \
-4.4200E-02, -1.2600E-02, -1.2800E-02, -4.1000E-03, -1.0000E-03, -3.2000E-03, \
-9.5000E-03, -9.0000E-03, -5.4000E-03, -2.2000E-03, -8.8000E-03, -5.3000E-03, \
-3.8000E-03, -2.3000E-03, -2.8400E-02, -8.8000E-03, -7.6000E-03, -5.0000E-03, \
-1.1400E-02, -2.0800E-02, -6.3000E-03, -1.0200E-02, -1.1100E-02, -8.0000E-03, \
-1.2400E-02, -1.1500E-02, -1.2100E-02, -9.7000E-03, -1.1000E-02, -1.0300E-02, \
-2.2000E-03, -1.8000E-03, -1.2000E-03, -8.4000E-03, -1.0100E-02, -8.1000E-03, \
-4.5500E-02, -6.0000E-04, -1.5400E-02, -5.5000E-03, -5.0000E-03, -1.0000E-03, \
-5.0000E-04, -1.0900E-02, -8.9000E-03, -6.4000E-03, -1.3300E-02, -1.2400E-02, \
-1.2700E-02, -4.3000E-03, -2.9100E-02, -1.4200E-02, -1.3100E-02, -9.6000E-03, \
-1.1400E-02, -2.5000E-02, -1.1600E-02, -1.3100E-02, -1.3500E-02, -1.2400E-02, \
-1.3600E-02, -1.2900E-02, -1.2400E-02, -1.0400E-02, -1.2200E-02, -9.3000E-03, \
-1.3100E-02, -6.8000E-03, -1.5000E-03, -1.6400E-02, -1.5500E-02, -1.4400E-02, \
-3.9800E-02, -4.9700E-02, -1.4600E-02, -6.3000E-03, -1.2800E-02, -5.8000E-03, \
-2.3000E-03, -1.8400E-02, -1.7800E-02, -1.3800E-02, -1.4000E-02, -1.4600E-02, \
-1.8300E-02, -1.2100E-02, -3.6000E-03, -1.8900E-02, -1.6200E-02, -1.2400E-02, \
-6.8000E-03, -2.1900E-02, -1.3900E-02, -1.4900E-02, -1.4900E-02, -1.4500E-02, \
-1.4600E-02, -1.3800E-02, -1.3400E-02, -1.1800E-02, -1.3200E-02, -1.1400E-02, \
-1.9200E-02, -1.5600E-02, -9.7000E-03, -2.1400E-02, -1.9700E-02, -1.7800E-02, \
-2.0500E-02, -1.9500E-02, -9.0000E-03, -1.1700E-02, -1.5800E-02, -1.7500E-02, \
-1.8100E-02, -1.8500E-02, -1.9900E-02, -2.0500E-02, -9.6000E-03, -1.2400E-02, \
-1.2800E-02, -1.4700E-02, -1.5800E-02, -1.4700E-02, -1.5700E-02, -1.6400E-02, \
-8.7000E-03, -1.0200E-02, -9.0000E-03, -9.3000E-03, -9.7000E-03, -1.0500E-02, \
-9.9000E-03, -1.0500E-02, -1.0900E-02, -1.1400E-02, -1.0700E-02, -1.1700E-02, \
-4.0000E-04, -1.4200E-02, -1.5400E-02, -1.4000E-02, -1.4800E-02, -1.5600E-02, \
-2.2500E-02, -2.1300E-02, -1.0000E-02, -1.2300E-02, -2.7000E-03, -2.2100E-02, \
-2.2600E-02, -2.3500E-02, -2.4300E-02, -2.4400E-02, -1.0500E-02, -1.3800E-02, \
-1.9400E-02, -2.0400E-02, -1.7000E-03, -1.9000E-02, -2.0600E-02, -2.2000E-02, \
-1.0400E-02, -1.1500E-02, -1.0600E-02, -1.1100E-02, -1.0900E-02, -1.1800E-02, \
-1.1000E-02, -1.1500E-02, -1.2000E-02, -1.3000E-02, -1.3000E-02, -1.2100E-02, -1.3900E-02, \
-1.9300E-02, -2.0300E-02, -1.9000E-03, -1.7500E-02, -1.8700E-02, -2.0000E-02, \
-4.4200E-02, -3.4200E-02, -8.9000E-03, -2.1200E-02, -4.8400E-02, -5.2100E-02, \
-3.7600E-02, -6.2300E-02, -5.8600E-02, -4.7600E-02, -6.6000E-03, -2.8300E-02, \
-3.8200E-02, -4.8300E-02, -5.1400E-02, -5.7400E-02, -6.2200E-02, -6.3000E-02, \
-1.9400E-02, -2.4500E-02, -1.3900E-02, -2.5900E-02, -2.7500E-02, -2.8100E-02, \
-3.2900E-02, -3.5300E-02, -3.7600E-02, -3.9600E-02, -3.7600E-02, -4.4100E-02, \
-3.5800E-02, -4.5800E-02, -5.1700E-02, -5.0300E-02, -5.6400E-02, -5.9500E-02, \
-6.4200E-02, -4.3900E-02, -2.0000E-03, -1.9200E-02, -3.9000E-02, -5.3200E-02, \
-3.7400E-02, -5.6800E-02, -5.9400E-02, -5.0100E-02, -2.2100E-02, -1.9800E-02, \
-3.4700E-02, -4.4100E-02, -4.9700E-02, -4.7200E-02, -5.3300E-02, -5.6300E-02, \
-2.2200E-02, -1.6100E-02, -5.0000E-03, -1.8200E-02, -2.0100E-02, -2.2400E-02, \
-2.6700E-02, -2.9500E-02, -3.2200E-02, -3.4500E-02, -3.3700E-02, -3.9700E-02, \
-3.3500E-02, -3.9900E-02, -4.7900E-02, -4.2800E-02, -4.8200E-02, -5.2900E-02, \
-2.7100E-02, -7.2000E-03, -8.0000E-04, -1.4000E-02, -1.8000E-02, -2.3000E-02, \
-1.7400E-02, -1.9600E-02, -1.7700E-02, -1.5400E-02, -2.4300E-02, -5.6000E-03, \
-2.2700E-02, -1.7800E-02, -1.9100E-02, -2.5600E-02, -2.3600E-02, -2.1000E-02, \
-2.3700E-02, -7.3000E-03, -2.5200E-02, -1.2300E-02, -1.4500E-02, -1.5000E-02, \
-2.0200E-02, -2.2100E-02, -2.4000E-02, -2.5000E-02, -2.3700E-02, -2.8600E-02, \
-2.4500E-02, -2.0200E-02, -1.8900E-02, -2.7300E-02, -2.5600E-02, -2.3500E-02, \
-4.6700E-02, -7.9900E-02, -1.0400E-02, -1.8800E-02, -2.4700E-02, -2.7900E-02, \
-3.8900E-02, -2.4100E-02, -2.4800E-02, -2.8000E-02, -3.0000E-04, -1.1900E-02, \
-2.9400E-02, -2.3900E-02, -2.4700E-02, -3.1200E-02, -2.8100E-02, -2.5000E-02, \
-2.2700E-02, -1.3500E-02, -1.0400E-02, -2.0100E-02, -2.1700E-02, -2.1900E-02, \

```

-2.4600E-02, -2.5800E-02, -2.6600E-02, -2.7300E-02, -2.6300E-02, -2.8500E-02, \\
-3.1500E-02, -2.6500E-02, -2.3600E-02, -3.2300E-02, -3.1000E-02, -2.8400E-02, \\
-1.0700E-02, -5.4300E-02, -2.3500E-02, -1.6000E-03, -7.8000E-03, -1.5000E-03, \\
-4.6200E-02, -9.1000E-03, -4.7000E-03, -4.6000E-02, -2.3800E-02, -2.7100E-02, \\
-1.3900E-02, -7.0000E-03, -2.6000E-03, -1.4000E-02, -1.2300E-02, -9.2000E-03, \\
-2.1700E-02, -2.5700E-02, -2.4900E-02, -2.6000E-02, -2.0000E-03, -2.7000E-02, \\
-9.7000E-03, -1.0900E-02, -1.3800E-02, -1.3800E-02, -1.2100E-02, -1.8400E-02, \\
-1.5700E-02, -1.0300E-02, -6.5000E-03, -1.5800E-02, -1.3900E-02, -1.2300E-02, \\
-5.3400E-02, -7.1400E-02, -2.8300E-02, -1.3900E-02, -1.0400E-02, -4.3500E-02, \\
-4.9800E-02, -4.0200E-02, -4.3700E-02, -4.7800E-02, -2.7400E-02, -3.3000E-03, \\
-2.1300E-02, -4.5000E-03, -4.2900E-02, -2.0400E-02, -1.1000E-02, -4.0200E-02, \\
-2.6100E-02, -5.0000E-03, -3.0200E-02, -5.9000E-03, -6.5000E-03, -6.6000E-03, \\
-1.1100E-02, -1.2200E-02, -1.2200E-02, -1.3300E-02, -1.2200E-02, -1.3800E-02, \\
-2.4000E-02, -1.1800E-02, -3.9700E-02, -2.4400E-02, -1.9700E-02, -1.2200E-02, \\
-5.2200E-02, -7.1500E-02, -2.8200E-02, -9.9000E-03, -3.2900E-02, -3.9600E-02, \\
-4.6100E-02, -3.8000E-02, -3.9000E-02, -4.5400E-02, -2.6900E-02, -2.8400E-02, \\
-1.3900E-02, -3.3200E-02, -4.1300E-02, -1.3000E-02, -1.7000E-03, -3.7400E-02, \\
-2.5700E-02, -2.6900E-02, -2.9900E-02, -2.9800E-02, -1.5000E-03, -3.0200E-02, \\
-6.8000E-03, -2.8300E-02, -1.0200E-02, -4.5000E-03, -8.6000E-03, -9.8000E-03, \\
-1.2500E-02, -5.3000E-03, -3.8100E-02, -1.7200E-02, -1.0000E-02, -5.3000E-03, \\
-1.6200E-02, -7.8100E-02, -2.9400E-02, -1.0300E-02, -7.0000E-04, -4.1200E-02, \\
-4.8300E-02, -4.0400E-02, -4.5200E-02, -4.8900E-02, -2.8200E-02, -2.9600E-02, \\
-1.3900E-02, -3.6400E-02, -4.1000E-02, -7.6000E-03, -3.6200E-02, -3.9000E-02, \\
-2.6300E-02, -3.0000E-02, -3.1200E-02, -3.1600E-02, -2.0000E-04, -3.1700E-02, \\
-4.1000E-03, -4.7000E-03, -5.5000E-03, -6.0000E-03, -5.3000E-03, -7.1000E-03, \\
-1.6800E-02, -2.7000E-03, -3.9000E-02, -1.3400E-02, -7.3000E-03, -4.0000E-04, \\
-7.3000E-03, -7.7200E-02, -4.2000E-03, -1.4800E-02, -8.0000E-03, -4.1900E-02, \\
-4.9100E-02, -3.8800E-02, -4.3300E-02, -4.8100E-02, -2.6500E-02, -1.9000E-03, \\
-1.9500E-02, -2.9000E-03, -4.1300E-02, -1.5700E-02, -6.3000E-03, -3.9000E-02, \\
-2.4500E-02, -3.8000E-03, -3.0000E-02, -7.2000E-03, -8.6000E-03, -7.8000E-03, \\
-1.1600E-02, -1.2400E-02, -1.3000E-02, -1.3400E-02, -1.2800E-02, -1.3800E-02, \\
-2.3200E-02, -1.0100E-02, -3.8500E-02, -2.0900E-02, -1.5100E-02, -7.6000E-03, \\
-6.4000E-03, -8.4600E-02, -2.9900E-02, -1.3000E-03, -3.7900E-02, -4.5900E-02, \\
-5.1500E-02, -4.4000E-02, -4.9200E-02, -5.4300E-02, -2.9600E-02, -3.1800E-02, \\
-6.4000E-03, -3.9900E-02, -4.5500E-02, -3.6000E-02, -3.9300E-02, -4.3500E-02, \\
-2.7500E-02, -3.1600E-02, -3.1300E-02, -3.2400E-02, -3.2200E-02, -3.2800E-02, \\
-3.2800E-02, -3.3300E-02, -1.1000E-03, -3.4500E-02, -3.3900E-02, -5.0000E-03, \\
-7.9000E-03, -3.7900E-02, -4.2900E-02, -4.6000E-03, -3.6300E-02, -3.9000E-02, \\
-1.4800E-02, -3.2700E-02, -1.7500E-02, -1.5600E-02, -2.4200E-02, -2.6800E-02, \\
-2.6000E-02, -2.0300E-02, -1.3100E-02, -1.6300E-02, -1.3400E-02, -1.5600E-02, \\
-1.9700E-02, -3.0200E-02, -3.2100E-02, -2.4800E-02, -2.9900E-02, -3.3500E-02, \\
-6.9000E-03, -1.9600E-02, -1.9900E-02, -2.5400E-02, -2.6500E-02, -2.3600E-02, \\
-2.9000E-02, -2.9600E-02, -2.9900E-02, -2.8400E-02, -2.9200E-02, -3.1200E-02, \\
-1.7800E-02, -2.7600E-02, -3.1900E-02, -2.3900E-02, -2.7400E-02, -3.1500E-02, \\
-4.6300E-02, -1.8700E-02, -1.4500E-02, -1.4800E-02, -3.2400E-02, -3.3700E-02, \\
-3.0800E-02, -3.2600E-02, -3.0200E-02, -3.0300E-02, -1.1700E-02, -1.5100E-02, \\
-2.2600E-02, -3.4700E-02, -3.6300E-02, -3.2900E-02, -3.6100E-02, -3.9400E-02, \\
-4.3000E-03, -1.7000E-02, -2.0400E-02, -2.7000E-02, -2.7900E-02, -2.6700E-02, \\
-3.0500E-02, -3.0600E-02, -3.1800E-02, -3.1300E-02, -3.1000E-02, -3.3200E-02, \\
-2.2100E-02, -3.3000E-02, -3.7300E-02, -2.8700E-02, -3.3700E-02, -3.6700E-02, \\
-5.9000E-02, -3.7900E-02, -1.2200E-02, -1.3500E-02, -3.6700E-02, -4.0900E-02, \\
-3.4100E-02, -4.3600E-02, -4.3800E-02, -4.4700E-02, -1.0300E-02, -1.4400E-02, \\
-2.7100E-02, -3.9100E-02, -4.1500E-02, -3.7300E-02, -4.3900E-02, -4.7900E-02, \\
-2.2000E-03, -1.7200E-02, -1.9500E-02, -2.6700E-02, -2.7300E-02, -2.6500E-02, \\
-3.1300E-02, -3.1700E-02, -3.2500E-02, -3.2700E-02, -3.2300E-02, -3.5600E-02, \\
-2.6600E-02, -3.7700E-02, -4.3200E-02, -3.0500E-02, -3.7600E-02, -4.2600E-02, \\
-3.6300E-02, -4.2600E-02, -2.2000E-03, -1.5300E-02, -2.3800E-02, -2.7100E-02, \\
-2.9600E-02, -2.7300E-02, -3.0800E-02, -3.0100E-02, -1.4100E-02, -1.6000E-02, \\
-1.9400E-02, -2.2900E-02, -2.4800E-02, -1.7000E-03, 0.0000E+00, -2.3300E-02, \\
-1.1300E-02, -5.9000E-03, -1.7100E-02, -1.7900E-02, -5.0000E-04, -1.8600E-02, \\
-3.7000E-03, -2.7000E-03, -4.1000E-03, -3.4000E-03, -2.3000E-03, -4.3000E-03, \\
-1.9500E-02, -2.2000E-02, -2.3500E-02, -1.0000E-03, -1.6000E-03, -5.0000E-04, \\
-3.6900E-02, -3.9800E-02, -7.1000E-03, -7.0000E-04, -2.4300E-02, -2.7400E-02, \

```

-2.8400E-02,-4.0000E-04,-3.0900E-02,-3.0600E-02,-2.6000E-03,-1.2000E-03, \
-4.0000E-03,-2.3200E-02,-2.5000E-02,-7.1000E-03,-5.8000E-03,-2.4000E-03, \
-2.5000E-03,-1.2100E-02,-1.6000E-03,-4.3000E-03,-5.3000E-03,-3.4000E-03, \
-5.9000E-03,-5.4000E-03,-5.5000E-03,-4.4000E-03,-5.0000E-03,-5.1000E-03, \
-3.3000E-03,-5.0000E-04,-2.3800E-02,-8.1000E-03,-8.3000E-03,-6.6000E-03, \
-3.6700E-02,-3.8700E-02,-9.9000E-03,-1.8000E-03,-3.0000E-04,-2.8800E-02, \
-2.9400E-02,-8.7000E-03,-3.7000E-03,-3.0100E-02,-6.8000E-03,-6.6000E-03, \
-9.4000E-03,-1.6000E-03,-2.4900E-02,-9.9000E-03,-7.3000E-03,-3.6000E-03, \
-4.8000E-03,-1.7200E-02,-5.9000E-03,-6.8000E-03,-7.0000E-03,-6.6000E-03, \
-6.8000E-03,-6.3000E-03,-6.0000E-03,-4.8000E-03,-5.9000E-03,-4.6000E-03, \
-1.0300E-02,-4.5000E-03,-2.3800E-02,-1.2500E-02,-1.0700E-02,-8.7000E-03 \
]
return alpha
# end of set_alpha

# set pairwise exponential decay parameters
def set_beta():
    beta = [
4.1330E-01,-1.9980E-01,1.4000E-01,1.7100E-02,-1.0940E-01,-1.8640E-01, \
-1.1330E-01,-2.8500E-02,-2.9000E-02,1.5930E-01,1.4930E-01,9.5700E-02, \
1.1800E-01,8.0400E-02,1.0320E-01,1.5450E-01,1.5190E-01,1.6850E-01, \
1.3140E-01,2.3260E-01,1.5240E-01,1.4560E-01,1.6240E-01,2.0940E-01, \
1.6420E-01,1.7640E-01,1.8730E-01,1.9630E-01,2.0830E-01,2.0050E-01, \
1.4400E-01,1.3180E-01,1.2700E-01,2.1410E-01,2.1570E-01,2.0140E-01, \
3.2180E-01,9.9440E-01,1.5900E-02,6.1000E-03,3.2030E-01,2.6780E-01, \
4.0300E-02,3.5650E-01,3.7840E-01,3.4670E-01,9.8000E-03,1.2050E-01, \
3.7880E-01,4.0890E-01,3.6950E-01,4.6010E-01,4.8580E-01,4.8560E-01, \
-1.4340E-01,1.9900E-01,2.3980E-01,3.0260E-01,2.6980E-01,2.2090E-01, \
3.1990E-01,2.7700E-01,2.6610E-01,2.4980E-01,2.5680E-01,2.6680E-01, \
3.9450E-01,3.8890E-01,3.9200E-01,4.3620E-01,4.6230E-01,4.7450E-01, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,-3.6470E-01,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
-8.9000E-02,-2.0200E-01,-3.9110E-01,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
-1.4550E-01,-1.0540E-01,-2.4370E-01,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,-3.1570E-01,5.0700E-02,-7.3300E-02,-4.5590E-01,-1.8430E-01, \
-2.9510E-01,-3.9770E-01,0.0000E+00,0.0000E+00,-3.1700E-02,-9.4000E-02, \
-3.2120E-01,-4.4430E-01,-5.1240E-01,-1.1160E-01,-1.4850E-01,-2.7700E-01, \
4.2100E-02,1.7880E-01,-4.3800E-02,-6.7800E-02,-7.8500E-02,-9.9200E-02, \
-9.4900E-02,-1.0590E-01,-1.1000E-01,-1.3160E-01,-1.1910E-01,-9.3200E-02, \
-2.8770E-01,-4.0430E-01,-4.6350E-01,-9.6700E-02,-9.3800E-02,-1.4490E-01, \
0.0000E+00,0.0000E+00,1.4450E-01,8.5900E-02,8.2700E-02,9.1400E-02, \
7.4000E-02,-5.9000E-03,-2.5840E-01,-3.0040E-01,5.9300E-02,1.4380E-01, \
1.4450E-01,1.7010E-01,1.7050E-01,3.3470E-01,3.2440E-01,3.0240E-01, \
-9.8600E-02,1.9440E-01,2.1630E-01,3.1650E-01,3.2210E-01,3.0770E-01, \
3.6410E-01,3.4870E-01,3.4650E-01,3.2540E-01,3.3760E-01,3.2430E-01, \
1.6230E-01,1.7670E-01,1.8350E-01,3.4720E-01,3.5550E-01,3.3030E-01, \
-1.8740E-01,-5.4300E-02,-4.8800E-02,-9.8000E-02,-1.6600E-02,2.1730E-01, \
3.0620E-01,-3.0500E-02,-3.8000E-02,1.2050E-01,-1.0120E-01,-2.0500E-02, \
1.7700E-01,2.4080E-01,3.4440E-01,1.7930E-01,3.3020E-01,4.3790E-01, \
-2.5840E-01,5.5500E-02,6.5000E-02,1.8200E-01,2.0020E-01,1.5380E-01, \
2.6090E-01,2.7630E-01,2.8130E-01,2.5630E-01,2.6940E-01,3.1030E-01, \
2.0540E-01,2.9680E-01,3.5960E-01,1.9150E-01,3.0930E-01,3.7340E-01, \
2.5100E-02,1.8900E-01,-2.0740E-01,-1.2360E-01,1.7390E-01,8.1500E-02, \
3.5950E-01,2.2970E-01,2.1460E-01,1.3710E-01,-2.5920E-01,-1.0840E-01, \
2.0900E-02,1.7700E-02,1.9540E-01,8.1300E-02,7.8200E-02,1.2750E-01, \
0.0000E+00,-1.1200E-01,-1.5650E-01,-2.0200E-02,8.0000E-03,-4.2600E-02, \
8.0600E-02,9.5900E-02,1.0490E-01,4.6800E-02,8.2600E-02,1.1410E-01, \
1.6800E-02,1.5200E-02,1.3290E-01,2.2600E-02,2.9800E-02,5.5400E-02, \
0.0000E+00,0.0000E+00,-9.6200E-02,-2.2450E-01,-1.1960E-01,-4.8810E-01, \
-5.6930E-01,-2.6140E-01,-3.0600E-01,-3.8480E-01,-7.9000E-02,-6.3000E-02, \
-9.0700E-02,-1.4970E-01,-3.8680E-01,5.7500E-02,-3.2100E-02,-1.1000E-01, \

```

```

-1.2650E-01,6.7000E-03,-1.8020E-01,-2.4110E-01,-2.4310E-01,-2.1500E-01, \
-2.8470E-01,-3.0490E-01,-3.2320E-01,-2.4640E-01,-3.1140E-01,-2.9080E-01, \
-1.0440E-01,-1.3470E-01,-3.1500E-01,4.2300E-02,-3.1900E-02,-7.3000E-02, \
-3.6180E-01,0.0000E+00,-1.5610E-01,-2.4090E-01,3.1400E-02,-2.9630E-01, \
-5.5730E-01,-3.2080E-01,-3.8320E-01,-2.2110E-01,-2.0550E-01,-1.9800E-01, \
-1.2980E-01,-9.2700E-02,-2.4030E-01,1.5890E-01,7.2500E-02,-8.0300E-02, \
-2.3920E-01,-1.6890E-01,-3.0430E-01,-3.6720E-01,-3.6210E-01,-2.7350E-01, \
-3.9980E-01,-4.1730E-01,-3.8600E-01,-2.3850E-01,-3.1870E-01,-2.8160E-01, \
-9.7200E-02,-1.1770E-01,-2.0650E-01,1.0090E-01,3.8800E-02,-4.7700E-02, \
7.8900E-02,0.0000E+00,-2.1170E-01,-3.2240E-01,-1.4870E-01,-9.9900E-02, \
-4.2300E-01,-1.5510E-01,-4.4570E-01,-3.1610E-01,-2.7740E-01,-3.1950E-01, \
-1.9010E-01,-2.2710E-01,-2.5930E-01,6.0700E-02,2.4600E-02,-1.5140E-01, \
-3.3240E-01,-2.8090E-01,-3.8780E-01,-4.3970E-01,-4.2270E-01,-3.5160E-01, \
-4.6230E-01,-4.6450E-01,-4.3950E-01,-3.0660E-01,-3.8190E-01,-3.5260E-01, \
-1.5050E-01,-2.8790E-01,-3.1270E-01,-3.4100E-02,-1.0200E-01,-1.2810E-01, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,-3.9990E-01,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,-3.6420E-01,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,-1.3570E-01,-2.1950E-01,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,-2.0940E-01,-2.9380E-01, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
-7.8900E-02,-6.7600E-02,-1.9050E-01,-7.9700E-02,-2.8600E-02,-1.2570E-01, \
9.7300E-02,1.3520E-01,1.5760E-01,7.7700E-02,9.8500E-02,1.2130E-01, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
1.9560E-01,-2.5060E-01,1.5220E-01,7.1100E-02,1.5130E-01,1.5950E-01, \
1.7960E-01,8.4600E-02,7.4500E-02,9.9500E-02,9.0100E-02,1.0680E-01, \
6.5800E-02,2.6060E-01,3.1280E-01,1.8460E-01,2.6650E-01,3.2200E-01, \
7.8000E-03,1.9040E-01,1.3650E-01,2.6480E-01,2.8710E-01,1.9540E-01, \
3.4570E-01,3.5030E-01,3.5070E-01,2.9990E-01,3.2000E-01,3.4810E-01, \
2.8200E-02,2.2640E-01,3.1830E-01,1.7260E-01,2.7300E-01,3.0670E-01, \
4.3890E-01,9.3800E-02,-4.2200E-02,-2.5000E-02,2.1620E-01,2.9800E-01, \
2.4440E-01,1.1300E-01,1.7270E-01,2.1830E-01,-3.1300E-02,2.2900E-02, \
5.9800E-02,3.0180E-01,4.0130E-01,2.1550E-01,3.3930E-01,4.1400E-01, \
-1.4840E-01,5.7000E-03,5.6700E-02,1.6990E-01,1.8720E-01,1.5220E-01, \
2.3530E-01,2.4380E-01,2.5450E-01,2.2650E-01,2.4450E-01,2.8750E-01, \
7.0800E-02,2.5260E-01,3.9900E-01,1.5120E-01,2.5750E-01,3.3970E-01, \
2.9070E-01,3.9480E-01,-6.1300E-02,-1.2700E-02,9.7400E-02,4.6440E-01, \
3.9180E-01,1.3530E-01,1.2780E-01,3.1610E-01,-8.7200E-02,2.2100E-02, \
1.1720E-01,2.4540E-01,4.8120E-01,1.5230E-01,2.4730E-01,4.3050E-01, \
-2.3850E-01,6.4100E-02,-2.0000E-03,6.2400E-02,7.5800E-02,6.3800E-02, \
1.1930E-01,1.3100E-01,1.3900E-01,1.1430E-01,1.3010E-01,1.6720E-01, \
1.4200E-01,1.7920E-01,4.1390E-01,1.5810E-01,1.7080E-01,2.4790E-01, \
0.0000E+00,-4.8720E-01,-8.6800E-02,-2.7550E-01,-4.7430E-01,-5.2970E-01, \
-4.6480E-01,-3.5580E-01,-4.4370E-01,-5.5730E-01,-1.4770E-01,-2.4990E-01, \
-3.5740E-01,-4.5450E-01,0.0000E+00,-2.8690E-01,-3.3920E-01,-4.1860E-01, \
-2.6300E-02,9.5100E-02,-2.8320E-01,-2.2250E-01,-2.0480E-01,-2.7580E-01, \
-1.8360E-01,-2.0990E-01,-1.9760E-01,-2.5120E-01,-2.2840E-01,-2.4010E-01, \
-3.8580E-01,-4.5560E-01,-5.0880E-01,-2.6680E-01,-2.5390E-01,-3.1470E-01, \
0.0000E+00,-7.0030E-01,-6.9400E-02,-2.9470E-01,-4.2660E-01,-5.9730E-01, \
-6.5150E-01,-3.7370E-01,-4.4750E-01,-5.2610E-01,-8.6900E-02,-1.3760E-01, \
-1.9250E-01,-4.3100E-01,0.0000E+00,-2.2450E-01,-2.6170E-01,-3.3640E-01, \
-6.5000E-02,1.4110E-01,-2.0600E-01,-1.9620E-01,-1.8980E-01,-2.2060E-01, \
-1.9680E-01,-2.1600E-01,-2.3370E-01,-2.8160E-01,-2.4010E-01,-3.0870E-01, \
-1.8680E-01,-3.6640E-01,-5.1340E-01,-1.5130E-01,-1.8670E-01,-2.1980E-01, \
0.0000E+00,0.0000E+00,-7.9400E-02,-2.7240E-01,-2.5210E-01,-4.7990E-01, \
-5.9960E-01,-2.3800E-01,-2.8200E-01,-3.7450E-01,-6.2900E-02,-8.4700E-02, \
-7.9200E-02,-2.5040E-01,-4.6430E-01,-1.3650E-01,-2.1240E-01,-2.9190E-01, \
-1.6990E-01,6.3000E-02,-1.4610E-01,-1.4910E-01,-1.5050E-01,-1.6520E-01, \
-1.7040E-01,-1.9180E-01,-2.0840E-01,-2.4970E-01,-2.1420E-01,-2.6570E-01, \
-6.6500E-02,-1.6670E-01,-3.1770E-01,-5.5700E-02,-1.0910E-01,-1.6550E-01, \

```

```

0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
-2.8500E-01,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,-3.6690E-01,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,-4.1720E-01,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,-4.0950E-01,0.0000E+00,0.0000E+00,0.0000E+00, \
1.5590E-01,0.0000E+00,-1.7270E-01,4.2100E-02,4.1780E-01,3.5490E-01, \
4.8000E-03,5.7240E-01,4.5490E-01,2.1940E-01,-2.4990E-01,1.7750E-01, \
3.1870E-01,4.1840E-01,3.8800E-01,6.5430E-01,6.8020E-01,6.3740E-01, \
0.0000E+00,1.4630E-01,-9.9900E-02,1.2250E-01,1.5290E-01,1.6320E-01, \
2.4580E-01,2.8420E-01,3.2010E-01,3.4920E-01,3.1850E-01,4.1210E-01, \
2.8620E-01,4.0230E-01,4.4360E-01,5.8300E-01,6.5280E-01,6.6580E-01, \
3.1550E-01,0.0000E+00,-3.7050E-01,-6.2800E-02,1.7110E-01,2.5940E-01, \
-1.0820E-01,3.7840E-01,3.4200E-01,1.3380E-01,0.0000E+00,-6.2500E-02, \
1.7370E-01,2.4540E-01,2.4990E-01,3.8900E-01,4.2830E-01,4.1120E-01, \
0.0000E+00,-1.0830E-01,-3.4840E-01,-1.0090E-01,-6.2800E-02,-1.5700E-02, \
5.3700E-02,9.8400E-02,1.4120E-01,1.7430E-01,1.7200E-01,2.4720E-01, \
1.6860E-01,2.0950E-01,2.7350E-01,3.5990E-01,4.1250E-01,4.4680E-01, \
-3.6550E-01,-8.1020E-01,-4.4910E-01,-2.1390E-01,-2.8640E-01,-3.8500E-01, \
-5.3870E-01,-3.3490E-01,-4.3330E-01,-5.4780E-01,0.0000E+00,-4.0970E-01, \
-1.3700E-01,-3.1530E-01,-3.9130E-01,-9.4400E-02,-1.8600E-01,-3.0900E-01, \
0.0000E+00,-3.6880E-01,0.0000E+00,-2.7840E-01,-2.3120E-01,-2.2700E-01, \
-1.2770E-01,-9.8500E-02,-7.1800E-02,-5.8200E-02,-7.5200E-02,-1.6800E-02, \
-8.5300E-02,-2.4410E-01,-3.4390E-01,-2.3000E-02,-9.6900E-02,-1.7850E-01, \
1.1100E-02,2.7070E-01,-2.5560E-01,-1.0690E-01,-1.2490E-01,-2.0470E-01, \
-1.4510E-01,-2.0440E-01,-2.5490E-01,-2.9200E-01,-4.8180E-01,-2.5360E-01, \
1.9100E-02,-1.5720E-01,-2.4290E-01,4.2000E-02,-6.0600E-02,-1.8740E-01, \
0.0000E+00,-2.2950E-01,-2.9660E-01,-1.1990E-01,-8.5300E-02,-8.4900E-02, \
-3.7000E-02,-1.9500E-02,-9.7000E-03,-2.0000E-03,-1.5300E-02,1.8000E-03, \
7.4600E-02,-8.5100E-02,-2.1500E-01,9.7000E-02,3.7600E-02,-4.6400E-02, \
-6.5630E-01,0.0000E+00,0.0000E+00,-4.4240E-01,-5.2830E-01,-7.6310E-01, \
0.0000E+00,-5.6810E-01,-6.9510E-01,0.0000E+00,0.0000E+00,0.0000E+00, \
-3.4150E-01,-5.6560E-01,-7.3610E-01,-3.5860E-01,-4.4310E-01,-5.6560E-01, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,-4.7980E-01,0.0000E+00, \
-3.3010E-01,-3.1630E-01,-2.6600E-01,-2.7570E-01,-3.0020E-01,-2.1240E-01, \
-2.9150E-01,-4.7760E-01,-6.2690E-01,-2.8600E-01,-3.6170E-01,-4.3440E-01, \
0.0000E+00,0.0000E+00,0.0000E+00,-3.1010E-01,-5.1150E-01,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,-5.5800E-01, \
-2.3520E-01,-6.7480E-01,0.0000E+00,-2.7980E-01,-5.2750E-01,0.0000E+00, \
0.0000E+00,-5.2910E-01,0.0000E+00,-5.2050E-01,-5.0200E-01,-5.0600E-01, \
-4.0050E-01,-3.8630E-01,-3.9650E-01,-3.7340E-01,-3.9410E-01,-3.9220E-01, \
-1.6800E-01,-4.8890E-01,0.0000E+00,-1.6820E-01,-2.9850E-01,-4.9870E-01, \
0.0000E+00,0.0000E+00,0.0000E+00,-3.8570E-01,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
-3.8490E-01,0.0000E+00,0.0000E+00,-4.3550E-01,-7.0360E-01,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,-5.9360E-01,0.0000E+00, \
-4.8280E-01,0.0000E+00,-4.3090E-01,-5.2340E-01,-4.5080E-01,-4.6120E-01, \
-3.8520E-01,-6.2040E-01,0.0000E+00,-3.1900E-01,-4.9460E-01,-6.3240E-01, \
-6.6850E-01,0.0000E+00,0.0000E+00,-4.1060E-01,-7.4430E-01,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
-4.2800E-01,0.0000E+00,0.0000E+00,-5.8510E-01,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,-6.5360E-01,0.0000E+00, \
-5.6260E-01,-5.5430E-01,-5.4130E-01,-5.3600E-01,-5.4440E-01,-5.2980E-01, \
-3.6080E-01,-7.1170E-01,0.0000E+00,-4.4420E-01,-5.9800E-01,-7.6460E-01, \
-7.7070E-01,0.0000E+00,-5.0490E-01,-2.9660E-01,-5.7200E-01,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,-5.9170E-01, \
-2.9430E-01,-7.0540E-01,0.0000E+00,-3.9280E-01,-6.2360E-01,0.0000E+00, \

```

```

0.0000E+00,-5.6800E-01,0.0000E+00,-4.9710E-01,-4.5690E-01,-4.8640E-01, \
-3.9240E-01,-3.8130E-01,-3.7420E-01,-3.6830E-01,-3.7720E-01,-3.8190E-01, \
-2.0830E-01,-5.3580E-01,0.0000E+00,-2.5990E-01,-4.0880E-01,-5.9640E-01, \
-8.2770E-01,0.0000E+00,0.0000E+00,-5.6070E-01,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
-5.7060E-01,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,-6.2600E-01,0.0000E+00,0.0000E+00,-5.8030E-01, \
-5.2540E-01,0.0000E+00,0.0000E+00,-6.0620E-01,0.0000E+00,0.0000E+00, \
-8.8500E-02,0.0000E+00,1.6910E-01,1.0380E-01,1.2870E-01,9.1700E-02, \
2.9900E-02,-8.0000E-04,-1.6190E-01,-1.4680E-01,1.0540E-01,1.3390E-01, \
1.1840E-01,2.5870E-01,2.3640E-01,1.9420E-01,2.5460E-01,3.0180E-01, \
-9.4000E-03,2.2970E-01,1.9930E-01,2.8400E-01,2.9870E-01,2.3380E-01, \
3.3270E-01,3.3450E-01,3.3360E-01,2.8900E-01,3.1600E-01,3.2870E-01, \
8.4100E-02,2.2260E-01,2.6210E-01,2.1760E-01,2.5930E-01,3.1300E-01, \
2.9750E-01,-2.3630E-01,2.4600E-02,6.1000E-03,2.2080E-01,1.5480E-01, \
3.4200E-02,1.4470E-01,5.8400E-02,4.5700E-02,-1.0000E-02,3.8300E-02, \
9.5800E-02,2.6820E-01,2.4690E-01,2.6350E-01,2.8400E-01,3.1940E-01, \
-1.3990E-01,8.5100E-02,1.2300E-01,2.1980E-01,2.3130E-01,2.0560E-01, \
2.6550E-01,2.5690E-01,2.7260E-01,2.5200E-01,2.5630E-01,2.6990E-01, \
9.4200E-02,2.4900E-01,2.9530E-01,2.2770E-01,2.9040E-01,3.1780E-01, \
3.9100E-01,6.8000E-03,-7.8400E-02,-7.9300E-02,2.0970E-01,2.1960E-01, \
2.3000E-02,2.3590E-01,2.0430E-01,1.9890E-01,-9.5300E-02,-3.7700E-02, \
1.0820E-01,2.5810E-01,2.6110E-01,2.6010E-01,3.2740E-01,3.6130E-01, \
-2.4240E-01,2.1600E-02,3.6600E-02,1.4010E-01,1.4390E-01,1.2800E-01, \
2.0260E-01,1.9930E-01,2.0430E-01,1.9690E-01,2.0050E-01,2.2900E-01, \
1.0770E-01,2.4950E-01,3.1250E-01,1.8430E-01,2.7480E-01,3.2750E-01, \
0.0000E+00,0.0000E+00,-2.6920E-01,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00,0.0000E+00, \
0.0000E+00,0.0000E+00,0.0000E+00,-3.8400E-01,-4.4260E-01,0.0000E+00, \
0.0000E+00,-1.9360E-01,0.0000E+00,0.0000E+00,-3.7290E-01,0.0000E+00, \
-3.0940E-01,-3.3220E-01,-3.0030E-01,-3.1980E-01,-3.4440E-01,-2.9900E-01, \
0.0000E+00,0.0000E+00,0.0000E+00,-3.7430E-01,-3.7970E-01,-4.2490E-01, \
0.0000E+00,0.0000E+00,-1.9660E-01,-3.3910E-01,0.0000E+00,0.0000E+00, \
0.0000E+00,-4.8540E-01,0.0000E+00,0.0000E+00,-2.8100E-01,-3.4020E-01, \
-3.2910E-01,0.0000E+00,0.0000E+00,-3.0690E-01,-3.4460E-01,-4.2680E-01, \
-2.3290E-01,-9.2600E-02,-3.6790E-01,-3.2520E-01,-3.0340E-01,-3.4970E-01, \
-2.9580E-01,-3.1230E-01,-3.1000E-01,-3.3630E-01,-3.2710E-01,-3.2390E-01, \
-3.4390E-01,-4.5040E-01,0.0000E+00,-2.6200E-01,-2.7340E-01,-3.1900E-01, \
0.0000E+00,0.0000E+00,-1.5520E-01,-3.4350E-01,-4.7050E-01,0.0000E+00, \
0.0000E+00,-3.3460E-01,-4.4780E-01,0.0000E+00,-2.0370E-01,-2.4090E-01, \
-2.3120E-01,-4.4350E-01,0.0000E+00,-2.7610E-01,-3.3860E-01,-4.1540E-01, \
-1.9870E-01,-6.9000E-03,-2.9650E-01,-2.9370E-01,-2.8940E-01,-3.0340E-01, \
-3.0110E-01,-3.1640E-01,-3.2650E-01,-3.5590E-01,-3.2990E-01,-3.6410E-01, \
-2.1590E-01,-3.7080E-01,0.0000E+00,-1.9870E-01,-2.4940E-01,-2.9970E-01 \
]
return beta
# end of set_beta

if __name__ == "__main__":
    main()

```

References

- (1) Zhou, H.-C.; Kitagawa, S. Metal-Organic Frameworks (MOFs). *Chem. Soc. Rev* **2014**, *43*, 5415–5418.
- (2) Ranocchiari, M.; Bokhoven, J. A. v. Catalysis by metal-organic frameworks: fundamentals and opportunities. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6388–6396.
- (3) Li, B.; Wen, H.-M.; Zhou, W.; Chen, B. Porous Metal-Organic Frameworks for Gas Storage and Separation: What, How, and Why? *The Journal of Physical Chemistry Letters* **2014**, *5*, 3468–3479.
- (4) Schrödinger, E. Quantisierung als Eigenwertproblem. *Annalen der Physik* **1926**, *385*, 437–490.
- (5) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.
- (6) Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *The Journal of Chemical Physics* **1966**, *45*, 4256–4266.
- (7) Crawford, T. D.; Schaefer, H. F., *An Introduction to Coupled Cluster Theory for Computational Chemists*; John Wiley & Sons, Inc.: 2007, pp 33–136.
- (8) Šimová, L.; Řezáč, J.; Hobza, P. Convergence of the Interaction Energies in Noncovalent Complexes in the Coupled-Cluster Methods Up to Full Configuration Interaction. *Journal of Chemical Theory and Computation* **2013**, *9*, 3420–3428.
- (9) Raghavachari, K; Trucks, G.; Pople, J. A.; Head-Gordon, M. A Fifth-Order Perturbation Comparison of Electron Correlation Theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.
- (10) Hohenberg, P; Kohn, W Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.
- (11) Kohn, W; Sham, L. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- (12) Slater, J. C., *The Self-Consistent Field for Molecules and Solids, Quantum Theory of Molecules and Solids*; McGraw-Hill: New York, 1974.

- (13) Perdew, J.; Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B* **1992**, *45*, 244–249.
- (14) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (15) Becke, A. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (16) Lee, C.; Yang, W.; Parr, R. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
- (17) Becke, A. Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals. *J. Chem. Phys.* **1997**, *107*, 8554–8560.
- (18) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.* **2003**, *119*, 12129.
- (19) 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.
- (20) 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.
- (21) Yu, H. S.; He, X.; Truhlar, D. G. MN15-L: A New Local Exchange-Correlation Functional for Kohn-Sham Density Functional Theory with Broad Accuracy for Atoms, Molecules, and Solids. *Journal of Chemical Theory and Computation* **2016**, *12*, 1280–1293.
- (22) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (23) Mardirossian, N.; Head-Gordon, M. ω B97X-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. ω B97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016**, *144*, 214110.
- (25) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other function. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

- (26) Yu, H. S.; He, X.; Li, S. L.; Truhlar, D. G. MN15: A Kohn-Sham global-hybrid exchange-correlation density functional with broad accuracy for multi-reference and single-reference systems and noncovalent interactions. *Chem. Sci.* **2016**, *7*, 5032–5051.
- (27) Perdew, J. P.; Ruzsinszky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E.; Csonka, G. I. Prescription for the design and selection of density functional approximations: more constraint satisfaction with fewer fits. *J. Chem. Phys.* **2005**, *123*, 062201.
- (28) Klimeš, J.; Michaelides, A. Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory. *J. Chem. Phys.* **2012**, *137*, 120901.
- (29) Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* **2004**, *25*, 1463–1473.
- (30) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.
- (31) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
- (32) Becke, A. D.; Johnson, E. R. Exchange-hole dipole moment and the dispersion interaction. *J. Chem. Phys.* **2005**, *122*, 154104.
- (33) Becke, A. D.; Johnson, E. R. A density-functional model of the dispersion interaction. *J. Chem. Phys.* **2005**, *123*, 154101.
- (34) Johnson, E. R.; Becke, A. D. A post-Hartree-Fock model of intermolecular interactions. *J. Chem. Phys.* **2005**, *123*, 24101.
- (35) Johnson, E. R.; DiLabio, G. A. Structure and binding energies in van der Waals dimers: Comparison between density functional theory and correlated ab initio methods. *Chem. Phys. Lett.* **2006**, *419*, 333–339.
- (36) 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.
- (37) 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.
- (38) Lee, K.; Murray, D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. A Higher-Accuracy van der Waals Density Functional. *Phys. Rev. B* **2010**, *82*, 081101.
- (39) Vydrov, O.; Van Voorhis, T. Nonlocal van der Waals Density Functional Made Simple. *Phys. Rev. Lett.* **2009**, *103*, 063004.
- (40) Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals density functional: the simpler the better. *J. Chem. Phys.* **2010**, *133*, 244103.

- (41) Witte, J.; Neaton, J. B.; Head-Gordon, M. Assessing electronic structure approaches for gas-ligand interactions in metal-organic frameworks: the CO₂-benzene complex. *J. Chem. Phys.* **2014**, *140*, 104707.
- (42) Witte, J.; Goldey, M.; Neaton, J. B.; Head-Gordon, M. Beyond Energies: Geometries of Nonbonded Molecular Complexes as Metrics for Assessing Electronic Structure Approaches. *J. Chem. Theory Comput.* **2015**, *11*, 1481–1492.
- (43) Witte, J.; Neaton, J. B.; Head-Gordon, M. Push it to the limit: Characterizing the convergence of common sequences of basis sets for intermolecular interactions as described by density functional theory. *J. Chem. Phys.* **2016**, *144*, 194306.
- (44) Witte, J.; Mardirossian, N.; Neaton, J. B.; Head-Gordon, M. Assessing DFT-D3 damping functions across widely-used density functionals: can we do better?, In-Press.
- (45) Witte, J.; Neaton, J. B.; Head-Gordon, M. Effective empirical corrections for basis set superposition error in the def2-SVPD basis: gCP and DFT-C., Submitted.
- (46) Li, J.-R.; Ma, Y.; McCarthy, M. C.; Sculley, J.; Yu, J.; Jeong, H.-K.; Balbuena, P. B.; Zhou, H.-C. Carbon dioxide capture-related gas adsorption and separation in metal-organic frameworks. *Coord. Chem. Rev.* **2011**, *255*, 1791–1823.
- (47) James, S. L. Metal-organic frameworks. *Chem. Soc. Rev.* **2003**, *32*, 276–288.
- (48) Britt, D.; Furukawa, H.; Wang, B.; Glover, T. G.; Yaghi, O. M. Highly efficient separation of carbon dioxide by a metal-organic framework replete with open metal sites. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 20637–40.
- (49) Caskey, S. R.; Wong-Foy, A. G.; Matzger, A. J. Dramatic tuning of carbon dioxide uptake via metal substitution in a coordination polymer with cylindrical pores. *J. Am. Chem. Soc.* **2008**, *130*, 10870–1.
- (50) Dietzel, P. D. C.; Johnsen, R. E.; Fjellvåg, H.; Bordiga, S.; Groppo, E.; Chavan, S.; Blom, R. Adsorption properties and structure of CO₂ adsorbed on open coordination sites of metal-organic framework Ni₂(dhtp) from gas adsorption, IR spectroscopy and X-ray diffraction. *Chem. Commun.* **2008**, *2*, 5125–7.
- (51) Bloch, E. D.; Britt, D.; Lee, C.; Doonan, C. J.; Uribe-Romo, F. J.; Furukawa, H.; Long, J. R.; Yaghi, O. M. Metal insertion in a microporous metal-organic framework lined with 2,2'-bipyridine. *J. Am. Chem. Soc.* **2010**, *132*, 14382–4.
- (52) Valenzano, L.; Civalleri, B.; Sillar, K.; Sauer, J. Heats of Adsorption of CO and CO₂ in Metal-Organic Frameworks: Quantum Mechanical Study of CPO-27-M (M= Mg, Ni, Zn). *J. Phys. Chem. C* **2011**, *115*, 21777–21784.
- (53) Poloni, R.; Smit, B.; Neaton, J. B. Ligand-assisted enhancement of CO₂ capture in metal-organic frameworks. *J. Am. Chem. Soc.* **2012**, *134*, 6714–6719.
- (54) Poloni, R.; Smit, B.; Neaton, J. B. CO₂ capture by metal-organic frameworks with van der Waals density functionals. *J. Phys. Chem. A* **2012**, *116*, 4957–4964.

- (55) Torrisi, A.; Mellot-Draznieks, C.; Bell, R. G. Impact of ligands on CO₂ adsorption in metal-organic frameworks: first principles study of the interaction of CO₂ with functionalized benzenes. I. Inductive effects on the aromatic ring. *J. Chem. Phys.* **2009**, *130*, 194703.
- (56) Torrisi, A.; Mellot-Draznieks, C.; Bell, R. G. Impact of ligands on CO(2) adsorption in metal-organic frameworks: First principles study of the interaction of CO(2) with functionalized benzenes. II. Effect of polar and acidic substituents. *J. Chem. Phys.* **2010**, *132*, 044705.
- (57) Mackie, I. D.; DiLabio, G. A. CO₂ adsorption by nitrogen-doped carbon nanotubes predicted by density-functional theory with dispersion-correcting potentials. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2780–2787.
- (58) Chen, L.; Cao, F.; Sun, H. Ab initio study of the π - π interactions between CO₂ and benzene, pyridine, and pyrrole. *Int. J. Quantum Chem.* **2013**, *113*, 2261–2266.
- (59) Shen, J.-W.; Domafski, K. B.; Kitao, O.; Nakanishi, K.; Carlo, M. Computer Simulation of Supercritical Carbon Dioxide Fluid: A potential model for the benzene-carbon dioxide system from ab initio calculations. *Fluid Phase Equilib.* **1995**, *104*, 375–390.
- (60) Kristyán, S.; Pulay, P. Can (semi)local density functional theory account for the London dispersion forces? *Chem. Phys. Lett.* **1994**, *229*, 175–180.
- (61) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.
- (62) Kong, J.; Gan, Z.; Proynov, E.; Freindorf, M.; Furlani, T. Efficient computation of the dispersion interaction with density-functional theory. *Phys. Rev. A* **2009**, *79*, 042510.
- (63) Becke, A. D.; Arabi, A. A.; Kannemann, F. O. Nonempirical density-functional theory for van der Waals interactions. *Can. J. Chem.* **2010**, *88*, 1057–1062.
- (64) Arabi, A. A.; Becke, A. D. Assessment of the PW86+PBE+XDM density functional on van der Waals complexes at non-equilibrium geometries. *J. Chem. Phys.* **2012**, *137*, 014104.
- (65) Eichkorn, K.; Treutler, O.; Ohm, H.; Marco, H.; Ahlrichs, R. Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.* **1995**, *240*, 283–290.
- (66) Weigend, F.; Haser, M.; Patzelt, H.; Ahlrichs, R. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem. Phys. Lett.* **1998**, *294*, 143–152.
- (67) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. Use of approximate integrals in ab initio theory. An application in MP2 energy calculations. *Chem. Phys. Lett.* **1993**, *208*, 359–363.
- (68) Jung, Y.; Sodt, A.; Gill, P. M. W.; Head-Gordon, M. Auxiliary basis expansions for large-scale electronic structure calculations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6692–6697.

- (69) Hobza, P.; Selzle, H.; Schlag, E. Potential Energy Surface for the Benzene Dimer. Results of ab Initio CCSD(T) Calculations Show Two Nearly Isoenergetic Structures: T-Shaped and Parallel-Displaced. *J. Phys. Chem.* **1996**, *100*, 18790–18794.
- (70) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence of correlated calculations on water. *J. Chem. Phys.* **1997**, *106*, 9639–9646.
- (71) Goldey, M.; Head-Gordon, M. Attenuating Away the Errors in Inter- and Intramolecular Interactions from Second-Order Møller-Plesset Calculations in the Small Aug-cc-pVDZ Basis Set. *J. Phys. Chem. Lett.* **2012**, *3*, 3592–3598.
- (72) Goldey, M.; Dutoi, A.; Head-Gordon, M. Attenuated second-order Møller-Plesset perturbation theory: performance within the aug-cc-pVTZ basis. *Phys. Chem. Chem. Phys.* **2013**, *15*, 15869–15875.
- (73) Besnard, M.; Cabaço, M. I.; Talaga, D.; Danten, Y Raman spectroscopy and ab initio investigations of transient complex formation in CO₂-benzene mixtures. *J. Chem. Phys.* **2008**, *129*, 224511.
- (74) Wolinski, K.; Pulay, P. Second-order Møller-Plesset calculations with dual basis sets. *J. Chem. Phys.* **2003**, *118*, 9497.
- (75) Liang, W.; Head-Gordon, M. Approaching the Basis Set Limit in Density Functional Theory Calculations Using Dual Basis Sets without Diagonalization. *J. Phys. Chem. A* **2004**, *108*, 3206–3210.
- (76) Steele, R. P.; DiStasio, R. A.; Shao, Y.; Kong, J.; Head-Gordon, M. Dual-basis second-order Møller-Plesset perturbation theory: A reduced-cost reference for correlation calculations. *J. Chem. Phys.* **2006**, *125*, 074108.
- (77) Steele, R. P.; DiStasio, R. A.; Head-Gordon, M. Non-Covalent Interactions with Dual-Basis Methods: Pairings for Augmented Basis Sets. *J. Chem. Theory Comput.* **2009**, *5*, 1560–1572.
- (78) Steele, R. P.; DiStasio, R. A.; Head-Gordon, M.; Li, Y.; Galli, G. The 1,4-phenylenedicyanide dimer: gas-phase properties and insights into organic self-assembled monolayers. *Phys. Chem. Chem. Phys.* **2010**, *12*, 82–96.
- (79) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (80) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- (81) Head-Gordon, M.; Pople, J.; Frisch, M. MP2 Energy Evaluation By Direct Methods. *Chem. Phys. Lett.* **1988**, *153*, 503–506.
- (82) Stephens, P.; Devlin, F.; Chabalowski, C.; Frisch, M. Ab Initio Calculation of Vibrational Absorption and Circular Dichromism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

- (83) Adamo, C.; Scuseria, G. E.; Barone, V. Accurate excitation energies from time-dependent density functional theory: Assessing the PBE0 model. *J. Chem. Phys.* **1999**, *111*, 2889.
- (84) Tao, J.; Perdew, J.; Staroverov, V.; Scuseria, G. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- (85) Peverati, R.; Truhlar, D. Improving the Accuracy of Hybrid Meta-GGA Density Functionals by Range Separation. *J. Phys. Chem. Lett.* **2011**, *2*, 2810–2817.
- (86) Chai, J.-D.; Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **2008**, *128*, 084106.
- (87) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–20.
- (88) Murray, E. D.; Lee, K.; Langreth, D. C. Investigation of Exchange Energy Density Functional Accuracy for Interacting Molecules. *J. Chem. Theory Comput.* **2009**, *5*, 2754–2762.
- (89) Perdew, J. P.; Wang, Y. Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation. *Phys. Rev. B* **1986**, *33*, 12–14.
- (90) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. High-Level ab Initio Calculations of Interaction Energies of C₂H₄-CH₄ and C₂H₆-CH₄ Dimers: A Model Study of CH/ π Interaction. *J. Phys. Chem. A* **1999**, *103*, 8265–8271.
- (91) Boys, S.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553–566.
- (92) Gill, P. M. W.; Johnson, B. G.; Pople, J. A.; Frisch, M. J. The performance of the Becke-Lee-Yang-Parr with various basis sets (B-LYP) density functional theory with various basis sets. *Chem. Phys. Lett.* **1992**, *197*, 499–505.
- (93) Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175–3183.
- (94) Shao, Y. et al. Advances in methods and algorithms in a modern quantum chemistry program package. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.
- (95) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **2012**, *4*, 1–17.
- (96) Szabo, A.; Ostlund, N. S. The correlation energy in the random phase approximation: Intermolecular forces between closed-shell systems. *J. Chem. Phys.* **1977**, *67*, 4351–4360.

- (97) Tsuzuki, S.; Uchamaru, T.; Mikami, M.; Tanabe, K. Intermolecular interaction potential of the carbon dioxide dimer. *J. Chem. Phys.* **1998**, *109*, 2169–2175.
- (98) Johnson, E. R.; Becke, A. D.; Sherrill, C. D.; DiLabio, G. A. Oscillations in meta-generalized-gradient approximation potential energy surfaces for dispersion-bound complexes. *J. Chem. Phys.* **2009**, *131*, 034111.
- (99) 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.
- (100) Berland, K.; Borck, Ø; Hyldgaard, P. van der Waals density functional calculations of binding in molecular crystals. *Comput. Phys. Commun.* **2011**, *182*, 1800–1804.
- (101) Lee, K.; Berland, K.; Yoon, M.; Andersson, S.; Schröder, E.; Hyldgaard, P.; Lundqvist, B. I. Benchmarking van der Waals density functionals with experimental data: potential energy curves for H₂ molecules on Cu(111), (100) and (110) surfaces. *J. Phys. Condens. Matter* **2012**, *24*, 424213.
- (102) Vydrov, O. A.; Van Voorhis, T. Benchmark assessment of the accuracy of several van der Waals density functionals. *J. Chem. Theory Comput.* **2012**, *8*, 1929–1934.
- (103) Rana, M. K.; Koh, H. S.; Hwang, J.; Siegel, D. J. Comparing van der Waals Density Functionals for CO₂ Adsorption in Metal Organic Frameworks. *J. Phys. Chem. C* **2012**, *116*, 16957–16968.
- (104) Hujo, W.; Grimme, S. Performance of the van der Waals Density Functional VV10 and (hybrid)GGA Variants for Thermochemistry and Noncovalent Interactions. *J. Chem. Theory Comput.* **2011**, *7*, 3866–3871.
- (105) Verma, P.; Xu, X.; Truhlar, D. G. Adsorption on Fe-MOF-74 for C₁-C₃ Hydrocarbon Separation. *J. Phys. Chem. C* **2013**, *117*, 12648–12660.
- (106) Sumida, K.; Stück, D.; Mino, L.; Chai, J.-D.; Bloch, E. D.; Zavorotynska, O.; Murray, L. J.; Dinc, M.; Chavan, S.; Bordiga, S.; Head-Gordon, M.; Long, J. R. Impact of metal and anion substitutions on the hydrogen storage properties of M-BTT metal-organic frameworks. *J. Am. Chem. Soc.* **2013**, *135*, 1083–91.
- (107) Roothaan, C. New developments in molecular orbital theory. *Rev. Mod. Phys.* **1951**, *23*, 69–89.
- (108) Hall, G. G. The Molecular Orbital Theory of Chemical Valency. VIII. A Method of Calculating Ionization Potentials. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1951**, *205*, 541–552.
- (109) Hehre, W. J.; Radom, L.; v.R. Schleyer, P.; Pople, J. A., *Ab Initio Molecular Orbital Theory*, 1st ed.; John Wiley and Sons: New York, 1986, pp 145–202.
- (110) Helgaker, T.; Jo, P. The prediction of molecular equilibrium structures by the standard electronic wave functions. *J. Chem. Phys.* **1997**, *106*, 6430–6440.

- (111) Sánchez Márquez, J.; Fernández Núñez, M. On the determination of bond lengths by ab initio methods: estimation of errors and some improvements. *J. Mol. Struct. THEOCHEM* **2003**, *624*, 239–249.
- (112) Riley, K.; Holt, B. O.; Merz, K. Critical assessment of the performance of density functional methods for several atomic and molecular properties. *J. Chem. Theory Comput.* **2007**, *3*, 407–433.
- (113) Guido, C. a.; Knecht, S.; Kongsted, J.; Mennucci, B. Benchmarking Time-Dependent Density Functional Theory for Excited State Geometries of Organic Molecules in Gas-Phase and in Solution. *J. Chem. Theory Comput.* **2013**, *9*, 2209–2220.
- (114) Tamblyn, I.; Refaely-Abramson, S.; Neaton, J. B.; Kronik, L. Simultaneous Determination of Structures, Vibrations, and Frontier Orbital Energies from a Self-Consistent Range-Separated Hybrid Functional. *J. Phys. Chem. Lett.* **2014**, *5*, 2734–2741.
- (115) Gráfová, L.; Pitonák, M.; ezáč, J.; Hobza, P. Comparative study of selected wave function and density functional methods for noncovalent interaction energy calculations using the extended S22 data set. *J. Chem. Theory Comput.* **2010**, *6*, 2365–2376.
- (116) Řezáč, 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.
- (117) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 4285–4292.
- (118) Goerigk, L.; Kruse, H.; Grimme, S. Benchmarking density functional methods against the S66 and S66x8 datasets for non-covalent interactions. *ChemPhysChem* **2011**, *12*, 3421–3433.
- (119) Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. Assessment of Theoretical Methods for Nonbonded Interactions: Comparison to Complete Basis Set Limit Coupled-Cluster Potential Energy Curves for the Benzene Dimer, the Methane Dimer, Benzene-Methane, and Benzene-H₂S. *J. Phys. Chem. A* **2009**, *113*, 10146–10159.
- (120) Vázquez-Mayagoitia, A.; Sherrill, C. D.; Aprà, E.; Sumpter, B. G. An Assessment of Density Functional Methods for Potential Energy Curves of Nonbonded Interactions: The XYG3 and B97-D Approximations. *J. Chem. Theory Comput.* **2010**, *6*, 727–734.
- (121) Takatani, T.; David Sherrill, C Performance of spin-component-scaled Møller-Plesset theory (SCS-MP2) for potential energy curves of noncovalent interactions. *Phys. Chem. Chem. Phys.* **2007**, *9*, 6106–6114.
- (122) Rao, L.; Ke, H.; Fu, G.; Xu, X.; Yan, Y. Performance of Several Density Functional Theory Methods on Describing Hydrogen-Bond Interactions. *J. Chem. Theory Comput.* **2009**, *5*, 86–96.

- (123) Thanthiriwatte, K. S.; Hohenstein, E. G.; Burns, L. a.; Sherrill, C. D. Assessment of the Performance of DFT and DFT-D Methods for Describing Distance Dependence of Hydrogen-Bonded Interactions. *J. Chem. Theory Comput.* **2011**, *7*, 88–96.
- (124) Hujo, W.; Grimme, S. Comparison of the performance of dispersion-corrected density functional theory for weak hydrogen bonds. *Phys. Chem. Chem. Phys.* **2011**, *13*, 13942–13950.
- (125) Remya, K.; Suresh, C. H. Which density functional is close to CCSD accuracy to describe geometry and interaction energy of small non-covalent dimers? A benchmark study using Gaussian09. *J. Comput. Chem.* **2013**, *34*, 1341–1353.
- (126) Gordon, M. S.; Truhlar, D. G. Scaling all correlation energy in perturbation theory calculations of bond energies and barrier heights. *J. Am. Chem. Soc.* **1986**, *108*, 5412–5419.
- (127) Grimme, S. Improved second-order Møller-Plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies. *J. Chem. Phys.* **2003**, *118*, 9095–910.
- (128) Jung, Y.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. Scaled opposite-spin second order Møller-Plesset correlation energy: an economical electronic structure method. *J. Chem. Phys.* **2004**, *121*, 9793–802.
- (129) Grimme, S. Density functional theory with London dispersion corrections. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 211–228.
- (130) Turney, J. M. et al. Psi4: an open-source ab initio electronic structure program. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 556–565.
- (131) Řezáč, J.; Hobza, P. Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the “Gold Standard,” CCSD(T) at the Complete Basis Set Limit? *J. Chem. Theory Comput.* **2013**, *9*, 2151–2155.
- (132) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N₂, and H₂O. *Chem. Phys. Lett.* **1998**, *286*, 243–252.
- (133) Gill, P. M.; Johnson, B. G.; Pople, J. A. A standard grid for density functional calculations. *Chem. Phys. Lett.* **1993**, *209*, 506–512.
- (134) Tang, K.; Toennies, J. New Combining Rules for Well Parameters and Shapes of the van der Waals Potential of Mixed Rare Gas Systems. *Zeitschrift für Phys. D Atoms, Mol. Clust.* **1986**, *1*, 91–101.
- (135) Miliordos, E.; Aprà, E.; Xantheas, S. S. Benchmark Theoretical Study of the π - π Binding Energy in the Benzene Dimer. *J. Phys. Chem. A* **2014**, *118*, 7568–7578.
- (136) Cohen, M.; Kelly, P. Hartree-fock wave functions for excited states II. simplification of the orbital equations. *Can. J. Phys.* **1966**, *44*, 3227–3240.

- (137) Janowski, T.; Ford, A. R.; Pulay, P. Accurate correlated calculation of the intermolecular potential surface in the coronene dimer. *Mol. Phys.* **2010**, *108*, 249–257.
- (138) Huang, Y.; Goldey, M.; Head-Gordon, M.; Beran, G. J. O. Achieving High-Accuracy Intermolecular Interactions by Combining Coulomb-Attenuated Second-Order Møller-Plesset Perturbation Theory with Coupled Kohn-Sham Dispersion. *J. Chem. Theory Comput.* **2014**, *10*, 2054–2063.
- (139) Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I. Binding energy curves from nonempirical density functionals II. van der Waals bonds in rare-gas and alkaline-earth diatomics. *J. Phys. Chem. A* **2005**, *109*, 11015–11021.
- (140) Wu, Q.; Yang, W. Empirical correction to density functional theory for van der Waals interactions. *J. Chem. Phys.* **2002**, *116*, 515.
- (141) Sedlak, R.; Janowski, T.; Pitoňák, M.; Řezáč, J.; Pulay, P.; Hobza, P. The accuracy of quantum chemical methods for large noncovalent complexes. *J. Chem. Theory Comput.* **2013**, *9*, 3364–3374.
- (142) Goldey, M.; Head-Gordon, M. Separate electronic attenuation allowing a spin-component-scaled second-order Møller-Plesset theory to be effective for both thermochemistry and noncovalent interactions. *J. Phys. Chem. B* **2014**, *118*, 6519–6525.
- (143) Boese, A. D.; Handy, N. C. A new parametrization of exchange-correlation generalized gradient approximation functionals. *J. Chem. Phys.* **2001**, *114*, 5497.
- (144) Boese, A. D.; Handy, N. C. New exchange-correlation density functionals: The role of the kinetic-energy density. *J. Chem. Phys.* **2002**, *116*, 9559.
- (145) Boese, A. D.; Martin, J. M. L. Development of density functionals for thermochemical kinetics. *J. Chem. Phys.* **2004**, *121*, 3405–16.
- (146) Schwerdtfeger, P. The Pseudopotential Approximation in Electronic Structure Theory. *ChemPhysChem* **2011**, *12*, 3143–3155.
- (147) Slater, J. C. Atomic shielding constants. *Phys. Rev.* **1930**, *36*, 57–64.
- (148) Boys, S. F. Electronic Wave Functions. I. A General Method of Calculation for the Stationary States of Any Molecular System. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1950**, *200*, 542–554.
- (149) Jensen, F. Atomic orbital basis sets. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 273–295.
- (150) Peterson, K. A.; Woon, D. E.; Dunning, T. H. Benchmark calculations with correlated molecular wave functions. IV. The classical barrier height of the $\text{H}+\text{H}_2\rightarrow\text{H}_2+\text{H}$ reaction. *J. Chem. Phys.* **1994**, *100*, 7410–7415.
- (151) Jensen, F. Polarization consistent basis sets: Principles. *J. Chem. Phys.* **2001**, *115*, 9113–9125.

- (152) Jensen, F. Polarization consistent basis sets: II. Estimating the Kohn-Sham basis set limit. *J. Chem. Phys.* **2002**, *116*, 7372–7379.
- (153) Jensen, F. Polarization consistent basis sets. III. The importance of diffuse functions. *J. Chem. Phys.* **2002**, *117*, 9234–9240.
- (154) 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.
- (155) Rappoport, D.; Furche, F. Property-optimized Gaussian basis sets for molecular response calculations. *J. Chem. Phys.* **2010**, *133*, 134105.
- (156) Schwartz, C. Importance of Angular Correlations between Atomic Electrons. *Phys. Rev.* **1962**, *126*, 1015–1019.
- (157) Hill, R. N. Rates of convergence and error estimation formulas for the Rayleigh-Ritz variational method. *J. Chem. Phys.* **1985**, *83*, 1173–1196.
- (158) Jansen, H. B.; Ros, P. Non-Empirical Molecular Orbital Calculations of Carbon Monoxide. *Chem. Phys. Lett.* **1969**, *3*, 140–143.
- (159) Sure, R.; Brandenburg, J. G.; Grimme, S. Small Atomic Orbital Basis Set First-Principles Quantum Chemical Methods for Large Molecular and Periodic Systems: A Critical Analysis of Error Sources. *ChemistryOpen* **2015**, n/a–n/a.
- (160) 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.
- (161) Collins, J.; Gallup, G. The full versus the virtual counterpoise correction for basis set superposition error in self-consistent field calculations. *Chem. Phys. Lett.* **1986**, *123*, 56–61.
- (162) Frisch, M. J.; Del Bene, J. E.; Binkley, J. S.; Schaefer III, H. F. Extensive theoretical studies of the hydrogen-bonded complexes water dimer, diaquahydrogen(1+)((H₂O)₂H⁺), hydrogen fluoride dimer, di(hydrofluoro)hydrogen(1+) ((HF)₂H⁺), fluoride (F₂H⁻), and ammonia dimer. *J. Chem. Phys.* **1986**, *84*, 2279–2289.
- (163) Gutowski, M.; Van Duijneveldt, F.; Chałasiński, G.; Piela, L. Does the boys and bernardi function counterpoise method actually overcorrect the basis set superposition error? *Chem. Phys. Lett.* **1986**, *129*, 325–328.
- (164) Cook, D. B.; Sordo, J. A.; Sordo, T. L. Some Comments on the Counterpoise Correction for the Basis-Set Superposition Error at the Correlated Level. *Int. J. Quantum Chem.* **1993**, *48*, 375–384.
- (165) 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.
- (166) Mentel, M.; Baerends, E. J. Can the counterpoise correction for basis set superposition effect be justified? *J. Chem. Theory Comput.* **2014**, *10*, 252–267.

- (167) Kalescky, R.; Kraka, E.; Cremer, D. Accurate determination of the binding energy of the formic acid dimer: the importance of geometry relaxation. *J. Chem. Phys.* **2014**, *140*, 084315.
- (168) Miliordos, E.; Xantheas, S. S. On the validity of the basis set superposition error and complete basis set limit extrapolations for the binding energy of the formic acid dimer. *J. Chem. Phys.* **2015**, *142*, 094311.
- (169) 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.
- (170) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724–728.
- (171) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- (172) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28*, 213–222.
- (173) Hariharan, P. C.; Pople, J. A. Equilibrium Geometries By Single Determinant Molecular Orbital Theory. *Mol. Phys.* **1974**, *27*, 209–214.
- (174) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980**, *72*, 650–654.
- (175) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77*, 3654–3665.
- (176) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li-F. *J. Comput. Chem.* **1983**, *4*, 294–301.
- (177) 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.
- (178) Woon, D. E.; Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties. *J. Chem. Phys.* **1994**, *100*, 2975–2988.
- (179) Woon, D. E.; Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. V. Core-valence basis sets for boron through neon. *J. Chem. Phys.* **1995**, *103*, 4572–4585.

- (180) Papajak, E.; Zheng, J.; Xu, X.; Leverentz, H. R.; Truhlar, D. G. Perspectives on basis sets beautiful: Seasonal plantings of diffuse basis functions. *J. Chem. Theory Comput.* **2011**, *7*, 3027–3034.
- (181) Jurecka, P.; Sponer, J.; Cerný, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.
- (182) Shao, Y. et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, 184–215.
- (183) Manzer, S.; Horn, P. R.; Mardirossian, N.; Head-Gordon, M. Fast, accurate evaluation of exact exchange: The occ-RI-K algorithm. *J. Chem. Phys.* **2015**, *143*, 024113.
- (184) Weigend, F. A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.
- (185) Weigend, F. Hartree-Fock exchange fitting basis sets for H to Rn. *J. Comput. Chem.* **2008**, *29*, 167–75.
- (186) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. Basis set convergence of the coupled-cluster correction, $\delta(\text{MP2})(\text{CCSD(T)})$: best practices for benchmarking non-covalent interactions and the attendant revision of the S22, NBC10, HBC6, and HSG databases. *J. Chem. Phys.* **2011**, *135*, 194102.
- (187) Vydrov, O. A.; Wu, Q.; Van Voorhis, T. Self-consistent implementation of a nonlocal van der Waals density functional with a Gaussian basis set. *J. Chem. Phys.* **2008**, *129*, 014106.
- (188) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. Effectiveness of Diffuse Basis Functions for Calculating Relative Energies by Density Functional Theory. *J. Phys. Chem. A* **2003**, *107*, 1384–1388.
- (189) Faver, J. C.; Zheng, Z.; Merz, K. M. Model for the fast estimation of basis set superposition error in biomolecular systems. *J. Chem. Phys.* **2011**, *135*, 144110.
- (190) Faver, J. C.; Zheng, Z.; Merz, K. M. Statistics-based model for basis set superposition error correction in large biomolecules. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7795–7799.
- (191) Kruse, H.; Grimme, S. A geometrical correction for the inter- and intra-molecular basis set superposition error in Hartree-Fock and density functional theory calculations for large systems. *J. Chem. Phys.* **2012**, *136*, 154101.
- (192) Grimme, S.; Brandenburg, J. G.; Bannwarth, C.; Hansen, A. Consistent structures and interactions by density functional theory with small atomic orbital basis sets. *J. Chem. Phys.* **2015**, *143*, 054107.

- (193) Dunning, T. H. A Road Map for the Calculation of Molecular Binding Energies. *J. Phys. Chem. A* **2000**, *104*, 9062–9080.
- (194) Bak, K. L.; Jørgensen, P.; Olsen, J.; Helgaker, T.; Klopper, W. Accuracy of atomization energies and reaction enthalpies in standard and extrapolated electronic wave function/basis set calculations. *J. Chem. Phys.* **2000**, *112*, 9229–9242.
- (195) Bak, K. L.; Gauss, J.; Jørgensen, P.; Olsen, J.; Helgaker, T.; Stanton, J. F. The accurate determination of molecular equilibrium structures. *J. Chem. Phys.* **2001**, *114*, 6548.
- (196) Shahbazian, S.; Zahedi, M. Towards a complete basis set limit of Hartree-Fock method: Correlation-consistent versus polarized-consistent basis sets. *Theor. Chem. Acc.* **2005**, *113*, 152–160.
- (197) Kupka, T.; Lim, C. Polarization-consistent versus correlation-consistent basis sets in predicting molecular and spectroscopic properties. *J. Phys. Chem. A* **2007**, *111*, 1927–1932.
- (198) Elsohly, A. M.; Tschumper, G. S. Comparison of polarization consistent and correlation consistent basis sets for noncovalent interactions. *Int. J. Quantum Chem.* **2009**, *109*, 91–96.
- (199) Eshuis, H.; Furche, F. Basis set convergence of molecular correlation energy differences within the random phase approximation. *J. Chem. Phys.* **2012**, *136*, 084105.
- (200) Christensen, K. A.; Jensen, F. The basis set convergence of the density functional energy for H₂. *Chem. Phys. Lett.* **2000**, *317*, 400–403.
- (201) Galano, A.; Alvarez-Idaboy, J. R. A new approach to counterpoise correction to BSSE. *J. Comput. Chem.* **2006**, *27*, 1203–1210.
- (202) Jensen, F. An atomic counterpoise method for estimating inter- and intramolecular basis set superposition errors. *J. Chem. Theory Comput.* **2010**, *6*, 100–106.
- (203) Perdew, J. P.; Schmidt, K. In *AIP Conf. Proc.* Ed. by Van Doren, V. E.; Van Alsenoy, K.; Geerlings, P, American Institute of Physics: Melville, NY, 2001; Vol. 577, pp 1–20.
- (204) Hepburn, J.; Scoles, G.; Penco, R. A simple but reliable method for the prediction of intermolecular potentials. *Chem. Phys. Lett.* **1975**, *36*, 451–456.
- (205) Ahlrichs, R.; Penco, R.; Scoles, G. Intermolecular forces in simple systems. *Chem. Phys.* **1977**, *19*, 119–130.
- (206) Von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. Optimization of Effective Atom Centered Potentials for London Dispersion Forces in Density Functional Theory. *Phys. Rev. Lett.* **2004**, *93*, 153004.
- (207) DiLabio, G. A. Accurate treatment of van der Waals interactions using standard density functional theory methods with effective core-type potentials: Application to carbon-containing dimers. *Chem. Phys. Lett.* **2008**, *455*, 348–353.

- (208) Schröder, H.; Creon, A.; Schwabe, T. Reformulation of the D3(Becke-Johnson) Dispersion Correction without Resorting to Higher than C 6 Dispersion Coefficients. *J. Chem. Theory Comput.* **2015**, *11*, 3163–3170.
- (209) 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.
- (210) Zhang, Y.; Yang, W. Comment on “Generalized Gradient Approximation Made Simple”. *Phys. Rev. Lett.* **1998**, *80*, 890–890.
- (211) Sun, J.; Haunschild, R.; Xiao, B.; Bulik, I. W.; Scuseria, G. E.; Perdew, J. P. Semilocal and hybrid meta-generalized gradient approximations based on the understanding of the kinetic-energy-density dependence. *J. Chem. Phys.* **2013**, *138*, 1–9.
- (212) Gillan, M. J.; Alfè, D.; Michaelides, A. Perspective: How good is DFT for water? *J. Chem. Phys.* **2016**, *144*, 130901.
- (213) Řezáč, J.; Riley, K. E.; Hobza, P. Extensions of the S66 Data Set: More Accurate Interaction Energies and Angular-Displaced Nonequilibrium Geometries. *J. Chem. Theory Comput.* **2011**, *7*, 3466–3470.
- (214) Boese, A. D. Assessment of coupled cluster theory and more approximate methods for hydrogen bonded systems. *J. Chem. Theory Comput.* **2013**, *9*, 4403–4413.
- (215) Boese, A. D. Basis set limit coupled-cluster studies of hydrogen-bonded systems. *Mol. Phys.* **2015**, *113*, 1–12.
- (216) Boese, A. D. Density functional theory and hydrogen bonds: Are we there yet? *ChemPhysChem* **2015**, *16*, 978–985.
- (217) Granatier, J.; Pitoňák, M.; Hobza, P. Accuracy of several wave function and density functional theory methods for description of noncovalent interaction of saturated and unsaturated hydrocarbon dimers. *J. Chem. Theory Comput.* **2012**, *8*, 2282–2292.
- (218) Lao, K. U.; Herbert, J. M. An improved treatment of empirical dispersion and a many-body energy decomposition scheme for the explicit polarization plus symmetry-adapted perturbation theory (XSAPT) method. *J. Chem. Phys.* **2013**, *139*, 034107.
- (219) Lao, K. U.; Schäffer, R.; Jansen, G.; Herbert, J. M. Accurate Description of Intermolecular Interactions Involving Ions Using Symmetry-Adapted Perturbation Theory. *J. Chem. Theory Comput.* **2015**, *11*, 2473–2486.
- (220) Bryantsev, V. S.; Diallo, M. S.; Van Duin, A. C. T.; Goddard, W. A. Evaluation of B3LYP, X3LYP, and M06-Class density functionals for predicting the binding energies of neutral, protonated, and deprotonated water clusters. *J. Chem. Theory Comput.* **2009**, *5*, 1016–1026.
- (221) Goerigk, L.; Grimme, S. A general database for main group thermochemistry, kinetics, and noncovalent interactions - Assessment of common and reparameterized (meta-)GGA density functionals. *J. Chem. Theory Comput.* **2010**, *6*, 107–126.

- (222) Fanourgakis, G. S.; Aprà, E.; Xantheas, S. S. High-level ab initio calculations for the four low-lying families of minima of (H₂O)₂₀. I. Estimates of MP2/CBS binding energies and comparison with empirical potentials. *J. Chem. Phys.* **2004**, *121*, 2655–2663.
- (223) Anacker, T.; Friedrich, J. New accurate benchmark energies for large water clusters: DFT is better than expected. *J. Comput. Chem.* **2014**, *35*, 634–643.
- (224) Kozuch, S.; Bachrach, S. M.; Martin, J. M. L. Conformational equilibria in butane-1,4-diol: A benchmark of a prototypical system with strong intramolecular H-bonds. *J. Phys. Chem. A* **2014**, *118*, 293–303.
- (225) Fogueri, U. R.; Kozuch, S.; Karton, A.; Martin, J. M. L. The melatonin conformer space: Benchmark and assessment of wave function and DFT methods for a paradigmatic biological and pharmacological molecule. *J. Phys. Chem. A* **2013**, *117*, 2269–2277.
- (226) Yoo, S.; Aprà, E.; Zeng, X. C.; Xantheas, S. S. High-level Ab initio electronic structure calculations of water clusters (H₂O)₁₆ and (H₂O)₁₇: A new global minimum for (H₂O)₁₆. *J. Phys. Chem. Lett.* **2010**, *1*, 3122–3127.
- (227) Crittenden, D. L. A Systematic CCSD(T) Study of Long-Range and Noncovalent Interactions between Benzene and a Series of First- and Second-Row Hydrides and Rare Gas Atoms. *J. Phys. Chem. A* **2009**, *113*, 1663–1669.
- (228) Hohenstein, E. G.; Sherrill, C. D. Effects of heteroatoms on aromatic π - π interactions: Benzene-pyridine and pyridine dimer. *J. Phys. Chem. A* **2009**, *113*, 878–886.
- (229) Mintz, B. J.; Parks, J. M. Benchmark Interaction Energies for Biologically Relevant Noncovalent Complexes Containing Divalent Sulfur. *J. Phys. Chem. A* **2012**, *116*, 1086–1092.
- (230) Faver, J. C.; Benson, M. L.; He, X.; Roberts, B. P.; Wang, B.; Marshall, M. S.; Kennedy, M. R.; Sherrill, C. D.; Merz, K. M. Formal estimation of errors in computed absolute interaction energies of protein-ligand complexes. *J. Chem. Theory Comput.* **2011**, *7*, 790–797.
- (231) Řezáč, J.; Hobza, P. Advanced corrections of hydrogen bonding and dispersion for semiempirical quantum mechanical methods. *J. Chem. Theory Comput.* **2012**, *8*, 141–151.
- (232) Copeland, K. L.; Tschumper, G. S. Hydrocarbon/water interactions: Encouraging energetics and structures from dft but disconcerting discrepancies for hessian indices. *J. Chem. Theory Comput.* **2012**, *8*, 1646–1656.
- (233) Smith, D. G. A.; Jankowski, P.; Slawik, M.; Witek, H. A.; Patkowski, K. Basis Set Convergence of the Post-CCSD(T) Contribution to Noncovalent Interaction Energies. *J. Chem. Theory Comput.* **2014**, *10*, 3140–3150.

- (234) Řezáč, J.; Huang, Y.; Hobza, P.; Beran, G. J. O. Benchmark Calculations of Three-Body Intermolecular Interactions and the Performance of Low-Cost Electronic Structure Methods. *J. Chem. Theory Comput.* **2015**, *11*, 3065–3079.
- (235) Li, S.; Smith, D. G. A.; Patkowski, K. An accurate benchmark description of the interactions between carbon dioxide and polyheterocyclic aromatic compounds containing nitrogen. *Phys. Chem. Chem. Phys.* **2015**, *17*, 16560–16574.
- (236) Lao, K. U.; Herbert, J. M. Accurate and Efficient Quantum Chemistry Calculations for Noncovalent Interactions in Many-Body Systems: The XSAPT Family of Methods. *J. Phys. Chem. A* **2015**, *119*, 235–252.
- (237) Temelso, B.; Archer, K. A.; Shields, G. C. Benchmark Structures and Binding Energies of Small Water Clusters with Anharmonicity Corrections. *J. Phys. Chem. A* **2011**, *115*, 12034–12046.
- (238) Mardirossian, N.; Lambrecht, D. S.; McCaslin, L.; Xantheas, S. S.; Head-Gordon, M. The Performance of Density Functionals for Sulfate-Water Clusters. *J. Chem. Theory Comput.* **2013**, *9*, 1368–1380.
- (239) Karton, A.; O'Reilly, R. J.; Chan, B.; Radom, L. Determination of Barrier Heights for Proton Exchange in Small Water, Ammonia, and Hydrogen Fluoride Clusters with G4(MP2)-Type, MP_n, and SCS-MP_n Procedures—A Caveat. *J. Chem. Theory Comput.* **2012**, *8*, 3128–3136.
- (240) Chan, B.; Gilbert, A. T. B.; Gill, P. M. W.; Radom, L. Performance of density functional theory procedures for the calculation of proton-exchange barriers: Unusual behavior of M06-type functionals. *J. Chem. Theory Comput.* **2014**, *10*, 3777–3783.
- (241) Tentscher, P. R.; Arey, J. S. Binding in Radical-Solvent Binary Complexes: Benchmark Energies and Performance of Approximate Methods. *J. Chem. Theory Comput.* **2013**, *9*, 1568–1579.
- (242) Kozuch, S.; Martin, J. M. L. Halogen bonds: Benchmarks and theoretical analysis. *J. Chem. Theory Comput.* **2013**, *9*, 1918–1931.
- (243) Bauzá, A.; Alkorta, I.; Frontera, A.; Elguero, J. On the reliability of pure and hybrid DFT methods for the evaluation of halogen, chalcogen, and pnictogen bonds involving anionic and neutral electron donors. *J. Chem. Theory Comput.* **2013**, *9*, 5201–5210.
- (244) Otero-de-la Roza, A.; Johnson, E. R.; DiLabio, G. A. Halogen Bonding from Dispersion-Corrected Density-Functional Theory: The Role of Delocalization Error. *J. Chem. Theory Comput.* **2014**, *10*, 5436–5447.
- (245) Steinmann, S. N.; Piemontesi, C.; Delachat, A.; Corminboeuf, C. Why are the Interaction Energies of Charge-Transfer Complexes Challenging for DFT? *J. Chem. Theory Comput.* **2012**, *8*, 1629–1640.

- (246) Karton, A.; Gruzman, D.; Martin, J. M. L. Benchmark thermochemistry of the C_nH_{2n+2} alkane isomers ($n = 2-8$) and performance of DFT and composite ab initio methods for dispersion-driven isomeric equilibria. *J. Phys. Chem. A* **2009**, *113*, 8434–8447.
- (247) Gruzman, D.; Karton, A.; Martin, J. M. L. Performance of Ab Initio and Density Functional Methods for Conformational Equilibria of C_nH_{2n+2} Alkane Isomers ($n = 4-8$). *J. Phys. Chem. A* **2009**, *113*, 11974–11983.
- (248) Wilke, J. J.; Lind, M. C.; Schaefer, H. F.; Csaszar, A. G.; Allen, W. D. Conformers of Gaseous Cysteine. *J. Chem. Theory Comput.* **2009**, *5*, 1511–1523.
- (249) Martin, J. M. L. What Can We Learn about Dispersion from the Conformer Surface of n -Pentane? *J. Phys. Chem. A* **2013**, *117*, 3118–3132.
- (250) Kesharwani, M. K.; Karton, A.; Martin, J. M. L. Benchmark ab Initio Conformational Energies for the Proteinogenic Amino Acids through Explicitly Correlated Methods. Assessment of Density Functional Methods. *J. Chem. Theory Comput.* **2016**, *12*, 444–454.
- (251) Yu, L.-J.; Karton, A. Assessment of theoretical procedures for a diverse set of isomerization reactions involving double-bond migration in conjugated dienes. *Chem. Phys.* **2014**, *441*, 166–177.
- (252) Karton, A.; Daon, S.; Martin, J. M. L. W4-11: A high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data. *Chem. Phys. Lett.* **2011**, *510*, 165–178.
- (253) Manna, D.; Martin, J. M. L. What Are the Ground State Structures of C_{20} and C_{24} ? An Explicitly Correlated Ab Initio Approach. *J. Phys. Chem. A* **2016**, *120*, 153–160.
- (254) Curtiss, L. A.; Raghavachari, K. Gaussian-2 theory for molecular energies of first- and second-row compounds. *J. Chem. Phys.* **1991**, *94*, 7221–7230.
- (255) Zhao, Y.; González-Garda, N.; Truhlar, D. G. Benchmark database of barrier heights for heavy atom transfer, nucleophilic substitution, association, and unimolecular reactions and its use to test theoretical methods. *J. Phys. Chem. A* **2005**, *109*, 2012–2018.
- (256) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. Multi-coefficient extrapolated density functional theory for thermochemistry and thermochemical kinetics. *Phys. Chem. Chem. Phys.* **2005**, *7*, 43–52.
- (257) Grimme, S.; Kruse, H.; Goerigk, L.; Erker, G. The mechanism of dihydrogen activation by frustrated Lewis pairs revisited. *Angew. Chemie - Int. Ed.* **2010**, *49*, 1402–1405.

- (258) Goerigk, L.; Grimme, S. Efficient and accurate double-hybrid-meta-GGA density functionals- evaluation with the extended GMTKN30 database for general main group thermochemistry, kinetics, and noncovalent interactions. *J. Chem. Theory Comput.* **2011**, *7*, 291–309.
- (259) Krieg, H.; Grimme, S. Thermochemical benchmarking of hydrocarbon bond separation reaction energies: Jacob’s ladder is not reversed! *Mol. Phys.* **2010**, *108*, 2655–2666.
- (260) O’Reilly, R. J.; Karton, A. A dataset of highly accurate homolytic NBr bond dissociation energies obtained by Means of W2 theory. *Int. J. Quantum Chem.* **2016**, *116*, 52–60.
- (261) Karton, A.; Martin, J. M. L. Explicitly correlated benchmark calculations on C₈H₈ isomer energy separations: how accurate are DFT, double-hybrid, and composite ab initio procedures? *Mol. Phys.* **2012**, *110*, 19–20.
- (262) Karton, A.; Schreiner, P. R.; Martin, J. M. L. Heats of formation of platonic hydrocarbon cages by means of high-level thermochemical procedures. *J. Comput. Chem.* **2016**, *37*, 49–58.
- (263) Karton, A.; Goerigk, L. Accurate reaction barrier heights of pericyclic reactions: Surprisingly large deviations for the CBS-QB3 composite method and their consequences in DFT benchmark studies. *J. Comput. Chem.* **2015**, *36*, 622–632.
- (264) Yu, L.-J.; Sarrami, F.; O’Reilly, R. J.; Karton, A. Reaction barrier heights for cycloreversion of heterocyclic rings: An Achilles’ heel for DFT and standard ab initio procedures. *Chem. Phys.* **2015**, *458*, 1–8.
- (265) Zheng, J.; Zhao, Y.; Truhlar, D. G. Representative Benchmark Suites for Barrier Heights of Diverse Reaction Types and Assessment of Electronic Structure Methods for Thermochemical Kinetics. *J. Chem. Theory Comput.* **2007**, *3*, 569–582.
- (266) Karton, A.; Tarnopolsky, A.; Lamère, J. F.; Schatz, G. C.; Martin, J. M. L. Highly accurate first-principles benchmark data sets for the parametrization and validation of density functional and other approximate methods. Derivation of a robust, generally applicable, double-hybrid functional for thermochemistry and thermochemical. *J. Phys. Chem. A* **2008**, *112*, 12868–12886.
- (267) Yu, L.-J.; Sarrami, F.; O’Reilly, R. J.; Karton, A. Can DFT and ab initio methods describe all aspects of the potential energy surface of cycloreversion reactions? *Mol. Phys.* **2016**, *114*, 21–33.
- (268) Schröder, H.; Hühnert, J.; Schwabe, T. Evaluation of DFT-D3 dispersion corrections for various structural benchmark sets. *J. Chem. Phys.* **2017**, *146*, 044115.
- (269) Tao, J.; Mo, Y. Accurate Semilocal Density Functional for Condensed-Matter Physics and Quantum Chemistry. *Phys. Rev. Lett.* **2016**, *117*, 73001.

- (270) Goerigk, L. How Do DFT-DCP, DFT-NL, and DFT-D3 Compare for the Description of London-Dispersion Effects in Conformers and General Thermochemistry? *J. Chem. Theory Comput.* **2014**, *10*, 968–980.
- (271) Tkatchenko, A.; DiStasio, R. A.; Car, R.; Scheffler, M. Accurate and Efficient Method for Many-Body van der Waals Interactions. *Phys. Rev. Lett.* **2012**, *108*, 236402.
- (272) Bowler, D. R.; Miyazaki, T. Calculations for millions of atoms with density functional theory: linear scaling shows its potential. *Journal of Physics: Condensed Matter* **2010**, *22*, 074207.
- (273) Skylaris, C.-K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. Introducing ONETEP: Linear-scaling density functional simulations on parallel computers. *The Journal of Chemical Physics* **2005**, *122*, 084119.
- (274) Skylaris, C.-K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. Recent progress in linear-scaling density functional calculations with plane waves and pseudopotentials: the ONETEP code. *Journal of Physics: Condensed Matter* **2008**, *20*, 064209.
- (275) Rayson, M. J.; Briddon, P. R. Highly efficient method for Kohn-Sham density functional calculations of 500~10 000 atom systems. *Phys. Rev. B* **2009**, *80*, 205104.
- (276) Rayson, M. Rapid filtration algorithm to construct a minimal basis on the fly from a primitive Gaussian basis. *Computer Physics Communications* **2010**, *181*, 1051 – 1056.
- (277) Mao, Y.; Horn, P. R.; Mardirossian, N.; Head-Gordon, T.; Skylaris, C.-K.; Head-Gordon, M. Approaching the basis set limit for DFT calculations using an environment-adapted minimal basis with perturbation theory: Formulation, proof of concept, and a pilot implementation. *The Journal of Chemical Physics* **2016**, *145*, 044109.
- (278) Khaliullin, R. Z.; Head-Gordon, M.; Bell, A. T. An efficient self-consistent field method for large systems of weakly interacting components. *J. Chem. Phys.* **2006**, *124*, 204105.
- (279) Horn, P. R.; Sundstrom, E. J.; Baker, T. A.; Head-Gordon, M. Unrestricted absolutely localized molecular orbitals for energy decomposition analysis: Theory and applications to intermolecular interactions involving radicals. *J. Chem. Phys.* **2013**, *138*, 134119.
- (280) Sure, R.; Grimme, S. Corrected small basis set Hartree-Fock method for large systems. *J. Comput. Chem.* **2013**, *34*, 1672–1685.
- (281) Brandenburg, J. G.; Caldeweyher, E.; Grimme, S. Screened exchange hybrid density functional for accurate and efficient structures and interaction energies. *Phys. Chem. Chem. Phys.* **2016**, *18*, 15519–15523.
- (282) Roothaan, C. C. J. Self-Consistent Field Theory for Open Shells of Electronic Systems. *Rev. Mod. Phys.* **1960**, *32*, 179–185.

- (283) 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.
- (284) Morrell, M. M.; Parr, R. G.; Levy, M. Calculation of ionization potentials from density matrices and natural functions, and the long-range behavior of natural orbitals and electron density. *J. Chem. Phys.* **1975**, *62*, 549–554.
- (285) Dutoi, A. D.; Head-Gordon, M. A study of the effect of attenuation curvature on molecular correlation energies by introducing an explicit cutoff radius into two-electron integrals. *J. Phys. Chem. A* **2008**, *112*, 2110–2119.
- (286) Tang, K. T.; Toennies, J. P. The van der Waals potentials between all the rare gas atoms from He to Rn. *J. Chem. Phys.* **2003**, *118*, 4976–4983.