Non-Covalent Interactions Atlas Benchmark Data Sets 5: London Dispersion in an Extended Chemical Space

Jan Řezáč*

Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, 166 10 Prague, Czech Republic

E-mail: rezac@uochb.cas.cz

January 11, 2022

Abstract

The Non-Covalent Interactions Atlas (www.nciatlas.org) has been extended with two data sets of benchmark interaction energies in complexes dominated by London dispersion. The D1200 data set of equilibrium geometries provides a thorough sampling of an extended chemical space, while the D442×10 set features dissociation curves for selected complexes. In total, they provide 5,178 new CCSD(T)/CBS data points of the highest quality. The new data have been combined with previous NCIA data sets in a comprehensive test of dispersion-corrected DFT methods, identifying the ones that achieve high accuracy in all types of non-covalent interactions in a broad chemical space. Additional tests of dispersion-corrected MP2 and semiempirical QM methods are also reported.

1 Introduction

The Non-Covalent Interactions Atlas (NCIA, `www.nciatlas.org`) is a collection of benchmark data sets of interaction energies in non-covalent complexes. This work introduces two new data sets covering complexes mainly driven by London dispersion, the D1200 set of equilibrium geometries and the D442×10 set of dissociation curves computed for selected systems from the D1200 set. These two data sets bring 5,178 interaction energies computed at a true benchmark CCSD(T)/CBS level, a substantial addition to the 11,445 data points provided by the earlier NCIA data sets covering hydrogen bonds[1,2] and repulsive contacts.[3]

Compared to earlier data sets used in the field, the NCIA data sets are not only larger and more accurate, but they also offer additional advantages such as more systematic construction, increased diversity of the model systems, quality geometries verified to be minima and, last but not least, the availability of all the data in an easy-to-use form.

Recently, two other large databases of benchmark non-covalent interactions have been introduced by other authors. The NENCI-2021 data set[4] combines the 66 dimers from the S66 data set[5,6] with another 35 complexes of organic molecules with water[7] and 40 new ion-π complexes, sampling both dissociation curves (up to a very repulsive region) and angular displacements, which results in 7,763 data points computed at a proper benchmark level. It is a well-rounded data set for e.g. the testing of wavefunction or density functional theory (DFT) methods as it extends the coverage of earlier data sets to more complicated systems; nevertheless, with only 141 unique dimers, it does not have enough diversity for the parameterization of more empirical methods. The DES370K database[8] is much larger, comprising 3,700 non-covalent complexes with an extensive sampling of non-equilibrium geometries, resulting in 370,000 data points. Here, the accuracy of the CCSD(T) benchmark is slightly lower as smaller basis sets are used in the larger systems. This data set is complemented by an even larger set of 5,000,000 data points computed using MP2 with a correction fitted to the core data set. Because of their size and extensive sampling of non-equilibrium geometries, these databases are well-suited for the development of machine learning (ML) models,

but they may be too large for applications in quantum chemistry.

One of the most important features of the current data sets and the whole NCIA database is the coverage of a wider chemical space. The D1200 and D442×10 sets cover not only the usual organic chemistry but also boron, sulfur, phosphorus, halogens up to iodine, and noble gases up to xenon. This is necessary for the development of more empirical methods which use element-specific parameters (such as molecular mechanics forcefields or semiempirical quantum-mechanical methods) that would cover this chemical space. It is, however, very important also for the testing and development of methods using only more general parameters, such as DFT functionals and dispersion corrections for them, where the diversity of the data sets helps to identify possible issues not found in a narrower chemical space. Moreover, the systematic construction of the NCIA database with data sets devoted to specific classes of non-covalent interactions as well as further grouping of the systems in each data set allow a straightforward interpretation of the results in line with the intuitive classification of the non-covalent interactions used in the field (but also with the actual physics of the interactions, as we provide interaction energy decomposition for all the systems).

Focusing on the London dispersion covered by the two data sets presented here, these data sets address two different requirements in method development and validation. The D1200 data set of 1,200 complexes in equilibrium geometries is large enough to provide a detailed, systematic coverage of the target chemical space. For each element, multiple atom types are considered, and the data set contains several representatives for each combination of these atom types. The D442×10 set is a subset of 442 systems from the D1200 set (where each combination of atom types is present only once), for which ten-point dissociation curves have been constructed. This data set covers the most important coordinate of the intermolecular potential energy surface crucial for method development. A combination of these two data sets provides both the coverage of non-equilibrium geometries and a high diversity of systems but keeps the size of the database manageable.

The D1200 and D442×10 data sets can be used on their own in applications focused solely

on London dispersion, but their main strength lies in their combination with the other NCIA data sets. Together with the already available data sets focused on hydrogen bonding,[1,2] repulsive contacts[3] and σ-hole interactions[9] (prepared simultaneously with this work), they provide a complete coverage of the common classes of non-covalent interactions in neutral organic molecules. We plan to extend the database further in the future, namely by covering ionic species including select metal ions important in biochemistry and by complementing the existing data sets with more non-equilibrium geometries.

The increased coverage of the chemical space enabled several interesting applications of the new data sets. Computational methods including empirically parameterized corrections for London dispersion were usually developed on smaller data sets of mostly organic molecules, and it is interesting to examine their performance in a broader chemical space. In this paper, this is applied to dispersion-corrected MP2 and DFT methods. A wide range of dispersion-corrected DFT methods have been tested not only in the newly introduced data sets but also in the other NCIA data sets covering different classes of interactions in order to find the most accurate as well as the most robust ones. The coverage of a broader chemical space combined with the availability of non-equilibrium geometries in the D442×10 data set also allows a deeper analysis of the role of the damping function in DFT-D3. Finally, several dispersion-corrected semiempirical quantum-mechanical (SQM) methods have been tested, although the errors of these methods themselves clearly outweigh the possible error of the dispersion corrections used.

2 Methods

2.1 The Construction of the D1200 Data Set

The data set is constructed by combining monomers featuring specific valence states of the elements studied. A single type of monomer, a free atom, is used for noble gases. For hydrogen and halogens, the data set uses only one class of monomers where the element

forms a single covalent bond. For the remaining elements, separate monomer classes are defined for compounds with the element of interest involved in single, double and triple bonds (where applicable), and for aromatic rings. Additionally, aliphatic hydrocarbons are subdivided into linear, cyclic and double-branched. The list of compounds representing each of these classes is provided in the Supporting Information (SI), Table S1.

Initially, these monomers were combined into dimers, retaining up to five complexes for each combination. This set of 2,175 dimers was subjected to geometry optimizations. Here, some systems had to be removed because of convergence problems, but it was ensured that each combination of monomer classes is still covered sufficiently. Subsequently, hydrogen bonds (with the XH–Y contact shorter than the sum of van der Waals radii and the XHY angle greater than 120°, with X and Y being O, N or S) were removed because they are covered by the existing HB375 and HB300SPX data sets. The same was applied to σ-hole interactions (close contacts between S, P, Cl, Br or I and N, O or F at a distance shorter than the sum of van der Waals radii within 30°of the estimated axis of the σ-hole); these are covered by the dedicated data set SH250.

Next, SAPT0 interaction energy decomposition[10,11] was computed for the remaining dimers, and the ratio of the dispersion energy to the remaining terms except for the first-order exchange repulsion (i.e. electrostatics and induction) was computed. Complexes with less than 40% dispersion were discarded.

This procedure yielded 1,350 complexes, from which the final data set was selected. The aim was to balance the coverage of different groups of elements. The first group is formed by systems comprising H, B, C, N and O elements (the HBCNO group). In the second group, named PS, there are interactions involving also sulfur and phosphorus (i.e. their mutual combinations as well as their interactions with the previous group). The third group (halogens) adds F, Cl, Br and I. Finally, the fourth group (noble gases) adds He, Ne, Ar, Kr and Xe. The HBCNO and PS groups were used without any further reduction, which means that there are up to five complexes representing each combination of monomer classes. For

halogens, each combination of monomer classes was limited to four members, or to three if the other monomer represented carbon. Similarly, in the noble gases group, the number of dimers representing each combination was reduced to three, or to two in complexes with carbon. The discarded systems were chosen randomly. Furthermore, four dimers from the most populated combinations of the halogen group (which is the largest) were removed to round the number of complexes in the data set to 1,200.

This selection led to four groups of approximately the same size, as listed in Table 1. In the groups of halogens and especially of noble gases, a larger number of element combinations is represented by fewer systems, but their diversity is lower, as a result of which no important information is lost (e.g. for a given noble-gas atom, it would probably be redundant to include more complexes with a specific class of hydrocarbons). The counts of systems representing each element combination are listed in SI, Table S2.

Table 1: Groups of the D1200 and D442×10 data sets and their sizes. Each group covers interactions involving the listed elements as well as their interactions with elements covered by all the previous groups.

#	Group	D1200	D442×10
1	HBCNO	308	105
2	PS	293	103
3	Halogens	310	94
4	Noble gases	289	140

2.2 The Construction of the D442×10 Data Set

The dissociation curves are computed only for a subset of the D1200 data set. This has mainly been motivated by the need to reduce the computational demands of the benchmark calculations, but it also keeps the size of the data set of the dissociation curves comparable to the other data sets from the NCIA series. To build this subset, we randomly selected one representative system for each combination of element valence states. Since this procedure yields 442 complexes, the resulting data set of ten-point dissociation curves is named D442×10.

It is divided into the same groups as the D1200 data set, and the numbers of complexes in these groups are listed in Table 1. The sizes of the HBCNO, PS and halogen groups are comparable in about 100 entries; the noble gases group is larger, consisting of 140 items, which is a result of the growing number of possible combinations. The counts of the systems representing each element combination are provided in SI, Table S3.

The D442×10 does not use a separate numbering – the identification of the systems is taken from the D1200 data set. This allows easy assignment of the dissociation curves to the corresponding entries of D1200, and it simplifies application where both data sets are used together.

Smaller predefined subsets are also selected from D442×10 dataset, which makes them applicable to D1200 as well. Like the other NCIA data sets, these subsets are defined by clustering analysis, which maximizes their diversity. As the clustering is based on the errors of all the methods tested in this work, the resulting subsets should capture as many differences between these methods as possible at a given set size. The clustering algorithm used has been described in detail in ref. 1. The subsets (with 20, 50, 100 and 200 systems) are listed in the SI, Table S4, and the assignment of the systems to the subsets is also provided in the data files. The use of a smaller subset from each NCIA data set would yield a general-purpose data set covering multiple classes of non-covalent interactions whose size can be tailored for a specific purpose. The rational selection of the systems ensures the maximum diversity of the set at any given size.

2.3 The Preparation of the Complexes

The non-covalent complexes were built from the monomers automatically, positioning the molecules at a van der Waals distance and orienting them so that their contact surface is maximized. These initial geometries were then optimized at the same level as all the previous NCIA data sets, namely using B3LYP-D3(BJ)/def2-QZVP. The choice of this setup had been justified in the papers on the previous data sets.[1,2] Each structure was verified to be a true

minimum (having real vibrational frequencies); when this condition was been satisfied, the mutual orientation of the molecules was displaced randomly and the procedure was repeated until a minimum was found. An exception were noble-gas dimers, for which experiment-derived van der Waals distances had been used.[12,13]

The quality of the resulting geometries can be tested by comparing the optimized intermolecular distance (defined as the closest contact between the monomers) with its counterpart interpolated from CCSD(T)/CBS dissociation curves. The average absolute difference is only 1.03 %, and the systematic error is even smaller – the average (signed) difference is 0.24 %. Out of the 442 complexes, 23 have an error over 2.5 % and two over 5 %, with the maximum of 7 % in the He–H_2 complex, where the minimum is very shallow.

2.4 Benchmark CCSD(T)/CBS Calculations

The benchmark CCSD(T)/CBS interaction energies in the D1200 and D442×10 data sets are computed using exactly the same setup as in the previous data sets from the NCIA series.[1,2] The CCSD(T)/CBS results are obtained using a composite scheme where MP2 correlation energy is extrapolated to the complete basis set limit from large basis sets,[14] and a CCSD(T) correction (the difference between CCSD(T) and MP2 correlation energies) is added. The Hartree-Fock energy is taken from the largest basis set used in this scheme without extrapolation. All the benchmark calculations also employ the counterpoise correction of the basis set superposition error[15] and frozen-core approximation.

This composite scheme employs the correlation-consistent basis sets from the Dunning family augmented with diffuse functions (aug-cc-pVXZ, with X being D, T, Q or 5).[16] For heavier elements (Br, Kr, I, Xe), effective core potentials (ECPs) and matching variants of the basis sets are used to account for the relativistic effects.[17] In addition, the sub-valence *d* orbitals on these elements are excluded from the frozen-core approximation, as it had been shown that they may significantly contribute to dispersion interactions.[18] This requires a polarized-core variant of the basis sets, specifically from the aug-cc-pwCVXZ-PP series.[19]

The resolution of identity approximation is applied in the calculation of correlation energy using the auxiliary basis sets optimized for the specific atomic orbital basis.[20,21]

The whole D1200 data set as well as the anchor points of D442×10 dissociation curves (the equilibrium and the closest point) are calculated at a "gold level" with MP2 correlation energy extrapolated from aug-cc-pVQZ and aug-cc-pV5Z basis sets and CCSD(T) correction computed in the heavy-aug-cc-pVTZ basis (cc-pVTZ for hydrogen, aug-cc-pVTZ for other elements). It should be noted again that in order to obtain a reliable benchmark, it is necessary to use a triple-ζ basis for the CCSD(T) correction.

The remaining points on the dissociation curves were computed at the "silver level" with MP2 extrapolated from aug-cc-pVTZ and aug-cc-pVQZ basis sets and the CCSD(T) correction calculated in the aug-cc-pVDZ basis. The interaction energies in these points were then rescaled to the gold level as described in ref. 1. This procedure yields only a negligible error, which was verified again on ten randomly chosen systems from the D442×10 data set (complexes no. 1.07.45, 1.11.12, 2.02.04, 2.03.37, 2.04.02, 2.05.13, 3.08.17, 3.11.08, 4.12.09 and 4.23.04), for which all the points were computed at the gold level. The RMSE between the actual gold-level results and the ones obtained by scaling the silver-level interaction energies was 0.003 kcal/mol, with the largest difference being 0.01 kcal/mol. This confirms that for all practical purposes, the D442×10 dissociation curves can be considered of gold-level quality.

The benchmark calculations have been carried out using the Turbomole program package version 7.3 or later.[22,23]

2.5 SAPT0 Interaction Energy Decomposition

To estimate the contribution of London dispersion to the interaction energies of the complexes studied, the approximate symmetry-adapted perturbation theory, SAPT0, was used.[10,11] The calculations were performed in the jun-cc-pVDZ basis,[24] which is recommended for this method as it offers favorable error compensation. This basis set is not available for iodine,

where it was substituted with a similarly-sized def2-TZVP basis coupled with the corresponding pseudopotential. The SAPT0 calculations were carried out in the Psi4 program.[25]

2.6 Density Functional Theory Calculations

Similarly to the previous NCIA data sets, D1200 and D442x10 were used to test a variety of DFT functionals and associated corrections for London dispersion. The DFT methods tested are summarized in Table 2. Stand-alone GGA, meta-GGA and hybrid DFT functional are combined with several dispersion corrections. These include the D3 correction in the original version with zero damping – D3(zero),[26] the revised version with Becke-Johnson damping – D3(BJ),[27] and the later "optimized power" extension D3(OP).[28] The next correction is D4,[29,30] which extends D3 with charge-dependent C_6 coefficients. Another approach tested is the non-local van der Waals correction,[31] applied as a post-SCF calculation to the B3LYP functional.[32] The D3 correction has been added to the common functionals using the Cuby framework[33,34] as it implements the OP damping and other experimental features used in this study. The D4 correction has been computed using the standalone program provided by the authors.[35]

The functionals from the (ω)B97 family include either the non-local (-V suffix) or D3(BJ) dispersion correction. Similarly, some of the double-hybrid functionals include either D3 or D4 dispersion. In these cases, the correction is considered as integral part of the functional, as it has been parameterized together with the rest of the method.

The DFT calculations have been carried out in Orca 4.2[36] using RI approximation. The large def2-QZVP basis set[37] has been used in all DFT calculations in order to minimize the errors caused by the basis set size. It is also the basis set for which the D3 and D4 dispersion corrections have been parameterized, so that their performance should be optimal in this basis.

Table 2: The DFT functionals tested on the presented data sets. The dispersion corrections used are listed separately if they have been developed separately and not parameterized as an integral part of the functional.

Functional	Ref.	Dispersion corrections
GGA functionals		
PBE	38	D3(BJ), D4
BLYP	39,40	D3(BJ), D3(zero), D3(OP), D4
BP	39,41	D3(BJ), D4
Meta-GGA functionals		
B97M-V	42	
B97M-D3(BJ)	43	
SCAN	44	D3(BJ), D4
TPSS	45	D3(BJ), D3(OP), D4
Hybrid functionals		
B3LYP	46,47	D3(BJ), D3(zero), D3(OP), D4, NL
BHLYP	47	D3(BJ), D4
PBE0	48	D3(BJ), D4
M06-2X	49	D3(zero)
TPSSH	45	D3(BJ), D3(OP), D4
Range-separated hybrid functionals		
ωB97M-V	50	
ωB97M-D3(BJ)	43	
ωB97X-V	51	
ωB97X-D3(BJ)	43	
Double-hybrid functionals		
DSD-BLYP	52	
DSD-BLYP-D3	53	
DSD-PBEP86	53	
DSD-PBEP86-D3	53	
revDSD-PBEP86-D3	54	
revDSD-PBEP86-D4	54	

2.7 Semiempirical QM Calculations.

The new data sets have also been used for testing several semiempirical QM methods with either built-in or standalone corrections for London dispersion. From the classic NDDO-based SQM methods, we have selected the PM6 method[55] with D3H4X (dispersion, hydrogen bonding[56] and halogen bonding[57]) corrections and the PM7 method, which already includes analogous corrections.[58]. These calculations have been performed in MOPAC 2016.[59]

The self-consistent-charge density-functional tight binding methodology is represented by DFTB3[60] with D3H4 and D3H5 corrections[61–63] and the 3OB parameter set.[64,65]. All DFTB3 calculations have been carried out in the DFTB+ program, which now implements all of these corrections as well.[66,67] The extended tight binding approach, GFN2-xTB,[68] is a related approach that has been designed with non-covalent interactions in mind (GFN stands for geometries, frequencies and non-covalent interactions) and includes a dispersion correction as well as other features needed for a good description of non-covalent interactions. The GFN2-xTB calculations have been carried out in the standalone code provided by the authors of the method.[69]

2.8 Other Data Sets Used in This Work

The testing of DFT methods presented here also uses four other data sets from the NCIA database. The HB375 data set covers hydrogen bonding in organic molecules as well as other interactions of polar systems that do not form hydrogen bonds.[1] The HB300SPX set features hydrogen bonds involving additional elements, namely halogens, S and P.[2] SH250 is a new data set covering σ-hole interactions – halogen, chalcogen and pnictogen bonds.[9] In this work, only the equilibrium geometries are used, which is denoted by dropping the $\times$10 suffix utilized for the corresponding data sets of ten-point dissociation curves. Finally, the R739 data set of repulsive contacts also employs one geometry per system from the R739$\times$5 set of dissociation curves.[3] Here, we use the first point of the curve that represents a moderately repulsive geometry.

Most of the data have been taken from previous publications, but some new DFT functionals have been calculated in order to provide a complete, up-to-date set. These results will be added to the data files published at the `www.nciatlas.org` website.

3 Data Availability

The D1200 and D442×10 data sets (molecular geometries, benchmark results as well as other results discussed in this paper and additional metadata) are available in the Supporting Information and at the NCIA website `www.nciatlas.org`. At the website, it is also possible to browse all the systems. The whole NCIA database is also available as the GitHub repository `https://github.com/Honza-R/NCIAtlas`.

The minimum package consists of geometries in .xyz format, also including the benchmark interaction energies, the definition of the monomers and selected additional metadata. The benchmark CCSD(T)/CBS results, the components used to construct them, the results of other calculations presented here, and more detailed metadata are available as plain text tables as well. The metadata describe the classification of the systems and their assignment to predefined groups and subsets.

Additionally, all this information is also provided in the form of a machine-readable YAML data file used to automate the calculations in the Cuby framework.[33,34] The data set will also be bundled with a future version of the framework. Cuby can be used to run calculations on the data set with a single input file and to process the results. Since YAML is a structured data format[70] that is human-readable and accessible from all common programming languages, these data can easily be used outside of Cuby as well.

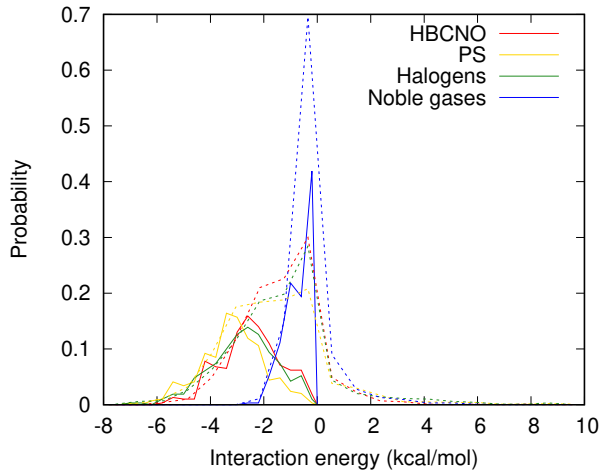
4 Results and Discussion

4.1 Benchmark Calculations

The benchmark CCSD(T)/CBS interaction energies in the D1200 data set range from -7 to nearly 0 kcal/mol. The distribution of the interaction energies in the groups of the data set, plotted in Fig. 1, is very similar in the HBCNO, PS and halogen groups; as expected, only the complexes of noble gases exhibit significantly weaker interactions. Together with similar sizes of all the groups (Table 1), this suggests that the whole data set is well balanced and does not overemphasize any group of elements.

In the D442×10 set, the distribution of the interaction energies, of course, shifts toward weaker interactions in the non-equilibrium geometries. An important feature is that there are only a few points with positive interaction energies, so there is no significant overlap with our R739×5 data set dedicated to repulsive interactions in a similar chemical space.[3]

Figure 1: The distribution of interaction energies in the groups of the D1200 (solid lines) and D442x10 (dashed lines) data sets.



4.2 MP2 and MP2D Calculations

The calculations needed to construct the composite CCSD(T)/CBS benchmark can also be used to look at the performance of MP2 and its various modifications. MP2 is notorious

for overestimating the strength of dispersion interactions. This observation had been rationalized down to the properties of uncoupled Hartree-Fock dispersion,[71] which was also exploited in several attempts to correct this behaviour.[72–75] Another, more empirical approach to correcting the interaction energies is scaling the spin components of the MP2 correlation energy in SCS-MP2 and SCS-MI-MP2 methods.[76,77] The results obtained in the whole D1200 data set are in line with previous findings; the errors of MP2 are rather large, and neither SCS-MP2 nor SCS-MI-MP2 addresses this problem satisfactorily when compared to e.g. dispersion-corrected DFT, which is also less expensive. Therefore, the MP2 results are listed only in the SI in Tables S5 and S6.

An interesting topic is, however, the discussion of MP2 errors in the broader chemical space covered by the D1200 data set. It is well known that MP2 overestimates London dispersion in some systems, and the large discrepancy between the description of saturated hydrocarbons and π – π interactions is a prime example of this effect. The D1200 set includes over 60 hydrocarbon–hydrocarbon complexes on which this issue can be analyzed with more statistical relevance. The results, summarized in Table 3, confirm all previous findings. Complexes of saturated hydrocarbons are described rather well with a relative error of about 5 %, while complexes featuring π–π interactions (i.e. both molecules featuring double or triple C–C bonds or an aromatic ring) have an error of 43 %.

What is less known is how this error behaves in interactions involving other elements. The D1200 set makes it possible to study two series of similar elements with growing proton number, specifically halogens and noble gases (neglecting helium, whose interactions are too weak for reliable analysis). For this purpose, subsets of complexes of hydrocarbons with molecules containing a specific element have been selected, and the errors in these subsets are reported in Table 3. Especially in the complexes of noble gases, the trends in the description of dispersion can be clearly related to the growing polarizability of the atoms as they are not a part of a molecule that can interfere with the studied effect. In both series, the error steadily grows and becomes very large for the heaviest element in the series, i.e. iodine and

xenon. This cannot be attributed merely to the increase in the strength of the interaction because this trend is also exhibited by the relative error (the RMSE divided by the average interaction energy). The relative errors for the heaviest elements, I and Xe, amount to 37 and 46 %, which is comparable to the large error in complexes of unsaturated hydrocarbons (43 %).

This effect can be explained based on interaction energy decomposition. In both series, the SAPT0 dispersion term grows faster than the total interaction energy, but it is compensated for by larger exchange repulsion (the data are provided in the SI, Table S7). As the majority of the error of MP2 calculations comes from the dispersion, it is proportional to the dispersion component rather than to the total interaction energy.

Table 3: The errors of MP2/CBS, MP2D/CBS and SCS-MP2D/CBS interaction energies with respect to the CCSD(T)/CBS benchmark for selected subsets of the D1200 data set. First, we consider the interaction between hydrocarbons, which are divided into saturated and unsaturated (labeled π). Next, we select complexes of hydrocarbons with molecules containing a selected element. (The RMSE is in kcal/mol, the relative error is calculated as the RMSE divided by the average magnitude of interaction energy.)

	MP2/CBS		MP2D/CBS		SCS-MP2D/CBS	
Subset	RMSE	rel. err.	RMSE	rel. err.	RMSE	rel. err.
Hydrocarbons						
sat. – sat.	0.132	5.3 %	0.050	2.0 %	0.027	1.1 %
sat. – π	0.596	25.0 %	0.139	5.8 %	0.104	4.4 %
π – π	1.285	42.6 %	0.207	6.9 %	0.172	5.7 %
Interactions involving a specific element						
F	0.450	20.1 %	0.145	6.5 %	0.169	7.5 %
Cl	0.736	28.5 %	0.224	8.7 %	0.208	8.0 %
Br	0.851	31.7 %	0.269	10.0 %	0.210	7.8 %
I	1.337	37.8 %	—	—	—	—
Ne	0.038	11.6 %	0.019	5.7 %	0.026	8.0 %
Ar	0.231	28.9 %	0.054	6.8 %	0.063	7.8 %
Kr	0.404	35.5 %	—	—	—	—
Xe	0.615	45.9 %	—	—	—	—

It is interesting to investigate how these errors are corrected in the MP2D approach, which replaces the problematic uncoupled dispersion term with the more appropriate coupled dispersion. Both of these terms are calculated empirically (analogously to the correction used

in DFT-D) from a corresponding set of C_6 coefficients (which are, however, not available for all the elements studied here).[74] SCS-MP2D then extends this model by scaling the spin components of the MP2 correlation energy.[75] The results are listed along with uncorrected MP2 in Table 3. Both of these approaches work very well in hydrocarbons, reducing the error in $\pi - \pi$ interactions from 43 to about 6 %. In the interactions of the other elements, the errors are similar in magnitude, around 0.2 kcal/mol, but due to the smaller interaction energies in these subsets, the relative error becomes slightly larger, up to 10 %. It is notable that in MP2D, the errors grow systematically when passing to heavier elements, but the relative error of SCS-MP2D is practically constant. These results show that these methods, parameterized on mostly organic molecules, are robust enough to be applicable to a wider chemical space.

4.3 DFT Calculations in the D1200 and D442×10 Data Sets

The results of DFT calculations in the D1200 data set are considerably different from those in the other NCIA data sets; this comparison is shown below. In dispersion-bound complexes, the intermolecular distances are larger than in e.g. hydrogen or halogen bonds, which means that the DFT methods are tested in a different regime. The description of the contacts at a short distance is more difficult both for DFT itself (where the description of exchange repulsion can be problematic) as well as for the dispersion corrections where the transition region strongly affected by the damping function is being sampled. In the complexes studied here, purposefully excluding any strong short-ranged interactions, we test only the ability of the method to describe "pure" dispersion. The results are thus less dependent on the overall quality of the method and should not be used on their own for ranking DFT functionals.

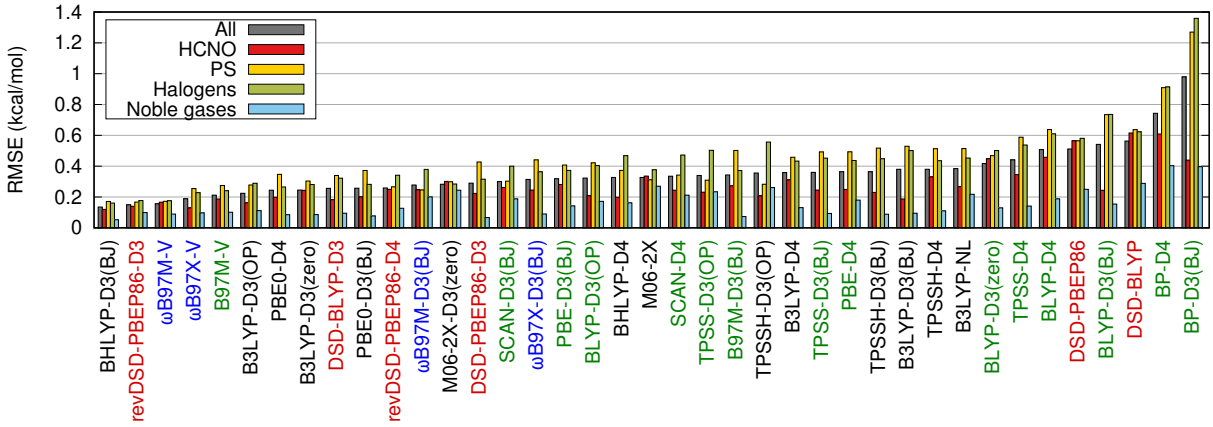
The errors of DFT methods are plotted in Fig. 2; an analogous plot of the systematic error (MSE) as well as tables of these results are provided in the SI (Fig. S1, Tables S8 and S9). The DFT functionals that perform well in other benchmarks are found at the top of the list here as well, but they are accompanied by some functionals that are not suitable

for general use. The best-ranking functional here is BHLYP-D3(BJ), which, despite its high contents of exact exchange, does not work well for e.g. hydrogen bonds. It is followed by revDSD-PBEP86-D3 and several functionals from the (ω)B97M/X family with non-local dispersion correction, and these functionals appear on top of the list in the other data sets as well. The ωB97M-V functional also exhibits a practically zero systematic error not only in the data set as a whole but also in each of its groups separately (see Fig. S1 in the SI). It is interesting that the variants of these (ω)B97M/X functionals with D3 dispersion perform considerably worse here. Another notable feature is the difference between revDSD-PBEP86-D3 and revDSD-PBEP86-D4 (having the RMSE of 0.15 and 0.26 kcal/mol, respectively). Apparently, the results in this data set are very sensitive to subtle details in the dispersion correction and its parameterization to a specific DFT functional. However, it must also be noted that the errors here are usually smaller than in other data sets, and the analysis of such small differences has only a limited practical value.

The methods with larger errors (still around 0.5 kcal/mol) exhibit one clear trend. Here, the systems limited to first-row elements show a significantly smaller error than PS and halogen groups. These are usually the classical DFT functionals with *a posteriori* parameterized dispersion, and it is very likely that the parameterization of the correction is biased toward organic molecules, which comprise the majority of the training set used.

The situation is different in the D442×10 data set of dissociation curves. When we compare the D1200 data set, its subset forming the D442 data set, and the full D442×10 data set, it is clear that the results in equilibrium geometries differ from those obtained on the curves (see Table S10 in the SI). At the level of equilibrium geometries, practically nothing changes when passing from D1200 to the smaller D442 data set. The errors stay practically the same (the average changes by 0.01 kcal/mol) and correlate almost perfectly ($R^2 = 0.999$). However, when we compare D442 and D442×10, the average error increases from 0.34 to 0.47 kcal/mol, and the ordering of the functionals becomes different – the correlation between their errors is now much worse with R^2 of 0.77 and practically random ($R^2 = 0.55$) when

Figure 2: The error (RMSE in kcal/mol) of DFT calculations in the D1200 data set and its groups. The DFT functionals are sorted by ascending RMSE in this data set. The labels of the functionals are colored as follows: double hybrids – red, range-separated hybrids – blue, hybrids – black, GGA and meta-GGA – green.



the two functionals with the largest error are removed from the comparison (these are BP with D3 and D4 corrections).

This increase in the errors clearly comes from the close points of the dissociation curves; the errors at the equilibrium and longer distances are smaller. The interactions at the short distances become repulsive, and it is no surprise that the overall results in the D442×10 correlate better with the errors in the R739 data set of repulsive contacts ($R^2 = 0.75$, with BP-D3 and BP-D4 excluded). The ordering of the functionals also becomes closer to what is observed in the other data sets. The errors of individual DFT methods in the D442×10 data set are available in the SI, Tables S11 and S12.

4.4 Benchmarking DFT across the NCIA Data Sets

As shown above, the DFT results in the D1200 data set must be viewed in the context of other classes of non-covalent interactions covered by the NCIAtlas. For this comparison, we use the HB375 and HB300SPX data sets of hydrogen bonds, the SH250 set of σ-hole interactions, and the R739 set of repulsive contacts. These are equilibrium geometries, with the exception of the R739 data set, which uses the first points of the dissociation curves,

moderately repulsive contacts. Only neutral molecules are considered here, with the IHB100 data set of ionic hydrogen bonds being excluded.

The results are summarized in Fig. 3 and also available as Table S13 in the Supporting Information. From a broader perspective, the performance of the studied DFT methods in the D1200, HB375 and R739 is similar, with errors rarely exceeding 0.5 kcal/mol. The functionals are differentriated by their performance in the other two data sets, HB300SPX (hydrogen bonds including halogens, S and P) and the SH250 data set (σ-hole interactions – halogen, chalcogen and pnictogen bonds). Here, only the modern double-hybrid functionals from the revDSD family and some range-separated ones from the ωB97 family achieve a RMSE smaller than 0.5 kcal/mol in each of these data sets. Another functional reaching this limit is M06-2X, both with and without D3 dispersion correction. All of these functionals involve a number of empirical parameters, and they have been optimized on large databases covering diverse properties in a broader chemical space.

In the remaining cases, the errors can be rather large in these two data sets. This can be attributed to the presence of strong interactions putting heavier atoms in close contact. In the HB300SPX, this applies to complexes of iodine, but the H-bonds involving sulfur and phosphorus are also often described poorly.[2] In the SH250 data set, the largest errors are observed in halogen bonds of chlorine, where especially GGA functionals artificially overestimate charge transfer.[9]

The overall performance of the DFT method can also be summarized as the average RMSE in the five data sets used here. Table 4 lists the functionals with an average error below 0.5 kcal/mol; full listing is available in the SI as Table S14. These tables order the functionals by the average error, but the maximum error (also listed) is equally important as discussed above. However, the few best functionals do well in both of these measures.

To go beyond the issues already described in the papers on the respective data sets, the data can be combined in other interesting ways. The most useful view of the whole database can be obtained when the systems are selected by individual chemical elements.

Figure 3: The errors (RMSE in kcal/mol) of DFT calculations in the D1200, HB375, HB300SPX and SH250 data sets (equilibrium geometries) and in the R739 set (initial close-contact geometries). The DFT functionals are listed in groups with decreasing computational complexity, from double-hybrid methods to pure GGA functionals.

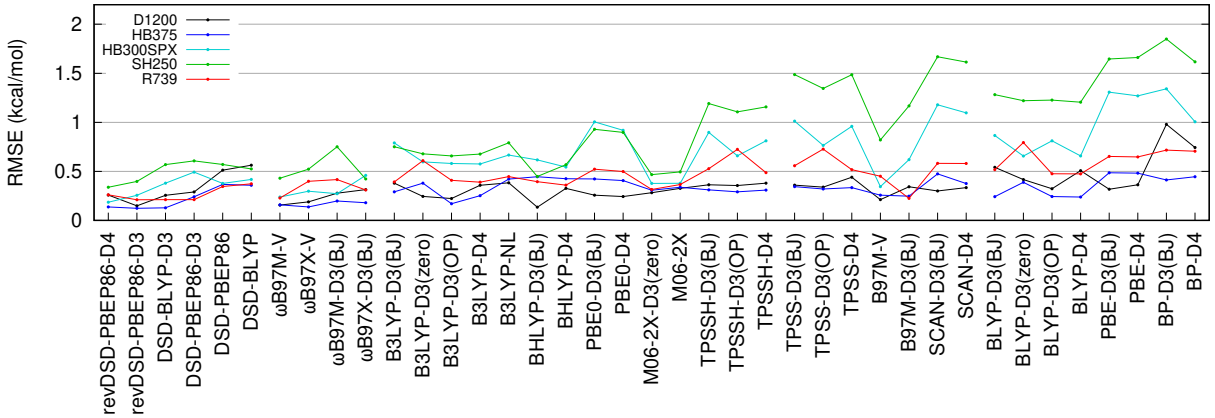


Table 4: The average and maximum errors (RMSE in kcal/mol) of the best-performing DFT methods in the D1200, HB375, HB300SPX, SH250 and R739 data sets. The table is sorted by the average error.

Functional	Avg. err.	Max. err.
revDSD-PBEP86-D3	0.227	0.397
revDSD-PBEP86-D4	0.236	0.338
ωB97M-V	0.243	0.431
DSD-BLYP-D3	0.309	0.569
ωB97X-V	0.309	0.522
ωB97X-D3(BJ)	0.338	0.46
M06-2X-D3(zero)	0.350	0.467
DSD-PBEP86-D3	0.368	0.607
M06-2X	0.381	0.495
ωB97M-D3(BJ)	0.383	0.751
BHLYP-D3(BJ)	0.408	0.617
B3LYP-D3(OP)	0.408	0.657
B97M-V	0.417	0.821
DSD-PBEP86	0.434	0.57
BHLYP-D4	0.445	0.568
DSD-BLYP	0.447	0.563
B3LYP-D4	0.450	0.677

More specifically, this partitioning of the database starts with pure organic molecules (the elements H,C,N,O) and defines further subsets for S, P, F, Cl, Br, I, As, Se, Ne, Ar, Kr and Xe selected as systems including this element and all lighter ones. This analysis excludes B and He, which are represented only by a small number of systems. Again, only equilibrium geometries have been selected from the D1200, HB375, HB300SPX and SH250 data sets along with the initial (moderately repulsive) points from the R739 set. The error in each subset is calculated as an average over the subsets from all the data sets that contain the respective element. A plot of relative errors is shown here as Fig. 4; an analogous plot of the RMSE is provided in the SI as Fig. S2. The errors are also listed there in Tables S16 and S15.

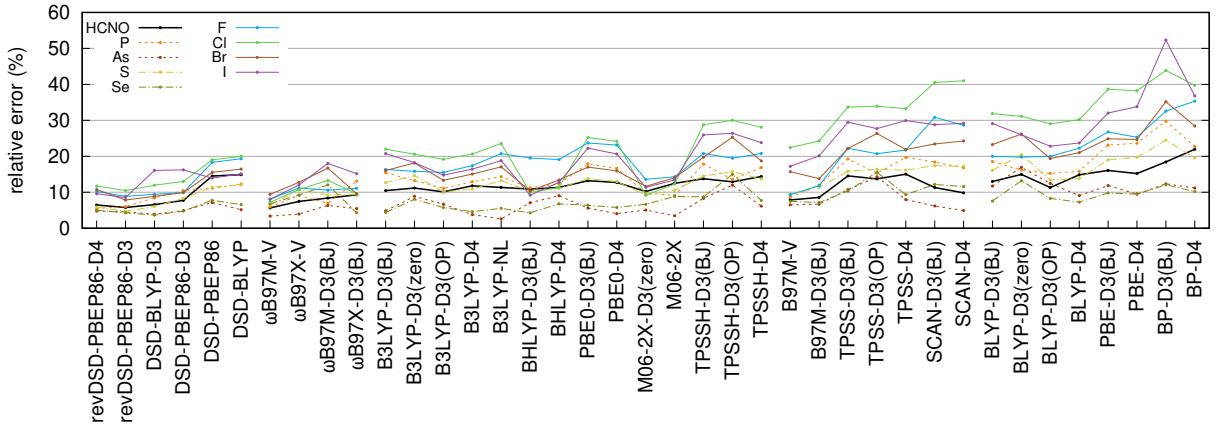
This analysis clearly illustrates the difference between simple organic molecules, where the error stays below 0.5 kcal/mol (or 15 %) for practically all the tested functionals and dispersion corrections, and complexes involving other elements, where the errors can be significantly larger. There are only few DFT functionals that perform consistently well for all the elements. The best is ωB97M-V with relative errors under 10 % and the RMSE under 0.5 kcal/mol for all the element groups. The double-hybrid functionals revDSD-PBEP86-D3 and revDSD-PBEP86-D4 perform only slightly worse. These functionals are also the ones with the highest accuracy in the organic molecules where the errors are as small as 0.16 kca/mol (6 %). A surprise here is the M06-2X functional, both with and without dispersion correction, which also describes all the elements consistently with good accuracy. It has, however, one downside – the error in the HCNO subset is about twice as large (0.3 kcal/mol) compared to the best functionals. The only remaining functional with no error exceeding 15 % is ωB97X-V.

In all other cases, which includes all the traditional DFT functionals with an *a posteriori* added dispersion correction, there is at least one element where the relative error is larger than 15 %, which translates approximately to a RMSE above 0.5 kcal/mol. Most frequently, the problematic elements are halogens, and these errors have already been discussed in the

analyses of the individual data sets.

In the complexes of noble gases, the relative errors can become rather large, but this is mostly caused by the small magnitude of these interactions. These systems are not included in Fig. 4 as it is better to look at the absolute values of the error (Fig. S2), where they rarely exceed 0.5 kcal/mol.

Figure 4: The relative error of DFT calculations involving a specific element across the D1200, HB375, HB300SPX, SH250 and R739 data sets. The relative error (the RMSE divided by the magnitude of interaction energies) averaged over the respective subsets of the datasets containing the element. The DFT functionals are listed in groups with decreasing computational complexity, from double-hybrid methods to pure GGA functionals.



4.5 The Analysis of the Relationships between the NCIA Data Sets

The testing of the same set of methods on multiple data sets yields information not only on the methods themselves – these results can also be used to analyze the relationship between the data sets quantitatively. Again, the five NCIA data sets of neutral complexes featuring different types of non-covalent interactions are considered. The similarity of two data sets is evaluated as a correlation between the errors of the set of 39 DFT methods used here. This measure does not include any direct information on the chemical composition of the data sets and should thus be free of arbitrary presumptions such as the assignment of the non-covalent interactions to named classes. On the other hand, it is possible to show whether

data sets covering seemingly different systems are truly independent or not. This analysis is, of course, fully valid only in the context of the methods tested here, but it can be expected that the findings obtained here can be extrapolated to other computational methods at a similar level of approximation.

The correlation between the DFT errors in each pair of data sets is quantified using the Pearson correlation coefficient R^2, and these values are listed in Table 5. In addition to the correlation between the data sets, we also evaluate the correlation of the errors in one data set with the errors averaged over all the sets, and this criterion is used to sort the sets in the following discussion. The data set that is the most different from the others is D1200 (with the R^2 on average being 0.63). This is likely to reflect the fact that it features weaker interactions at van der Walls distances, while all the other sets include significantly closer contacts, either because of the nature of the interactions or by design. The second most different data set is HB375 – it is distinguished by the limited chemical space that it covers (only H,C,N and O), although it correlates rather well with the HB300SPX set of hydrogen bonds involving other elements. The remaining data sets, R739, HB300SPX and SH250, are more similar to each other and thus also to the average (with the R^2 of 0.84, 0.94 and 0.96, respectively). In addition to having short intermolecular contacts, they also cover a similar chemical space, where the errors are dominated by heavier elements.

4.6 Benchmarking DFT in Representative Subsets of the NCIA Database

The testing of many computational methods on the complete NCIA database, such as the benchmarking of DFT methods presented above, is computationally demanding. When working with robust methods that do not yield unexpected outliers, the database may be reduced to a more manageable size. For each NCIA dataset, we provide predefined representative subsets of several different sizes. These subsets are obtained by a clustering analysis (see the Methods) designed to capture as much diversity of the whole data set as possible.

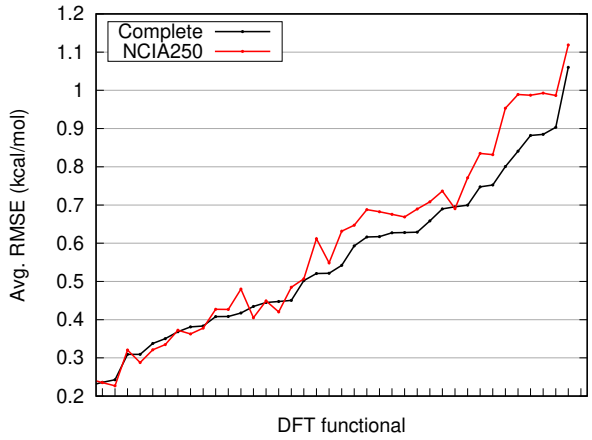
Table 5: The correlation (expressed as R^2) between the errors in the five NCIA data sets, and the error averaged over these data sets (the first part of the table) and the correlation of the errors in the complete data sets and their smaller subsets.

Correlation between data sets						
	D1200	HB375	HB300SPX	SH250	R739	Average
D1200						
HB375	0.35					
HB300SPX	0.46	0.70				
SH250	0.54	0.53	0.89			
R739	0.45	0.62	0.71	0.77		
Average	0.63	0.70	0.94	0.96	0.84	
Correlation with smaller subsets						
	D1200	HB375	HB300SPX	SH250	R739	Average
Subsets 50	0.98	0.93	0.98	0.99	0.97	0.99
Subsets 20	0.94	0.92	0.96	0.99	0.90	0.99

Here, this approach is tested by comparing the DFT results obtained in the D1200, HB375, HB300SPX, SH250 and R739 data sets (2,846 data points) with their equivalents computed in the respective subsets with 50 entries (250 data points). The resulting data set will be labeled NCIA250. For each DFT method, the RMSE is computed in the whole dataset and its subset; these are reported in Table S17 in the SI. The average RMSE in the five complete data sets and the NCIA250 subset is then plotted in Fig. 5. It is clear that all the trends in the ranking of the functionals are reproduced very well even in the smaller-scale calculation. If there is any difference, the error in NCIA250 is higher than in the complete data set – the subsets are built to be as diverse as possible, and they are thus a more difficult test case.

This can also be evaluated quantitatively as the correlation between the errors in the full and reduced data sets. This correlation is very good, with the correlation coefficient R^2 ranging from 0.94 in the HB375 data set to 0.99 in the SH250 data set; for the average RMSE in all five data sets, the R^2 is 0.99 (all the values are listed in Table 5). If we use this ranking to select the best three (all with the RMSE < 0.25 kcal/mol) or the best ten (the RMSE < 0.4 kcal/mol) functionals, this selection will yield the same set in both the full and reduced databases, although the ordering within these groups may differ slightly.

Figure 5: The RMSE of DFT methods average over the five NCIA datasets discussed here evaluated in the complete data sets (black) and their subsets with 50 entries (NCIA250, red). The DFT functionals are sorted by increasing error in the complete calculations; labels are not shown.



In this application, practically equivalent results can be obtained at 10 % of the cost of computing the whole database. This may be generalized to the benchmarking of all less approximate computational methods with a solid physical basis and a limited number of empirical parameters. On the other hand, using the complete database is advisable when working with more empirical methods, such as semiempirical quantum chemistry, where more specific errors can occur. Similarly, the complete database would be preferred for the parameterization of empirical methods whenever it is computationally feasible.

Practically the same correlation is achieved when the database is reduced even more, using subsets with only 20 entries that form the NCIA100 database of only 100 non-covalent complexes (see Table 5, source data in Table S18 in the SI). This, of course, increases the chance that an error specific to some kind of systems would escape unnoticed, but the NCIA100 set is still much more diverse than e.g. the commonly used S66 data set. The correlation of the DFT errors between the NCIA100 database and the S66 data set has the R^2 of 0.38, which can be interpreted as a failure of the S66 set (focused on common interactions in organic molecules) to identify errors occurring in more diverse interactions and a broader chemical space.

4.7 DFT-D3 Damping Function Parameterization

The work on the HB300SPX data set revealed a frequent problem of DFT-D3 with describing hydrogen bonds involving iodine.[2] It was hypothesized that it may be caused by the atomic radii used in the damping function. However, the lack of other complexes involving iodine than hydrogen bonds prevented the making of any strong conclusions. Here, this issue is revisited, parameterizing the damping function on the D442×10 data set or its combination with HB300SPX×10. Again, the B3LYP functional is taken as an example because it can be expected to be rather robust, often performing the best among the classical functionals to which the dispersion correction is added *a posteriori*.

The D3 dispersion correction with the Becke-Johnson damping, D3(BJ), can be rewritten as a function of the interatomic distances R_{AB} and the dispersion coefficients $C_{6,AB}$ and $C_{8,AB}$:

$$E_{disp} = \sum_{A<B} \left(f_d(R_{AB}, 6) \frac{C_{6,AB}}{R_{AB}^6} + s_8 f_d(R_{AB}, 8) \frac{C_{8,AB}}{R_{AB}^8} \right) \tag{1}$$

with damping functions of orders $\beta = 6$ and 8 taking the form

$$f_d(R_{AB}, \beta) = \frac{R_{AB}^\beta}{R_{AB}^\beta + (a_1 R_{0,AB} + a_2)^\beta}. \tag{2}$$

The damping radii $R_{0,AB}$ are pairwise parameters derived from the $C_{6,AB}$ coefficients.[27] This dispersion correction has three empirical parameters, a_1 and a_2, used in the damping function, and the scaling of the 8th order term, s_8. The "optimized power" damping, denoted D3(OP), makes the exponent β in the 6th order term a free parameter and uses $\beta + 2$ in the 8th order.[28]

Three versions of the damping function are considered here: the original D3(BJ), its modification using independently computed damping radii taken from the "zero" damping[26], labeled as D3(BJ'), and the D3(OP) version that offers additional flexibility in the parameter β. These were refitted on either the D442×10 or the D442×10 data set combined with HB375×10 and HB300SPX×10 (where the objective function minimized is a

linear combination of the RMSE in these data sets with the weights of 1.0, 0.5 and 0.5, respectively). The resulting errors in D442×10 and HB300SPX×10 are presented in Table 6 (equilibrium geometries only; the results on the whole dissociation curves are similar – see Table S19 in the SI). The optimized values of the parameters are listed in the SI, Table S20. Note, however, that the reparameterization has been performed to explore the properties of the damping function rather than generally for actual applications.

Table 6: The errors of B3LYP-D3 with different versions of the D3 correction in the D442 and HB300SPX data sets, and in the XH–Br and XH–I groups of the latter.

		RMSE, kcal/mol			
Version	Fitted to	D442	HB300SPX	XH–Br	XH–I
D3(BJ)	original parameters	0.37	0.79	0.47	0.74
	D442×10	0.21	0.85	0.54	0.74
	D442×10 + H-bonds	0.22	0.75	0.45	0.68
D3(BJ')	D442×10	0.20	0.93	0.68	0.87
	D442×10 + H-bonds	0.21	0.80	0.57	0.81
D3(OP)	original parameters	0.22	0.58	0.29	0.52
	D442×10	0.22	0.77	0.48	0.68
	D442×10 + H-bonds	0.23	0.63	0.33	0.53
D3(BJ, scaled C_6)	D442×10	0.15	0.85	0.64	0.76
D3(BJ, scaled R_0)	D442×10	0.15	0.83	0.63	0.74

The results indicate that using an alternative set of damping radii does not bring any improvement, and the hydrogen bonds of iodine are described significantly worse than the ones of bromine, regardless of the version of the D3(BJ) correction. It is also clear that the parameterization on the D442×10 set itself does not yield a transferable correction, and other non-covalent interactions such as hydrogen bonds must be used to balance the narrow focus of D442×10. In the latter case, the resulting D3(BJ) parameters are similar to the original ones. A combination of D442×10 with R739×5 does not make a significant difference, probably because of the redundancy of the repulsive points on the D442×10 curves with the R739×5 set. The D3(OP) approach brings some improvement, both in its original form and when reparameterized to D442×10 and H-bonds, which can be attributed to its additional flexibility.

The next step is the introduction of more empiricism into the damping functions. Both the dispersion coefficients and damping radii come with some uncertainty, and modifying them empirically may improve the accuracy further, but this requires more reference data. The size of the D442×10 data set allows such an experiment. To keep the number of parameters reasonable, elementwise scaling factors s_A and s_B are introduced, and pairwise scaling factors are constructed using combination rules previously utilized for the construction of $C_{6,AB}$ or $R_{0,AB}$ from elementwise parameters. Now the new dispersion coefficients and damping radii are computed as

$$C'_{6,AB} = C_{6,AB} \frac{2 s_A s_B}{s_A + s_B} \tag{3}$$

and

$$R'_{0,AB} = R_{0,AB} \frac{s_A + s_B}{2}. \tag{4}$$

The scaling factors s are optimized along with the three global parameters in the dispersion correction (with the exception of carbon in the case of radii scaling because there is a redundancy with the global parameter a_1). The results (the last two rows in Table 6) show that the scaling of either the C_6 coefficients or damping radii leads to practically identical results. In both cases, the error in the D442 data set drops to 0.15 kcal/mol. However, the description of the hydrogen bonds in the HB300SPX set remains rather poor. If this error were caused by the values of the C_6 coefficients or by the damping function, it would be reduced with the introduction of the elementwise parameters. As this is not the case, the only remaining explanation is that the large errors in e.g. the XH–I halogen bonds come from the DFT calculation itself, likely related to overdelocalization of the electron density in the complex.

The last question to answer is the source of the error removed by the empirical scaling of the C_6 coefficients or damping radii. Both approaches yield seemingly equivalent results, but a look at the scaling factors (Table S21 in the SI) shows that they are complementary –

the same effect is achieved by reducing the C_6 coefficient or increasing the damping radius and vice versa. This can be determined by looking only at the most distant points of the D442×10 dissociation curves – here the damping function should have no effect, and the correction depends solely on the C_6 coefficients. The scaling of the C_6 coefficients increases the error there, so it can be concluded that the scaling of the radii is the physically more sound solution. This suggests that the next step in the possible development of the dispersion corrections would be the refinement of these radii.

4.8 Semiempirical QM calculations

Only few semiempirical methods are applicable to all the elements included in the D1200 data set. Therefore, a subset of 778 systems, excluding boron and noble gases, is used here instead. This has made it possible to add DFTB3 calculations (DFTB3 does not have parameters for boron and noble gases) to the generally applicable methods PM6, PM7 and GFN2-xTB. For PM6 and DFTB3, several corrections for non-covalent interactions have been applied: D3 dispersion correction, H4 and H5 hydrogen-bonding corrections and the X correction for halogen bonds. The H and X corrections are not needed in this data set, but it is useful to determine whether they affect the results at all.

The resulting errors (RMSE and MSE) are plotted in Figures 6 and 7. The actual values of the errors are listed in the SI, Tables S22 and S23. In the simple organic molecules comprising no more than H, C, N and O atoms, all the methods with a dispersion correction yield a satisfactory RMSE of about 0.5 kcal/mol. The systematic part of the error, MSE, is usually also small. It is the largest in GFN2-xTB which is otherwise a rather accurate and robust method. The large systematic error indicates that there may be a room for further improvement in this direction.

The additional corrections beyond dispersion have only a negligible effect here because the D1200 data set has been designed to exclude hydrogen- and halogen-bonded complexes. These results confirm that these corrections work as expected, acting only in the systems

for which they have been designed. The largest residual effect has been observed in the X correction for DFTB3, which leads to a slight increase in the RMSE, although it notably reduces the systematic part of the error.

Figure 6: The errors of semiempirical QM methods in the subset of the D1200 data set to which they are applicable. RMSE in kcal/mol.

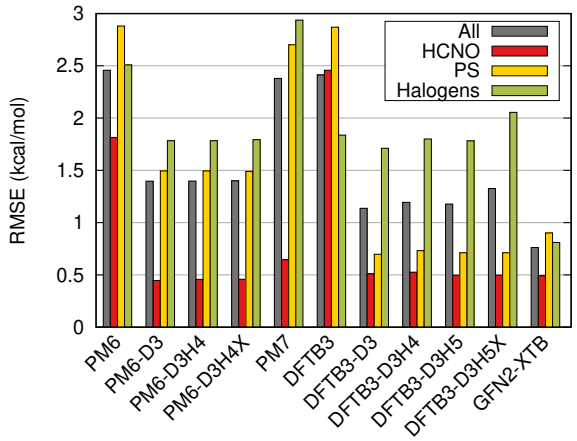
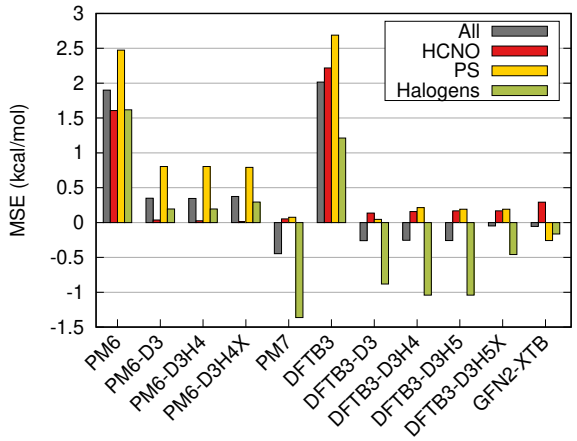


Figure 7: The systematic errors of semiempirical QM methods in the subset of the D1200 data set to which they are applicable. MSE in kcal/mol.



Moving to other elements, the situation becomes worse. PM6 and PM7 have problems already in the complexes with S and P, and these methods along with DFTB3 also fail to describe halogens accurately. Only GFN2-xTB yields the RMSE below 1 kcal/mol in all the groups. It is clear that this lack of robustness originates from the methods themselves, not from the dispersion correction used. Moreover, the dispersion interactions studied here are

rather weak and do not involve any large changes in the electronic structure of the molecules. It is thus likely that these problems stem from the deficiencies in the parameterization of these methods, rather than from the limitations of the semiempirical formalism used. Also the results of GFN2-xTB suggest that the accuracy of 1 kcal/mol can be achieved in dispersion-bound complexes in a broader chemical space.

5 Conclusions

The D1200 and D442×10 data sets represent a substantial step in the development of the Non-Covalent Interactions Atlas database. Together with the other NCIA data sets, they complete the coverage of the most common classes of non-covalent interactions in neutral systems. Their size, diversity and coverage of chemical space significantly exceed the previous state of the art. All the data, including not only the geometries and benchmark results but also the results of our calculations on these data sets, are openly available, both here and at the `www.nciatlas.org` website.

The enhanced diversity of the D1200 data set has been exploited in the analysis of the errors of MP2. Besides the well-known failure in π–π interactions of hydrocarbons, the new data show a clear trend of increasing error when passing to heavier elements. This can be explained using the SAPT interaction energy decomposition, where the contribution of London dispersion grows faster than the final interaction energy in the analyzed series. The MP2D and SCS-MP2D methods, which replace the problematic dispersion term with a correct one calculated in the spirit of DFT-D, work rather well in correcting all these errors despite being parameterized for a narrower chemical space.

The D1200 data set also highlights very specific features of dispersion-corrected DFT methods because it isolates the dispersion well from other interactions. Along with the D442×10 data set, it would be very useful for the development of dispersion corrections; for the general testing of DFT methods, however, it should be used together with other

data sets. In this paper, we analyze the performance of DFT methods across five datasets from the NCIA database, covering a wide range of types of non-covalent interactions. The analysis has made it possible to identify not only the functionals that are the most accurate for a specific problem, but, more importantly, those that are the most robust and have the highest chance of success in more difficult systems. The highest-ranking functionals are revDSD-PBEP86-D4, revDSD-PBEP86-D3, ωB97M-V, ωB97X-D3(BJ), M06-2X-D3(zero) and M06-2X. All the GGA or meta-GGA functionals have more or less serious problems in strong interactions of heavier elements in either the HB300SPX or the SH250 data set.

A comparison of the DFT results in the different NCIA data sets has also enabled the quantification of the relationship between them. The D1200 data set has been found to be the most dissimilar to the rest of the database, likely because it covers only weaker interactions and avoids any close contacts where the errors of DFT become larger and more diverse. It has also been shown that the smaller subsets of all of these data sets, built using a clustering analysis, represent the complete data sets well and together they form a smaller but still highly diverse test set useful for applications. The NCIA100 and NCIA250 databases, extracted from the five NCIA data sets of neutral systems, provide very diverse test sets, which reduce the cost of computing the whole database in many applications.

Several experiments with the reparameterization of the damping function in the D3 correction have indicated that in the case of standalone DFT functionals, such as B3LYP, used as an example here, the overall error in the NCIA data sets is dominated by the DFT part of the calculation itself. In the correction, a small improvement can be achieved when a more flexible damping function is used. It is possible to achieve further gain in accuracy by refining the atomic radii used in the damping function. However, these effects are rather small and may be practically important only in the development of more accurate DFT functionals including the dispersion correction.

Several semiempirical QM methods with dispersion corrections have also been tested on the D1200 data set. All of them describe the HCNO chemical space with good accuracy, but

they often fail in systems with additional elements. These problems are caused neither by the description of the London dispersion itself nor by the semiempirical approximations used, but most likely by issues in the parameterization of the methods. The only SQM method to yield errors below 1 kcal/mol in the whole data set is the empirical tight binding method GFN2-xTB.

This paper introduces the new D1200 and D442×10 data sets and demonstrates their use in benchmarking various approximate computational methods. Nevertheless, the data sets have mainly been designed for the development of future methods that could overcome at least some of the issues discussed here. These data could help especially in making semiempirical QM methods more accurate in a broader chemical space.

6 Acknowledgements

The author acknowledges the support from the Czech Science Foundation, Grant No. 19-13905S, and from the European Regional Development Fund, OP RDE, Project: Chemical Biology for Drugging Undruggable Targets (Chem-BioDrug, No. CZ.02.1.01/0.0/0.0/16_019/0000729). This work has also been supported by the Ministry of Education, Youth and Sports from the Large Infrastructures for Research, Experimental Development and Innovations project IT4Innovations National Supercomputing Center, LM2015070.

Supporting Information Available

The Supporting Information contains: 1) additional tables and figures referenced in the main text, including tables of the data used in the plots featured here, 2) geometries of all the systems and tables of benchmark interaction energies needed for the reproduction of the results presented here, and 3) a machine-readable definition of the data set containing the metadata describing the categorization of the systems.

References

(1) Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets: Hydrogen Bonding. *J. Chem. Theory Comput.* **2020**, *16*, 2355–2368.

(2) Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets 2: Hydrogen Bonding in an Extended Chemical Space. *J. Chem. Theory Comput.* **2020**, *16*, 6305–6316.

(3) Kříž, K.; Nováček, M.; Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets 3: Repulsive Contacts. *J. Chem. Theory Comput.* **2021**, *17*, 1548–1561.

(4) Sparrow, Z. M.; Ernst, B. G.; Joo, P. T.; Lao, K. U.; DiStasio, R. A. NENCI-2021. I. A large benchmark database of non-equilibrium non-covalent interactions emphasizing close intermolecular contacts. *J. Chem. Phys.* **2021**, *155*, 184303.

(5) 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.

(6) 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.

(7) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardere, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 2609–2618.

(8) Donchev, A. G.; Taube, A. G.; Decolvenaere, E.; Hargus, C.; McGibbon, R. T.; Law, K.-H.; Gregersen, B. A.; Li, J.-L.; Palmo, K.; Siva, K.; Bergdorf, M.; Klepeis, J. L.; Shaw, D. E. Quantum chemical benchmark databases of gold-standard dimer interaction energies. *Sci Data* **2021**, *8*, 55.

(9) Kříž, K.; Řezáč, J. *Manuscript in preparation*

(10) Hohenstein, E. G.; Sherrill, C. D. Density fitting of intramonomer correlation effects in symmetry-adapted perturbation theory. *J. Chem. Phys.* **2010**, *133*, 014101.

(11) Hohenstein, E. G.; Sherrill, C. D. Wavefunction methods for noncovalent interactions. *WIREs Comput. Mol. Sci.* **2012**, *2*, 304–326.

(12) Ogilvie, J. F.; Wang, F. Y. H. Potential-energy functions of diatomic molecules of the noble gases I. Like nuclear species. *Journal of Molecular Structure* **1992**, *273*, 277–290.

(13) Ogilvie, J. F.; Wang, F. Y. H. Potential-energy functions of diatomic molecules of the noble gases: II. Unlike nuclear species. *Journal of Molecular Structure* **1993**, *291*, 313–322.

(14) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence of correlated calculations on water. *The Journal of Chemical Physics* **1997**, *106*, 9639–9646.

(15) Boys, S.; Bernardi, F. Calculation of Small Molecular Interactions by Differences of Separate Total Energies - Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, *19*, 553–566.

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

(17) Peterson, K. A.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. Systematically convergent basis sets with relativistic pseudopotentials. II. Small-core pseudopotentials and correlation consistent basis sets for the post-d group 16-18 elements. *J. Chem. Phys.* **2003**, *119*, 11113–11123.

(18) Kesharwani, M. K.; Manna, D.; Sylvetsky, N.; Martin, J. M. L. The X40x10 Halogen

Bonding Benchmark Revisited: Surprising Importance of (n–1)d Subvalence Correlation. *J. Phys. Chem. A* **2018**, *122*, 2184–2197.

(19) Peterson, K. A.; Dunning, T. H. Accurate correlation consistent basis sets for molecular core-valence correlation effects: The second row atoms Al-Ar, and the first row atoms B-Ne revisited. *J. Chem. Phys.* **2002**, *117*, 10548–10560.

(20) Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *The Journal of Chemical Physics* **2002**, *116*, 3175–3183.

(21) Hill, J. G. Auxiliary basis sets for density fitting second-order Møller-Plesset perturbation theory: Weighted core-valence correlation consistent basis sets for the 4d elements Y–Pd. *Journal of Computational Chemistry* **2013**, *34*, 2168–2177.

(22) TURBOMOLE v7.3. 2018; `http://www.turbomole.com`.

(23) Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. Turbomole. *WIREs Comput. Mol. Sci.* **2014**, *4*, 91–100.

(24) 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.

(25) Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D. Psi4: an open-source ab initio electronic structure program. *WIREs Comput Mol Sci* **2012**, *2*, 556–565.

(26) 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.

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

(28) Witte, J.; Mardirossian, N.; Neaton, J. B.; Head-Gordon, M. Assessing DFT-D3 Damping Functions Across Widely Used Density Functionals: Can We Do Better? *J. Chem. Theory Comput.* **2017**, *13*, 2043–2052.

(29) Caldeweyher, E.; Bannwarth, C.; Grimme, S. Extension of the D3 dispersion coefficient model. *The Journal of Chemical Physics* **2017**, *147*, 034112.

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

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

(32) 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.

(33) Řezáč, J. Cuby: An integrative framework for computational chemistry. *J. Comput. Chem.* **2016**, *37*, 1230–1237.

(34) Řezáč, J. Cuby 4, software framework for computational chemistry. 2015; `http://cuby4.molecular.cz/`.

(35) DFTD4 program. 2019; `https://github.com/dftd4/dftd4`.

(36) Neese, F. Software update: the ORCA program system, version 4.0. *WIREs Computational Molecular Science* **2018**, *8*, e1327.

(37) 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.

(38) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(39) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(40) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.

(41) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33*, 8822–8824.

(42) 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.

(43) Najibi, A.; Goerigk, L. The Nonlocal Kernel in van der Waals Density Functionals as an Additive Correction: An Extensive Analysis with Special Emphasis on the B97M-V and ωB97M-V Approaches. *J. Chem. Theory Comput.* **2018**, *14*, 5725–5738.

(44) Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **2015**, *115*, 036402.

(45) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.

(46) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(47) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

(48) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

(49) Zhao, Y.; Truhlar, D. G. Hybrid meta density functional theory methods for thermochemistry, thermochemical kinetics, and noncovalent interactions: The MPW1B95 and MPWB1K models and comparative assessments for hydrogen bonding and van der Waals interactions. *J. Phys. Chem. A* **2004**, *108*, 6908–6918.

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

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

(52) Kozuch, S.; Gruzman, D.; Martin, J. M. L. DSD-BLYP: A General Purpose Double Hybrid Density Functional Including Spin Component Scaling and Dispersion Correction. *J. Phys. Chem. C* **2010**, *114*, 20801–20808.

(53) Kozuch, S.; Martin, J. M. L. DSD-PBEP86: in search of the best double-hybrid DFT with spin-component scaled MP2 and dispersion corrections. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20104–20107.

(54) Santra, G.; Sylvetsky, N.; Martin, J. M. L. Minimally Empirical Double-Hybrid Functionals Trained against the GMTKN55 Database: revDSD-PBEP86-D4, revDOD-PBE-D4, and DOD-SCAN-D4. *J. Phys. Chem. A* **2019**, *123*, 5129–5143.

(55) Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J Mol Model* **2007**, *13*, 1173–1213.

(56) Řezáč, J.; Hobza, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *J. Chem. Theory Comput.* **2012**, *8*, 141–151.

(57) Řezáč, J.; Hobza, P. A halogen-bonding correction for the semiempirical PM6 method. *Chem. Phys. Lett.* **2011**, *506*, 286–289.

(58) Stewart, J. J. P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J Mol Model* **2013**, *19*, 1–32.

(59) Stewart, J. J. P. MOPAC 2016. 2016; `http://openmopac.net/`.

(60) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7*, 931–948.

(61) Miriyala, V. M.; Řezáč, J. Description of non-covalent interactions in SCC-DFTB methods. *J. Comput. Chem.* **2017**, *38*, 688–697.

(62) Grimme, S. Towards First Principles Calculation of Electron Impact Mass Spectra of Molecules. *Angew. Chem. Int. Ed.* **2013**, *52*, 6306–6312.

(63) Řezáč, J. Empirical Self-Consistent Correction for the Description of Hydrogen Bonds in DFTB3. *J. Chem. Theory Comput.* **2017**, *13*, 4804–4817.

(64) Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2013**, *9*, 338–354.

(65) Kubillus, M.; Kubař, T.; Gaus, M.; Řezáč, J.; Elstner, M. Parameterization of the DFTB3 Method for Br, Ca, Cl, F, I, K, and Na in Organic and Biological Systems. *J. Chem. Theory Comput.* **2015**, *11*, 332–342.

(66) Hourahine, B.; Aradi, B.; Blum, V.; Bonafé, F.; Buccheri, A.; Camacho, C.; Cevallos, C.; Deshaye, M. Y.; Dumitrică, T.; Dominguez, A.; Ehlert, S.; Elstner, M.; van der Heide, T.; Hermann, J.; Irle, S.; Kranz, J. J.; Köhler, C.; Kowalczyk, T.; Kubař, T.; Lee, I. S.; Lutsker, V.; Maurer, R. J.; Min, S. K.; Mitchell, I.; Negre, C.; Niehaus, T. A.; Niklasson, A. M. N.; Page, A. J.; Pecchia, A.; Penazzi, G.; Persson, M. P.; Řezáč, J.; Sánchez, C. G.; Sternberg, M.; Stöhr, M.; Stuckenberg, F.; Tkatchenko, A.; Yu, V. W.-z.; Frauenheim, T. DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *J. Chem. Phys.* **2020**, *152*, 124101.

(67) DFTB+, general package for performing fast atomistic simulations. `http://www.dftb-plus.info/`.

(68) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.

(69) XTB, Semiempirical Extended Tight-Binding Program Package. 2019; `https://github.com/grimme-lab/xtb`.

(70) The Official YAML Web Site. `http://yaml.org/`.

(71) Cybulski, S. M.; Lytle, M. L. The origin of deficiency of the supermolecule second-order Moller-Plesset approach for evaluating interaction energies. *J. Chem. Phys.* **2007**, *127*, 141102.

(72) Heßelmann, A. Improved supermolecular second order Møller–Plesset intermolecular

interaction energies using time-dependent density functional response theory. *J. Chem. Phys.* **2008**, *128*, 144112.

(73) Pitoňák, M.; Heßelmann, A. Accurate Intermolecular Interaction Energies from a Combination of MP2 and TDDFT Response Theory. *J. Chem. Theory Comput.* **2010**, *6*, 168–178.

(74) Řezáč, J.; Greenwell, C.; Beran, G. J. O. Accurate Noncovalent Interactions via Dispersion-Corrected Second-Order Møller–Plesset Perturbation Theory. *J. Chem. Theory Comput.* **2018**, *14*, 4711–4721.

(75) Greenwell, C.; Řezáč, J., Jan; Beran, G. Spin-component-scaled and dispersion-corrected second-order Møller-Plesset perturbation theory: A path toward chemical accuracy. **2021**, Preprint, DOI: 10.33774/chemrxiv-2021-jdv3c.

(76) 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.

(77) DiStasio, R.; Head-Gordon, M. Optimized spin-component scaled second-order Moller-Plesset perturbation theory for intermolecular interaction energies. *Mol. Phys.* **2007**, *105*, 1073–1083.