

# Non-Covalent Interactions Atlas Benchmark

## Data Sets 4: $\sigma$ -Hole Interactions

Kristian Kříž and Jan Řezáč\*

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

\*e-mail: rezac@uochb.cas.cz

February 21, 2022

### Abstract

The SH250 $\times$ 10 dataset presented here extends the Non-Covalent Interactions Atlas database ([www.nciatlas.org](http://www.nciatlas.org)) to complexes bound by  $\sigma$ -hole interactions – halogen, chalcogen and pnictogen bonds. It comprises 250 complexes where Cl, Br, I, S, Se, P and As interact with diverse electron donors. An accurate CCSD(T)/CBS benchmark is provided for ten points along a dissociation curve of each complex. The SH250 $\times$ 10 set is used in testing a wide variety of DFT functionals and semiempirical quantum-mechanical methods. In DFT calculations, the new data set exposes large errors of some functionals related to exaggerated charge transfer. The size and diversity of the data set have also been exploited in the reparametrization of a halogen-bond correction for the PM6 semiempirical method.

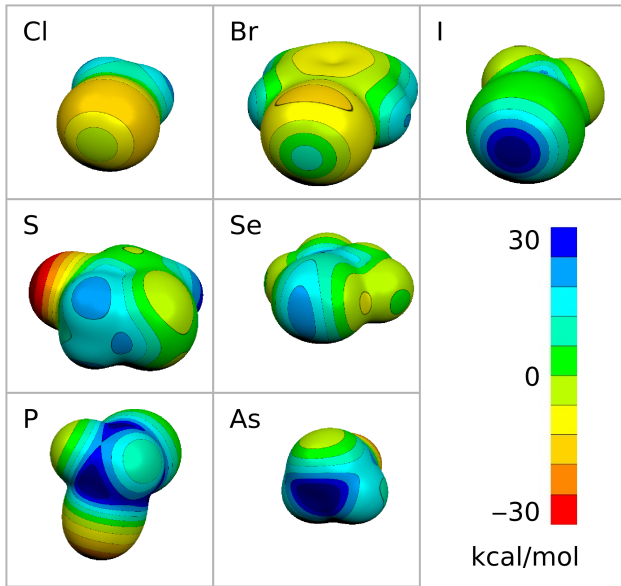
## 1 Introduction

The Non-Covalent Interactions Atlas (NCIA, [www.nciatlas.org](http://www.nciatlas.org)) is a collection of data sets of accurate benchmark interaction energies in non-covalent complexes. It aims to map

different classes of non-covalent interactions in a chemical space larger than previously considered. The database currently covers hydrogen bonds,<sup>1,2</sup> repulsive contacts<sup>3</sup> and London dispersion<sup>4</sup> with more than 16,000 data points.

In this work, we present the SH250 $\times$ 10 data set, extending the NCIA database to  $\sigma$ -hole interactions – halogen, chalcogen and pnictogen bonds featuring Cl, Br, I, S, Se, P and As. These specific non-covalent interactions result from the existence of a  $\sigma$ -hole, a region of positive electrostatic potential on an otherwise electronegative atom.<sup>5</sup> The  $\sigma$ -hole is typically located in the extension of a covalent  $\sigma$  bond to another atom, which leads to a simple linear arrangement in halogens but also to more complex electrostatic potentials in polyvalent atoms (for an illustration, see Fig. 1). The positive  $\sigma$ -hole can interact with negatively charged sites (electron donors) on other molecules, forming a  $\sigma$ -hole bond.

Figure 1: Electrostatic potentials of representative  $\sigma$ -hole featuring molecules selected from the SH250 data set. These are chloromethane (Cl), bromobenzene (Br), trifluoriodomethane (I), thiophene-3(2H)-one (S), tetrafluoroselenourea (Se), fluorobromocyanophosphine (P) and fluoroarsine (As). The potentials are constructed at the  $0.001 e \text{ Bohr}^{-3}$  density iso-surface.



The family of  $\sigma$ -hole interactions is large and has grown to cover more and more elements since the discovery of the phenomena. Halogens, chalcogens, pnictogens, tetrels and even

noble gases have been shown to feature  $\sigma$ -holes under some conditions.<sup>6,7</sup> The first three groups are most prominent, forming strong  $\sigma$ -hole interactions found in diverse chemical systems of practical importance, such as materials or biomolecules. The significance of halogen and chalcogen bonding in biological systems has been investigated earlier.<sup>8-11</sup>

The  $\sigma$ -hole interactions have several unique properties that distinguish them from other classes of non-covalent interactions. They form very close intermolecular contacts, often substantially shorter than the sum of the van der Waals radii of the participating atoms.<sup>12,13</sup> This makes other contributions to the interaction such as London dispersion or exchange repulsion very strong as well.<sup>13</sup> Since they originate from the local inhomogeneity of electron density, they are sensitive to the orientation of the molecules. In addition, the  $\sigma$ -hole interactions are accompanied by a varying degree of charge transfer, whose role is a subject of an ongoing debate.<sup>13-17</sup> These properties make  $\sigma$ -hole interactions challenging for approximate computational methods, and their testing and development thus require quality benchmark data.

Several data sets covering  $\sigma$ -hole interactions have been introduced previously. The X40 $\times$ 10 data set covers halogen bonds as well as other interactions of halogens with 40 dissociation curves. Two datasets, XB18 and XB51 by Kozuch et al.,<sup>18</sup> cover the geometries and interaction energies of halogen-bonded complexes. Another 30 benchmark complexes, also including chalcogen and pnictogen bonds, were prepared by Bauzá et al. for their analysis of DFT functional performance in  $\sigma$ -hole complexes.<sup>19</sup>

The SH250 $\times$ 10 data set presented here aims to unite the coverage of these interactions with a more comprehensive benchmark. It comprises 250 complexes of molecules featuring  $\sigma$ -holes interacting with various electron donors. For each complex, the interaction energy is computed at ten distances along the dissociation curve, which results in 2,500 benchmark data points. We consider the halogen bonds of chlorine, bromine and iodine, the chalcogen bonds of sulfur and selenium, and the pnictogen bonds of phosphorus and arsenic. These elements form  $\sigma$ -hole interactions strong enough to define the geometry of the complex.

The first-row elements from these groups (oxygen, nitrogen and fluorine) are weak  $\sigma$ -hole binders<sup>5</sup> and usually prefer other modes of interactions. On the other hand, the elements heavier than arsenic, selenium or iodine are of little practical interest.

Other classes of interactions involving the elements defining the SH250 $\times$ 10 data set (except for Se and As) are now covered by the other NCIA data sets spanning a similar chemical space – hydrogen bonds by the HB300SPX $\times$ 10 set,<sup>2</sup> repulsive contacts by the R739 $\times$ 5 set<sup>3</sup> and London dispersion by the D1200 and D442 $\times$ 10 data sets.<sup>4</sup> Since these data sets have been constructed to exclude  $\sigma$ -hole interactions, they provide information complementary to SH250 $\times$ 10. This is important e.g. in method development – it should target not only a single prominent class of interactions but any general mode of interaction in which an element may be involved. This area has previously been sampled only sparsely, for example by some systems in the X40 $\times$ 10 data set.

The benchmark interaction energies are calculated using the composite CCSD(T)/CBS scheme based on MP2 and CCSD(T) calculations in very large basis sets. The specific requirements of the heavier elements are handled appropriately. Together, this computational setup ensures a true “gold standard” benchmark as accurate as possible in a data set of this size. Analogously to the other NCIA data sets, the benchmark data are freely available in an easy-to-use form.

We use the SH250 $\times$ 10 data set to benchmark a wide range of density functional theory (DFT) functionals and semiempirical quantum mechanical (SQM) methods applicable to this chemical space. The results show that this data set is a rather challenging test. In DFT, it reveals some interesting errors not encountered in other data sets. Specifically, in some complexes with strong charge transfer, the strength of the interaction may be overestimated as much as twofold. The size of the SH250 $\times$ 10 set also enables a the reparametrization of the halogen bond correction for PM6, which should improve its robustness.<sup>20,21</sup>

## 2 Methods

### 2.1 The Construction of the SH250×10 Data Set

**The composition of the data set.** For each element considered, representative compounds featuring a  $\sigma$ -hole have been selected to cover a range of  $\sigma$ -hole magnitudes. This has been achieved by varying the substituents on the atom. For light chalcogens and pnictogens, strong electron-withdrawing substituents are needed to form a  $\sigma$ -hole that could compete with other modes of interaction. These compounds are listed in Table S1 in the Supporting Information (SI).

The dataset is built by combining these compounds with electron donors interacting via nitrogen, oxygen or sulfur, a halogen atom (dihalogen bonds<sup>13,22</sup>), or an aromatic ring forming a  $\sigma$ -hole –  $\pi$  interaction. These are listed in Table S2 in the SI. All the monomers used to construct the complexes are closed-shell, neutral molecules up to 13 atoms large.

**The preparation of the complexes.** The monomers described above were combined into 168 halogen bond complexes, 119 chalcogen- and 70 pnictogen-bond complexes – 357 in total. The initial geometry of the complex was determined by the position of the  $\sigma$ -hole and the electron-donor group in the other molecule, and the intermolecular distance was set to be equal to the sum of the van der Waals radii of the interacting atoms.

These structures were refined in two steps of DFT-D geometry optimization. The initial optimization was performed in a smaller basis at the BLYP-D3(BJ)/DZVP-DFT level.<sup>23</sup> If the  $\sigma$ -hole interaction motif was conserved, the geometry was refined using a larger basis in B3LYP-D3(BJ)/def2-QZVP calculation; a setup identical to the one used in all the previous NCIA data sets. Finally, a vibrational analysis was performed. If imaginary frequencies were found, the geometry was perturbed and the last step of the optimization was repeated until a true minimum was found.

The initial optimization was performed in Turbomole 7.3.<sup>24</sup> The final step and the vi-

brational analysis were done in Gaussian 09,<sup>25</sup> using tight optimization criteria and the ultra-fine grid for the numerical integration.

From the resulting geometries, only those featuring a well-defined  $\sigma$ -hole interaction were kept. The following criteria were used: 1) the distance between the atom with the  $\sigma$ -hole and the electron-donor site had to be shorter than the sum of the respective van der Waals radii, 2) the interacting sites had to represent the closest contact between the molecules (evaluated in percentage of the corresponding van der Waals radius sums). Finally, several systems from the most abundant element combinations were removed to round the size of the data set to 250  $\sigma$ -hole complexes.

**Dissociation curves.** For each of the 250 complexes (the SH250 data set), ten-point dissociation curves (the SH250 $\times$ 10 set) have been constructed by scaling the closest intermolecular distances (which corresponds to the  $\sigma$ -hole interaction) by a factor of 0.8, 0.85, 0.9, 0.95, 1.05, 1.10, 1.25, 1.50 and 2.00. The same scaling procedure is also used in the previous NCIA data sets. The molecules are displaced along the vector defined by the interacting atoms that is the most appropriate coordinate for studying the distance dependence of the  $\sigma$ -hole interaction.

For the verification of the DFT-D3-optimized geometries of the SH250 data set, each actual intermolecular distance has been compared with a minimum interpolated from the dissociation curve computed at the benchmark level. In this data set, the average absolute difference is 1.3% and the largest one is 4.3%, which is a rather good result; there were larger differences observed in several other NCIA data sets.

An overestimation of the interaction energy in some complexes of molecular chlorine by several DFT functionals including B3LYP-D3(BJ) (see Section 3.2) raised a concern about the legitimacy of the geometries optimized at this level. In a model complex where this error was large (trimethylamine – chlorine), we performed an alternative optimization with the same protocol but using the M06-2X functional, which does not exhibit this overbinding. The

difference between B3LYP-D3(BJ) and M06-2X geometries is minimal – the N–Cl contact distance differs only by 0.04 Å (1.2%), which is in line with the validation of the geometries by interpolating the CCSD(T)/CBS dissociation curves.

**The organization of the dataset.** The SH250×10 data set is divided into seven groups defined by the element featuring the  $\sigma$ -hole. The sizes of these groups as well as the key statistics on the benchmark interaction energies in these groups are listed in Table 1. Although one of the criteria when choosing the complexes was to keep the numbers of the systems balanced between the groups, the Cl, S and P groups are somewhat smaller; the weaker  $\sigma$ -hole interactions of lighter elements less often meet the conditions used here to define a  $\sigma$ -hole interaction. The numbers of systems representing each element-pair combination are listed in Table S3 in the SI.

Analogously to the other NCIA data sets, the complexes are assigned a code consisting of the number of the group (corresponding to the  $\sigma$ -hole element), subgroup (electron-donor element), and the number of the system within the subgroup.

Table 1: The counts of complexes in each group of the SH250 data set. For each of the groups, the average of the benchmark interaction energy  $\Delta E_{avg}$  and the most negative interaction energy  $\Delta E_{max}$  are also listed (values in kcal/mol).

Group	Count	$\Delta E_{avg}$	$\Delta E_{max}$
1 Cl	29	-3.4	-14.5
2 Br	36	-4.2	-17.1
3 I	42	-4.6	-15.6
4 S	31	-5.1	-17.1
5 Se	44	-6.1	-22.8
6 P	33	-8.5	-19.2
7 As	35	-10.7	-22.2

**Subsets obtained by clustering analysis.** In each NCIA data set, we define smaller subsets intended for applications that require less data. To maximize the diversity of these subsets, they are constructed systematically using a clustering analysis described in detail in ref. 1. The similarity of the system is evaluated from the errors of the DFT and SQM

methods tested on the data set. A set of cluster representatives thus forms the most difficult test set of a given size. The assignment of the systems to subsets of 20, 50, 100 and 200 entries is listed in Table S4 in the SI as well as in the data files there.

## 2.2 Benchmark Calculations

The benchmark interaction energies in the SH250 $\times$ 10 data set are computed using the same composite CCSD(T)/CBS scheme as all the other NCIA data sets to ensure their seamless compatibility.<sup>1,2</sup>

The CCSD(T)/CBS interaction energies are assembled from three components. The MP2 correlation energy is extrapolated to the complete basis set limit from calculations in two basis sets.<sup>26</sup> The Hartree–Fock term is taken from the calculation in the larger basis and is not extrapolated. Finally, the  $\Delta$ CCSD(T) correction, the difference between CCSD(T) and MP2 correlation energies, is computed in a third, smaller basis set.

For all of the calculations used in the CCSD(T)/CBS benchmark, we have used the correlation-consistent basis sets with diffuse functions from the aug-cc-pVXZ family (X = D, T, Q, 5), introduced by Dunning et al.<sup>27</sup> For the  $\Delta$ CCSD(T) calculations in the triple- $\zeta$  basis set, we use the “heavy-augmented” variant of the basis set with diffuse functions dropped for hydrogen (labeled as heavy-aug-cc-pVTZ). Since the sub-valence orbitals of bromine and iodine have been shown to be important for the London dispersion,<sup>28</sup> we have employed the polarizable-core version of the basis sets<sup>29</sup> and excluded the last sub-valence shell from the frozen core for Br, I as well as As and Se. Furthermore, since relativistic effects may be important in these elements,<sup>18</sup> the relativistic effective-core potentials along with the corresponding version of the basis sets, aug-cc-pwCVXZ-PP,<sup>30</sup> are used for these atoms.

Only a CCSD(T)/CBS setup with the  $\Delta$ CCSD(T) computed in at least a triple- $\zeta$  basis set can be considered a true “gold standard” benchmark.<sup>31,32</sup> Here, we extrapolate the MP2 energy from aug-cc-pVQZ and aug-cc-pV5Z, and the  $\Delta$ CCSD(T) correction is calculated in



the heavy-aug-cc-pVTZ basis.

Most of the points on the SH250×10 dissociation curves (all but the equilibrium and the closest point) are computed at the “silver level” and then then rescaled to the “gold level” using a procedure described and verified in the other papers on the NCIA data sets.<sup>1</sup> This procedure has been shown to yield a negligible error in the order of units of cal/mol, but it makes it possible to accelerate the benchmark calculations considerably.<sup>1,4</sup> At the “silver level”, MP2 is extrapolated from aug-cc-pVTZ and aug-cc-pVQZ, and  $\Delta\text{CCSD(T)}$  is computed in aug-cc-pVDZ.

All the calculations of interaction energies employ the counterpoise correction<sup>33</sup> for the basis set superposition error. In addition, frozen-core approximation and RI approximation (density fitting) are used in all the calculations of correlation energy. In the RI approximation, the auxiliary basis sets optimized for the respective AO basis have been used.<sup>34</sup>

The majority of the benchmark calculations have been carried out in Turbomole.<sup>24,35</sup> Some MP2 calculations were carried out in the MRCC program<sup>36,37</sup> after it was verified to yield the same results as Turbomole (up to insignificant numerical differences).

## 2.3 Density Functional Theory Calculations

Selected DFT methods ranging from generalized gradient approximation (GGA) functionals to double hybrids have been tested on the SH250×10 data set. They are listed in Table 2. All the DFT calculations have been performed using ORCA 4.2.0 and 5.0.1<sup>38</sup> in the def2-QZVP basis set.<sup>39</sup> The RI approximation has been used in all calculations except for range-separated functionals.

More recent DFT functionals already contain some form of a dispersion correction; for the remaining ones, we have tested several standalone corrections available. For most functionals, we have tested the D3 correction<sup>59</sup> with Becke-Johnson (BJ) damping<sup>60</sup> and the more recent D4 correction, which adds charge-dependence of dispersion coefficients.<sup>61</sup> For B3LYP, we have also tested the D3 correction with either the original zero damping, D3(zero), or the

Table 2: The DFT functionals tested on the SH250×10 data set.

GGA	Meta-GGA	Hybrid	Range-separated	Double-hybrid
BLYP <sup>40,41</sup>	SCAN <sup>42</sup>	B3LYP <sup>43,44</sup>	$\omega$ B97X-D3(BJ) <sup>45</sup>	DSD-BLYP <sup>46</sup>
BP <sup>40,47</sup>	TPSS <sup>48</sup>	BHLYP <sup>49</sup>	$\omega$ B97M-D3(BJ) <sup>45</sup>	DSD-PBEP86 <sup>50</sup>
PBE <sup>51</sup>	B97-D3(BJ) <sup>52</sup>	PBE0 <sup>53</sup>	$\omega$ B97X-V <sup>54</sup>	DSD-PBEP86-D3(BJ) <sup>50</sup>
	B97M-V <sup>55</sup>	TPSSH <sup>56</sup>	$\omega$ B97M-V <sup>57</sup>	revDSD-PBEP86-D3(BJ) <sup>50</sup>
		M06-2X <sup>58</sup>		revDSD-PBEP86-D4 <sup>50</sup>

“optimized-power” damping D3(OP),<sup>62</sup> and the non-local dispersion correction used in the B3LYP-NL method.<sup>63</sup>

The variants of D3 dispersion corrections have been added to the interaction energies using the Cuby framework,<sup>64,65</sup> the D4 correction has been calculated using a program provided by the authors of the method.<sup>66</sup> The B3LYP-NL method is available in Orca.

## 2.4 Constrained DFT Calculations of Charge-Transfer Energy

In some systems, we investigate also the contribution of charge transfer to the interaction energy. To quantify this term, we use a variant of the constrained DFT (cDFT) method<sup>67–70</sup> designed specifically for this purpose. The component of the interaction energy associated with charge transfer, also called charge-transfer energy ( $\Delta E_{CT}$ ), is computed as the energy difference between an unconstrained calculation and a cDFT calculation in which intermolecular charge transfer is prohibited. The constraint is defined using relaxed electron densities of free molecules (a fragment-based Hirshfeld partitioning scheme), which represent a natural charge-transfer-free reference state. Details on the methodology were presented in our earlier publications.<sup>17,71</sup>

To minimize the artificial overestimation of the charge transfer by DFT, we use the BHLYP functional, which includes a high ratio (50%) of exact exchange, as our reference. This functional also yields accurate interaction energies in the SH250 data set.

The cDFT calculations have been performed in the deMon program, version 6.02, using the def2-QZVP atomic-orbital basis set and density fitting with the GEN-A3\* auxiliary basis.

The constrained DFT with the ‘‘Hirshfeld fragment’’ population scheme has been applied with the convergence limit of  $10^{-4}$  a.u.

## 2.5 Semiempirical Quantum-Mechanical Methods

On the newly developed data set, we have also evaluated the performance of several SQM methods. Here the choice of methods is limited as many SQM methods do not have parameters for all the elements studied here. The PM6 method<sup>20</sup> is applicable to almost any element, and we also test its versions including additional empirical corrections: dispersion-corrected PM6-D3, PM6-D3H4 with dispersion and hydrogen-bond correction<sup>72</sup> and PM6-D3H4X, also adding halogen-bond correction.<sup>21</sup> Its successor PM7<sup>73</sup> has these corrections built in. These calculations were performed in MOPAC 2016.<sup>74</sup>

Finally, the empirical tight-binding method GFN-xTB<sup>75</sup> and its newer version GFN2-xTB<sup>76</sup> are also applicable to the whole chemical space of the SH250 data set. The xtb software package has been used to carry out the calculations.<sup>77</sup>

## 2.6 Data Availability

The SH250×10 data set is available in several formats in the Supporting Information of this paper, at the [www.nciatlas.org](http://www.nciatlas.org) website and at the GitHub repository <https://github.com/Honza-R/NCIAtlas>.

The geometries in the .xyz format also include the benchmark interaction energies, the definition of the monomers and the categorization of the systems, which makes them all that is needed for performing the calculations. The benchmark interaction energies, the results of the individual calculations used to construct them, the results of the tested methods and additional metadata are also provided as plain-text tables.

Additionally, all these data are bundled in a structured YAML<sup>78</sup> file, which can be used to automate the calculations on the data set with the Cuby framework.<sup>64</sup> This file will also be distributed as a part of Cuby at <http://cuby4.molecular.cz/>.

## 3 Results and Discussion

### 3.1 Characterization of the SH250×10 Data Set

The benchmark interaction energies in the equilibrium geometries range from -0.7 to -22.8 kcal/mol, with the average of -6.1 kcal/mol. In each group of elements (halogens, chalcogens, pnictogens), the average interaction energy increases with the weight of the element as expected (see Table 1). The strongest average interactions are in the arsenic group.

The  $\sigma$ -hole interactions are known for very short equilibrium intermolecular distances. This can be demonstrated by a comparison with the R739×5 data set, which also covers interactions of halogens, S and P with other atoms. There, the initial geometries of the R739×5 dissociation curves (referred to as R739) are designed to be slightly repulsive, targeting the interaction energy of approx. +2.5 kcal/mol. Table 3 compares intermolecular distances in R739, relative to the sum of van der Waals radii, with equilibrium distances measured in the SH250 data set. In most element pairs, and also on average, the equilibrium geometry with an attractive  $\sigma$ -hole interaction has a closer intermolecular distance than a repulsive contact in another geometrical arrangement. The largest difference has been observed in systems with heavier atoms, where the  $\sigma$ -hole is usually larger.

As a result, already the equilibrium geometries in the SH250 data set represent a difficult test case for many approximate methods, and the closer points on the SH250×10 dissociation curves are even more challenging, with extremely short intermolecular distances. The interaction energies are moderate even in the repulsive part of the dissociation curves, but it has to be kept in mind that these numbers are a sum of individual contributions larger than in more common non-covalent interactions closer to the van der Waals contact distance.

### 3.2 Density Functional Theory Calculations

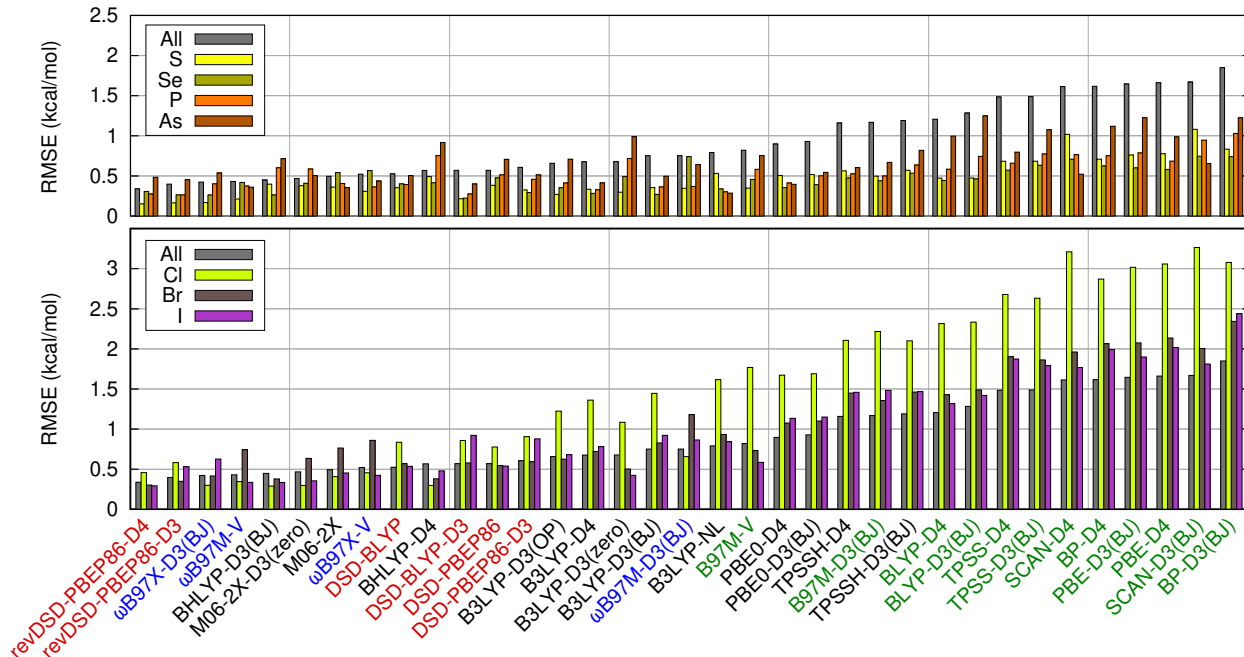
First, we discuss DFT calculations in the equilibrium geometries, the SH250 data set. The results are summarized in a plot of the root-mean-square error, RMSE, in the whole data set

Table 3: The average relative intermolecular contact distance for identical element pairs in the SH250 data set of  $\sigma$ -hole interactions and in the R739 data set of artificial repulsive contacts (atom–atom distances relative to the sum of the respective van der Waals radii). Molecules with selenium and arsenic are not present in R739.

Element pair	SH250	R739
Cl-O	0.88	0.85
Cl-N	0.86	0.87
Cl-S	0.81	0.87
Br-O	0.88	0.86
Br-N	0.85	0.89
Br-S	0.81	0.88
I-O	0.86	0.87
I-N	0.83	0.89
I-S	0.83	0.89
S-O	0.86	0.86
S-N	0.81	0.88
P-O	0.80	0.98
P-N	0.79	1.00
Average	0.84	0.89

and its groups in Fig. 2. A corresponding plot of the systematic error (mean signed error, MSE) is provided in the SI as Fig. S1, and the source data used to construct these plots are available in the SI as Tables S5 and S7.

Figure 2: The RMSE of DFT methods in the SH250 data set (equilibrium geometries only) and its groups by the interacting element. The DFT-functional labels are colored by the rung of the functional: double hybrids (red), range-separated hybrids (blue), hybrids (black) and GGA and meta-GGA (green).



The double-hybrid functional revDSD-PBEP86 with either D4 or D3 dispersion correction has been found to be the most accurate one of the SH250 data set. The range-separated functionals  $\omega$ B97X-D3(BJ),  $\omega$ B97M-V, and even the hybrids BHLYP-D3(BJ) and M06-2X follow closely and outperform even the other double-hybrid functionals studied.

All GGA and meta-GGA, but also some hybrid functionals have noticeably larger errors in the chlorine group (see Fig. 2). This issue can be traced down mainly to several complexes of molecular chlorine, where the interaction energy is severely exaggerated. Three of these complexes with the largest errors, along with one example of the same element combination where this problem does not occur, are listed in Table 4. In the series of the functional with increasing amount of exact exchange, the error decreases; in BHLYP, it becomes negligible. The good performance of the revDSD-PBEP86 functionals can also be explained by the high content of exact exchange, which amounts to 69 %. This indicates that the issue is connected to the charge-delocalization error of DFT.<sup>79,80</sup> Similar, albeit smaller, errors have also been

found in some complexes of molecular bromine and iodine.

Table 4: Selected complexes of molecular chlorine – DFT-D3(BJ) interaction energies, their error to the CCSD(T)/CBS benchmark, and charge-transfer energies computed using constrained DFT. All values in kcal/mol.

No.	El. donor	CCSD(T)/CBS	BLYP	B3LYP	BHLYP
Interaction energy					
1.2.01	acetonitrile	-3.34	-3.68	-3.48	-3.28
1.2.08	trimethylamine	-14.52	-19.39	-17.33	-13.85
1.4.03	dimethylthioether	-9.06	-16.02	-13.55	-9.74
1.4.05	thioacetone	-7.33	-14.84	-12.13	-8.25
Interaction energy, error to CCSD(T)/CBS					
1.2.01	acetonitrile		-0.35	-0.14	0.05
1.2.08	trimethylamine		-4.87	-2.82	0.67
1.4.03	dimethylthioether		-6.96	-4.49	-0.68
1.4.05	thioacetone		-7.51	-4.80	-0.92
Charge-transfer energy					
1.2.01	acetonitrile		-1.03	-0.88	-0.72
1.2.08	trimethylamine		-14.64	-13.20	-11.05
1.4.03	dimethylthioether		-10.15	-8.96	-7.06
1.4.05	thioacetone		-7.34	-7.10	-5.76

To investigate this issue in more depth, we have calculated the contribution of charge transfer to the interaction energy of the three most problematic complexes using constrained DFT. Here, the effect of the functional is less pronounced, which confirms that this variant of cDFT is a robust method of the estimation of the charge transfer. The charge-transfer calculations in the BHLYP functional, which also yields accurate interaction energies here, can be considered a good estimate of this component of the interaction energy.

The well-behaved chlorine–acetonitrile complex exhibits rather weak charge transfer, while in the three problematic systems, the charge-transfer term is approximately as large as the interaction energy itself. This confirms our hypothesis on the source of the error observed in many of the lower-rung functionals, which are more prone to the overdelocalization of the charge. The strength of this effect in some of the hybrid functionals including B3LYP is, however, still surprising.

On the other hand, some of the functionals that do not suffer from this problem, such as

M06-2X and all the range-separated functionals except for  $\omega$ B97X-D3(BJ), tend to underestimate the interaction energy in halogen-bonded complexes (see the plot of the systematic errors, Fig. S1 in the SI and Table S7 therein); this effect is the strongest in the complexes of bromine.

In contrast to the other NCIA data sets, the differences between the variants of the dispersion correction (where available) at the equilibrium distances are rather small and there are no clear trends. These results suggest that in this data set, the DFT errors are governed mainly by the exchange part of the functional, and the details in the dispersion correction become significant only if the fundamental description of the system is correct. The best-performing functionals also usually include dispersion correction as an integral part that was parametrized together with other terms, so there is a room for a favorable error cancellation.

Next, we discuss the entire SH250 $\times$ 10 data set, namely including also non-equilibrium geometries. These results are summarized in the SI in Figures S2 and S3 and Tables S6 and S8. The two rev-DSD-PBEP86 functionals are still the two best DFT methods tested here, except that they exchanged places and the -D3 variant outperforms the -D4 one by a small margin. This is caused by the increase of the RMSE of rev-DSD-PBEP86-D4 from 0.34 kcal/mol at equilibrium distances to 0.45 kcal/mol in the entire SH250 $\times$ 10 data set. On the other hand, the other double-hybrid “DSD-” functionals show a minimal drop in accuracy (of about 0.02 kcal/mol) when passing to the whole dissociation curves, and they move up in ranks over other higher-performing functionals.

In GGA and meta-GGA functionals, two different modes of failure can be observed at the shortest distances of the SH250 $\times$ 10 data set. In the complexes of halogens, the interaction is systematically overestimated, which likely stems from the charge-delocalization error discussed above. On the other hand, there are also rather large positive systematic errors (overestimated repulsion), most pronounced in the arsenic group, followed by the complexes of phosphorus. Again, an overestimated repulsion can also be attributed to the



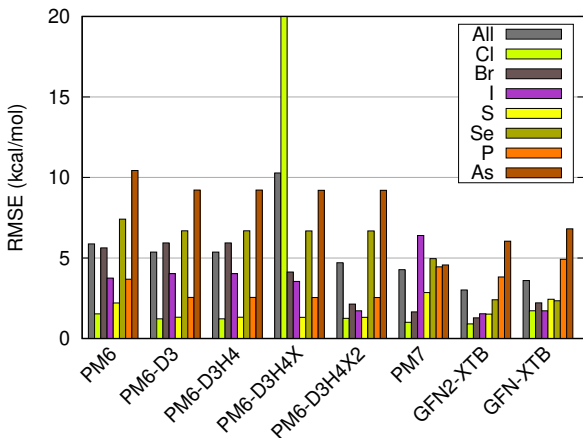
exchange part of the functional, and the SH250 $\times$ 10 seems to expose these different errors.

At the short intermolecular distances, the differences between alternative dispersion corrections are more pronounced as well. For example, B3LYP-D3(zero) is clearly hindered by the form of the dispersion corrections and performs considerably worse than in the equilibrium. The inadequacy of the zero-damping function at shorter distances has already been discussed elsewhere and it was a reason for its replacement with Becke-Johnson damping.<sup>60</sup> On the other hand, NL non-local correction, which is in the equilibrium the least accurate correction paired with B3LYP, becomes the most accurate one in SH250 $\times$ 10.

### 3.3 Semiempirical Quantum-Mechanical Methods

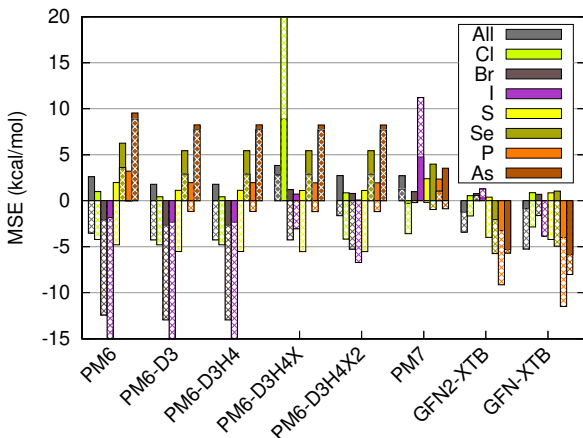
Few SQM methods are applicable to the whole chemical space spanned by the SH250 data set. Among these, we have selected PM6-D3H4, PM6-D3H4X, PM7 and GNF2-xTB for more detailed analysis as they have all the features necessary for the description of non-covalent interactions. We have also introduced a reparameterization of the halogen-bond correction for the PM6 method, labeled as PM6-D3H4X2, which is described in detail later. The errors in the SH250 data set (equilibrium geometries) as a whole and in the groups by the  $\sigma$ -hole element are plotted in Fig. 3. Tables of both the RMSE and the MSE of these methods are provided in the SI as Tables S9 and S10.

Figure 3: The RMSE of semiempirical quantum-mechanical methods in the SH250 data set and its groups by the interacting element (equilibrium geometries only).



In general, all the SQM methods yield larger errors than in other NCIA data sets. Halogens and sulfur are described satisfactorily, but none of these methods describes the interactions of pnictogens well. Since all the methods considered use elementwise or even pairwise parameters, it is worth discussing them by the element groups, where we will also examine the whole dissociation curves of the SH250 $\times$ 10 data set. The errors averaged over multiple systems often do not capture specific failures of the SQM methods, which leads to a wrong shape of the dissociation curve. We thus provide plots of all the curves in the SI (Figures S4 to S10) and selected examples are discussed here. In terms of quantitative measures, the most important one is the behavior at the shortest distances – these errors are listed in the SI in Tables S11 and S12. For an easy comparison, we combine the plots of the systematic error, MSE, in the equilibrium and at the shortest distance in Fig. 4, which carries all the essential information needed to judge the methods.

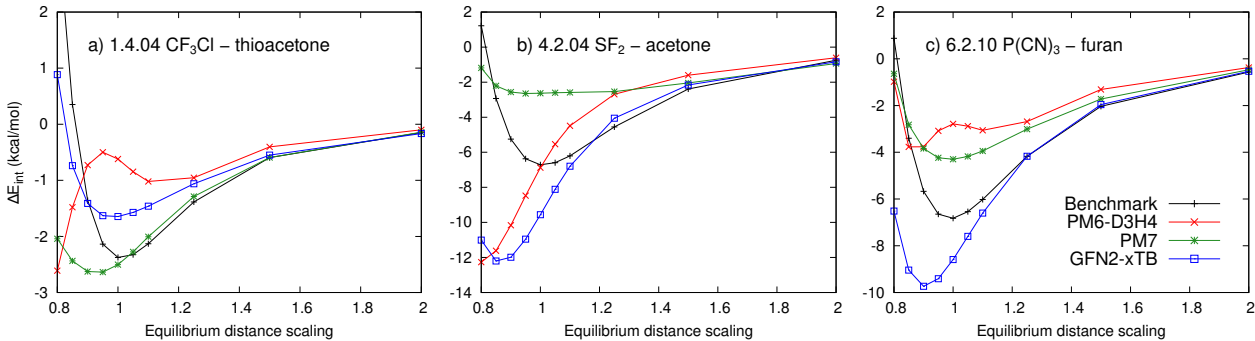
Figure 4: The RMSE of the systematic error (MSE) of semiempirical quantum-mechanical methods in the SH250 $\times$ 10 data set and its groups by the interacting element at the equilibrium (solid) and the shortest distance (cross-hatched).



**Halogens.** The X in PM6-D3H4X stands for a dedicated correction for halogen bonds that corrects a serious lack of repulsion in PM6. It has been parametrized on only a few prototypical halogen bonds, but it has been found to be transferable to other similar systems.<sup>21</sup> The more diverse SH250 set highlights the deficiencies of this parametrization, namely too strong

repulsion in complexes with a short intermolecular distance, which is the most pronounced in the complexes of chlorine (with the RMSE of 26 kcal/mol). The error grows even larger at closer points of the dissociation curves. This issue is easily fixed by reparametrization in the PM6-D3H4X2 method, discussed below. There are, however, some remaining issues that cannot be solved by a simple repulsive correction, mostly in less common element combinations such as halogen–sulfur and halogen–halogen interactions. Here, PM6 may yield a wrong shape of the dissociation curve with a false minimum at a longer distance and a severe lack of repulsion at the shortest distances (for an example of Cl–S interaction, see Fig. 5a). PM7, which has been parametrized on data including the X40 data set, works rather well for chlorine and bromine, but it severely overestimates repulsion in the complexes of iodine already in the equilibrium geometries (the RMSE of 6.4, the MSE of 4.7 kcal/mol) and even more at the shortest distances. GFN2-xTB yields the best results here, although the error in the iodine group is still rather large (the RMSE of 1.5 kcal/mol). Here the error has a random character with the MSE of only 0.3 kcal/mol, and the systematic error remains small even at shorter distances.

Figure 5: Dissociation curves of selected complexes demonstrating large errors of semiempirical QM methods.



**Chalcogens.** At the equilibrium, both PM6-D3H4 and PM7 underestimate the strength of the interactions of sulfur slightly and of selenium severely. Nevertheless, the problem is more complex – the dissociation curves often have a wrong shape, with a deep minimum at short distances (only PM7 does not suffer from this, but it often underestimates the inter-

action) as illustrated in Fig. 5b. PM6-D3H4 also often yields a second shallow minimum at longer distances. Since these errors are rather complex and cannot be addressed by a simple correction, a reparametrization of the method itself would be needed. For selenium, this is reversed because PM6-D3H4 now underestimates the interaction more strongly while PM7 often provides an unreasonably attractive interaction. GFN2-xTB consistently overestimates the strength of the interactions of selenium (with the MSE of -2.1 kcal/mol), which is in contrast with its earlier version GFN-xTB, which has a positive MSE of 1.0 kcal/mol. This seems to be a systematic error, which can be corrected by a more careful parametrization of this element.

**Pnictogens.** In some pnictogen-bonded complexes of phosphorus, the repulsion at shorter separations is still underestimated by SQM methods and PM6-D3H4 again yields multiple minima in several cases (an example is shown Fig. 5c). Phosphorus is, however, one of the elements described rather well by these methods, which also applies to the short distances. On the other hand, the GFN2-xTB overestimates the interactions of phosphorus significantly with the MSE of -3.2 kcal/mol already in the equilibrium and -9.1 kcal/mol at the shortest distances. The latter issue is the largest problem encountered in GFN2-xTB in this study. Passing to arsenic, PM6-D3H4 has a large positive systematic error here (MSE 8.2 kcal/mol in equilibrium) while PM7 describes it similarly to phosphorus. Also GFN2-xTB has a large error in the arsenic group, with the RMSE of 6.0 kcal/mol at the equilibrium being the largest error among the elements studied. Nevertheless, the issues at short distances are weaker than in the case of phosphorus.

Overall, the GFN2-xTB is the most accurate SQM method when tested on equilibrium geometries, but it is not free of the problems plaguing all the SQM methods – the dissociation curves are often wrong, with a deeper minimum at shorter distances. PM7 and PM6-D3H4 are even less balanced, with large errors in several elements. The description of selenium or arsenic interactions by SQM methods is troublesome in general. Considering

the dissociation curves and thus the ability of the methods to predict correct geometries, only PM6-D3H4X2 (with reparametrized halogen-bonding correction) and GFN2-xTB are applicable to all halogen bonds. PM6-D3H4 needs a correction to describe sulfur well but works reasonably for phosphorus. GFN2-xTB yields too short intermolecular distances in all chalcogen and pnictogen bonds. It is clear that all SQM methods have serious problems in describing short-ranged non-covalent interactions of less common elements, which is likely to stem from a lack of relevant model systems in the data used to parameterize them. It can be hoped that using the data set presented here would improve many of the systematic errors discussed here.

### 3.4 The Reparametrization of the Halogen-Bonding Correction for PM6

The results in the previous section clearly show that the X correction in PM6-D3H4X fails in halogen bonds farther from the few prototypical systems used in its development. The most obvious issue is the severe overestimation of repulsion in the stronger halogen bonds of chlorine. We have thus reparametrized it on the SH250×10 data set, which yields a more robust method labeled here as PM6-D3H4X2. The correction keeps the original form of a simple exponential repulsion  $E_X = k \cdot e^{-ar}$  with two parameters  $k$  and  $a$ , and with  $r$  being the atom–atom distance.<sup>21</sup> The parameters are derived for each element pair, and the correction is applied to every matching pair of atoms except for those separated by one to three covalent bonds.

The parameters were optimized for the halogen bonds of Cl, Br and I with nitrogen and oxygen considered previously, and the correction was also extended to Cl–S, Br–S and I–Br interactions, which exhibited large but systematic errors in the present data set. The parametrization was carried out separately for each element pair on the relevant subset of the SH250×10 data set, with the RMSE being the objective function minimized. The two shortest points of the dissociation curves (with the scaling factors 0.8 and 0.85) were

not considered in order to improve the description of the minimum at the cost of worse description of unphysically short distances.

The resulting parameters are listed in Table 5. With this modification, the method now describes halogen bonds of all the elements rather well (with the largest RMSE of 2.1 kcal/mol for bromine – see Fig. 3), and the systematic error at the equilibrium distances is also controlled well (below 1 kcal/mol in all the elements, plotted in Fig. 4). A good sign is that this has been achieved by minimizing the RMSE, not by targeting the systematic error itself. This complements our previous findings that PM6 describes the  $\sigma$ -hole rather well and its problems in the description of halogen bonds stem from the description of repulsion.<sup>81</sup> The new correction is still too weak at the shortest distances, but it should yield a reasonable intermolecular distance, which is a more important objective.

Table 5: New parameters for the halogen-bonding correction in PM6-D3H4X. The parameter  $k$  in kcal/mol,  $a$  in  $1/\text{\AA}$ .

Elements	$k$	$a$
Cl-N	990079	6.915
Cl-O	100891	5.217
Cl-S	516740	4.537
Br-N	23215	2.792
Br-O	1485982	5.131
Br-S	131621890	6.471
I-N	13473085	5.594
I-O	1484563	4.738
I-Br	12638	2.432

## 4 Conclusions

The SH250 $\times$ 10 data set extends the Non-Covalent Interaction Atlas to  $\sigma$ -hole interactions – halogen, chalcogen and pnictogen bonds featuring Cl, Br, I, S, Se, P and As. The accurate CCSD(T)/CBS benchmark interaction energies are compatible with the remaining NCIA data sets, and together they provide the most comprehensive coverage of non-covalent interactions of neutral molecules in a chemical space significantly wider than that provided by

the data sets available previously.

In this work, this data set has been applied to benchmarking a variety of semiempirical quantum-mechanical and density functional theory methods. It has been found that already the equilibrium geometries, the SH250 data set, present a challenging test set for DFT. In contrast to the other NCIA data sets, we have identified large errors that can be attributed to exaggerated charge transfer due to the overdelocalization of the electron density. This error plagues all the GGA and meta-GGA functionals tested, but it is surprisingly strong even in some common hybrid functionals such as B3LYP or PBE0. It is mitigated by using a larger fraction of exact exchange in BHLYP or by utilizing range-separated hybrid functionals. The most accurate DFT methods tested are the double-hybrid functionals rev-DSD-PBEP86-D4 and rev-DSD-PBEP86-D3, which also employ a large fraction of exact exchange, followed by  $\omega$ B97X-D3(BJ) and  $\omega$ B97M-V.

The situation is even worse in SQM methods – none of those tested describes all the  $\sigma$ -hole interactions without serious issues. Overall, GFN2-xTB ranked best, but even this rather robust method has problems e.g. in describing interactions of phosphorus and arsenic. We have also reparametrized the halogen-bonding correction in the PM6-D3H4X method, which is now able to describe common halogen bonds rather well, but it still exhibits severe problems in the less common combinations of elements. These problems are mostly attributable to incomplete parametrization of these methods, and the present benchmark data can be used to improve the future versions of these methods.

With all the data openly available, we hope that the SH250 $\times$ 10 data set will contribute to the development of a more robust computational method able to describe these challenging non-covalent interactions better.

## 5 Acknowledgements

We acknowledge 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) Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets 5: London Dispersion in an Extended Chemical Space. **2022**, Preprint, DOI: 10.26434/chemrxiv-2022-pl3r8.



- (5) Politzer, P.; Murray, J. S.; Clark, T.; Resnati, G. The  $\sigma$ -hole revisited. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32166–32178.
- (6) Oliveira, V.; Kraka, E. Systematic Coupled Cluster Study of Noncovalent Interactions Involving Halogens, Chalcogens, and Pnicogens. *J. Phys. Chem. A* **2017**, *121*, 9544–9556.
- (7) Bauzá, A.; Frontera, A. Aerogen Bonding Interaction: A New Supramolecular Force? *Angew. Chem. Int. Ed.* **2015**, *54*, 7340–7343.
- (8) Wilcken, R.; Zimmermann, M. O.; Lange, A.; Joerger, A. C.; Boeckler, F. M. Principles and Applications of Halogen Bonding in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **2013**, *56*, 1363–1388.
- (9) Iwaoka, M.; Babe, N. Mining and Structural Characterization of S...X Chalcogen Bonds in Protein Database. *Phosphorus Sulfur Silicon Relat Elem* **2015**, *190*, 1257–1264.
- (10) Riveras, J. A. F.; Frontera, A.; Bauzá, A. Selenium chalcogen bonds are involved in protein–carbohydrate recognition: a combined PDB and theoretical study. *Phys. Chem. Chem. Phys.* **2021**, *23*, 17656–17662.
- (11) Czarny, R. S.; Ho, A. N.; Shing Ho, P. A Biological Take on Halogen Bonding and Other Non-Classical Non-Covalent Interactions. *Chem. Rec.* **2021**, *21*, 1240–1251.
- (12) Mohajeri, A.; Pakiari, A. H.; Bagheri, N. Theoretical studies on the nature of bonding in  $\sigma$ -hole complexes. *Chem. Phys. Lett.* **2009**, *467*, 393–397.
- (13) Riley, K. E.; Hobza, P. The relative roles of electrostatics and dispersion in the stabilization of halogen bonds. *Phys. Chem. Chem. Phys.* **2013**, *15*, 17742–17751.
- (14) Tsuzuki, S.; Wakisaka, A.; Ono, T.; Sonoda, T. Magnitude and Origin of the Attraction and Directionality of the Halogen Bonds of the Complexes of C<sub>6</sub>F<sub>5</sub>X and C<sub>6</sub>H<sub>5</sub>X (X=I, Br, Cl and F) with Pyridine. *Chem. Eur. J* **2012**, *18*, 951–960.

- (15) Wolters, L. P.; Bickelhaupt, F. M. Halogen Bonding versus Hydrogen Bonding: A Molecular Orbital Perspective. *ChemistryOpen* **2012**, *1*, 96–105.
- (16) Wang, C.; Danovich, D.; Mo, Y.; Shaik, S. On The Nature of the Halogen Bond. *J. Chem. Theory Comput.* **2014**, *10*, 3726–3737.
- (17) Řezáč, J.; Lande, A. d. l. On the role of charge transfer in halogen bonding. *Phys. Chem. Chem. Phys.* **2016**, *19*, 791–803.
- (18) Kozuch, S.; Martin, J. M. L. Halogen Bonds: Benchmarks and Theoretical Analysis. *J. Chem. Theory Comput.* **2013**, *9*, 1918–1931.
- (19) Bauzá, A.; Alkorta, I.; Frontera, A.; Elguero, J. On the Reliability of Pure and Hybrid DFT Methods for the Evaluation of Halogen, Chalcogen, and Pnicogen Bonds Involving Anionic and Neutral Electron Donors. *J. Chem. Theory Comput.* **2013**, *9*, 5201–5210.
- (20) 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.
- (21) Řezáč, J.; Hobza, P. A halogen-bonding correction for the semiempirical PM6 method. *Chem. Phys. Lett.* **2011**, *506*, 286–289.
- (22) Grabowski, S. J.; Sadlej, A. J.; Sokalski, W. A.; Leszczynski, J. Attractive halogen–halogen interactions: F3CCl...FH and F3CCl...FCH3 dimers. *Chem. Phys.* **2006**, *327*, 151–158.
- (23) Hostaš, J.; Řezáč, J. Accurate DFT-D3 Calculations in a Small Basis Set. *J. Chem. Theory Comput.* **2017**, *13*, 3575–3585.
- (24) TURBOMOLE v7.3. 2018; <http://www.turbomole.com>.
- (25) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.;

Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian09 Revision D.01*.

- (26) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence of correlated calculations on water. *J. Chem. Phys.* **1997**, *106*, 9639–9646.
- (27) 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.
- (28) 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.
- (29) 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.
- (30) 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.
- (31) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. Basis set convergence of the coupled-cluster correction,  $\delta$ MP2CCSD(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.
- (32) Kodrycka, M.; Patkowski, K. Platinum, gold, and silver standards of intermolecular interaction energy calculations. *J. Chem. Phys.* **2019**, *151*, 070901.
- (33) 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.
- (34) 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.
- (35) Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. Turbomole. *WIREs Comput. Mol. Sci.* **2014**, *4*, 91–100.
- (36) Kállay, M.; Nagy, P. R.; Mester, D.; Rolik, Z.; Samu, G.; Csontos, J.; Csóka, J.; Szabó, P. B.; Gyevi-Nagy, L.; Hégyel, B.; Ladjánszki, I.; Szegedy, L.; Ladóczki, B.; Petrov, K.; Farkas, M.; Mezei, P. D.; Ganyecz, Á. The MRCC program system: Accurate quantum chemistry from water to proteins. *J. Chem. Phys* **2020**, *152*, 074107.
- (37) MRCC, a quantum chemical program suite written by M. Kállay, P. R. Nagy, D. Mester, Z. Rolik, G. Samu, J. Csontos, J. Csóka, P. B. Szabó, L. Gyevi-Nagy, B. Hégyel, I. Ladjánszki, L. Szegedy, B. Ladóczki, K. Petrov, M. Farkas, P. D. Mezei, and Á. Ganyecz. 2017; [www.mrcc.hu](http://www.mrcc.hu).

- (38) Neese, F. The ORCA program system. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73–78.
- (39) 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.
- (40) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (41) 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.
- (42) Sun, J.; Ruzsinszky, A.; Perdew, J. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **2015**, *115*, 036402.
- (43) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (44) Raghavachari, K. Perspective on “Density functional thermochemistry. III. The role of exact exchange”. *Theor. Chem. Acc.* **2000**, *103*, 361–363.
- (45) 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  $\omega$ B97M-V Approaches. *J. Chem. Theory Comput.* **2018**, *14*, 5725–5738.
- (46) 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.
- (47) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33*, 8822–8824.

- (48) 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.
- (49) Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377.
- (50) 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.
- (51) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (52) 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.
- (53) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- (54) Mardirossian, N.; Head-Gordon, M.  $\omega$ 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.
- (55) 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.
- (56) 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*, 12129–12137.
- (57) Mardirossian, N.; Head-Gordon, M.  $\omega$ B97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016**, *144*, 214110.
- (58) 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 functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
- (59) 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.
- (60) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.
- (61) 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.
- (62) 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.
- (63) 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.

- (64) Řezáč, J. Cuby: An Integrative Framework for Computational Chemistry. *J. Comput. Chem.* **2016**, *37*, 1230–1237.
- (65) Řezáč, J. Cuby 4, software framework for computational chemistry. 2015; <http://cuby4.molecular.cz/>.
- (66) Eike Caldeweyher, C. B.; Grimme, S. D4 - A Generally Applicable Atomic-Charge Dependent London Dispersion Correction. 2017; <https://www.chemie.uni-bonn.de/pctc/mulliken-center/software/dftd4>.
- (67) Dederichs, P. H.; Blügel, S.; Zeller, R.; Akai, H. Ground States of Constrained Systems: Application to Cerium Impurities. *Phys. Rev. Lett.* **1984**, *53*, 2512–2515.
- (68) Wu, Q.; Van Voorhis, T. Direct Optimization Method to Study Constrained Systems within Density-Functional Theory. *Phys. Rev. A* **2005**, *72*, 024502.
- (69) Kaduk, B.; Kowalczyk, T.; Van Voorhis, T. Constrained Density Functional Theory. *Chem. Rev.* **2012**, *112*, 321–370.
- (70) Řezáč, J.; Lévy, B.; Demachy, I.; de la Lande, A. Robust and Efficient Constrained DFT Molecular Dynamics Approach for Biochemical Modeling. *J. Chem. Theory Comput.* **2011**, *8*, 418–427.
- (71) Řezáč, J.; de la Lande, A. Robust, Basis-Set Independent Method for the Evaluation of Charge-Transfer Energy in Noncovalent Complexes. *J. Chem. Theory Comput.* **2015**, *11*, 528–537.
- (72) Řezáč, J.; Hobza, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *J. Chem. Theory Comput.* **2012**, *8*, 141–151.



- (73) 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.
- (74) Stewart, J. J. P. MOPAC 2016. 2016; <http://openmopac.net/>.
- (75) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements ( $Z = 1-86$ ). *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.
- (76) 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.
- (77) XTB, Semiempirical Extended Tight-Binding Program Package. 2019; <https://github.com/grimme-lab/xtb>, original-date: 2019-09-30T12:40:09Z.
- (78) The Official YAML Web Site. <http://yaml.org/>.
- (79) Lundberg, M.; Siegbahn, P. E. M. Quantifying the Effects of the Self-Interaction Error in DFT: When Do the Delocalized States Appear? *J. Chem. Phys.* **2005**, *122*, 224103.
- (80) Polo, V.; Kraka, E.; Cremer, D. Electron Correlation and the Self-Interaction Error of Density Functional Theory. *Mol. Phys.* **2002**, *100*, 1771–1790.
- (81) Řezáč, J. Description of halogen bonding in semiempirical quantum-mechanical and self-consistent charge density-functional tight-binding methods. *J. Comput. Chem.* **2019**, *40*, 1633–1642.