JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules

Jan Řezáč,*[,†] Kevin E. Riley,[‡] and Pavel Hobza[†,§]

[†]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, 166 10 Prague, Czech Republic

[‡]Department of Chemistry, Xavier University of Louisiana, 1 Drexel Drive, New Orleans, Louisiana 70125, United States

[§]Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Palacky University, 771 46 Olomouc, Czech Republic

Supporting Information

ABSTRACT: We present a set of 40 noncovalent complexes of organic halides, halohydrides, and halogen molecules where the halogens participate in a variety of interaction types. The set, named X40, covers electrostatic interactions, London dispersion, hydrogen bonds, halogen bonding, halogen−π interactions, and stacking of halogenated aromatic molecules. Interaction energies at equilibrium geometries were calculated using a composite CCSD(T)/CBS scheme where the CCSD(T) contribution is calculated using triple-ζ basis sets with diffuse functions on all atoms but hydrogen. For each complex, we also provide 10 points along the dissociation curve calculated at the CCSD(T)/CBS level. We use this accurate reference to assess the accuracy of selected post-HF methods.

INTRODUCTION

Accurate quantum mechanical (QM) calculations often serve as a benchmark for the testing and development of more approximate computational methods. They are especially important in the study of noncovalent interactions because interaction energies in noncovalent complexes cannot be easily measured experimentally.[1] Computationally, accurate results can be obtained when high-quality correlated QM methods are used. Among them, CCSD(T) is used most often as a benchmark, because it is the most accurate nonempirical method that is applicable to molecules large enough to be of practical interest.

For the development and validation of computational methods, statistics over large numbers of calculations is required. Although it is possible to collect multiple isolated results from the literature, using a self-contained benchmark data set has multiple advantages.[2] A well-prepared data set should provide balanced, statistically meaningful coverage of the problem, providing results obtained with a consistent setup. Once a data set is established, it allows comparison of results across different studies that use it. Finally, using an existing data set is more convenient.

A large body of data concerning interactions of small organic molecules is already available, especially for biomolecular building blocks and model molecules describing the most common noncovalent interactions. Several contributions to reference binding energy data have been made by our group, starting with the S22 data set.[3] Recently, we have published the S66 data set[4] and its extensions to nonequilibrium geometries (S66 × 8 covering varying intermolecular distances[4] and S66a8 covering angular displacements[5]). These data sets were built using our experience with previous reference data sets and were designed to overcome their limitations, such as missing or imbalanced coverage of some interaction types, inconsistent methods used within a single set and others (see the paper on S66 set for details[4]). Our long-term goal is to expand the database of benchmark calculations further by developing new data sets compatible with the S66 data set.

In this work, we present a new set of 40 complexes, named X40, describing noncovalent interactions of molecules containing halogens. The data set was constructed to cover a wide range of interaction types; for each of them, complexes featuring all relevant halogens (fluorine to iodine) were included. The set covers electrostatic interactions, London dispersion, hydrogen bonds, halogen bonding, halogen−π interactions, and stacking of halogenated aromatic molecules. Some of these interactions are specific to (and directly involve) halogens while others are more general, but all of them should be taken into account when complete coverage of this part of chemical space is sought.

Special attention is paid to the phenomena of halogen bonding.[6−8] A halogen bond is an interaction of a region of positive charge on the surface of a halogen atom (called the σ-hole) with an electronegative atom (most often oxygen or nitrogen). The σ-hole lies in the axis of the covalent bond of the halogen atom and is caused by reduced occupation of the respective p orbital (p_z) involved in the σ-bond (compared to p_x and p_y orbitals that do not share electrons in the bond). This electrostatic interaction is complemented by London dispersion. The strength of both these components increases with the size of the halogen due to the increasing polarizability of the atom. Halogen bonds involving fluorine are very rare, and some other interaction mode is preferred in most cases. Halogen bonds of chlorine are also weak, but stable complexes exist. Bromine and iodine halogen bonds are most prominent and can be as strong as hydrogen bonds. In the past decade, halogen bonds have been studied extensively using computational methods,[8−11] and some CCSD(T) results are available in the literature.[12] Although these interactions might seem

Received: July 25, 2012
Published: September 17, 2012

   dx.doi.org/10.1021/ct300647k | J. Chem. Theory Comput. 2012, 8, 4285−4292

Table 1. List of the Complexes in the X40 Data Set, Their Categorization (Interaction Type and Halogen Element Involved) and the Benchmark CCSD(T)/CBS Interaction Energies (in kcal/mol)[a]

	complex	interaction, halogen	$\Delta E^{CCSD(T)/CBS}$
1	methane−F_2	dispersion, F	−0.49
2	methane−Cl_2	dispersion, Cl	−1.08
3	methane−Br_2	dispersion, Br	−1.30
4	methane−I_2	dispersion, I	−1.35
5	fluoromethane−methane	induction, F	−0.75
6	chloromethane−methane	induction, Cl	−0.98
7	trifluoromethane−methane	induction, F	−0.69
8	trichloromethane−methane	induction, Cl	−1.15
9	fluoromethane dimer	dipole−dipole, F	−1.65
10	chloromethane dimer	dipole−dipole, Cl	−1.34
11	trifluorobenzene−benzene	stack, F	−4.40
12	hexafuorobenzene−benzene	stack, F	−6.12
13	chloromethane−formaldehyde	X bond, Cl	−1.17
14	bromomethane−formaldehyde	X bond, Br	−1.72
15	iodomethane−formaldehyde	X bond, I	−2.38
16	trifluorochloromethane−formaldehyde	X bond, Cl	−2.25
17	trifluorobromomethane−formaldehyde	X bond, Br	−3.10
18	trifluoroiodomethane−formaldehyde	X bond, I	−4.08
19	chlorobenzene−acetone	X bond, Cl	−1.49
20	bromobenzene−acetone	X bond, Br	−2.43
21	iodobenzene−acetone	X bond, I	−3.46
22	chlorobenzene−trimethylamine	X bond, Cl	−2.11
23	bromobenzene−trimethylamine	X bond, Br	−3.78
24	iodobenzene−trimethylamine	X bond, I	−5.81
25	bromobenzene−methanethiol	X bond, Br	−2.32
26	iodobenzene−methanethiol	X-bond, I	−3.08
27	bromomethane−benzene	X−π, Br	−1.81
28	iodomethane−benzene	X−π, I	−2.48
29	trifluorobromomethane−benzene	X−π, Br	−3.11
30	trifluoroiodomethane−benzene	X−π, I	−3.91
31	trifluoromethanol−water	H bond (OH−O), F	−9.67
32	trichloromethanol−water	H bond (OH−O), Cl	−10.41
33	HF−methanol	H bond (XH−O), F	−9.59
34	HCl−methanol	H bond (XH−O), Cl	−6.30
35	HBr−methanol	H bond (XH−O), Br	−5.36
36	HI−methanol	H bond (XH−O), I	−3.97
37	HF−methylamine	H bond (XH−N), F	−14.32
38	HCl−methylamine	H bond (XH−N), Cl	−11.42
39	methanol−fluoromethane	H bond (OH−XC), F	−3.89
40	methanol−chloromethane	H bond (OH−XC), Cl	−3.78

[a]Structures of the complexes are shown in Figure 1.

improbable, halogen bonds have been found in various crystalline materials as well as in biomolecular (protein−ligand) complexes, attracting considerable attention.[13,14] Another class of interaction that have been studied only recently are halogen−π interactions, where the σ-hole interacts with π electrons of an aromatic molecule.[15,16]

The X40 data set is constructed analogously to the S66 data set. Both geometries and interaction energies are calculated using the same setup, which ensures full compatibility of these data sets. The X40 × 10 set features 10 points along the dissociation curve of each of the 40 complexes of the X40 data set. Again, it is built analogously to the S66 × 8 data set, but two more points at the shortest distances have been added. These extra points have been very important for testing semiempirical QM methods and in the development of corrections (for halogen bonds) for them.

In this paper, we use the newly developed data set to test the same set of methods that were tested on the S66 data set. The overlap between these data sets is very small, because the main carriers of the interaction are always the halogen atoms, elements not present in the S66 set. A direct comparison of results obtained for the X40 and S66 data sets is therefore a very good indicator of the robustness of the tested methods. This comparison is even more important in the case of empirically parametrized methods; these have been developed on organic molecules (S22, S66 data sets or similar), and their validation on different elements and interaction types is a necessary step toward their general application. For most of the parametrized methods included in the study, such a test was performed for the first time here.

X40 SET

The 40 complexes in the X40 data set are listed in Table 1 and shown in Figure 1. We divide the set into groups of systems featuring noncovalent interactions of different natures. These groups are not based on any quantitative measures but follow common physicochemical classification of the mode of interaction.

London Dispersion. The heavier halogens have large polarizability, and therefore, they can take part in strong dispersion interactions. In order to isolate the dispersion from other effects as much as possible, we use the least polar models, halogen molecules (F_2, Cl_2, Br_2, and I_2) interacting with methane (complexes **1**−**4**).

Induction. The large dipole moment of a halogenated molecule can induce electrostatic response in anther nonpolar molecule. We illustrate this type of interaction on fluoromethane−methane, chloromethane−methane, trifluoromethane−methane, and trichloromethane−methane complexes (complexes **5**−**8**).

Dipole−Dipole Interaction. The large dipole moment of halides leads to a strong interaction in their homodimers. We include fluoromethane and chloromethane dimers (complexes **9** and **10**).

Stacking. A special case of dispersion interaction is stacking of aromatic rings. Fluorinated benzene (we use 1,3,5-trifluorobenzene and hexafluorobenzene) has the quadrupole moment reversed when compared to benzene, which makes its interaction with benzene more favorable in both dispersion and electrostatic components. We include these two complexes in the set as systems **11** and **12**.

Halogen Bonds. Halogen bonds are unique interactions, and we provide model systems for all possible combinations of halogens (Cl, Br, I) and acceptor atoms (O, N, S) (complexes **13**−**26**). The effect of chemical environment on the strength of a halogen bond is demonstrated in the series of complexes with oxygen acceptors (here, formaldehyde and acetone). The weakest interaction is observed in halomethanes while the same halogen bond in aromatic halides (halobenzenes) is stronger. In trifluorohalomethane complexes, the withdrawal of electrons from the atom involved in the halogen bond and a reversal of the dipole moment make the interaction much stronger. Halogen bonds involving nitrogen are demonstrated on complexes of

dx.doi.org/10.1021/ct300647k | J. Chem. Theory Comput. 2012, 8, 4285−4292

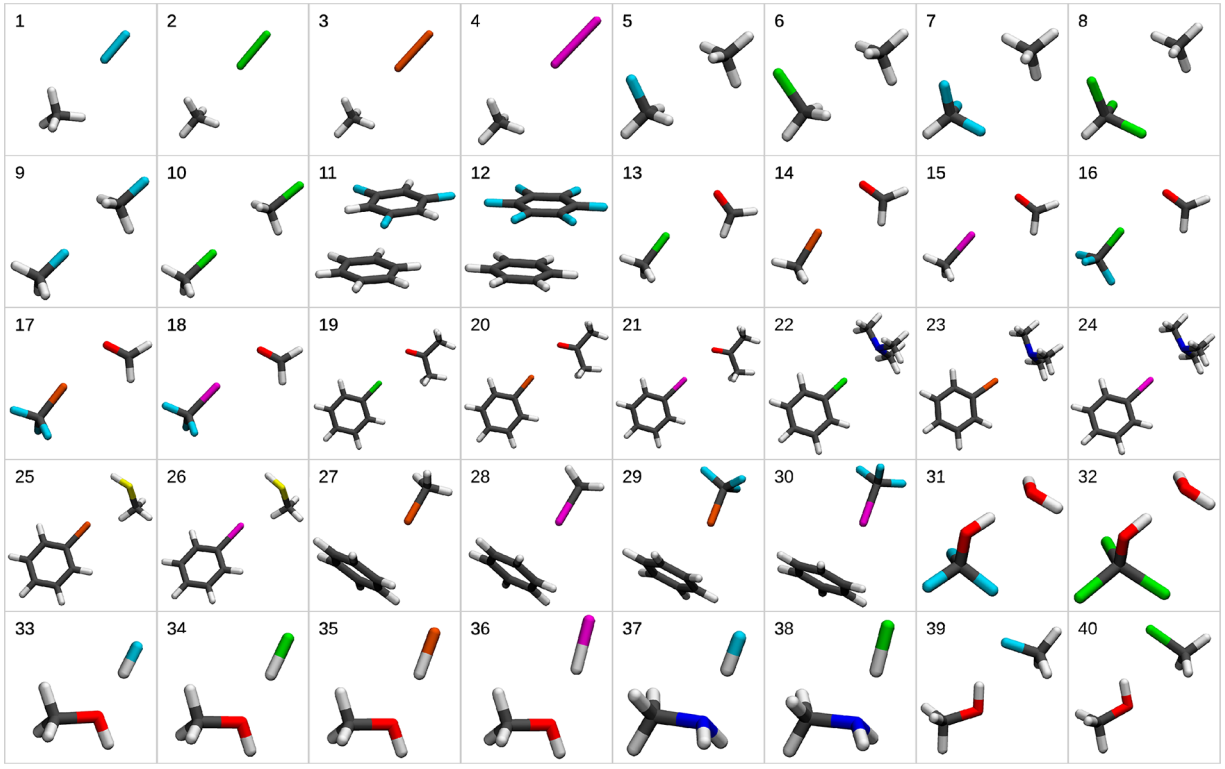

Figure 1. Noncovalent complexes of organic halides, halohydrides, and halogen molecules included in the X40 data set. For names of the complexes and more details see Table 1.

halobenzenes with trimethylamine. Halogen bond with sulfur are shown in complexes of methanethiol with bromo- and iodobenzene; no stable complex with chlorobenzene was found.

Halogen–π Interactions. Heavier halogens (Br and I) interact with aromatic systems via both dispersion and interaction of the σ-hole with the π electrons. These interactions are shown in bromo- and iodomethane–benzene complexes (**27**, **28**). The latter effect can be strengthened by enlarging the σ-hole, such as in trifluorobromomethane and trifluorochloromethane interacting with benzene (complexes **29** and **30**).

Hydrogen Bonds. Halogens can participate in hydrogen bonds or affect the strength of a hydrogen bond indirectly. Several examples of hydrogen-bonding complexes including halogens are included in the set. The first two complexes, trifluoro- and trichloromethanol interacting with water (**31** and **32**) demonstrate the strengthening of a hydrogen bond when the electron density is withdrawn from the hydroxyl group. Hydrogen bonds of halohydrides with organic molecules (methanol and methylamine) are listed as complexes **33**–**38**. Note that bromo- and iodohydride do not form stable hydrogen bonds with methylamine; other modes of interaction are preferred. Finally, hydrogen bonds between methanol and fluoro- and chloromethane are included as complexes **39** and **40**.

X40 × 10 SET

For each of the X40 complexes, a dissociation curve is constructed by scaling the intermolecular distance in the equilibrium geometry obtained with MP2. Eight points are constructed analogously to the S66 × 8 set, but two more points were added at the shortest distance. We found that these additional data are very important for parametrization of the repulsive potential in the development of semiempirical methods. These 10 points (including the MP2 minimum) for each complex form the X40 × 10 set (400 different geometries in total).

METHODS

Basis Sets. The correlation-consistent basis sets of Dunning[17] augmented with diffuse functions[18] are used thoroughout the paper (aug-cc-pVZX, denoted here aXZ, where X = D, T, Q). The largest basis set applicable to the CCSD(T) energies is the "heavy augmented" cc-pVTZ basis sets (denoted haTZ) where the diffuse functions are used for all atoms but hydrogen (cc-pVTZ is used for hydrogen). Such a basis set has been shown to provide results very close to those obtained using the fully augmented one,[19] which is especially true in the present systems where the interactions largely involve heavier atoms.

Pseudopotentials are used for bromine and iodine (with basis sets aug-cc-pVXZ-PP) in order to include relativistic effects and to reduce the size of the calculations.[20] All calculations presented here use the frozen core approximation for the calculation of the correlation energy.

X40 × 10 Geometries. The 40 complexes were optimized at the MP2/cc-pVTZ (cc-pVTZ-PP on Br and I) level with counterpoise correction using tight convergence criteria (energy change 3×10^{-4} kcal/mol (5×10^{-7} au), maximum gradient component 0.06 kcal/mol/Å (5×10^{-5} au), and root-mean-square (rms) gradient 0.03 kcal/mol/Å (2.5×10^{-5} au)). This basis set was found to provide the most favorable error compensation when used with MP2, and the resulting geometries are in good agreement with the CCSD(T)/CBS reference.[21] These geometries then serve for the construction of the dissociation curves. The displacement vector is chosen so that the mutual orientation of the molecules is conserved in the displaced geometries. In most cases, the axis between two

dx.doi.org/10.1021/ct300647k | J. Chem. Theory Comput. 2012, 8, 4285–4292

interacting atoms is used; when a larger moiety contributes to the interaction equally (e.g., benzene ring in halogen−π interactions), the center of mass is used. Along this vector, the closest intermolecular distance in the complex is scaled by factors of 0.8, 0.85, 0.9, 0.95, 1.0, 1.05, 1.1, 1.25, 1.5, and 2.0.

X40 × 10 Benchmark Calculations. On the 400 geometries of the X40 × 10 set, CCSD(T)/CBS interaction energies are calculated as the sum of HF energy calculated with the aQZ basis set, MP2 correlation energy extrapolated[22] from the aTZ and aQZ basis sets, and a ΔCCSD(T) term ($E^{CCSD(T)} - E^{MP2}$) calculated with aDZ. Counterpoise corrections are used for all calculations, and density fitting is used to accelerate MP2 calculations.

X40 Geometries. For the X40 set itself, accurate geometries are obtained by interpolation of the minimum of the CCSD(T)/CBS dissociation curves. Five points around the minimum are fitted with a fifth-order polynomial to obtain the CCSD(T)/CBS intermolecular distance, and the MP2 geometry is scaled to obtain approximate CCSD(T)/CBS equilibrium.[4] Note that the X40 geometry of a complex is not identical but often very close to the MP2 minimum in the X40 × 10 set (scaling factor 1.00). Therefore, merging the X40 and X40 × 10 sets would lead to a redundancy.

X40 Benchmark Calculations. For the core X40 set, we use a more accurate CCSD(T)/CBS scheme. The HF and MP2 energies are calculated as described above, but the ΔCCSD(T) correction is calculated using the haTZ basis sets. The improvement of accuracy toward the CCSD(T) complete basis set limit (compared to the scheme used in X40 × 10 set) over the X40 set is 0.06 kcal/mol (root-mean-square difference). Note that a more complex setup was used in the S66 data set, where this term had been extrapolated from the haDZ and haTZ bais sets. Recently, we have found that identical or even slightly better results are obtained without the extrapolation.[23] These differences are negligible for the comparison between these data sets (rms difference to extrapolated results is 0.03 kcal/mol over the X40 data set).

Interaction Energy Decomposition, Molecular Properties. Interaction energy decomposition is performed for selected complexes by the means of density functional theory based symmetry adapted perturbation theory[24,25] (DFT-SAPT). Calculations were performed at the PBE0AC/aTZ level using density fitting and frozen core approximation. As we are mainly interested in the amount of dispersion in a given interaction, we compare the dispersion energy E^{D} defined as a sum of the dispersion term and the associated exchange with the sum of the remaining terms E^{noD} (electrostatics, induction, exchange, and its higher order components in the δHF term). Electric quadrupole moments of selected monomers were calculated on symmetrized geometries at the MP2/aTZ level.

Tested Methods. Where applicable, the results were extrapolated to the complete basis set limit analogously to the CCSD(T)/CBS calculations: the HF energy is calculated with the aQZ basis set and the MP2 correlation energy is extrapolated from aTZ and aQZ basis sets. Higher order terms of the correlation energy (defined as $E^{corr} - E^{MP2}$) are calculated in the aDZ basis set. A major part of the methods have been tested on the S66 data set in the original paper,[4] please refer to this for more details. These methods are MP2 (including calculations in smaller basis sets where there is favorable error cancellation), spin component scaled MP2[26] (SCS-MP2) and its variant optimized for noncovalent interaction SCS-MI-MP2,[27] MP3 and its scaled variant MP2.5,[28] and CCSD and its spin component scaled versions SCS-CCSD[29] and SCS-MI-CCSD.[30] Additionally, we include two examples of more efficient methods. Among density functional theory (DFT) methods, we have chosen Grimme's advanced dispersion-corrected approach[31] (DFT-D3) with the Becke−Johnson damping function[32] in a large basis set def2-QZVP. Various DFT-D approaches have been tested on the S66 data set, and this setup based on BLYP performed best among the standard functionals.[33] We also briefly test two common hybrid functionals, B3LYP and PBE0, with the same dispersion correction. Finally, we include a semiempirical method PM6[34] with empirical correction for noncovalent interactions[35] (D3 dispersion, H4 hydrogen bonding) including a correction for halogen bonds[36] with updated parameters.[37] It yields very good results for the interactions for which it was corrected, and the X40 data set provides further testing on different systems.

Computational Details. All the correlated and DFT-SAPT calculations presented here were carried out using the Molpro 2010 package.[38] Turbole 6.3[39] with in-house implementation of the D3 dispersion correction was used for the DFT calculations.

Data Availability. Geometries of the complexes, the benchmark CCSD(T)/CBS interaction energies, as well as the results of the tested methods are available in the Supporting Information and in the online database BEGDB (www.begdb.com).[40]

RESULTS AND DISCUSSION

Benchmark Calculations. Geometries and benchmark interaction energies for both X40 and X40 × 10 data sets are provided in the Supporting Information and published online in the BEGDB database. Table 1 lists the complexes in the S40 data set and the benchmark interaction energies.

We will discuss several halogen-specific phenomena that can be observed in the table. The complexes of Br and I in the dispersion group (complexes **3** and **4**) exhibit rather large interaction energies for such small, neutral systems (the interaction energy for the methane dimer is −0.5 kcal/mol). This is caused by the large polarizability of the halogen atoms.

The series of stacked benzene dimers (complex **24** in the S66 data set, interaction energy −2.72 kcal/mol) and the benzene−trifluoro- and hexafluorobenzene (complexes **11** and **12**) demonstrates the effect of the first nonzero electric moment, a quadrupole, on stacking (the quadrupole moments of benzene, trifluorobenzene, and hexafluorobenzene are −6.6, 0.6, and 7.4 au, respectively).[41,42] The most obvious feature is the increase of the strength of the interaction. It has been discussed previously on a symmetric (D_{6h} or D_{3h}) geometry of the complexes,[43] where the electrostatic interaction in the benzene dimer is repulsive and becomes attractive only after the quadrupole of one of the molecules is reversed. The fully optimized structures used in our data set are all parallel-displaced, which makes the electrostatic interaction weakly attractive even in the benzene dimer. Nevertheless, the interaction gets stronger in the fluorinated complexes. Also, the parallel displacement decreases as in the series. Finally, the intermolecular distance (measured as a distance between a benzene ring plane and a closest carbon atom of the other ring) decreases in the series (3.50, 3.34, and 3.33 Å). The strong interaction in the fluorinated complexes forces the molecules to distances that would normally be repulsive (the sum of the C−C van der Waals radii is 3.5 Å).

Among the halogen bonds, there are two main trends in the strength of the interaction. First, it depends on the halogen element. With increasing size (and thus polarizability), both the most important components of the interaction, dispersion and

electrostatics (due to the σ-hole), get stronger. Second, the electrostatic component is also affected by the properties of the rest of the molecule. Electron-withdrawing moieties (benzene and trifluoromethane in our set) enlarge the σ-hole, making the interaction stronger, while electron-donating groups (methane in our set) act in the opposite direction. In the extreme case of trifluoromethane, the dipole moment of the molecule is reversed and thus aligned with the dipole of the other molecule in the complex, which leads to further stabilization of the complex. Halogen$-\pi$ interactions (complexes **27**$-$**30**) are very similar to halogen bonds; the halogen atom interacts via the σ-hole with an electronegative region, the π-electron system.[15,16] Effects analogous to the ones described above apply here as well. However, the interactions are weaker, and stable complexes of this type are formed only by bromine and iodine. In the case of chlorine, optimization yields a dispersion-bound complex where the halogen does not form a halogen$-\pi$ bond.

The group of hydrogen bonds features three different classes of interaction. In the first one, the effect of halogen substitution is demonstrated on OH$-$O hydrogen bonds (complexes **31** and **32**). The halogens (F and Cl) withdraw electrons from the OH group, making the interaction substantially stronger (interaction energy in the methanol$-$water complex in the S66 data set is -5.1 kcal/mol). Surprisingly, the effect of chlorine is larger than that of fluorine, although the electronegativity of the elements suggests otherwise. This is caused by secondary interactions of the larger chlorine atoms with the water molecule, which are of a dispersion character (the DFT-SAPT dispersion energy E^{D} grows from -3.8 to -5.0 kcal/mol while the remaining terms E^{noD} only change from -5.3 to -4.9 kcal/mol when passing from F to Cl). There are two series of hydrogen bonds where a halohydride serves as a hydrogen donor in an H bond. In the first one, it interacts with oxygen in methanol (complexes **33**$-$**36**). The strength of the interaction decreases from HF to HI as expected. The same applies for the two hydrogen bonds with methylamine acceptors. DFT-SAPT shows that these hydrogen bonds are substantially stronger than their counterparts with oxygen receptors mainly due to the increase of the induction term. Finally, there are two hydrogen bonds where F and Cl bound to carbon serve as a hydrogen-bond acceptor (complexes **39** and **40**). Here, the interaction with fluorine is stronger as it is more electronegative.

Methods Tested on the X40 Set. Table 2 lists root-mean-square error (RMSE) and mean unsigned error (MUE) values for all the tested methods. As a more universal measure of accuracy, we also list a relative error calculated as RMSE divided by the average magnitude of interaction energy in the data set, expressed in percent. The RMSE values of the same methods for the S66 are listed for comparison. The relative errors in X40 and S66 data sets are plotted in Figure 2. Interaction energies calculated by all the methods tested are listed in the Supporting Information in Table S1.

We also list the errors separately for the groups of interactions of the same type. These interactions often vary in strength; to make the results comparable, we use RMSE error in percent, relative to the average interaction energy in the group. One group named "electrostatic" is used for both dipole$-$dipole and dipole$-$induced dipole interactions. These results are summarized in Figure 3, and particular cases are highlighted in the following discussion. It must be noted that, in the case of weaker interactions, these relative errors translate to rather small absolute

Table 2. Root-Mean-Square Error (in kcal/mol and in % of Average Interaction Energy in the Set) and Mean Unsigned Error (in kcal/mol) of the Tested Methods in the X40 Data Set. RMSE in the S66 data set is listed for comparison.

	X40 set			
method	RMSE	RMSE (%)	MUE	S66 set RMSE
SCS-MI-CCSD/CBS	0.06	2	0.05	0.08
SCS-CCSD/CBS	0.19	5	0.12	0.25
CCSD/CBS	0.48	13	0.40	0.70
MP2.5/CBS	0.18	5	0.12	0.16
MP3/CBS	0.45	12	0.32	0.62
SCS-MI-MP2/CBS	0.39	10	0.26	0.38
SCS-MP2/CBS	0.40	11	0.35	0.87
MP2/CBS	0.70	19	0.43	0.69
MP2/TZ	0.58	15	0.50	0.70
MP2/aDZ	0.55	15	0.43	0.79
BLYP-D3(BJ)/def2-QZVP	0.39	10	0.29	0.25
PM6-D3H4X[a]	0.62	26	0.48	0.68
PM6-D3H4X	2.31	61	1.11	0.68

[a]Hydrogen bonds removed from the X40 set.

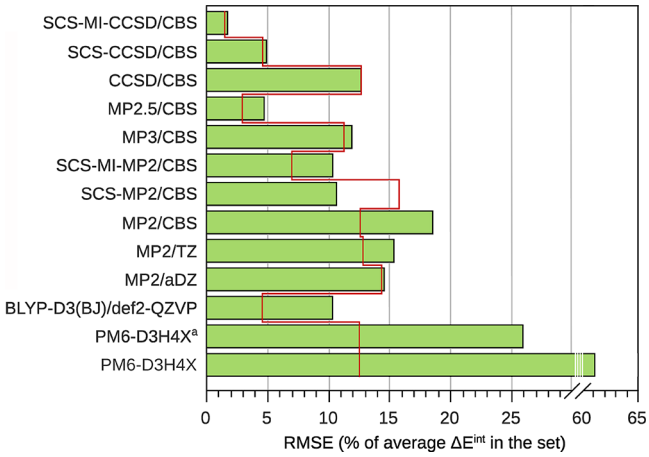

Figure 2. RMSEs (relative, in % of average interaction energy in the set) of the tested methods over the X40 data set (plotted as bars). S66 data set results (shown as a red line) are provided for comparison. Superscript "a" indicates that hydrogen bonds were removed from the X40 set.

ones (average interaction energies in the groups are as follows: dispersion, -1.05; electrostatic, -1.09; stacking, -5.26; halogen bonds, -2.80; halogen$-\pi$, -2.83; and hydrogen bonds, -7.87).

The most computationally demanding class of the tested methods is based on a CCSD calculation. Pure CCSD is known to underestimate the strength of noncovalent interactions, and here, it yields rather large errors (RMSE 0.48 kcal/mol, 13%). Spin component scaled CCSD, SCS-CCSD,[29] improves the results significantly (RMSE 0.19 kcal/mol, 5%), but the best results can be achieved when the scaling is optimized for noncovalent interactions. The SCS-MI-CCSD method[30] yields an RMSE of 0.06 kcal/mol (2%); such a small error is close to the estimated accuracy of the CCSD(T)/CBS reference. The relative errors of the CCSD-based methods are practically identical to their counterparts in the S66 set. This is especially important in the case of the SCS-MI-CCSD method, which was parametrized on the S22 set (similar to S66) that covers only H, C, N, and O elements. Successful validation on the X40 set

dx.doi.org/10.1021/ct300647k | J. Chem. Theory Comput. 2012, 8, 4285−4292

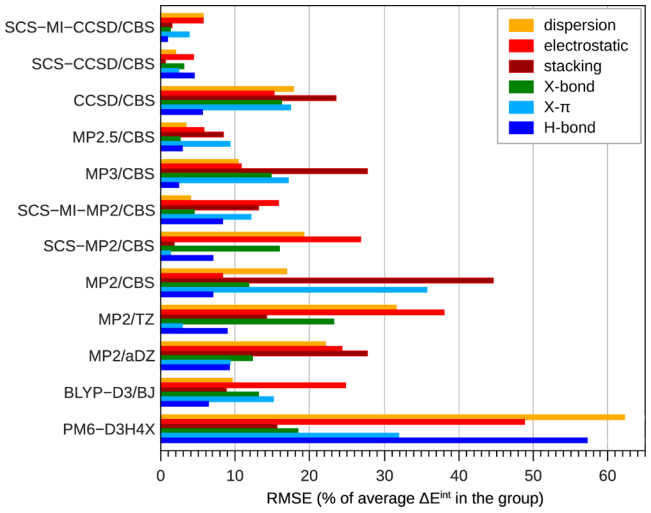

Figure 3. Relative errors for different interaction types in the X40 data set.

that features very different systems is an important proof of robustness and transferability of the method.

One step lower in terms of calculation complexity are methods based on MP3. Again, MP3 itself does not perform very well (RMSE 0.45 kcal/mol, 12%). Especially the error in stacked complexes is very large (see Figure 3), but in contrast to MP2, the strength of the interaction is underestimated. When the MP3 term ($E^{MP3} - E^{MP2}$) is scaled by one-half in the MP2.5 method,[28] the results are much better: RMSE decreases to 0.18 kcal/mol, which is 5% of the average interaction energy. Compared to S66, MP2.5 performs slightly worse here, but it remains the second most accurate method in our test after SCS-MI-CCSD. As the computational expense is significantly reduced compared to CCSD(T) and CCSD, MP2.5 still offers very good accuracy to cost ratio and can be recommended for many applications.

MP2 exhibits the largest errors from the correlated methods. In the complete basis set limit, it overestimates the strength of the interactions with RMSE 0.70 kcal/mol (19%). The error is largest in stacked complexes and halogen−π interactions. This effect is known from stacking of organic molecules, but our calculations suggest that it is more common feature of interactions involving aromatic π systems. When a smaller basis set is used, error compensation reduces the errors slightly to about 15%. When diffuse functions are used (in the aDZ basis set), stacking is still overestimated. When a basis set of comparable size but without diffuse functions is used (TZ), the nature of the errors changes, and now, the largest relative errors are found in the weakly bound dispersion and electrostatic complexes. Spin component scaling (both SCS-MP2 and SCS-MI-MP2) improves the results more to about 10%. Here, we observe the largest difference to the S66 set, where SCS-MI-MP2, a method optimized for interactions on the S22 set, performs substantially better than SCS-MP2 that had been optimized for thermochemistry applications. SCS-MP2 almost eliminates the error in stacks, but SCS-MI-MP2 provides a more balanced description of all the groups.

In applications where correlated wave function calculations are not feasible, density functional theory augmented with dispersion correction (DFT-D) is the most reliable tool providing a reasonable description of noncovalent interactions. There are many options to choose from, as the selection of DFT functional, basis set, and dispersion correction scheme affect the results. We have chosen the D3 correction by Grimme,[31,32] as it is currently the most advanced dispersion correction available. It has been tested extensively on the S66 and S66 × 8 data sets[33] with very good results. From the functionals listed in this study, we use B-LYP in a large def2-QZVP basis. It yields the lowest error in the S66 set among the GGA functionals tested and is very close to the best more advanced functionals. However, the error in the X40 set (0.39 kcal/mol, 10%) is larger than in S66 (0.25 kcal/mol, 5%). In some of the complexes, a GGA functionals might introduce artificial charge transfer resulting in overestimated interaction energy due to the self-interaction error. We have tested hybrid functionals that should be more robust in such cases, but no improvement is observed. B3LYP-D3 yields the same RMSE as BLYP-D3 (0.39 kcal/mol), and PBE0 performs worse with error of 0.54 kcal/mol.

The most efficient quantum-mechanical method tested is the semiempirical method PM6. Used alone, it does not describe noncovalent interactions well, but with corrections for dispersion and hydrogen bonding[35] (PM6-D3H4), it yields surprisingly accurate results for organic molecules (RMSE in S66 is 0.68 kcal/mol). It has also been corrected for halogen bonding,[36] resulting in the PM6-D3H4X method used here. The question is how this method describes other interactions involving halogens and variants of halogen bonds it was not trained on. The overall error in the X40 data set is very large (2.3 kcal/mol). A major source of this error is hydrogen bonds involving halogens, because these are not corrected by the H4 correction that is applied only to oxygen and nitrogen atoms. Extension of this correction is possible but beyond the scope of this paper. Instead, we report the error in the X40 set with hydrogen bonds removed (listed under PM6-D3H4X[a] in Table 2 and Figure 2). This error is acceptably small (0.62 kcal/mol), but in relative measure, it amounts to 26%, which is about twice as much as in the S66 data set (12%). Halogen bonds are described well due to the specific correction, as is stacking. The other interactions where halogen atoms are directly involved exhibit large relative errors. Nevertheless, the efficiency of this method combined with reasonable accuracy for commonly encountered interactions makes corrected PM6 a useful tool for study of very large systems that cannot be studied by more demanding methods. We have applied it successfully in computer-aided drug design, where this method yields results that agree well with experimental evidence.[37,44]

X40 × 10 Set. We have not tested different methods on the X40 × 10 set. Based on the analogy with S66 and S66 × 8 sets,[4] we do not expect substantial differences for the high-level methods. The main purpose of the X40 × 10 set is parametrization of novel methods; our work in this direction will be published separately. Here, we will only discuss general findings based on the analysis of the CCSD(T)/CBS dissociation curves.

First, we can compare the optimized MP2 geometries with the CCSD(T)/CBS intermolecular distances obtained from interpolation of the X40 × 10 dissociation curves. The closest intermolecular distances compared here are provided in the Supporting Information, Table S2. Surprisingly, the largest errors are found in the weakest complexes. In the trifluoromethane−methane complex, the distance is overestimated by 0.23 Å (7% of the reference intermolecular distance). This is connected with a rather large error in the interaction energy, which is −0.69 kcal/mol at the CCSD(T)/CBS level but only −0.23 kcal/mol in the MP2/TZ calculation. One would expect large errors in the stacked complexes where MP2 has problems, but the MP2 distance in the trifluorobenzene−benzene complex is shorter by only 0.06 Å (1.8%), and the error in hexafluorobenzene−benzene

complex is even smaller. This can be attributed only to error compensation in the TZ basis set; MP2/CBS distances would be shorter as shown in the S66 × 8 data set.[21]

When the intermolecular distance is scaled by 0.9, which is the shortest distance used in the S66 × 8 set, none of the interactions in the X40 complexes is repulsive (has positive interaction energy). For our work on semiempirical methods (which generally fail to describe the interaction at short distances), we had to add the extra two points with scaling 0.85 and 0.8. At the latter point, all the complexes but hydrogen bonds became repulsive. The highest energies were found in the halogen bonds with sulfur acceptors (complexes **25** and **26**).

CONCLUSIONS

The X40 and X40 × 10 data sets represent an important extension of available benchmark data on noncovalent interactions. The sets are compatible with the recently published S66 database and its extensions and can be used either together when coverage of halogen-specific interaction is needed or independently as a validation set with very small overlap with S66 or other data on organic molecules.

Both geometries and interaction energies in the X40 set are calculated at state-of-the-art levels: intermolecular distances are interpolated from CCSD(T)/CBS dissociation curves and interaction energies on this geometry are calculated using a CCSD(T)/CBS scheme that includes a CCSD(T) calculation with a heavy-augmented cc-pVTZ basis set, which in the case of the complex in the set, is close to the limit of available computers.

Multiple computational methods have been tested on the X40 set. It is the first study where the methods are tested on a large and diverse set of complexes that is very different from the commonly used data on organic molecules. The results are positive as all the methods perform similarly here.

These findings are especially important in the case of parametrized methods, such as SCS-MI-CCSD[30] and MP2.5.[28] These methods offer the best accuracy to cost ratio and significantly outperform their parameter-free counterparts. As these methods were parametrized to reproduce interaction energies in complexes of organic molecules not containing halogens, it is important proof of their transferability and robustness that they perform equally well on different classes of interactions.

The X40 and X40 × 10 data sets are intended as tools for further development of efficient methods describing noncovalent interactions in a wide variety of systems. To make the use of these sets more convenient, we have published all our results in the BEGDB database (www.begdb.com).[40]

ASSOCIATED CONTENT

Supporting Information

Geometries of the X40 and X40 × 10 complexes and tables of the benchmark CCSD(T)/CBS interaction energies, interaction energies obtained with all the tested methods on the X40 data set, and the closest intermolecular distances in the MP2 optimized structures and the values interpolated from the CCSD(T)/CBS dissociation curves. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Fax: +420 220 410 320; e-mail: rezac@uochb.cas.cz.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was a part of Research Project RVO:61388963 of the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, and was supported by grant P208/12/G016 from the Czech Science Foundation. It was also supported by the operational program Research and Development for Innovations of European Social Fund (CZ.1.05/2.1.00/03.0058). The support of Praemium Academiae, Academy of Sciences of the Czech Republic, awarded to P.H. in 2007 is also acknowledged.

REFERENCES

(1) Hobza, P.; MüllerDethlefs, K. *Non-Covalent Interactions: Theory and Experiment*; Royal Society of Chemistry: Cambridge, U.K., 2010.
(2) Hobza, P. *Acc. Chem. Res.* **2012**, *45*, 663−672.
(3) Jurečka, P.; Šponer, J.; DČerný, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.
(4) Řezáč, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2011**, *7*, 2427−2438.
(5) Řezáč, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2011**, *7*, 3466−3470.
(6) Metrangolo, P.; Resnati, G. *Chem.—Eur. J.* **2001**, *7*, 2511−2519.
(7) Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16789−16794.
(8) Politzer, P.; Murray, J. S.; Clark, T. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7748−7757.
(9) Brinck, T.; Murray, J.; Politzer, P. *Int. J. Quantum Chem.* **1992**, 57−64.
(10) Clark, T.; Hennemann, M.; Murray, J. S.; Politzer, P. *J. Mol. Model.* **2007**, *13*, 291−296.
(11) Murray, J. S.; Riley, K. E.; Politzer, P.; Clark, T. *Aust. J. Chem.* **2010**, *63*, 1598−1607.
(12) Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2008**, *4*, 232−242.
(13) Metrangolo, P.; Neukirch, H.; Pilati, T.; Resnati, G. *Acc. Chem. Res.* **2005**, *38*, 386−395.
(14) Lu, Y.; Shi, T.; Wang, Y.; Yang, H.; Yan, X.; Luo, X.; Jiang, H.; Zhu, W. *J. Med. Chem.* **2009**, *52*, 2854−2862.
(15) Matter, H.; Nazaré, M.; Güssregen, S.; Will, D. W.; Schreuder, H.; Bauer, A.; Urmann, M.; Ritter, K.; Wagner, M.; Wehner, V. *Angew. Chem., Int. Ed.* **2009**, *48*, 2911−2916.
(16) Lu, Y.; Wang, Y.; Zhu, W. *Phys. Chem. Chem. Phys.* **2010**, *12*, 4543.
(17) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.
(18) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1994**, *100*, 2975.
(19) ElSohly, A. M.; Tschumper, G. S. *Int. J. Quantum Chem.* **2009**, *109*, 91−96.
(20) Peterson, K. A.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. *J. Chem. Phys.* **2003**, *119*, 11113−11123.
(21) Řezáč, J.; Riley, K. E.; Hobza, P. *J. Comput. Chem.* **2012**, *33*, 691−694.
(22) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243−252.
(23) Řezáč, J.; Hobza, P., in preparation.
(24) Williams, H. L.; Chabalowski, C. F. *J. Phys. Chem. A* **2001**, *105*, 646−659.
(25) Hesselmann, A.; Jansen, G. *Chem. Phys. Lett.* **2002**, *357*, 464−470.
(26) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095.
(27) Distasio, R.; Head-Gordon, M. *Mol. Phys.* **2007**, *105*, 1073−1083.
(28) Pitoňák, M.; Neogrády, P.; Černý, J.; Grimme, S.; Hobza, P. *ChemPhysChem* **2009**, *10*, 282−289.

(29) Takatani, T.; Hohenstein, E. G.; Sherrill, C. D. *J. Chem. Phys.* **2008**, *128*, 124111.
(30) Pitoňák, M.; Řezáč, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 9611.
(31) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
(32) Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(33) Goerigk, L.; Kruse, H.; Grimme, S. *ChemPhysChem* **2011**, *12*, 3421−3433.
(34) Stewart, J. J. P. *J. Mol. Model.* **2007**, *13*, 1173−1213.
(35) Řezáč, J.; Hobza, P. *J. Chem. Theory Comput.* **2012**, *8*, 141−151.
(36) Řezáč, J.; Hobza, P. *Chem. Phys. Lett.* **2011**, *506*, 286−289.
(37) Brahmkshatriya, P.; Dobeš, P.; Přenosil, O.; Řezáč, J.; Fanfrlík, J.; Paruch, K.; Lepšík, M.; Hobza, Pavel *J. Comput.-Aided Mol. Des.* **2012**, under revision.
(38) Werner, H.-J.; Knowles, P. J.; Manby, F. R.; Schütz, M.; et al. *MOLPRO*, version 2010.1, a package of ab initio programs; University College Cardiff Consultants Limited: Cardiff, Wales, U.K., 2010.
(39) *TURBOMOLE v6.3*; University of Karlsruhe and Forschungszentrum Karlsruhe GmbH: Karlsruhe, Germany, 2011.
(40) Řezáč, J.; Jurečka, P.; Riley, K. E.; Černý, J.; Valdes, H.; Pluháčková, K.; Berka, K.; Řezáč, T.; Pitoňák, M.; Vondrášek, J.; Hobza, P. *Collect. Czech. Chem. Commun.* **2008**, *73*, 1261−1270.
(41) Williams, J. H. *Acc. Chem. Res.* **1993**, *26*, 593−598.
(42) West, A. P.; Mecozzi, S.; Dougherty, D. A. *J. Phys. Org. Chem.* **1997**, *10*, 347−350.
(43) Pluháčková, K.; Jurečka, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2007**, *9*, 755.
(44) Dobeš, P.; Řezáč, J.; Fanfrlík, J.; Otyepka, M.; Hobza, P. *J. Phys. Chem. B* **2011**, *115*, 8581−8589.