A simple, efficient and universal energy decomposition analysis method based on dispersion-corrected density functional theory

Tian Lu,[a],* Qinxue Chen[a]

[a] Beijing Kein Research Center for Natural Sciences, Beijing 100024, P. R. China (http://www.keinsci.com)

* Correspondence author. E-mail: sobereva@sina.com

Author information

Tian Lu ORCiD: 0000-0002-1822-1229 E-mail: sobereva@sina.com

Qinxue Chen: ORCiD: 0000-0003-0155-2387 E-mail: qinxue_chen@sina.com

Abstract: Energy decomposition analysis (EDA) is an important method to explore the nature of interaction between fragments in a chemical system. It can decompose the interaction energy into different physical components to understand the factors that play key roles in the interaction. This work proposes an energy decomposition analysis strategy based on dispersion-corrected density functional theory (DFT), called sobEDA. This method is fairly easy to implement and very universal. It can be used to study weak interactions, chemical bond interactions, open-shell systems, and interactions between multiple fragments. The total time consumption of sobEDA is only about twice that of conventional DFT calculation for the entire system. This work also proposes a variant of the sobEDA method named sobEDAw, which is designed specifically for decomposing weak interaction energies. Through a proper combination of DFT correlation energy and dispersion correction term, sobEDAw gives a ratio between dispersion energy and electrostatic energy that is highly consistent with the symmetry-adapted perturbation theory (SAPT), which is quite popular and robust in studying weak

interactions but expensive. We present a shell script *sobEDA.sh* to implement the methods proposed in this work based on the very popular Gaussian quantum chemistry program and Multiwfn wavefunction analysis code. Via the script, theoretical chemists can use the sobEDA and sobEDAw methods very conveniently in their study. Through a series of examples, the rationality of the new methods and their implementation are verified, and their great practical values in the study of various chemical systems are demonstrated.

Keywords: Energy decomposition analysis; Density functional theory; intermolecular interaction; Quantum chemistry; Gaussian; Multiwfn

1. Introduction

Energy decomposition analysis (EDA) is a very important class of methods to investigate interaction between different fragments in a chemical system. By decomposing the total interaction energy into different easily understandable physical components, chemists can readily recognize which factors play a major and minor role in the interaction of interest, this is evidently very important in understanding the nature of the interaction. EDA methods have been widely employed in theoretical chemical studies.[1-6]

There are many EDA methods in the field of quantum chemistry. The present article will not make a comprehensive overview, interested readers are recommended to read relevant reviews.[7-12] Briefly speaking, interfragment EDA can be mainly divided into two categories. The methods in the first category are dependent of variationally solved orbitals of every fragment and entire system, such as Kitaura-Morokuma,[13] LMOEDA,[14] GKS-EDA[15], NEDA[16] and ALMO-EDA.[17,18] The second category includes the methods that only depend on fragment wavefunctions, the most representative one is the symmetry-adapted perturbation theory (SAPT) method,[19,20] which rigorously derive various interaction components between molecules through intermolecular perturbation theory. There are also many other forms of EDA. For example, Mayer's energy decomposition calculates various interaction terms between atoms based on electron integrals and density matrix of the entire system.[21] The EDA-FF method proposed by us provides interaction information between atoms or fragments based on forcefields and atomic charges.[22] ETS-NOCV method can further decompose orbital interaction into components corresponding to different features and make them visible.[23] The scheme proposed by Liu decomposes total energy into steric, electrostatic, and fermionic quantum effects to explore the physical essence that causes the difference in total energy between different status.[24]

Although, as mentioned above, there are many interfragment EDA methods, each of them has its own limitations. For example, Kitaura-Morokuma is defined based on the now obsolete Hartree-Fock (HF) method. The use of LMOEDA and GKS-EDA

mainly depends on the modified version of the GAMESS-US program,[25,26] whose use is relatively complicated, and the source code needs to be compiled by users, which is somewhat difficult for junior researchers in quantum chemistry. ALMO-EDA is only implemented in the commercial Q-Chem program,[27] thus limiting its accessibility. The well-known SAPT is mostly suitable for the study of weak interactions between two molecules but can hardly study the nature of chemical bonds. In addition, the accurate high-order SAPT realizations such as SAPT2+(3)[20] are fairly expensive and implementation for open-shell systems are not available in most codes.

In this article, we introduce a new energy decomposition method, called sobEDA, which is defined based on the density functional theory (DFT) with dispersion correction. This method is easy to implement, low in time consumption, and has very good universality. In addition, we also propose sobEDAw method for the study of weakly interacting systems, which redistributes the DFT correlation and dispersion correction terms in sobEDA via a carefully designed switching function. It was found that sobEDAw can ideally reproduce the ratio between dispersion and electrostatic interactions produced by the very popular and widely accepted SAPT method. We will also introduce a shell script named *sobEDA.sh*, which was developed by us to implement sobEDA and sobEDAw based on the Gaussian quantum chemistry package[28] and our freely available Multiwfn wavefunction analysis program[29]. Given that this script is convenient to use, and the Gaussian program is fast, robust, and has a very wide user base, we believe the methods proposed in this work will become popular; and to a certain extent, they will make up for the shortcomings of other EDA methods mentioned above. A series of application examples will be given later to prove the rationality and important practical values of the sobEDA and sobEDAw methods.

2 sobEDA method

First we stress that sobEDA is fully defined on the top of dispersion-corrected Kohn-Sham DFT (KS-DFT) energy, which is composed of non-interacting electron kinetic energy, classic electrostatic energy, exchange energy (by exchange functional

and possibly with some portion of HF exchange), DFT correlation energy (by correlation functional), and dispersion correction energy (via DFT-D3,[30] DFT-D4,[31] VV10[32] and so on). Note that double-hybrid functionals are not considered in the present work. The definition of physical components of interfragment interaction in sobEDA comes from a simple physical picture of step-by-step formation of wavefunction of the final state of the entire system from wavefunctions of isolated fragments. Fig. 1 portrays full variation of energy of the entire system during the formation procedure, and the definition of sobEDA terms is directly exhibited in the figure. It can be seen that the formation of the final state consists of three stages:

1. The individual fragments (in complex geometry) first form a promolecule state of the entire system, whose occupied orbitals simply correspond to union of occupied fragment orbitals, and wavefunction of the promolecule state can be viewed as Hartree product of the fragment wavefunctions. The variation of electrostatic energy, exchange energy, correlation energy and dispersion correction in this stage directly defines ΔE_{els}, ΔE_{x}, ΔE_{DFTc} and ΔE_{dc} terms of sobEDA method, respectively. Note that in this stage kinetic part of system energy keeps unchanged, because KS-DFT framework evaluates kinetic energy of non-interacting electrons as sum of expectation values of kinetic operator for occupied orbitals, which are completely unperturbed in this stage.
2. Promolecule state transforms to frozen state to fulfill Pauli exclusion principle, in which all orbitals form an orthonormal set. Frozen state wavefunction can be easily generated by orthogonalization of occupied orbitals in promolecule wavefunction. The total energy variation in this stage defines Pauli repulsion term (ΔE_{rep}) in sobEDA.
3. In the last stage, by allowing occurrence of intrafragment electron polarization and charge transfer between fragments, frozen state relaxes to final state. The corresponding total energy variation corresponds to orbital interaction term (ΔE_{orb}) in sobEDA.

In the present implementation of sobEDA, only the mostly popular DFT-D3(BJ)[33]

dispersion correction[33] is considered. Since it is only dependent of geometry and fully irrelevant to wavefunction, therefore change in dispersion correction energy in the last two stages, namely ΔD_2 and ΔD_3 in Fig. 1, is exactly zero. Other dispersion correction scheme such as DFT-D4[31] and VV10[32] are also in principle compatible with sobEDA. Another worth-mentioning point is that the definition of sobEDA is directly applicable to decomposition of interaction among more than two fragments.

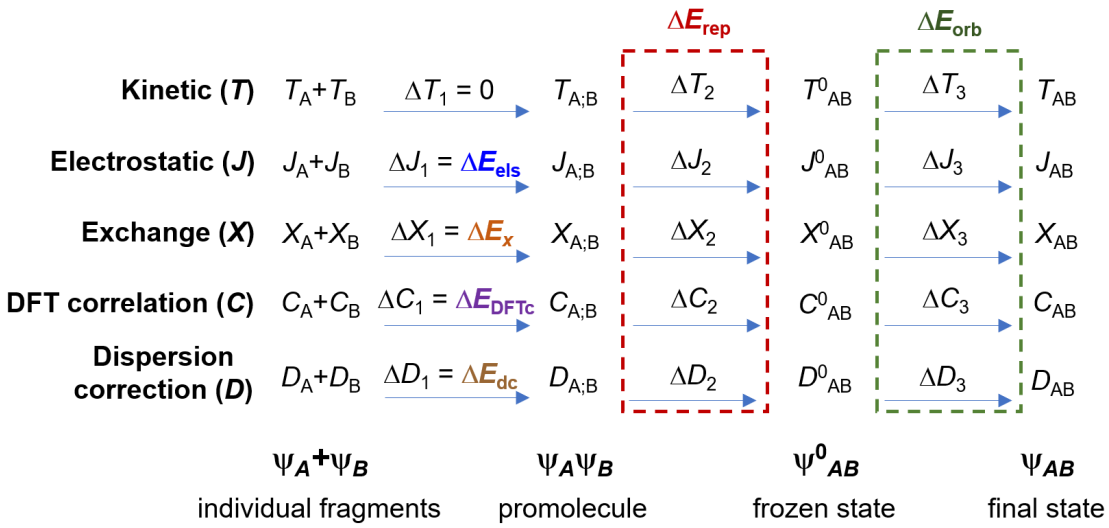

Fig. 1 Variation of various terms of dispersion-corrected DFT energy during formation of final state wavefunction of the entire system from individual fragment wavefunctions. Two fragments are assumed here for simplicity. The colored texts indicate the physical components of interaction between fragments *A* and *B* defined by sobEDA method.

As shown above, the complete form of sobEDA represent interaction energy between two or more fragments as six terms, that is

$$\Delta E_{\mathrm{int}} = \Delta E_{\mathrm{els}} + \Delta E_{\mathrm{x}} + \Delta E_{\mathrm{rep}} + \Delta E_{\mathrm{orb}} + \Delta E_{\mathrm{DFTc}} + \Delta E_{\mathrm{dc}} \quad (1)$$

In order to simplify discussion, it is suggested to merge ΔE_{x} and ΔE_{rep} together as exchange-repulsion term, namely $\Delta E_{\mathrm{xrep}} = \Delta E_{\mathrm{x}} + \Delta E_{\mathrm{rep}}$. This term occurs in many EDA methods such as SAPT, Kitaura-Morokuma and ALMO-EDA. Furthermore, if dispersion interaction is not of interest, ΔE_{DFTc} and ΔE_{dc} may be combined as Coulomb correlation term, that is $\Delta E_{\mathrm{c}} = \Delta E_{\mathrm{DFTc}} + \Delta E_{\mathrm{dc}}$. Note that dispersion correction essentially describes the portion of Coulomb correlation that cannot be captured by DFT correlation functionals.

Assume that a quantum chemistry code is able to provide detailed components of total energy, such as Gaussian program, then only five steps are needed to obtain all information needed by deriving the terms in sobEDA:

1. Performing a conventional self-consistent field (SCF) calculation for each fragment until convergence
2. Combining all occupied fragment orbitals to yield promolecule wavefunction
3. Calculating energy of promolecule state
4. Orthogonalizing occupied orbitals in promolecule state to obtain the frozen state wavefunction
5. Using frozen state wavefunction as initial guess to perform conventional SCF calculation for the entire system until convergence. The total energy of the first round of SCF corresponds to the frozen state energy, that of the last round corresponds to final state energy.

The time cost in steps (2) and (4) is negligible. The time consumed in (3) just corresponds to one SCF iteration and hence the cost is also trivial. Typically, the overall cost of sobEDA is merely about two times of regular single point calculation for the entire system. The low cost of sobEDA enabling it to be easily employed for systems consisting of hundreds of and even more than one thousand atoms with a 2-zeta basis set plus polarization functions.

By properly setting up the calculation of fragments and combining their occupied orbitals as promolecule wavefunction in a suitable way, sobEDA can be used for a rich variety of systems, including studying interaction between open-shell fragments. For example, by using sobEDA to study the interaction between CH_3 and NH_2 free radical fragments in CH_3NH_2, the neutral doublet state of the two fragments should be calculated separately. When combining the fragment wavefunctions, the information of alpha and beta orbitals of one of the fragments should be exchanged with each to generate the appropriate promolecule wavefunction. After that, when constructing the frozen state wavefunction, the orthogonalization should be performed separately for occupied orbitals of each spin.

In the calculation of fragments, either the basis functions carried by the fragment atoms (monomer basis, MB) or those carried by all atoms in the entire system (complex basis, CB) may be used. The latter is evidently more expensive and hence MB is suggested to be used by default; however, as will be shown later, the use of CB is important for reliably studying intermolecular weak interactions, especially when the basis set is not diffuse enough. When CB is used instead of MB, the total interaction energy derived by sobEDA is equal to the counterpoise corrected interaction energy,[34] that is $\Delta E_{\mathrm{int}} = E_{\mathrm{complex}}(\mathrm{CB}) - \sum_A E_A(\mathrm{CB})$, and at the mean time, all sobEDA terms except for the DFT-D3(BJ) dispersion correction term will be affected to a certain extent.

The sobEDA method described above does not include effect of geometry distortion of fragments due to interaction, does not taking solvation effect into account, and ignores any thermodynamic effects. However, in some situations these factors play nonnegligible roles. For example, distortion energy of fragments is critical in the analysis of interfragment interaction at a series of points along the path of chemical reactions, as highlighted by the known distortion-interaction analysis;[35] polar solvent environment significantly weakening interaction between solute molecules with obvious polarity; Entropy penalty effect markedly hinders complexation between molecules when temperature is not quite low. Since these factors can be straightforwardly evaluated via any mainstream quantum chemistry as follows by users, they are not explicitly considered in the standard definition of sobEDA.

- Fragment geometry distortion effect: For each fragment, such as fragment *A*, its distortion energy is calculated as $\Delta E_A^{\mathrm{dist}} = E_A@\mathrm{complex} - E_A@\mathrm{isolated}$, where $E_A@\mathrm{complex}$ and $E_A@\mathrm{isolated}$ stand for fragment energy calculated in optimized complex geometry and in isolated fragment geometry, respectively. It is noted that adding the distortion energies of all fragments to interaction energy gives the binding energy in common sense.
- Solvation effect: Influence of solvation effect on interaction energy is evaluated as $\Delta E_{\mathrm{int}}^{\mathrm{solv}} = \left(E_{\mathrm{complex}}^{\mathrm{soln}} - \sum_A E_A^{\mathrm{soln}}\right) - \left(E_{\mathrm{complex}}^{\mathrm{vac}} - \sum_A E_A^{\mathrm{vac}}\right)$, where $E_{\mathrm{complex}}^{\mathrm{soln}}$ and E_A^{soln} denote energy of complex and that of fragment *A* calculated under solvation

model such as SMD,[36] respectively, while $E^{\mathrm{vac}}_{\mathrm{complex}}$ and E^{vac}_{A} denote the counterparts calculated in vacuum. In fact, $\Delta E^{\mathrm{solv}}_{\mathrm{int}}$ can be further decomposed to polar and nonpolar contributions to provide a deeper insight, as SMD defines them separately.

- Thermodynamic effect: For example, free energy correction to interaction is calculated as $\Delta\Delta G^{\mathrm{corr}} = \Delta G^{\mathrm{corr}}_{\mathrm{complex}} - \sum_A \Delta G^{\mathrm{corr}}_{\mathrm{A}}$, where $\Delta G^{\mathrm{corr}}_{\mathrm{complex}}$ and $\Delta G^{\mathrm{corr}}_{A}$ denote thermal correction to free energy of complex and fragment A evaluated at their respective equilibrium geometries, respectively. It is noticed that Gaussian program employs rigid-rotor harmonic oscillator (RRHO) model to evaluate entropy, while we strongly suggest using Grimme's quasi-RRHO model implemented in our Shermo code[37] instead to calculate the free energy correction when low-frequencies occur, as in this case RRHO may severely overestimate their entropy contribution.[38]

If taking all terms mentioned above into account, binding free energy at solvent environment can be represented as $\Delta G_{\mathrm{bind}} = \Delta E_{\mathrm{int}} + \sum_A \Delta E^{\mathrm{dist}}_{A} + \Delta E^{\mathrm{solv}}_{\mathrm{int}} + \Delta\Delta G^{\mathrm{corr}}$, where ΔE_{int} collects all sobEDA terms described earlier evaluated under vacuum and at complex geometry.

3 sobEDAw method

In this section, we will describe the variant of sobEDA, namely sobEDAw, where "w" indicates it is designed specifically for studying intermolecular weak interactions.

Before introducing the definition of sobEDAw, one should first pay great attention to the concept of "dispersion interaction". Contribution of dispersion interaction to total intermolecular interaction can only be separated from other effects ambiguously in the case that there is no overlap between electron density of the interacting molecules,[39] and it is well-known that dispersion effect sources from long-range Coulomb correlation between the electrons in the respective molecules. However, chemists are mostly interested in molecular complexes in equilibrium configuration. In this case

intermolecular electron density overlap is never negligible, not only long-range but also middle-range Coulomb correlation exist, and hence there is no unique way to divide the dispersion interaction from the total interaction. The dispersion term defined by SAPT method fully represents intermolecular Coulomb correlation, the middle-range part is also included, so it may be more appropriate to be referred to as "dispersion-like" term.[40] Because DFT correlation functional can largely represent middle-range Coulomb correlation, therefore the accompanied dispersion correction energy mainly accounts for long-range Coulomb correlation. If an exchange-correction functional does not have any ability to represent the long-range Coulomb correlation, such as B3LYP,[41] then the dispersion correction contribution to intermolecular interaction energy could be regarded as "pure" dispersion interaction energy. Due to the difference in nature, the dispersion correction term E_{dc} in sobEDA is always notably or even significantly smaller than the SAPT dispersion term, and hence direct comparison between them is evidently unfair. A similar viewpoint and finding has been mentioned in Ref. [40]. It should be noted in passing that the dispersion correction term for DFT functionals that can partially represent long-range Coulomb correlation (such as M06-2X[42]) does not have any clear physical meaning. Therefore, these functionals should not be used in combination with sobEDA if dispersion effect will be analyzed separately.

Although the definition of dispersion term in SAPT somewhat exaggerates the magnitude of the dispersion effect, the SAPT and its variant DFT-SAPT[43,44] are quite popular and widely employed in literatures in exploring the nature of intermolecular interactions. The ratio between SAPT dispersion term and electrostatics term (d/e ratio) is even more useful in practice, which has been extensively employed to classify a wide variety of molecular complexes.[3,45] Therefore, we believe it is quite valuable to design a special form of sobEDA method to maximally reproduce the d/e ratio of SAPT so that the result can be compared with SAPT analysis, and this form may also be utilized as an inexpensive substitute for SAPT. To this aim, we defined sobEDAw, which decomposes intermolecular interaction energy as follows

$$\Delta E_{\mathrm{int}} = \Delta E_{\mathrm{els}} + \Delta E_{\mathrm{xrep}} + \Delta E_{\mathrm{orb}} + \Delta E_{\mathrm{disp}} \quad (2)$$

where ΔE_{els} and ΔE_{orb} have the same definition as sobEDA, while exchange-repulsion term and dispersion term are defined as follows

$$\Delta E_{\mathrm{xrep}} = \Delta E_{\mathrm{xrep}}^{\mathrm{sobEDA}} + (1-w)\Delta E_{\mathrm{DFTc}} \quad (3)$$

$$\Delta E_{\mathrm{disp}} = \Delta E_{\mathrm{dc}} + w\Delta E_{\mathrm{DFTc}} \quad (4)$$

in which $\Delta E_{\mathrm{xrep}}^{\mathrm{sobEDA}}$ specifically denotes the exchange-repulsion term obtained by sobEDA. It can be seen that the key difference between sobEDA and sobEDAw is that the latter divides the DFT correlation (ΔE_{DFTc}) by a factor *w*, some portion of ΔE_{DFTc} is combined with dispersion correction term to define the dispersion term that comparable with SAPT, while rest of ΔE_{DFTc} is simply incorporated into exchange-repulsion term.

The factor *w* should be carefully defined so that d/e ratio of sobEDAw will be close to that of SAPT as much as possible for wide variety of intermolecular interactions. The famous S66 is a well-balanced set of small molecule dimers containing C, H, N, O elements,[45] the 66 dimers in the set spans d/e ratio from very small to very large (0.29 to 5.42); in other words, this set involves interactions from highly electrostatics-dominated to highly dispersion-dominated. The d/e ratios in S66 paper were calculated using DFT-SAPT method at PBE0AC/aug-cc-pVDZ level with empirical enhancement for dispersion component, they provide a very useful reference of determining the way of evaluating *w* parameter. We performed sobEDA for the S66 set at B3LYP-D3(BJ)/aug-cc-pVTZ level, then we plotted scatter map between d/e ratios of DFT-SAPT and $\Delta E_{\mathrm{dc}}/\Delta E_{\mathrm{els}}$ of sobEDA, as well as between d/e ratios of DFT-SAPT and $(\Delta E_{\mathrm{dc}}+\Delta E_{\mathrm{DFTc}})/\Delta E_{\mathrm{els}}$, see Figs. 2(a) and 2(b), respectively. It can be seen that $\Delta E_{\mathrm{dc}}/\Delta E_{\mathrm{els}}$ is consistently too small compared to d/e ratios of DFT-SAPT, while $(\Delta E_{\mathrm{dc}}+\Delta E_{\mathrm{DFTc}})/\Delta E_{\mathrm{els}}$ is relatively too large in most range. However, as shown in Fig. 2(c), $(\Delta E_{\mathrm{dc}}+\Delta E_{\mathrm{DFTc}})/\Delta E_{\mathrm{els}}$ is fairly close to d/e ratios of DFT-SAPT in the range of small $\Delta E_{\mathrm{dc}}/\Delta E_{\mathrm{els}}$. This observation clearly indicates that *w* should be designed as a function that dependent of $\Delta E_{\mathrm{dc}}/\Delta E_{\mathrm{els}}$ and thereby allowing larger portion of ΔE_{DFTc} be incorporated into ΔE_{disp} in the situation of small $\Delta E_{\mathrm{dc}}/\Delta E_{\mathrm{els}}$. After many attempts, we

found an ideal form of w function for realizing this purpose:

$$w = \exp\left[-r\left(\frac{\Delta E_{\mathrm{dc}}}{\Delta E_{\mathrm{els}}} - a\right)\right](1-c) + c \tag{5}$$

where a, r, c are fitting parameters for specific DFT functional and basis set, c corresponds to limiting value of w at large $\Delta E_{\mathrm{dc}}/\Delta E_{\mathrm{els}}$ ratio. In order to ensure physical meaning, w is set to 1 when it is greater than 1 in very rare cases. At B3LYP-D3(BJ)/aug-cc-pVTZ level, the fitted values are c = 0.472, a = 0.024 and r = 2.478, the corresponding map showing how w function varies with $\Delta E_{\mathrm{dc}}/\Delta E_{\mathrm{els}}$ is shown in Fig. S1. It can be seen that when ΔE_{dc} is obviously lower than ΔE_{els}, w is large and thus most part of DFT correlation term will be mixed with ΔE_{dc} to yield dispersion term of sobEDAw. By employing the w defined in this way, we plotted $\Delta E_{\mathrm{disp}}/\Delta E_{\mathrm{els}}$ of sobEDAw and that of DFT-SAPT for the S66 set as Fig. 2(d). It can be seen that the former nicely match the latter for almost all complexes, and there is no systematical deviation in any range. One may find that in the region of very large $\Delta E_{\mathrm{disp}}/\Delta E_{\mathrm{els}}$, the inconsistency between sobEDAw and DFT-SAPT is seemingly more evident, actually this is not an issue since the relative deviation is still small, and more importantly, the d/e ratio derived by different SAPT implementations are not very consistent with each other in this case. As an instance, for the dispersion-dominated benzene···ethene complex, DFT-SAPT gives a d/e ratio of 4.57 in the original paper of S66 set, while its value given by SAPT2+/CBS calculation is merely 3.27.[39]

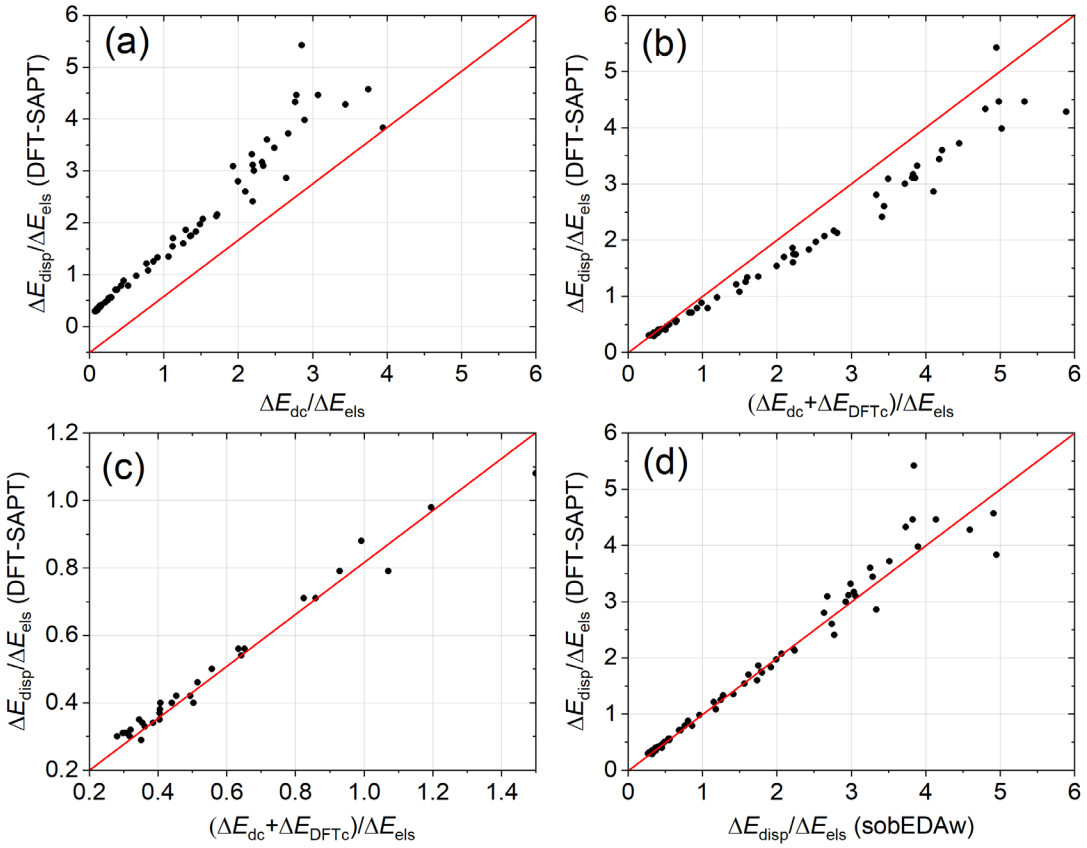

Fig. 2 Correlation between terms defined in sobEDAw and DFT-SAPT for S66 set. The sobEDAw data were obtained at B3LYP-D3(BJ)/aug-cc-pVTZ level, ΔE_{disp} was evaluated using Eqs. 4 and 5 with fitted parameters c = 0.472, a = 0.024 and r = 2.478. DFT-SAPT data were taken from S66 original paper.[45] The red lines highlight the diagonals in the figures for ease of comparison.

It is noteworthy that the electrostatics terms derived by sobEDA at B3LYP/aug-cc-pVTZ level for the S66 test set is in very good agreement with those derived by SAPT2+/CBS given in Ref. [39], see Fig. S2 for comparison. This finding demonstrates the reasonableness of sobEDAw, and meantime implies that the absolute value of ΔE_{disp} of sobEDAw should be very close to that of DFT-SAPT.

Not all exchange-correlation functionals are compatible with sobEDAw, the prerequisite of compatibility is that the functional should fully or basically lack ability to represent dispersion effect. To better illustrate this point, Fig. 3 plotted interaction energy curves for Ar···Ar calculated by several popular functionals as well as HF. It is known that only dispersion effect is the driving force of the binding between Ar atoms. From the figure it can be seen that there is no well depth in the case of HF, BLYP,[46,47] B3LYP[41] and BHandHLYP,[48] therefore they are well-suited to be used in conjunction

with sobEDAw. TPSS[49] and TPSSh[50] show very shallow well, they may be still usable with sobEDAw. In contrast, the representation of dispersion attraction by PBE,[51] PBE0,[52] ωB97X[53] and especially M06-2X[42] is evident, therefore they are in principle incompatible with the idea of sobEDAw.

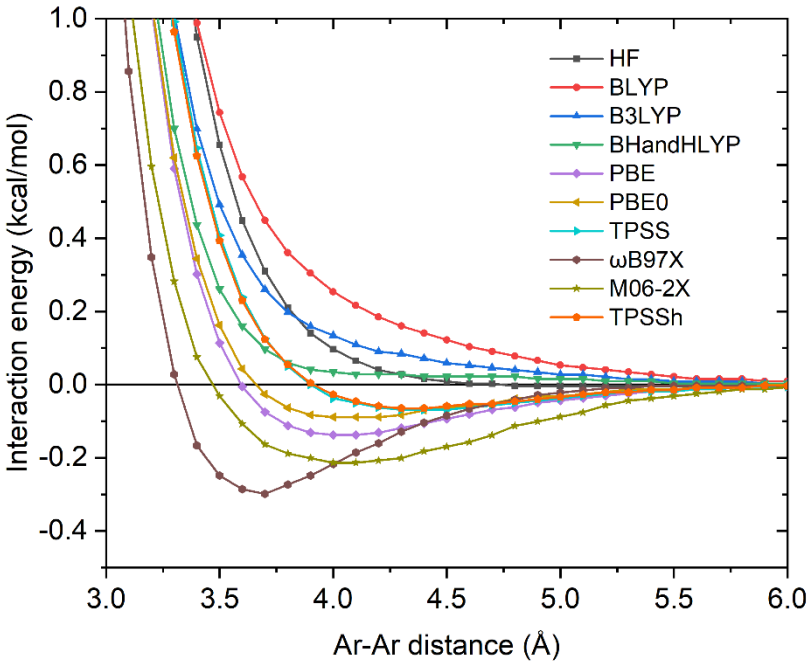

Fig. 3 Interaction energy calculated by various methods in combination with aug-cc-pVTZ basis set for Ar···Ar dimer at different interatomic distances.

An ideal basis set for sobEDAw analysis should be able to give accurate total interaction energy, a reasonable dispersion/electrostatics ratio, and the cost should be as low as possible. In order to find such basis sets, we combined different basis sets with the very popular B3LYP-D3(BJ) functional to calculate the S66 set. The error of the total interaction energy relative to the high-precision data of CCSD(T)/CBS level given in Ref. [39], the percentage error of dispersion/electrostatics ratio relative to the DFT-SAPT data given in original paper of S66 set, and the fitted parameters used for evaluating *w*, are all listed in Table 1. The basis sets we considered comprehensively cover from the small 2-zeta basis set 6-31G*[54] to the large 4-zeta basis set def2-QZVP.[55] The cases with and without diffusion functions, and the cases of using own and complex basis functions in monomer calculations, are all taken into consideration. In addition, we also counted the wall times consumed in sobEDAw analysis for a small

complex (adenine···thymine base pair) and a fairly large host-guest complex (C_{18}@OPP[56]) under different basis sets. From Table 1 one can find that both employing diffuse functions for non-hydrogen atoms and using complex basis functions are crucial for obtaining accurate interaction energies, also they are helpful in yielding $\Delta E_{disp}/\Delta E_{els}$ ratios closer to the DFT-SAPT reference. Among the tested basis sets, we found 6-31+G** with complex basis functions (6-31+G** w.CB) is a very ideal choice, not only the result is good enough (even slightly better than the largest def2-QZVP due to error cancellation with theoretical method), but also the cost is low. Even for the C_{18}@OPP complex containing as many as 244 atoms, the sobEDAw analysis with 6-31+G** w.CB only takes about 1 hour on a 96-cores server. 6-311+G(2d,p) w.CB is a even better choice in accuracy, but much more expensive. Another basis set that worth to mention is def2-TZVP w.CB, although it does not show a superior performance than 6-311+G(2d,p) w.CB and meantime more costly, its unique advantage is that def2 series of basis set is available for almost entire periodic table.[55] It is fully acceptable to remove *f* polarization functions of main group elements from def2-TZVP to reduce cost, this does not detectably sacrifice accuracy, see def2-TZVP(-f) w.CB data in Table 1. It is also a good idea to only employs def2-TZVP to represent the very heavy elements that not defined by the inexpensive 6-31+G** and 6-311+G(2d,p).

Table 1 Error in interaction energies and $\Delta E_{disp}/\Delta E_{els}$ ratios, fitted parameters for evaluating w, and time cost of sobEDAw for B3LYP-D3(BJ) with different basis sets.

Basis set [a]	Interaction energy [b]			dispersion/electrostatics ratio [c]					cost (sec.) [d]	cost (min.) [d]
	ME	MAE	maxE	*c*	*a*	*r*	MAPE	maxAPE	14+16 atoms	18+224 atoms
def2-QZVP	-0.30	0.31	-1.20	0.507	0.041	3.032	6.5	13.5	703	
aug-cc-pVTZ	-0.32	0.33	-1.23	0.472	0.024	2.478	7.6	16.6	452	
may-cc-pVTZ	-0.31	0.31	-1.08	0.554	0.056	2.887	6.2	12.1	143	
aug-cc-pVDZ	-0.86	0.86	-1.82	0.469	0.004	2.396	8.2	20.9	66	
def2-TZVPD	-0.28	0.29	-0.91	0.439	0.029	2.351	7.8	20.2	166	
ma-def2-TZVP	-0.41	0.41	-0.97	0.611	0.097	3.450	5.0	10.4	127	
ma-def2-TZVP(-f) w.CB	-0.19	0.20	-0.74	0.538	0.050	2.690	6.3	13.7	161	
ma-def2-TZVP(-f)	-0.44	0.45	-1.04	0.623	0.097	3.506	4.8	10.4	80	
def2-TZVP w.CB	-0.19	0.20	-0.85	0.538	0.066	2.724	5.8	13.3	131	
def2-TZVP(-f) w.CB	-0.22	0.23	-0.88	0.546	0.064	2.594	5.6	15.6	75	
def2-TZVP	-0.61	0.61	-1.51	0.422	0.065	2.260	4.9	17.3	74	
6-311+G(2d,p) w.CB	-0.16	0.18	-0.58	0.550	0.037	2.750	6.3	13.5	79	171
6-311+G(2d,p)	-0.52	0.52	-1.35	0.550	0.088	2.531	5.2	16.3	43	104
6-311+G(d,p) w.CB	-0.22	0.24	-0.90	0.553	0.058	2.633	6.2	12.5	54	
6-311+G(d,p)	-0.68	0.68	-1.54	0.507	0.116	2.198	5.2	20.4	32	
6-31+G** w.CB	-0.23	0.26	-1.13	0.575	0.071	2.571	6.0	14.0	39	71
6-31++G(d,p)	-0.76	0.76	-1.94	0.643	0.124	3.581	5.6	22.5	26	
6-31+G(d,p)	-0.72	0.72	-1.91	0.638	0.124	3.523	5.7	22.7	25	47
6-31G(d,p) w.CB	-0.28	0.41	-1.36	0.424	0.152	0.829	6.4	25.0	23	
6-31G(d,p)	-2.13	2.13	-5.39	-1.617	-0.327	0.223	12.1	50.2	17	

[a] w.CB means with complex basis functions in monomer calculations. (-f) means polarization functions of *f* angular moment of main group elements are removed for reducing cost. ma- prefix denotes the minimally augmented strategy of adding diffuse functions.[57]

[b] Mean error (ME), mean absolute error (MAE) and maximum error (maxE) of interaction energies calculated for S66 set with respect to the reference data at CCSD(T)/CBS level.[58] Data are in kcal/mol.

[c] Mean absolute percentage error (MAPE, in %) and maximum absolute percentage error (maxAPE, in %) of $\Delta E_{disp}/\Delta E_{els}$ derived by sobEDAw for S66 set with respect to the reference data derived by DFT-SAPT in S66 original paper. The fitted *c*, *a*, *r* parameters used in evaluating ΔE_{disp} of sobEDAw are given together.

[d] Wall time consumed for performing sobEDAw analysis for adenine···thymine base pair (14+16 atoms) and for C_{18}@OPP complex (18+224 atoms) via *sobEDA.sh* script on dual AMD EPYC 7R32 server (96 cores).

The widely used B3LYP-D3(BJ) is not a well-suited choice for all situations. For example, we have shown that B3LYP is unable to qualitatively correctly represent geometry of cyclo[18]carbon, and thus B3LYP-D3(BJ) cannot be used with sobEDAw for studying interactions containing this species;[59,60] in contrast, functionals with high HF composition like BHandHLYP and M06-2X works well for cyclo[18]carbon. Moreover, B3LYP is also known to poorly represent pure metals.[61] In addition, many quantum chemistry codes such as Gaussian and ORCA[62] are able to utilize Coulomb fitting auxiliary bases[63] to significantly accelerate calculation of pure functionals with

large basis sets, the cost in this case is often much lower than using B3LYP. Due to these considerations, we also provide fitted parameters and error statistics for combinations between four popular DFT functionals and three aforementioned basis sets, see Table 2. From the data one can find that all of the combinations work reasonably, and the foregoing conclusions about the behavior of basis sets still hold true. The reason of considering BHandHLYP-D3(BJ) here is that it shows insignificant self-interaction error (SIE) problem due to large (50%) HF exact exchange composition and thus useful in sobEDAw analysis to deal with the systems sensitive to SIE. The TPSSh-D3(BJ) with 10 % HF composition is a good choice for transition metal systems.[64,65] The GGA functional BLYP-D3(BJ) and meta-GGA functional TPSS-D3(BJ) may be used for sobEDAw when pure functional is favorable due to computational efficiency consideration. We do no take additional popular functionals such as PBE, PBE0, and M06-2X into account in this work. On the one hand, the functionals examined above are already sufficient to cover almost all research scenarios about weak interactions. On the other hand, as mentioned earlier, these functionals have obvious ability to describe dispersion effect, and thus are not suitable for combining with sobEDAw.

Table 2 Error in interaction energies and $\Delta E_{disp}/\Delta E_{els}$ ratios, and fitted parameters of evaluating w, for different DFT functionals with different basis sets. See Table 1 for meaning of the data.

Functional	Basis set	Interaction energy			dispersion/electrostatics ratio				
		ME	MAE	maxE	*c*	*a*	*r*	MAPE	maxAPE
BHandHLYP-D3(BJ)	def2-TZVP w.CB	-0.21	0.33	-1.69	0.680	-0.050	1.216	5.6	10.7
	6-311+G(2d,p) w.CB	-0.20	0.31	-1.40	0.595	-0.284	0.582	6.4	11.3
	6-31+G** w.CB	-0.30	0.44	-1.43	0.634	-0.065	0.702	5.8	13.7
TPSSh-D3(BJ)	def2-TZVP w.CB	0.09	0.25	0.75	0.583	0.629	0.934	4.5	21.3
	6-311+G(2d,p) w.CB	0.05	0.24	0.68	0.531	0.567	0.806	3.6	20.6
	6-31+G** w.CB	-0.06	0.26	-1.13	0.519	0.693	0.730	3.8	19.6
BLYP-D3(BJ)	def2-TZVP w.CB	0.07	0.19	0.56	0.304	-0.026	2.800	6.7	16.3
	6-311+G(2d,p) w.CB	0.11	0.22	0.61	0.310	-0.047	2.860	7.3	15.6
	6-31+G** w.CB	0.06	0.24	-0.87	0.338	-0.033	2.783	7.1	14.9
TPSS-D3(BJ)	def2-TZVP w.CB	0.17	0.29	1.03	0.698	0.488	1.429	4.5	20.4
	6-311+G(2d,p) w.CB	0.13	0.27	0.87	0.677	0.422	1.345	3.6	20.1
	6-31+G** w.CB	0.03	0.29	-1.18	0.675	0.556	1.125	3.8	19.4

At the end of this section, we would like to emphasize that it is not suggested to use sobEDAw in studying interactions containing great covalent character, because the dispersion term in sobEDAw may be severely overestimated due to excessive mixing

of DFT correlation term. Let us consider an extreme case, H_2 molecule at equilibrium distance. Because its DFT-D3(BJ) dispersion correction term is nearly zero and thus $\Delta E_{dc}/\Delta E_{els}$ ratio is also very close to zero, which in turn makes w close to 1, the dispersion contribution to interaction energy is almost equivalent to the contribution from DFT correlation according to the sobEDAw definition. This is obviously contrary to the general understanding of the dispersion effect. We found ΔE_{disp} of sobEDAw at B3LYP-D3(BJ)/6-311+G(2d,p) w.CB in this case is as large as -16.9 kcal/mol, which strikingly exaggerates the role of dispersion effect in stabilizing the H_2.

4 Script of realizing sobEDA and sobEDAw

To make the sobEDA and sobEDAw proposed in this work easily accessible by quantum chemistry researchers, we developed a Bash shell script *sobEDA.sh*. The script file, a detailed tutorial and numerous example files can be freely downloaded at http://sobereva.com/soft/sobEDA_tutorial.zip.

sobEDA.sh is able to produce various terms defined in sobEDA and sobEDAw with minimal preparation works for a given system. This script can also produce all terms involved in Fig. 1 to understand all details of energy changes. *sobEDA.sh* invokes Gaussian program to conduct all quantum chemistry calculations needed in the analyses. Multiwfn is also needed by the script to generate promolecule wavefunction based on the wavefunctions of isolated fragments. *sobEDA.sh* was written in a clear way with sufficient comments, hence users can easily modify the script to fine tune the calculation details to realize special purposes.

Note that fragment distortion, solvation and thermodynamic effects mentioned earlier are not considered in *sobEDA.sh*. They should be manually calculated by users themselves.

5. Application examples

In this section we will present several application examples of sobEDA and sobEDAw to illustrate their reliability and usefulness. Gaussian 16 A.03[28] and Multiwfn 3.8(dev)[29] updated on 2023-Jun-24 were used for sobEDA and sobEDAw analyses, VMD 1.9.3[66] and ChimeraX 1.6.1[67] were employed for visualization.

5.1 Adenine···Thymine complex

Adenine and thymine molecules can form two representative dimer configurations, which are respectively dominated by H-bond and π-π stacking effects. Fig. 4 shows their geometries, which were optimized at MP2/cc-pVTZ level in S22 set.[68] Based on these structures, we performed sobEDAw analysis at B3LYP-D3(BJ)/6-311+G(2d,p) w.CB level, the results are given in Table 3, the corresponding SAPT2+3(CCD)/aug-cc-pVTZ data provided in BioFragment Database (BFDb)[69] are also listed for comparison purpose. The reference interaction energies for H-bond and stacked configurations evaluated using CCSD(T)/CBS are -16.37 and -12.23 kcal/mol, respectively.[68] From the table it can be seen that the total interaction energies of sobEDAw are quite close to the reference values, and $\Delta E_{disp}/\Delta E_{els}$ ratios given by sobEDAw are in excellent agreement with those by the robust while fairly expensive SAPT2+3(CCD) calculations. Furthermore, we find that the values of the four physical components ΔE_{els}, ΔE_{xrep}, ΔE_{orb} and ΔE_{disp} of sobEDAw well correspond to the ΔE_{els}, ΔE_{exch} (exchange), ΔE_{ind} (induction) and ΔE_{disp} of the SAPT2+3(CCD), respectively. This not only further demonstrates the rationality of sobEDAw, but also highlights that the inexpensive sobEDAw is a very good alternative to the high-order SAPT method for weak interaction analyses.

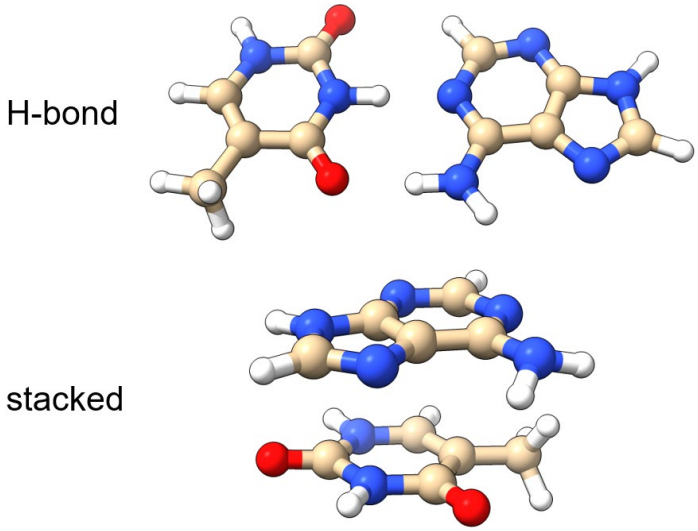

Fig. 4 Geometry of H-bond and stacked configurations of adenine and thymine dimer. White, blue, red, and yellow spheres correspond to hydrogen, nitrogen, oxygen and carbon atoms, respectively, similarly hereinafter.

Table 3 sobEDAw and SAPT result for H-bond and stacked configurations of adenine and thymine dimer

configuration	sobEDAw B3LYP-D3(BJ)/6-311+G(2d,p) w.CB					
	ΔE_{int}	ΔE_{els}	ΔE_{xrep}	ΔE_{orb}	ΔE_{disp}	$\Delta E_{disp}/\Delta E_{els}$
H-bond	-16.9	-26.9	34.4	-14.7	-9.8	0.36
stacked	-11.6	-10.9	20.5	-3.4	-17.9	1.64
configuration	SAPT2+3(CCD)/aug-cc-pVTZ					
	ΔE_{int}	ΔE_{els}	ΔE_{exch}	ΔE_{ind}	ΔE_{disp}	$\Delta E_{disp}/\Delta E_{els}$
H-bond	-17.5	-26.6	31.8	-11.9	-10.8	0.41
stacked	-12.5	-10.7	18.3	-2.5	-17.6	1.66

5.2 $O_2\cdots C_{18}$ triplet complex

The cyclo[18]carbon was very recently determined by experimental observation in condensed phase,[70] it shows quite unusual geometry and electronic characteristics.[71-73] We have extensively studied its molecular complexes with a wide variety of small molecules.[2] Here we employ sobEDAw to investigate the intermolecular interaction for the open-shell triplet complex $O_2\cdots C_{18}$, the structure optimized at ωB97XD/def2-TZVP level is shown in Fig. 5. Because B3LYP-D3(BJ) is found to be completely failed to represent the basic characteristics of cyclo[18]carbon,[59] BHandHLYP is used instead in this instance, we have shown it is

adequate to correctly capture the geometry character of this unique system.[59] At BHandHLYP/6-311+G(2d,p) w.CB level, sobEDAw gives ΔE_{els} = -0.67 kcal/mol, ΔE_{xrep} = 2.97 kcal/mol, ΔE_{orb} = -0.13 kcal/mol and ΔE_{disp} = -5.37 kcal/mol, suggesting that only dispersion effect dominates the adsorption of O_2 into the ring center of the cyclo[18]carbon. This conclusion is also fully in line with chemical intuition as both O_2 and cyclo[18]carbon show very low polarity and there does not form a chemical bond. In addition, the total interaction energy at this level (-3.21 kcal/mol) is in good agreement with that estimated at a much more expensive level (-3.6 kcal/mol at ωB97X-V/def2-QZVPP).[2] In previous study we have also applied SAPT0/aug-cc-pVTZ for this system, which shows ΔE_{els} = -0.68 kcal/mol, ΔE_{exch} = 2.09 kcal/mol, ΔE_{ind} = -0.08 kcal/mol and ΔE_{disp} = -5.83 kcal/mol.[2] From the comparison we can again find that there is a very strong one-to-one correspondence between the terms of sobEDAw and SAPT.

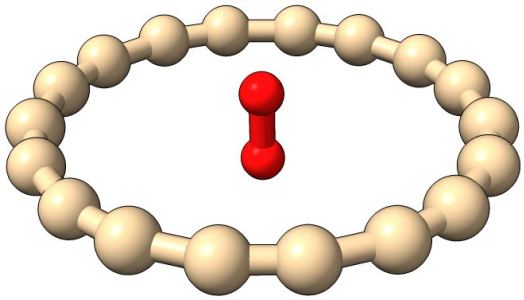

Fig. 5 Geometry of optimized triplet complex $O_2\cdots C_{18}$ at ωB97XD/def2-TZVP level.

5.3 Vinylene carbonate trimer

This example illustrates the ability of sobEDAw in studying molecular multimer. The structure of vinylene carbonate trimer given in 3B-69 trimer set[74] is shown in Fig. 6, which was extracted from optimized molecular crystal. Independent gradient model based on Hirshfeld partition (IGMH) is a popular method of visually exhibiting interfragment interactions,[75,76] IGMH isosurfaces are also appended to the trimer structure map to reveal all notable interactions between the three molecules. From the

IGMH map one can find that the isosurfaces between the molecules are basically in green, indicating that the interactions can hardly be regarded as typical H-bonds. Indeed, C-H···O type of interaction is known to be much weaker compared to common H-bonds.[3] It is difficult to judge whether the interaction between the three molecules is electrostatics or dispersion dominated based on common knowledge. In order to find a definitive answer, we analyzed this system with sobEDAw, the result is ΔE_{els} = -14.53 kcal/mol, ΔE_{xrep} = 15.14 kcal/mol, ΔE_{orb} = -4.30 kcal/mol and ΔE_{disp} = -9.77 kcal/mol, indicating that electrostatic effect plays a more important role than dispersion effect for the trimerization, while dispersion also contributes greatly. Furthermore, although the orbital interaction is relatively insignificant, it is not so small that can be ignored. All terms derived by sobEDAw for this system are in very reasonable ranges, which shows that this method is fully applicable to the interactions between more than two molecules.

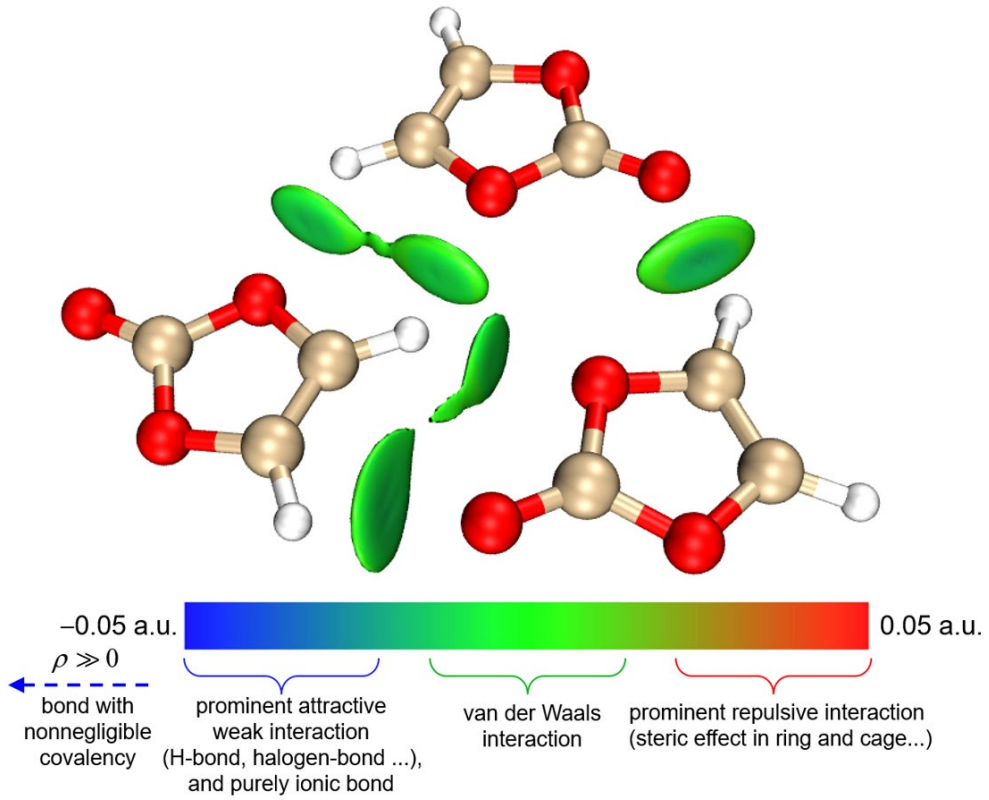

Fig. 6 Structure of vinylene carbonate trimer with IGMH isosurfaces to visually exhibit interaction characteristics. Isovalue of δg^{inter} function of IGMH is set to 0.005 a.u. Standard assignment of various colors in IGMH method is also shown.

5.4 Coordination bond of $(CO)_5Cr=CH_2$

In this and the following examples, we will employ sobEDA to study chemical bonds. To illustrate the relevant capacity of sobEDA, we performed the analysis for the coordination bond between CH_2 and $(CO)_5Cr$ fragments in transition metal complex $(CO)_5CrCH_2$. The entire system and the fragments are set to be in neutral singlet state during the calculations. It should be noted that the actual ground state of CH_2 carbene is triplet, if one hopes to take triplet as reference state of the fragment, one may manually calculate its single-triplet energy gap as preparation energy of the electronic structure, which could then be treated as an individual term of the energy decomposition. The sobEDA calculation for this system was conducted using TPSSh-D3(BJ) with 6-311G* basis set[77] for ligands and Stuttgart pseudopotential basis set[78] for Cr atom, the result is ΔE_{els} = -113.84 kcal/mol, ΔE_{xrep} = 162.45 kcal/mol, ΔE_{orb} = -111.49 kcal/mol, ΔE_c = -25.46 kcal/mol (in which ΔE_{dc} contributes -3.87 kcal/mol) and total interaction energy of -88.34 kcal/mol. The data indicates that electrostatic and orbital interactions play nearly equal roles in stabilizing the coordinate bond, while exchange-repulsion effect cancels most of the attractive effects. Besides, the Coulomb correlation makes a small contribution to the formation of the bond.

5.5 Diamantane-diamantane

It was reported that there is an highly elongated and labile C-C bond in the diamantane-diamantane system,[79] as shown in Fig. 7. The uncommon long length of the bond is believed to be caused by strong steric effect between the two diamantane moieties, and it is also argued that the dispersion effect between the caged alkanes significantly enhances the bonding strength, thus making the system less prone to spontaneous dissociation. Systems containing unusual interactions like this are well suited for study by energy decomposition analysis. Based on the structure of diamantane-diamantane optimized at M06-2X/6-31G(d,p) level,[79] we carried out sobEDA analysis based on B3LYP-D3(BJ)/6-311+G(2d,p) level, each diamantane is treated as a doublet radical, and spin of electrons is flipped for one of the two

diamantanes when constructing promolecule wavefunction. The result is ΔE_{els} = -143.96 kcal/mol, ΔE_{xrep} = 258.62 kcal/mol, ΔE_{orb} = -156.07 kcal/mol, ΔE_{DFTc} = -30.42 kcal/mol and ΔE_{dc} = -17.83 kcal/mol. Compared to the last example $(CO)_5Cr{=}CH_2$, although the magnitude of ΔE_{els} and ΔE_{orb} in this system is evidently increased, the increase in ΔE_{xrep} is significantly higher, which fully reflects that the steric hindrance effect in the diamantane-diamantane system is particularly significant. Another important point is that the ΔE_{dc} is quite notable in this system. Recall that this dispersion correction term for B3LYP can be simply viewed as representing pure or long-range dispersion interaction, one can conclude that the dispersion effect between the two diamantanes indeed greatly stabilizes the elongated C-C bond. By the way, it only took 6 minutes to finish the present sobEDA analysis on a 96-core server, which demonstrates the high efficiency of the implementation of the sobEDA approach.

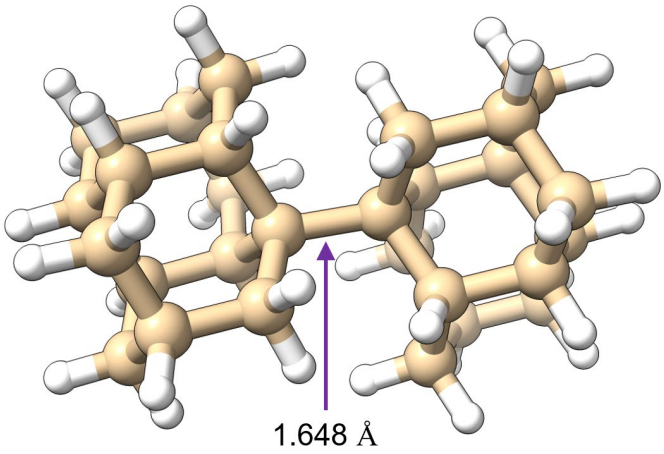

Fig. 7 Structure of diamantane-diamantane optimized at M06-2X/6-31G(d,p) level.

5.6 Cation-π and anion-π complexes: $C_6H_6\cdots Na^+$ and $C_6F_6\cdots Cl^-$

This example illustrates using sobEDA to study charged interactions. It is easy to understand that the benzene containing π electrons can form a stable complex with Na^+ cation, because the Lewis basic π electrons can have significant electrostatic attraction with the positively charged Na^+. The interaction of anions with π systems is relatively less known. It was found that when the six hydrogens of benzene are replaced by

fluorine with strong electron-withdrawing ability, the region above the center of the ring will show Lewis acidity, which allows the C_6F_6 to combine with anions.[80] In this example, we use sobEDA to investigate the difference in the interaction strength and nature of $C_6H_6\cdots Na^+$ and $C_6F_6\cdots Cl^-$. Their structures optimized at BHandHLYP/6-311+G(2d,p) level are shown in Fig. 8, and the sobEDA result is given in Table 4. From the data it can be seen that the overall interaction of $C_6H_6\cdots Na^+$ is remarkably stronger than that of $C_6F_6\cdots Cl^-$. The secondary reason is that the electrostatic interaction of $C_6H_6\cdots Na^+$ is modestly larger, while the main reason is that the orbital interaction between C_6H_6 and Na^+ is conspicuously more intense, which reflects that the binding between Na^+ and benzene can cause much more evident charge transfer and electronic polarization phenomena. As for exchange-repulsion and Coulomb correlation, they show about the same magnitude for the interactions in the two systems.

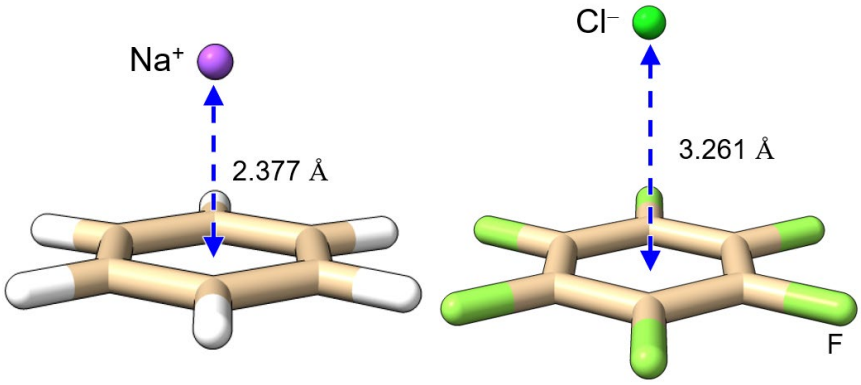

Fig. 8 Structure of $C_6H_6\cdots Na^+$ and $C_6F_6\cdots Cl^-$ optimized at BHandHLYP/6-311+G(2d,p) level.

Table 4 Various sobEDA terms for $C_6H_6\cdots Na^+$ and $C_6F_6\cdots Cl^-$ interactions in kcal/mol calculated at BHandHLYP/6-311+G(2d,p) level.

	ΔE_{int}	ΔE_{els}	ΔE_{xrep}	ΔE_{orb}	ΔE_c
$C_6H_6\cdots Na^+$	-28.14	-17.54	11.20	-16.42	-5.37
$C_6F_6\cdots Cl^-$	-14.25	-14.09	12.15	-6.03	-6.29

5.7 Single, double, and triple bonds in ethane, ethene and acetylene

This example uses sobEDA to study the C-C bond in ethane, ethylene, and acetylene, so as to test the capability of sobEDA to explore the multiple-bond

interaction between fragments. Table 5 lists the interaction energies and their physical components among the fragments in these systems, geometry optimizations and sobEDA analyses were performed at B3LYP-D3(BJ)/def2-TZVP level. Notice that the ground state of the CH fragment that constitutes acetylene is doublet state, but its quartet excited state with three unpaired electrons is closer to the actual status of CH in acetylene, so we respectively considered both spin states in the sobEDA analyses. If quartet state of CH is employed, then from Table 5 one can clearly find orbital interaction strength sharply increases in the order of ethane, ethene and acetylene, which is fully expected according to the bond multiplicities. In addition, it can be found that change in ΔE_{orb} has the closest correlation with variation of ΔE_{int} compared to other terms, indicating that the relative strength of C-C bond in these systems is mainly determined by orbital interactions rather than other reasons.

When doublet CH is used instead of quartet CH in the sobEDA analysis for acetylene, from the corresponding data in Table 5 one can see that there is a reduction of 40 kcal/mol in the interfragment interaction energy, which is due to the fact that the energy of the doublet CH is 20 kcal/mol lower than that of the quartet state. In addition, the exchange-repulsion term (1299.5 kcal/mol) calculated with doublet CH is significantly larger than that with quartet CH (249.0 kcal/mol), this is because there is a significant overlap between the lone pair electrons of the two doublet CH fragments in their bonding region, thus bringing a huge Pauli repulsion effect. At the same time, the ΔE_{orb} term -1065.9 kcal/mol obtained with doublet CH is also significantly more negative than the counterpart (-336.8 kcal/mol) with quartet CH, clearly the significantly stronger orbital relaxation effect greatly cancels out the oversized exchange-repulsion effect. The analysis conclusion of sobEDA in this example is fully consistent with common chemical knowledge.

Table 5 sobEDA analysis result for ethane, ethene and acetylene at B3LYP-D3(BJ)/def2-TZVP. The spin multiplicities used in the fragment calculations are marked in the upper left corner of the fragments

system		ΔE_{int}	ΔE_{els}	ΔE_{xrep}	ΔE_{orb}	ΔE_c
ethane	$^2(H_3C) - ^2(CH_3)$	-112.1	-135.4	212.7	-167.7	-21.7
ethene	$^3(H_2C) - ^3(CH_2)$	-186.7	-182.2	283.3	-255.9	-32.0
acetylene	$^4(HC) - ^4(CH)$	-275.9	-143.3	249.0	-336.8	-44.8
	$^2(HC) - ^2(CH)$	-235.9	-448.9	1299.5	-1065.9	-20.7

5.8 Variation of energy components of CH_3-CN with respect to interfragment distance

In the final example, we study variation of various terms of sobEDA with respect to change in the spacing between CH_3 and CN radical fragments in CH_3CN. The data obtained at B3LYP-D3(BJ)/def2-TZVP level are collectively plotted in Fig. 9. It is noteworthy that starting from elongation of C-C bond of 1.2 Å, the converged wavefunction of final state of sobEDA analysis is in broken-symmetry state, which makes the dissociation curve have a correct asymptotic behavior. From Fig. 9 we can see magnitude of each energy component gradually decreases as the spacing between the two fragments increases. Orbital interaction plays a major role in the binding, followed by electrostatic interaction, and then exchange interaction. Among them, the orbital interaction decays relatively slowly. When the distance between the fragments is more than 0.5 Å relative to equilibrium position, the interfragment attraction is almost solely contributed by the orbital interaction. Regardless of the interfragment distance, the contribution of the DFT correlation energy to the binding is quite inapparent, and the dispersion correction term is fully negligible relative to the magnitude of the other terms, suggesting that dispersion effect little affects the formation of CH_3-CN. It can also be seen from the figure that the Pauli repulsion is the only source of mutual exclusion between CH_3 and CN. From this example, it is clear that sobEDA is completely suitable for studying the interaction at non-equilibrium structures and may also be used to reveal the driving forces and hindering factors in chemical or physical processes to provide an in-depth perspective.

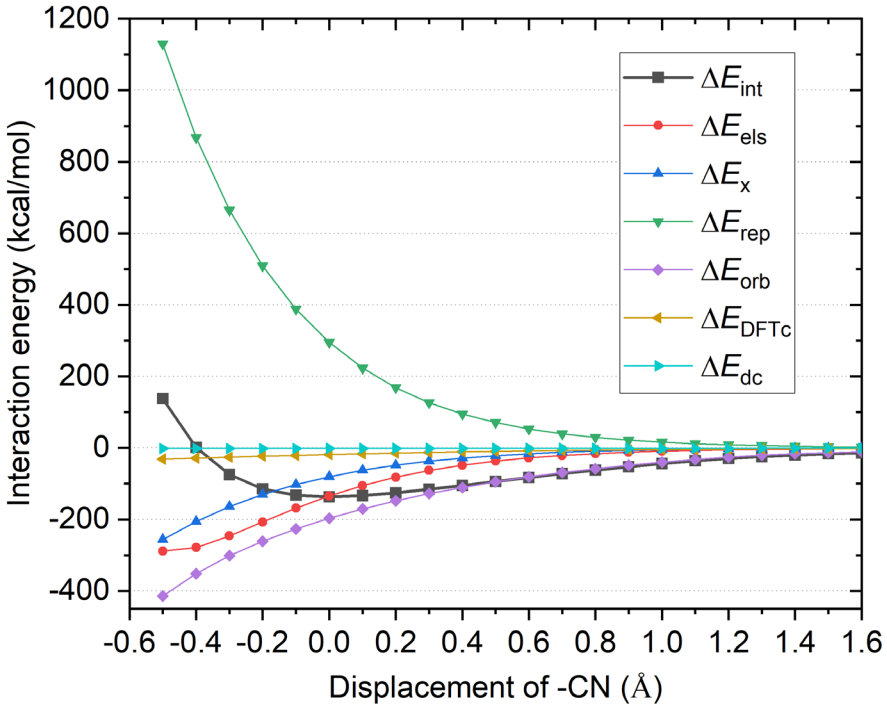

Fig. 9 Changes of various components of interaction energy between CH_3 and CN fragments in CH_3CN calculated by sobEDA as displacement of CN fragment with respect to equilibrium position. Data were calculated at B3LYP-D3(BJ)/def2-TZVP level. The structures of the two fragments are kept fixed during the scan.

6. Summary

In this work, sobEDA is proposed to realize energy decomposition analysis based on the very popular dispersion-corrected density functional theory. sobEDA is easy to implement, high-efficient, universally applicable, and has a clear physical meaning. In addition, a variant of sobEDA named sobEDAw is specifically designed to study intermolecular weak interactions. By means of a carefully defined expression of dispersion term, sobEDAw is able to provide a dispersion to electrostatics ratio that is fairly close to the much more sophisticated and evidently more expensive SAPT method. We developed a Bash shell script *sobEDA.sh* to make the new methods readily accessible to Gaussian and Multiwfn users, by which only very few steps are needed to carry out the analyses for a new system. A tutorial of using this script along with input files of rich application examples is available at

http://sobereva.com/soft/sobEDA_tutorial.zip. Although currently sobEDA is realized on the top of Gaussian program, we emphasize that in principle it can be very easily incorporated into any other DFT-based codes.

Currently sobEDA and sobEDAw are only examined for molecular systems, they are also very promising to be applied to periodic systems based on first-principle programs such as CP2K[81] to study nature of interactions involving solids, polymers, 2D layered materials, and so on. There is no evident difficulty in theory and technology to implement sobEDA and sobEDAw in these codes especially when only **k**=Γ point is considered.

Supplemental information

Variation of *w* with respect to ratio between dispersion correction and electrostatics terms; Electrostatics terms derived by SAPT2+/CBS and those by sobEDA at B3LYP/aug-cc-pVTZ level for S66 set.

Acknowledgments

This research was not funded.

CRediT authorship contribution statement

Tian Lu: Conceptualization, Methodology, Software, Investigation, Formal analysis, Writing - Original Draft

Qinxue Chen: Validation, Writing - Review & Editing, Investigation

Conflicts of interest

The authors declare no conflict of interest.

References

[1] T. Lu, F. Chen, Revealing the nature of intermolecular interaction and configurational preference of the nonpolar molecular dimers (H2)2, (N2)2, and (H2)(N2). *J. Mol. Model.* **2013**, *19*, 5387.

[2] Z. Liu, T. Lu, Q. Chen, Intermolecular interaction characteristics of the all-carboatomic ring, cyclo[18]carbon: Focusing on molecular adsorption and stacking. *Carbon* **2021**, *171*, 514.

[3] S. Emamian, T. Lu, H. Kruse, H. Emamian, Exploring Nature and Predicting Strength of Hydrogen Bonds: A Correlation Analysis Between Atoms-in-Molecules Descriptors, Binding Energies, and Energy Components of Symmetry-Adapted Perturbation Theory. *J. Comput. Chem.* **2019**, *40*, 2868.

[4] Y. Jiao, Y. Liu, W. Zhao, Z. Wang, et al., Theoretical study on the interactions of halogen-bonds and pnicogen-bonds in phosphine derivatives with Br2, BrCl, and BrF. *Int. J. Quantum Chem.* **2017**, *117*, e25443.

[5] P. Jerabek, G. Frenking, Comparative bonding analysis of N2 and P2 versus tetrahedral N4 and P4. *Theor. Chem. Acc.* **2014**, *133*, 1447.

[6] P. Jerabek, P. Schwerdtfeger, G. Frenking, Dative and electron-sharing bonding in transition metal compounds. *J. Comput. Chem.* **2019**, *40*, 247.

[7] M. J. S. Phipps, T. Fox, C. S. Tautermann, C.-K. Skylaris, Energy decomposition analysis approaches and their evaluation on prototypical protein–drug interaction patterns. *Chem. Soc. Rev.* **2015**, *44*, 3177.

[8] P. Su, Z. Tang, W. Wu, Generalized Kohn-Sham energy decomposition analysis and its applications. *WIREs Comput. Mol. Sci.* **2020**, *10*, e1460.

[9] S. Grimme, J. Antony, T. Schwabe, C. Mück-Lichtenfeld, Density functional theory with dispersion corrections for supramolecular structures, aggregates, and complexes of (bio)organic molecules. *Org. Biomol. Chem.* **2007**, *5*, 741.

[10] G. Bistoni, Finding chemical concepts in the Hilbert space: Coupled cluster analyses of noncovalent interactions. *WIREs Comput. Mol. Sci.* **2020**, *10*, e1442.

[11] M. v. Hopffgarten, G. Frenking, Energy decomposition analysis. *WIREs Comput. Mol. Sci.* **2012**, *2*, 43.

[12] L. Pecher, R. Tonner, Deriving bonding concepts for molecules, surfaces, and solids with energy decomposition analysis for extended systems. *WIREs Comput. Mol. Sci.* **2019**, *9*, e1401.

[13] K. Kitaura, K. Morokuma, A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation. *Int. J. Quantum Chem.* **1976**, *10*, 325.

[14] P. Su, H. Li, Energy decomposition analysis of covalent bonds and intermolecular interactions. *J. Chem. Phys.* **2009**, *131*, 014102.

[15] P. Su, Z. Jiang, Z. Chen, W. Wu, Energy Decomposition Scheme Based on the Generalized Kohn–Sham Scheme. *J. Phys. Chem. A* **2014**, *118*, 2531.

[16] E. D. Glendening, Natural Energy Decomposition Analysis: Explicit Evaluation of Electrostatic and Polarization Effects with Application to Aqueous Clusters of Alkali Metal Cations and Neutrals. *J. Am. Chem. Soc.* **1996**, *118*, 2473.

[17] P. R. Horn, Y. Mao, M. Head-Gordon, Probing non-covalent interactions with a second generation energy decomposition analysis using absolutely localized molecular orbitals. *Phys. Chem. Chem. Phys.* **2016**, *18*, 23067.

[18] R. Z. Khaliullin, E. A. Cobar, R. C. Lochan, A. T. Bell, et al., Unravelling the Origin of Intermolecular Interactions Using Absolutely Localized Molecular Orbitals. *J. Phys. Chem. A* **2007**, *111*, 8753.

[19] K. Szalewicz, Symmetry-adapted perturbation theory of intermolecular forces. *WIREs Comput. Mol. Sci.* **2012**, *2*, 254.

[20] E. G. Hohenstein, C. D. Sherrill, Wavefunction methods for noncovalent interactions. *WIREs*

Comput. Mol. Sci. **2012**, *2*, 304.
[21] I. Mayer, Improved chemical energy component analysis. *Phys. Chem. Chem. Phys.* **2012**, *14*, 337.
[22] T. Lu, Z. Liu, Q. Chen, Comment on "18 and 12 – Member carbon rings (cyclo[n]carbons) – A density functional study". *Mat. Sci. Eng. B* **2021**, *273*, 115425.
[23] M. P. Mitoraj, A. Michalak, T. Ziegler, A Combined Charge and Energy Decomposition Scheme for Bond Analysis. *J. Chem. Theory Comput.* **2009**, *5*, 962.
[24] X. Cao, S. Liu, C. Rong, T. Lu, et al., Is there a generalized anomeric effect? Analyses from energy components and information-theoretic quantities from density functional reactivity theory. *Chem. Phys. Lett.* **2017**, *687*, 131.
[25] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, et al., General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14*, 1347.
[26] Z. Tang, Y. Song, S. Zhang, W. Wang, et al., XEDA, a fast and multipurpose energy decomposition analysis program. *J. Comput. Chem.* **2021**, *42*, 2341.
[27] E. Epifanovsky, A. T. B. Gilbert, X. Feng, J. Lee, et al., Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *J. Chem. Phys.* **2021**, *155*.
[28] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, et al., Gaussian 16 A.03. Wallingford, CT, **2016**.
[29] T. Lu, F. Chen, Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012**, *33*, 580.
[30] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, 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.
[31] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, et al., A generally applicable atomic-charge dependent London dispersion correction. *J. Chem. Phys.* **2019**, *150*.
[32] W. Hujo, S. Grimme, 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.
[33] S. Grimme, S. Ehrlich, L. Goerigk, Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456.
[34] S. F. Boys, F. Bernardi, The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553.
[35] F. M. Bickelhaupt, K. N. Houk, Analyzing Reaction Rates with the Distortion/Interaction-Activation Strain Model. *Angew. Chem. Int. Ed.* **2017**, *56*, 10070.
[36] A. V. Marenich, C. J. Cramer, D. G. Truhlar, Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378.
[37] T. Lu, Q. Chen, Shermo: A general code for calculating molecular thermochemistry properties. *Comput. Theor. Chem.* **2021**, *1200*, 113249.
[38] S. Grimme, Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chem.-Eur. J.* **2012**, *18*, 9955.
[39] Q. Wang, J. A. Rackers, C. He, R. Qi, et al., General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 2609.
[40] O. A. Stasyuk, R. Sedlak, C. F. Guerra, P. Hobza, Comparison of the DFT-SAPT and Canonical EDA Schemes for the Energy Decomposition of Various Types of Noncovalent Interactions. *J. Chem. Theory Comput.* **2018**, *14*, 3440.

[41] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623.

[42] Y. Zhao, D. Truhlar, 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.

[43] A. Heßelmann, G. Jansen, Intermolecular dispersion energies from time-dependent density functional theory. *Chem. Phys. Lett.* **2003**, *367*, 778.

[44] A. Heßelmann, G. Jansen, The helium dimer potential from a combined density functional theory and symmetry-adapted perturbation theory approach using an exact exchange-correlation potential. *Phys. Chem. Chem. Phys.* **2003**, *5*, 5010.

[45] J. Řezáč, K. E. Riley, P. Hobza, S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427.

[46] A. D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098.

[47] C. Lee, W. Yang, R. G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785.

[48] A. D. Becke, A New Mixing of Hartree--Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372.

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

[50] V. N. Staroverov, G. E. Scuseria, J. Tao, J. P. Perdew, Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.* **2003**, *119*, 12129.

[51] J. P. Perdew, K. Burke, M. Ernzerhof, Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.

[52] C. Adamo, V. Barone, Toward Reliable Density Functional Methods Without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110*, 6158.

[53] J.-D. Chai, M. Head-Gordon, Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.

[54] P. C. Hariharan, J. A. Pople, The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chem. Acc.* **1973**, *28*, 213.

[55] F. Weigend, R. Ahlrichs, 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.

[56] Z. Liu, X. Wang, T. Lu, J. Wang, et al., Molecular assembly with a figure-of-eight nanohoop as a strategy for the collection and stabilization of cyclo[18]carbon. *Phys. Chem. Chem. Phys.* **2023**.

[57] J. Zheng, X. Xu, D. G. Truhlar, Minimally augmented Karlsruhe basis sets. *Theor. Chem. Acc.* **2011**, *128*, 295.

[58] J. Řezáč, K. E. Riley, P. Hobza, Extensions of the S66 Data Set: More Accurate Interaction Energies and Angular-Displaced Nonequilibrium Geometries. *J. Chem. Theory Comput.* **2011**, *7*, 3466.

[59] Z. Liu, T. Lu, Q. Chen, An sp-hybridized all-carboatomic ring, cyclo[18]carbon: Bonding character, electron delocalization, and aromaticity. *Carbon* **2020**, *165*, 468.

[60] Z. Liu, T. Lu, Q. Chen, Comment on "Theoretical investigation on bond and spectrum of cyclo[18]carbon (C18) with sp-hybridized". *J. Mol. Model.* **2021**, *27*, 42.
[61] J. Paier, M. Marsman, G. Kresse, Why does the B3LYP hybrid functional fail for metals? *J. Chem. Phys.* **2007**, *127*.
[62] F. Neese, F. Wennmohs, U. Becker, C. Riplinger, The ORCA quantum chemistry program package. *J. Chem. Phys.* **2020**, *152*, 224108.
[63] F. Weigend, A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285.
[64] M. Bühl, H. Kabrede, Geometries of Transition-Metal Complexes from Density-Functional Theory. *J. Chem. Theory Comput.* **2006**, *2*, 1282.
[65] S. Dohm, A. Hansen, M. Steinmetz, S. Grimme, et al., Comprehensive Thermochemical Benchmark Set of Realistic Closed-Shell Metal Organic Reactions. *J. Chem. Theory Comput.* **2018**, *14*, 2596.
[66] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33.
[67] E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, et al., UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **2021**, *30*, 70.
[68] P. Jurečka, J. Šponer, J. Černý, P. Hobza, Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.
[69] L. A. Burns, J. C. Faver, Z. Zheng, M. S. Marshall, et al., The BioFragment Database (BFDb): An open-data platform for computational chemistry analysis of noncovalent interactions. *J. Chem. Phys.* **2017**, *147*.
[70] K. Kaiser, L. M. Scriven, F. Schulz, P. Gawel, et al., An sp-hybridized molecular carbon allotrope, cyclo[18]carbon. *Science* **2019**, *365*, 1299.
[71] Z. Liu, T. Lu, Q. Chen, An sp-hybridized all-carboatomic ring, cyclo[18]carbon: Electronic structure, electronic spectrum, and optical nonlinearity. *Carbon* **2020**, *165*, 461.
[72] Z. Liu, T. Lu, Q. Chen, Vibrational spectra and molecular vibrational behaviors of all-carboatomic rings, cyclo[18]carbon and its analogues. *Chem. – Asian J.* **2021**, *16*, 56.
[73] T. Lu, Q. Chen, Ultrastrong Regulation Effect of the Electric Field on the All-Carboatomic Ring Cyclo[18]Carbon. *ChemPhysChem* **2021**, *22*, 386.
[74] J. Řezáč, Y. Huang, P. Hobza, G. J. O. Beran, Benchmark Calculations of Three-Body Intermolecular Interactions and the Performance of Low-Cost Electronic Structure Methods. *J. Chem. Theory Comput.* **2015**, *11*, 3065.
[75] T. Lu, Q. Chen, Visualization Analysis of Weak Interactions in Chemical Systems. In Reference Module in Chemistry, Molecular Sciences and Chemical Engineering, Elsevier, **2023**.
[76] T. Lu, Q. Chen, Independent gradient model based on Hirshfeld partition: A new method for visual study of interactions in chemical systems. *J. Comput. Chem.* **2022**, *43*, 539.
[77] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980**, *72*, 650.
[78] M. Dolg, U. Wedig, H. Stoll, H. Preuss, Energy-adjusted ab initio pseudopotentials for the first row transition elements. *J. Chem. Phys.* **1987**, *86*, 866.
[79] P. R. Schreiner, L. V. Chernish, P. A. Gunchenko, E. Y. Tikhonchuk, et al., Overcoming lability of extremely long alkane carbon–carbon bonds through dispersion forces. *Nature* **2011**, *477*, 308.
[80] B. L. Schottel, H. T. Chifotides, K. R. Dunbar, Anion-π interactions. *Chem. Soc. Rev.* **2008**, *37*, 68.

[81] T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, et al., CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **2020**, *152*, 194103.