Accurate Interaction Energies by Spin Component Scaled Möller-Plesset Second Order Perturbation Theory Calculations with Optimized Basis Sets (SCS-MP2^{mod}): development and application to aromatic heterocycles.

Ivo Cacelli,[a,b,*] Filippo Lipparini,[a] Leandro Greff da Silveira,[c]
Matheus Jacobs,[d,e] Paolo Roberto Livotto[c] and Giacomo Prampolini[b]

[a]*Dipartimento di Chimica e Chimica Industriale, Università di Pisa*,
Via G. Moruzzi 13, I-56124 Pisa, Italy
[b]*Istituto di Chimica dei Composti OrganoMetallici* (ICCOM-CNR),
Area della Ricerca, via G. Moruzzi 1, I-56124 Pisa, Italy
[c]*Instituto de Química, Universidade Federal do Rio Grande do Sul*,
Avenida Bento Gonçalves 9500, CEP 91501-970 Porto Alegre, Brazil
[d]*Institut für Physik, Humboldt-Universität zu Berlin*,
Newtonstr. 15, 12489, Berlin, Germany
[e]*IRIS Adelrshof, Humboldt-Universität zu Berlin*,
Zum Großen Windkanal 6, 12489, Berlin, Germany

April 18, 2019

Abstract

The Spin Component Scaled (SCS) MP2 method using a reduced and optimized basis set (SCS-MP2^{mod}) is employed to compute the interaction energies of nine homo-dimers, formed by aromatic hetero-cyclic molecules (pyrrole, furan, thiophene, oxazole, isoxazole, pyridine, pyridazine, pyrimidine and pyrazine). The coefficients of the same-spin and opposite-spin correlation energies and the GTO polarization exponents of the 6-31G** basis set are simultaneously optimized in order to minimize the energy differences with respect to the CCSD(T) reference interaction energies, extrapolated to complete basis set. It is demonstrated that the optimization of the spin scale factors leads to a noticeable improvement of the accuracy with a root mean square deviation less than 0.1 kcal/mol and a largest unsigned deviation smaller than 0.25 kcal/mol. The pyrrole dimer provides an exception, with slightly higher deviation from the reference data. Given the high benefit in terms of computational time with respect to the CCSD(T) technique and the small loss of accuracy, the SCS-MP2^{mod} method appears to be particularly indicated for extensive sampling of intermolecular potential energy surfaces at a quantum mechanical level. Within this framework, the results of exponents and scaling factors optimization for the whole set of molecules are again accurate, showing a good level of transferability to this class of molecules.

1 Introduction

Intermolecular interactions are responsible for the existence of any type of condensed matter, and therefore they have received much attention in the fields of physics, chemistry, biology, nano-sciences and engineering.[1–4] An accurate determination of the forces taking place when two or more molecules approach each other, allows in principle for a rationalization of a number of experimental observations, including bulk phase properties. This task is however rather difficult, as the most reliable methods, as those based on the Coupled Cluster (CC) Ansatz, are limited to medium-sized molecules and to a moderate number of points of the interaction potential energy surface (IPES), due to their high computational cost. For this reason a number of approaches based on cheaper methods exploiting both wave function[5–11] and DFT families[12–16]were developed in recent years (see also Ref. [17] for an excellent review). In this paper we give a contribution to this research line with the aim to improve a general method to compute bulk phase properties, as detailed in the following.

Several years ago we developed in our group a protocol to calculate the bulk properties of pure substances by Molecular Dynamics (MD) and/or Monte Carlo (MC) simulation methods using a Force Field (FF) completely derived by *ab initio* quantum mechanical calculations.[18–25]The parameters entering the FF were derived separately for the intra- and inter-molecular interactions respectively using the Joyce [26–28] and Picky [22–24] protocols. The former task can be accomplished with a negligible computational cost, as it only requires the equilibrium geometry and Hessian as derived for the isolated target molecules, which can be computed with DFT. On the contrary, a robust determination of the intermolecular FF parameters requires a large number of reference calculations at hundreds of points on the IPES. This step is very time consuming and a manageable although accurate quantum mechanical method is mandatory for the feasibility of the whole protocol. In a first attempt performed on the benzene molecule[18] we have used the MP2 method with a small 6-31G* basis set, by setting the exponents of the polarization GTO functions of the carbon atom to 0.25, as suggested in previous works.[29–31] Further experience was gained in a challenging calculation of nematogenic molecule in its condensed phases[20] and of IPES portions of the quinhydrone dimer and eumelanin building blocks.[32,33] Since then, the name MP2^{mod} was proposed to indicate MP2 calculations with modified basis sets, suitable for intermolecular calculations. In a very recent work,[25,34] we applied the same method to a variety of homo-dimers, formed by hetero-cyclic aromatic molecules. The GTO exponents of the polarization functions (d for N,C,O and p for H atoms) of a 6-31G** basis set were optimized by fitting the MP2^{mod} results to CCSD(T) energies, extrapolated at in the complete basis set (CBS) limit, at a number of dimer geometries for several aromatic hetero-cycles, namely pyrrole, furan, thiophene, oxazole, isoxazole, pyridine, pyridazine, pyrimidine and pyrazine.[34] The resulting system specific basis sets were then used in MP2^{mod} calculations for about 500 points of the IPES for each system, sorted with the PICKY protocol. The FF parameters were then determined exploiting this extended MP2^{mod} database and successively used in MD to compute the bulk properties.[25]

In this paper, we calculate the intermolecular energies for several homo- and hetero-dimers of the aforementioned species, using again the MP2^{mod}/6-31G** model, but refining it with the use of Spin Component Scaled (SCS) correlation energies. By this method, first proposed by Grimme,[5] the correlation energy as obtained by standard MP2 calculations is recomputed by weighting the same-spin (SS) and opposite-spin (OS) components with two independent scaling factors

$$E_{corr}(SCS-MP2) = C_{SS}E_{SS} + C_{OS}E_{OS}$$

Although this procedure was designed to improve the MP2 results on isolated molecules, SCS-MP2 was successively considered in intermolecular calculations, in several attempts aimed to find the optimal scaling coefficients for binding energy calculations.[6,17,35–38] The main conclusions of these studies may be summarized as follows.

1. By exploiting the advantage of tuning two independent parameters, SCS-MP2 performs better than MP2, as the root mean square deviation from accurate data roughly halves.[35,36]

2. The optimal scaling factors are very different from those found for single molecule calculations and, specifically, the same-spin component is enhanced whereas the opposite-spin component is reduced, and in some cases may become even negative. For specific systems only the same-spin component was used.[38]

In this paper we study the aforementioned hetero-aromatic dimers in several conformations with the aim to improve the previous results by optimizing both the exponents of the polarization GTO functions and the MP2 spin scaling factors. The parameters for SCS-MP2 were first optimized specifically for each considered homo-dimer, aiming to reach the best possible accuracy with respect to reference data. In a attempt to check the transferability of the exponents and the spin scaling factors further calculations were performed. In the first one all the five membered ring molecules were constrained to take the same parameters; in the second one the same was done for the six membered ring systems and finally, all molecules here considered takes the same exponents and scaling factors through a fitting using all the reference data.

2 Method and Computational Details

The same-spin (SS) and opposite-spin (OS) components of the MP2 correlation energy of a closed shell system have the following expression

$$E_{SS} = -\sum_{ijab} \frac{\langle ij|ab\rangle(\langle ij|ab\rangle - \langle ij|ba\rangle)}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \tag{1}$$

$$E_{OS} = -\sum_{ijab} \frac{|\langle ij|ab\rangle|^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \tag{2}$$

where i, j and a, b run on the occupied and virtual spatial orbitals, respectively. The two-electron integrals are always written in the Dirac notation. The Spin Component Scaled (SCS) correlation energy is a linear combination of both components through the scaling factors

$$E_{corr} = C_{SS}E_{SS} + C_{OS}E_{OS} \tag{3}$$

The optimization of both the exponents of the GTO polarization functions of the 6-31G** basis set and the scaling factors C_{SS} and C_{OS}, was performed by fitting the resulting interaction energies with a reference database determined by CCSD(T)/CBS results, often defined as quantum chemical gold standard.[39]

In most of its current implementations,[40] the CCSD(T)/CBS interaction energy can be recovered from the MP2/CBS limit (ΔE_{CBS}^{MP2}), corrected with the difference ($\Delta CCSD(T)$) between the MP2 and CCSD(T) interaction energy, computed with a smaller X basis set:

$$\Delta E_{CBS}^{CCSD(T)} = \Delta E_{CBS}^{MP2} + \Delta E_{X}^{CCSD(T)} - \Delta E_{X}^{MP2} \tag{4}$$

Although this procedure has been extensively used by several groups (see references [39–45], just to cite a few), the best route to correctly estimate both ΔE_{CBS}^{MP2} and $\Delta CCSD(T)$ values has not yet been uniquely assessed. In consideration of the large number of CCSD(T) calculations required, in this paper the MP2 interaction energy at CBS was estimated by the Halkier extrapolation scheme[46] employing the aug-cc-pVDZ and aug-cc-pVTZ basis sets, whereas the $\Delta CCSD(T)$ correction was computed exploiting the aug-cc-pVDZ basis set. Indeed, in previous work,[34] it was found, at least for the benzene dimer, that this choice did not introduce appreciable differences (0.01 kcal/mol) with respect to more refined schemes,[44,45] which employed aug-cc-pVTZ and aug-cc-pVQZ basis set in the Halkier extrapolation, and heavy-aug-cc-pVDZ/heavy-aug-cc-pVTZ two-point extrapolation basis set for $\Delta CCSD(T)$.

The interaction energy ΔE of a dimer $A...B$ is defined as

$$\Delta E = E_{AB} - E_{A} - E_{B} \tag{5}$$

where the E's are the absolute energies of the super-molecule or of the single molecules for any type of calculation. In all CCSD(T) and MP2 calculations the basis set superposition error (BSSE) was taken into account by the standard Counterpoise (CP) correction[47] so that all energies entering the above expression are computed with the basis set of the AB molecule at the considered geometry. The complete interaction energy of a SCS-MP2 calculation is evaluated by the Hartree-Fock, same-spin and opposite-spin interaction energies

$$\Delta E^{SCS-MP2} = \Delta E^{HF} + C_{SS}\Delta E_{SS} + C_{OS}\Delta E_{OS} \tag{6}$$

The equilibrium geometry of each isolated monomer was taken from Ref. [34], where it was obtained by geometry optimization at DFT level, using the B3LYP functional joined with the Dunning's correlation consistent cc-pVTZ basis set. All the calculations concerning monomers and homo-dimers were performed with the GAUSSIAN09 suite of programs.[48] All the calculations concerning hetero-dimers were performed with a locally modified version of the CFOUR suite of programs.[49,50] Geometry optimizations for the hetero-dimers at the SCS-MP2^{mod} level of theory were performed with CFOUR using analytical gradients.[51]

In system-specific parameterizations, the GTO exponents and the scaling factors were optimized separately for each system and the reference dimer database is formed by 4 or 5 curves, corresponding to the most common arrangements for the considered homo-dimer reported in literature: face-to-face (both parallel and anti-parallel), T-shaped, parallel displaced and other system specific arrangements which span attractive regions of the IPES. A complete collection of the dimer geometries is shown in the Supporting Information, along with the detailed instructions to get the geometrical arrangements from the translational vectors and the Euler angles defined therein in Figure A. Three additional optimizations were attempted, with the aim to obtain more general (yet probably less accurate) parameters, possibly suitable for interaction energy SCS-MP2 calculations in any pair of hetero-cycles. Namely, two parameterizations were performed with respect to the database built with all the CCSD(T)/CBS interaction energies computed for either the five- or six-membered ring systems, whereas a final one was performed over the database comprising the whole set of computed CCSD(T)/CBS data. In all cases, all molecules pertaining to the chosen training set were equipped with the same exponents and scaling factors, thus offering in principle the possibility to use SCS-MP2^{mod} for any pair of five-membered, six-membered or generic hetero-cyclic aromatic molecules, respectively.

For each parameterization, the GTO exponents and scaling factors (herein also referred to as the parameters) were optimized by minimizing the Root Mean Square Deviation (RMSD) of the MP2^{mod} energies with respect to the CCSD(T)/CBS ones,

$$\text{RMSD}(\alpha, C_{SS}, C_{OS}) = \sum_{i=1}^{N_p} [E_i^{MP2}(\alpha, C_{SS}, C_{OS}) - E_i^{(CCSD(T)/CBS]}]^2 / N_p \qquad (7)$$

where N_p is the number of geometrical arrangements of the system (or systems) under scrutiny and α indicates the collection of exponents of the polarization GTOs. The best-fit parameters were determined using the "in house" program EXOPT. At each optimization cycle one exponent is varied, the new MP2^{mod} energy components are computed for all dimer arrangements and the scaling factors are optimized to give the new value of the RMSD. The C values were constrained to be positive or null, whereas the GTO exponents were confined in the 0.05 - 1.5 Bohr^{-2} range.

3 Results

3.1 System-specific SCS-MP2^{mod}

In this section we will present the results of the optimization of both basis set exponents and spin component scaling factor carried out separately for each molecule, i.e. using as reference CCSD(T)/CBS interaction energies of a specific homo-dimer in different arrangements. For each considered species, different classes of homo-dimer geometrical dispositions were hence devised, by displacing one monomer with respect to the other (see Supporting Information for a detailed definition). First, a preliminary interaction energy curve was obtained at MP2^{mod} level[34] for each arrangement, by displacing one monomer along the translation vector $\vec{R}$. Next, a number of geometries was selected along the curve, and their interaction energy computed at CCSD(T)/CBS level. The latter values were then used as a reference for the optimization procedure, performed separately for each species. An example of the resulting curves is displayed for pyridine in Figure 1, whereas all the remaining curves are shown in Figures **B-J** in the Supporting Information.

The main results of the present work are graphically summarized in Figure 2 and collected in Table 1. The very good agreement between SCS-MP2^{mod}energies and CC reference data clearly appears from the bottom panel of Figure 2, where the correlation plot between the two set of energies is displayed for all compounds. A more detailed analysis, taking in consideration the different geometrical classes for each dimer, is displayed in the Supporting Information (Figure K). In general, as evidenced from the top panel of Figure 2, all SCS-MP2^{mod} energies are less than 0.4 kcal/mol different from the reference CCSD(T)/CBS counterparts.

On the same foot, the statistical analysis reported in Table 1 confirms the close agreement between the representation offered by the current SCS-MP2^{mod} model and the reference CCSD(T)/CBS database. In fact, the noticeable improvement provided by tuning the SCS scaling factors is apparent: all the usual statistical descriptors show a markedly reduced error, with a particular improvement of the LUD, which on average becomes 1/3 of that given by the standard MP2^{mod} scaling factors. As already found in previous

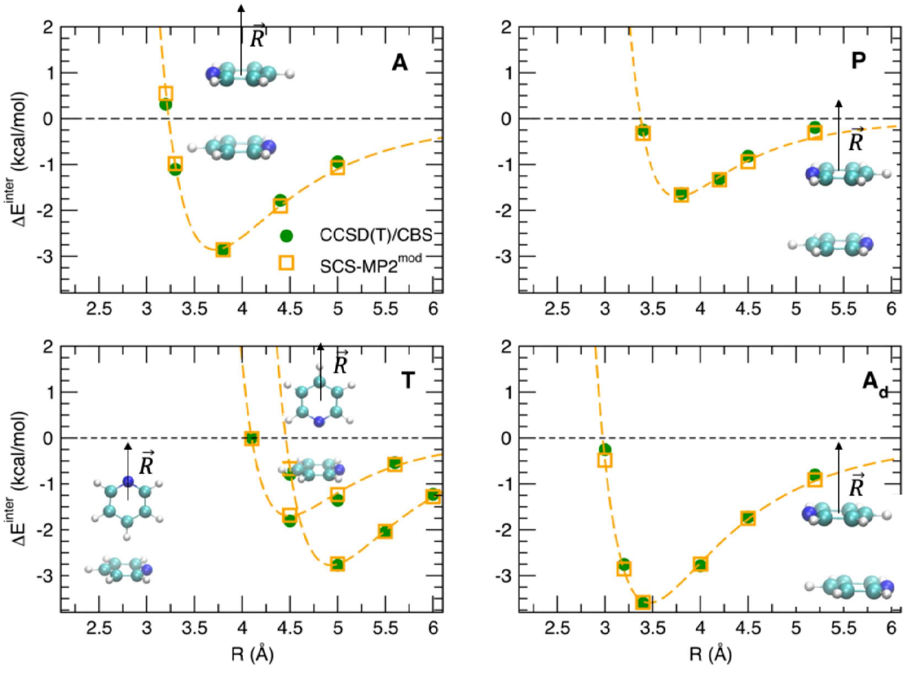

Figure 1: CCSD(T)/CBS and SCS-MP2^{mod} interaction energies computed for the pyridine dimer arrangements shown in the insets.

papers,[35,36] this means that the two scaling factors are significant parameters that allow for better flexibility and thus an improved agreement of the SCS-MP2^{mod} energy to the CCSD(T)/CBS reference.

The ratio $R = \Delta E_{SS}/\Delta E_{OS}$ averaged along the curves of the several arrangement classes, reported in Table 2, indicates a dependence on the relative orientation of the two molecules. Although it is hard to find a well defined rationale, it seems that the most relevant geometrical feature affecting R is given by the closest atoms in the dimer belonging to different monomers. The main conclusion is that in the cases where one or two H atoms of one monomer approach the π cloud of the other one, ΔE_{SS} becomes larger than ΔE_{OS} and R reaches values up to 1.3. In all other cases R is smaller than one. Moreover, the ratio is almost constant along the abscissa of the single curves, showing a

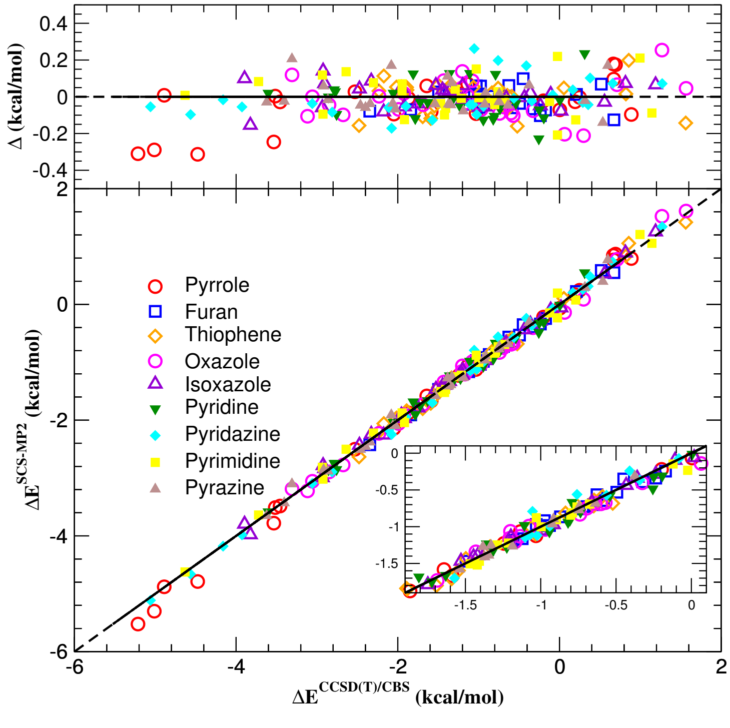

Figure 2: Top: Signed error ($\Delta = \Delta E^{SCS-MP2}$ - $\Delta E^{CCSD(T)}$) (kcal/mol) between SCS-MP2^{mod} computed values and CCSD(T)/CBS reference energies. Bottom: Correlation plot between SCS-MP2^{mod} energies and reference CCSD(T)/CBS data for the investigated aromatic hetero-cycles. The correlation of the [-2 – 0] kcal/mol range is evidenced in the inset.

		NO-SCS			SCS		
	N.pt	RMSD	MAD	LUD	RMSD	MAD	LUD
Pyrrole	23	0.410	0.230	0.980	0.141	0.101	0.313
Furan	24	0.198	0.142	0.601	0.061	0.052	0.125
Thiophene	26	0.204	0.129	0.704	0.085	0.068	0.198
Oxazole	26	0.229	0.163	0.642	0.101	0.079	0.254
Isoxazole	28	0.428	0.332	1.120	0.069	0.059	0.153
Pyridine	25	0.216	0.154	0.658	0.104	0.083	0.235
Pyridazine	26	0.237	0.159	0.716	0.098	0.077	0.262
Pyrimidine	27	0.250	0.168	0.862	0.102	0.080	0.220
Pyrazine	26	0.285	0.215	0.591	0 .082	0.065	0.206

Table 1: Results of the MP2^{mod} GTO exponents and scaling factors from individual optimization in terms of: Root Mean Square Deviation (RMSD), Mean Absolute Deviation (MAD) and Largest Unsigned Deviation (LUD) with respect to interaction energies computed at CCSD(T)/CBS level. The MP2^{mod} basis set is the modified 6-31G**. All energies are in kcal/mol.

dimer geometrical arrangement	$R = \Delta E_{SS}/\Delta E_{OS}$
parallel	0.7 - 0.9
parallel + lateral displacement	0.9 - 1.1
T-shaped (H pointing towards π cloud)	1.1 - 1.3
T-shaped (N,O,S pointing towards π cloud)	0.9 - 0-95

Table 2: Mean value of the $R = \Delta E_{SS}/\Delta E_{OS}$ along the several geometrical arrangement considered in the present work.

marked dependence on the relative orientation, rather than on the distance. By looking at equations (1) and (2), the difference between the two spin component interaction energies is roughly given by

$$\Delta E_{OS} - \Delta E_{SS} \simeq \sum_{ia}^{A} \sum_{jb}^{B} \frac{\langle ij|ab\rangle\langle ij|ba\rangle}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \tag{8}$$

where A and B indicates the set of orbitals of the two monomers. This formula may be a reasonable approximation in the hypothesis that the molecular orbitals are strongly localized on one monomer and that they resemble those of the single molecule. This approximate expression includes a number of terms with opposite sign and is rather difficult to be exploited in order to rationalize the observed R values.

In general, the energy residues are rather equally distributed along the dimer arrangements, meaning that different dimer classes are represented with the same quality by SCS-MP2^{mod}, as also evident from a visual inspection of Figure 2. With few exceptions, the short distance repulsive points are those showing the largest differences with respect to CCSD(T)/CBS energies. However, for such points the correlation energy is rather large and the extrapolation scheme might not be fully applicable in such a regime, making the reference energies less reliable. The least accurate performances of SCS-MP2^{mod} are observed for pyrrole, in all the three statistical indicators. As this system is able to give rise to one or two hydrogen bonds and arrangements that allow such an interaction were included in the database, pyrrole dimers were further investigated. Concretely, the GTO p exponent of the H atom bonded to Nitrogen and thus able to participate in hydrogen bonds was optimized as a further, independent parameter. The results were not encouraging as the RMSD goes from 0.141 to 0.136 and the LUD goes from 0.313 to 0.300 kcal/mol. Therefore, the proposed strategy for improving the results is ineffective and the origin of the unusual higher deviations from CCSD(T)/CBS found for pyrrole with respect to the other systems seems to call for further investigations.

Nonetheless, despite the slightly worse performance registered for the pyrrole dimer, it is evident from the top panel of Figure 2 that the SCS-MP2^{mod} procedure is able to deliver a very accurate estimate ($<$ 0.4 kcal/mol) of the interaction energies of the investigated dimers in a wide range of values (from -7 to +1 kcal/mol), as well as a balanced description of the relative stability of different arrangements. A further point of strength, which becomes crucial in an extensive sampling of dimer IPESs, consists in the computational benefit with respect to high level techniques. Indeed, as it appears from Table 3, the cost in terms of CPU time is decreased by three orders of magnitude with respect to a corresponding CCSD(T)/CBS calculation, thus allowing for the computation of hundreds of IPES points.

The optimal exponents and scaling factors for SCS-MP2^{mod} are reported in Table 4 separately for each investigated species. When SCS is not activated (left data in Table 4), both C_{SS} and C_{OS} scaling factors are fixed to 1.0 as in the traditional MP2^{mod}. It was

Method	basis set	CPU time (m)	
		five membered rings	six membered rings
MP2	aug-cc-pVDZ	45	90
MP2	aug-cc-pVTZ	1200	1700
CCSD(T)	aug-cc-pVDZ	4100	7800
CCSD(T)	CBS	5300	9600
SCS-MP2^{mod}	6-31G**	5	10

Table 3: Average CPU times on a single Xeon(R) 2.6 GHz processor employed in the calculation of dimer interaction energies with different methods.

	NO-SCS				SCS					
	α_C	α_N	$\alpha_{O/S}$	α_H	α_C	α_N	$\alpha_{O/S}$	α_H	C_{SS}	C_{OS}
Pyrrole	0.244	0.364		0.546	0.216	0.455		0.946	1.259	0.941
Furan	0.217		0.326	0.187	0.598		0.195	0.399	1.874	0.633
Thiophene	0.230		0.218	0.341	0.263		0.231	1.500	1.515	0.639
Oxazole	0.247	0.266	0.368	0.176	0.690	0.562	0.405	0.588	1.771	0.692
Isoxazole	0.207	0.312	0.350	0.324	0.904	0.645	0.237	1.500	2.570	0.157
Pyridine	0.366	0.184		0.206	1.216	0.205		0.593	2.629	0.000
Pyridazine	0.482	0.266		0.207	1.240	0.439		0.184	2.161	0.180
Pyrimidine	0.415	0.301		0.219	0.903	0.653		0.328	2.444	0.000
Pyrazine	0.465	0.330		0.180	1.101	0.731		0.175	1.945	0.420

Table 4: Orbital exponents of the polarization GTOs and scaling factors of the MP2^{mod} optimization of the 6-31** basis set. All energies are in kcal/mol.

noticed that the parameters corresponding to the GTO exponents show a certain degree of redundancy, so that similar values for $RMSD$ in equation (7) may sometimes be found at different values of the parameters, i.e. there might be some local minima rather close to each other. This implies that the specific GTO exponents have no particular meaning when considered independently of each other: it is only the full set of exponents that defines the 6-31G** basis set as a whole that makes it suitable for computing molecular interactions with the SCS-MP2^{mod} scheme.

A further interesting result of the present work concerns with the scaling factors. In all cases the SS coefficient is higher than the OS one, but the behavior is different for the two sets of dimers formed by five member rings and six member rings. With the exception of isoxazole, in the first group C_{SS} is always in the 1.2 - 1.9 range and C_{OS} takes values greater than 0.6, whereas in the second group C_{SS} is even more dominant with values close or greater than 2 whereas C_{OS} is rather small and would become even negative after removal of the imposed constraint to be positive or null. For all systems the sum of the scaling factors is in the 2.1 - 2.7 range. Apart from some specificities connected both with the type of dimer and with the geometrical arrangement, a conclusion of the present study is that in all cases C_{SS} is the most relevant coefficient. This finding is in line with practically all the authors who studied the intermolecular interactions by the SCS-MP2 method. In particular, Antony and Grimme[35] compared results given by several scaling protocols of MP2 for the S22 data set, and found that the SCS-MP2[37,38] ($C_{OS} = 0$ and $C_{SS} = 1.76$) is the best performing one. The present results are also in line with the conclusions of Distasio and Head-Gordon[36] who found the optimal SCS-MP2 scaling factors to be $C_{SS} = 1.29$ and $C_{OS} = 0.40$ for the S22 data set with CCSD(T)/CBS reference data. In the case of cc-pVTZ basis set the same authors found that $C_{SS} = 1.75$ and $C_{OS} = 0.17$ are the best scaling factors, that are similar to the optimal values found in the present work. The systematic higher contribution of ΔE_{SS} to the intermolecular energies at SCS-MP2 level is rather difficult to be rationalized. The considerations of Grimme et al.[13] concerning a connection of E_{SS} and E_{OS} with the long-range and short-range correlation effects, respectively, are stimulating and can furnish a sort of rationale, although they seem more adequate for intra-molecular correlation energies. In fact the dispersion energy (8) originates from simultaneous excitations of the two monomers and the short-range long-range distinction appears to be rather weak. For our purposes, in line with the empirical motivations of the SCS strategy, this systematic higher contribution of ΔE_{SS} can be considered for the moment a fortuitous feature which allows to greatly improve the accuracy of the intermolecular energies with no additional computational cost.

One more comment on the results of Table 4 concerns with the high values of the sum of the scaling factors, joined with the general increase of the GTO exponents when using SCS. The optimal GTO exponents without using SCS are rather low for C,N,O atoms and higher for the p functions of the Hydrogen atoms. The former are in line with the first attempt to evaluate the intermolecular energy of dispersion driven complexes by MP2 with modified basis sets,[18,22,29–31] where the GTO exponent of the d function of the Carbon atoms was set to 0.25, below the one employed in standard single molecule calculations. On the contrary, when optimized for SCS-MP2, the GTO exponents tend in general to increase and in some cases they reach values comparable with those optimized for single molecule calculations. High exponents are suitable to describe the dynamic electronic correlation but can not correctly represent the polarization induced by the second molecule and give a small contribution to the molecular polarizability, which is a crucial ingredient in the evaluation of the dispersion energy. Therefore such exponent rise corresponds to a decrease of the MP2 interaction energies for both the same-spin and opposite-spin components. For example, the mean value of the same-spin and opposite-spin interaction energies of Pyridazine are -2.33 and -2.58 kcal/mol with the NO-SCS basis set and -2.13 and -2.37 kcal/mol for the SCS basis set, respectively. Thus it is apparent that in the SCS-MP2 calculations the interaction energy without scaling (i.e. using $C_{SS} = C_{OS} = 1$) would be underestimated. As a consequence, the scaling factors need to be increased in order to match the reference values. In turn this implies that their sum ranges values between 2.1 and 2.4, in contrast with the general prescription, based on physical considerations of long range molecular interactions, that for intermolecular interactions $C_{SS} + C_{OS} = 2$.[10,17,52]

3.2 Transferable SCS-MP2^{mod}

While the MP2^{mod}and SCS-MP2^{mod} protocols were originally devised to find the best set of parameters for a specific homo-dimer, it is interesting to investigate to what extent the SCS-MP2^{mod} method can be applied to a wider range of targets using the same basis set and scaling factors for all the nine molecules here considered. Indeed, due to the limited computational cost of the SCS-MP2^{mod} calculations, it would be greatly beneficial to employ previously optimized basis sets, without performing any additional tuning through CCSD(T) reference data. Therefore, in this section, we give up system specific basis set and scaling factors and check the possibility of using the same parameter set for a wide range of systems. To this end, a new reference CCSD(T)/CBS database, consisting in all dimer interaction energies with the exception of the more repulsive points, was employed to obtain a unique set of SCS-MP2^{mod} parameters, i.e. five exponents for the 6-31G** polarization functions and two SCS scaling factors, which can be in principle used for any pair of hetero-cyclic here considered. All parameters, as well as the RMSD and LUD indicators are reported in Table 5.

C_{SS}	C_{OS}	α_C	α_N	α_O	α_S	α_H	RMSD	LUD
2.061	0.018	0.274	0.481	0.348	0.230	0.183	0.199	0.707

Table 5: Orbital exponents of the polarization GTOs and scaling factors of the global SCS-MP2^{mod} optimization of the 6-31G** basis set. RMSD and LUD are in kcal/mol.

The exponents take reasonable values and also the scaling factors are similar to those obtained for the system specific optimizations. As expected, the statistical descriptors take less favorable values than the previous ones, although the accuracy is still satisfactory. In particular by looking at the data of Table 1 it is apparent that both the RMSD and the LUD are still better than the system specific optimizations MP2^{mod} , i.e. without tuning the scaling factors.

To test the reliability of this global parameter set, a benchmark set of aromatic hetero-dimers was created by picking two different species among the compounds investigated in this work. Concretely, twelve different pairs were considered, namely Furan – Isoxazole, Furan – Oxazole, Thiophene – Isoxazole, Thiophene – Furan, Furan – Pyrimidine, Thiophene – Pyridazine, Oxazole – Pyridine, Isoxazole – Pyrazine, Pyridine – Furan, Pyridine – Pyrimidine, Pyridine – Pyridazine and Pyrazine – Pyrimidine. For each dimer, two attractive geometries were considered, one in a stacked arrangement and one in a T-shaped one. All considered geometries are shown in detail in Figures L and M in the Supporting Information. The interaction energy of each of the resulting 24 conformers was computed at CCSD(T)/CBS and SCS-MP2^{mod} level. Additionally, to better evaluate the gain in accuracy while maintaining the same computational cost, the interaction energy of

	MP2			SCS-MP2^{mod}		
Set	RMSD	MAD	LUD	RMSD	MAD	LUD
two 5-membered rings	2.30	2.23	3.07	0.13	0.11	0.22
5-membered and 6-membered rings	2.16	2.04	2.78	0.23	0.06	0.52
two 6-membered rings	2.04	2.02	2.37	0.40	0.20	0.67
Stacked arrangements	2.58	2.57	3.07	0.32	0.14	0.67
T-shaped arrangements	1.68	1.63	1.95	0.17	0.07	0.39
all heterodimers	2.17	2.10	3.07	0.26	0.04	0.67

Table 6: Comparison, with respect of with reference CCSD(T)/CBS values, between standard MP2 calculations carried out with the 6-31G** basis set and SCS-MP2^{mod} calculations performed with the modified 6-31G** basis set . The 24 dimer arrangements discussed in the text and shown in Figures L and M in the Supporting Information are considered as reference database. RMSD, MAD and LUD are all reported in kcal/mol.

all arrangements were also computed using the standard MP2 method with the original 6-31G** basis set. Results are reported in Figure 3 and summarized in Table 6.

Apart from the impressive improvement of the present results with respect to MP2/6-31G** calculations, we may notice that the accuracy is rather balanced within the several types of dimer and geometries. More in detail, the accuracy is well balanced among the hetero-dimers formed by either two 5-membered rings or one 5- and one 6-membered, whereas slightly less accurate prediction were obtained for the 6-membered rings hetero-dimers. As far as the geometrical arrangement is concerned, the performance is slightly better for the T-shaped than for the stacked arrangements, even if the limited number of dimer arrangements here considered does not allow any precise conclusion about the quality of the results as related to a specific geometry. Yet, this last test seems to confirm that the transferable basis set and scaling factors are able to deliver rather accurate estimates of the interaction energy. Considering the that choice of the pairs reported in Table 6 was random, a similar accuracy can reasonably be expected for any different pair of aromatic hetero-cycles at any geometry. This is of valuable importance in all situations where a relevant portion of the IPES should be explored for a wide range of molecular pairs.

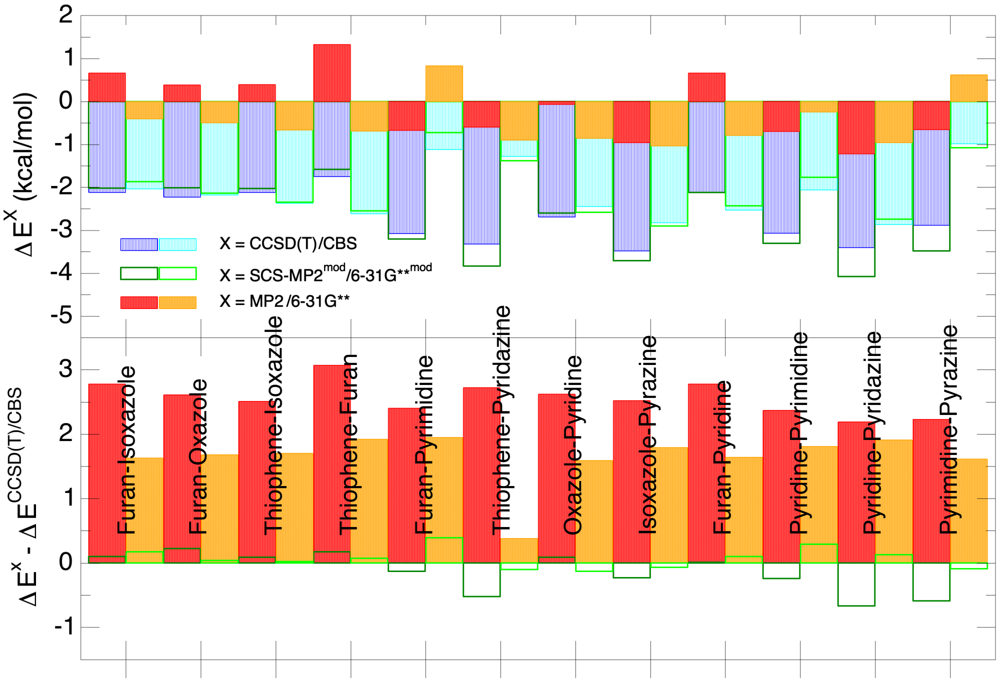

Figure 3: CCSD(T)/CBS (blueish bars), SCS-MP2^{mod} (greenish) and MP2/6-31G** (reddish) interaction energies computed for the considered hetero-dimers in stacked (blue, green and red) and T-shaped (cyan, light green and orange) geometries (see Figure L and M in Supporting Information). Top: interaction energies; bottom: relative error with respect to CCSD(T)/CBS. All energies are in kcal/mol.

4 Conclusions

In this paper we have presented a method to calculate the intermolecular energy of an important class of molecules. The main goal of this work is devising a cheap theoretical method, capable of delivering results at a level of accuracy sufficient to be profitably used in exploring large portions of the IPES. The computational convenience relies on joining the perturbative MP2 method with a modified basis set 6-31G**, with the further inexpensive expedient of tuning the scaling factors of the the same-spin and opposite-spin components of the correlation energy. The main conclusion of the present effort is that the optimization of these scaling factors in the SCS-MP2^{mod} has very relevant effects on the accuracy of the results. The most accurate results were obtained for system specific optimization of the basis set and scaling factors, although in our opinion, even the use of a unique set of parameters for all the molecules here considered leads to satisfactory results. This generalization opens the way to the possibility of computing IPES for any pair of molecules so enlarging the usefulness of SCS-MP2^{mod} to a rather large chemical situations.

Supporting Information

Additional data and several details about the reported calculations not included in this paper are available: dimer geometries, SCS-MP2^{mod} interaction energy curves, CCSD(T)/SCS-MP2^{mod} correlation plots and starting and optimized SCS-MP2^{mod} geometries of the considered hetero-dimers.

References

[1] Stone, A. J. The Theory of Intermolecular Forces; Oxford University Press: Oxford, 1996.

[2] Parsegian, V. A. Van der Waals Forces: A Handbook for Biologists, Chemists, Engineers, and Physicists; Cambridge University Press: Cambridge, 2005.

[3] Kaplan, I. Intermolecular Interactions: Physical Picture, Computational Methods and Model Potentials; Wiley: Chichester, West Sussex, England, 2006.

[4] Israelachvili, J. Intermolecular and Surface Forces. Third Edition; Academic Press: London, 2010.

[5] Grimme, S. Improved second-order MøllerPlesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies *J. Chem. Phys.* **2003**, *118*, 9095.

[6] Marchetti, O.; Werner, H.-J. Accurate calculations of intermolecular interaction energies using explicitly correlated wave functions *Phys. Chem. Chem. Phys.* **2008**, *10*, 3400–3409.

[7] Pitonák, M.; Neogrády, P.; Cerný, J.; Grimme, S.; Hobza, P. Scaled MP3 noncovalent interaction energies agree closely with accurate CCSD(T) benchmark data. *ChemPhysChem.* **2009**, *10*, 282–9.

[8] Pitonak, M.; Heßelmann, A. Accurate Intermolecular Interaction Energies from a Combination of MP2 and TDDFT Response Theory *J. Chem. Theory Comput.* **2010**, *6*, 168–178.

[9] Riley, K. E.; Platts, J. A.; Rezáč, J.; Hobza, P.; Hill, J. G. Assessment of the Performance of MP2 and MP2 Variants for the Treatment of Noncovalent Interactions. *J. Phys. Chem. A* **2012**, *116*, 4159–69.

[10] Grimme, S.; Goerick, L.; Fink, R. F. Spin-component-scaled electron correlation methods *Advanced Review* **2012**, *2*, 886–906.

[11] Goldey, M. B.; Belzunces, B.; Head-Gordon, M. Attenuated MP2 with a Long-Range Dispersion Correction for Treating Nonbonded Interactions *J. Chem. Theory Comput.* **2015**, *11*, 4159–4168.

[12] Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-class Functionals and 12 Other Function *Theor. Chem. Accounts* **2008**, *120*, 215–241.

[13] Grimme, S. Density Functional Theory with London Dispersion Corrections *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 211.

[14] Burns, L. a.; Vázquez-Mayagoitia, A.; Sumpter, B. G.; Sherrill, C. D. Density-Functional Approaches to Noncovalent Interactions: a Comparison of Dispersion Corrections (DFT-D), Exchange-Hole Dipole Moment (XDM) Theory, and Specialized Functionals. *J. Chem. Phys.* **2011**, *134*, 084107.

[15] Antony, J.; Sure, R.; Grimme, S. Using Dispersion-Corrected Density Functional Theory to Understand Supramolecular Binding Thermodynamics. *Chem. Comm.* **2014**, *51*, 1764–1774.

[16] Brandenburg, J. G.; Hochheim, M.; Bredow, T.; Grimme, S. Low-Cost Quantum Chemical Methods for Noncovalent Interactions *J. Phys. Chem. Lett.* **2014**, *5*, 4275–4284.

[17] Tan, S.; Acevedo, S. B.; Izgorodina, E. I. Generalized spin-ratio scaled MP2 method for accurate prediction of intermolecular interactions for neutral and ionic species *J. Chem. Phys.* **2017**, *146*, 064108.

[18] Cacelli, I.; Cinacchi, G.; Prampolini, G.; Tani, A. Computer Simulation of Solid and Liquid Benzene with an Atomistic Interaction Potential Derived from Ab Initio Calculations. *J. Am. Chem. Soc.* **2004**, *126*, 14278–86.

[19] Prampolini, G. Parametrization and Validation of Coarse Grained Force-Fields Derived from ab Initio Calculations *J. Chem. Theory Comput.* **2006**, *2*, 556–567.

[20] Cacelli, I.; Lami, C. F.; Prampolini, G. Force-field Modeling through Quantum Mechanical Calculations: Molecular Dynamics Simulations of a Nematogenic Molecule in its Condensed Phases. *J. Comp. Chem.* **2009**, *30*, 366–378.

[21] Cacelli, I.; Cimoli, A.; De Gaetani, L.; Prampolini, G.; Tani, A. Chemical Detail Force Fields for Mesogenic Molecules *J. Chem. Theory Comput.* **2009**, *5*, 1865–1876.

[22] Cacelli, I.; Cimoli, A.; Livotto, P. R.; Prampolini, G. An Automated Approach for the Parameterization of Accurate Intermolecular Force-Fields: Pyridine as a Case Study. *J. Comp. Chem.* **2012**, *33*, 1055.

[23] Prampolini, G.; Livotto, P. R.; Cacelli, I. Accuracy of Quantum Mechanically Derived Force-Fields Parameterized from Dispersion-Corrected DFT Data: The Benzene Dimer as a Prototype for Aromatic Interactions. *J. Chem. Theory Comput.* **2015**, *11*, 5182–96.

[24] Prampolini, G.; Campetella, M.; De Mitri, N.; Livotto, P. R.; Cacelli, I. Systematic and Automated Development of Quantum Mechanically Derived Force Fields: The Challenging Case of Halogenated Hydrocarbons *J. Chem. Theory Comput.* **2016**, *12*, 5525–5540.

[25] Greff da Silveira, L.; Jacobs, M.; Prampolini, G.; Livotto, P. R.; Cacelli, I. Development and Validation of Quantum Mechanically Derived Force-Fields: Thermodynamic, Structural, and Vibrational Properties of Aromatic Heterocycles *J. Chem. Theory Comput.* **2018**, *14*, 4884–4900.

[26] Cacelli, I.; Prampolini, G. Parametrization and Validation of Intramolecular Force Fields Derived from DFT Calculations *J. Chem. Theory Comput.* **2007**, *3*, 1803–1817.

[27] Barone, V.; Cacelli, I.; De Mitri, N.; Licari, D.; Monti, S.; Prampolini, G. Joyce and Ulysses: Integrated and User-Friendly Tools for the Parameterization of Intramolecular Force Fields from Quantum Mechanical Data. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3736–51.

[28] Cerezo, J.; Prampolini, G.; Cacelli, I. Developing accurate intramolecular force fields for conjugated systems through explicit coupling terms *Theor. Chem. Accounts* **2018**, *137*, 80.

[29] Kroon-Batenburg, L.; Van Duijneveldt, F. The Use of a Moment-Optimized DZP Basis Set for Describing the Interaction in the Water Dimer *J. Mol. Struct. THEOCHEM* **1985**, *22*, 185–199.

[30] Hobza, P.; Zahradnik, R. Intermolecular Interactions between Medium-sized Systems. Nonempirical and Empirical Calculations of Interaction Energies. Successes and Failures *Chem. Rev.* **1988**, *88*, 871–897.

[31] Sponer, J.; Leszczynski, J.; Hobza, P. Base Stacking in Cytosine Dimer. A Comparison of Correlated Ab Initio Calculations with Three Empirical Potential Models and Density Functional Theory Calculations *J. Comput. Chem.* **1996**, *17*, 841–850.

[32] Barone, V.; Cacelli, I.; Crescenzi, O.; D'Ischia, M.; Ferretti, A.; Prampolini, G.; Villani, G. Unraveling the Interplay of Different Contributions to the Stability of the Quinhydrone Dimer *RSC Adv.* **2014**, *4*, 876.

[33] Prampolini, G.; Cacelli, I.; Ferretti, A. Intermolecular Interactions in Eumelanins: a Computational Bottom-up Approach. I. Small Building Blocks *RSC Adv.* **2015**, *5*, 38513–38526.

[34] Jacobs, M.; Greff da Silveira, L.; Prampolini, G.; Livotto, P. R.; Cacelli, I. Interaction Energy Landscapes of Aromatic Heterocycles through a Reliable yet Affordable Computational Approach *J. Chem. Theory Comput.* **2018**, 543–556.

[35] Antony, J.; Grimme, S. Is spin-component scaled second-order Möller-Plesset perturbation theory an appropriate method for the study on noncovalent interaction in molecules? *J. Phys. Chem. A* **2007**, *111*, 4862–4868.

[36] Distasio Jr., R. A.; Head-Gordon, M. Optimized spin-component scaled second-order Möller-Plesset theory for intermolecular energies *Mol. Phys.* **2007**, *105*, 1073.

[37] Hill, J. G.; Platts, J. A. Spin-Component Scaling Methods for Weak and Stacking Interactions *J. Chem. Theory Comput.* **2007**, *3*, 80–85.

[38] Hill, J. G.; Platts, J. A. Calculating stacking interactions in nucleic acid base-pair steps using spin-component scaling and local second order MllerPlesset perturbation theory *Phys. Chem. Chem. Phys.* **2008**, *10*, 2785–2791.

[39] Řezáč, J.; Hobza, P. Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the Gold Standard, CCSD(T) at the Complete Basis Set Limit? *J. Chem. Theory Comput.* **2013**, *9*, 2151–2155.

[40] Řezáč, J.; Hobza, P. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. *Chem. Rev.* **2016**, *116*, 5038–5071.

[41] Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. Origin of Attraction and Directionality of the π/π Interaction: Model Chemistry Calculations of Benzene Dimer Interaction *J. Am. Chem. Soc.* **2002**, *124*, 104–112.

[42] Sinnokrot, M. O.; Sherrill, C. D. Highly Accurate Coupled Cluster Potential Energy Curves for the Benzene Dimer: Sandwich, T-Shaped, and Parallel-Displaced Configurations *J. Phys. Chem. A* **2004**, *108*, 10200–10207.

[43] Hobza, P.; Zahradník, R.; Müller-Dethlefs, K. The World of Non-Covalent Interactions: 2006 *Collect. Czech. Chem. Comm.* **2006**, *71*, 443–531.

[44] Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. An Assessment of Theoretical Methods for Nonbonded Interactions: Comparison to Complete Basis Set Limit Coupled-Cluster Potential Energy Curves for the Benzene Dimer, the Methane Dimer, Benzene-Methane, and Benzene-H_2S *J. Phys. Chem. A* **2009**, *113*, 10146–10159.

[45] Vázquez-Mayagoitia, A.; Sherrill, C. D.; Sumpter, B. G.; Aprá, E. An Assessment of Density Functional Methods for Potential Energy Curves of Nonbonded Interactions: The XYG3 and B97-D Approximations *J. Chem. Theory Comput.* **2010**, *6*, 727–734.

[46] Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. Basis-set Convergence in Correlated Calculations on Ne, N_2 and H_2O *Chem. Phys. Lett.* **1998**, *286*, 243. ——-

[47] Boys, S.; Bernardi, F. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors *Mol. Phys.* **1970**, *19*, 553–566.

[48] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J.; Iyengar, S. S.; Tomasi, J.; ; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.;

Dannenberg, J. J.; Dapprich, S.; Parandekar, P. V.; Mayhall, N. J.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fo, D. J.; Gaussian09, Revision C.01; Gaussian, Inc.; Wallingford CT; 2009.

[49] Stanton, J. F.; Gauss, J.; Cheng, L.; Harding, M. E.; Matthews, D. A.; Szalay, P. G.; CFOUR, Coupled-Cluster techniques for Computational Chemistry, a quantum-chemical program package; With contributions from A.A. Auer, R.J. Bartlett, U. Benedikt, C. Berger, D.E. Bernholdt, Y.J. Bomble, O. Christiansen, F. Engel, R. Faber, M. Heckert, O. Heun, M. Hilgenberg, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, T. Kirsch, K. Klein, W.J. Lauderdale, F. Lipparini, T. Metzroth, L.A. Mück, D.P. O'Neill, D.R. Price, E. Prochnow, C. Puzzarini, K. Ruud, F. Schiffmann, W. Schwalbach, C. Simmons, S. Stopkowicz, A. Tajti, J. Vázquez, F. Wang, J.D. Watts and the integral packages MOLECULE (J. Almlöf and P.R. Taylor), PROPS (P.R. Taylor), ABACUS (T. Helgaker, H.J. Aa. Jensen, P. Jørgensen, and J. Olsen), and ECP routines by A. V. Mitin and C. van Wüllen. For the current version, see http://www.cfour.de.

[50] Harding, M. E.; Metzroth, T.; Gauss, J.; Auer, A. A. Parallel Calculation of CCSD and CCSD(T) Analytic First and Second Derivatives *J. Chem. Theory Comput.* **2008**, *4*, 64–74.

[51] Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. Derivative studies in hartree-fock and møller-plesset theories *Int. J. Quant. Chem.* **1979**, *16*, 225–241.

[52] Lochan, R. C.; Jung, Y.; Head-Gordon, M. Scaled Opposite Spin Second Order Mller-Plesset Theory with Improved Physical Description of Long-Range Dispersion Interactions *J. Phys. Chem. A* **2005**, *109*, 7598–7605.