

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Extension of the "cheap" Composite Approach to Noncovalent Interactions: The jun-ChS Scheme

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Alessandrini S., Barone V., Puzzarini C. (2020). Extension of the "cheap" Composite Approach to Noncovalent Interactions: The jun-ChS Scheme. JOURNAL OF CHEMICAL THEORY AND COMPUTATION, 16(2), 988-1006 [10.1021/acs.jctc.9b01037].

Availability:

This version is available at: <https://hdl.handle.net/11585/728327> since: 2020-02-21

Published:

DOI: <http://doi.org/10.1021/acs.jctc.9b01037>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

S. ALESSANDRINI, V. BARONE, C. PUZZARINI. EXTENSION OF THE “CHEAP” COMPOSITE APPROACH TO NON-COVALENT INTERACTIONS: THE JUN-CHS SCHEME. J. CHEM. THEORY COMP., 16 (2020) 988.

The final published version is available online at:

<https://doi.org/10.1021/acs.jctc.9b01037>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)

When citing, please refer to the published version.

Extension of the “cheap” composite approach to non-covalent interactions: the jun-ChS scheme

Silvia Alessandrini,^{*,†,‡} Vincenzo Barone,^{*,†} and Cristina Puzzarini^{*,‡}

*Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy, and Dipartimento di
Chimica “Giacomo Ciamician”, Università di Bologna, Via F. Selmi 2, 40126 Bologna, Italy*

E-mail: silvia.alessandrini@sns.it; vincenzo.barone@sns.it; cristina.puzzarini@unibo.it

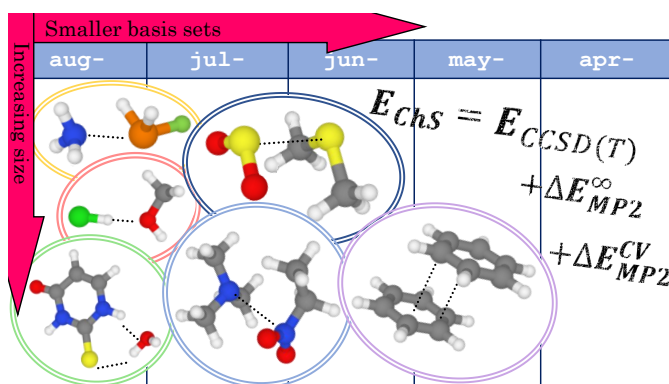
^{*}To whom correspondence should be addressed

[†]Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy

[‡]Dipartimento di Chimica “Giacomo Ciamician”, Università di Bologna, Via F. Selmi 2, 40126 Bologna,
Italy

Abstract

A new variant of the so-called “cheap composite scheme” has been purposely developed for the evaluation of the interaction energy of non-covalent molecular complexes, with its various contributions being tested for a set of 15 systems using the accurate interaction energies reported in ref. [Řezáč, J.; Dubecký, M.; Jurečka, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2015**, *17*, 19268-19277] as reference values. The modified scheme, starting from the CCSD(T) method in conjunction with a triple-zeta-quality basis set augmented by diffuse functions, includes two additional terms computed at the MP2 level: (i) the extrapolation to the complete basis set (CBS) limit; (ii) the contribution of core-valence correlation effects (CV term). Various families of basis sets including diffuse functions have been tested for the CCSD(T) model as well as for the extrapolation to the CBS limit, with a mean absolute error of about 1% (below $0.2 \text{ kJ}\cdot\text{mol}^{-1}$ in absolute terms) obtained with the jun-cc-pVnZ and the jul-cc-pVnZ families. As far as the CV term is concerned, the cc-pCVTZ and cc-pwCVTZ basis sets provide comparable contributions, which are non-negligible in several cases. While the benchmark analysis has been carried out using accurate structures available in the literature, geometrical effects due to the use of reference B2PLYP(-D3BJ) geometries, optimized in conjunction with a triple-zeta-quality basis set, have been investigated, thus pointing out their suitability. Finally, the modified scheme has been applied to a number of test cases for which interaction energies were already available in the literature; among these, a number of molecular complexes bearing second-row atoms have been considered.



1 Introduction

The quantitative evaluation of non-covalent interactions, the determination of the number and type of linkages, and the disentanglement of the different contributions ruling the structure and interaction energy of molecular complexes are of paramount importance for a number of processes of scientific as well as technological interest.^{1–5} In this connection, investigations in condensed phases are of limited use due to the concomitant effects of local and bulk contributions. Furthermore, two-body and many-body contributions should be disentangled for a better understanding of the overall description. This has stimulated in the last years huge efforts and has led to several gas-phase spectroscopic investigations of non-covalent molecular complexes (see, e.g., refs. 6–10), with rotational spectroscopy being one of the most promising techniques. While the experimental challenge of generating and detecting molecular complexes can be overcome with the last-generation microwave spectrometers (with broadband chirped-pulse spectrometers and laser ablation vaporization techniques having opened the way to the investigation of molecular adducts of increasing complexity^{6,7,11}), more sophisticated approaches and implementations are required on the computational side in order to exploit the interplay of experiment and theory at best. Indeed, the qualitative and quantitative description of non-covalent interactions is still a great challenge in computational chemistry, mainly due to the important role played by dispersion interactions as well as the size of the systems and the flatness of the potential energy surface (PES) governing intermolecular degrees of freedom. Therefore, investigating the structure and the spectroscopic properties of these systems might be non-trivial and several aspects have to be considered.

Aiming at defining a general protocol to exploit the interplay of experiment and theory in the field of the spectroscopic characterization of molecular complexes (in particular of biological or prebiotic relevance), our focus will be the interaction energy at the equilibrium structure of medium-sized molecular systems containing up to 20-30 atoms (also belonging to the second row of the periodic table). Rotational spectroscopy is able to

provide accurate information on structural parameters,^{6–10} but it requires to be strongly complemented by theory when moving to the quantitative analysis of interaction energies. In order to reach our goal, it will be necessary not only to set up a reliable and accurate model chemistry protocol, but it will be important to define a suitable level of theory for determining the equilibrium structure with sufficient accuracy, in order to avoid additional significant errors, at a reduced computational cost.

Since the accuracy obtained for intermolecular interactions should not come at the price of impairing the results for standard covalent bonds, the starting point for our new developments is the so-called “cheap” composite scheme,^{12–14} whose reference is the coupled-cluster method including full account of single and double excitations, CCSD,¹⁵ and perturbative estimate of triple excitations (CCSD(T))¹⁶ in conjunction with a triple-zeta quality basis set within the frozen-core (fc) approximation. To improve the accuracy of this level of theory, the “cheap” scheme requires the incorporation of the extrapolation to complete basis set (CBS), the effect of diffuse functions, and the core-valence correlation (CV) contribution, with all of them taken into account using Møller-Plesset theory to second order (MP2).¹⁷ While this approach was initially developed for accurate molecular structure determinations¹² and then extended to energetic evaluations,¹⁴ when focusing on the latter, geometry optimizations can be performed –without significantly worsening the accuracy– using double-hybrid functionals (e.g. B2PLYP^{18,19}), augmented by empirical dispersion contributions (e.g. the D3BJ model by Grimme^{20–23}), in conjunction with partially augmented triple-zeta basis sets (e.g. maug-cc-pVTZ-*d*H^{24–26}). Note that, although more recent functionals provide improved energetic quantities, the B2PLYP-D3BJ functional remains among the best performers for geometries and second energy derivatives.^{27,28}

While the “cheap” scheme is well tested for semi-rigid molecules^{12,13} and also rather flexible systems,^{14,29} its extension to molecular complexes^{26,30–33} cannot be taken for granted, and indeed the additive inclusion of the effect of diffuse functions has casted some doubts and led some instabilities when dealing with systems containing second-row

atoms. In this respect, it has been recently claimed that hierarchical basis sets consistently including diffuse functions (e.g. the aug-cc-pVnZ family³⁴) are mandatory for reaching the so-called “gold standard” for intermolecular complexes,^{35,36} also suggesting to avoid any additivity approximation in this respect. Since in the “cheap” approach the limiting steps are the geometry optimizations and energy evaluations at the CCSD(T) level of theory, the replacement in this scheme of the cc-pVTZ basis set with aug-cc-pVTZ would worsen significantly the situation from a computational-cost point of view. Therefore, we decided to check the performances of the above model in conjunction with basis sets containing a reduced number of diffuse functions, with the final aim of obtaining a balanced compromise between accuracy and computational cost.

As a matter of fact, although the extrapolation to the CBS limit can be much faster when using explicitly correlated (F12) coupled-cluster ansatzes,^{37,38} reduction of the basis set below the aug-cc-pVTZ quality reduces the accuracy of the model chemistry down to the level denoted as “silver standard”.³⁹ Since the MP2 extrapolation to the CBS limit is quite effective (and can be further fastened by resolution of identity and analogous techniques⁴⁰), the bottleneck remains the CCSD(T) calculation, whose (T) part does not benefit from current F12 implementations. This aspect will be discussed in some detail after the derivation of the most effective modified “cheap” scheme. Local correlation models can become the methods of choice for large systems,⁴¹ but the robustness of their implementation and their general applicability for non-covalent complexes are beyond the scope of the present study. On the other hand, we will pay special attention to the so-called basis set superposition error (BSSE), which plagues the calculations of intermolecular complexes whenever basis sets of limited size are employed. While we mostly resort to its removal by means of full inclusion of the counterpoise (CP) correction,⁴² non-CP-corrected (NCP) results will be addressed –for comparison purposes– for both energies and equilibrium geometries. The partial inclusion of CP corrections (thus leading to half-CP values) will be also considered.

On the premise that the experimental accuracy reachable for the interaction energy of

molecular complexes is $0.012 \text{ kJ}\cdot\text{mol}^{-1}$,⁴³ the last point that requires to be addressed is the definition of the target accuracy for the new model chemistry protocol. Non-covalent bondings span a huge range of interactions: from the relatively strong electrostatic, hydrogen bonds ($10\text{-}30 \text{ kJ}\cdot\text{mol}^{-1}$) to the much weaker dispersion interactions ($1\text{-}10 \text{ kJ}\cdot\text{mol}^{-1}$). Therefore, discussions based on absolute errors might be questionable: a target error of $0.2 \text{ kJ}\cdot\text{mol}^{-1}$, which –according to ref. 43– corresponds to the “gold standard”, is fully satisfactory in the first case, while it could lead to a large uncertainty for the latter. As a consequence, we prefer to avoid different thresholds for different kinds of interactions, resorting instead to percentage errors with a target of $\sim 1\%$, possibly without any outlier above 3% and, in any case, with absolute deviations within $0.2 \text{ kJ}\cdot\text{mol}^{-1}$. Indeed, a mean error of a few percent is generally reachable for molecular complexes of about ten atoms,⁴³ and in refs. 35,44 the mean error of CCSD(T)/CBS values is actually estimated to be $\sim 1\text{-}2\%$ with respect to more accurate computations, such as those used as reference in the present work.

In the following section, the computational details are provided, starting from a short description of the dataset chosen for benchmarking interaction energies, moving to the illustration of the original “cheap” scheme and of the different improvements considered in the present study. Next, case studies involving larger molecules and/or systems containing second-row elements are analyzed in detail by means of the most promising modified “cheap” scheme. Additional considerations and conclusions are sketched in the last two sections.

2 Computational details

In the present study, quantum-chemical calculations have been performed at the CCSD(T),¹⁶ MP2,¹⁷ and density functional theory (DFT) levels. For the latter, the double-hybrid B2PLYP¹⁸ functional including the D3 dispersion correction²² with Becke-Johnson (BJ) damping function²³ has been considered. All calculations have been carried out using the

Gaussian package,⁴⁵ always employing very tight thresholds in both geometry optimizations and energy evaluations.

Partial or full addition of diffuse functions has been investigated by means of the basis sets denoted as “seasonal”:⁴⁶ starting from the aug-cc-pVnZ set,³⁴ removal of diffuse functions on H and He atoms leads to the jul-cc-pVnZ basis set (also known as heavy-augmented basis set, ha); then, successive removal of high-angular momentum diffuse functions produces the jun-, may-, and apr- sets. Finally, when only *s* and *p* diffuse functions are added irrespective of the cardinal number, the minimally augmented (maug) family of basis sets is obtained, which corresponds to the jun-cc-pVDZ/may-cc-pVTZ/apr-cc-pVQZ sequence.²⁴

2.1 Interaction energy and dataset

By definition, the CP-corrected interaction energy, $E_{\text{CP}}^{\text{int}}$, is computed as follows:

$$E_{\text{CP}}^{\text{int}} = E_{\text{AB}} - (E_{\text{A}}^{\text{AB}} + E_{\text{B}}^{\text{AB}}), \quad (1)$$

where E_{AB} is the molecular complex energy, while E_{A}^{AB} and E_{B}^{AB} are the energies of the monomers, A and B, calculated using the basis set of the molecular complex (AB) at the geometry they assume in the latter. The corresponding NCP energy value is obtained with respect to the two fragments in their basis set and at the corresponding geometry in the molecular adduct (E_{A}^{A} and E_{B}^{B}). Straightforwardly, half-CP energies are derived by the arithmetic mean of CP and NCP energies.

The computational schemes for the determination of accurate interaction energies presented in this work have been tested against a subset of the A24 dataset,^{35,44} which includes 24 molecular complexes, covering the most important types of non-covalent interactions: from hydrogen bonds to dispersion-dominated interactions. With the aim to develop a general and robust scheme able to describe all of them, 15 molecular complexes from the A24 dataset have been selected for a statistical analysis. These systems

are: the dimers of H_2O , NH_3 , HF , HCN , H_2CO , CH_4 and C_2H_4 , together with mixed clusters: $\text{NH}_3 \cdots \text{H}_2\text{O}$, $\text{HF} \cdots \text{CH}_4$, $\text{NH}_3 \cdots \text{CH}_4$, $\text{H}_2\text{O} \cdots \text{CH}_4$, $\text{H}_2\text{O} \cdots \text{C}_2\text{H}_4$, $\text{NH}_3 \cdots \text{C}_2\text{H}_4$, $\text{CH}_4 \cdots \text{C}_2\text{H}_4$ and $\text{H}_2\text{CO} \cdots \text{C}_2\text{H}_4$. We have decided not to include the adducts involving Ar, borane, ethyne, and ethane because the former three systems are not particularly representative of bio-molecular building blocks, while ethane introduces the problem of intramolecular flexibility without adding significant additional information with respect to the methane complexes. Furthermore, as mentioned in the introduction, our interest is focused on molecular complexes at the equilibrium structure. This led, for instance, to the exclusion of the systems involving only π - π stacking interactions such as the “parallel” ethene dimer, which does not present any energy minimum on the corresponding PES. The selected systems define the A15 dataset and are shown in Figure 1 together with the corresponding intermolecular distances (evaluated at different levels of theory), which will be discussed later in the text.

After the most promising model has been defined together with a critical investigation of the role of the reference geometry, the study has extended to treat four small molecular complexes containing second-row atoms of interest for biological systems (P, S, and Cl), namely $\text{FH}_2\text{P} \cdots \text{NH}_3$, $\text{FH}_2\text{P} \cdots \text{SH}_2$, $\text{FHS} \cdots \text{OFH}$ and $\text{CH}_3\text{OH} \cdots \text{HCl}$, before moving to address 7 larger test cases, characterized by different types of interactions, also including the π - π stacking. The basis sets used for second-row atoms are analogous to those employed for first-row elements, but we have adopted the variant including additional tight d -functions (cc-pV(n +d)Z).⁴⁷

2.2 The reference: geometries and energies

The reference geometries used in our benchmark analysis are those reported in ref. 35. These were obtained by means of a composite scheme that starts from the HF-SCF/aug-cc-pVQZ level of theory and incorporates two corrections: (i) the extrapolation to the CBS limit at the MP2 level using the two-point formula by Helgaker et al.⁴⁸ applied to the aug-cc-pVTZ and aug-cc-pVQZ basis sets and (ii) the high-order contribution, evaluated

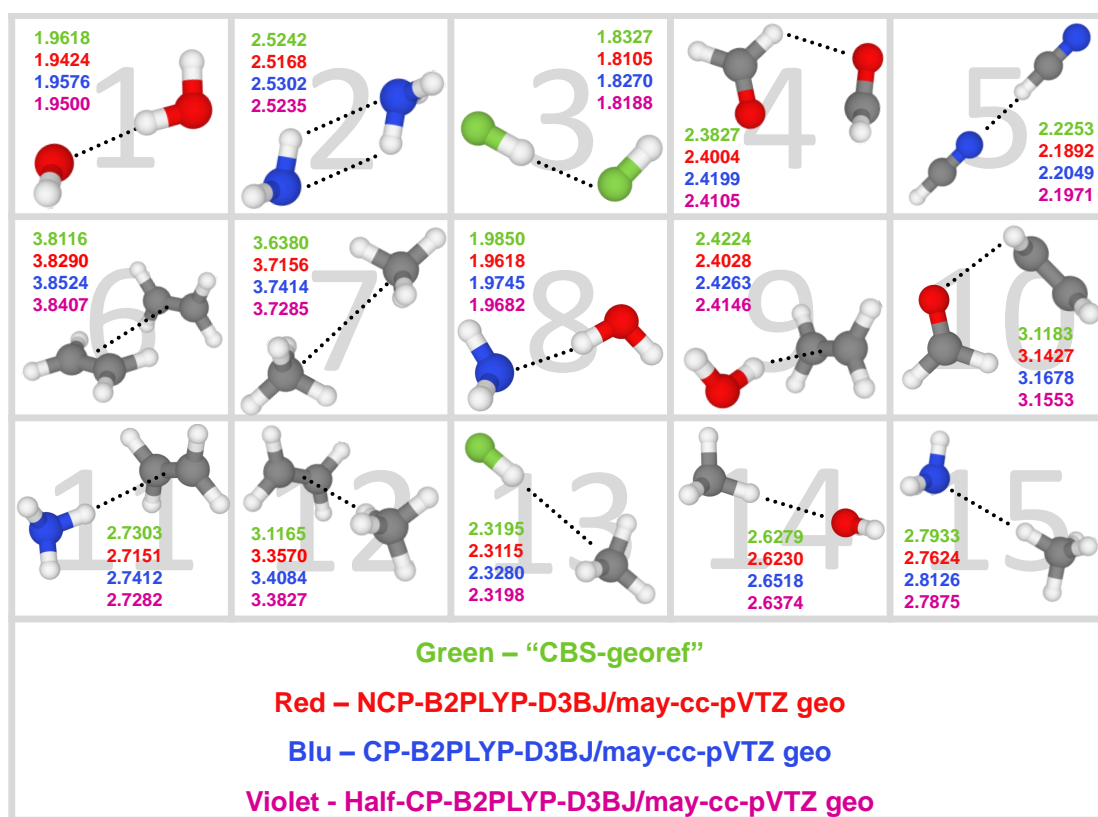


Figure 1: The A15 dataset: the molecular complexes employed in the benchmark analysis. The leading intermolecular distance (in Å), evaluated at different levels of theory, is reported for each molecular complex (see text).

as the difference between CCSD(T) and MP2 when employing the aug-cc-pVDZ basis set. The electronic energy resulting from this composite scheme is then employed in the geometry optimization performed using numerical gradients. Furthermore, to rule out the BSSE effect, the CP correction⁴² is included at each iteration step. The resulting geometry will be denoted as "CBS-georef" in the following.

The interaction energies, which have been taken as reference for this part of our analysis (and denoted as "ref" in the following), have been taken from ref. 44 and were obtained by means of a composite approach that incorporates the extrapolation to the CBS limit of both the fc-CCSD(T) and all-electron (ae) CCSD(T) terms:

1. CCSD(T) correlation energy, within the fc approximation, extrapolated to the CBS limit taken from the exhaustive analysis performed by Burns et al.⁴⁹;
2. core-valence and core-core contributions, calculated as the difference between ae-

and fc-CCSD(T) computations using the aug-cc-pCVTZ and aug-cc-pCVQZ basis sets, this contribution being extrapolated to the CBS limit with a two-point formula.⁴⁸

The reason behind the choice of these energies as reference (denoted in the text and tables as “ref”) is that reliable computations of quadruple excitations in the cluster expansion are not available for all the molecules belonging to the A15 dataset and we want to avoid any further reduction of the number and representativeness of the analyzed systems. In a later section, however, once the new computational model is set up, to further investigate the accuracy obtainable, a subset of 13 molecules will be considered using as reference (denoted as “best-ref”) the scheme that includes, in addition to the two terms introduced above, the following terms:

3. contribution due to quadruple excitations, computed as the difference between the ae-CCSDT(Q)⁵⁰ and the ae-CCSD(T) calculation using the aug-cc-pVDZ basis set;
4. inclusion of the scalar relativistic effects, calculated using CCSD(T) in conjunction with the aug-cc-pCVQZ-DK⁵¹ basis set and the fourth-order Douglas-Kroll-Hess (DKH) Hamiltonian.⁵²

2.3 The “cheap” scheme

As evident from the list above, the two approaches (“ref” and “best-ref”) include several extremely expensive computational steps, such as the fc-CCSD(T)/aug-cc-pV5Z and ae-CCSDT(Q)/aug-cc-pVDZ calculations, and they are therefore affordable only for small molecular systems. However, the interaction energies provided by both of them represent high-level references for testing the performances of less expensive computational methodologies. Indeed, a great effort has been put in the derivation of composite schemes able to provide similar results, even if relying on a lower scaling of the computational cost with the basis set. The corresponding strategies usually consider the use of MP2 in conjunction with small- to medium-sized basis sets, also exploiting the fact that, when

the MP2 method is used with such bases, a systematic error compensation occurs in the case of weakly-bounded compounds and the obtained energy is closer to the CBS limit than expected.⁴³ However, to fulfil the aimed accuracy, composite approaches have to be considered. The so-called “cheap” scheme has been introduced in this regard, with the extrapolation to the CBS limit as well as the effect of core-valence correlation and diffuse functions in the basis set being incorporated at the MP2 level. For the evaluation of energetics, the “cheap” scheme, shortly denoted as ChS in the following, can be summarized as follows:

- The starting point is the fc-CCSD(T)/cc-pVTZ level of theory.
- The extrapolation to the CBS limit is performed in two steps: the HF-SCF energies, evaluated in conjunction with the cc-pVTZ, cc-pVQZ and cc-pV5Z basis sets, are extrapolated using the 3-point exponential formula by Feller⁵³, while fc-MP2 correlation energies, computed employing the cc-pVTZ and cc-pVQZ sets, are extrapolated by means of the n^{-3} formula by Helgaker et al.⁴⁸. This term is denoted as ΔE_{MP2}^{∞} . In the present work, ΔE_{MP2}^{∞} is however computed in a single step by extrapolating the MP2 energy with the 2-point formula⁴⁸ (i.e. without extrapolating HF and MP2 energies separately). This is to avoid problems related to the cc-pV5Z basis set for which convergence of the HF-SCF energy can be troublesome, especially in the case of non-covalent complexes. Furthermore, we have enough experience to claim that both procedures lead to very similar results, the differences being negligible.
- The third contribution is the inclusion of CV effects, thus accounting for (i) the inter-shell correlation, i.e. the core-valence contribution and (ii) the intra-shell contribution due to the core-core correlation, with the latter being less important compared to the former. This term is obtained as energy difference between ae- and fc-MP2 calculations with the cc-pCVTZ basis set. This term is shortly indicated as ΔE_{MP2}^{CV} . In this work, the use of the cc-pwCVTZ set⁵⁴ will be also tested.

To this general scheme, a further term is often added to account for the effect of diffuse functions, which is here denoted as $\Delta\alpha$:

- The $\Delta\alpha$ contribution is calculated as the difference between fc-MP2/aug-cc-pVTZ and fc-MP2/cc-pVTZ energies.

Since this term is not justified by any theoretical consideration when extrapolation to the CBS limit is performed, a proper analysis of its effects is needed. In the case of structure determinations, its addition tends to lead to relevant improvements because it compensates the overestimation of the MP2 extrapolation with small- to medium-sized basis sets,^{13,14} while for energy evaluations $\Delta\alpha$ has often been removed because of erratic results.^{26,32} Furthermore, even if the ChS approach has been employed in the characterization of various non-covalent interactions,^{26,30–32} its performance for molecular complexes has never been thoroughly tested against other accurate computational strategies.

In this work, different families of basis sets have been tested by incorporating their contribution in the ChS approach using two different strategies:

- Diffuse functions are included directly in the scheme using different diffuse-augmented basis sets to compute the ΔE_{MP2}^{∞} term and the CCSD(T) energy: the extrapolation is carried out using the $n=T,Q$ combination, while the CCSD(T) calculations are always performed in conjunction with the corresponding triple-zeta set.
- The effect of diffuse functions is introduced by means of the previously mentioned $\Delta\alpha$ term, the latter being calculated with the aug-cc-pVTZ, the jul-cc-pVTZ, or the jun-cc-pVQZ basis set. In the case of the fully diffuse-augmented basis set, the “traditional” ChS approach is obtained.

To remove any potential bias affecting the overall conclusions due to the use of different equilibrium geometries, the investigation and test of the different approaches and families of basis sets have been carried out using the “CBS-georef” structures reported in ref. 35.

3 Results and discussion

3.1 Extension of the “cheap” scheme to non-covalent interactions

In the following, a new ChS-type scheme able to correctly describe non-covalent interactions is developed. First of all, the most convenient way to introduce diffuse functions in the ChS approach is analyzed for both the extrapolation to the CBS limit and CCSD(T) energy calculations. For these initial considerations, the CV term will be set at the value computed with the cc-pCVTZ basis set, while –later in the manuscript– the results obtained with the cc-pwCVTZ basis set will be also addressed. While initially the reference geometries employed are the “CBS-georef” structures, the use of B2PLYP-D3BJ structures will be subsequently addressed. As already mentioned, CP-corrected energies are mainly considered, but NCP and half-CP results will be also discussed.

3.1.1 Extrapolation to the CBS limit and CCSD(T) calculations: do we need diffuse functions?

The starting point is the composite approach with the MP2 extrapolation to the CBS limit carried out using the cc-pV(T,Q)Z basis sets on top of the CCSD(T)/cc-pVTZ calculation, with the ΔE_{MP2}^{CV} term evaluated using the cc-pCVTZ basis. The resulting mean absolute error (MAE) is 2.92% and, for each molecular complex, the contributions to the interaction energy are detailed in Table 1. It is noted that not only the MAE is well outside our target accuracy, but also the maximum errors ($\sim 9\%$) are too large. In Table 1, in addition to relative errors, the absolute errors are also given. The latter are rather small, ranging –in unsigned terms– from $0.02 \text{ kJ}\cdot\text{mol}^{-1}$ to $0.76 \text{ kJ}\cdot\text{mol}^{-1}$. However, these values are not satisfactory because more than half of the molecular complexes considered are characterized by small interaction energies and, therefore, a discussion in terms of relative errors is more sound. Furthermore, it is noteworthy that CP results are slightly better than their NCP counterparts and essentially equivalent to half-CP values.

The first step is the introduction of additional diffuse functions in the extrapolation

Table 1: ChS energies ($\text{kJ}\cdot\text{mol}^{-1}$) using the cc-pVnZ basis sets.

Complex	"ref"	fc-CCSD(T)/cc-pVTZ	$\Delta E_{MP2}^{\infty}/\text{cc-pV(T,Q)Z}$	$\Delta E_{MP2}^{\text{CV}}$	Energy	Relative Error	Error	Error _{NCP}	Error _{half-CP}
H ₂ O-H ₂ O	-21.0832	-18.3868	-1.9570	-0.1354	-20.4792	-2.86%	-0.60	-0.29	-0.44
NH ₃ -NH ₃	-13.2131	-11.3498	-1.6052	-0.0733	-13.0283	-1.40%	-0.18	-0.40	-0.29
HF-HF	-19.2213	-17.2997	-1.2275	-0.1049	-18.6321	-3.07%	-0.59	-0.65	-0.62
H ₂ CO-H ₂ CO	-18.9284	-13.5169	-5.2213	-0.0987	-18.8369	-0.48%	-0.09	-0.55	-0.32
HCN-HCN	-19.9828	-18.2582	-1.3439	-0.0733	-19.6754	-1.54%	-0.31	-0.41	-0.36
C ₂ H ₄ -C ₂ H ₄	-4.5647	-3.0318	-1.5070	-0.0439	-4.5827	0.39%	0.02	0.07	0.05
CH ₄ -CH ₄	-2.2301	-1.2588	-0.8478	-0.0068	-2.1134	-5.23%	-0.12	-0.08	-0.10
H ₂ O-NH ₃	-27.3759	-23.5124	-2.9325	-0.1677	-26.6126	-2.79%	-0.76	-0.40	-0.58
H ₂ O-C ₂ H ₄	-10.7696	-8.0903	-2.4127	-0.0875	-10.5905	-1.66%	-0.18	0.37	0.09
C ₂ H ₄ -H ₂ CO	-6.7948	-4.5478	-2.1112	-0.0600	-6.7190	-1.12%	-0.08	-0.36	-0.22
C ₂ H ₄ -NH ₃	-5.7865	-4.1166	-1.5012	-0.0488	-5.6666	-2.07%	-0.12	0.23	0.06
C ₂ H ₄ -CH ₄	-2.1255	-1.4393	-0.5964	-0.0258	-2.0615	-3.01%	-0.06	-0.10	-0.08
HF-CH ₄	-6.9162	-4.7205	-2.0688	-0.0860	-6.8753	-0.59%	-0.04	0.07	0.02
H ₂ O-CH ₄	-2.8242	-2.1376	-0.4039	-0.0286	-2.5701	-9.00%	-0.25	-0.39	-0.32
NH ₃ -CH ₄	-3.2175	-2.7172	-0.1872	-0.0352	-2.9396	-8.64%	-0.28	-0.19	-0.23
MAE						2.92%	0.25	0.30	0.25

term and, for consistency, also in CCSD(T) calculations. Resorting to the Dunning and coworkers hierarchical basis sets, namely the aug-cc-pVnZ family,³⁴ leads to a large improvement, the results being collected in Table 2. Indeed, the MAE, being 1.52% for the CP-corrected energy, nearly halves with respect to the starting approach. Particularly remarkable is the reduction of the errors associated to the molecular complexes involving methane, but a general improvement of the relative discrepancies is observed for most complexes. This is mainly due to a better balance between the extrapolated CBS energies and the CCSD(T) term. Also in this case, CP results are significantly or slightly better than their NCP or half-CP counterparts. However, while the composite scheme based on the aug-cc-pVnZ family of basis sets can represent an accurate approach to describe any type of non-covalent interactions, the full inclusion of diffuse functions strongly affects its applicability to medium- and large-sized systems. Furthermore, for systems like the HF dimer and the H₂CO dimer the absolute deviation is not satisfactory yet, being larger than $0.2 \text{ kJ}\cdot\text{mol}^{-1}$.

It has to be noted that the scheme results in energies that are overestimated for almost all complexes, thus suggesting that full inclusion of diffuse functions is not the best choice for an effective extrapolation. This overestimation is probably due to the well-known overshooting of intermolecular correlation by the MP2 model. Since, on the contrary, the interaction energies are systematically underestimated (with the exception

Table 2: ChS energies ($\text{kJ}\cdot\text{mol}^{-1}$) using the aug-cc-pVnZ basis sets.

Complex	"ref"	fc-CCSD(T)/aug-cc-pVTZ	ΔE_{MP2}^{∞} /aug-cc-pV(T,Q)Z	ΔE_{MP2}^{CV}	Energy	Relative Error	Error	Error _{NCP}	Error _{half-CP}
H ₂ O-H ₂ O	-21.0832	-19.9250	-1.1894	-0.1354	-21.2498	0.79%	0.17	0.44	0.30
NH ₃ -NH ₃	-13.2131	-12.5692	-0.6682	-0.0733	-13.3107	0.74%	0.10	0.12	0.11
HF-HF	-19.2213	-17.9928	-1.3570	-0.1049	-19.4547	1.21%	0.23	0.75	0.49
H ₂ CO-H ₂ CO	-18.9284	-17.6285	-1.5474	-0.0987	-19.2746	1.83%	0.35	0.43	0.39
HCN-HCN	-19.9828	-19.3043	-0.5778	-0.0733	-19.9554	-0.14%	-0.03	-0.11	-0.07
C ₂ H ₄ -C ₂ H ₄	-4.5647	-4.3068	-0.3787	-0.0439	-4.7294	3.61%	0.16	0.06	0.11
CH ₄ -CH ₄	-2.2301	-2.0849	-0.1707	-0.0068	-2.2624	1.45%	0.03	-0.02	0.004
H ₂ O-NH ₃	-27.3759	-26.1081	-1.2329	-0.1677	-27.5087	0.49%	0.13	0.36	0.25
H ₂ O-C ₂ H ₄	-10.7696	-10.1895	-0.6334	-0.0875	-10.9104	1.31%	0.14	0.07	0.11
C ₂ H ₄ -H ₂ CO	-6.7948	-6.3544	-0.5124	-0.0600	-6.9268	1.94%	0.13	0.13	0.13
C ₂ H ₄ -NH ₃	-5.7865	-5.4694	-0.3740	-0.0488	-5.8922	1.83%	0.11	-0.06	0.02
C ₂ H ₄ -CH ₄	-2.1255	-1.9806	-0.1742	-0.0258	-2.1805	2.59%	0.06	-0.04	0.01
HF-CH ₄	-6.9162	-6.4952	-0.4980	-0.0860	-7.0792	2.36%	0.16	0.56	0.36
H ₂ O-CH ₄	-2.8242	-2.6011	-0.2057	-0.0286	-2.8354	0.40%	0.01	0.01	0.01
NH ₃ -CH ₄	-3.2175	-3.0165	-0.2344	-0.0352	-3.2861	2.13%	0.07	0.06	0.06
MAE						1.52%	0.13	0.21	0.16

of the ethene dimer) when diffuse functions are completely neglected (see Table 1), the systematic removal of some sub-shells of diffuse functions present in the aug-cc-pVnZ basis sets might provide improved results at a reduced computational cost. For this reason, we have investigated systematically the "seasonal" set of bases developed by Papajak et al.⁴⁶. In particular, the jul-cc-pVnZ basis sets are commonly used for the description of weakly-bound compounds (see, e.g., ref. 43). The results obtained by employing the jul-cc-pV(T,Q)Z combination in the extrapolation to the CBS limit and the jul-cc-pVTZ set for the CCSD(T) energy evaluations are summarized in Table 3. As expected, this type of basis sets describes uniformly all the complexes considered, the relative errors being on average lower than those obtained with the fully augmented basis set (the MAE is as low as 0.84%). Focusing on methane-containing complexes, which are among the most challenging systems to correctly describe, an improvement is observed for three of them, while the relative error worsens for the H₂O-CH₄ adduct and it remains nearly unchanged for the methane dimer. Inclusion of CP correction delivers once again better results than the NCP or half-CP alternatives. As a matter of fact, jul-cc-pVnZ basis sets (better known as heavy-augmented) have become the *de facto* standard for the most refined composite methods.^{55,56} The latest HEAT variant⁵⁷ employs the opposite strategy of retaining diffuse functions on hydrogens, but not on other atoms. Partial inclusion of diffuse functions thus appears to be preferred with respect to full augmentation.

Table 3: ChS energies ($\text{kJ}\cdot\text{mol}^{-1}$) using the jul-cc-pVnZ basis sets.

Complex	"ref"	fc-CCSD(T)/jul-cc-pVTZ	ΔE_{MP2}^{∞} /jul-cc-pV(T,Q)Z	$\Delta E_{MP2}^{\text{CV}}$	Energy	Relative Error	Error	Error _{NCP}	Error _{half-CP}
H ₂ O-H ₂ O	-21.0832	-19.6966	-1.2103	-0.1354	-21.0423	-0.19%	-0.04	0.09	0.02
NH ₃ -NH ₃	-13.2131	-12.2108	-0.9287	-0.0733	-13.2128	0.00%	-0.001	0.09	0.04
HF-HF	-19.2213	-17.8393	-1.2781	-0.1049	-19.2223	0.01%	0.001	0.17	0.09
H ₂ CO-H ₂ CO	-18.9284	-17.4291	-1.6128	-0.0987	-19.1406	1.12%	0.21	0.33	0.27
HCN-HCN	-19.9828	-19.1640	-0.6978	-0.0733	-19.9351	-0.24%	-0.05	-0.15	-0.10
C ₂ H ₄ -C ₂ H ₄	-4.5647	-4.0271	-0.5995	-0.0439	-4.6705	2.32%	0.11	0.10	0.11
CH ₄ -CH ₄	-2.2301	-1.8353	-0.3506	-0.0068	-2.1927	-1.67%	-0.04	-0.18	-0.11
H ₂ O-NH ₃	-27.3759	-25.6680	-1.5765	-0.1677	-27.4122	0.13%	0.04	0.10	0.07
H ₂ O-C ₂ H ₄	-10.7696	-9.8531	-0.8998	-0.0875	-10.8404	0.66%	0.07	0.07	0.07
C ₂ H ₄ -H ₂ CO	-6.7948	-6.1293	-0.6630	-0.0600	-6.8523	0.85%	0.06	0.15	0.11
C ₂ H ₄ -NH ₃	-5.7865	-5.1967	-0.5967	-0.0488	-5.8422	0.96%	0.06	0.08	0.07
C ₂ H ₄ -CH ₄	-2.1255	-1.8073	-0.3215	-0.0258	-2.1545	1.36%	0.03	0.01	0.02
HF-CH ₄	-6.9162	-6.1143	-0.7931	-0.0860	-6.9934	1.12%	0.08	0.08	0.08
H ₂ O-CH ₄	-2.8242	-2.4392	-0.3136	-0.0286	-2.7814	-1.52%	-0.04	-0.03	-0.04
NH ₃ -CH ₄	-3.2175	-2.7902	-0.4066	-0.0352	-3.2320	0.45%	0.01	0.02	0.02
MAE						0.84%	0.06	0.11	0.08

In view of the promising results obtained using the jul-cc-pVnZ sets and with the aim of further reducing the computational cost, the jun-cc-pVnZ family of basis sets has been considered. The modified scheme thus involves the extrapolation to the CBS limit using the jun-cc-pVTZ/jun-cc-pVQZ combination of bases and CCSD(T) calculations in conjunction with the jun-cc-pVTZ basis set.

The results are reported in Table 4, where it is noted that the MAE of 1.20% places the scheme in between the previous one with the jul-cc-pVnZ basis sets (MAE 0.84%) and the one obtained with the fully augmented aug-cc-pVnZ basis sets (MAE 1.52%). Furthermore, the so-called "jun-ChS" approach fulfils the target accuracy we set in the Introduction. However, we note one outlier with a relative deviation larger than 3%, namely the ethene dimer (4.04%). This suggests the importance of all diffuse functions on non-hydrogen atoms for pushing the accuracy to the limit when the interaction involves hydrogen atoms on one side and a π bond on the other. Interestingly, in absolute terms the maximum error is below $0.2 \text{ kJ}\cdot\text{mol}^{-1}$, with the only exception of the formaldehyde dimer ($0.37 \text{ kJ}\cdot\text{mol}^{-1}$), which –however– has some intrinsic problems (see, e.g., ref. 44). In view of the good accuracy accompanied by a remarkable reduction of the computational cost (due to the decreased number of diffuse functions in the basis sets), the jun-ChS scheme seems to be the most promising approach for the description of larger non-covalent molecular adducts. Furthermore, for the methane-containing complexes,

small relative errors are noted. Although the core-valence correlation contributions are generally small, they cannot be neglected when aiming at a quantitative accuracy because their value is larger than the overall absolute error for 8 systems (over a total of 15) within the jul-ChS approach and for 6 molecular complexes when applying the jun-ChS scheme. Once again, CP results are better than NCP or half-CP counterparts.

Table 4: ChS energies ($\text{kJ}\cdot\text{mol}^{-1}$) using the jun-cc-pVnZ basis sets.

Complex	"ref"	fc-CCSD(T)/jun-cc-pVTZ	ΔE_{MP2}^{∞} /jun-cc-pV(T,Q)Z	ΔE_{MP2}^{CV}	Energy	Relative Error	Error	Error _{NCP}	Error _{half-CP}
H ₂ O-H ₂ O	-21.0832	-19.2986	-1.5414	-0.1354	-20.9754	-0.51%	-0.11	0.17	0.03
NH ₃ -NH ₃	-13.2131	-11.8413	-1.3507	-0.0733	-13.2653	0.40%	0.05	0.09	0.07
HF-HF	-19.2213	-17.3511	-1.7673	-0.1049	-19.2233	0.01%	0.002	0.23	0.12
H ₂ CO-H ₂ CO	-18.9284	-16.1715	-3.0240	-0.0987	-19.2942	1.93%	0.37	0.56	0.46
HCN-HCN	-19.9828	-18.8230	-1.0026	-0.0733	-19.8989	-0.42%	-0.08	-0.23	-0.16
C ₂ H ₄ -C ₂ H ₄	-4.5647	-3.7570	-0.9481	-0.0439	-4.7489	4.04%	0.18	0.24	0.21
CH ₄ -CH ₄	-2.2301	-1.6317	-0.6187	-0.0068	-2.2572	1.22%	0.03	-0.003	0.01
H ₂ O-NH ₃	-27.3759	-25.2160	-1.9219	-0.1677	-27.3056	-0.26%	-0.07	0.07	0.002
H ₂ O-C ₂ H ₄	-10.7696	-9.5029	-1.2704	-0.0875	-10.8608	0.85%	0.09	0.25	0.17
C ₂ H ₄ -H ₂ CO	-6.7948	-5.6830	-1.2112	-0.0600	-6.9543	2.35%	0.16	0.31	0.23
C ₂ H ₄ -NH ₃	-5.7865	-4.9695	-0.8644	-0.0488	-5.8828	1.66%	0.10	0.20	0.15
C ₂ H ₄ -CH ₄	-2.1255	-1.7123	-0.4288	-0.0258	-2.1668	1.94%	0.04	0.05	0.05
HF-CH ₄	-6.9162	-5.8334	-1.0948	-0.0860	-7.0142	1.42%	0.10	0.12	0.11
H ₂ O-CH ₄	-2.8242	-2.3273	-0.4490	-0.0286	-2.8049	-0.68%	-0.02	-0.005	-0.01
NH ₃ -CH ₄	-3.2175	-2.6783	-0.5145	-0.0352	-3.2280	0.33%	0.01	0.01	0.01
MAE						1.20%	0.09	0.17	0.12

In Figure 2, the interaction energies obtained with the different basis-set families considered are graphically represented for the 5 complexes involving methane in terms of the relative differences with respect to the corresponding "ref" values (a similar figure based on absolute errors is reported in the Supporting Information, SI). It is observed that the schemes based on the use of the jul- and jun-cc-pVnZ basis sets provide similar results and very close (indeed, the closest) to the reference values. Furthermore, as already mentioned above, the lack of diffuse functions in the basis sets leads to underestimated results, while those obtained by the approach exploiting the fully diffuse-augmented basis sets are always overestimated.

Going a step further in removing diffuse functions from basis set, one can resort to the minimal augmented set of bases (maug-cc-pVnZ²⁴). The composite scheme employing the latter type of basis has also been tested; however, the results worsen for all complexes, with a maximum relative error of $\sim 21\%$ in the case of the CH₄...CH₄ complex and a MAE of 7.1%. The detailed energetic contributions for this scheme can be found in

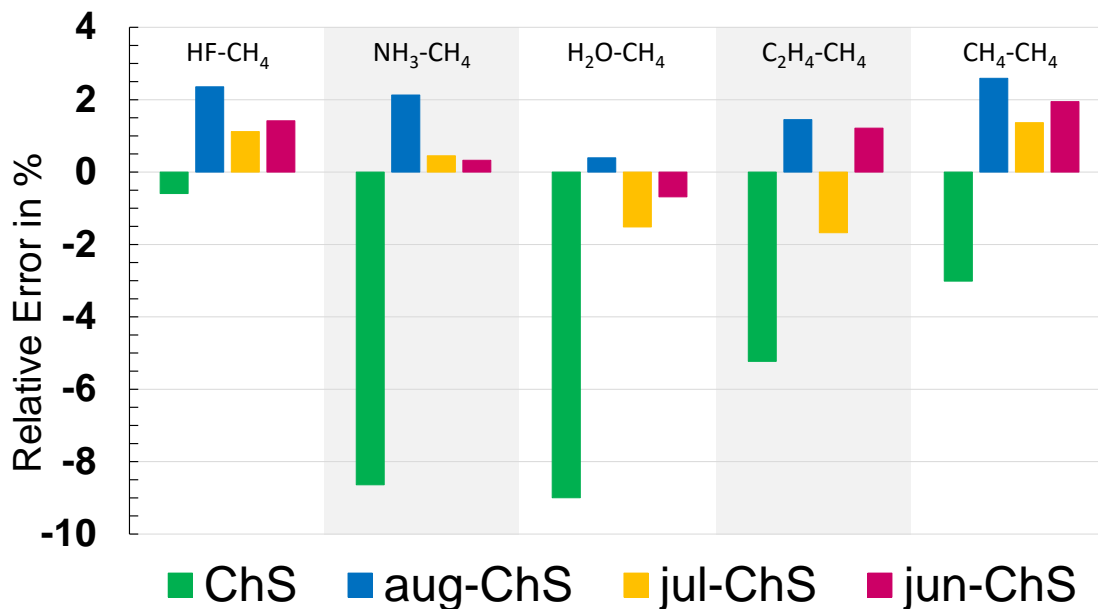


Figure 2: Interaction energy differences (%) with respect to the “ref” values for 5 methane-containing complexes from the A24 dataset (see text). Comparison of different ChS approaches: green bars refer to the scheme using the cc-pVnZ basis sets (ChS), blue bars to that employing the aug-cc-pVnZ basis sets (aug-ChS), yellow bars to that exploiting the jul-cc-pVnZ basis sets (jul-ChS), and pink bars to that using the jun-cc-pVnZ basis sets (jun-ChS).

Table S1 of the SI. This marked enhancement of relative errors could be due to the fact that the maug-cc-pVnZ basis sets are not smoothly convergent in the diffuse functions space (i.e. they cannot be considered a hierarchy of bases), as pointed out in ref. 46. Indeed, the maug-cc-pVTZ basis is equivalent to may-cc-pVTZ, while maug-cc-pVQZ corresponds to the apr-cc-pVQZ basis set, but the diffuse space is instead convergent within the “same month of the calendar” basis sets. In view of these considerations, the combination may-cc-pVTZ/may-cc-pVQZ has also been tested, with the results being reported in Table S2. It is observed that the MAE, 2.4%, lowers with respect to the use of the maug-cc-pVnZ series of bases, with a general improvement of all relative errors being noted. However, the MAE obtained is larger than that for the schemes that make use of the jun- or jul-cc-pVnZ combinations. In analogy to the extrapolation with the maug-cc-pVnZ basis sets, the extrapolation to the CBS limit using the jun-cc-pVTZ/may-cc-pVQZ combination has also been tested. The results are collected in the SI (Table S3)

and represent an improvement with respect to those obtained for the maug-cc-pV(T,Q)Z extrapolation scheme (the MAE is 2.9%). Despite this improvement, the MAE is larger than our target accuracy and, for three molecular complexes, large relative errors (in the 6-9% range) are noted; therefore, the scheme with the last combination of basis sets has not been further considered.

Since it has been demonstrated the fundamental role played by diffuse functions in the accurate description of non-covalent interactions and in order to avoid their use in CCSD(T) calculations (thus lowering the computational cost), one might also consider the possibility to include the effect of diffuse functions in a separate contribution (i.e. the $\Delta\alpha$ term introduced above), thus performing both the MP2 extrapolation to the CBS limit and CCSD(T) energy evaluations using the cc-pV n Z basis sets. This would thus imply that the diffuse functions contribution follows the additivity approximation, which is at the basis of composite schemes. For this purpose, the $\Delta\alpha$ term has been computed using the aug-cc-pVTZ and jul-cc-pVTZ basis sets, the results of the former scheme being shown in Table 5, while for the latter they are reported in Table S4.

Table 5: ChS energies ($\text{kJ}\cdot\text{mol}^{-1}$) using the aug-cc-pVTZ basis set for the $\Delta\alpha$ term.

Complex	fc-CCSD(T)/cc-pVTZ	$\Delta E_{MP2}^{\infty}/\text{cc-pV(T,Q)Z}$	ΔE_{MP2}^{CV}	$\Delta\alpha$	Energy	"ref"	Relative Error	Error
H ₂ O-H ₂ O	-18.3968	-1.9570	-0.1354	-1.0176	-21.4968	-21.0832	1.96%	0.41
NH ₃ -NH ₃	-11.3498	-1.6052	-0.0733	-0.9211	-13.9494	-13.2131	5.57%	0.74
HF-HF	-17.2997	-1.2275	-0.1049	-0.0722	-18.7043	-19.2213	-2.69%	-0.52
H ₂ CO-H ₂ CO	-13.5169	-5.2213	-0.0987	-3.7534	-22.5903	-18.9284	19.36%	3.66
HCN-HCN	-18.2582	-1.3439	-0.0733	-0.8891	-20.5645	-19.9828	2.91%	0.58
C ₂ H ₄ -C ₂ H ₄	-3.0318	-1.5070	-0.0439	-1.1774	-5.7600	-4.5647	26.19%	1.20
CH ₄ -CH ₄	-1.2588	-0.8478	-0.0068	-0.6666	-2.7801	-2.2301	24.67%	0.55
H ₂ O-NH ₃	-23.5124	-2.9325	-0.1677	-2.1569	-28.7695	-27.3759	5.09%	1.39
H ₂ O-C ₂ H ₄	-8.0903	-2.4127	-0.0875	-1.8544	-12.4449	-10.7696	15.56%	1.68
C ₂ H ₄ -H ₂ CO	-4.5478	-2.1112	-0.0600	-1.6154	-8.3344	-6.7948	22.66%	1.54
C ₂ H ₄ -NH ₃	-4.1166	-1.5012	-0.0488	-1.1854	-6.8521	-5.7865	18.42%	1.07
C ₂ H ₄ -CH ₄	-1.4393	-0.5964	-0.0258	-0.4504	-2.5119	-2.1255	18.18%	0.39
HF-CH ₄	-4.7205	-2.0688	-0.0860	-1.5040	-8.3793	-6.9162	21.16%	1.46
H ₂ O-CH ₄	-2.1376	-0.4039	-0.0286	-0.2821	-2.8522	-2.8242	0.99%	0.03
NH ₃ -CH ₄	-2.7172	-0.1872	-0.0352	-0.0938	-3.0334	-3.2175	-5.72%	-0.18
MAE							12.74%	0.90

The relative errors obtained in both cases are considerably higher than those without the $\Delta\alpha$ term (see Table 1). Focusing on Table 5, the inclusion of the diffuse functions term decreases –on average– the energy by about $1.2 \text{ kJ}\cdot\text{mol}^{-1}$. Particularly sensitive to this

term is the formaldehyde dimer, for which the relative error increase up to 19%, which –in absolute terms– corresponds to a deviation of $3.7 \text{ kJ}\cdot\text{mol}^{-1}$. Similar large relative discrepancies (18-26%) are also observed for most of the non-covalent systems for which dispersion interactions play an important role, with the only exception of the $\text{H}_2\text{O}\cdots\text{CH}_4$ and $\text{NH}_3\cdots\text{CH}_4$ adducts. Rigorous conclusions on the effect of diffuse functions and on the way to correctly account for them can be obtained from the comparison between Tables 2 and 5 since in both tables the aug-cc-pV n Z basis sets are used: if large discrepancies are not observed, then the inclusion of the separated $\Delta\alpha$ term can be considered a powerful alternative to incorporate the effect of diffuse functions at a reduced computational cost. However, this is not the case and, indeed, large differences are noticed. First of all, the MAE increases significantly: from 1.52% to 12.74%, with the latter referring to the composite scheme including the $\Delta\alpha$ term. To conclude, the “ $\Delta\alpha$ approach” is thus not recommended for any application to energies. Very similar results and conclusions are obtained with the evaluation of the $\Delta\alpha$ term using the jul-cc-pVTZ basis set: the MAE is as large as 10.6% and for 9 molecular complexes large relative errors (in the range 12-21%) are observed. Furthermore, employing a quadruple-zeta basis set (namely, the jun-cc-pVQZ set) leads to a significant but limited improvement. In fact, the MAE lowers to 3.94% (see Table S5), which is however larger than that obtained without the $\Delta\alpha$ term (see Table 1).

Once established that the effect of diffuse functions needs to be incorporated in the evaluation of the CBS extrapolation and coupled-cluster terms, a possible strategy to further reduce the computational cost is to develop an effective decomposition of the latter in order to employ the cc-pVTZ basis set (instead of the jun-cc-pVTZ set) in CCSD(T) calculations. This leads to a two-step procedure: (i) CCSD computations using the jun-cc-pVTZ basis set to incorporate the effect of diffuse functions and (ii) inclusion of the triple excitations contribution as difference between CCSD(T) and CCSD energy evaluations in conjunction with the cc-pVTZ basis set. The detailed contributions of this approach are reported in Table 6. The results are very promising: the MAE is as small as 1.08%,

thus lying in between the jul-ChS and jun-ChS schemes. The only outlier is the $\text{H}_2\text{O}-\text{CH}_4$ complex, for which the relative error is -3.3%. As last note, we point out that this type of scheme systematically leads to an underestimation of the energy. If we consider NCP values, an overestimation is noted instead in most cases, with the MAE worsening to 1.89%. Based on these considerations, the half-CP values are expected to improve the agreement with the reference energies. Indeed, moving from CP to half-CP the MAE for the absolute error improve from $0.10 \text{ kJ}\cdot\text{mol}^{-1}$ to $0.08 \text{ kJ}\cdot\text{mol}^{-1}$, which means – in relative terms – an improvement of the MAE from 1.08% to 0.86%. Therefore, a scaling of the triple contribution (akin the one usually employed in F12 approaches) could lead to improved results.

Table 6: ChS energies ($\text{kJ}\cdot\text{mol}^{-1}$) with separated evaluation of triple excitations contribution, $\Delta(\text{T})^a$ (see text).

Complex	"ref"	fc-CCSD/jun-cc-pVTZ	$\Delta E_{MP2}^{\infty}/\text{jun-cc-pV(T,Q)Z}$	$\Delta(\text{T})/\text{cc-pVTZ}$	$\Delta E_{MP2}^{\text{CV}}$	Energy	Relative Error	Error	Error _{NCP}	Error _{half-CP}
$\text{H}_2\text{O}-\text{H}_2\text{O}$	-21.0832	-18.4548	-1.5414	-0.6543	-0.1354	-20.7859	-1.42%	-0.30	0.31	0.007
NH_3-NH_3	-13.2131	-10.9234	-1.3507	-0.7717	-0.0733	-13.1191	-0.71%	-0.09	0.18	0.04
$\text{HF}-\text{HF}$	-19.2213	-16.8437	-1.7673	-0.3872	-0.1049	-19.1031	-0.61%	-0.12	0.41	0.15
$\text{H}_2\text{CO}-\text{H}_2\text{CO}$	-18.9284	-14.5750	-3.0240	-1.2011	-0.0987	-18.8989	-0.16%	-0.03	0.44	0.21
$\text{HCN}-\text{HCN}$	-19.9828	-18.3744	-1.0026	-0.3348	-0.0733	-19.7851	-0.99%	-0.20	-0.29	-0.25
$\text{C}_2\text{H}_4-\text{C}_2\text{H}_4$	-4.5647	-2.5842	-0.9481	-1.0209	-0.0439	-4.5971	0.71%	0.03	0.13	0.08
CH_4-CH_4	-2.2301	-1.1622	-0.6187	-0.4034	-0.0068	-2.1911	-1.75%	-0.04	-0.04	-0.04
$\text{H}_2\text{O}-\text{NH}_3$	-27.3759	-23.9719	-1.9219	-0.9919	-0.1677	-27.0534	-1.18%	-0.32	0.08	-0.12
$\text{H}_2\text{O}-\text{C}_2\text{H}_4$	-10.7696	-8.4607	-1.2704	-0.8411	-0.0875	-10.6597	-1.02%	-0.11	0.21	0.05
$\text{C}_2\text{H}_4-\text{H}_2\text{CO}$	-6.7948	-4.7535	-1.2112	-0.7336	-0.0600	-6.7583	-0.54%	-0.04	0.25	0.11
$\text{C}_2\text{H}_4-\text{NH}_3$	-5.7865	-4.1678	-0.8644	-0.6632	-0.0488	-5.7443	-0.73%	-0.04	0.16	0.06
$\text{C}_2\text{H}_4-\text{CH}_4$	-2.1255	-1.2402	-0.4288	-0.4094	-0.0258	-2.1042	-1.00%	-0.02	0.03	0.004
$\text{HF}-\text{CH}_4$	-6.9162	-5.0638	-1.0948	-0.6432	-0.0860	-6.8878	-0.41%	-0.03	0.15	0.06
$\text{H}_2\text{O}-\text{CH}_4$	-2.8242	-1.9594	-0.4490	-0.2942	-0.0286	-2.7312	-3.29%	-0.09	0.04	-0.03
NH_3-CH_4	-3.2175	-2.2244	-0.5145	-0.3913	-0.0352	-3.1653	-1.62%	-0.05	0.04	-0.005
MAE							1.08%	0.10	0.18	0.08

^a $\Delta(\text{T})$ term evaluated as the CCSD(T)-CCSD energy difference employing the cc-pVTZ basis set.

3.1.2 The core-valence correlation contribution

After the detailed analysis of the effect of diffuse functions, the investigation of the contribution due to the correlation of core electrons is deserved. Although CV contributions are often neglected in the models that focus on the characterization of intermolecular interactions, sub-valence correlation effects are by no means negligible for covalent bonds and/or for atoms belonging to the second and successive rows of the periodic table (*vide infra* and ref. 55). Since the aim of the new ChS models is to be applicable to all kinds of

interactions along the whole periodic table, we think that this contribution (which never represents the computational bottleneck) should be consistently incorporated in the model. Due to a fortuitous error compensation, the results obtained using the aug-cc-pV n Z basis sets for systems containing only first-row atoms (see Table 2) are indeed improved when neglecting the CV term (see Table S8), the MAE lowering from 1.52% to 0.89%. However, the results delivered by the jul-cc-pV n Z and jun-cc-pV n Z basis sets (see Tables S9 and S10, respectively) show comparable errors with and without CV contributions: 0.71% vs 0.84% (see Table 3) and 0.97% vs 1.20% (see Table 4), respectively. It is interesting to note that if diffuse functions are entirely neglected (see Table 1), the MAE worsens when the CV term is excluded from 2.92% to 3.65% (see Table S11). Since the additional cost of full augmented basis sets (which enter the most demanding CCSD(T) computations) does not justify the ‘ad hoc’ procedure of removing CV contributions, for the sake of consistency, our models are based on the jul-cc-pV n Z and jun-cc-pV n Z alternatives and include CV contributions.

While the cc-pCVTZ basis set is typically used, the cc-pwCVTZ set is expected to provide a better description of the intershell correlation, i.e. the most relevant contribution to the CV term. The results for the jun-ChS approach when using cc-pwCVTZ for evaluating ΔE_{MP2}^{CV} are reported in Table 7 and should be compared with those of Table 4. It is noted that the two basis sets provide similar terms, with an overall small difference, about 0.05%, between the respective MAEs. Since the use of the cc-pwCVTZ basis set does not increase the computational cost with respect to cc-pCVTZ, the former set should be preferred because the cc-pwCV n Z family of basis sets has been demonstrated to improve structural and molecular properties with respect to the non-weighted counterparts^{54,58} as well as to better describe core-valence effects when second-row atoms are involved.⁵⁴

For completeness, the CV term computed with the cc-pwCVTZ basis set has also been considered in the composite schemes using the jul-cc-pV n Z basis sets and in the one considering the separated $\Delta(T)$ term computed with the cc-pVTZ basis set. Comparing

Table 7: jun-ChS energies ($\text{kJ}\cdot\text{mol}^{-1}$) using the cc-pwCVTZ basis set for the CV term.

Complex	fc-CCSD(T)/jun-cc-pVTZ	ΔE_{MP2}^{∞} /jun-cc-pV(T,Q)Z	ΔE_{MP2}^{CV}	Energy	“ref”	Relative Error	Error
H ₂ O-H ₂ O	-19.2986	-1.5414	-0.1444	-20.9844	-21.0832	-0.47%	-0.10
NH ₃ -NH ₃	-11.8413	-1.3507	-0.0801	-13.2721	-13.2131	0.45%	0.06
HF-HF	-17.3511	-1.7673	-0.1106	-19.2290	-19.2213	0.04%	0.01
H ₂ CO-H ₂ CO	-16.1715	-3.0240	-0.0895	-19.2850	-18.9284	1.88%	0.36
HCN-HCN	-18.8230	-1.0026	-0.0652	-19.8908	-19.9828	-0.46%	-0.09
C ₂ H ₄ -C ₂ H ₄	-3.7570	-0.9481	-0.0495	-4.7546	-4.5647	4.16%	0.19
CH ₄ -CH ₄	-1.6317	-0.6187	-0.0069	-2.2573	-2.2301	1.22%	0.03
H ₂ O-NH ₃	-25.2160	-1.9219	-0.1814	-27.3193	-27.3759	-0.21%	-0.06
H ₂ O-C ₂ H ₄	-9.5029	-1.2704	-0.0963	-10.8696	-10.7696	0.93%	0.10
C ₂ H ₄ -H ₂ CO	-5.6830	-1.2112	-0.0656	-6.9598	-6.7948	2.42%	0.17
C ₂ H ₄ -NH ₃	-4.9695	-0.8644	-0.0545	-5.8884	-5.7865	1.76%	0.10
C ₂ H ₄ -CH ₄	-1.7123	-0.4288	-0.0297	-2.1708	-2.1253	2.13%	0.05
HF-CH ₄	-5.8334	-1.0948	-0.0976	-7.0258	-6.9162	1.59%	0.11
H ₂ O-CH ₄	-2.3273	-0.4490	-0.0323	-2.8086	-2.8242	-0.55%	-0.02
NH ₃ -CH ₄	-2.6783	-0.5145	-0.0401	-3.2329	-3.2175	0.48%	0.02
MAE						1.25%	0.09

Tables 3 and 6 with Tables S6 and S7 of the SI, respectively, the conclusions already drawn for the previous scheme can be confirmed. Indeed, the cc-pwCVTZ basis set leads again to similar errors, with the MAE increasing from 0.84% to 0.89% for the jul-ChS scheme and decreasing from 1.08% to 1.02% for the Δ (T) composite scheme.

3.2 The modified “cheap” scheme: the jun-ChS approach

Based on the results and discussions reported above, the best compromise between accuracy and computational cost for the modified “cheap” scheme is the jun-ChS approach, which can be summarized as follows:

- The starting term is the fc-CCSD(T)/jun-cc-pVTZ level of theory.
- The extrapolation to the CBS limit of the HF-SCF and MP2 energies is performed in single step with the 2-point n^{-3} expression,⁴⁸ the jun-cc-pVnZ family of basis sets being employed, with $n=T,Q$.
- The ΔE_{MP2}^{CV} term is calculated at the MP2 level using the cc-pwCVTZ basis set.

As well known from the literature (see, e.g. refs. 43,59) and already mentioned, the MP2 method tends to overestimate the interaction energies. As a consequence, an effective error compensation can occur, thus –hopefully– leading to a better agreement with

reference results when they also include the effect of quadruple excitations. To further test this aspect, the results of the jun-ChS and its analogous employing the jul-cc-pVnZ family of basis sets (jul-ChS scheme) have been compared with the best available interaction energies, which –as mentioned in the Computational Details– include higher excitation terms (up to CCSDT(Q)) and scalar relativistic corrections. This comparison required the removal from our A15 set of the formaldehyde and HCN dimers because of the lack of converged (and thus reliable) post-CCSD(T) contributions. The results are reported in Table 8. Furthermore, since the scheme that involves the separate contribution of the perturbative treatment of triples at the CCSD(T)/cc-pVTZ (denoted as $\Delta(T)$ -scheme, see Table 6) was found to be very promising, this comparison has been carried out also for this approach, the results being collected in Table 8 as well.

Table 8: jul-ChS and jun-ChS energies ($\text{kJ}\cdot\text{mol}^{-1}$) together with those from the $\Delta(T)$ -scheme.^a

Complex	"best-ref"	jul-ChS			jun-ChS			$\Delta(T)$ -scheme		
		Energy	Relative Error	Error	Energy	Relative Error	Error	Energy	Relative Error	Error
H ₂ O-H ₂ O	-21.0706	-21.0513	-0.09%	-0.02	-20.9844	-0.41%	-0.09	-20.7950	-1.31%	-0.28
NH ₃ -NH ₃	-13.2591	-13.2196	-0.30%	-0.04	-13.2721	0.10%	0.01	-13.1259	-1.01%	-0.13
HF-HF	-19.1836	-19.2280	0.23%	0.04	-19.2290	0.24%	0.05	-19.1088	-0.44%	-0.09
C ₂ H ₄ -C ₂ H ₄	-4.6442	-4.6761	0.69%	0.03	-4.7546	2.38%	0.11	-4.6027	-0.89%	-0.04
CH ₄ -CH ₄	-2.2677	-2.1928	-3.30%	-0.07	-2.2573	-0.46%	-0.01	-2.1912	-3.37%	-0.08
H ₂ O-NH ₃	-27.3885	-27.4259	0.14%	0.04	-27.3193	-0.25%	-0.07	-27.0671	-1.17%	-0.32
H ₂ O-C ₂ H ₄	-10.8198	-10.8492	0.27%	0.03	-10.8696	0.46%	0.05	-10.6685	-1.40%	-0.15
C ₂ H ₄ -H ₂ CO	-6.8367	-6.8579	0.31%	0.02	-6.9598	0.92%	0.12	-6.7639	-1.06%	-0.08
C ₂ H ₄ -NH ₃	-5.8409	-5.8479	0.11%	0.007	-5.8884	0.81%	0.05	-5.7499	-1.56%	-0.09
C ₂ H ₄ -CH ₄	-2.1673	-2.1584	-0.41%	-0.01	-2.1708	0.16%	-0.004	-2.1081	-2.73%	-0.06
HF-CH ₄	-6.9538	-7.0050	0.74%	0.04	-7.0258	1.04%	0.07	-6.8994	-0.78%	-0.05
H ₂ O-CH ₄	-2.8493	-2.7851	-2.25%	-0.06	-2.8086	-1.43%	-0.04	-2.7349	-4.02%	-0.11
NH ₃ -CH ₄	-3.2593	-3.2369	-0.69%	-0.02	-3.2329	-0.81%	-0.03	-3.1703	-2.73%	-0.09
MAE			0.73%	0.03		0.73%	0.05		1.32%	0.12

^a The CV term is computed using the cc-pwCVTZ basis set.

It is first of all noted that, for both the jul-ChS and jun-ChS approaches, the MAE decreases: from 0.89% to 0.73% for the former and from 1.20% to 0.73% for the latter. These results suggest that, once more accurate interaction energies are used as reference, the jul-cc-pVnZ and jun-cc-pVnZ families of basis sets lead to nearly equivalent schemes in terms of accuracy and, moving from the jul-ChS to the jun-ChS approach, only a small, negligible worsening of the MAE in terms of absolute error (from $0.03 \text{ kJ}\cdot\text{mol}^{-1}$ to 0.05

$\text{kJ}\cdot\text{mol}^{-1}$) is observed. This outcome further strengthens our choice of proposing the jun-ChS scheme as accurate model chemistry for treating with high accuracy (at a quite limited computational cost) non-covalent interactions. Furthermore, the role of the CV term can be highlighted. If this is neglected, the MAEs of the jun-ChS and jul-ChS approaches worsen to 1.26% and 0.96%, respectively.

Moving to the $\Delta(\text{T})$ -scheme, a relevant worsening is observed: the MAE increases from 1.02% to 1.32% and 2 outliers (4 if we lower the threshold to 2%) are present, all belonging to the methane-containing complexes. However, from the inspection of Tables 6 and 8, it is noted that the interaction energies are systematically underestimated. Since the $\Delta(\text{T})$ term is always negative, a scaling factor (greater than 1) can be introduced to correct this behavior when comparing to the “best-ref” energies. By doing so, a scaling factor of 1.1825 is obtained, which –in turn– leads to a significant reduction (from 1.32% to 0.82%) of the MAE.

3.3 B2PLYP-D3BJ geometries and CP corrections

The ChS model for energies is thought to be applied on top of equilibrium geometries evaluated using the B2PLYP-D3BJ functional in conjunction with a triple-zeta quality basis set.^{26,31,32} To investigate the effect of different reference structures and possible errors associated to their choice, a comparison has been carried between NCP and CP B2PLYP-D3BJ/may-cc-pVTZ energies obtained on top of the “CBS-georef” structures and at the following optimized geometries:

- NCP-B2PLYP-D3BJ/may-cc-pVTZ;
- CP-B2PLYP-D3BJ/may-cc-pVTZ;
- CP-B2PLYP-D3BJ/jun-cc-pVTZ.

We are fully aware of the limited accuracy of B2PLYP-D3BJ energies, but the focus is here only on the effect of the reference structures and not on the interaction energies them-

selves. Indeed, B2PLYP-D3BJ interaction energies are not directly used in the models under development and more recent functionals can surely provide better results. However, B2PLYP-D3BJ energy values are reasonably good in view of discussing the reliability of computed geometries.

Noted is that all B2PLYP-D3BJ optimized structures belong to the same symmetry point group as reported in Řezáč and Hobza, with the comparison of the “CBS-georef” and B2PLYP-D3BJ intermolecular distances being reported in Figure 1. It is apparent that the intermolecular distances obtained at the NCP-B2PLYP-D3BJ/may-cc-pVTZ level are shorter than the “CBS-georef” ones, while CP-B2PLYP-D3BJ geometry optimizations, employing either the may- or jun-cc-pVTZ basis set, lead to longer distances, sometimes closer to the reference. In this respect, consideration of half-CP correction could slightly improve the results, but already the more physically sound CP results, or their cheaper NCP counterparts, obtained with the relatively small may-cc-pVTZ basis set are close enough to the reference values to be considered sufficiently reliable. The only exceptions are the $\text{C}_2\text{H}_4\text{-CH}_4$ and $\text{CH}_4\text{-CH}_4$ complexes for which the B2PLYP-D3BJ intermolecular distances are strongly overestimated irrespective of the basis set and of the counterpoise correction. However, the PES of these complexes around the equilibrium structure is rather flat and sensitive to the level of theory; therefore, reference geometries might be not fully satisfactory as well. To better inspect the intermolecular distance in the methane dimer, its determination has been carried out by at the CCSD(T)/CBS+CV level^{29,60–62} by using the aug-cc-pVnZ sets ($n=\text{T,Q}$) for the extrapolation to the CBS limit. The CP intermolecular distance obtained is 3.6431 Å, thus being 0.005 Å longer than the “CBS-georef” value and 0.07 Å shorter than the NCP-B2PLYP/may-cc-pVTZ distance. This points out some limitations of current DFT models in accurately describing intermolecular interactions dominated by dispersive forces. While this issue deserves to be investigated in the future, all further analyses have been carried out by excluding the $\text{C}_2\text{H}_4\text{-CH}_4$ and $\text{CH}_4\text{-CH}_4$ complexes from the dataset considered.

The complete list of results is reported in Tables S12-S15, whereas a general summary is provided by Figure 3, which graphically reports both the CP- and NCP-B2PLYP-D3BJ/may-cc-pVTZ energy differences with respect to the corresponding “ref” values. For NCP energies, the MAEs are 3.6%, 3.9%, 3.7%, and 4.0% when using the “CBS-georef”, NCP-B2PLYP-D3BJ/may-cc-pVTZ, CP-B2PLYP-D3BJ/may-cc-pVTZ, and CP-B2PLYP-D3BJ/jun-cc-pVTZ references geometries, respectively. These MAEs increase to 7.6%, 7.6%, 7.4%, and 6.1%, respectively, when CP energies are considered. In general, the relative error of NCP and CP energies are of opposite sign (see Figure 3), thus suggesting that half-CP energies might lead to an improvement. In any case, the worsening noted when moving from NCP- to CP-energies is clearly due to the limited accuracy of the B2PLYP-D3BJ/may-cc-pVTZ level, and it should not be used to draw any conclusions about the importance of CP corrections in energy evaluations. However, these results are informative because they allow us to conclude that the use of NCP or CP-corrected reference structures seems to have a negligible effect on the determination of interaction energies. Furthermore, they also point out that the B2PLYP-D3BJ/may-cc-pVTZ and B2PLYP-D3BJ/jun-cc-pVTZ structures lead to nearly equivalent results for energetics, thus suggesting to use the geometries optimized with the smaller basis set in the jun-ChS scheme, which permits a reduction of the computational cost when large-sized systems are considered. This outcome is further reinforced by the fact that the MAEs for the “CBS-georef” structures are very similar to those obtained with the B2PLYP-D3BJ geometries, the overall conclusion being the suitability of B2PLYP-D3BJ for computing reference structures.

3.4 The final jun-ChS model

As discussed above (see section 3.3), the use of B2PLYP-D3BJ/may-cc-pVTZ optimized structures (instead of “CBS-georef”) has a very small effect on the energetics. Furthermore, the jun-cc-pVTZ basis set as well as the may-cc-pVTZ set (and, even, its variant without *d*-type functions on hydrogens) are demonstrated to perform well in the predic-

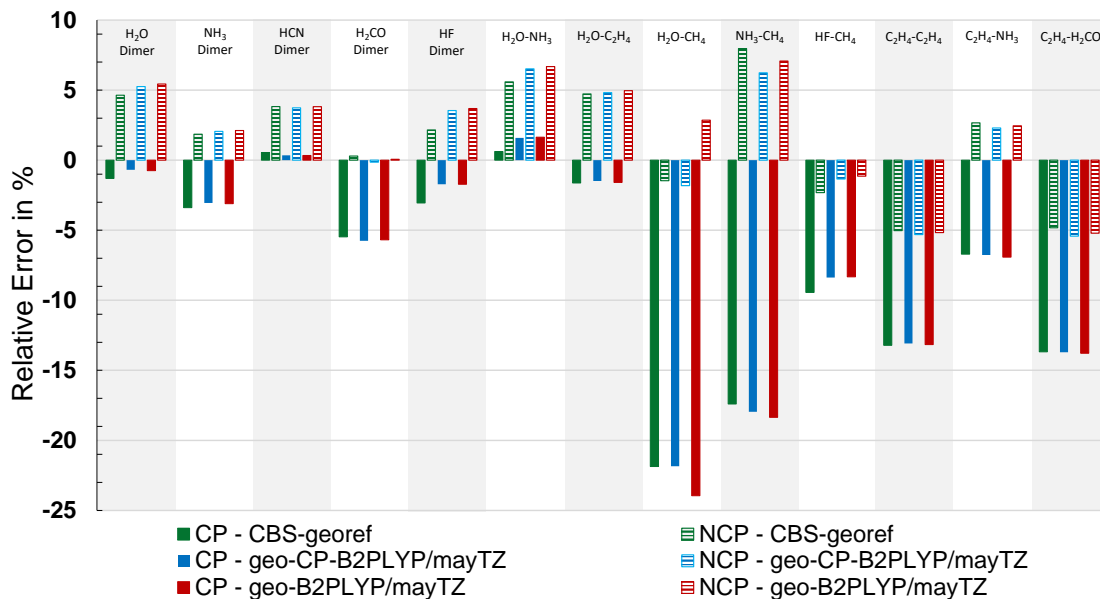


Figure 3: CP and NCP B2PLYP-D3BJ/may-cc-pVTZ energy differences (full and dashed bars, respectively) with respect to “ref” values computed at: (i) “CBS-georef” geometries (green bars), (ii) CP-B2PLYP-D3BJ/may-cc-pVTZ structures (blue bars), and (iii) B2PLYP-D3BJ/may-cc-pVTZ geometries (red bars).

tion of molecular proprieties and structures in conjunction with hybrid and double hybrid DFT functionals.^{46,63} For all these reasons, the jun-ChS scheme has been applied to both the CP-B2PLYP-D3BJ/jun-cc-pVTZ and CP-B2PLYP-D3BJ/may-cc-pVTZ reference geometries and compared to the “best-ref” values, the results being collected in Table 9. While similar trends are observed for both reference structures, it has to be noted that CP-B2PLYP-D3BJ geometries lead to a slight worsening of the MAE: from 0.73% (see Table 8; “geo-ref”) to 0.95% and 0.87% for CP-B2PLYP-D3BJ/jun-cc-pVTZ and CP-B2PLYP-D3BJ/may-cc-pVTZ reference geometries, respectively. In terms of absolute errors, the MAE increases from 0.05 kJ·mol⁻¹ to 0.08/0.09 kJ·mol⁻¹, uncertainties still well within our target accuracy.

3.4.1 The jun-ChS approach at work: Case studies

The A24 set does not include molecular complexes containing second-row atoms and, more generally, accurate interaction energies for this class of systems are very limited in

Table 9: jun-ChS energies ($\text{kJ}\cdot\text{mol}^{-1}$) at the CP-B2PLYP-D3BJ/jun-cc-pVTZ and CP-B2PLYP-D3BJ/may-cc-pVTZ geometries.

Complex	"best-ref"	CP-B2PLYP-D3BJ/jun-cc-pVTZ geometry					CP-B2PLYP-D3BJ/may-cc-pVTZ geometry				
		fc-CCSD(T)	ΔE_{MP2}^∞	ΔE_{MP2}^{CV}	Energy	Relative Error	fc-CCSD(T)	ΔE_{MP2}^∞	ΔE_{MP2}^{CV}	Energy	Relative Error
H ₂ O-H ₂ O	-21.0706	-19.4234	-1.5614	-0.1457	-21.1305	0.28%	-19.4081	-1.5543	-0.1453	-21.1077	0.18%
NH ₃ -NH ₃	-13.2591	-11.9003	-1.3133	-0.0768	-13.2904	0.24%	-19.9066	-1.3264	-0.0770	-13.3100	0.38%
HF-HF	-19.1836	-17.5817	-1.8116	-0.1129	-19.5062	1.68%	-17.5628	-1.7959	-0.1131	-19.4718	1.50%
C ₂ H ₄ -C ₂ H ₄	-4.6442	-3.7742	-0.8809	-0.0458	-4.7009	1.22%	-3.7847	-0.8614	-0.0448	-4.6909	1.01%
H ₂ O-NH ₃	-27.3885	-25.4564	-1.9690	-0.1861	-27.6115	0.81%	-25.4307	-1.9506	-0.1842	-27.5655	0.65%
H ₂ O-C ₂ H ₄	-10.8198	-9.4814	-1.2440	-0.0951	-10.8205	0.01%	-9.4793	-1.2400	-0.0948	-10.8141	-0.05%
C ₂ H ₄ -H ₂ CO	-6.8367	-5.7133	-1.1033	-0.0595	-6.8761	0.58%	-5.7124	-1.0807	-0.0584	-6.8515	0.22%
C ₂ H ₄ -NH ₃	-5.8409	-4.9671	-0.8372	-0.0533	-5.8576	0.29%	-4.9768	-0.8282	-0.0525	-5.8576	0.29%
HF-CH ₄	-6.9538	-5.9098	-1.0851	-0.0972	-7.0921	1.99%	-5.9039	-1.0786	-0.0964	-7.0789	1.80%
H ₂ O-CH ₄	-2.8493	-2.3397	-0.4301	-0.0310	-2.8008	-1.70%	-2.3388	-0.4291	-0.0309	-2.7988	-1.77%
NH ₃ -CH ₄	-3.2593	-2.6717	-0.4964	-0.0388	-3.2069	-1.61%	-2.6707	-0.4948	-0.0387	-3.2042	-1.69%
MAE						0.95%					0.87%

the literature.^{55,64,65} For these reasons, we have decided to study four small non-covalent complexes bearing at least one second-row element for which reasonably accurate computations are available, namely FH₂P \cdots NH₃, FH₂P \cdots SH₂, FHS \cdots OFH and CH₃OH \cdots HCl (see Figure 4). Their investigation paves the way for subsequently addressing two larger systems containing one and two sulfur atoms (the thiouracil-water and dimethylsulfide-sulfur dioxide adducts, respectively). Even if not explicitly stated, as mentioned in the Computational Details section, for second-row elements the tight *d*-functions-augmented jun-cc-p-V(*n*+*d*)Z sets⁴⁷ are used. Furthermore, the CP correction is always incorporated.

To compare our results with those reported in the literature when dissociation energies (D_e) are provided, the jun-ChS interaction energies (computed using B2PLYP-D3BJ/jun-cc-pVTZ optimized structures) have to be corrected for the deformation contribution, Δ^{def} , which is defined as:

$$\Delta^{\text{def}} = E_A^{A,A} + E_B^{B,B} - E_A^{A,AB} - E_B^{B,AB} \quad (2)$$

where the superscript "Basis,Geo" indicates the basis set and the geometry used, respectively, while the subscript refers to the monomer considered. Therefore, to give an example, $E_B^{B,AB}$ is the energy of the monomer B computed using its basis set at the geometry the monomer assumes in the complex. Needless to say, the energy of the isolated fragments has been computed using the jun-ChS scheme at the B2PLYP-D3BJ/jun-cc-pVTZ

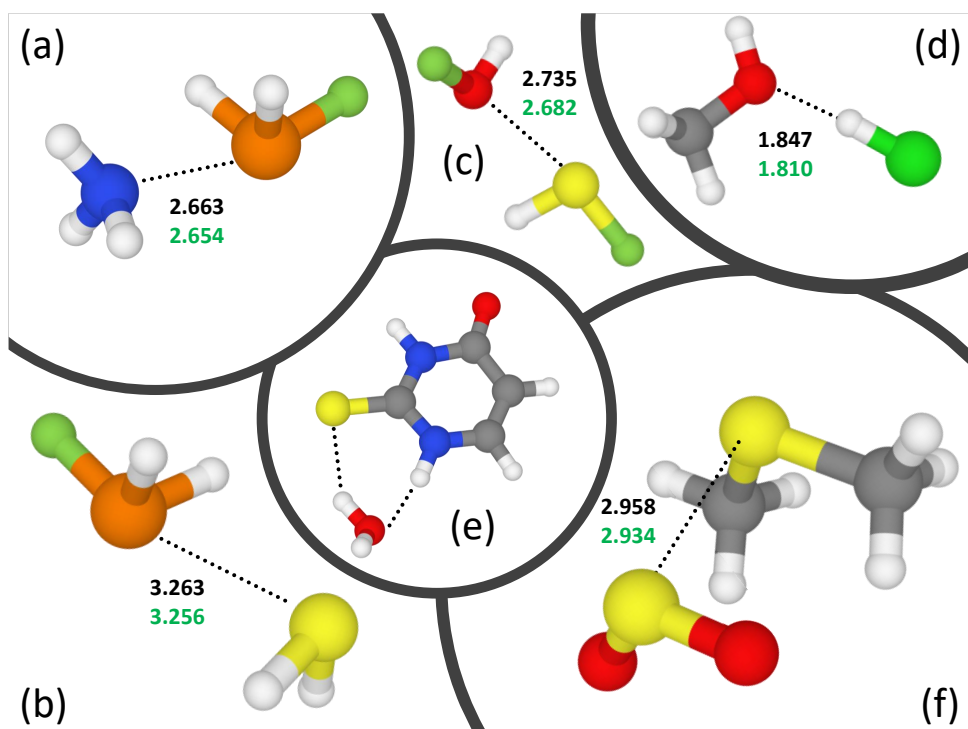


Figure 4: Test cases containing second-row atoms: (a) $\text{FH}_2\text{P} \cdots \text{NH}_3$, (b) $\text{FH}_2\text{P} \cdots \text{SH}_2$, (c) $\text{FHS} \cdots \text{OFH}$, (d) $\text{CH}_3\text{OH} \cdots \text{HCl}$, (e) thiouracil-water, (f) dimethylsulfide-sulfur dioxide complexes. The intermolecular distances (in Å) reported in the literature (see text) are in black, while in green are those optimized at the CP-B2PLYP-D3BJ/jun-cc-pV(T+d)Z level.

optimized geometry. Once the deformation energy is computed, the equilibrium dissociation energy is straightforwardly derived as $D_e = E_{\text{CP}}^{\text{int}} - \Delta^{\text{def}}$, the detailed results being reported in Table 10.

The first three complexes ($\text{FH}_2\text{P} \cdots \text{NH}_3$, $\text{FH}_2\text{P} \cdots \text{SH}_2$, and $\text{FHS} \cdots \text{OFH}$) are representative of pnictogen/chalcogen bonds with linear arrangements, whose interaction is ruled by electrostatic rather than dispersive forces. Reference values for geometries and interaction energies have been obtained at the CCSD(T)/aug-cc-pVTZ level, and reported in ref. 65. The B2PLYP-D3BJ intermolecular distances (in green in Figure 4) are in close agreement (within 0.01 Å) with the CCSD(T)/aug-cc-pVTZ counterparts (in black in Figure 4) for the first two complexes, whereas the difference increases to about 0.05 Å for $\text{FHS} \cdots \text{OFH}$. Moving to energetics, it is noted that our results lie in between the CP and NCP binding energies reported in ref. 65, with deviations from the corresponding half-CP

Table 10: Case studies: jun-ChS energies ($\text{kJ}\cdot\text{mol}^{-1}$) at the CP-B2PLYP-D3BJ/jun-cc-pVTZ reference geometries.

Complex	fc-CCSD(T)/jun-cc-pVTZ	ΔE_{MP2}^{∞} /jun-cc-pV(T,Q)Z	ΔE_{MP2}^{CV}	E_{CP}^{int}	Δ^{def}	D_e	Literature	
							D_e	E^{int}
FH ₂ P-NH ₃	-24.8311	-4.4465	-0.2669	-29.5445	-2.1238	-27.42	-28.49 ^a ; -25.52 ^b	
FH ₂ P-SH ₂	-10.7386	-3.9130	-0.2836	-14.9352	-0.5120	-14.42	-15.40 ^a ; -12.89 ^b	
FHS-OFH	-11.3921	-2.6460	-0.0930	-14.1311	-0.3666	-13.36	-15.15 ^a ; -12.55 ^b	
CH ₃ OH-HCl	-23.7406	-3.1925	-0.3500	-27.2831	-1.3241	-25.96		-26.61 ^c
Bz-Bz	-8.8894	-2.5160	-0.1859	-11.5913	-	-		-11.44 ^d
Formamide dimer	-62.5877	-4.9149	-0.3146	-67.8172	-5.6435	-62.17	-61.7 ^e (-62.34 ^f)	
py-H ₂ O	-27.4756	-2.1073	-0.1938	-29.7767	-1.0866	-28.69		-28.99 ^d (-29.16 ^g)
TBA-DFM	-16.1906	-1.3903	-0.1220	-17.7029	-0.6170	-17.09	-17.2 ^h	
NE-TMA	-25.0525	-2.5467	-0.1668	-27.7660	-0.2523	-27.51	-27.1 ⁱ	
ThioUra-H ₂ O	-43.2283	-4.1817	-0.3261	-47.7361	-2.5989	-45.14	-41.06 ^l	
DMS-SO ₂	-25.3454	-8.1692	-0.5300	-34.0446	-2.9672	-31.08	-28.51 ^m	

^a NCP results from Ref. 65. ^b CP results from Ref. 65. ^c Ref. 55. ^dRef. 56. ^e Ref. 66. ^f Ref. 67 ^g Ref. 68 ^h Ref. 32. ⁱ Ref. 26. ^l Ref. 30. ^m Ref. 31.

values smaller than $0.5 \text{ kJ}\cdot\text{mol}^{-1}$. The differences increase when our results are compared with either NCP or CP values, with discrepancies of about $1.5\text{-}1.9 \text{ kJ}\cdot\text{mol}^{-1}$ for the latter. This first example tends therefore to support the idea that half-CP results provide the most reliable estimates when quantum-chemical calculations do not account for the convergence to the CBS limit. For FHS \cdots OFH, a more accurate value, $13.56 \text{ kJ}\cdot\text{mol}^{-1}$ at the CP-DLPNO-CCSD(T)/aug-cc-pV5Z level, is available in the literature (see supplementary material of ref. 65), which is in nice agreement with our jun-ChS result ($13.36 \text{ kJ}\cdot\text{mol}^{-1}$), the difference being only $0.2 \text{ kJ}\cdot\text{mol}^{-1}$ and thus well within the error bar of about $0.4 \text{ kJ}\cdot\text{mol}^{-1}$ introduced by the DLPNO approximation. In any case, we can consider the present results as the most accurate and reliable among those available in the literature and they suggest that improved benchmark values should be compiled for non-covalent interactions involving second-row atoms (possibly using the jun-ChS approach). The last note concerns deformation energies; they are always non-negligible (even $\sim 7\%$ for FH₂P \cdots NH₃) and therefore should be included in any quantitative study aiming at the accurate evaluation of binding energies.

For the CH₃OH \cdots HCl adduct, the reference interaction energy, $-26.61 \text{ kJ}\cdot\text{mol}^{-1}$ (to be compared with our value of $-27.28 \text{ kJ}\cdot\text{mol}^{-1}$), is taken from ref. 55 and was obtained by means of the composite approach defined as “Gold2”, which assumes as best result the half-CP corrected MP2-F12/cc-pV(T,Q)Z-F12 value combined with a high-order term

derived from the difference between conventional, half-CP corrected CCSD(T) and MP2 results extrapolated to the CBS limit using the jul-cc-pVTZ and jul-cc-pVQZ basis sets.⁵⁵ Taking into account that the core-valence correlation correction amounts to about $-0.35 \text{ kJ}\cdot\text{mol}^{-1}$ (a value not negligible in view of the aimed accuracy) and the effect of the tight *d*-functions leads to contribution of $\sim 0.06 \text{ kJ}\cdot\text{mol}^{-1}$ (see discussion later in the text), a remaining difference of about $0.4 \text{ kJ}\cdot\text{mol}^{-1}$ is noted, which can be ascribed to the intrinsic limits of our scheme, to the fact that we consider CP-corrected energies (and not half-CP), and to possible differences in the reference geometries. Indeed, the CP-B2PLYP-D3BJ/jun-cc-pVTZ intermolecular distance (in green in Figure 4) is 0.037 \AA shorter than the CCSD(T)/CBS⁶⁴ reference value (in black in Figure 4) used in ref. 55. Furthermore, the missing effect of diffuse functions in the MP2-F12 CBS extrapolation can also play a role, as demonstrated in ref 69. To move from the interaction to dissociation energy, the Δ^{def} term should be added, which is not negligible at all ($-1.32 \text{ kJ}\cdot\text{mol}^{-1}$, i.e. $\sim 5\%$).

A note is deserved on the effect of tight *d*-functions-augmented jun-cc-p-V(*n*+*d*)Z sets for second-row elements, which can be derived by comparing the results of Table 10 with those collected in Table S16. It is apparent that, in absolute terms, the differences in the interaction energy are small and nearly negligible: ~ 0.11 , ~ 0.02 , ~ 0.02 , and $\sim 0.06 \text{ kJ}\cdot\text{mol}^{-1}$ for $\text{FH}_2\text{P}\cdots\text{NH}_3$, $\text{FH}_2\text{P}\cdots\text{SH}_2$, $\text{FHS}\cdots\text{OHF}$ and $\text{CH}_3\text{OH}\cdots\text{HCl}$, respectively. However, it has to be noted that such basis sets do not significantly affect the computational cost, while their employment is important in view of defining a general composite scheme for energetics of molecular systems involving both covalent and non-covalent interactions.

To further test the performance of the newly developed jun-ChS model for systems larger than those of the benchmark set, seven case studies have been selected from previous works reported in the literature. These are: the benzene and formamide dimers, the pyridine-water (py-H₂O) complex, the difluoromethane-tertbutylalcohol (DFM-TBA) complex, and the molecular adduct formed by nitroethane and trimethylamine (NE-TMA) (see Figure 5) together with systems containing second-row elements already mentioned,

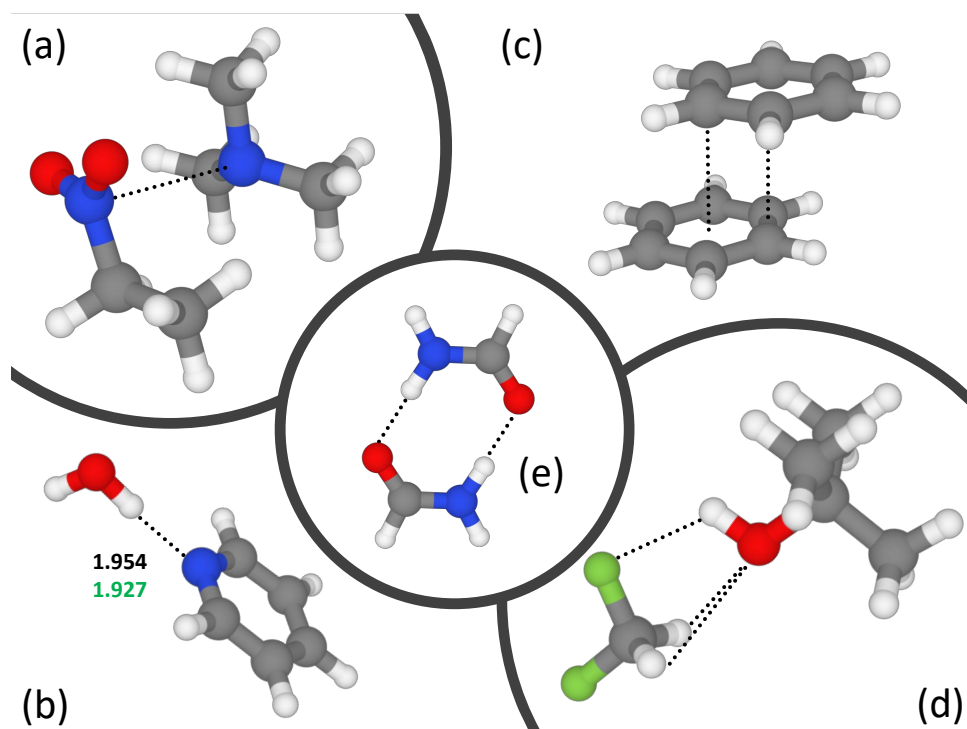


Figure 5: Test cases containing first-row atoms: (a) nitroethane-trimethylamine complex, (b) pyridine-water complex (the intermolecular distance is given for two different levels of theory in Å, see text), (c) benzene dimer, (d) difluoromethane-tertbutylalcohol complex, (e) formamide dimer.

i.e., the thiouracil-water (ThioUra-H₂O) complex and the adduct formed by dimethylsulfide and sulfur dioxide (DMS-SO₂) (see Figure 4). If different isomers are possible, only the most stable one has been investigated in the present study (see Figures 4 and 5). As above, for all these systems, the reference geometry in the jun-ChS approach has been optimized at the CP-B2PLYP-D3BJ/jun-cc-pVTZ level, with the only exception being the benzene dimer, for which the structure reported in the literature has been used. In the case of the py-H₂O adduct, the geometries reported in refs. 70,71 have also been considered to further address the effect of the reference structure.

The parallel displaced benzene dimer (Bz-Bz) can be considered the prototype of the π - π stacking interaction. Among the several results available in the literature, we selected as reference that reported in ref. 56, which has been obtained from a composite approach that accounts for the MP2-F12 extrapolation to the CBS limit (cc-pV n Z-F12; n =T,Q) and a contribution for higher order excitations from CCSD-(F12*)/cc-pVDZ-F12 calculations

to which a scaled (T^*) factor, computed with the cc-pVDZ-F12 set, was also added. The latter interaction energy, $-11.40 \text{ kJ}\cdot\text{mol}^{-1}$, agrees well with our value of $-11.59 \text{ kJ}\cdot\text{mol}^{-1}$. The resulting discrepancy, $\sim 0.2 \text{ kJ}\cdot\text{mol}^{-1}$, is completely recovered if the CV term ($-0.19 \text{ kJ}\cdot\text{mol}^{-1}$ at the MP2/cc-pwCVTZ level), which is missing in the reference value of ref. 56, is taken into account. Indeed, in view of the target accuracy such a contribution cannot be neglected. Overall, this example demonstrates the reliability and accuracy of the jun-ChS approach in describing also the π - π stacking interaction.

The second system considered from the set of larger case studies is the formamide dimer for which the dissociation energy at the CCSD(T)/CBS+CV level was reported in ref. 66. The latter composite scheme is entirely based on CCSD(T) calculations, with the extrapolation to the CBS limit performed using the two-step procedure already mentioned in the Computational Details section (with cc-pVTZ and cc-pVQZ basis sets employed for the CCSD(T) correlation energy) and the CV contribution evaluated at the CCSD(T)/cc-pCVTZ level. CCSD(T)/CBS+CV energies were computed at the corresponding optimized geometry. According to ref. 66 and the literature on this topic,^{29,72–74} the equilibrium CCSD(T)/CBS+CV dissociation energy, $-61.7 \text{ kJ}\cdot\text{mol}^{-1}$, is expected to have an accuracy of about $1\text{--}2 \text{ kJ}\cdot\text{mol}^{-1}$. The jun-ChS value ($-62.17 \text{ kJ}\cdot\text{mol}^{-1}$) is in good agreement with this result, the observed discrepancy of $0.5 \text{ kJ}\cdot\text{mol}^{-1}$ being well within the statistical error affecting the CCSD(T)/CBS+CV scheme, which is – however – considerably more expensive than the jun-ChS model. Furthermore, the present result is in extremely good agreement with the benchmark value of ref. 67, $-62.34 \text{ kJ}\cdot\text{mol}^{-1}$. The latter value was obtained as the average between the CP and NCP dissociation energies computed by extrapolating the MP2 energy (aug-cc-pVDZ, aug-cc-pVTZ and aug-cc-pVQZ basis sets employed for a three point extrapolation) and a CCSD(T) correction computed with the aug-cc-pVTZ basis set.

The third molecular complex of this series is the py- H_2O adduct. For this system, different interaction energies were reported in literature: starting from the original work in the S66 database,^{70,71} which predicted a value of $-28.59 \text{ kJ}\cdot\text{mol}^{-1}$, a revised interac-

tion energy of $-29.16 \text{ kJ}\cdot\text{mol}^{-1}$ was later obtained.⁶⁸ In addition, in ref. 56 a value of $-28.99 \text{ kJ}\cdot\text{mol}^{-1}$ was derived as described above. Our value, computed using the jun-ChS scheme (CP-B2PLYP-D3BJ/jun-cc-pVTZ optimized geometry depicted in Figure 5), is $-29.78 \text{ kJ}\cdot\text{mol}^{-1}$, which agrees within $0.8 \text{ kJ}\cdot\text{mol}^{-1}$ with the most accurate value reported in the literature. However, if one computes the jun-ChS interaction energy at the geometry reported in the literature^{70,71} (CCSD(T)/CBS value, in black in Figure 5), then the discrepancy reduces to $0.22 \text{ kJ}\cdot\text{mol}^{-1}$, which is within the uncertainty affecting the F12-composite scheme of ref. 56 and the error essentially vanishes if one considers the fact that our CV term leads to an additional correction of $-0.19 \text{ kJ}\cdot\text{mol}^{-1}$.

By further increasing the complexity of the system, the next case study considered is the M1 isomer of the TBA-DFM adduct (see ref. 32 for all details on stable isomers). The dissociation energy reported in the literature³² was computed at the B2PLYP-D3BJ/may-cc-pVTZ-*dH* optimized geometry using the “cheap” scheme, which means that the cc-pVnZ basis sets were employed and the $\Delta\alpha$ term was not taken into account. From the inspection of Table 10, it is noted that the present result and dissociation energy from ref. 32 are in good agreement, with the jun-ChS model predicting the molecular complex to be less stable by about $0.1 \text{ kJ}\cdot\text{mol}^{-1}$ ($\sim 0.58\%$ in relative terms). To further discuss this outcome, it is necessary to recall that the M1 isomer of the TBA-DFM cluster involves the formation of one $\text{O}-\text{H}\cdots\text{F}$ and two $\text{C}-\text{H}\cdots\text{O}$ weak hydrogen bonds. With this in mind, the good agreement between the ChS and jun-ChS results suggests that, when the molecular complex is characterized by an interaction dominated by electrostatic forces, the effect of diffuse functions is less important.

The subsequent molecular complex considered is the NE-TMA adduct, which has been recently studied by means of a joint experimental-computational investigation that led to the first characterization of a nitrogen-nitrogen non-covalent interaction.²⁶ Also in this case, the dissociation energy was computed by means of the traditional ChS scheme, as above for the TBA-DFM complex. However, for this system, the discrepancy between the ChS and jun-ChS results is more pronounced, with the new composite scheme predict-

ing the molecular complex more stable by about $0.4 \text{ kJ}\cdot\text{mol}^{-1}$, which means –in relative terms– a difference of 1.5%. This worsening is in line with the preceding discussion concerning the proper inclusion of the effect of diffuse functions when dispersive forces are involved. In fact, according to ref. 26, the intermolecular interaction is equally ruled by electrostatic and dispersive forces.

The two last complexes involve second-row atoms. The first one is the thiouracil-water complex previously studied employing the original ChS model including the $\Delta\alpha$ correction for diffuse functions and neglecting tight d functions on the sulfur atom.³⁰ First of all, the comparison between the ChS and B2PLYP-D3BJ equilibrium geometries (see columns 2 and 4 of Table 1 of ref. 30) confirms the reliability and accuracy of B2PLYP-D3BJ structural parameters. However, the new jun-ChS dissociation energy is –in absolute terms– about $4 \text{ kJ}\cdot\text{mol}^{-1}$ (i.e. roughly 10%) larger than the value given in ref. 30, thus confirming the unsuitability of the $\Delta\alpha$ term for incorporating the effect of diffuse functions and pointing out the significant improvement brought by a more consistent treatment of diffuse functions (also including the use of tight d functions for sulfur). Noted is that the discrepancy in relative term ($\sim 10\%$) is in line with the MAE (12.7 %) obtained for the ChS approach with the separated $\Delta\alpha$ term (see Table 5).

The last complex is the prototype of chalcogen interactions recently studied by means of a joint microwave-theoretical investigation.³¹ The result issuing from the traditional ChS approach without the $\Delta\alpha$ term shows a relevant discrepancy from the jun-ChS value, the latter D_e being nearly 10% larger. In the present case, the difference in the reference geometries can contribute significantly to the slight disagreement noted. The intermolecular distances, obtained at the ChS and CP-B2PLYP-D3BJ/jun-cc-pV(T+ d)Z levels, are reported in Figure 4, thus showing that the ChS value is about 0.024 \AA longer than the B2PLYP-D3BJ one.

4 Additional Aspects

Having defined and validated the new jun-ChS model, in this section we address three additional aspects, namely (i) explicitly correlated (F12) models, (ii) the computational cost of the jun-ChS model with respect to the original ChS approach and its aug-ChS variant, and (iii) many-body effects.

Although the F12 ansatz is very effective in reducing the number of high angular momentum basis functions for converged calculations, it requires specific F12 basis sets,⁷⁵ which are significantly larger than the corresponding (i.e. same cardinal number) conventional basis sets. Furthermore, diffuse functions must be included for an effective treatment of non-covalent complexes irrespective of the presence or not of F12 terms (see ref. 69) and intermolecular correlation benefits from explicit correlation significantly less than short-range intramolecular correlation. At the level of triple-zeta basis sets, diffuse functions on hydrogen atoms have a marginal role and have been neglected in all the most sophisticated composite approaches.^{55,56} Therefore, since explicit correlation increases the computer cost, requires specifically optimized basis sets, and is of no real benefit for the contribution of triple excitations, the trade-off between accuracy and basis-set dimension is different for conventional and explicitly-correlated approaches.

Table 11 reports in the first three columns the number of basis functions for some selected basis sets as well as, in the last two columns (for naphthalene and anthracene), their relative computer time, calculations being carried out at the CCSD(T) level for standard basis sets and at the CCSD(T)-F12b level for the others.⁷⁶ Table 11 clearly shows that the aug-cc-pVDZ-F12 and jun-cc-pVTZ (or jul-cc-pVTZ) basis sets are comparable in terms of both dimensions and computer times. Therefore, a ChS approach based on CCSD(T)-F12/aug-cc-pVDZ-F12 does not improve the computational cost (while probably worsening the contribution of triple excitations) with respect to the jun-ChS (or even jul-ChS) protocol, thus not justifying, in our opinion, the replacement of conventional approaches in favor of their explicitly correlated counterparts. On the other hand, a com-

posite approach based on CCSD(T)-F12/aug-cc-pVTZ-F12 would increase significantly the cost and, in view of the accuracy of the jun-ChS model, such a worsening in terms of computational requirement does not seem to be justified. On the other hand, for large systems a better solution can be represented by local correlation treatments, as briefly addressed in the Introduction.

Table 11: Selected (conventional and F12) basis sets: number of basis functions and relative computer time.

Basis set	H ^a	first-row ^a	second-row ^a	naphtalene ^b	anthracene ^b
aug-cc-pVDZ ^c	9	23	27	1	1
aug-cc-pVDZ-F12 ^d	13	35	44	31	20
jun-cc-pVTZ ^e	14	39	48	25	22
jul-cc-pVTZ ^e	14	46	55	25	22
aug-cc-pVTZ-F12 ^d	27	65	74	191	138

^a Number of basis functions. ^b Relative timings without diffuse functions from ref. 76. ^c Refs. 34,77. ^d Ref. 69. ^e Ref. 46.

Having addressed the size of the conventional basis sets with respect to those purposely tailored for the F12 method and the computer time of the latter, it is now important to compare the computational cost of the jun-ChS model with respect to the original ChS approach and the fully augmented aug-ChS variant. A straightforward way to perform such comparison is to analyze the computer time required for the bottleneck step, which is the CCSD(T) computation in conjunction with the triple-zeta basis sets required by the different schemes, i.e cc-pVTZ, jun-cc-pVTZ and aug-cc-pVTZ. In detail, Table 12 reports the computer time needed to evaluate the CCSD(T) energy of the pyridine-water complex with the basis sets mentioned above. Furthermore, the total running time (including the CCSD(T) energies for the two monomers, both of them using their basis set, as well as that for the complex) is also reported. The results collected in Table 12 show that the computer time required for the CCSD(T)/jun-cc-pVTZ calculation nearly doubles that at the CCSD(T)/cc-pVTZ level. However, this increase is counterbalanced by the gain in terms of accuracy of the scheme: MAE of 1.2% for jun-ChS vs 2.9% for ChS. The use of a fully augmented basis set leads to a significant increase of the computational cost (about

eight and four times higher with respect to the cc-pVTZ and jun-cc-pVTZ basis sets, respectively) which is, however, accompanied by a decrease in accuracy with respect to the jun-ChS scheme. This result gives further support to our choice of using the jun-cc-pV n Z family as the best compromise to account for diffuse functions without any huge increase of the computational cost while retaining a fully satisfactory accuracy in the description of non-covalent systems.

Table 12: Pyridine-water complex: computer time for the CCSD(T) calculation in conjunction with triple-zeta basis sets.

Basis set	Number of basis functions			Running time ^a	
	Pyridine	H ₂ O	Total	Total	Complex
cc-pVTZ	250	58	308	7h 47min 53s	4h 00min 11s
jun-cc-pVTZ	304	67	371	16h 12min 42s	8h 03min 32s
aug-cc-pVTZ	391	92	483	59h 52min 24s	28h 49min 12s

^a Computed using a node with 64 cpus and 120GB of memory.

Although the present study is devoted to bi-molecular complexes, a comment is deserved for more complex situations, i.e. molecular adducts containing more than 2 fragments. A recent study on water clusters,⁷⁸ including up to six molecules, allows for estimating the deviations between the jul-ChS approach and the “platinum” CCSD(T)/CBS composite scheme (see ref. 78). The root mean square (RMS) deviation of the MP2/CBS contribution issuing from the extrapolation of heavy-augmented triple- and quadruple-zeta (i.e. jul-cc-pV n Z) basis sets amounts to 0.12 kJ·mol⁻¹, irrespective of the inclusion or not of counterpoise corrections (see Table 1 of ref. 78), whereas the corresponding RMS error for the CCSD(T) - MP2 difference contribution computed with the heavy-augmented triple-zeta basis set (i.e. jul-cc-pVTZ) is below 0.1 kJ·mol⁻¹ (see Table 5 of ref. 78). As a matter of fact, both these contributions are of the same order of magnitude as the CV correction (neglected in ref. 78) and show that the performance of the jul-ChS (or jun-ChS, which delivers equivalent results for the water dimer, see Tables 3 and 4) model does not worsen significantly for complexes containing more than two fragments. Moreover, the

total energy of a cluster containing n individual fragments can be conveniently expressed in terms of a many-body expansion. The corresponding analysis of water complexes indicates that high-level (post-MP2) contributions are essential for three-body terms, desirable for four-body terms and negligible for higher terms. Therefore, the jun-ChS approach can be considered fully adequate also for treating n -molecular fragments (with $n > 2$).

5 Concluding remarks

By employing the benchmark CCSD(T)/CBS+CV interaction energies⁴⁴ for a set of 15 non-covalent molecular complexes and the CCSDT(Q)/CBS+CV+rel (‘rel’ standing for scalar relativistic effect) interaction energies⁴⁴ for a subset of 13 molecular complexes, taken from the A24 dataset,³⁵ the so-called “cheap” composite scheme has been investigated in great detail with the final aim of deriving an improved model for accurately describing non-covalent interaction without any huge increase of the overall computational cost with respect to the original approach.^{12–14} Our work confirms the need of incorporating diffuse functions in quantitative studies and introduces a modified version of the “cheap” composite scheme, here denoted as jun-ChS, as a very good compromise between accuracy and computational cost.

While it has been demonstrated that the inclusion of a separate term ($\Delta\alpha$) is not able to properly recover such an effect, the employment of the “seasonal” basis sets developed by Papajak et al.⁴⁶ in the MP2 extrapolation to the CBS limit as well as in CCSD(T) energy evaluations leads to encouraging results. Once employed in conjunction with the CCSD(T)/CBS reference structures from ref. 35, the jun-ChS model provides results that show an average deviation of $\sim 0.73\%$ from the benchmark CCSDT(Q)/CBS+CV+rel interaction energies. A limited worsening in its accuracy has been observed when CP-B2PLYP-D3BJ (in conjunction with triple-zeta quality basis set) optimized geometries are used as reference structures, the MAE however remaining below 1%. The obtained results are therefore fully satisfactory, also because –in terms of absolute errors– the agreement

is well within $0.2 \text{ kJ}\cdot\text{mol}^{-1}$.

By applying the jun-ChS model chemistry to small complexes including second-row atoms, we have shown that previous benchmark studies are not sufficiently accurate and reliable, also because of the lack of extrapolation to the CBS limit and/or core-valence correlation effects. Seven larger complexes have also been investigated leading to reliable results and confirming the effectiveness of the jun-ChS model.

Finally, it has to be noted that the approach accounting for a separate contribution of triple excitations was found to be very promising. In fact, if the CCSD(T)/CBS+CV energy is taken as reference, this scheme, with a MAE of 1.08%, meets our target accuracy. Furthermore, it has been observed that the inclusion of higher order terms in the CC expansion can be improved using a scaling factor, which allows for reproducing the CCSDT(Q)/CBS+CV+rel interaction energies with a MAE of 0.82%. These results are particularly encouraging in view of extending the “cheap” protocol to larger systems than those considered as test cases. Work along this direction is under consideration.

Acknowledgement

This work has been supported by MIUR “PRIN 2015” funds (Grant Number 2015F59J3R) and by the University of Bologna (RFO funds). The SMART@SNS Laboratory (<http://smart.sns.it>) is acknowledged for providing high-performance computer facilities. The authors also thank Dr. Lorenzo Spada for useful discussions.

References

- (1) Riley, K. E.; Hobza, P. Noncovalent interactions in biochemistry. *WIREs Comput. Mol. Sci.* **2011**, *1*, 3–17.
- (2) Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. Biomolecular Simulations: Recent Developments in Force Fields, Simulations of Enzyme Catalysis, Protein-Ligand, Protein-Protein, and Protein-Nucleic Acid Noncovalent Interactions. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211–243.
- (3) Knowles, R. R.; Jacobsen, E. N. Attractive noncovalent interactions in asymmetric catalysis: Links between enzymes and small molecule catalysts. *Proc. Nat. Acad. Sci. USA* **2010**, *107*, 20678–20685.
- (4) Tang, W.; Johnston, S.; Iggo, J. A.; Berry, N. G.; Phelan, M.; Lian, L.; Bacsá, J.; Xiao, J. Cooperative Catalysis through Noncovalent Interactions. *Angew. Chem. Int. Ed.* **2013**, *52*, 1668–1672.
- (5) Beno, B. R.; Yeung, K.-S.; Bartberger, M. D.; Pennington, L. D.; Meanwell, N. A. A Survey of the Role of Noncovalent Sulfur Interactions in Drug Design. *J. Med. Chem.* **2015**, *58*, 4383–4438.
- (6) Alonso, J. L.; López, J. C. *Microwave Spectroscopy of Biomolecular Building Blocks*; Top. Curr. Chem.; Springer Berlin Heidelberg, 2015; Vol. 364; pp 335–401.
- (7) Perez, C.; Neill, J. L.; Muckle, M. T.; Zaleski, D. P.; Peña, I.; Lopez, J. C.; Alonso, J. L.; Pate, B. H. Water-Water and Water-Solute Interactions in Microsolvated Organic Complexes. *Angew. Chem. Int. Ed.* **2015**, *54*, 979–982.
- (8) Becucci, M.; Melandri, S. High-Resolution Spectroscopic Studies of Complexes Formed by Medium-Size Organic Molecules. *Chem. Rev.* **2016**, *116*, 5014–5037.
- (9) Zwier, T. S. The spectroscopy of solvation in hydrogen-bonded aromatic clusters. *Ann. Rev. Phys. Chem.* **1996**, *47*, 205–241.

- (10) Bernhard, D.; Fatima, M.; Poblitzki, A.; Steber, A. L.; Perez, C.; Suhm, M. A.; Schnell, M.; Gerhards, M. Dispersion-controlled docking preference: multi-spectroscopic study on complexes of dibenzofuran with alcohols and water. *Phys. Chem. Chem. Phys.* **2019**, *21*, 16032–16046.
- (11) Lesarri, A.; Mata, S.; López, J. C.; Alonso, J. L. A laser-ablation molecular-beam Fourier-transform microwave spectrometer: The rotational spectrum of organic solids. *Rev. Sci. Instrum.* **2003**, *74*, 4799–4804.
- (12) Puzzarini, C.; Barone, V. Extending the Molecular Size in Accurate Quantum-Chemical Calculations: the Equilibrium Structure and Spectroscopic Properties of Uracil. *Phys. Chem. Chem. Phys.* **2011**, *13*, 7189–7197.
- (13) Puzzarini, C.; Biczysko, M.; Barone, V.; Peña, I.; Cabezas, C.; Alonso, J. L. Accurate molecular structure and spectroscopic properties of nucleobases: a combined computational-microwave investigation of 2-thiouracil as a case study. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16965–16975.
- (14) Puzzarini, C.; Biczysko, M.; Barone, V.; Largo, L.; Peña, I.; Cabezas, C.; Alonso, J. L. Accurate Characterization of the Peptide Linkage in the Gas Phase: a Joint Quantum-chemical and Rotational Spectroscopy Study of the Glycine Dipeptide Analogue. *J. Phys. Chem. Lett.* **2014**, *5*, 534–540.
- (15) Shavitt, I.; Bartlett, R. J. *Many-Body Methods in Chemistry and Physics*, 1st ed.; Cambridge University Press, 2009.
- (16) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.
- (17) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.

- (18) Grimme, S. Semiempirical Hybrid Density Functional with Perturbative Second-order Correlation. *J. Chem. Phys.* **2006**, *124*, 034108.
- (19) Schwabe, T.; Grimme, S. Towards chemical accuracy for the thermodynamics of large molecules: new hybrid density functionals including non-local correlation effects. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4398–4401.
- (20) Grimme, S. Density Functional Theory with London Dispersion Corrections. *WIREs Comput. Mol. Sci.* **2011**, *1*, 211–228.
- (21) Caldeweyher, E.; Bannwarth, C.; Grimme, S. Extension of the D3 dispersion coefficient model. *J. Chem. Phys.* **2017**, *147*, 034112.
- (22) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
- (23) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.
- (24) Papajak, E.; Leverentz, H. R.; Zheng, J.; Truhlar, D. G. Efficient Diffuse Basis Sets: cc-pVxZ+ and maug-cc-pVxZ. *J. Chem. Theory Comput.* **2009**, *5*, 1197–1202.
- (25) Fornaro, T.; Biczysko, M.; Bloino, J.; Barone, V. Reliable vibrational wavenumbers for CO and NH stretchings of isolated and hydrogen-bonded nucleic acid bases. *Phys. Chem. Chem. Phys.* **2016**, *18*, 8479–8490.
- (26) Li, W.; Spada, L.; Tasinato, N.; Rampino, S.; Evangelisti, L.; Gualandi, A.; Cozzi, P. G.; Melandri, S.; Barone, V.; Puzzarini, C. Theory Meets Experiment for Noncovalent Complexes: The Puzzling Case of Pnictogen Interactions. *Angew. Chem. Int. Ed.* **2018**, *57*, 13853–13857.
- (27) Biczysko, M.; Panek, P.; Scalmani, G.; Bloino, J.; Barone, V. Harmonic and Anharmonic Vibrational Frequency Calculations with the Double-Hybrid B2PLYP

- Method: Analytic Second Derivatives and Benchmark Studies. *J. Chem. Theory Comput.* **2010**, *6*, 2115–2125.
- (28) Penocchio, E.; Piccardo, M.; Barone, V. Semi-Experimental Equilibrium Structures for Building Blocks of Organic and Biological Molecules: the B2PLYP Route. *J. Chem. Theory Comput.* **2015**, *11*, 4689–4707.
- (29) Barone, V.; Biczysko, M.; Bloino, J.; Puzzarini, C. Accurate structure, thermodynamic and spectroscopic parameters from CC and CC/DFT schemes: the challenge of the conformational equilibrium in glycine. *Phys. Chem. Chem. Phys.* **2013**, *15*, 10094–10111.
- (30) Puzzarini, C.; Biczysko, M. Microsolvation of 2-Thiouracil: Molecular Structure and Spectroscopic Parameters of the Thiouracil-Water Complex. *J. Phys. Chem. A* **2015**, *119*, 5386–5395.
- (31) Obenchain, D. A.; Spada, L.; Alessandrini, S.; Rampino, S.; Herbers, S.; Tasinato, N.; Mendolicchio, M.; Kraus, P.; Gauss, J.; Puzzarini, C.; et al., Unveiling the Sulfur-Sulfur bridge: accurate structural and energetic characterization of a homo chalcogen inter-molecular bond. *Angew. Chem. Int. Ed.* **2018**, *57*, 15822–15826.
- (32) Spada, L.; Tasinato, N.; Bosi, G.; Vazart, F.; Barone, V.; Puzzarini, C. On the competition between weak $\text{OH} \cdots \text{F}$ and $\text{CH} \cdots \text{F}$ hydrogen bonds, in cooperation with $\text{CH} \cdots \text{O}$ contacts, in the difluoromethane tert-butyl alcohol cluster. *J. Mol. Spectrosc.* **2017**, *337*, 90–95.
- (33) Spada, L.; Tasinato, N.; Vazart, F.; Barone, V.; Caminati, W.; Puzzarini, C. Non-covalent Interactions and Internal Dynamics in Pyridine-Ammonia: A Combined Quantum-Chemical and Microwave Spectroscopy Study. *Chem. Eur. J.* **2017**, *23*, 4876–4883.
- (34) Kendall, R.; Dunning Jr., T.; Harrison, R. Electron Affinities of the First-row Atoms

- Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96*, 6769–6806.
- (35) 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.
- (36) Dutta, N. N.; Patkowski, K. Improving “Silver-Standard” Benchmark Interaction Energies with Bond Functions. *J. Chem. Theory Comput.* **2018**, *14*, 3053–3070.
- (37) Tew, D. P.; Klopper, W.; Neiss, C.; Hättig, C. Quintuple- ζ quality coupled-cluster correlation energies with triple- ζ basis sets. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1921–1930.
- (38) Hättig, C.; Tew, D. P.; Köhn, A. Communications: Accurate and efficient approximations to explicitly correlated coupled-cluster singles and doubles, CCSD-F12. *J. Chem. Phys.* **2010**, *132*, 231102.
- (39) Kodrycka, M.; Patkowski, K. Platinum, gold, and silver standards of intermolecular interaction energy calculations. *J. Chem. Phys.* **2019**, *151*, 070901.
- (40) Schurkus, H. F.; Luenser, A.; Ochsenfeld, C. Communication: Almost error-free resolution-of-the-identity correlation methods by null space removal of the particle-hole interactions. *J. Chem. Phys.* **2017**, *146*, 211106.
- (41) Nagy, P. R.; Kállay, M. Approaching the Basis Set Limit of CCSD(T) Energies for Large Molecules with Local Natural Orbital Coupled-Cluster Methods. *J. Chem. Theory Comput.* **2019**, *in press*, DOI:10.1021/acs.jctc.9b00511.
- (42) 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.

- (43) Řezáč, J.; Hobza, P. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. *Chem. Rev.* **2016**, *116*, 5038–5071.
- (44) Řezáč, J.; Dubecký, M.; Jurečka, P.; Hobza, P. Extensions and applications of the A24 data set of accurate interaction energies. *Phys. Chem. Chem. Phys.* **2015**, *17*, 19268–19277.
- (45) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; et al., Gaussian 16 Revision A.03. 2016; Gaussian Inc. Wallingford CT.
- (46) Papajak, E.; Zheng, J.; Xu, X.; Leverentz, H. R.; Truhlar, D. G. Perspectives on Basis Sets Beautiful: Seasonal Plantings of Diffuse Basis Functions. *J. Chem. Theory Comput.* **2011**, *7*, 3027–3034.
- (47) Dunning, T. H.; Peterson, K. A.; Wilson, A. K. Gaussian basis sets for use in correlated molecular calculations. X. The atoms aluminum through argon revisited. *J. Chem. Phys.* **2001**, *114*, 9244–9253.
- (48) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-Set Convergence of Correlated Calculations on Water. *J. Chem. Phys.* **1997**, *106*, 9639.
- (49) Burns, L. A.; Marshall, M. S.; Sherrill, C. D. Comparing Counterpoise-Corrected, Uncorrected, and Averaged Binding Energies for Benchmarking Noncovalent Interactions. *J. Chem. Theory Comput.* **2014**, *10*, 49–57.
- (50) Bomble, Y. J.; Stanton, J. F.; Kállay, M.; Gauss, J. Coupled-cluster methods including noniterative corrections for quadruple excitations. *J. Chem. Phys.* **2005**, *123*, 054101.
- (51) De Jong, W. A.; Harrison, R. J.; Dixon, D. A. Parallel Douglas–Kroll energy and gradients in NWChem: estimating scalar relativistic effects using Douglas–Kroll contracted basis sets. *J. Chem. Phys.* **2001**, *114*, 48–53.

- (52) Reiher, M.; Wolf, A. Exact decoupling of the Dirac Hamiltonian. II. The generalized Douglas-Kroll-Hess transformation up to arbitrary order. *J. Chem. Phys.* **2004**, *121*, 10945–10956.
- (53) Feller, D. The Use of Systematic Sequences of Wave Functions for Estimating the Complete Basis Set, Full Configuration Interaction Limit in Water. *J. Chem. Phys.* **1993**, *98*, 7059–7071.
- (54) Peterson, K. A.; Dunning Jr., T. H. Accurate correlation consistent basis sets for molecular core-valence correlation effects: The second row atoms Al-Ar, and the first row atoms B-Ne revisited. *J. Chem. Phys.* **2002**, *117*, 10548–10560.
- (55) Kesharwani, M. K.; Manna, D.; Sylvetsky, N.; Martin, J. M. L. The X4010 Halogen Bonding Benchmark Revisited: Surprising Importance of (n1)d Subvalence Correlation. *J. Phys. Chem. A* **2018**, *122*, 2184–2197.
- (56) Brauer, B.; Kesharwani, M. K.; Kozuch, S.; Martin, J. M. The S66x8 benchmark for noncovalent interactions revisited: Explicitly correlated ab initio methods and density functional theory. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20905–20925.
- (57) Thorpe, J. H.; Lopez, C. A.; Nguyen, T. L.; Baraban, J. H.; Bross, D. H.; Ruscic, B.; Stanton, J. F. High-accuracy extrapolated ab initio thermochemistry. IV. A modified recipe for computational efficiency. *J. Chem. Phys.* **2019**, *150*, 224102.
- (58) Coriani, S.; Marchesan, D.; Gauss, J.; Hättig, C.; Helgaker, T.; Jørgensen, P. The accuracy of ab initio molecular geometries for systems containing second-row atoms. *J. Chem. Phys.* **2005**, *123*, 184107.
- (59) Goldey, M.; Dutoi, A.; Head-Gordon, M. Attenuated second-order MøllerPlesset perturbation theory: performance within the aug-cc-pVTZ basis. *Phys. Chem. Chem. Phys.* **2013**, *15*, 15869–15875.

- (60) Stanton, J. F.; Gauss, J.; Harding, M. E.; Szalay, P. G. CFour. A quantum chemical program package. 2016; with contributions from A. A. Auer, R. J. Bartlett et al. 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>.
- (61) Heckert, M.; Kállay, M.; Tew, D. P.; Klopper, W.; Gauss, J. Basis-Set Extrapolation Techniques for the Accurate Calculation of Molecular Equilibrium Geometries Using Coupled-Cluster Theory. *J. Chem. Phys.* **2006**, *125*, 044108.
- (62) Puzzarini, C.; Stanton, J. F.; Gauss, J. Quantum-chemical calculation of spectroscopic parameters for rotational spectroscopy. *Int. Rev. Phys. Chem.* **2010**, *29*, 273–367.
- (63) Fornaro, T.; Biczysko, M.; Monti, S.; Barone, V. Dispersion Corrected DFT Approaches for Anharmonic Vibrational Frequency Calculations: Nucleobases and their Dimers. *Phys. Chem. Chem. Phys.* **2014**, *16*, 10112–10128.
- (64) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 4285–4292.
- (65) Oliveira, V.; Kraka, E. Systematic Coupled Cluster Study of Noncovalent Interactions Involving Halogens, Chalcogens, and Pnicogens. *J. Phys. Chem. A* **2017**, *121*, 9544–9556.
- (66) Alessandrini, S.; Puzzarini, C. Structural and Energetic Characterization of Prebiotic Molecules: The Case Study of Formamide and Its Dimer. *J. Phys. Chem. A* **2016**, *120*, 5257–5263.
- (67) Frey, J. A.; Leutwyler, S. An ab Initio Benchmark Study of Hydrogen Bonded Formamide Dimers. *J. Phys. Chem. A* **2006**, *110*, 12512–12518.

- (68) Řezáč, J.; Riley, K. E.; Hobza, P. Extensions of the S66 Data Set: More Accurate Interaction Energies and Angular-Displaced Nonequilibrium Geometries. *J. Chem. Theory. Comput.* **2011**, *7*, 3466–3470.
- (69) Sylvetsky, N.; Kesharwani, M. K.; Martin, J. M. L. The aug-cc-pVnZ-F12 basis set family: Correlation consistent basis sets for explicitly correlated benchmark calculations on anions and noncovalent complexes. *J. Chem. Phys.* **2017**, *147*, 134106.
- (70) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory. Comput.* **2011**, *7*, 2427–2438.
- (71) Řezáč, J.; Riley, K. E.; Hobza, P. Erratum to S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory. Comput.* **2014**, *10*, 1359–1360.
- (72) Puzzarini, C. Accurate thermochemistry and spectroscopy of the oxygen-protonated sulfur dioxide isomers. *Phys. Chem. Chem. Phys.* **2011**, *13*, 21319–21327.
- (73) Puzzarini, C.; Penocchio, E.; Biczysko, M.; Barone, V. Molecular Structure and Spectroscopic Signatures of Acrolein: Theory Meets Experiment. *J. Phys. Chem. A* **2014**, *118*, 6648–6656.
- (74) Barone, V.; Biczysko, M.; Bloino, J.; Cimino, P.; Penocchio, E.; Puzzarini, C. CC/DFT Route toward Accurate Structures and Spectroscopic Features for Observed and Elusive Conformers of Flexible Molecules: Pyruvic Acid as a Case Study. *J. Chem. Theory Comput.* **2015**, *11*, 4342–4363.
- (75) Peterson, K. A.; Adler, T. B.; Werner, H.-J. Systematically convergent basis sets for explicitly correlated wavefunctions: The atoms H, He, B-Ne, and Al-Ar. *J. Chem. Phys.* **2008**, *128*, 084102.

- (76) Spackman, P. R.; Jayatilaka, D.; Karton, A. Basis set convergence of CCSD(T) equilibrium geometries using a large and diverse set of molecular structures. *J. Chem. Phys.* **2016**, *145*, 104101.
- (77) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371.
- (78) Manna, D.; Kesharwani, M. K.; Sylvetsky, N.; Martin, J. M. L. Conventional and Explicitly Correlated ab Initio Benchmark Study on Water Clusters: Revision of the BEGDB and WATER27 Data Sets. *J. Chem. Theory Comput.* **2017**, *13*, 3136–3152.