

The Performance of Dunning, Jensen and Karlsruhe Basis Sets on Computing Relative Energies and Geometries

Karl N. Kirschner[†], Dirk Reith[‡], and Wolfgang Heiden[†]

[†]Department of Computer Science, University of Applied Sciences Bonn-Rhein-Sieg,
Grantham-Allee 20, 53757 Sankt Augustin, Germany

[‡]Department of Electrical Engineering, Mechanical Engineering and Technical
Journalism, University of Applied Sciences Bonn-Rhein-Sieg, Grantham-Allee 20,
53757 Sankt Augustin, Germany

December 20, 2019

Abstract

In an effort to assist researchers in choosing basis sets for quantum mechanical modeling of molecules (i.e. balancing calculation cost versus desired accuracy), we present a systematic study on the accuracy of computed conformational relative energies and their geometries in comparison to MP2/CBS and MP2/AV5Z data, respectively. In order to do so, we introduce a new nomenclature to unambiguously indicate how a CBS extrapolation was computed. Nineteen minima and transition states of buta-1,3-diene, propan-2-ol and the water dimer were optimized using forty-five different basis sets. Specifically, this includes one Pople (i.e. 6-31G(d)), eight Dunning (i.e. VXZ and AVXZ, X=2-5), twenty-five Jensen (i.e. pc-n, pcseg-n, aug-pcseg-n, pcSseg-n and aug-pcSseg-n, n=0-4) and nine Karlsruhe (e.g. def2-SV(P), def2-QZVPPD) basis sets. The molecules were chosen to represent both common and electronically diverse molecular systems. In comparison to MP2/CBS relative energies computed using the largest Jensen basis sets (i.e. n=2,3,4), the use of smaller sizes (n=0,1,2 and n=1,2,3) provides results that are within 0.11–0.24 and 0.09–0.16 kcal·mol⁻¹. To practically guide researchers in their basis set choice, an equation is introduced that ranks basis sets based on a user-defined balance between their accuracy and calculation cost. Furthermore, we explain why the aug-pcseg-2, def2-TZVPPD and def2-TZVP basis sets are very suitable choices to balance speed and accuracy.

1 Introduction

In projects involving quantum mechanics (QM) calculations a compromise must often be reached between the computational cost – theory level and basis set size – and the desired precision of the resulting data. The Dunning series of correlation-consistent polarized basis sets (i.e. VXZ and AVXZ)^{3,25,45,47} are frequently used and considered among the best for generating accurate data. However, they are also among the most costly to use. In an effort to reduce calculation cost while maintaining good data accuracy, alternative basis sets have been proposed. Two such basis sets are Jensen’s polarization-consistent (e.g. pc-n, aug-pcseg-n)^{16–21} and response property-optimized Karlsruhe (i.e. def2)^{13,32,34,35,41,43} basis sets.

The Jensen basis sets were developed for use in Hartree-Fock (HF) and density functional theory (DFT) calculations. These basis sets were designed to converge faster than Dunning

basis sets without the loss of accuracy in certain observables (e.g. absolute energies, electron affinity, dipole moment, polarizability). Two groups have investigated how Jensen’s pc-n basis set family perform for correlated wave function (i.e. MP2 and CCSD(T)) optimizations. For isolated molecules, Kupka and Lim showed that MP2 and CCSD(T) calculations using pc-n basis sets performed favorably in comparison to Dunning basis sets and at a reduced computational cost.²⁹ More recently, Kupka and coworkers analyzed the aug-pc-n basis sets for predicting the water dimer minimum’s interaction energy using CCSD(T) theory, and their convergence to the complete basis set (CBS) limit using different fitting approaches²⁸. ElSohly and Tschumper extended this work by investigating weakly-bonded systems. Using modified pc-n basis sets that include diffuse functions that better model weak forces; they found that the basis sets could reproduce benchmark values, but at a higher cost than the Dunning basis sets.⁴

The early Karlsruhe basis sets (e.g. SVP, QZV) were developed to reproduce energetic target data^{34,35,41,43}. The more recent basis sets that include diffuse functions were optimized for reproducing dipole polarizabilities computed by HF, DFT, MP2 and their corresponding post-SCF resolution-of-the-identity (RI) calculations.^{13,32} The most recent of these basis sets were shown to calculate interaction energies (using the S22 data set²³) electron affinity, dipole moments and polarizabilities of atoms and relatively inflexible molecules (e.g. He₂, Be(CH₃)₂, Mo(CO)₆) that are competitive with the larger Dunning basis set.¹³ For organic compounds (e.g. 2-fluoroethylamine, aspirin, butane-1,4-diol), def2 basis sets have been used to investigate geometries and conformational energies using DFT and post-SCF calculations, some of which made use of the RI approximation.^{8,15,27} In particular, one study computed the relative energy profile for a torsion rotation that included transition states (TS), but did not compare the resulting data to higher theories levels.¹⁵

To our knowledge, there have been two studies that include a comparison between Jensen and Karlsruhe basis sets. Johansson and Olsen computed the transition states for rotating about the central bond of biphenyl using pc-n (n=1-4), aug-pc-n (n=1-3), def2-SVP, def2-TZVPP, and def2-QZVPP basis sets using HF and DFT theories.²² In a larger study, Witte et al examined the basis sets’ performances when used with DFT theory for reproducing intermolecular interactions of the S22 data set.⁴⁶ They concluded that once basis set superposition error has corrected for, def2-SVPD performs well to similarly sized Dunning and Jensen basis sets, while def2-QZVPD is a practical alternative to pc-4.

Herein, we present an extension of the above papers by comparing an extensive collection of Jensen, Karlsruhe and Dunning basis sets for computing MP2 geometries and relative conformational energies of minima and transition states. The choice of MP2 theory was motivated by the fact that it is a well established theory whose behavior is systematic as a function of basis sets size (i.e. variational-like). While modern DFT methods can offer better performance at a reduced cost, we feel that the rapid progress and development of new functionals makes the longevity of a given theory difficult to predict. The results herein will be of interest to researchers who model larger and more flexible molecules (e.g. carbohydrates, drugs), whose potential energy surfaces are not easily explored using more rigorous theories (e.g. CCSD(T)), but want to include some level of electron correlation into their modeling.

The segmented contracted basis sets by Jensen (i.e. pcSseg-n and aug-pcSseg-n) were specifically created to calculate nuclear magnetic shielding constants²¹. Consequently, one might not expect them to provide highly accurate relative energies and geometries. Never-the-less, we include them into this study in hopes that they might fortuitously provide good results at a reasonable expense.

The target benchmark data includes MP2/AV5Z optimized geometries and MP2 CBS relative energies for buta-1,3-diene, propan-2-ol and the water dimer. These molecules were chosen to represent diverse organic systems (i.e. containing H, C and O) that comprise significant electron delocalization, non-negligible dipole moments and hydrogen bonds. By including both minima and transition states, we further enrich the variety of electronic configurations that are studied. Finally, we also compare two schemes for computing CBS energies and introduce

an equation that ranks theory levels according to their calculation cost, accuracy and a user specified weighting factor.

2 Methods

2.1 Optimizations and Frequency Calculations

PyMol² was used to create the buta-1,3-diene and propanol’s initial conformations, which were then fully optimized using MP2/AV5Z theory level. The resulting geometries were subsequently used as input for the other theory levels herein. For the water dimer, all input configurations were taken from supplementary information of Tschumper et al.,³⁶ which were computed at the CCSD(T)/TZ2P(f,d)+dif theory level. To simplify the writing, water configurations will be referred to as conformations. To characterize the structures’ positions on their potential energy surface, MP2/AVTZ//MP2/AVTZ frequency calculations were computed using finite-differences of gradients.

In all MP2 calculations, only the valence electrons were correlated, and the RI approximation^{24,37} with auxiliary basis sets^{10,40,42–44} were used (see SI material for full listing). Exceptions to this include calculations involving correlation-consistent polarized core-valence basis sets (i.e. ACVXZ) and two CCSD(T)/TZ2P(f,d)+dif calculation (i.e. water dimer **2** and **7**), where core electrons were included into the correlation (see Table S1 and S2). All optimizations were performed using C₁ molecular symmetry unless otherwise stated, with a maximum force of 1.5E-5, a root-mean-squared (RMS) force of 1.0E-5, a maximum displacement of 6.0E-5 and an RMS displacement of 4.0E-5 convergence criteria specified.

2.2 Complete Basis Set Extrapolation

In practice, extrapolating energies to the complete basis set limit can be done using two different schemes. These differ by the choice in the molecular geometry used for determining the component energy computed by each basis set. For clarification, a generalized representation for computing a CBS energy is:

$$E_{\text{CBS}}^{\text{THEORY}}(R^{\text{source}}) = E_{\text{CBS}}^{\text{SCF}}(R^{\text{source}}) + E_{\text{CBS}}^{\text{CORR}}(R^{\text{source}}) \quad (1)$$

where $E_{\text{CBS}}^{\text{THEORY}}$ is the CBS limit of the total energy computed using a given theory level sequence, $E_{\text{CBS}}^{\text{SCF}}$ is the CBS limit of the Hartree-Fock energy, $E_{\text{CBS}}^{\text{CORR}}$ is the CBS limit of the correlation energy and R^{source} indicates the source of the geometries used in the calculations.

In Scheme 1, a single source (i.e. reference) geometry is used for computing the required energies for an extrapolation. Generally, the geometry comes from either experimental spectroscopy or a high-level ab initio optimization. This scheme reduces the overall calculational expense since multiple geometry optimizations are not performed. An example for noting this scheme that uses an MP2/AV5Z optimized geometry is:

$$E_{\text{CBS}}^{\text{MP2/AV[D,T,Q//T,Q]Z}}(R^{\text{MP2/AV5Z}}) = E_{\text{CBS}}^{\text{SCF}}(R^{\text{MP2/AV5Z}}) + E_{\text{CBS}}^{\text{CORR}}(R^{\text{MP2/AV5Z}}) \quad (2)$$

where the numbers of the sequential basis sets used are given within the square brackets. The double slash notation within these brackets signify that three and two basis sets (see below) were used to obtain the CBS limits of the Hartree-Fock and correlation energies, respectively.

In Scheme 2, multiple geometries are used in computing the component energies. Specifically, these are geometries that were optimized using each basis set specified for the extrapolation workflow. An example for an extrapolation that uses MP2/AVDZ, MP2/AVTZ and MP2/AVQZ optimized geometries is:

$$E_{\text{CBS}}^{\text{MP2/AV[D,T,Q//T,Q]Z}}(R^{\text{MP2/AV[D,T,Q]Z}}) = E_{\text{CBS}}^{\text{SCF}}(R^{\text{MP2/AV[D,T,Q]Z}}) + E_{\text{CBS}}^{\text{CORR}}(R^{\text{MP2/AV[T,Q]Z}}) \quad (3)$$

One advantage of this scheme is that it can enable a CBS limit to be obtained for internal coordinates.¹¹

In all CBS limit calculations, the $E_{\text{CBS}}^{\text{SCF}}$ was extrapolated to using three sequential basis sets and Feller’s exponential equation.^{5,6} Conversely, the $E_{\text{CBS}}^{\text{CORR}}$ was extrapolated to using the two largest basis sets within the sequence and Helgaker’s power equation with $\alpha = 3.0$.^{9,12} The difference in the basis set involvement for computing $E_{\text{CBS}}^{\text{SCF}}$ and $E_{\text{CBS}}^{\text{CORR}}$ gives rise to the notation within the square brackets (e.g. MP2/AV[D,T,Q//T,Q]). In the current study, the individual component E^{SCF} values were extracted from the MP2 optimizations rather than performing additional Hartree-Fock optimizations. Additional details of the equations used are given in supplementary information.

For an unambiguous shorthand nomenclature of how a CBS limit was computed one can extend the double slash notation (e.g. MP2/AVDZ//MP2/AV5Z) to include the numbers of the basis set triads (e.g. [T,Q,5]) used in the extrapolation workflow. For both schemes, this can be illustrated in the following examples:

Scheme 1 MP2/AV[T,Q,5//Q,5]Z//MP2/AV5Z

Scheme 2 MP2/AV[T,Q,5//Q,5]Z//MP2/AV[T,Q,5]Z

2.3 Indexing of Jensen Basis Sets

In order to make comparisons as easy to read and understandable as possible, the Dunning and Jensen basis set nomenclature needed be adjusted such that the indexing becomes consistent between them. Therefore, we chose an index range from 1–5 instead of 0–4, which also allows for clear CBS extrapolations. Consequently, the Jensen index will be adjusted by +1 (i.e. $X = n + 1$, $n=0,1,2,3,4$) throughout this work, and will be referred to as such (e.g. pc-X, $X=1,2,3,4,5$) unless specifically indicated otherwise.

2.4 Software and Analysis

All QM calculations were done using Psi4 (v. 1.1a2.dev170)³⁰. Root-mean-squared deviations (RMSD) were computed using pytraj, a Python package that uses the cpptraj program.³³ Python3^{1,38} was used to execute pytraj, perform statistical analysis and generate plots using Matplotlib (v. 1.3.1)¹⁴ and Seaborn (v. 0.7.1).³⁹ A total of 860 full optimizations and 80 single-point calculations were used in the analysis. All averages presented were computed using absolute values (i.e. $\frac{|x_i - x_j|}{n}$) and exclude the global minima (i.e. 0.000 kcal·mol⁻¹) for each molecular system. Data normalization was done using max-min feature scaling implemented in the Python scikit-learn library.³¹

3 Results and Discussion

A total of four, seven and eight conformations of buta-1,3-diene, propan-2-ol and the water dimer were optimized by nearly all basis sets used herein, and are shown in Figure 1. The specific instances where an optimization was unable to converge include water dimer **4** (MP2/aug-pcSseg-2, MP2/def2-SVPD) and **5** (MP2/aug-pcSseg-1 and MP2/aug-pcSseg-2, MP2/def2-SVPD). None of the theory levels employed were able to converge to the water dimers **2** and **7** that were reported in reference 36.

In an attempt to reproduce the water dimers **2** and **7** as stationary points, we performed full optimizations at the CCSD(T)/TZ2P(f,d)+diff theory level (i.e. the same theory level

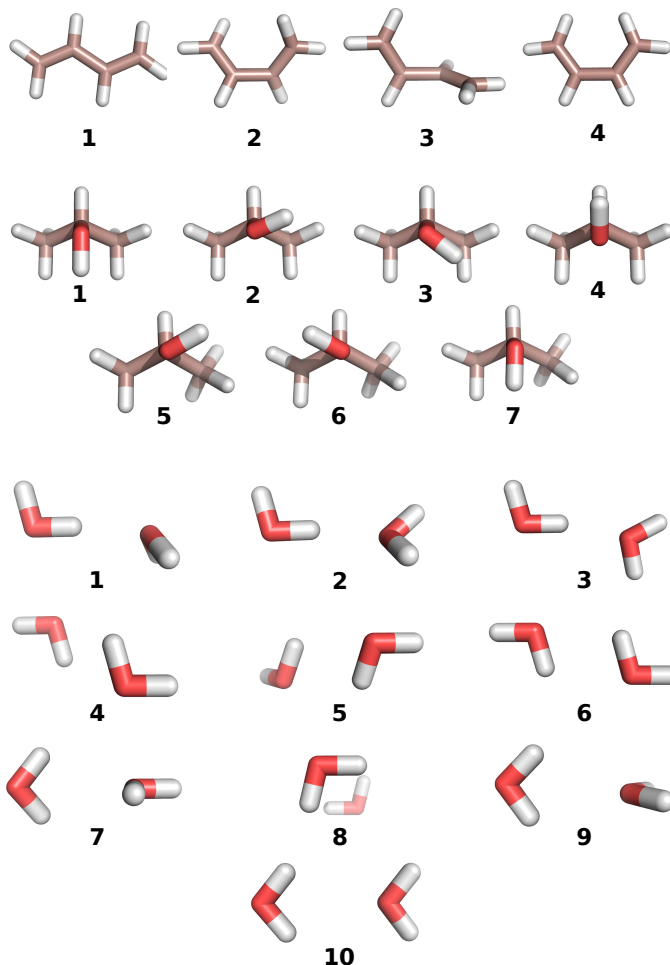


Figure 1: The MP2/AV5Z//MP2/AV5Z optimized conformations for buta-1,3-diene, propan-2-ol and the water dimer. The water dimer conformation numbers are consistent with those used in reference 36.

reported in reference 36), using our specified convergence criteria (see Methods), enforcing C_s symmetry for **7** and C_1 for **2**, and correlating all electrons. Using this theory level, we were able to optimize water dimer **7**. A full optimization at MP2/TZ2P(f,d)+diff (valence electrons correlated) also found **7** as a stationary point. Finally, we added the three diffuse function (i.e. $\alpha_s(\text{H})=0.03016$, $\alpha_s(\text{O})=0.08993$, and $\alpha_p(\text{O})=0.05840$.) to the VTZ basis set (i.e. VTZ+diff), which also resulted in being able to optimize **7** using MP2 theory. An MP2/VTZ optimization of **7** was also performed with the molecular symmetry of C_s enforced, which resulted in the structure optimizing to water dimer **1**. Thus for water dimer **7**, the presence of these diffuse functions is critical for its characterization as a stationary point. In all of these additional calculations, the water dimer **2** remained elusive. Additional information can be found in SI material.

3.1 CBS Extrapolation Using Schemes 1 and 2

As briefly mentioned in the introduction, one goal of this study is to evaluate the impact of the more severe approximations made in Scheme 1 in comparison to the more rigorous approach of Scheme 2. Due to their wide-spread and long usage, the AVXZ ($X=\text{D,T,Q,5}$) basis set family was used for this part of the study. Table 1 provides the relative energies for all conformations investigated, computed using both CBS schemes. Included in this table are previously published CCSD(T) relative energies, computed using a variety of approaches,^{7,26,36} which are the current

state-of-the-art results.

On average, the use of Scheme 1 and 2 results in values that are within $0.002 \text{ kcal}\cdot\text{mol}^{-1}$ of one another when using [D,T,Q] or [T,Q,5] basis sets triads. The largest deviation observed is $|0.007| \text{ kcal}\cdot\text{mol}^{-1}$, computed using the [T,Q,5] triad for water dimer **6**. Thus, for the augmented Dunning basis sets and the molecules studied here, the two extrapolation schemes yield nearly equivalent MP2/CBS values results. Note that this result is based on the fact that a very reliable geometry (i.e. MP2/AV5Z//MP2/AV5Z) was used for Scheme 1. It is reasonable to assume that if one would alternatively use a less reliable geometry (e.g. computed at a lower theory), then implementing Scheme 1 and 2 would likely yield less equivalent results.

Within each scheme, increasing the basis sets triad used (i.e. [D,T,Q] \rightarrow [T,Q,5]) alters the relative energies by values ranging from -0.030 (water dimer **6**) to $+0.062$ (buta-1,3-diene **3**) $\text{kcal}\cdot\text{mol}^{-1}$. An average absolute difference is computed to be 0.015 and $0.014 \text{ kcal}\cdot\text{mol}^{-1}$ for Scheme 1 and 2, respectively. In comparison to the small differences observed when changing the CBS scheme used, the size of the basis sets triad has a noticeable impact on the resulting MP2/CBS relative energy values.

3.2 CBS Trends

3.2.1 Electronic Energies

Observing data trends and why they occur improves our understanding of basis set behavior and helps to identify questionable data in future calculations. Both the Dunning and Jensen basis set families allow for straightforward CBS limit extrapolations. As expected due to the variational-like behavior of MP2, all MP2 electronic energies (Table S3) asymptotically approach a CBS limit (Figures S1–S3). Using these energies, the CBS SCF and MP2 energies were computed and are graphically shown as a function of their triad sequence in Figures S4–S6. For both Dunning basis set families (i.e. VXZ, AVXZ), more negative MP2/CBS energies are always computed when using a larger basis set triad. This trend is also seen for the pcseg-X and aug-pcseg-X Jensen basis set families.

However, exceptions occur for pc-X, pcSseg-X and aug-pcSseg-X where the use of a [X=3,4,5] triad results in more positive MP2/CBS energies and is dependent upon the chemical system. As a mathematical property of the extrapolation, the MP2/CBS electronic energy computed using a specific triad results in a more negative value in comparison to the component energies that were used in the extrapolation process (e.g. $\text{AVDZ} > \text{AVTZ} > \text{AVQZ} > \text{A}[\text{D,T,Q}]\text{Z}$). Therefore, the fact that some [X=3,4,5] CBS energies are more positive than those computed by the [X=2,3,4] triad is a consequence of how their respective component energies are distributed (i.e. the shape of the curve formed by the three points). And recall that the pcSseg basis sets were optimized for nuclear magnetic shielding constants, not for determining energies.

Decomposing the MP2/CBS energies into CBS SCF and correlation energies reveal that the SCF component consistently becomes more positive as the triad increases (e.g. $[2,3,4] \rightarrow [3,4,5]$), with the single exception of the VXZ basis set family. Conversely, the correlation energies do not follow a consistent trend. By examining Figures S4–S6, it is clear that the MP2/CBS energy trends follow those of the CBS correlation energies as the triad sequence increases.

3.2.2 Relative Energies

Considering that multiple CBS values can be computed for a given basis set family — for example, a pc-X CBS energy can be computed using [X=1,2,3], [X=2,3,4], or [X=3,4,5] triad — we also examined how the extrapolated relative energies computed using the smaller triads compare to those computed using the largest triad. Here we make the assumption that the CBS values computed using the largest triad are the most accurate within a given family. Figures 2–4 plot the MP2/CBS relative energies (Scheme 2, Table S4) for the three systems studied as a function of the Dunning and Jensen triad sequence. What is reassuring, in comparison to the

MP2/CBS electronic energies that show significant changes as the triad size increases (Figures S4–S6), is that their computed relative energies show little variation. Consequently, while the ultimate goal here is to obtain the best CBS relative energies through the use of the largest size triad, one can get reasonably close values through the use of a smaller triad (e.g. [1,2,3]) as elaborated upon in the following paragraphs.

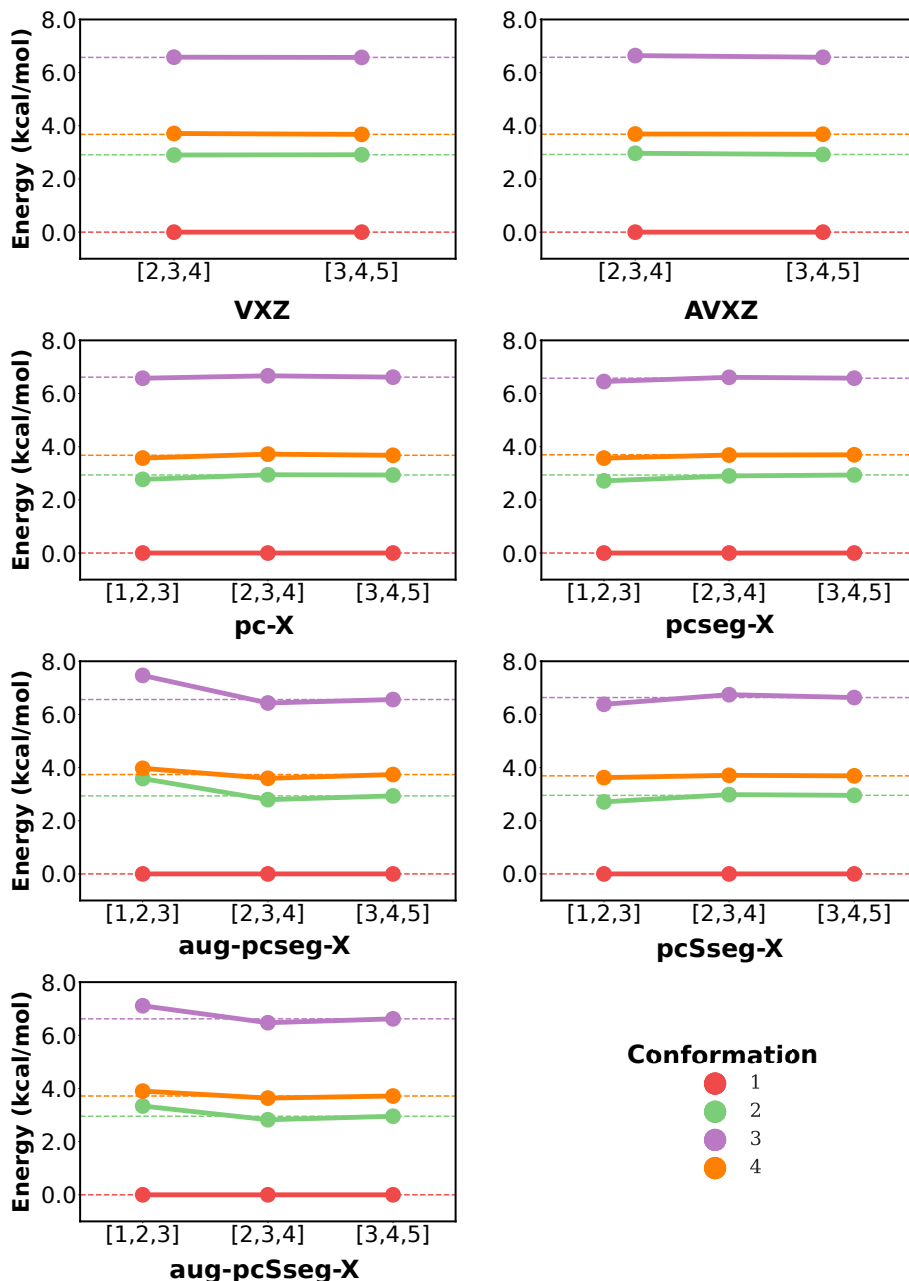


Figure 2: The MP2/CBS relative energies (Scheme 2) computed from the possible sequential Dunning and Jensen basis set triad combination (e.g. [2,3,4]) for buta-1,3-diene conformations. The horizontal dashed lines are added as visual guides and indicate the CBS limit computed using the [3,4,5] triad within the each basis set family.

The mean absolute errors of the MP2/CBS relative energies computed using the [1,2,3] and [2,3,4] triads, relative to a basis set family’s largest [3,4,5] triad extrapolated value, are given in Table 2. The largest error ($0.243 \text{ kcal}\cdot\text{mol}^{-1}$) is seen for pc-[X=1,2,3], while the smallest error ($0.015 \text{ kcal}\cdot\text{mol}^{-1}$) is seen for AV[D,T,Q]Z. Both Dunning basis set families (i.e. VXZ and AVXZ) generate [2,3,4] extrapolated values that are more accurate than those computed using the equivalent Jensen basis sets.

On average, the MP2/CBS relative energies computed using the $[X=1,2,3]$ Jensen triads result in values that are within $0.11\text{--}0.24\text{ kcal}\cdot\text{mol}^{-1}$ of those computed using their substantially larger $[X=3,4,5]$ triad (Table 2). Similarly, the use of $[X=2,3,4]$ triads provides improved agreement with values within $0.09\text{--}0.16\text{ kcal}\cdot\text{mol}^{-1}$ of the $[X=3,4,5]$ CBS computed limit. These values are similar to the ones reported by Kupka and coworkers’ investigation of the water dimer minimum using CCSD(T) theory level²⁸. One would expect these agreements to become slightly worse if further compared to extrapolated values computed using $[X=4,5,6]$ and larger triads.

Highlighting specific results, the aug-pcseg- $[X=1,2,3]$ triad CBS extrapolations perform the worst for buta-1,3-diene, with a mean absolute differences of $0.598\text{ kcal}\cdot\text{mol}^{-1}$. A significant improvement occurs when computing the aug-pcseg- $[X=2,3,4]$ extrapolations, with a resulting value of $0.138\text{ kcal}\cdot\text{mol}^{-1}$ obtained. For buta-1,3-diene, this improvement is mirrored within each of the Jensen basis set families. However, this is not consistently seen for propan-2-ol or the water dimer. Notable are the pcseg- X , aug-pcseg- X and pcSseg- X families where their $[X=1,2,4]$ triad results in better average agreements to $[X=3,4,5]$ triad CBS extrapolated relative energies for both propan-2-ol and the water dimer.

Finally, as seen in previous studies³⁶ the CBS’s SCF component is largely responsible for the general distribution of conformer stability (see Figures S7–S9). Including correlation energy slightly modulates the separation and order of the relative conformer stability.

3.2.3 Including Core Electron Correlation

The correlation of the valence electrons only introduces some error into the resulting relative energies. Table 3 provides the MP2/CBS relative energies computed using both the standard augmented valence Dunning basis sets (i.e. AVXZ) and the augmented Dunning basis sets that were specifically created for including core electron correlation into a calculation (i.e. correlation-consistent polarized core-valence basis sets: ACVXZ).⁴⁸ If one were to include electron correlation into a CBS extrapolation involving valence basis set (e.g. AVXZ), then an absolute difference of $0.12\text{ kcal}\cdot\text{mol}^{-1}$ would be, on average, expected based on the three molecules reported herein.

Alternatively, if the core-valence basis sets (i.e. ACVXZ) were used, including the core electrons into the correlation only changes the relative energies by an average absolute value of $0.02\text{ kcal}\cdot\text{mol}^{-1}$ in comparison to the equivalent valence basis sets (i.e. ACV $[2,3,4]$ Z versus AV $[2,3,4]$ Z) without core-electron correlation. However, this value reflects both the addition of core electron correlation and the basis set enlargement since the ACVXZ basis sets contain more functions than the AVXZ basis sets. Thus from a practical viewpoint, the optimizations that correlate only the valence electrons using the valence basis sets provide relative energies that are very close to those that correlate both core and valence electrons using the core-valence basis sets. While one might say that neglecting core-electron correlation in the CBS extrapolation results in $0.12\text{ kcal}\cdot\text{mol}^{-1}$ error, an apparent fortuitous cancellation of errors suggests that such a statement would be slightly misleading.

3.3 Individual Basis Set Performance

If computing MP2/CBS energies is not possible or desirable, then knowing how individual basis sets perform becomes important. For the remainder of the analysis the MP2/AV $[T,Q,5]$ Z//MP2/AV $[T,Q,5]$ Z CBS values (Table 1, Scheme 2) will be used as target reference values. For geometry comparison, the MP2/AV5Z//MP2/AV5Z structures will be used as benchmark targets.

3.3.1 Relative Energies

The relative conformational energies computed using each basis set is given in Table S5. Table S6 provides the individual absolute error, the mean absolute errors for each molecule and the overall mean absolute errors. The overall mean absolute errors of the MP2 relative energies computed

using the Dunning, Jensen and Karlsruhe basis sets are plotted in Figure 5. As expected, the general trend is that smallest basis within each family yields the least accurate relative energies, with the unpolarized Jensen basis sets being the least accurate. In the release of the pc-n basis sets, Jensen stated that pc-0 (herein labeled as pc-X, X=1) was expected to be inaccurate¹⁶. The addition of a polarization function (e.g. pc-n, n=1) reduces the error to under 0.3 kcal·mol⁻¹.

Interestingly, there are two instances where using a sequentially larger basis set results in a slightly larger overall mean absolute error for the relative energy: aug-pcseg-3 (0.03 kcal·mol⁻¹) → aug-pcseg-4 (0.05 kcal·mol⁻¹) and def2-SV(P) (0.37 kcal·mol⁻¹) → def2-SVP (0.39 kcal·mol⁻¹). Decomposing the overall mean error as a function of the molecule reveals that these results originate from propan-2-ol and the water dimer (see Table S6; Figures S10–S12). Concerning the Karlsruhe basis sets, def2-SVP (0.48 kcal·mol⁻¹) performs worse than def2-SV(P) (0.36 kcal·mol⁻¹) only for the water dimer even though it includes an additional polarizing p function on hydrogen atoms.⁴³ Part of this comes from def2-SVP incorrectly finding the water dimer’s conformer **4**, a cyclic structure, to be more stable than the true global conformer **1** minimum.

Apart from VQZ (water dimer: 0.111 kcal·mol⁻¹), pcseg-4 (water dimer: 0.083 kcal·mol⁻¹; buta-1,3-diene: 0.059 kcal·mol⁻¹) and aug-pcseg-4 basis set (water dimer: 0.080 kcal·mol⁻¹), all quadruple zeta basis sets perform well, with mean absolute relative energies that are within 0.05 kcal·mol⁻¹ of the MP2/AV[T,Q,5]Z//MP2/AV[T,Q,5]Z CBS values. As expected for the Dunning and Jensen basis set families, enlarging the basis set to a quintuple size improved values further, with the largest mean absolute error of 0.034 kcal·mol⁻¹ (water dimer) occurring for V5Z.

If one would like to model a large molecule, HF/6-31G(d) optimizations are considered a good theory due to fortuitous cancellation of errors and its small size. If one can afford a slightly more expensive level, then MP2/def2-TZVP theories provide relative energies and geometries (Figure 6) that are significantly better than HF/6-31G(d). Also notable is the smaller MP2/def2-SVPD theory level. However, as seen in Figures S10–S18, its improvement is system dependent. For buta-1,3-diene, MP2/def2-SVPD provides nearly equivalent results as HF/6-31G(d), while for the water dimer it shows a significant improvement.

3.3.2 Geometries

In regards to reproducing MP2/AV5Z geometries, all-atom and heavy-atom RMSD values for each conformation are provided in Table S7, and whose mean values computed for each molecule are plotted in Figures S13–S18. Figure 6 shows the overall mean all-atom RMSD values for all conformations investigated. As seen in this plot, the RMSD improves as the number of basis functions increases within each basis set family, with the exception of def2-SVP (0.146 Å versus def2-SV(P)’s 0.124 Å). This exception is due to the RMSD contributions from the water dimer (see Figure S15), and mirrors the increased error seen in def2-SVP’s relative energies. Furthermore, through a comparison to the heavy-atom RMSD values (Table S7 and Figure S18), one can conclude that a significant amount of def2-SVP’s error comes from the water dimer’s hydrogen atoms’ positions. An examination of the mean all-atom RMSD values as a function of the molecule finds one additional minor inconsistency worth noting - MP2/def2-TZVPP results in a 0.003 Å worse geometry than MP2/def2-TZVP for the water dimer.

On average, the geometries optimized using triple zeta basis sets provide structures that are between 0.005–0.034 Å, with the best and worst being aug-pcseg-3 and VTZ, respectively. Similarly, geometries optimized using quadruple zeta basis sets are within 0.005 Å of the AV5Z target geometries, with the exception of VQZ (0.017 Å). For both VTZ and VQZ basis sets, the water dimer contributes the most to the overall mean RMSD error. Consequently, one should preferentially use the AVDZ and AVTZ over the VTZ and VQZ basis sets since they provide better geometries and relative energies, which also holds true for propan-2-ol. However interestingly, the VTZ and VQZ basis sets perform better for buta-1,3-diene. And finally, AVQZ and all of the Jensen quintuple basis sets (i.e. $X = 5$) produce geometries that are within 0.002 Å of the target geometries.

3.4 Timing

Since one goal of the Jensen and Karlsruhe basis sets is to reduce the overall calculation cost while maintaining accuracy, we present the relative calculation time for computing a single point energy calculation for propan-2-ol in Table 4 as a general guide for the cost of basis sets when used with MP2 theory. Note that the following results, analysis and rankings could change if one were to use an alternative software (e.g. Gaussian, Molpro) or optimization algorithm. Consequently, we encourage researchers to conduct their own timing study if they are wish to identify cost-effective theory levels using their software and workflow.

Using Psi4, three theory levels are equivalently fast: HF/6-31G(d), MP2/pcseg-1 and MP2/pcSseg-1. The most time-consuming theory is MP2/aug-pcSseg-5, which is more than 600 times slower than the fastest theories. The most accurate theory for modeling propan-2-ol is pcseg-5, with a mean absolute relative energy error of 0.003 kcal·mol⁻¹ (see Table S6), and is about 200 times slower.

Choosing a theory level that balances the calculation cost and accuracy is often difficult and subjective. To provide a quantitative guide for evaluating a theory level as a function of both cost and accuracy, we propose the following ranking equation:

$$\text{Rank}(nC, nE) = [f * nC] + [(1 - f) * nE] \quad (4)$$

where nC is a normalized cost, f is a scaling factor (i.e. 0→1) and nE is the normalized error. The smallest result of this equation (i.e. min(Rank)) will be the best ranked theory level for a given scaling factor, representing a user-desired balance between the theory’s cost and its error.

Since timing data is available, nC was chosen to be normalized relative times (Table 4) and nE the normalized mean absolute error in the relative energies (e.g. Table S6). Table 5 gives the top three theories for five different scaling factors. An expanded version of this table can be found in the SI material (Table S9) that provides the five best theories for scaling factors given in 0.1 intervals. At the scaling factor extremes, the equation correctly ranks MP2/pcseg-5 as the most accurate theory relative to the MP2/CBS target values (i.e. $f = 0.0$); while HF/6-31G(d), MP2/pcseg-1 and MP2/pcSseg-1 are the fastest theories (i.e. $f = 1.0$). For all other scaling factors, different theories arise that represent different cost and error balances. Notable is that for scaling factors 0.20–0.90, the augmented Dunning basis sets (i.e. AVTZ and AVDZ) are ranked bests. If one is willing to give up some accuracy for speed improvement, then aug-pcseg-3 and def2-TZVPPD often appears to be reasonable options (Table S9).

Alternatively, nC could represent a different normalized observable. For example, a reasonable observable would be the number of uncontracted basis functions (UCBF) for a given basis set. Consequently, a notable difference in theory ranking occurs, as seen in Table 5. A Pearson correlation analysis between the relative calculation time and the number of UCBF results in a value of 0.87 ($p=7.2\text{E-}14$). While it is often generally used, the number of basis functions (contracted or uncontracted) is not a perfect predictor of a theory’s speed. As noted above, this conclusion is drawn for Psi4’s algorithm, and different software could yield different results.

4 Conclusions

Performance comparisons of different basis sets are an important means to aid researchers in choosing the optimal molecular orbital representation for their investigations. Usually, this choice is driven by a desire for high data accuracy and confined by how costly the calculation becomes. In this paper we compared the performance of Dunning, Jensen and Karlsruhe basis sets, forty-five in total, for computing relative electronic energies and geometries of nineteen stationary points across three electronically diverse molecular systems (i.e. buta-1,3-diene, propan-2-ol and the water dimer). For benchmark data, geometries were computed at the MP2/AV5Z//MP2/AV5Z theory level and relative energies were extrapolated to the CBS limit

via MP2/AV[T,Q,5//Q,5]Z//MP2/AV[T,Q,5]Z. Two different CBS extrapolation schemes with different approximation levels were also compared. Finally, to practically guide researchers in their basis set choice, an equation was presented that ranks basis sets based on a user-defined balance between their accuracy and calculation cost.

In summary, the highlights from this work are the following:

1. In computing CBS relative energies using the augmented Dunning basis sets (i.e. AVXZ), one can reduce the calculation cost by using Scheme 1 without significantly reducing the results' accuracies. Note however, that the use of a less reliable geometry in Scheme 1 (e.g. originating from a lower theory level) can have a significant impact on the resulting extrapolated energies – and consequently the use of Scheme 1 and 2 would provide different results.
2. The MP2/CBS relative energies computed using the Dunning basis set [2,3,4] triads were, on average, more accurate than those computed using the corresponding Jensen [X=2,3,4] triads (Table 2). This is reasonable since the Dunning basis sets were optimized for use with electron correlated theories, while the Jensen basis sets were optimized for use in DFT theory.
3. Including core-electron correlation into the CBS extrapolations using the valence basis sets alters, on average, the relative energies by 0.12 kcal·mol⁻¹. However, the valence-only correlations using the standard Dunning basis sets (i.e. AVXZ) result in extrapolations that are in close agreement to the more rigorous calculations that include core electrons and correlation-consistent polarized core-valence basis sets Dunning basis sets (i.e. ACVXZ).
4. On average for the Jensen basis sets, the MP2/CBS relative energies computed using [X=1,2,3] triads resulted in values that were within 0.11–0.24 kcal·mol⁻¹ of those computed using its substantially larger [X=3,4,5] triads (Table 2). Similarly, the use of Jensen [X=2,3,4] triads provided CBS results that were within 0.09–0.16 kcal·mol⁻¹ of those computed by [X=3,4,5] triads. However, the use of [X=2,3,4] triads does not automatically guarantee that the results will be closer to the [X=3,4,5] CBS limits. For propan-2-ol and the water dimer, pcseg-X, aug-pcseg-X and pcSseg-X provided [X=1,2,3] triad extrapolated CBS values that were in closer agreement to the [X=3,4,5] CBS values.
5. Caution should be employed when using def2-SVP since it incorrectly computed the most stable water dimer conformation as the cyclic structure (i.e. conformation 4). Caution should also be extended to def2-SVPD, aug-pcSseg-1 (i.e. aug-pcSseg-n, n=0) and aug-pcSseg-2 (i.e. aug-pcSseg-n, n=1) for their inability to optimize the water dimer's 4 and 5 conformations.
6. Of the triple zeta basis sets, aug-pcseg-3 (i.e. aug-pcseg-n, n=2) provides the best overall geometries in comparison to MP2/AV5Z structures. When Dunning basis sets are desired for studying polar molecules, one might consider using AVDZ and AVTZ rather than the VTZ and VQZ basis sets, respectively, since they provide better geometries and relative energies using less functions.
7. Utilizing a cost/error equation provides a quantitative way for evaluating which theories should be preferentially explored when considering theory choice early within a study. In exploring different weighting factors, the aug-pcseg-3 and def2-TZVPPD basis sets appear to be reasonable options that balance calculational cost and accuracy when using Psi4.
8. When possible, MP2/def2-TZVP theory level seems to be a better choice for optimization of large molecules than HF/6-31G(d). The cheaper MP2/def2-SVPD theory is also worth investigating, but shows more dependency on the system being optimized.
9. We have introduced a new nomenclature (e.g. MP2/AV[T,Q,5//Q,5]Z//MP2/AV[T,Q,5]Z), an extension of the double-slash notation, that attempts to unambiguously indicate how a complete basis set extrapolation are computed.

Acknowledgement

We would like to thank Rudolf Berrendorf and Javed Razzaq for their continuous development and support of the university’s high performance computing cluster, and we highly value the feedback given by anonymous reviewers for improving the manuscript. We would also like to acknowledge the tremendous and inspiring contributions of Florian Müller-Plathe who acted as a trend-setter for a broad number of topics within the soft matter field and to whom we dedicate this work. Computer hardware was supported by the Federal Ministry for Education and Research and by the Ministry for Innovation, Science, Research, and Technology of the state Northrhine-Westfalia (research grant 13FH156IN6).

References

- [1] Python software foundation. Python language reference, version 3.4.3. <http://www.python.org>.
- [2] W. L. DeLano and Schrödinger, LLC. The PyMOL molecular graphics system, version 1.7.0.0. <https://pymol.org>, 2015. <https://pymol.org>.
- [3] T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90(2):1007–1023, 1989.
- [4] A. M. ElSohly and G. S. Tschumper. Comparison of polarization consistent and correlation consistent basis sets for noncovalent interactions. *Int. J. Quantum Chem.*, 109(1):91–96, 2009.
- [5] D. Feller. Application of systematic sequences of wave functions to the water dimer. *J. Chem. Phys.*, 96(8):6104–6114, 1992.
- [6] D. Feller. The use of systematic sequences of wave functions for estimating the complete basis set, full configuration interaction limit in water. *J. Chem. Phys.*, 98(9):7059–7071, 1993.
- [7] D. Feller and N. C. Craig. High level ab initio energies and structures for the rotamers of 1,3-butadiene. *J. Phys. Chem. A*, 113(8):1601–1607, 2009. PMID: 19199679.
- [8] S. Grimme and M. Steinmetz. Effects of London dispersion correction in density functional theory on the structures of organic molecules in the gas phase. *Phys. Chem. Chem. Phys.*, 15:16031–16042, 2013.
- [9] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson. Basis-set convergence in correlated calculations on Ne, N₂, and H₂O. *Chem. Phys. Lett.*, 286(3–4):243–252, 1998.
- [10] C. Hättig. Optimization of auxiliary basis sets for RI-MP2 and RI-CC2 calculations: Core-valence and quintuple- ζ basis sets for H to Ar and QZVPP basis sets for Li to Kr. *Phys. Chem. Chem. Phys.*, 7:59–66, 2005.
- [11] Y. He and D. Cremer. Molecular geometries at sixth order Møller-Plesset perturbation theory. At what order does MP theory give exact geometries? *J. Phys. Chem. A*, 104(32):7679–7688, 2000.
- [12] T. Helgaker, W. Klopper, H. Koch, and J. Noga. Basis-set convergence of correlated calculations on water. *J. Chem. Phys.*, 106(23):9639–9646, 1997.

- [13] A. Hellweg and D. Rappoport. Development of new auxiliary basis functions of the karlsruhe segmented contracted basis sets including diffuse basis functions (def2-SVPD, def2-TZVPPD, and def2-QVPPD) for RI-MP2 and RI-CC calculations. *Phys. Chem. Chem. Phys.*, 17:1010–1017, 2015.
- [14] J. D. Hunter. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.*, 9(3):90–95, 2007.
- [15] I. Hyla-Kryspin, S. Grimme, S. Hruschka, and G. Haufe. Conformational preferences and basicities of monofluorinated cyclopropyl amines in comparison to cyclopropylamine and 2-fluoroethylamine. *Org. Biomol. Chem.*, 6:4167–4175, 2008.
- [16] F. Jensen. Polarization consistent basis sets: Principles. *J. Chem. Phys.*, 115(20):9113–9125, 2001.
- [17] F. Jensen. Erratum: "Polarization consistent basis sets: Principles" [j. chem. phys. 115, 9113 (2001)]. *J. Chem. Phys.*, 116(8):3502–3502, 2002.
- [18] F. Jensen. Polarization consistent basis sets. III. The importance of diffuse functions. *J. Chem. Phys.*, 117(20):9234–9240, 2002.
- [19] F. Jensen. Polarization consistent basis sets. IV. The basis set convergence of equilibrium geometries, harmonic vibrational frequencies, and intensities. *J. Chem. Phys.*, 118(6):2459–2463, 2003.
- [20] F. Jensen. Unifying general and segmented contracted basis sets. Segmented polarization consistent basis sets. *J. Chem. Theory Comput.*, 10(3):1074–1085, 2014.
- [21] F. Jensen. Segmented contracted basis sets optimized for nuclear magnetic shielding. *J. Chem. Theory Comput.*, 11(1):132–138, 2015.
- [22] M. P. Johansson and J. Olsen. Torsional barriers and equilibrium angle of biphenyl: Reconciling theory with experiment. *J. Chem. Theory Comput.*, 4(9):1460–1471, 2008.
- [23] P. Jurečka, J. Šponer, J. Černý, and P. Hobza. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.*, 8:1985–1993, 2006.
- [24] A. R. Kendall and A. H. Früchtl. The impact of the resolution of the identity approximate integral method on modern ab initio algorithm development. *Theor. Chem. Acc.*, 97(1):158–163, 1997.
- [25] R. A. Kendall, T. H. Dunning, and R. J. Harrison. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.*, 96(9):6796–6806, 1992.
- [26] K. N. Kirschner, W. Heiden, and D. Reith. Small alcohols revisited: CCSD(T) relative potential energies for the minima, first- and second-order saddle points, and torsion-coupled surfaces. *ACS Omega*, 3(1):419–432, 2018.
- [27] S. Kozuch, S. M. Bachrach, and J. M. Martin. Conformational equilibria in butane-1,4-diol: A benchmark of a prototypical system with strong intramolecular H-bonds. *J. Phys. Chem. A*, 118(1):293–303, 2014.
- [28] T. Kupka, A. Buczek, M. A. Broda, A. Mnich, and T. Kar. Performance of polarization-consistent vs. correlation-consistent basis sets for CCSD(T) prediction of water dimer interaction energy. *J. Mol. Model.*, 25(10):313, 2019.

- [29] T. Kupka and C. Lim. Polarization-consistent versus correlation-consistent basis sets in predicting molecular and spectroscopic properties. *J. Phys. Chem. A*, 111(10):1927–1932, 2007.
- [30] R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill. Psi4 1.1: An open-source electronic structure program emphasizing automation, advanced libraries, and interoperability. *J. Chem. Theory Comput.*, 13(7):3185–3197, 2017.
- [31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, 12:2825–2830, 2011.
- [32] D. Rappoport and F. Furche. Property-optimized gaussian basis sets for molecular response calculations. *J. Chem. Phys.*, 133(13):134105, 2010.
- [33] D. R. Roe and T. E. Cheatham. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.*, 9(7):3084–3095, 2013.
- [34] A. Schäfer, H. Horn, and R. Ahlrichs. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.*, 97(4):2571–2577, 1992.
- [35] A. Schäfer, C. Huber, and R. Ahlrichs. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys.*, 100(8):5829–5835, 1994.
- [36] G. S. Tschumper, M. L. Leininger, B. C. Hoffman, E. F. Valeev, H. F. Schaefer III, and M. Quack. Anchoring the water dimer potential energy surface with explicitly correlated computations and focal point analyses. *J. Chem. Phys.*, 116(2):690–701, 2002.
- [37] O. Vahtras, J. Almlöf, and M. Feyereisen. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.*, 213(5):514 – 518, 1993.
- [38] G. van Rossum. Python tutorial. Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, 1995.
- [39] M. Waskom, O. Botvinnik, drewokane, P. Hobson, David, Y. Halchenko, S. Lukauskas, J. B. Cole, J. Warmenhoven, J. de Ruiter, S. Hoyer, J. Vanderplas, S. Villalba, G. Kunter, E. Quintero, M. Martin, A. Miles, K. Meyer, T. Augspurger, T. Yarkoni, P. Bachant, M. Williams, C. Evans, C. Fitzgerald, Brian, D. Wehner, G. Hitz, E. Ziegler, A. Qalieh, and A. Lee. Seaborn: v0.7.1 (june 2016), 2016.
- [40] F. Weigend. A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.*, 4:4285–4291, 2002.
- [41] F. Weigend. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.*, 8(9):1057–1065, 2006.
- [42] F. Weigend. Hartree-Fock exchange fitting basis sets for H to Rn. *J. Comput. Chem.*, 29(2):167–175, 2008.

- [43] F. Weigend and R. Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, 7:3297–3305, 2005.
- [44] F. Weigend, A. Köhn, and C. Hättig. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.*, 116(8):3175–3183, 2002.
- [45] A. K. Wilson, T. van Mourik, and T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. VI. Sextuple zeta correlation consistent basis sets for boron through neon. *J. Mol. Struct. Theochem*, 388:339–349, 1996.
- [46] J. Witte, J. B. Neaton, and M. Head-Gordon. Push it to the limit: Characterizing the convergence of common sequences of basis sets for intermolecular interactions as described by density functional theory. *J. Chem. Phys.*, 144(19):194306, 2016.
- [47] D. E. Woon and T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.*, 98(2):1358–1371, 1993.
- [48] D. E. Woon and T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. V. Core-valence basis sets for boron through neon. *J. Chem. Phys.*, 103(11):4572–4585, 1995.

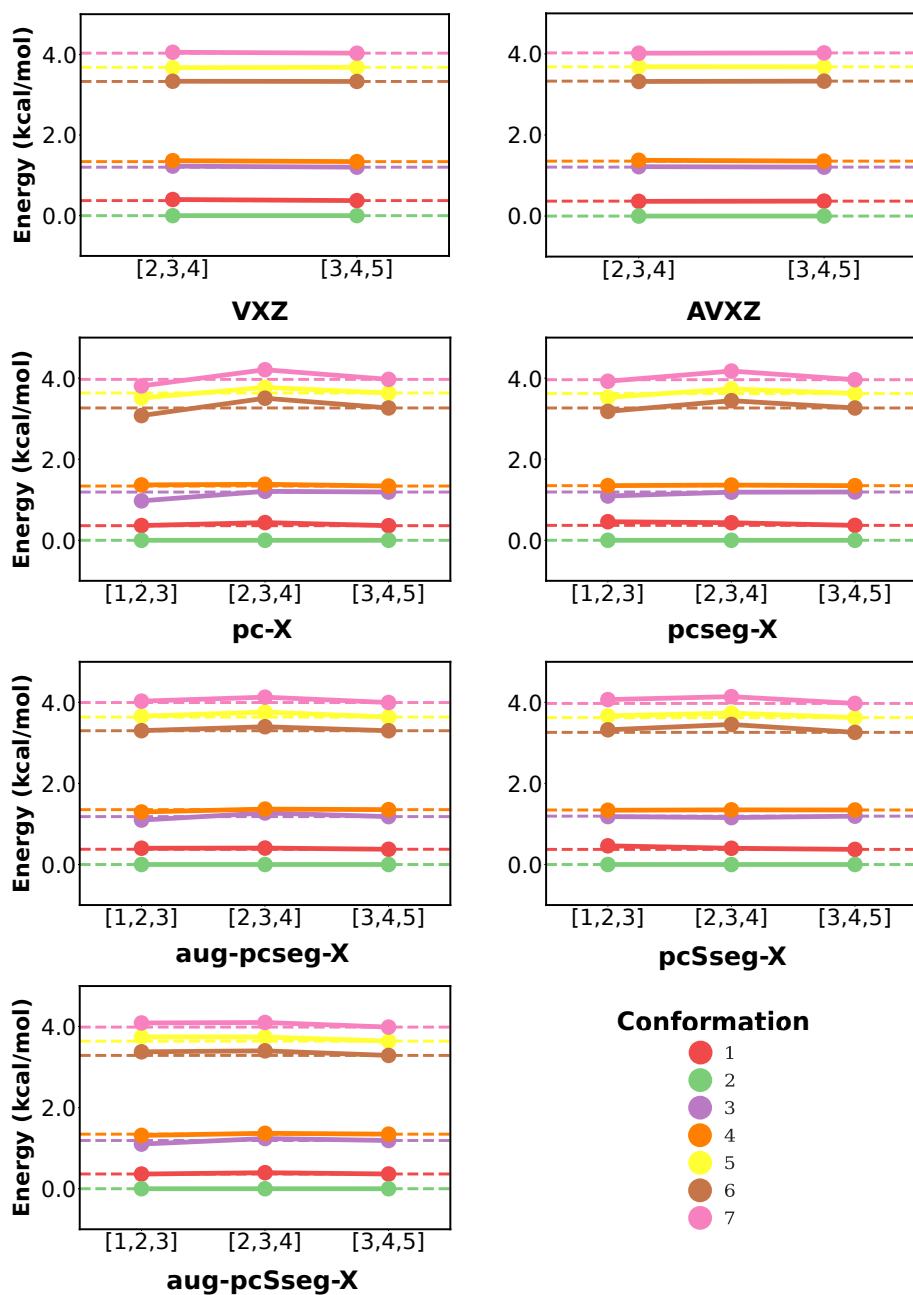


Figure 3: The MP2/CBS relative energies (Scheme 2) computed from the possible sequential Dunning and Jensen basis set triad combination for propan-2-ol conformations.

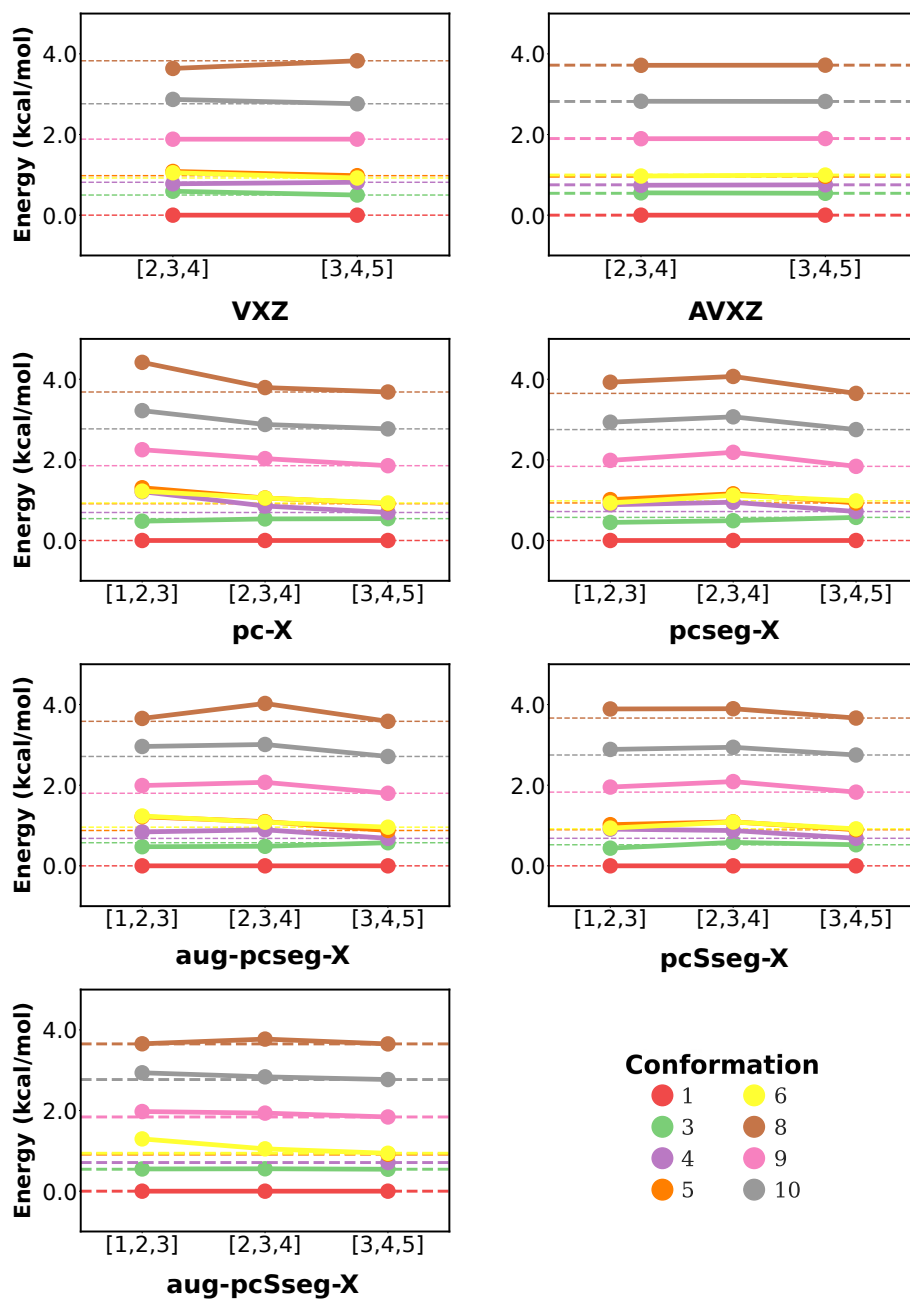


Figure 4: The MP2/CBS relative energies (Scheme 2) computed from the possible sequential Dunning and Jensen basis set triad combination for the water dimer conformations.

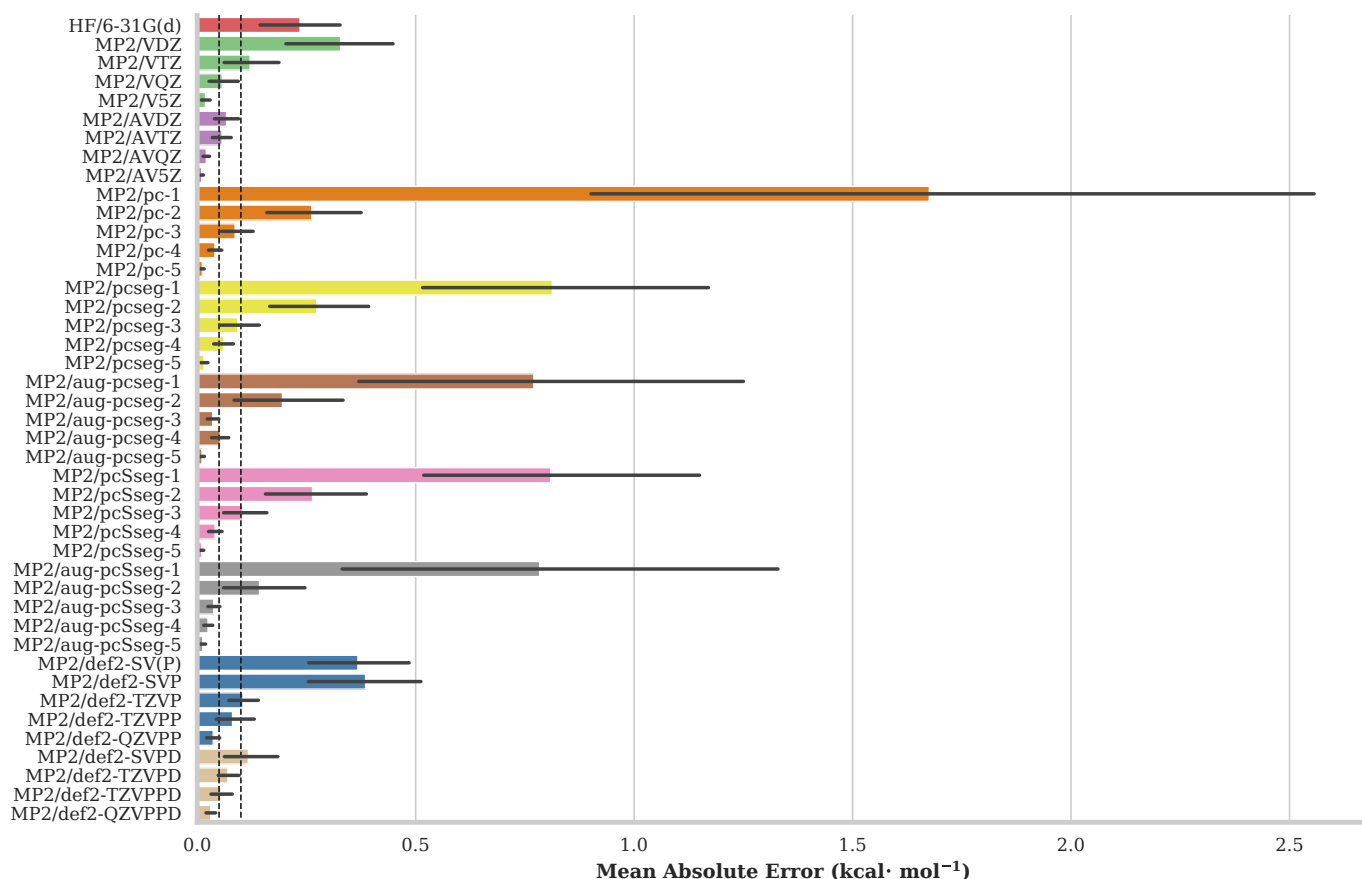


Figure 5: The mean absolute error (kcal·mol⁻¹) in the relative energies of the propan-2-ol, buta-1,3-diene and water dimer conformations and configurations, using MP2/CBS (i.e. MP2/AV[T,Q,5//Q,5]Z//MP2/AV[T,Q,5]Z, Scheme 2) energies as reference values. The solid black lines indicate the standard errors of the mean. The two dashed lines indicate values at 0.05 and 0.10 kcal·mol⁻¹. The Jensen basis sets are noted using X=1–5 indexing.

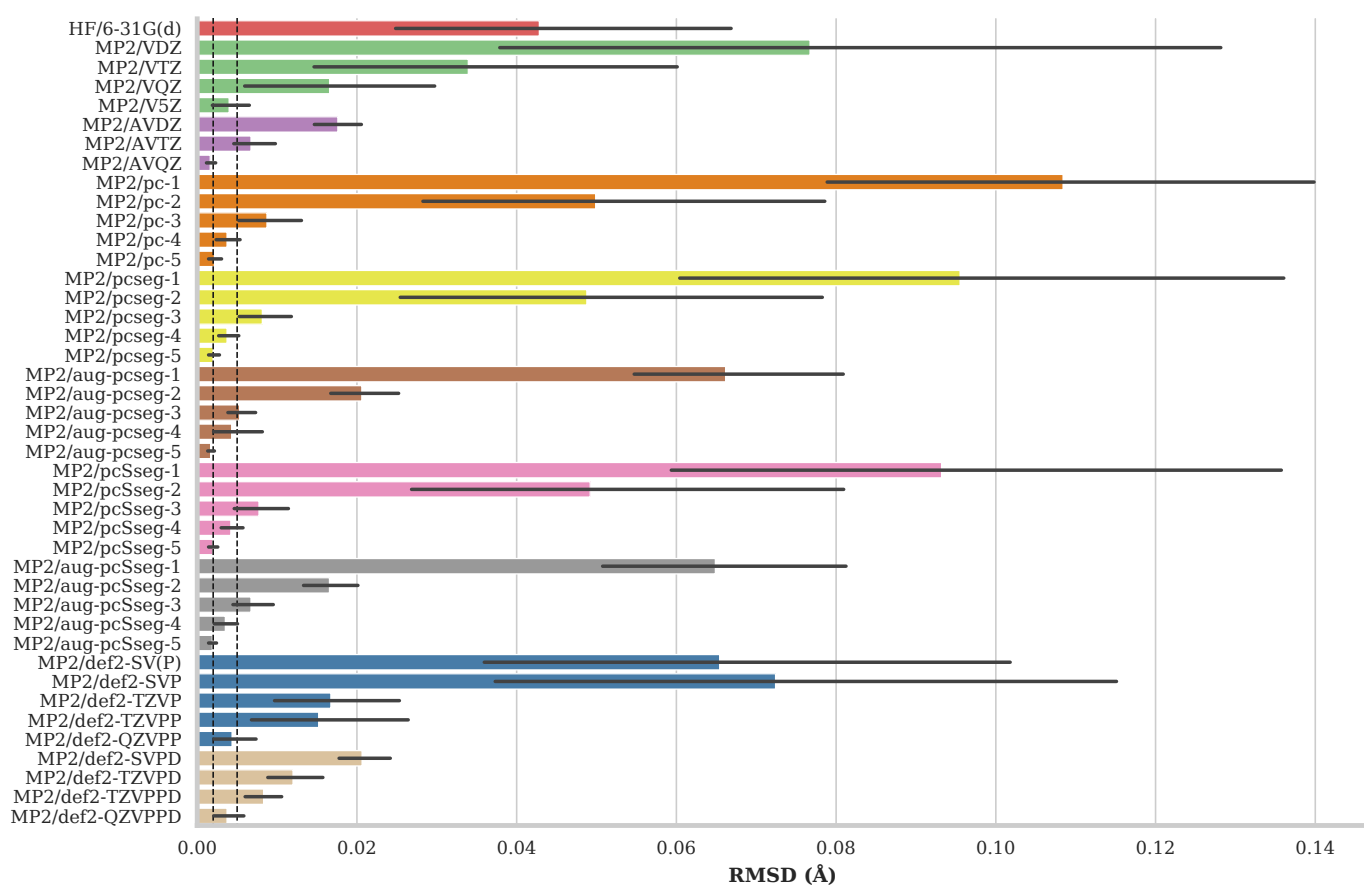


Figure 6: The mean all-atom RMSD (Å) of the propan-2-ol, buta-1,3-diene and water dimer conformations in comparison to MP2/AV5Z geometries. The two dashed lines indicate values at 0.002 and 0.005 Å. The Jensen basis sets are noted using X=1–5 indexing.

Table 1: MP2/CBS relative energies (kcal·mol⁻¹) of propan-2-ol, buta-1,3-diene and water dimer conformations and configurations using two different extrapolation schemes. CCSD(T) values from literature are provided for comparison

Conf.	Description	Scheme 1		Scheme 2			
		MP2/AV[D,T,Q]//T,Q[Z]//MP2/AV5Z	MP2/AV[T,Q,5]//Q,5[Z]//MP2/AV5Z	MP2/AV[D,T,Q]//T,Q[Z]//MP2/AV[D,T,Q]Z	MP2/AV[T,Q,5]//Q,5[Z]//MP2/AV[T,Q,5]Z	CCSD(T)	
buta-1,3-diene							
1	minimum	0.000	0.000	0.000	0.000	0.00 ^a	
2	minimum	2.970	2.923	2.967	2.922	3.01	
3	1 st -order TS	6.639	6.578	6.641	6.579	6.40	
4	1 st -order TS	3.693	3.684	3.692	3.684	3.49	
propan-2-ol							
2	minimum	0.000	0.000	0.000	0.000	0.000 ^b	
1	minimum	0.361	0.367	0.363	0.368	0.257	
3	1 st -order TS	1.219	1.207	1.219	1.207	1.142	
4	1 st -order TS	1.375	1.355	1.377	1.355	1.274	
5	1 st -order TS	3.681	3.682	3.684	3.682	3.482	
6	1 st -order TS	3.316	3.329	3.320	3.329	3.148	
7	1 st -order TS	4.013	4.023	4.018	4.025	3.728	
water dimer							
1	minimum	0.000	0.000	0.000	0.000	0.00 ^c	
2 ^d	1 st -order TS					0.52	
3	2 nd -order TS	0.552	0.546	0.553	0.542	0.57	
4	1 st -order TS	0.745	0.755	0.742	0.752	0.70	
5	2 nd -order TS	0.948	0.966		0.964	0.95	
6	3 rd -order TS	0.972	1.002	0.971	0.995	0.99	
7 ^d	2 nd -order TS					1.81	
8	3 rd -order TS	3.715	3.717	3.716	3.720	3.57	
9	1 st -order TS	1.893	1.901	1.896	1.900	1.79	
10	2 nd -order TS	2.820	2.821	2.824	2.820	2.71	

^a The relative energies were computed from a combination of CCSD(T)(FC)/CBS, CCSD(T)(CV)/cc-pwCVQZ and CCSDT(Q)(FC)/VDZ calculations⁷.

^b The relative energies were computed at CCSD(T)/AVTZ//MP2/AV5Z²⁶.

^c The relative energies were computed at the CCSD(T)/CBS//CCSD(T)/TZ2P(f,d)+dif theory level with core correlation and relativistic corrections³⁶.

^d Conformers 2 and 7 from reference 36 could not be optimized using any theory levels studied herein.

Table 2: The mean absolute error of the MP2/CBS relative energies ($\text{kcal}\cdot\text{mol}^{-1}$, Scheme 2) relative to the values computed by the [3,4,5] triad within each basis set family^a

	VXZ [2,3,4]	AVXZ [2,3,4]	pc-X [1,2,3] [2,3,4]	pcseg-X [1,2,3] [2,3,4]	aug-pcseg-X [1,2,3] [2,3,4]	pcSseg-X [1,2,3] [2,3,4]	aug-pcSseg-X [1,2,3] [2,3,4]					
Buta-1,3-diene	0.017	0.039	0.103	0.034	0.157	0.026	0.598	0.138	0.189	0.052	0.354	0.119
Propan-2-ol	0.020	0.009	0.120	0.126	0.068	0.099	0.038	0.081	0.051	0.090	0.070	0.072
Water Dimer	0.095	0.010	0.407	0.118	0.148	0.252	0.198	0.237	0.135	0.186	0.133	0.081
Overall	0.052	0.015	0.243	0.105	0.120	0.152	0.213	0.160	0.114	0.125	0.153	0.085

^a All individual MP2/CBS relative energies and absolute errors computed by the different basis set triads are given in Table S4.

Table 3: MP2/CBS relative energies ($\text{kcal}\cdot\text{mol}^{-1}$) computed with both valence and core electron correlated using Scheme 2. Differences are provided relative to values computed when correlating only the valence electrons (see Table 1)

Conf.	MP2/AV[D,T,Q]//T,Q[Z]// MP2/AV[D,T,Q]Z	Diff. ^a	MP2/AV[T,Q,5]//Q,5[Z]// MP2/AV[T,Q,5]Z	Diff. ^b	MP2/ACV[D,T,Q]//T,Q[Z]// MP2/ACV[D,T,Q]Z	Diff. ^a
Buta-1,3-ene						
1	0.000	0.000	0.000	0.000	0.000	0.000
2	3.026	0.059	2.738	0.184	2.989	0.022
3	6.841	0.200	6.157	0.422	6.665	0.024
4	3.705	0.013	3.517	0.167	3.717	0.025
Average		0.091		0.258		0.024
Propan-2-ol						
2	0.000	0.000	0.000	0.000	0.000	0.000
1	0.457	0.094	0.312	0.056	0.364	0.001
3	1.282	0.063	1.176	0.031	1.217	0.002
4	1.466	0.089	1.342	0.013	1.379	0.002
6	3.899	0.215	3.698	0.016	3.715	0.031
7	3.530	0.210	3.337	0.008	3.353	0.033
8	4.300	0.282	3.987	0.038	4.053	0.035
Average		0.159		0.027		0.017
Water dimer						
1	0.000	0.000	0.000	0.000	0.000	0.000
3	0.577	0.024	0.565	0.023	0.541	0.012
4	0.631	0.111	0.899	0.147	0.744	0.002
5			1.116	0.152		
6	0.861	0.110	1.133	0.138	0.970	0.001
8	3.603	0.113	3.911	0.191	3.751	0.035
9	1.774	0.122	2.057	0.157	1.913	0.017
10	2.708	0.116	2.992	0.172	2.838	0.014
Average		0.099		0.140		0.014
Overall Average		0.121		0.120		0.017

^a Absolute difference computed in comparison to MP2/AV[D,T,Q]//T,Q[Z]//MP2/AV[D,T,Q]Z Scheme 2 values of Table 1.

^b Absolute difference computed in comparison to MP2/AV[T,Q,5]//Q,5[Z]//MP2/AV[T,Q,5]Z Scheme 2 values of Table 1.

Table 4: Average relative calculation times for computing a single point energy calculation.^{a,b} The raw timings can be found in Table S8

Pople Basis Set HF/6-31G(d)	Dunning Basis Sets	
	MP2/VDZ	MP2/VTZ
1.0 ^c	1.7	6.2
	MP2/AVDZ	MP2/AVTZ
	3.7	19.4
		MP2/AVQZ
		116.3
		MP2/V5Z
		166.1
		MP2/AV5Z
		497.6
Jensen Basis Sets ^d	Jensen Basis Sets ^d	
	MP2/pc-2	MP2/pc-3
MP2/pc-1	2.8	6.9
1.9		
MP2/pcseg-1	MP2/pcseg-2	MP2/pcseg-3
1.0 ^c	1.6	5.0
MP2/pcseg-1	MP2/pcseg-2	MP2/pcseg-3
1.2	2.9	12.2
MP2/pcSseg-1	MP2/pcSseg-2	MP2/pcSseg-3
1.0 ^c	1.7	7.0
MP2/pcSseg-1	MP2/pcSseg-2	MP2/pcSseg-3
1.3	3.2	13.6
		MP2/pc-4
		28.5
		MP2/pcseg-4
		25.9
		MP2/pcseg-4
		57.4
		MP2/pcSseg-4
		30.2
		MP2/pcSseg-4
		64.8
		MP2/pc-5
		88.9
		MP2/pcseg-5
		200.4
		MP2/pcseg-5
		551.3
		MP2/pcSseg-5
		237.0
		MP2/pcSseg-5
		610.7
Karlsruhe Basis Sets	Karlsruhe Basis Sets	
	MP2/def2-SVP	MP2/def2-TZVP
MP2/def2-SV(P)	1.6	3.3
1.2		
	MP2/def2-SVPD	MP2/def2-TZVPD
	2.5	4.9
	MP2/def2-TZVPP	MP2/def2-TZVPP
	5.6	22.7
	MP2/def2-TZVPPD	MP2/def2-TZVPPD
	7.8	29.4

^a A total of six separate SCF calculations were performed using the same propan-2-ol input geometry to generate the average values.

^b All calculations were performed using one core of an AMD FX-8350 processor within a Linux Mint 17.1 desktop computer that contained 31.5 GB of RAM.

^c The average raw time was 1.8 seconds.

^d The Jensen basis sets are noted using X=1–5 indexing.

Table 5: Top three theory levels^a ranked by Equation 4 as a function of the scaling factor (f) used. Two sets of ranking are presented that is based on the normalized mean absolute energy error (nE), and either the use of normalized relative mean time or the normalized number of uncontracted basis set functions (nC).^b

f	Theory	nC = Relative Mean Time			nC = Number of Cartesian Functions					
		Relative Mean Time	Cartesian Functions	Mean Absolute Energy Error	Rank	Theory	Relative Mean Time	Cartesian Functions	Mean Absolute Energy Error	Rank
0.00	MP2/pcseg-5	200.4	1160	0.003	0.000	MP2/pcseg-5	200.4	1160	0.003	0.000
	MP2/AV5Z	497.6	1568	0.004	0.002	MP2/AV5Z	497.6	1568	0.004	0.002
	MP2/V5Z	166.1	1064	0.005	0.004	MP2/V5Z	166.1	1064	0.005	0.004
0.25	MP2/AVTZ	19.4	420	0.016	0.025	MP2/AVDZ	3.7	172	0.036	0.061
	MP2/VQZ	33.6	560	0.016	0.030	MP2/AVTZ	19.4	420	0.016	0.073
	MP2/def2-QZVPP	22.7	568	0.021	0.032	MP2/def2-TZVPPD	7.8	319	0.033	0.079
0.50	MP2/AVTZ	19.4	420	0.016	0.027	MP2/AVDZ	3.7	172	0.036	0.065
	MP2 aug-pcseg-3	12.2	420	0.025	0.028	MP2/def2-SVPD	2.5	155	0.069	0.088
	MP2/AVDZ	3.7	172	0.036	0.031	MP2/pcseg-3	5.0	260	0.041	0.096
0.75	MP2/AVDZ	3.7	172	0.036	0.018	MP2/AVDZ	3.7	172	0.036	0.068
	MP2/def2-TZVPPD	7.8	319	0.033	0.021	MP2/def2-SVPD	2.5	155	0.069	0.075
	MP2/pcseg-3	5.0	260	0.041	0.021	MP2/aug-pcSseg-1	1.3	76	0.155	0.076
1.00	HF/6-31G(d)	1.0	76	0.181	0.000	MP2/pc-1	1.9	52	0.581	0.000
	MP2/pcseg-1	1.0	52	0.315	0.000	MP2/pcseg-1	1.0	52	0.315	0.000
	MP2/pcSseg-1	1.0	52	0.334	0.000	MP2/pcSseg-1	1.0	52	0.334	0.000

^a The Jensen basis sets are noted using X=1–5 indexing.

^b A more complete list that contains 0.1 incremented scaling factors can be found in Table S9.