

Quantifying Intramolecular Basis Set Superposition Errors

Quentin Pitteloud^a, Peter Wind^a, Stig Rune Jensen^a, Luca Frediani^a, and Frank Jensen^{*b}

^aHylleraas Centre, Department of Chemistry, UiT The Arctic University of Norway, Tromsø N-9037, Norway

^bDepartment of Chemistry, Aarhus University, Aarhus DK-8000, Denmark, Email: frj@chem.au.dk

April 6, 2024

Abstract

We show that medium sized Gaussian basis sets lead to significant intramolecular basis set superposition errors at Hartree-Fock and Density Functional levels of theory, with artificial stabilization of compact over extended conformations for a 186 atom deca-peptide. Errors of ~ 80 kJ/mol and ~ 10 kJ/mol are observed, with polarized double zeta and polarized triple zeta quality basis sets, respectively. Two different procedures for taking the basis set superposition error into account are tested: While both reduce the error, it appears that polarized quadruple zeta basis sets are required to reduce the error below a few kJ/mol. Alternatively, the basis set superposition error can be eliminated by using multiresolution methods, based on Multiwavelets.

1 Introduction

Electronic structure calculations using density functional theory (DFT) are popular for probing a variety of properties for both molecular and extended systems. The typical approach for practical applications involves expanding the orbitals in a Gaussian basis set for molecules and plane waves for extended systems. Gaussian basis sets can also be used for extended systems, but special tailored basis sets are required to achieve numerical stability,^{1,2} although progress in using conventional basis sets has been made lately.³ On the other hand, plane waves can be used for molecular systems, but require quite large unit cells⁴ and a special treatment of the core electrons to remain computationally feasible, for example by replacing them with a pseudopotential.⁵

Recent developments have considered methods based on multiresolution analysis (MRA),^{6,7} with the basis set consisting of polynomials in adaptive regions of the three-dimensional space, and these methods are becoming competitive with traditional methods in terms of computational efficiency, especially if high precision is required.⁸

A large number of Gaussian basis sets have been proposed, and these can be considered as different compromises between accuracy and computational efficiency.^{9,10} Modern basis sets are usually defined in families, with the cardinal number representing a systematic progression towards the complete basis set (CBS) limit. As the convergence towards the latter depends on the underlying method, the optimum convergence requires basis sets that are optimized towards a specific type of method, such as the cc-pVXZ for wave function correlated methods¹¹ and the pcseg-*n* for DFT.¹² Although this allows a rigorous control over basis set errors, it may be infeasible to reach the desired accuracy, especially for large systems.

A particular problem arises for calculating the interaction between fragments, as incomplete basis sets will generally lead to an overestimation of the interaction energy. The reason is qualitatively easy to understand, as the basis set for the interacting fragments is more complete than for the separated species, and this effect is denoted Basis-Set Superposition Error (BSSE).¹³ A number of different schemes have been proposed to correct for BSSE, with the most commonly used being the counter-poise (CP),^{14,15} where the fragment energies are calculated in the full basis set of all fragments. It is recognized that the CP is only an estimate of the BSSE, and often overestimates the effect, but interaction energies are often found to converge faster towards the CBS limit with CP correction,¹⁶⁻²² than without this correction. The overestimation of the CP correction has led to approaches where only half the calculated correction is included,^{23,24} or only the virtual orbitals of the other fragment are included.²⁵ These conclusions are based on the difference relative to the CBS limit, but it should be recognized that this includes also the basis set incompleteness effect.²⁶

When the fragments are isolated molecules, the resulting intermolecular CP is well-defined. However, the same effect is present for calculating differences in conformational energies for larger systems, and this has been coined intramolecular BSSE:²⁷⁻²⁹ spatially compact conformations are artificially stabilized relative to extended conformations.^{30,31} The analogy of the intermolecular CP procedure can be used to estimate the intramolecular BSSE, but this requires a manual and somewhat arbitrary decision of how to define fragments within a molecule. Nevertheless, it has been shown that correction for intramolecular BSSE can have a large effect of for example rotational energy profiles for a tripeptide.³²⁻³⁴ A version based on atomic fragments removes the user involvement,³⁵ but it has been argued that only the CP effect for atoms beyond the close bonding region should be included, and we have suggested an Atomic Counter Poise (ACP) model where only atoms separated by at least three bonds are included.³⁶ By discarding atoms beyond a cutoff distance where the overlap is negligible, this can be made computationally tractable.³⁶ An alternative approach, labelled geometry Counter Poise (gCP), involves calculating only over-

lap between model atomic orbitals, and it has been parameterized for a selection of combined methods and basis sets.³⁷ The gCP requires negligible computational time, while the ACP method scales linearly with the system size in the large system limit, and for the present systems typically requires a computational time an order of magnitude larger than calculating the energy itself.

Alternative approaches, such as the chemical Hamiltonian approach^{38,39} or same number of optimized parameters,^{40,41} have been much less used than the CP, and are only defined for intermolecular fragments. A valence bond approach with restrictions on which basis functions are allowed to contribute to a given localized orbital can be used to estimate both inter- and intra-molecular BSSE,⁴² but this cannot account for possible charge-transfer between fragments.

Calculating conformational energies of medium sized systems, such as small peptides with a few hundred atoms, by electronic structure methods is a valuable tool in parameterizing force field methods.⁴³⁻⁴⁵ Such systems, however, are clearly susceptible to artifacts from intramolecular BSSE, but there are virtually no systematic studies of how large such errors can be, as calculations close to the CBS limit using standard basis sets are difficult. BSSE is absent when using plane wave basis functions, but replacing the core electrons by e.g. a pseudopotential⁵ introduces an inherent limitation that is difficult to quantify, and the requirement of using large unit cells for medium sized molecules make this unattractive. Recent developments in the use of MRA applied to quantum chemistry,^{7,46} however, have made it possible to generate precise reference values⁴⁷ at both Hartree-Fock (HF) and DFT levels at computational accessible costs.^{8,48} In the present work we quantify the magnitude of intramolecular BSSE for a selection of spatially different conformations for a deca-peptide at HF and DFT levels using a selection of standard Gaussian basis sets against reference multiresolution values, and evaluate the performance of proposed correction schemes. Calibration studies of BSSE have often focused on small model systems, but the present 186 atom system can be considered as a real-case application.

We note that BSSE at correlated levels of theory usually are larger than for independent particle models, like HF and DFT, and the present results can thus be taken as lower bounds for e.g. MP2 or double hybrid methods. As the BSSE is not expected to depend significantly on the specific exchange-correlation functional, we have elected to use only the popular B3LYP model⁴⁹ in addition to HF. Environmental effects are likewise not expected to be important, and all calculations have therefore been done in vacuum. The present work only addresses the effects of intramolecular BSSE on conformational energies, but including dispersion will be essential for calculating accurate values.

2 Methods

Conformations of teixobactin were generated by a Monte-Carlo approach and structures optimized by the MMFF force field,⁵⁰ yielding ~ 4000 conformations. The sampling statistics suggests that this is far from a complete sampling, but

it is sufficient for selecting a range of extended to compact structures. The conformations were sorted according to the difference between the largest and smallest eigenvalue of their moment of inertia tensor, which is used as a proxy for the elongation of the molecule. 20 structures were selected according to a uniform sampling of the sorted values. There is a strong correlation between the difference between the largest and smallest eigenvalue of their moment of inertia and the average moment of inertia, as shown in Figure 1.

HF and DFT calculations using Gaussian basis sets were done using the Gaussian-16 program package,⁵¹ while reference values were calculated using the MRChem⁵² program. The multiwavelet (MW) basis of MRA employs a basis of polynomials⁵³ on a real-space grid which is adaptively refined based on the norm of the wavelet functions, to make sure that the error is below a predefined threshold.⁶ The threshold is defined by a single accuracy parameter, and the corresponding results are labeled as MW n , where n indicates a 10^{-n} precision. The ACP calculations employed the ACP(4) option,³⁶ where 4 indicates that only atoms separated by 4 bonds or more are included in the CP procedure, and a cutoff of 10 Å was employed.

3 Results

We have selected the teixobactin peptide, which is a 186-atom decapeptide of biological interest as an antibiotic agent.⁵⁴ The system has a net charge of +1 and a cyclic structure at one end of the chain. From a force field generated conformational search, we have selected 20 conformations representing a range of structures from extended to compact, and Figure 2 shows the two most different conformations.

The hypothesis is that extended structures will have the least amount of intramolecular BSSE, and small basis sets will thus artificially stabilize the compact conformations. By investigating how the relative conformational energies depend on the quality of the basis set, we can quantify the magnitude of the intramolecular BSSE. A precise quantification requires a valid reference: In the present case we will make use of the the relative energies obtained at the MW7 level, which yield errors well below 0.01 kJ/mol, with respect to the CBS limit.

The results using Gaussian basis sets can be extrapolated to the CBS limit, and mean field methods like HF and DFT are expected to converge exponentially with the square root of the basis set cardinal number X .^{55,56} We have used the three-point extrapolation in Eq. 1 based on the pcseg-2,-3,-4 results. We note that this extrapolation provides results closer to the reference MW7 than a pure exponential and also better than functions with the cardinal number as a prefactor for the exponential.⁵⁷

$$E(X) = E(\infty) + A \exp -B\sqrt{X} \quad (1)$$

The pcseg-2,-3,-4 extrapolated absolute energies agree with the MW7 reference results to within 16 and 7 milli-Hartree at the HF and B3LYP levels

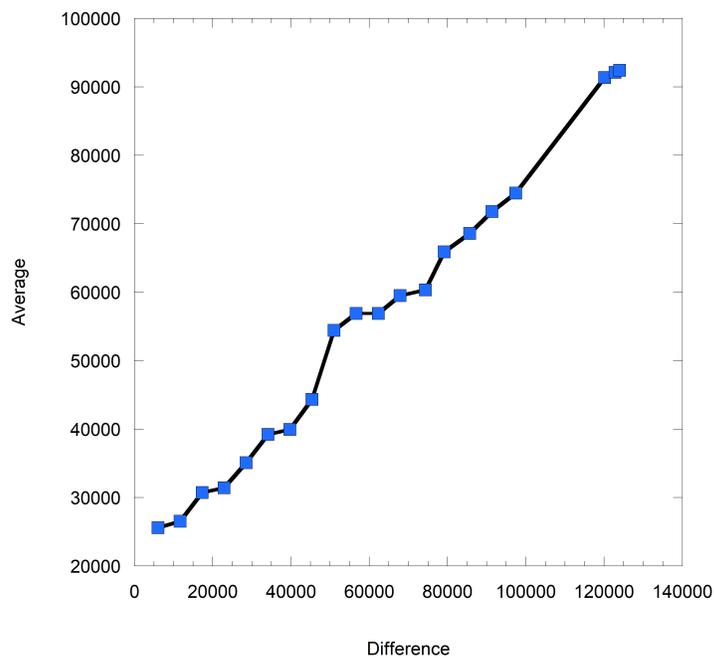


Figure 1: Correlation between the average value and the difference between highest and lowest eigenvalue of the inertia tensor (atomic units), the selected conformations denoted by points are labeled 1-20 left to right.

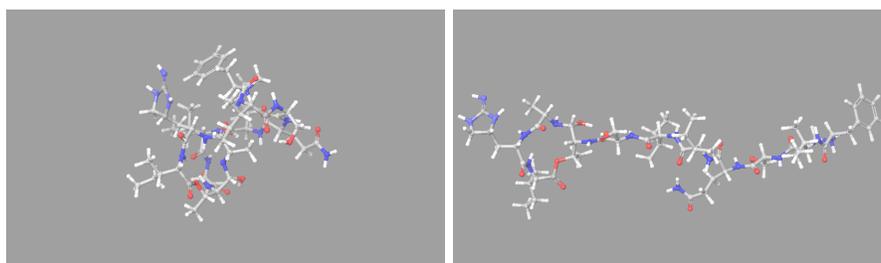


Figure 2: Most compact (left) and most extended (right) selected conformations of the teixobactin peptide, labeled as conformation 1 and 20, respectively

respectively (corresponding to 42 and 18 kJ/mol), and relative energies agree to within 0.04 kJ/mol on average, with the maximum deviation being 0.15 kJ/mol,

at both levels.

The MW7 relative energies of the 20 conformations at the HF and B3LYP levels are shown in Table 1, where the order label reflects the degree of compactness according to Figure 1, 1 being the most compact and 20 being the most extended, shown in the left and right panel in Figure 2, respectively. The energy of the most extended conformation is defined as zero, as this is expected to be least affected by intramolecular BSSE. The conformations span an energy range of ~ 200 kJ/mol, but there is no correlation between the relative energies and the degree of compactness. Neither HF nor B3LYP account for dispersion effects which are very important for predicting accurate conformational energies, and the last column in Table 1 shows the dispersion correction calculated with the D3 parameterization of the B3LYP method.⁵⁸ There is a significant correlation between the degree of compactness and the predicted differential stabilization by dispersion.

Label	Rel. E [HF]	Rel. E [B3LYP]	Rel. ΔE [D3]
1	-24.9	-56.7	-166.0
2	-25.1	-55.2	-150.1
3	86.5	47.9	-140.0
4	-25.0	-46.5	-131.9
5	-74.7	-114.6	-90.1
6	3.0	-11.0	-85.6
7	-60.6	-96.6	-121.9
8	79.8	41.8	-101.0
9	36.8	24.6	-28.5
10	110.7	89.5	-50.4
11	78.8	59.5	-28.3
12	35.2	24.6	-25.9
13	131.3	112.8	-64.2
14	75.6	52.3	-51.7
15	135.4	124.2	-39.3
16	-2.9	-18.7	-50.6
17	84.1	70.2	-50.4
18	-10.4	-12.1	2.4
19	-4.1	-3.9	2.3
20	(0.0)	(0.0)	(0.0)

Table 1: Reference relative energies in kJ/mol of conformers at MMFF optimized geometries at the MW7 HF and B3LYP levels with respect to the most extended conformer. The column labeled Rel. ΔE [D3] is the correction from dispersion calculated with the D3 parameterization of the B3LYP method. The labels 1-20 refer to the order of the degree of compactness, with 1 being the most compact and 20 the most extended, as shown in Figure 2

Figure 3 shows the absolute error in the conformational energy for the pcseg-

n basis sets as a function of the compactness order. The left panel refers to HF calculations, whereas the right panel contains the B3LYP results. Note the logarithmic energy scale and that only values for conformations 1-19 are shown, as conformation 20 is taken as a reference. Figure 3 shows a correlation between the conformer elongation and the error relative to the reference conformer, demonstrating the effect of intramolecular BSSE. Note that the conformations labeled 18, 19 and 20 according to Figure 1 are very similar, serving as an internal reference for our hypothesis, and their relative energies are indeed insensitive to the quality of the basis set. The magnitude of the BSSE error and its correlation with the elongation of the structure are both progressively smaller along the pcseg- n series, both for HF and B3LYP. In particular for B3LYP calculations with large pcseg- n bases ($n=3,4$) the errors are small, and not correlated with the elongation. The structural compactness measured by the moments of inertia is expected to provide a qualitative correlation with the number of atom pairs in close contact, and thus the intramolecular BSSE error. The correlation, however, is only qualitative, and for example conformations 9 and 15 have lower errors than expected from their moments of inertia. Table 1 shows that conformations 9 and 15 also have lower predicted dispersion corrections, which is consistent with these conformations having fewer atom pairs in close contact than expected from the moments of inertia.

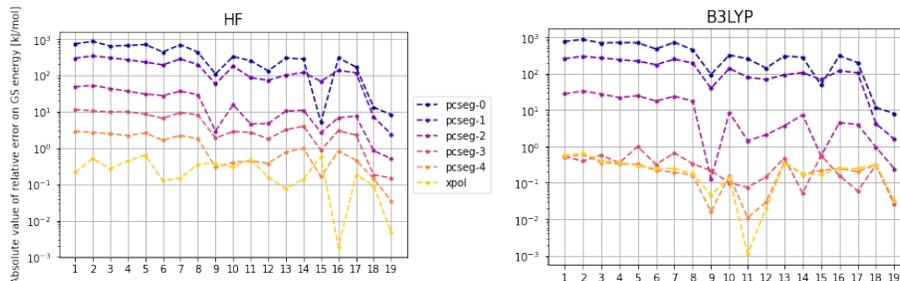


Figure 3: Absolute errors for the conformational energy of the 19 conformers relative to conformer 20. Each line in the graph refer to a different pcseg- n basis, with the MW7 conformational energies taken as a reference. HF results in the left panel and B3LYP results in the right panel. Logarithmic scale

The corresponding values for the MW results are collected in Figure 4. The MW results are less correlated with the conformer elongation, as would be expected for a numerically complete basis: the resulting error is mostly a consequence of the imposed precision and should be regarded mostly as numerical noise.

Figure 5 compares the behavior of the Gaussian basis set and multiresolution approaches, and shows that the errors on conformational energies computed with Gaussian basis sets approach the reference values from below (overstabilized) and they clearly show a correlation with their elongation (larger overstabilization for more compact conformers). The opposite is true for lower precision

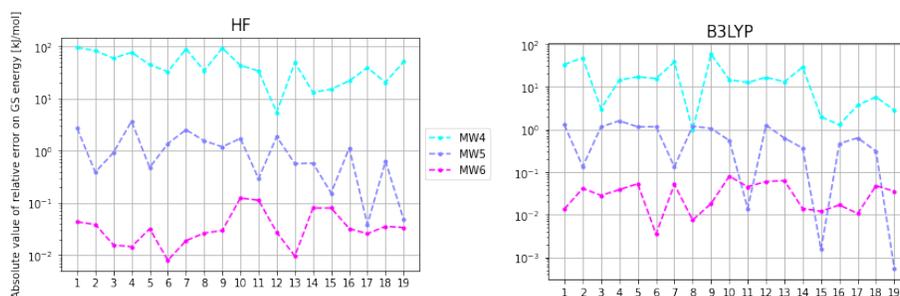


Figure 4: Absolute error for the conformational energy of the 19 conformers relative to conformer 20. Each line in the graph refer to a different precision (MW4, 5 and 6), with the MW7 conformational energies taken as a reference. HF results in the left panel and B3LYP results in the right panel. Logarithmic scale.

multiwavelet bases (MW4 and MW5), as most conformational energies are underestimated, although the correlation with the elongation is not obvious and a different reference choice might lead to a different outcome. This is indeed confirmed by looking at conformers 18, 19 and 20, which give practically identical results for the pcseg- n basis sets, but some differences are observed for MW calculations, showing that below the requested precision, the calculations are strongly affected by random noise.

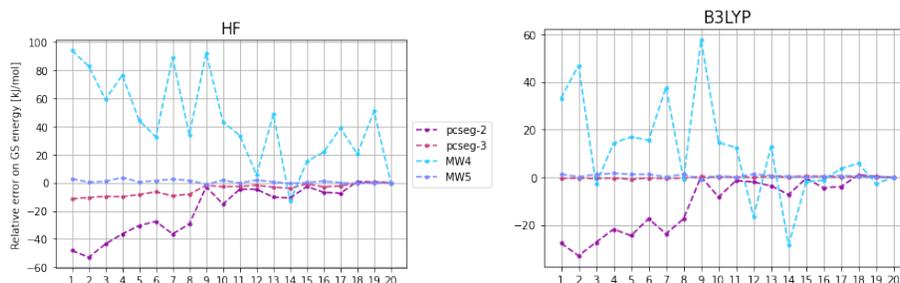


Figure 5: Error for the conformational energy of the 19 conformers relative to conformer 20. Each line in the graph refer to a different basis set or precision, with the MW7 conformational energies taken as a reference. The results at HF are shown in the left panel and the results at B3LYP level are in the right panel.

Table 2 shows the Mean Absolute Deviation, Maximum Absolute Deviation and Mean Signed Deviation (MAD, MaxAD, MSD) of the conformational energies relative to the MW7 reference results as a function of basis set quality, at the HF and B3LYP levels. The MAD and MSD are in essentially all cases of nearly the same magnitude, showing that the errors are systematic. For the whole set of conformations there is a clear trend that small basis sets strongly over-

stabilize the conformations according to their degree of compactness, as shown in Figure 3, and the MaxAD effectively measures the error between the most extended and most compact structures. The three basis sets of DZP quality, pcseg-1, Def2-SVP⁵⁹ and 6-31G(d),⁶⁰ display similar performance with MaxAD ~ 80 kJ/mol. The MaxAD is reduced to ~ 10 kJ/mol with the TZP quality pcseg-2 and Def2-TVZP⁵⁹ basis sets, while the 6-311G(2df,2pd)⁶¹ has much larger MaxAD values. The latter reflects the well-known fact that the 6-311G basis set is effectively only of DZP quality, despite it formally is a TZP style basis set.^{48,62} Note that the pcseg and Def2 basis sets perform similar at the HF level, but the pcseg has lower errors at the B3LYP, likely reflecting that the pcseg basis sets have been optimized specifically for DFT methods. Note also that the pcseg-3 basis set, being of QZP quality, is required to reduce the MaxAD below 1 kJ/mol at the B3LYP level. Including diffuse functions in the basis set is often found to reduce BSSE, but in the present case there is little difference between the results with the pcseg-1 and aug-pcseg-1, with the MaxAD only reduced from 82 to 77 kJ/mol at the HF level, and from 71 to 57 kJ/mol at the B3LYP level. The MW4 results are slightly less accurate than the pcseg-2 ones, while the MW5 results are comparable to those with the pcseg-3.

Method	HF			B3LYP		
Basis	MAD	MaxAD	MSD	MAD	MaxAD	MSD
pcseg-0	84.52	205.24	-84.20	87.33	205.70	-87.14
pcseg-1	36.60	81.72	-36.38	32.38	70.81	-32.24
Def2-SVP	37.32	86.25	-37.18	44.45	102.44	-44.34
6-31G(d)	33.59	74.78	-33.54	41.84	91.99	-41.76
pcseg-2	4.44	12.67	-4.41	2.70	7.89	-2.67
Def2-TVZP	3.50	8.72	-3.50	5.25	12.10	-5.25
6-311G(2df,2pd)	21.26	48.99	-21.17	25.02	58.47	-24.95
pcseg-3	1.16	2.72	-1.15	0.08	0.24	-0.03
pcseg-4	0.28	0.60	-0.27	0.05	0.14	0.04
MW4	10.69	22.46	10.38	3.90	13.75	2.58
MW5	0.26	0.87	0.19	0.16	0.38	0.13
MW6	0.01	0.03	-0.01	0.01	0.02	0.00

Table 2: Mean and maximum absolute deviation, and mean signed derivation (MAD, MaxAD, MSD) of relative energies in kJ/mol as a function of basis set quality, relative to MW7 reference values

The quantification of the BSSE as the difference relative to the CBS limit also includes a basis set incompleteness component, i.e. the change in relative energies due to details in the calculated electron distribution that would be present in the hypothetical situation where BSSE was absent. The energy differences in the present systems are between different conformations, where it is likely that basis set incompleteness effects are very similar for all structures, and thus differences in relative energies as a function of basis set mainly reflect intramolecular BSSE. The strong dependence on the compactness of the confor-

mations shown in Figure 3 clearly supports this view. Basis set incompleteness effects should show up as basis set dependence on inherent molecular properties not related to the energy, and we have in the present case chosen atomic charges calculated by the minimal basis set iterative Stockholder (MBIS) approach.⁶³ The basis set incompleteness is quantified by calculating the absolute change in atomic charges for each step up in basis set quality, and averaged over all 186 atoms in a given conformation. The average of this value over all 20 conformers and its standard deviation are computed, with the results shown in Table 3. The standard deviation is in all cases only a few percent of the average change, which implies that the change in atomic charges due to a change in basis set is nearly the same in all conformations. This result supports the assumption that changes in basis set incompleteness across the conformers is only a minor component in the strong dependence of the conformational relative energies on basis set quality in Table 2.

Basis change	HF		B3LYP	
	Average	Std. Dev.	Average	Std. Dev.
pcseg-0 to -1	0.01827	0.00048	0.01077	0.00042
pcseg-1 to -2	0.02247	0.00029	0.02264	0.00060
pcseg-2 to -3	0.00204	0.00010	0.00142	0.00009
pcseg-3 to -4	0.00042	0.00001	0.00015	0.00003

Table 3: Average and standard deviation over the 20 conformations between basis sets for the average absolute change in 186 atomic charges.

As mentioned in the Introduction, only a few procedures have been proposed to correct for intramolecular BSSE, and we here consider the ACP³⁶ and gCP³⁷ methods. The gCP procedure involves fitting four parameters for each combination of method and basis set to the CP corrections calculated for the $S66 \times 8$ benchmark set, and parameters have only been published for a small selection of commonly used combinations.³⁷ Table 4 shows the MAD, MaxAD and MSD for the DZ, DZP and TZP quality basis sets in their uncorrected form, and with the ACP and the gCP corrections applied. We have here employed the gCP HF and DFT versions in combinations with basis sets: SV for pcseg-0, SVP for pcseg-1 and Def2-SVP, 6-31G(d) for 6-31G(d), and TZVP for pcseg-2, Def2-TZVP and 6-311G(2df,2pd). The ACP and gCP corrections in all but one cases improve the performance, in some cases substantially, but with significant dependence on the specific basis set. The results based on pcseg-0 are, not surprisingly, still unacceptable. The ACP procedure is overall better at reducing the MaxAD than the gCP, especially at the B3LYP level, but neither can improve the performance of DZP quality basis sets to a chemical accuracy of a few kJ/mol. The MSD values show that both correction procedures tend to underestimate the BSSE effect at the HF level, but tend to overestimate the effect at the B3LYP level. The consistent improvement by the ACP and gCP procedures further strengthen the assessment that intramolecular BSSE is the leading cause of the basis set errors in relative conformational energies. The

gCP procedure has a negligible computational cost and can be used as a qualitative indicator of the amount of intramolecular BSSE, but as shown in Table 4, this should only be taken as a guideline.

Method	BSSE Corr	HF			B3LYP		
Basis		MAD	MaxAD	MSD	MAD	MaxAD	MSD
pcseg-0	none	84.52	205.24	-84.20	87.33	205.70	-87.14
pcseg-0	ACP	37.19	100.83	-33.64	16.28	58.91	-6.62
pcseg-0	gCP	50.34	125.03	-48.10	40.81	104.59	-39.15
pcseg-1	none	36.60	81.72	-36.38	32.38	70.81	-32.24
pcseg-1	ACP	15.53	37.90	-15.35	8.16	20.79	-8.08
pcseg-1	gCP	3.84	15.58	-2.39	16.52	46.98	16.48
Def2-SVP	none	37.32	86.25	-37.18	44.45	102.44	-44.34
Def2-SVP	ACP	14.51	38.08	-14.40	4.47	11.74	2.72
Def2-SVP	gCP	4.00	8.76	-3.19	4.93	17.56	4.38
6-31G(d)	none	33.59	74.78	-33.54	41.84	91.99	-41.76
6-31G(d)	ACP	4.15	10.21	-1.85	13.45	39.06	12.98
6-31G(d)	gCP	4.27	10.25	-1.11	4.61	16.11	1.60
pcseg-2	none	4.44	12.67	-4.41	2.70	7.89	-2.67
pcseg-2	ACP	0.91	2.24	0.32	1.03	2.97	-0.43
pcseg-2	gCP	0.88	2.27	0.72	5.12	11.38	5.08
Def2-TZVP	none	3.50	8.72	-3.50	5.25	12.10	-5.25
Def2-TZVP	ACP	1.12	3.19	-0.94	1.4	5.30	1.29
Def2-TZVP	gCP	1.68	5.05	1.63	2.64	8.47	2.50
6-311G(2df,2pd)	none	21.26	48.99	-21.17	25.02	58.47	-24.95
6-311G(2df,2pd)	ACP	1.93	6.79	-1.25	8.15	21.69	8.08
6-311G(2df,2pd)	gCP	16.08	36.81	-16.03	17.20	39.91	-17.19

Table 4: Mean and maximum absolute deviation, and mean signed derivation (MAD, MaxAD, MSD) of relative energies in kJ/mol as a function of basis set quality and BSSE corrections, relative to MW7 reference values

The results in Table 2 show that intramolecular BSSE can be significant for basis set of DZP quality, and may even be non-negligible at the TZP level. The error with DZP quality basis sets is strongly correlated with and of the same magnitude as the empirical D3 dispersion correction in Table 1 (correlation coefficient of 0.97 against the pcseg-1 results). Methods that do not account for dispersion, like HF and B3LYP, will thus by virtue of the BSSE include an effect that resembles dispersion. The geometry dependence of intramolecular BSSE and dispersion, however, is different, and the BSSE decreases with larger basis sets while the dispersion effect often increases. A DZP quality basis set may thus represent a sweet spot in terms of overall performance, but calculating accurate conformational energies will require accounting for both BSSE and dispersion. For DFT models with explicit dispersion corrections, the BSSE will lead to an artificial stabilization of structures with geometrically close fragments, and if used for structure optimization, will lead to optimized geometries being too

compact. As mentioned in the Introduction, the effects of intramolecular BSSE are likely larger with methods based on correlated wave functions, like MP2 and double hybrid DFT, and the present results can thus be taken as a lower bound guideline.

Figure 6 shows the absolute value of the relative error at both HF and DFT levels of theory, averaged over all conformers for each basis set and resolution. This figure allows the comparison between the overall precision offered by each basis set/resolution, as the averaging serves as a first correction to remove the superposition error inherent to Gaussian basis sets. For each increase in precision, multiwavelets bases expectedly improves precision by roughly one order of magnitude, whereas Gaussian bases offer a slower decrease in the error for each step up in cardinal number. The convergence with respect to cardinal number is exponential, as expected from Eq. 1. The precision offered by pcseg-2 is comparable to MW4, while MW5 is comparable to pcseg-4. As shown also in a recent benchmark by P. Wind *et al.*,⁸ for low/moderate precision Gaussian basis sets of TZP quality will likely be more efficient both complexity- and runtime-wise, compared to MW4, which yield similar results for relative energies. However, if precision requirements are tighter, MWs become competitive compared to modern atomic orbital bases.

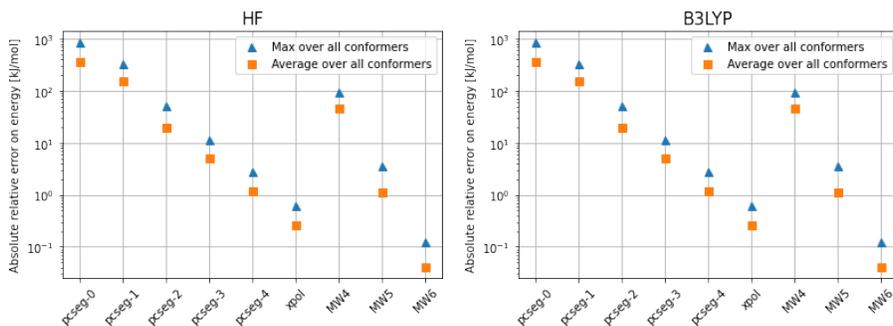


Figure 6: Absolute value of the relative error on ground state energy at HF (left) and DFT (right) levels of theory, averaged over all conformers for both Gaussian-type (pcseg- n series) and Multiwavelets (MW n series) basis sets. MW7 values taken as reference.

4 Summary

The common procedure of assessing conformational relative energies by DFT calculations using DZP or TZP quality basis sets is prone to systematic errors due to intramolecular BSSE which artificially favors compact structures. We show that the basis set dependence of relative conformational energies is strongly correlated with the molecular structure classified in terms of compactness, and the near-constant change in atomic charges between different conformations

suggests that basis set incompleteness is only a minor component. CP based correction procedures systematically improve the results, strongly suggesting that intramolecular BSSE is the leading factor in the basis set dependence of relative conformational energies. The results for a 186 atom decapeptide show that DZP quality basis set may lead to errors up to 80 kJ/mol in relative conformational energies, TZP quality basis sets bring the error down to ~ 10 kJ/mol, and QZP quality basis sets are required to bring the error down to ~ 1 kJ/mol. The overstabilization of compact structures may be further aggravated by the common procedure of optimizing structures with a medium sized basis set. The ACP and gCP procedures for estimating the intramolecular BSSE are shown to improve the results systematically, but both fail to reduce the error from DZP quality basis sets to a few kJ/mol, and only lead to a slight error reduction with TZP quality basis sets. The ACP procedure requires a non-negligible computational effort when used as *a posteriori*, while the gCP procedure employing an empirical parameterization based on overlap integrals calculated from atomic orbital models requires negligible computational time, and can be incorporated into geometry optimizations. Improved versions based on the gCP method may offer a pragmatic solution. Alternatively, replacing traditional Gaussian basis set with multiresolution methods may provide a computationally efficient way of eliminating intramolecular BSSE.

5 Author contributions

We use the CRediT taxonomy of contributor roles.^{64,65} The “Investigation” role also includes the “Methodology”, “Software”, and “Validation” roles. The “Analysis” role also includes the “Formal analysis” and “Visualization” roles. The “Funding acquisition” role also includes the “Resources” role. We visualize contributor roles in the following authorship attribution matrix, as suggested in Ref. [66].

Table 5: Levels of contribution: [major](#), [support](#). See Author Contributions for details.

	QP	PW	SRJ	LF	FJ
Conceptualization	Support			Support	Major
Investigation	Major	Support	Support	Support	Major
Data curation	Major			Support	Major
Analysis	Major			Major	Major
Supervision		Support	Support	Major	Support
Writing – original draft					Major
Writing – revisions	Major	Support	Support	Major	Major
Funding acquisition				Major	Major
Project administration					Major

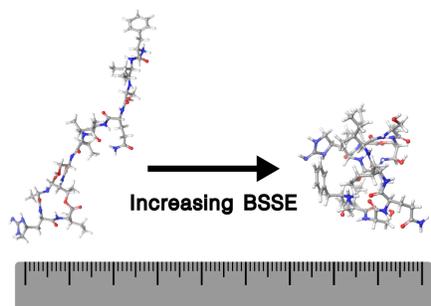
6 Acknowledgement

The numerical results using the Gaussian basis sets were obtained at the Centre for Scientific Computing, Aarhus <https://phys.au.dk/forskning/faciliteter/csaa/>. We acknowledge support from the Research Council of Norway through its Centres of Excellence scheme (262695), and from NOTUR - The Norwegian Metacenter for Computational Science through grant of computer time (nn4654k).

7 Supplementary Information

Supporting Information. Geometries of the 20 conformation (pdf) and all calculated energies (Excel). This material is available free of charge via the Internet at <http://pubs.acs.org>

8 TOC



References

- (1) Daga, L. E.; Civalleri, B.; Maschio, L. Gaussian Basis Sets for Crystalline Solids: All-Purpose Basis Set Libraries vs System-Specific Optimizations. *Journal of Chemical Theory and Computation* **2020**, *16*, 2192–2201.
- (2) Zhou, Y.; Gull, E.; Zgid, D. Material-Specific Optimization of Gaussian Basis Sets against Plane Wave Data. *Journal of Chemical Theory and Computation* **2021**, *17*, 5611–5622.
- (3) Kadek, M.; Wang, B.; Joosten, M.; Chiu, W.-C.; Mairesse, F.; Repisky, M.; Ruud, K.; Bansil, A. Band structures and Z_2 invariants of two-dimensional transition metal dichalcogenide monolayers from fully relativistic Dirac-Kohn-Sham theory using Gaussian-type orbitals. *Physical Review Materials* **2023**, *7*, 064001.

- (4) Tosoni, S.; Tuma, C.; Sauer, J.; Civalleri, B.; Ugliengo, P. A comparison between plane wave and Gaussian-type orbital basis sets for hydrogen bonded systems: Formic acid as a test case. *The Journal of Chemical Physics* **2007**, *127*.
- (5) Willand, A.; Kvashnin, Y. O.; Genovese, L.; Vázquez-Mayagoitia, Á.; Deb, A. K.; Sadeghi, A.; Deutsch, T.; Goedecker, S. Norm-conserving pseudopotentials with chemical accuracy compared to all-electron calculations. *The Journal of Chemical Physics* **2013**, *138*, 104109.
- (6) Alpert, B.; Beylkin, G.; Gines, D.; Vozovoi, L. Adaptive Solution of Partial Differential Equations in Multiwavelet Bases. *Journal of Computational Physics* **2002**, *182*, 149–190.
- (7) Harrison, R. J.; Fann, G. I.; Yanai, T.; Gan, Z.; Beylkin, G. Multiresolution quantum chemistry: Basic theory and initial applications. *The Journal of Chemical Physics* **2004**, *121*, 11587–11598.
- (8) Wind, P.; Bjørgve, M.; Brakestad, A.; S., G. A. G.; Jensen, S. R.; Eikås, R. D. R.; Frediani, L. MRChem Multiresolution Analysis Code for Molecular Electronic Structure Calculations: Performance and Scaling Properties. *Journal of Chemical Theory and Computation* **2023**, *19*, 137–146.
- (9) Jensen, F. Atomic orbital basis sets. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2013**, *3*, 273–295.
- (10) Nagy, B.; Jensen, F. Reviews in Computational Chemistry. *Reviews in Computational Chemistry* **2018**, 93–149.
- (11) Dunning, T. H. Gaussian-basis sets for use in correlated molecular calculations. 1. The atoms boron through neon and hydrogen. *Journal of Chemical Physics* **1989**, *90*, 1007–1023.
- (12) Jensen, F. Unifying General and Segmented Contracted Basis Sets. Segmented Polarization Consistent Basis Sets. *Journal of Chemical Theory and Computation* **2014**, *10*, 1074–1085.
- (13) Liu, B.; McLean, A. D. Accurate calculation of attractive interaction of two ground-state helium atoms. *Journal of Chemical Physics* **1973**, *59*, 4557–4558.
- (14) Duijneveldt, F. B. v.; Rijdt, J. G. C. M. v. D.-v. d.; Lenthe, J. H. v. State-of-the-art in counterpoise theory. *Chemical Reviews* **1994**, *94*, 1873–1885.
- (15) Boys, S. F.; Bernardi, F. Calculation of small molecular interactions by differences of separate total energies - some procedures with reduced errors. *Molecular Physics* **1970**, *19*, 553–&.
- (16) Mentel, L. M.; Baerends, E. J. Can the Counterpoise Correction for Basis Set Superposition Effect Be Justified? *Journal of Chemical Theory and Computation* **2014**, *10*, 252–267.

- (17) Alvarez-Idaboy, J. R.; Galano, A. Counterpoise corrected interaction energies are not systematically better than uncorrected ones: comparison with CCSD(T) CBS extrapolated values. *Theoretical Chemistry Accounts* **2010**, *126*, 75–85.
- (18) Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. An Assessment of Theoretical Methods for Nonbonded Interactions: Comparison to Complete Basis Set Limit Coupled-Cluster Potential Energy Curves for the Benzene Dimer, the Methane Dimer, Benzene-Methane, and Benzene-H₂S. *Journal of Physical Chemistry A* **2009**, *113*, 10146–10159.
- (19) Miliordos, E.; Apra, E.; Xantheas, S. S. Benchmark Theoretical Study of the pi-pi Binding Energy in the Benzene Dimer. *Journal of Physical Chemistry A* **2014**, *118*, 7568–7578.
- (20) Rezac, J.; Hobza, P. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. *Chemical Reviews* **2016**, *116*, 5038–5071.
- (21) Halkier, A.; Klopper, W.; Helgaker, T.; Jorgensen, P.; Taylor, P. R. Basis set convergence of the interaction energy of hydrogen-bonded complexes. *Journal of Chemical Physics* **1999**, *111*, 9157–9167.
- (22) Elm, J.; Kristensen, K. Basis set convergence of the binding energies of strongly hydrogen-bonded atmospheric clusters. *Physical Chemistry Chemical Physics* **2017**, *19*, 1122–1133.
- (23) Burns, L. A.; Marshall, M. S.; Sherrill, C. D. Comparing Counterpoise-Corrected, Uncorrected, and Averaged Binding Energies for Benchmarking Noncovalent Interactions. *Journal of Chemical Theory and Computation* **2014**, *10*, 49–57.
- (24) Brauer, B.; Kesharwani, M. K.; Martin, J. M. L. Some Observations on Counterpoise Corrections for Explicitly Correlated Calculations on Noncovalent Interactions. *Journal of Chemical Theory and Computation* **2014**, *10*, 3791–3799.
- (25) Daudey, J. P.; Claverie, P.; Malrieu, J. P. Perturbative ab-initio calculations of intermolecular energies .1. method. *International Journal of Quantum Chemistry* **1974**, *8*, 1–15.
- (26) Witte, J.; Neaton, J. B.; Head-Gordon, M. Push it to the limit: Characterizing the convergence of common sequences of basis sets for intermolecular interactions as described by density functional theory. *The Journal of Chemical Physics* **2016**, *144*, 194306.
- (27) Jensen, F. The magnitude of intramolecular basis set superposition error. *Chemical Physics Letters* **1996**, *261*, 633–636.
- (28) Reiling, S.; Brickmann, J.; Schlenkrich, M.; Bopp, P. A. Theoretical investigations on 1,2-ethanediol: The problem of intramolecular hydrogen bonds. *Journal of Computational Chemistry* **1996**, *17*, 133–147.

- (29) Senent, M. L.; Wilson, S. Intramolecular basis set superposition errors. *International Journal of Quantum Chemistry* **2001**, *82*, 282–292.
- (30) Balabin, R. M. Enthalpy difference between conformations of normal alkanes: Effects of basis set and chain length on intramolecular basis set superposition error. *Molecular Physics* **2011**, *109*, 943–953.
- (31) Sladek, V.; Holka, F.; Tvaroska, I. Ab initio modelling of the anomeric and exo anomeric effects in 2-methoxytetrahydropyran and 2-methoxythiane corrected for intramolecular BSSE. *Physical Chemistry Chemical Physics* **2015**, *17*, 18501–18513.
- (32) Holroyd, L. F.; Mourik, T. v. Insufficient description of dispersion in B3LYP and large basis set superposition errors in MP2 calculations can hide peptide conformers. *Chemical Physics Letters* **2007**, *442*, 42–46.
- (33) Shields, A. E.; Mourik, T. v. Comparison of ab initio and DFT electronic structure methods for peptides containing an aromatic ring: Effect of dispersion and BSSE. *Journal of Physical Chemistry A* **2007**, *111*, 13272–13277.
- (34) Valdes, H.; Klusak, V.; Pitonak, M.; Exner, O.; Stary, I.; Hobza, P.; Rulisek, L. Evaluation of the intramolecular basis set superposition error in the calculations of larger molecules: n Helicenes and Phe-Gly-Phe tripeptide. *Journal of Computational Chemistry* **2008**, *29*, 861–870.
- (35) Galano, A.; Alvarez-Idaboy, J. R. A new approach to counterpoise correction to BSSE. *Journal of Computational Chemistry* **2006**, *27*, 1203–1210.
- (36) Jensen, F. An Atomic Counterpoise Method for Estimating Inter- and Intramolecular Basis Set Superposition Errors. *Journal of Chemical Theory and Computation* **2010**, *6*, 100–106.
- (37) Kruse, H.; Grimme, S. A geometrical correction for the inter- and intramolecular basis set superposition error in Hartree-Fock and density functional theory calculations for large systems. *Journal of Chemical Physics* **2012**, *136*, 154101.
- (38) Hamza, A.; Vibok, A.; Halasz, G. J.; Mayer, I. BSSE-free SCF theories: a comment. *Journal of Molecular Structure-Theochem* **2000**, *501*, 427–434.
- (39) Mayer, I.; Vibok, A. BSSE-free second-order intermolecular perturbation theory. *Molecular Physics* **1997**, *92*, 503–510.
- (40) Rasmussen, T. H.; Wang, Y. M.; Kjaergaard, T.; Kristensen, K. The explicitly correlated same number of optimized parameters (SNOOP-F12) scheme for calculating intermolecular interaction energies. *Journal of Chemical Physics* **2016**, *144*, 204102.
- (41) Kristensen, K.; Ettenhuber, P.; Eriksen, J. J.; Jensen, F.; Jorgensen, P. The same number of optimized parameters scheme for determining intermolecular interaction energies. *Journal of Chemical Physics* **2015**, *142*, 114116.

- (42) Jensen, F. Using valence bond methods to estimate intramolecular basis set superposition errors. *Journal of Chemical Physics* **2017**, *146*, 184109.
- (43) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. Accurate ab Initio Quantum Chemical Determination of the Relative Energetics of Peptide Conformations and Assessment of Empirical Force Fields. *Journal of the American Chemical Society* **1997**, *119*, 5908–5920.
- (44) DiStasio, R. A.; Jung, Y.; Head-Gordon, M. A Resolution-Of-The-Identity Implementation of the Local Triatomics-In-Molecules Model for Second-Order Møller-Plesset Perturbation Theory with Application to Alanine Tetrapeptide Conformational Energies. *Journal of Chemical Theory and Computation* **2005**, *1*, 862–876.
- (45) Culka, M.; Kalvoda, T.; Gutten, O.; Rulíšek, L. Mapping Conformational Space of All 8000 Tripeptides by Quantum Chemical Methods: What Strain Is Affordable within Folded Protein Chains? *The Journal of Physical Chemistry B* **2021**, *125*, 58–69.
- (46) Yanai, T.; Fann, G. I.; Gan, Z.; Harrison, R. J.; Beylkin, G. Multiresolution quantum chemistry in multiwavelet bases: Hartree–Fock exchange. *The Journal of Chemical Physics* **2004**, *121*, 6680–6688.
- (47) Jensen, S. R.; Saha, S.; Flores-Livas, J. A.; Huhn, W.; Blum, V.; Goedecker, S.; Frediani, L. The Elephant in the Room of Density Functional Theory Calculations. *The Journal of Physical Chemistry Letters* **2017**, *8*, 1449–1457.
- (48) Brakestad, A.; Wind, P.; Jensen, S. R.; Frediani, L.; Hopmann, K. H. Multiwavelets applied to metal–ligand interactions: Energies free from basis set errors. *The Journal of Chemical Physics* **2021**, *154*, 214302.
- (49) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab-initio calculation of vibrational absorption and circular-dichroism spectra using density-functional force-fields. *Journal of Physical Chemistry* **1994**, *98*, 11623–11627.
- (50) Halgren, T. A. Merck molecular force field .1. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry* **1996**, *17*, 490–519.
- (51) 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.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams, J.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Jr., J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.;

- Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*.
- (52) Bast, R.; Bjørgve, M.; Brakestad, A.; Di Remigio, R.; Frediani, L.; Gerez, G.; Jensen, S. R.; Wind, P. MRChem: MultiResolution Chemistry, <https://github.com/MRChemSoft/mrchem/tree/v1.1.0>, version v1.1.0, 2022.
- (53) Alpert, B. K. A Class of Bases in \mathbb{L}^2 for the Sparse Representation of Integral Operators. *SIAM Journal on Mathematical Analysis* **1993**, *24*, 246–262.
- (54) Öster, C.; Walkowiak, G. P.; Hughes, D. E.; Spoering, A. L.; Peoples, A. J.; Catherwood, A. C.; Tod, J. A.; Lloyd, A. J.; Herrmann, T.; Lewis, K.; Dowson, C. G.; Lewandowski, J. R. Structural studies suggest aggregation as one of the modes of action for teixobactin. *Chemical Science* **2018**, *9*, 8850–8859.
- (55) Klopper, W.; Kutzelnigg, W. Gaussian basis sets and the nuclear cusp problem. *Journal of Molecular Structure: THEOCHEM* **1986**, *135*, 339–356.
- (56) McKemmish, L. K.; Gill, P. M. W. Gaussian Expansions of Orbitals. *Journal of Chemical Theory and Computation* **2012**, *8*, 4891–4898.
- (57) Jensen, F. Estimating the Hartree–Fock limit from finite basis set calculations. *Theoretical Chemistry Accounts* **2005**, *113*, 267–273.
- (58) 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. *The Journal of Chemical Physics* **2010**, *132*, 154104.
- (59) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- (60) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *The Journal of Chemical Physics* **2003**, *56*, 2257–2261.
- (61) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *The Journal of Chemical Physics* **2008**, *72*, 650–654.
- (62) Grev, R. S.; Schaefer, H. F. 6-311G is not of valence triple-zeta quality. *The Journal of Chemical Physics* **1989**, *91*, 7305–7306.
- (63) Verstraelen, T.; Vandenbrande, S.; Heidar-Zadeh, F.; Vanduyfhuys, L.; Van Speybroeck, V.; Waroquier, M.; Ayers, P. W. Minimal Basis Iterative Stockholder: Atoms in Molecules for Force-Field Development. *Journal of Chemical Theory and Computation* **2016**, *12*, 3894–3912.

- (64) Allen, L.; Scott, J.; Brand, A.; Hlava, M.; Altman, M. Publishing: Credit where credit is due. *Nature* **2014**, *508*, 312–313.
- (65) Brand, A.; Allen, L.; Altman, M.; Hlava, M.; Scott, J. Beyond authorship: attribution, contribution, collaboration, and credit. *Learned Publishing* **2015**, *28*, 151–155.
- (66) Tay, A. Researchers are embracing visual tools to give fair credit for work on papers. <https://www.natureindex.com/news-blog/researchers-embracing-visual-tools-contribution-matrix-give-fair-credit-authors-scientific-papers>.