

RESEARCH ARTICLE | JUNE 23 2023

A quantitative assessment of deformation energy in intermolecular interactions: How important is it? ✓

Caroline T. Sargent ; Raina Kasera ; Zachary L. Glick ; C. David Sherrill ; Daniel L. Cheney  



J. Chem. Phys. 158, 244106 (2023)

<https://doi.org/10.1063/5.0155895>



CrossMark

500 kHz or 8.5 GHz?
And all the ranges in between.

Lock-in Amplifiers for your periodic signal measurements



Find out more



A quantitative assessment of deformation energy in intermolecular interactions: How important is it?

Cite as: J. Chem. Phys. 158, 244106 (2023); doi: 10.1063/5.0155895

Submitted: 24 April 2023 • Accepted: 1 June 2023 •

Published Online: 23 June 2023



Caroline T. Sargent,¹  Raina Kasera,²  Zachary L. Glick,¹  C. David Sherrill,¹  and Daniel L. Cheney^{2,a)} 

AFFILIATIONS

¹ Center for Computational Molecular Science and Technology, School of Chemistry and Biochemistry, School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0400, USA

² Molecular Structure and Design, Bristol Myers Squibb Company, P.O. Box 5400, Princeton, New Jersey 08543, USA

^{a)} Author to whom correspondence should be addressed: danlcheney@gmail.com

ABSTRACT

Dimer interaction energies have been well studied in computational chemistry, but they can offer an incomplete understanding of molecular binding depending on the system. In the current study, we present a dataset of focal-point coupled-cluster interaction and deformation energies (summing to binding energies, D_e) of 28 organic molecular dimers. We use these highly accurate energies to evaluate ten density functional approximations for their accuracy. The best performing method (with a double- ζ basis set), B97M-D3BJ, is then used to calculate the binding energies of 104 organic dimers, and we analyze the influence of the nature and strength of interaction on deformation energies. Deformation energies can be as large as 50% of the dimer interaction energy, especially when hydrogen bonding is present. In most cases, two or more hydrogen bonds present in a dimer correspond to an interaction energy of -10 to -25 kcal mol⁻¹, allowing a deformation energy above 1 kcal mol⁻¹ (and up to 9.5 kcal mol⁻¹). A lack of hydrogen bonding usually restricts the deformation energy to below 1 kcal mol⁻¹ due to the weaker interaction energy.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0155895>

I. INTRODUCTION

The interaction energy of a dimer is defined as the energy difference between a dimer, AB , and its two monomers, A and B ,

$$\Delta E_{\text{int}}(AB) = E(AB) - E(A) - E(B). \quad (1)$$

Calculating interaction energies between molecules has led to advances in many areas, such as materials science and medicine.^{1–3} For example, knowledge of how a potential drug molecule engages a target protein can be leveraged to optimize its strength of binding and therapeutic activity.^{2,3} Often, interaction energies are calculated by Eq. (1). This is referred to as the supermolecular approach and is typically used due to the flexibility in the choice of methods. Any quantum chemical method (density functional or wavefunction) can be used for the three energy calculations, depending on the size of the system and the desired accuracy. Another approach is to compute the interaction energy directly with the Symmetry-Adapted Perturbation Theory (SAPT).^{4–6} An advantage of this method is that

it returns not only the interaction energy but also its four physically relevant components: dispersion, electrostatics, exchange, and induction.

There is often a trade-off between accuracy and computational cost, and computing interaction energies is not an exception. A number of studies have produced highly accurate interaction energies on small systems, and there now exist a handful of datasets such as S22,⁷ S66,⁸ X40,⁹ SSI,¹⁰ HSG,¹¹ HBC6,¹² and NBC10.^{13–15} These datasets consist of small, organic dimers with both hydrogen bonding and van der Waals intermolecular interactions. In addition to calculating accurate interaction energies, many groups have analyzed the performance of approximate, lower cost methods with these datasets.^{16–19} These studies inform chemists of the best lower cost method to use for their system and the confidence they can hold in their answer.

Strictly speaking, an interaction energy is defined by evaluating $E(A)$ and $E(B)$ at the geometries monomers A and B take within the dimer. In many cases, monomers deform their geometries (to a less energetically favorable state relative to their isolated, relaxed geometries) in order to achieve a stronger intermolecular interaction

in the dimer. The energy due to this change in geometry is often referred to as the “deformation energy,”

$$\Delta E_{\text{def}}(A) = E_{AB}(A) - E_A(A), \quad (2)$$

where the symbol in parentheses denotes the system considered, and the subscript denotes the geometry. ΔE_{def} must be included if one wishes to compute the entire energetic change as two isolated monomers form a dimer.

If we include the deformation energy, we obtain the electronic binding energy,

$$D_e = -[\Delta E_{\text{int}}(AB) + \Delta E_{\text{def}}(A) + \Delta E_{\text{def}}(B)]. \quad (3)$$

Note that the electronic binding energy D_e is considered to be a positive quantity. Because the interaction energy for a favorable interaction in Eq. (1) is negative, a minus sign is required in Eq. (3). The electronic binding energy is an energy that must be overcome to dissociate the dimer; therefore, it can also be considered to be the electronic component of the dissociation energy. To match an experimental dissociation energy at 0 K, we must also include the contribution from the change in the zero-point vibrational energy (ZPVE) caused by dissociating the dimer into monomers. Adding this contribution $\Delta ZPVE$ to the electronic binding energy D_e yields the ZPVE-corrected dissociation energy,

$$D_0 = D_e + \Delta ZPVE, \quad (4)$$

where

$$\Delta ZPVE = ZPVE(A) + ZPVE(B) - ZPVE(AB). \quad (5)$$

Note that $\Delta ZPVE$ has been defined from the point of view of the dissociation process (i.e., the final state is the dissociated products and the initial state is the dimer), whereas the interaction energy is defined in the opposite sense in Eq. (1), consistent with ΔE_{int} and D_0 having opposite signs.

The Cremer group highlighted the case of a formic acid dimer, where the most accurate experimental dissociation energy (D_0) differed from the best theoretical interaction energy (ΔE_{int}) by as much as 4.5 kcal mol^{−1} because of the lack of deformation and $\Delta ZPVE$ terms.²⁰ Consequently, they calculated a binding energy using coupled-cluster with single, double, and perturbative triple excitations [CCSD(T)]²¹ and an extrapolation using augmented, correlation-consistent quadruple and pentuple- ζ basis sets, aug-cc-pVQZ and aug-cc-pV5Z, using geometries and frequencies obtained from CCSD(T) with basis set extrapolation. In addition, anharmonic corrections were computed with density functional theory (DFT) at the B3LYP/aug-cc-pVTZ level. Their resulting value was within the error bars of the best experimental value.²⁰

Theoretically determined dissociation energies (D_0) have been published for a number of other dimers, such as H₂O:HF, (HF)₂, 1-naphthol complexes, (H₂S)₂, and HCN:HF.^{22–26} Dimers such as these are chosen as test cases because they include some interesting interaction motifs, like hydrogen bonding, while their small size allows the use of highly accurate levels of theory. While high-accuracy studies are useful and important, other methods that are lower cost should be tested, especially on larger systems. Very recently, Czerneková and co-workers studied mid-sized organic dimers and found that CCSD(T) energies [within the complete basis set (CBS)

limit] of structures optimized with second-order Møller–Plesset perturbation theory (MP2)²⁷/aTZ produced dissociation energies, D_0 , within 0.36 kcal mol^{−1} of the experiment for most cases studied.²⁸ Dissociation energies were assessed with DFT methods and DFT-based SAPT (DFT-SAPT).^{29,30} The DFT-SAPT/a[TQ]Z interaction energies with MP2/aTZ $\Delta ZPVE$ and deformation energies produced results in reasonable agreement to experiment—the maximum error was 0.26 kcal mol^{−1}.²⁸ In addition, Halder *et al.* published CCSD(T)/CBS dissociation energies (D_0) of 11 hydrogen-bound and 11 dispersion-bound dimers with up to 23 atoms. Dissociation energies of the hydrogen-bound dimers were also calculated with a resolution of the identity (RI)-MP2, and the mean relative error, relative to the experiment, was 12.3%, compared to 6.2% with CCSD(T)/CBS. For the dispersion-bound systems, using B97-D3 returned a mean relative error of 7.7%, slightly worse than 6.2% with CCSD(T)/CBS.³¹ Among the conclusions, it was recommended that calculating the $\Delta ZPVE$ is absolutely necessary to compare computational results to experimental measurements.³¹

Unfortunately, the $\Delta ZPVE$ term is the most computationally expensive step in calculating D_0 , if geometries and vibrational frequencies are obtained using the same level of theory as the interaction energy. Because ΔE_{int} itself can be difficult to compute accurately, due to significant electron correlation and basis set requirements,¹³ numerous theoretical studies have considered ΔE_{int} only, not D_e or D_0 . Such studies (e.g., Refs. 7, 10, 17, and 32–37) have shed light on the accuracy of various theoretical methods for computing the ΔE_{int} component of D_0 , which one might expect to be the most theoretically demanding component: in computational quantum chemistry, it is a common practice to obtain geometries and vibrational frequencies at lower levels of theory than that used to obtain electronic energy differences. Numerous studies focusing primarily on ΔE_{int} have also provided insight into the relative strength of different types of intermolecular interactions,^{10,13,38–41} how they depend on intermolecular geometry,^{13,14,33,42,43} and how they are affected by substituent effects and heteroatoms.^{44–50} Despite these advances and insights, however, direct comparison against experiment remains elusive in the absence of a consideration of deformation energies and ZPVEs.

By comparison to the many theoretical studies focusing on ΔE_{int} , very few studies of van der Waals dimers have explored what theoretical methods perform well for computing geometries of van der Waals clusters, deformation energies, and ZPVEs, and how large the latter two quantities are for typical dimers or how they depend on the chemical nature of the monomers. Regarding geometry optimization of van der Waals dimers, Řezáč and Hobza used a high-level CCSD(T)/CBS scheme to optimize 24 dimers that were either hydrogen bound, dispersion dominated, or a mixture of the two, and had up to four heavy atoms.⁵¹ Following the publication, a subset of 21 of these dimers (excluding three non-equilibrium geometries) was used to evaluate approximate DFT methods with aDZ⁵² and aTZ⁵³ basis sets. B97-D3/aDZ⁵² and ω B97X-V/aTZ⁵³ provided the lowest average root-mean squared deviations (RMSDs) for this dataset, relative to the many other density functional approximations (DFAs) tested. Datasets that include larger systems have also been studied.^{18,53,54} These studies agree that while some DFAs yield reasonable geometries compared to reference values, there is no one DFA that most accurately optimizes dimers of various system sizes and interactions.^{18,53,54} Beyond these studies, Halder

et al. have shown that dispersion-corrected DFT (DFT-D) can provide reasonable values for D_e at a relatively low computational cost for a quantum chemistry method.³¹

The aim of this work is to better understand the accuracy of various quantum chemistry methods for geometry optimization of van der Waals complexes by computing highly accurate geometries and binding energies for 28 organic dimers with up to 28 atoms. With these reference values, we examine many density functional approximations for their ability to optimize geometries and calculate interaction and deformation energies accurately and efficiently. Finally, using one of the well-performing DFAs, we obtain deformation energies for a diverse set of 104 van der Waals dimers and analyze the results to identify structural properties that result in larger deformation energies.

II. ANALYSIS OF DENSITY FUNCTIONAL APPROXIMATIONS FOR BINDING ENERGIES

A. Reference calculations

1. Focal-point methods

High-accuracy data, to be used as a reference when evaluating the density functional approximations, was obtained for the 28 dimers shown in Fig. 1. This dataset is composed of subsets of the S22,⁷ S66,⁸ and X40⁹ datasets.

Initially, dimers were optimized with a focal-point^{55,56} (FP) scheme meant to approximate coupled-cluster with single, double, and perturbative triple excitations [CCSD(T)]²¹ in the complete basis set (CBS) limit. This technique has been successful with correlation-consistent basis sets for noncovalent interactions.^{17,57,58} CCSD(T) in a large basis set may be approximated by

$$E(\text{CCSD(T)}/\text{Large}) \approx E(\text{MP2}/\text{Large}) + \delta_{\text{MP2}}^{\text{CCSD(T)}}/\text{Small}, \quad (6)$$

because smaller basis sets are often sufficient to capture higher order electron correlation effects. In Eq. (6), “large” and “small” refer to the relative sizes of the basis sets, and the delta term is

$$\delta_{\text{MP2}}^{\text{CCSD(T)}}/\text{Small} = E(\text{CCSD(T)}/\text{Small}) - E(\text{MP2}/\text{Small}). \quad (7)$$

A two-point extrapolation of the correlation energy typically provides a sufficient estimate of the CBS limit.⁵⁹ The large basis in Eq. (6) uses a two-point extrapolation of Dunning’s augmented, correlation-consistent quadruple and pentuple- ζ basis sets, aug-cc-pVQZ and aug-cc-pV5Z,^{60,61} abbreviated aQZ and a5Z, respectively. We tested two small basis sets for $\delta_{\text{MP2}}^{\text{CCSD(T)}}/\text{Small}$: aug-cc-pVDZ (aDZ) and cc-pVDZ (DZ). For brevity, we will refer to these specific CCSD(T)/CBS schemes as CCSD(T)/CBS(a[Q5]Z; δ :aDZ) and CCSD(T)/CBS(a[Q5]Z; δ :DZ).

After dimers were optimized with an FP scheme, interaction energies were calculated directly with a high-level symmetry-adapted perturbation theory, SAPT2+(3) δ MP2,⁶² in Dunning’s aug-cc-pVTZ basis set. SAPT2+(3) is usually reliable for dispersion-bound complexes, but for electrostatically bound systems, an MP2 correction (δ MP2) is added to account for higher order coupling of induction and dispersion. This correction is defined as

$$\delta E_{\text{MP2}} = E_{\text{int}}^{\text{MP2}} - E_{\text{int}}^{\text{SAPT2}}. \quad (8)$$

SAPT2+(3) δ MP2/aug-cc-pVTZ has previously been described as the “gold-standard” of SAPT.⁶² This level of SAPT, similar in quality to CCSD(T) for intermolecular interactions,⁶² uses an aTZ basis set and, thus, should be of similar or possibly better quality than the focal point approaches used for geometry optimization because the coupled-cluster part of those computations was limited to the DZ or aDZ basis sets. Although not analyzed in detail here, SAPT has the advantage over CCSD(T) of providing physically meaningful energy components, such as electrostatics, induction/polarization, etc.

After optimizing the dimers and calculating their interaction energies, the deformation term for each monomer was calculated according to Eq. (2). The starting geometries for the monomer relaxations were the monomer geometries in the optimized dimer configuration. Thus, the optimized, relaxed monomers are in a nearby local minimum, which might not be the global minimum, and the resulting deformation energies may be underestimates. The same focal-point CCSD(T)/CBS scheme that was used for the dimer optimization was used for all parts of the monomer deformation calculation.

As mentioned previously, two FP schemes were used. We initially tried to optimize the dimer geometries with CCSD(T)/CBS(a[Q5]Z; δ :aDZ) and the counterpoise (CP) correction of Boys and Bernardi.^{63,64} This high-level method was extremely computationally expensive, and we could only calculate binding energies of 22/28 dimers. Thus, we calculated binding energies (and optimized geometries) with a slightly cheaper level of theory: CCSD(T)/CBS(a[Q5]Z; δ :DZ), without CP correction. The omission of augmented basis sets in the coupled-cluster correction term and lack of CP correction during the dimer optimization allowed for the calculation of binding energies for all 28 dimers. Then, interaction energies were calculated with SAPT2+(3) δ MP2/aTZ. Finally, the deformation terms were calculated with CCSD(T)/CBS(a[Q5]Z; δ :DZ). A development version of Psi4 1.4⁶⁵ was used for all computations.

2. Results: Comparing the two focal-point schemes

To compare the dimer geometries optimized via CCSD(T)/CBS(a[Q5]Z; δ :DZ) without CP correction vs CCSD(T)/CBS(a[Q5]Z; δ :aDZ) with CP correction, root-mean squared deviations (RMSDs) were calculated by aligning heavy atoms and polar hydrogens. The optimal RMSDs are listed in the supplementary material. The RMSDs are generally quite low (≤ 0.02 Å). Methylamine:peptide dimer is representative, wherein the geometries are nearly identical and the RMSD is 0.02 Å (Fig. 2). The largest RMSD is observed for the benzene:hydrogen cyanide dimer (0.09 Å), reflecting deviations in the intermolecular distance (0.6 Å) and the NHC angle (20°).

Figure 3 shows how these differences in geometry translate to errors in binding energies and SAPT interaction energies when the interaction energy part of the binding energy is computed at the fixed SAPT2+(3) δ MP2/aTZ level of theory. The absolute errors in the binding energies when using CCSD(T)/CBS(a[Q5]Z; δ :DZ) (without CP) are less than 0.06 kcal mol⁻¹ relative to CCSD(T)/CBS(a[Q5]Z; δ :aDZ) (with CP) in all cases and generally less than 0.02 kcal mol⁻¹. The largest errors are for acetic acid (0.05 kcal mol⁻¹), formic acid (0.06 kcal mol⁻¹), and formamide (0.06 kcal mol⁻¹). Compared to the other dimers in the dataset, these have relatively large binding energies, given in Table I. For

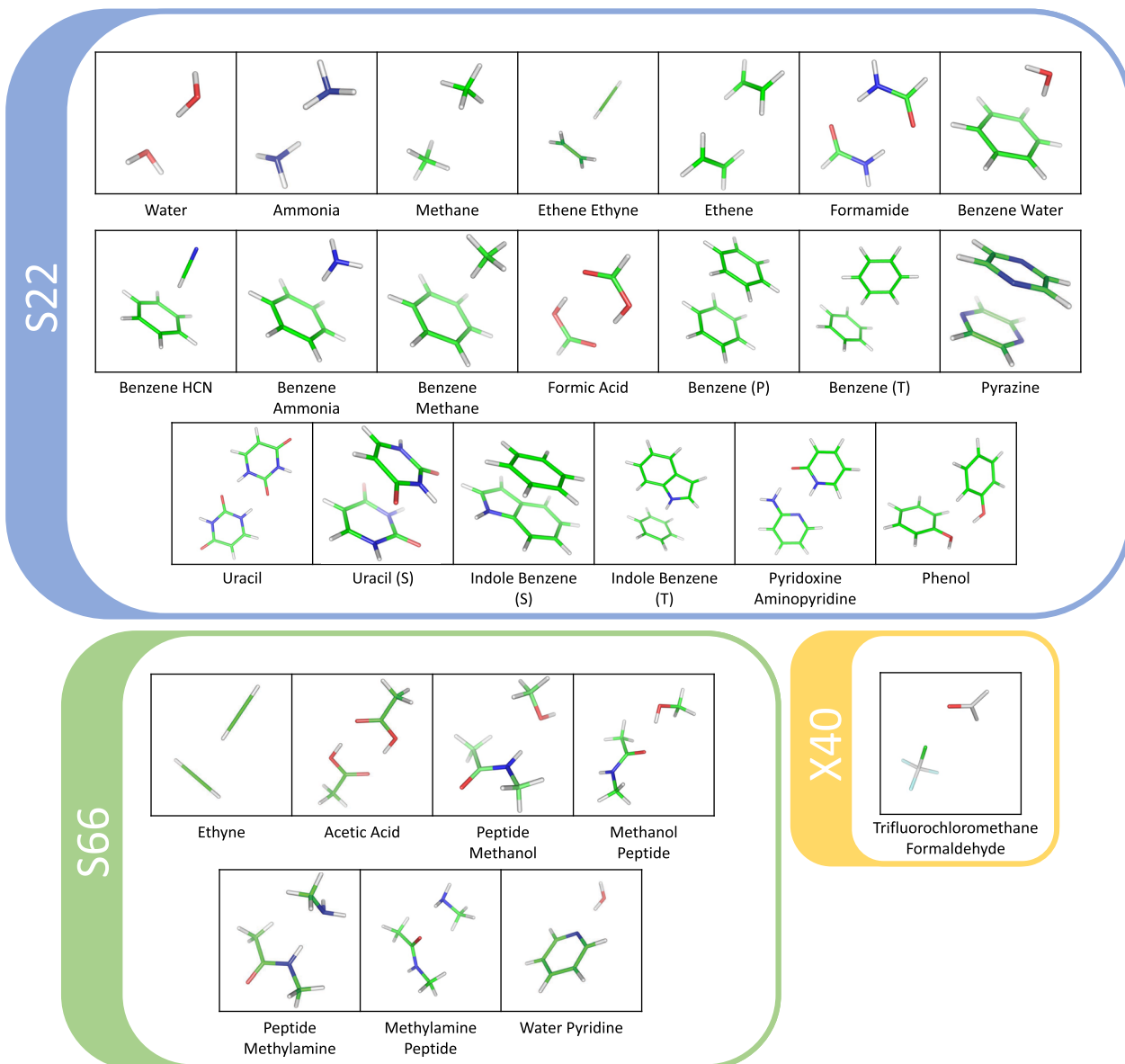


FIG. 1. Dataset of 28 dimers, composed of subsets of S22, S66, and X40, used to analyze the ability of different density functionals to compute binding energies accurately and efficiently.

each dimer, the errors in the binding energies result mainly from the errors in interaction energies caused by using different dimer geometries.

Due to only slight differences in binding energies and geometries between the two focal-point schemes, we elected to omit the counterpoise correction and augmented basis functions (for $\delta_{MP2}^{CCSD(T)}$) for geometry optimizations and binding energy calculations. The reduced computational cost enabled us to expand our dataset from the initial set of 22 dimers to 28. Table 1 reports the interaction energies (FP and SAPT), along with the deformation and

binding energies for each of the 28 dimers. For a breakdown of the SAPT components, see the supplementary material.

B. Density functional approximations

1. Methods

A variety of density functional approximations was analyzed regarding their ability to obtain accurate binding energies relative to high-quality reference data for the 28 dimers in Fig. 1. This reference data, as discussed in the prior section, is

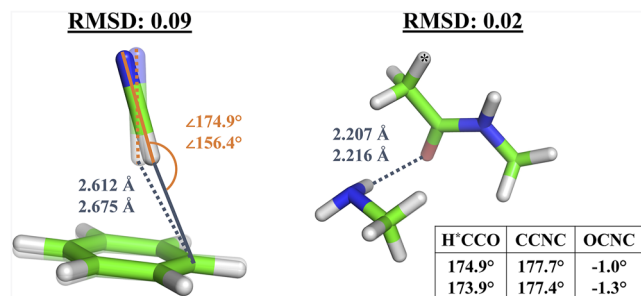


FIG. 2. Overlay of dimers optimized with CCSD(T)/CBS(a[Q5]Z; δ :DZ) without counterpoise correction (opaque) and CCSD(T)/CBS(a[Q5]Z; δ :aDZ) with counterpoise correction (transparent). Largest changes in the geometries are noted.

the combination of a SAPT2+(3) δ MP2/aTZ interaction energy and CCSD(T)/CBS(a[Q5]Z; δ :DZ) deformation energies, with all geometry optimizations at the latter level of theory. No counterpoise correction is used. The density functional approximations considered were mainly chosen based on existing studies that revealed reasonable results for binding or conformational energies.^{66–71} In this study, we include B3LYP-D3MBJ,⁷² B97-D3MBJ,⁷³ B97M-D3BJ,⁷⁴ M06-2X,⁷⁵ PBE0,⁷⁶ PBE0-D3MBJ,⁷⁶ ω B97M-D3BJ,⁷⁷ ω B97X-D3BJ,^{78,79} ω PBE,^{80,81} and ω PBE-D3MBJ. An M appended to a standard functional denotes a meta-GGA, and an X represents a hybrid functional. A prefix of ω symbolizes range-separation of exact (Hartree-Fock) and density functional exchange. D3BJ is the D3 dispersion correction of Grimme *et al.* that includes the Becke-Johnson damping function.⁸² D3MBJ modifies D3BJ by fitting the parameters using an expanded training set that includes more interaction types and a more balanced coverage of intermolecular distances.⁸³ M06-2X is the only method for which a dispersion correction has not been included. Adding a dispersion correction to Minnesota functionals (the family to which M06-2X belongs) can lead to, and sometimes worsen, overbinding in hydrogen-bound systems.⁸⁴ It has also been shown that adding a dispersion correction (D3) to PBE0 can sometimes produce greater errors in conformational and interaction energies than using PBE0 alone.^{12,85} Therefore, PBE0 and ω PBE are included with and without dispersion corrections.

Dunning's correlation-consistent augmented double and triple- ζ basis sets (aug-cc-pVDZ and aug-cc-pVTZ) have been used with each functional, and these will be abbreviated aDZ and aTZ, respectively. A development version of Psi4 1.4 was used for all computations, as well as Psi4's default grid (75 radial and 302 spherical points).

2. Results: Performance of the density functionals

Figure 4 reports the wall times of calculating single-point energies of pyridine polymers with each approximate level of theory. B97, PBE0, and B3LYP are the three fastest methods with both basis sets, and ω B97M, M06-2X, and B97M are the most computationally expensive. The results are not too surprising as the three cheaper methods are GGAs and hybrid GGAs, whereas the three most expensive are meta-GGAs. The range-separated functionals appear to be more sensitive to basis set size than other DFAs. ω PBE-D3MBJ

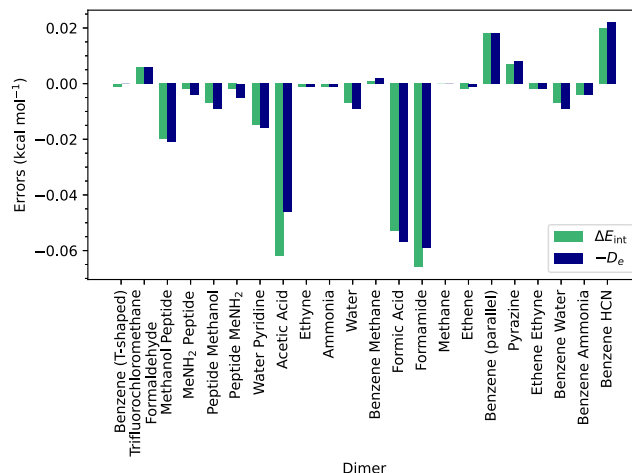


FIG. 3. Errors in interaction and binding energies. Interaction energy errors are of SAPT2+(3) δ MP2/aTZ // CCSD(T)/CBS(a[Q5]Z; δ :DZ) (without counterpoise correction) relative to SAPT2+(3) δ MP2/aTZ // CCSD(T)/CBS(a[Q5]Z; δ :aDZ) (with counterpoise correction). Binding energies include deformation energies calculated with the method of optimization.

behaves similarly to GGAs with an aDZ basis set, but similar to meta-GGAs with an aTZ basis set. ω B97M-D3BJ is the most expensive method for both basis sets, and the use of aTZ seems to promote especially poor timings.

Absolute errors in binding energies calculated with the various density functional approximations, relative to the reference method, are shown in the top left panel of Fig. 5. To compute each point, the DFT method was used to optimize each dimer, calculate the interaction energy, then optimize each monomer and calculate the deformation energies. After summing the interaction energy and deformation energies, the difference between this DFT binding energy and the reference binding energy was plotted. Individual errors for each of the 28 dimers are represented by dots, and squares show mean absolute errors. For a comprehensive list of errors, see the supplementary material.

Especially large errors in the binding energy occur when using PBE0 and ω PBE without dispersion corrections. More surprising is that both the maximum and mean absolute errors increase for these functionals when aTZ basis sets are used rather than aDZ. These DFAs return mean absolute errors of 0.84 and 1.11 kcal mol⁻¹ with aDZ basis sets and MAEs of 1.11 and 1.47 kcal mol⁻¹ with aTZ basis sets. The maximum magnitude errors for PBE0/aDZ and PBE0/aTZ are -3.25 and -3.67 kcal mol⁻¹, respectively. These both occur for pyrazine. Similarly, the largest errors for ω PBE are due to pyrazine. This is especially concerning as pyrazine has a binding energy of just 4.55 kcal mol⁻¹; these large errors will be explored further later. Other large errors are present, and for PBE0, the absolute errors greater than 1.5 kcal mol⁻¹ are mainly for dispersion-bound dimers, but the phenol dimer (hydrogen-bound) does incur an error around -2 kcal mol⁻¹ for both basis sets. ω PBE/aDZ produces absolute errors greater than 1.5 kcal mol⁻¹ for mainly dispersion-bound dimers, but using an aTZ basis set causes errors above 1.5 kcal mol⁻¹ for both dispersion and electrostatically bound dimers. Adding the dispersion correction D3MBJ to PBE0 reduces the errors for aTZ

TABLE I. Reference energy data for the dimers (and monomers) optimized with CCSD(T)/CBS(a[Q5]Z; δ :DZ): interaction energies with CCSD(T)/CBS(a[Q5]Z; δ :DZ) and no CP correction, interaction energies with SAPT2+(3) δ MP2/aug-cc-pVTZ, monomer A and B deformation energies with CCSD(T)/CBS(a[Q5]Z; δ :DZ), and binding energies using the SAPT interaction energies. All energies are in kcal mol^{−1}.

Dimer	$\Delta E_{\text{int}}^{\text{FP}}$	$\Delta E_{\text{int}}^{\text{SAPT}}$	$\Delta E_{\text{def}}^{\text{A}}$	$\Delta E_{\text{def}}^{\text{B}}$	D_e
Acetic acid dimer	−19.18	−19.45	1.35	1.35	16.75
Ammonia dimer	−2.93	−3.00	0.01	0.01	2.99
Benzene dimer (parallel)	−3.06	−2.66	0.01	0.01	2.65
Benzene dimer (T-shaped)	−2.78	−2.77	0.00	0.00	2.77
Benzene:ammonia	−2.29	−2.25	0.00	0.00	2.25
Benzene:hydrogen cyanide	−4.49	−4.57	0.01	0.01	4.56
Benzene:methane	−1.49	−1.38	0.00	0.00	1.38
Benzene:water	−3.16	−3.27	0.00	0.01	3.26
Ethene dimer	−1.38	−1.41	0.00	0.00	1.41
Ethene:ethyne	−1.43	−1.48	0.00	0.00	1.48
Ethyne dimer	−1.49	−1.53	0.00	0.00	1.53
Formic acid dimer	−18.21	−18.48	1.17	1.17	16.14
Formamide dimer	−15.78	−15.92	0.55	0.55	14.82
Indolebenzene dimer (stacked)	−5.10	−4.90	0.02	0.08	4.81
Indolebenzene dimer (T-shaped)	−5.55	−5.81	0.01	0.02	5.78
Methane dimer	−0.50	−0.49	0.00	0.00	0.49
Methanol:peptide	−8.35	−8.57	0.12	0.12	8.33
Methylamine:peptide	−5.31	−5.49	0.04	0.06	5.39
Peptide:methanol	−6.01	−6.24	0.07	0.03	6.15
Peptide:methylamine	−7.23	−7.38	0.18	0.02	7.18
Phenol dimer	−7.06	−7.26	0.07	0.09	7.11
Pyrazine dimer	−4.58	−4.57	0.01	0.01	4.55
Pyridoxine:aminopyridine	−16.84	−17.08	0.76	0.88	15.44
Trifluorochloromethane:formaldehyde	−2.28	−2.05	0.02	0.00	2.04
Uracil dimer (H-bonded)	−20.54	−21.21	0.85	0.85	19.51
Uracil dimer (stacked)	−10.39	−10.62	0.52	0.52	9.57
Water dimer	−4.74	−4.92	0.03	0.00	4.89
Water:pyridine	−6.74	−6.85	0.11	0.03	6.72

only, but using D3MBJ for ω PBE reduces the errors for both basis sets. For the PBE family of functionals tested, ω PBE-D3MBJ/aTZ proved to be the most accurate functional for calculating binding energies.

Some of the more expensive methods, ω B97M-D3BJ, ω B97X-D3BJ, M06-2X, and B97M-D3BJ, do result in accurate answers that are competitive with ω PBE-D3MBJ/aTZ, and B97-D3MBJ/aTZ produces surprisingly low errors for its reduced cost. The best-performing methods with aTZ are B97M-D3BJ, ω B97M-D3BJ, and ω B97X-D3BJ. Their mean absolute binding energy errors are 0.24, 0.26, and 0.26 kcal mol^{−1}, respectively.

In the case of 2-pyridoxine:2-aminopyridine, a dimer with two hydrogen bonds and a focal-point binding energy of 15.44 kcal mol^{−1}, maximum errors in binding energies are observed for half of the DFT methods with the aDZ basis set: B3LYP-D3MBJ (3.3), B97-D3MBJ (3.1), PBE0-D3MBJ (3.3), ω B97X-D3BJ (1.4), and ω PBE-D3MBJ (2.1). Maximum errors for the stacked uracil dimer (reference D_e = 9.57 kcal mol^{−1}) are observed for M06-2X (1.2) and B97M-D3BJ (1.3). Finally, with a triple- ζ basis set, the binding energy of the hydrogen-bonded uracil dimer (D_e = 19.51 kcal mol^{−1}) is the hardest to calculate for B97M-D3BJ (−0.8), M06-2X (−1.3), ω B97M-D3BJ (−0.8), and ω B97X-D3BJ (−0.9).

The aTZ basis set generally leads to more accurate results than the aDZ basis set but at a significant cost (Fig. 4). The double- ζ level of theory with both the lowest maximum and mean absolute binding energy errors is B97M-D3BJ/aDZ. Its maximum absolute error, 1.3 kcal mol^{−1}, occurs for uracil (stacked), and its mean absolute error is 0.5 kcal mol^{−1} out of an average reference binding energy of 6.4 kcal mol^{−1} across all dimers. Other dimers with binding energy errors greater in magnitude than 1 kcal mol^{−1} include benzene (parallel, 1.2 kcal mol^{−1}) and indole benzene (stack, 1.3 kcal mol^{−1}).

To further analyze each level of theory, the absolute errors in binding energies have been separated into absolute errors in interaction and deformation energies. Additionally, we calculated RMSDs to compare the dimers that were optimized with DFT vs geometries from the focal-point method.

PBE0 and ω PBE geometry optimizations result in RMSDs between 0.6 and 2.0 Å for two or three dimers (depending on the method/basis set) in the bottom left panel of Fig. 5, whereas all other methods return RMSDs below $\sim$ 0.5 Å. The stacked uracil dimer optimized with PBE0 and ω PBE has an RMSD approaching 2 Å, and this is due to the stacked dimer optimizing into the hydrogen-bonded uracil dimer. (Images of this and the following

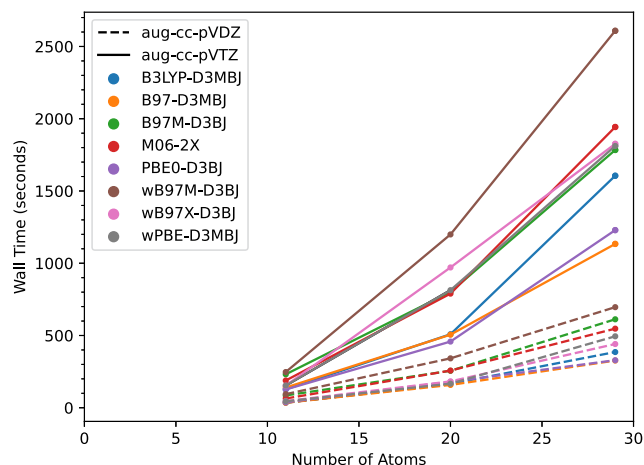


FIG. 4. Wall times (seconds) of single-point energy calculations of pyridine polymers (one, two, and three pyridine subunits) on two cores of an Intel Xeon E7-8867 v4 processor @ 2.40 GHz with Psi4 1.4. PBE0 and ω PBE are also methods used, but their timings can be considered equivalent to those of PBE0-D3MBJ and ω PBE-D3MBJ.

examples are provided in the supplementary material.) The non-dispersion corrected PBE0 and ω PBE also return RMSDs above 0.5 Å for the parallel-displaced benzene and methylamine:peptide dimers. In the parallel-displaced benzene dimer, the benzene rings remain parallel but move farther away from each other. In methylamine:peptide, the methyl group of the methylamine rotates away from the peptide, resulting in a larger RMSD. Methods other than PBE0 and ω PBE struggle the most when optimizing the methanol:peptide dimer, represented by points along 0.5 Å in the RMSD vs method plot of Fig. 5. While the closest contact distance stays within tenths of Å of the reference, rotation of methanol's methyl group contributes the most to this RMSD. Many MAEs are similar for all but the non-dispersion corrected PBE0 and ω PBE, and the choice of basis set does not seem to largely affect the RMSDs.

Figure 5 indicates that PBE0/aTZ and ω PBE/aTZ have larger average RMSDs than PBE0/aDZ and ω PBE/aDZ, as reflected in the larger errors in binding energies with the aTZ basis set noted above. (The geometries were optimized at the same level of theory used to compute the binding energies.) With poorer geometries, we expect that the interaction energy errors may be larger for the PBE functionals with aTZ basis sets vs aDZ basis sets. This is especially obvious for uracil (stacked) and benzene (parallel), which both have larger RMSDs and binding energy errors. Pyrazine, discussed before as having the largest binding energy errors with PBE0 and ω PBE, has RMSDs (Å) of 0.17 (PBE0/aDZ), 0.24 (PBE0/aTZ), 0.13 (ω PBE/aDZ), and 0.24 (ω PBE/aTZ). There are certainly some RMSDs larger than these for the PBE methods, and so the large binding energy errors may not be fully due to the geometry errors.

While many methods perform similarly when optimizing the dimer geometries, this is not the case for calculating interaction energies. The top right panel of Fig. 5 shows that only two levels of theory keep all interaction energy errors, relative to

the reference values, below 1 kcal mol⁻¹: B97M-D3BJ/aTZ and ω B97M-D3BJ/aTZ. All GGAs and hybrid GGAs show larger errors, especially with aDZ basis sets. Dimers with errors larger than 3 kcal mol⁻¹ (among various DFAs) include acetic acid, formic acid, uracil (H-bonded), 2-pyridoxine-2-aminopyridine, benzene (parallel), and pyrazine dimers. While most methods keep the interaction energy error of the pyrazine dimer below 1 kcal mol⁻¹, PBE0 and ω PBE return errors of 3.2–3.6 kcal mol⁻¹, and this is the cause of most of the error in the binding energy discussed earlier. B97M-D3BJ has the lowest mean absolute error of interaction energies for methods with the aDZ basis set. Its maximum error is just higher than two other methods, M06-2X and ω B97M-D3BJ, but these methods have MAEs higher than B97M-D3BJ/aDZ.

Figure 5 (bottom right) also plots absolute errors in deformation energies calculated with DFT relative to the reference method. Many methods have errors in deformation energies at or above 1 kcal mol⁻¹, but those that do not are the well-performing B97M-D3BJ, ω B97M-D3BJ, and ω B97X-D3BJ methods. Methods with errors approaching 1 kcal mol⁻¹ are B97-D3MBJ and M06-2X (aDZ only). Every clustered pair of dots plotted above 1 kcal mol⁻¹ represents acetic acid and formic acid, and these molecules also have the largest focal-point coupled-cluster deformation energies. From this benchmarking study, we conclude that B97M-D3BJ/aDZ is a reasonable choice that balances time and accuracy when calculating binding energies.

III. DEFORMATION ENERGY ANALYSIS

In this section, we study the binding energies of 104 dimers. We compute these with DFT and report interaction, deformation, and binding energies. We then analyze cases that have especially large deformation energies in an attempt to identify properties that warrant a binding energy calculation rather than solely an interaction energy calculation in order to better approximate the binding strength.

A. Methods

B97M-D3BJ with an aug-cc-pVDZ basis set was chosen to study the binding energies of a larger set of molecules due to its accuracy for binding energies and geometries while maintaining a relatively low computational cost (studied in Sec. II). This method was used for all geometry optimizations and energy calculations on a new, larger dataset. This dataset was taken from the Splinter database.⁸⁶ Dimers were selected on the basis of structural diversity, having relatively low (favorable) energies of interaction, and representing a variety of neutral, non-covalent interactions, including many commonly observed in protein-ligand binding.⁸⁶ Figures of all 104 dimers, as well as their SMILES strings (generated using RDKit⁸⁷), are in the supplementary material.

In addition to the geometry optimizations and energy calculations with B97M-D3BJ/aDZ, the lowest level of SAPT, SAPT0, was used to calculate physically relevant components of the interaction energy. These are listed in the supplementary material. The test set comprises a range of interaction types, from electrostatics-dominated to dispersion-dominated, as illustrated in Fig. 6. A development version of Psi4 1.4 and the default DFT grid were used for all of these computations.

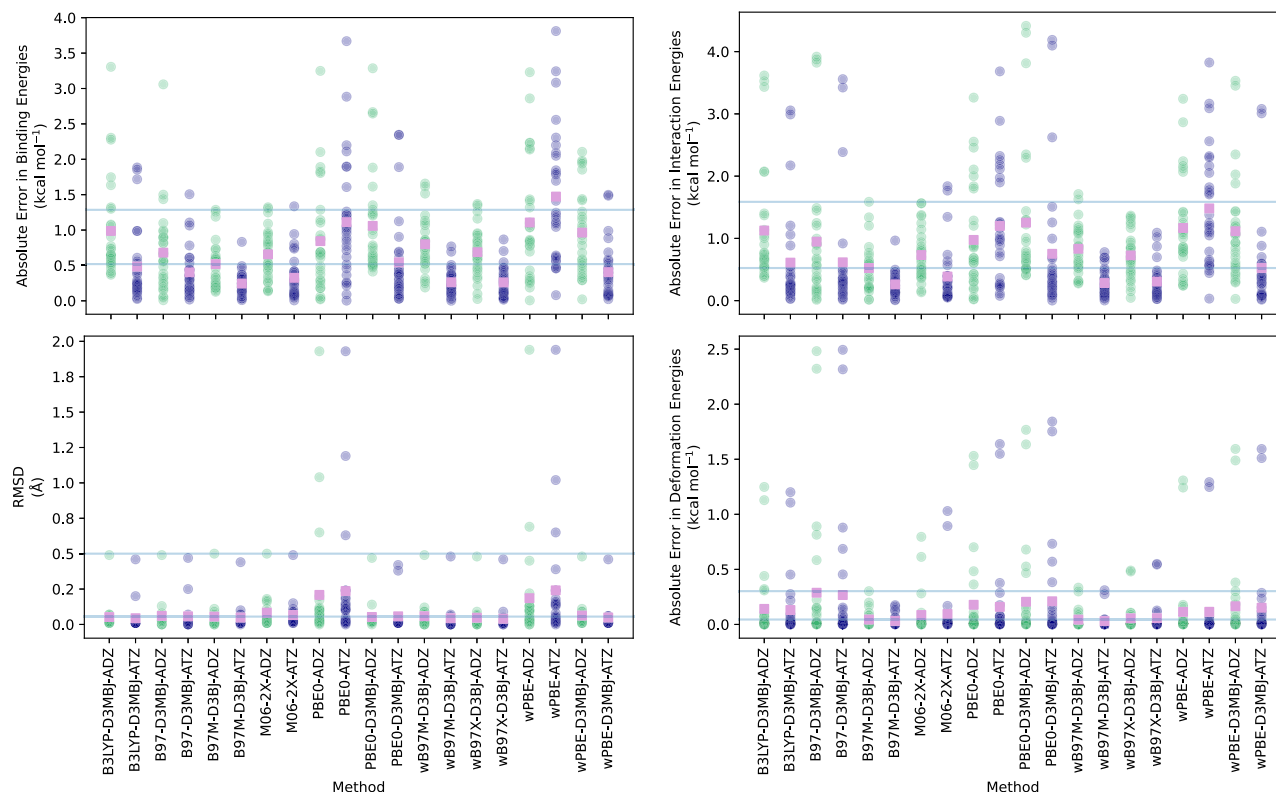


FIG. 5. Absolute errors in binding energies (D_e , top left), interaction energies (top right), and deformation energies (bottom right), and RMSDs of geometries (bottom left). Reference interaction energies used are computed with SAPT2+(3) δ MP2/aTZ. CCSD(T)/CBS(a[Q5]Z; δ :DZ) is used to compute reference geometries and deformation energies. DFT results are reported at geometries obtained using the same DFT method. Pink boxes are mean absolute errors (or mean RMSDs) for each level of theory, and the blue horizontal lines indicate the maximum and mean absolute errors for B97M-D3MBJ/aTZ.

B. Results and discussion

The interaction, deformation, and binding energies for the dataset of 104 dimers are available in the supplementary material. The binding energies range from 1.15 to 22.44 kcal mol⁻¹. Total deformation energy, the sum of the deformation energies of the two monomers in each dimer, is plotted against the dimer interaction energy in Fig. 7. As one may expect, these are somewhat related. As the interaction energy becomes more negative (stronger), deformation energy generally increases, indicating that greater deformations can occur when a stronger interaction is possible. Dimers with interaction energies more negative (stronger) than -20 kcal mol⁻¹ can have deformation energies varying in magnitude, ranging from 1 to 10 kcal mol⁻¹. Five of the 104 dimers have deformation energies larger than 5 kcal mol⁻¹. In these cases, the total deformation energy can be as much as 25%–50% of the interaction energy. When interaction energies are between -10 and -20 kcal mol⁻¹, the deformation energies tend to not exceed 2.5 kcal mol⁻¹. Aside from two outliers, dimers with interaction energies above -10 kcal mol⁻¹ (weaker) have deformation energies less than 1 kcal mol⁻¹.

The three dimers with the highest deformation energies are shown in Fig. 8. (See the supplementary material for structural details of the complete dataset.). Transparent structures are those where the monomer is relaxed to its nearest local minimum. Opaque

structures are the dimer in its optimized form. Apparent in each rendering is the rotation of bonds to form hydrogen bonds. This aligns with existing observations that higher deformation energies will occur in hydrogen bonded systems, rather than dispersion bound, due to the stronger interaction afforded by hydrogen bonding.^{31,88}

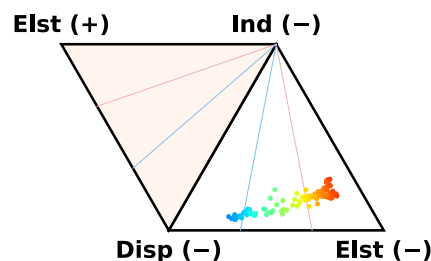


FIG. 6. Ternary diagram for the 104-dimer test set, with energy components obtained at the SAPT0/aDZ level of theory. Each dimer is represented by one colored dot on the diagram, whose color and placement are determined by the relative sizes of the three usually attractive components of the interaction energy: electrostatics (red), dispersion (blue), or induction. Points close to a vertex have an interaction energy dominated by that energy component.

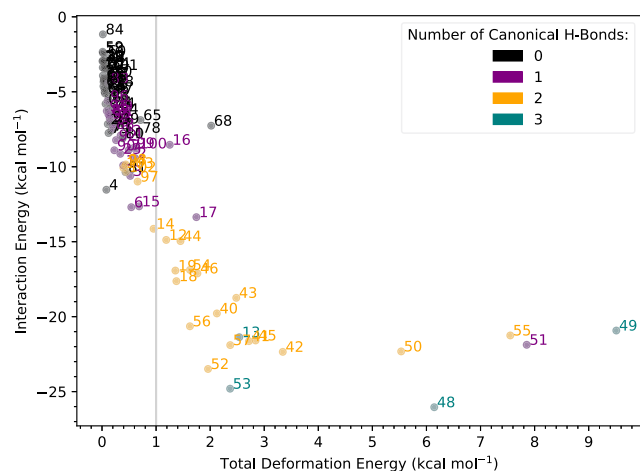


FIG. 7. Total deformation energies and interaction energies (both kcal mol⁻¹) calculated with B97M-D3BJ/aDZ for 104 dimers. All necessary optimizations also used B97M-D3BJ/aDZ. Dots are indexed by dimer and color-coded by the number of canonical hydrogen bonds present in the dimer.

To study the role of hydrogen bonding in the data, the canonical hydrogen bonds in each dimer were counted using the “Workspace Interaction” utility in Maestro.⁸⁹ This generally defines hydrogen bonds as involving polar hydrogens in close contact with nitrogen or oxygen atoms with available lone pairs of electrons. The 104 dimer data points in Fig. 7 are color-coded by the number of canonical hydrogen bonds.

Figure 7 shows that, generally, deformation energy increases with the number of hydrogen bonds due to the increase in the magnitude of the interaction energy. Dimers with two or three hydrogen bonds can have deformation energies as large as 9.5 kcal mol⁻¹, but when hydrogen bonding is absent or there is only one hydrogen bond, deformation energies tend to stay below 1 kcal mol⁻¹. Dimer 68 is the one case of a dimer with no hydrogen bonding and a deformation energy above 1 kcal mol⁻¹. Its deformation energy is around 2 kcal mol⁻¹. Here, a hydroxy group engages in favorable dipolar interactions with methane thiol, but in the absence of the latter, rotates to gain stabilizing intramolecular polar interactions. Increasing the hydrogen bond count to one slightly increases

the deformation energy, and two dimers (16 and 17) have deformation energies above 1 kcal mol⁻¹. In dimer 16, the largest strain comes from propanamide, where the CH₃CH₂-C(=O)(NH₂) bond rotates almost 90° in order for the amine proton to form a hydrogen bond with nitrogen on the opposing, bicyclic monomer. This rotation causes a monomeric deformation energy of 1.17 kcal mol⁻¹. For dimer 17, the largest deformation energy (1.45 kcal mol⁻¹) arises from a rotation of the benzamide sigma bond, allowing the oxygen to bond with the amine proton. (Illustrations of these monomer deformations are shown in the supplementary material.) Similarly, dimers containing two hydrogen bonds resulting from bond rotation can have deformation energies above 1 kcal mol⁻¹. Otherwise, deformation energies of dimers with two hydrogen bonds are generally below 1 kcal mol⁻¹ (dimers 24, 97, 98, 102, and 103). Finally, having three canonical hydrogen bonds, in this dataset, guarantees a total deformation energy greater than 2 kcal mol⁻¹.

A cluster of points appears to the right of Fig. 7. Five dimers have deformation energies between 5.5 and 9.5 kcal mol⁻¹, while the maximum for the other 99 dimers is 3.3 kcal mol⁻¹. Dimers 48–51 (four of the five) all have one monomer in common that itself contributes 4.5–6.8 kcal mol⁻¹ to the total deformation energy. This monomer, shown as part of Fig. 8, has an intramolecular hydrogen bond with an H–O bond distance of 1.96 Å. In dimers 48–51, this intramolecular bond is broken to form the dimer. Dimer 47 also contains this monomer, but it has a low deformation energy (0.2 kcal mol⁻¹). Further inspection reveals that the intramolecular hydrogen bond mentioned previously is retained when forming dimer 47. Dimer 55 (also in Fig. 8) is the remaining member of the cluster. 5.3 kcal mol⁻¹ of its total deformation energy is due to the breaking of an intramolecular hydrogen bond within monomer A. The O and H are 2.18 Å apart in the relaxed monomer, but 2.96 Å apart in the dimer. Other dimers include this monomer, but the hydrogen bond is not broken to form the dimer, and consequently, they return lower deformation energies.

In summary, for this dataset, deformation energies can be above 1 kcal mol⁻¹ in many cases, and even up to 9.5 kcal mol⁻¹. Hydrogen bonding promotes stronger intermolecular interactions, which in turn allow larger monomer deformations. Typically, a dimer with zero or one hydrogen bond will have a deformation energy below 1 kcal mol⁻¹ while higher deformation energies become more likely as the number of hydrogen bonds increases. Deviations from this principle occur, especially when intramonomer hydrogen bonds are

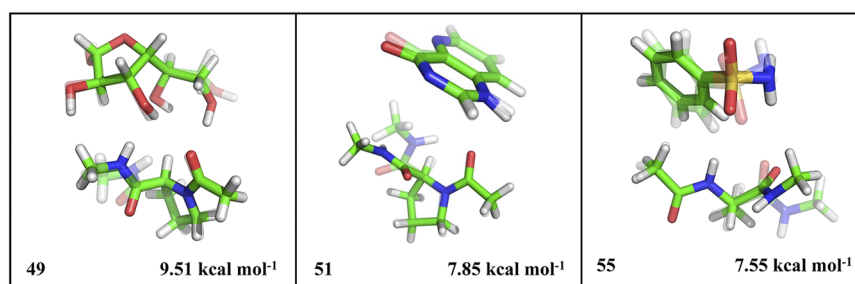


FIG. 8. The three dimer cases with the highest deformation energies. Dimer indices and total deformation energies are noted. Transparent molecules are monomers relaxed to their local minima, and opaque structures are the optimized dimer. B97M-D3BJ/aDZ was used for geometry optimizations.

broken to form intermolecular hydrogen bonds (leading to a very high deformation energy) or when no bond rotations are needed to form the hydrogen bonds (leading to a low deformation energy).

IV. CONCLUSIONS

Benchmark-quality binding energies (D_e), accounting for both interaction and deformation energies, have been presented for 28 organic dimers. Dimers were optimized with CCSD(T)/CBS(a[Q5]Z; δ :DZ) without counterpoise correction, and their interaction energies were calculated with SAPT2+(3) δ MP2/aTZ. Monomer optimizations and deformation energies utilized CCSD(T)/CBS(a[Q5]Z; δ :DZ) without CP correction. For 22 of the dimers, we have shown that using CCSD(T)/CBS(a[Q5]Z; δ :DZ) without CP correction returns similar results to using CCSD(T)/CBS(a[Q5]Z; δ :aDZ) with a CP correction. The binding energies (using the fixed SAPT method for the interaction energy contribution) differ by less than 0.07 kcal mol⁻¹. Various density functionals, with aug-cc-pVDZ and aug-cc-pVTZ basis sets, have been used to test their accuracy and efficiency in optimizing the 28 dimers and calculating their interaction and deformation energies relative to the reference data. From the 20 approximate levels of theory considered, B97M-D3BJ/aDZ was chosen for additional studies of deformation energies due to its accuracy in computing binding energies (D_e) and optimum geometries, in addition to having a relatively low computational cost. For this method, the mean absolute error in binding energies, relative to CCSD(T)/CBS(a[Q5]Z; δ :DZ), was 0.5 kcal mol⁻¹, and the maximum absolute error was 1.3 kcal mol⁻¹. The zero-point vibrational energy corrections are not studied here, and evaluating the accuracy of methods like B97M-D3BJ/aDZ for computing the ZPVE contributions to dissociation energies (D_0) should be explored in the future.

With B97M-D3BJ/aDZ, binding energies (D_e) of 104 dimers were calculated. 26 dimers had deformation energies larger than 1 kcal mol⁻¹ and 5 of those reached deformation energies larger than 5 kcal mol⁻¹. Deformation energies of this magnitude should certainly not be ignored when considering interactions of two molecules, thus highlighting the difference between interaction energies (which are easier to compute) and binding energies. Results were analyzed to find a principle that would help computational chemists discern, on the basis of the properties of the dimer, when it is important to include the deformation energy. The largest deformation energies occur when a hydrogen bond is broken within a monomer to form a hydrogen bonded dimer when there is a dramatic rotation to form a hydrogen bond and/or when two or more hydrogen bonds form. The hydrogen bonds increase the magnitude of the interaction energy, allowing for larger deformations to occur.

SUPPLEMENTARY MATERIAL

The supplementary material contains errors in the CCSD(T)/CBS(a[Q5]Z; δ :DZ) interaction energies without counterpoise correction relative to those of CCSD(T)/CBS(a[Q5]Z; δ :aDZ) with counterpoise correction for 22 of the 28 dimers in the DFT method benchmarking study. RMSDs of the geometries optimized with CCSD(T)/CBS(a[Q5]Z; δ :DZ) without CP

correction, relative to CCSD(T)/CBS(a[Q5]Z; δ :aDZ) with CP correction, are also tabulated for the 22 dimers. For the complete set of 28 dimers, SAPT2+(3) δ MP2/aTZ components are listed. Also for this set, the following energetic errors are tabulated: interaction energy errors of each DFT method relative to SAPT2+(3) δ MP2/aTZ, deformation energy errors of each DFT method relative to CCSD(T)/CBS(a[Q5]Z; δ :DZ) without CP correction, and binding energy errors of each DFT method relative to the CCSD(T)/CBS(a[Q5]Z; δ :DZ) deformation energies added to the SAPT2+(3) δ MP2/aTZ interaction energies. The necessary geometry optimizations for these calculations use the DFT method for each DFT datapoint and CCSD(T)/CBS(a[Q5]Z; δ :DZ) without CP correction for each reference datapoint. Also, RMSDs of the 28 dimers optimized with each DFT method, relative to CCSD(T)/CBS(a[Q5]Z; δ :DZ) without CP correction, are provided. Finally, the supplementary material includes example input files used to compare timings of each DFT method.

For the set of 104 dimers, indices matched to their SMILES strings and illustrations are provided, in addition to XYZ and SDF files of these dimers optimized with B97M-D3BJ/aDZ. Also, the interaction, deformation, and binding energies calculated with B97M-D3BJ/aDZ are listed, as well as the SAPT0/aDZ components, with geometries from optimizations using B97M-D3BJ/aDZ. Finally, the supplementary material contains illustrations of some dimer optimizations and deformations explained in the text.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge financial support from the U.S. National Science Foundation (Grant No. CHE-1955940) and from Bristol Myers Squibb.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Caroline T. Sargent: Formal analysis (lead); Investigation (supporting); Visualization (lead); Writing – original draft (lead). **Raina Kasera:** Data curation (lead); Formal analysis (supporting); Investigation (lead). **Zachary L. Glick:** Methodology (supporting); Software (lead). **C. David Sherrill:** Conceptualization (supporting); Funding acquisition (equal); Methodology (equal); Project administration (equal); Supervision (equal); Writing – review & editing (equal). **Daniel L. Cheney:** Conceptualization (lead); Funding acquisition (equal); Methodology (equal); Project administration (equal); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available in the article and its supplementary material.

REFERENCES

- ¹C. Sutton, C. Risko, and J.-L. Brédas, "Noncovalent intermolecular interactions in organic electronic materials: Implications for the molecular packing vs electronic properties of acenes," *Chem. Mater.* **28**, 3–16 (2016).
- ²O. A. Arodola and M. E. Soliman, "Quantum mechanics implementation in drug-design workflows: Does it really help?," *Drug Des., Dev. Ther.* **11**, 2551–2564 (2017).
- ³D. Mucs and R. A. Bryce, "The application of quantum mechanics in structure-based drug design," *Expert Opin. Drug Discovery* **8**, 263–276 (2013).
- ⁴B. Jeziorski, R. Moszynski, and K. Szalewicz, "Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes," *Chem. Rev.* **94**, 1887–1930 (1994).
- ⁵K. Szalewicz, "Symmetry-adapted perturbation theory of intermolecular forces," *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2**, 254–272 (2012).
- ⁶E. G. Hohenstein and C. D. Sherrill, "Wavefunction methods for noncovalent interactions," *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2**, 304–326 (2012).
- ⁷P. 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).
- ⁸J. Řezáč, K. E. Riley, and P. Hobza, "S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures," *J. Chem. Theory Comput.* **7**, 2427–2438 (2011).
- ⁹J. Řezáč, K. E. Riley, and P. Hobza, "Benchmark calculations of noncovalent interactions of halogenated molecules," *J. Chem. Theory Comput.* **8**, 4285–4292 (2012).
- ¹⁰L. A. Burns, J. C. Faver, Z. Zheng, M. S. Marshall, D. G. A. Smith, K. Vanommeslaeghe, A. D. MacKerell, K. M. Merz, and C. D. Sherrill, "The BioFragment Database (BFDdb): An open-data platform for computational chemistry analysis of noncovalent interactions," *J. Chem. Phys.* **147**, 161727 (2017).
- ¹¹J. C. Faver, M. L. Benson, X. He, B. P. Roberts, B. Wang, M. S. Marshall, M. R. Kennedy, C. D. Sherrill, and K. M. Merz, "Formal estimation of errors in computed absolute interaction energies of protein-ligand complexes," *J. Chem. Theory Comput.* **7**, 790–797 (2011).
- ¹²K. S. Thanthiriwatt, E. G. Hohenstein, L. A. Burns, and C. D. Sherrill, "Assessment of the performance of DFT and DFT-D methods for describing distance dependence of hydrogen-bonded interactions," *J. Chem. Theory Comput.* **7**, 88–96 (2011).
- ¹³C. D. Sherrill, T. Takatani, and E. G. Hohenstein, "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," *J. Phys. Chem. A* **113**, 10146–10159 (2009).
- ¹⁴T. Takatani and C. David Sherrill, "Performance of spin-component-scaled Møller-Plesset theory (SCS-MP2) for potential energy curves of noncovalent interactions," *Phys. Chem. Chem. Phys.* **9**, 6106–6114 (2007).
- ¹⁵E. G. Hohenstein and C. D. Sherrill, "Effects of heteroatoms on aromatic π - π interactions: Benzene-pyridine and pyridine dimer," *J. Phys. Chem. A* **113**, 878–886 (2009).
- ¹⁶P. Hobza, "Calculations on noncovalent interactions and databases of benchmark interaction energies," *Acc. Chem. Res.* **45**, 663–672 (2012).
- ¹⁷L. A. Burns, M. S. Marshall, and C. D. Sherrill, "Comparing counterpoise-corrected, uncorrected, and averaged binding energies for benchmarking noncovalent interactions," *J. Chem. Theory Comput.* **10**, 49–57 (2014).
- ¹⁸Y. Zhao and D. G. Truhlar, "Benchmark databases for nonbonded interactions and their use to test density functional theory," *J. Chem. Theory Comput.* **1**, 415–432 (2005).
- ¹⁹J. Antony and S. Grimme, "Density functional theory including dispersion corrections for intermolecular interactions in a large benchmark set of biologically relevant molecules," *Phys. Chem. Chem. Phys.* **8**, 5287–5293 (2006).
- ²⁰R. Kalescky, E. Kraka, and D. Cremer, "Accurate determination of the binding energy of the formic acid dimer: The importance of geometry relaxation," *J. Chem. Phys.* **140**, 084315 (2014).
- ²¹K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, "A fifth-order perturbation comparison of electron correlation theories," *Chem. Phys. Lett.* **157**, 479–483 (1989).
- ²²T. M. Sexton, J. C. Howard, and G. S. Tschumper, "Dissociation energy of the H₂O...HF dimer," *J. Phys. Chem. A* **122**, 4902–4908 (2018).
- ²³J. C. Howard, J. L. Gray, A. J. Hardwick, L. T. Nguyen, and G. S. Tschumper, "Getting down to the fundamentals of hydrogen bonding: Anharmonic vibrational frequencies of (HF)₂ and (H₂O)₂ from *ab initio* electronic structure computations," *J. Chem. Theory Comput.* **10**, 5426–5435 (2014).
- ²⁴R. Knochenmuss, R. K. Sinha, and S. Leutwyler, "Face, notch, or edge? Intermolecular dissociation energies of 1-naphthol complexes with linear molecules," *J. Chem. Phys.* **150**, 234303 (2019).
- ²⁵K. H. Lemke, "Structure and binding energy of the H₂S dimer at the CCSD(T) complete basis set limit," *J. Chem. Phys.* **146**, 234301 (2017).
- ²⁶T. M. Sexton, W. Z. Van Benschoten, and G. S. Tschumper, "Dissociation energy of the HCN...HF dimer," *Chem. Phys. Lett.* **748**, 137382 (2020).
- ²⁷C. Møller and M. S. Plesset, "Note on an approximation treatment for many-electron systems," *Phys. Rev.* **46**, 618–622 (1934).
- ²⁸J. Czernek, J. Brus, and V. Czerneková, "A computational inspection of the dissociation energy of mid-sized organic dimers," *J. Chem. Phys.* **156**, 204303 (2022).
- ²⁹A. Hesselmann, G. Jansen, and M. Schütz, "Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies," *J. Chem. Phys.* **122**, 014103 (2005).
- ³⁰A. J. Misquitta, R. Podeszwa, B. Jeziorski, and K. Szalewicz, "Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations," *J. Chem. Phys.* **123**, 214103 (2005).
- ³¹S. Halder, R. Gnanasekaran, and P. Hobza, "A comparison of *ab initio* quantum-mechanical and experimental *D*₀ binding energies of eleven H-bonded and eleven dispersion-bound complexes," *Phys. Chem. Chem. Phys.* **17**, 26645–26652 (2015).
- ³²R. Podeszwa, K. Patkowski, and K. Szalewicz, "Improved interaction energy benchmarks for dimers of biological relevance," *Phys. Chem. Chem. Phys.* **12**, 5974–5979 (2010).
- ³³J. Řezáč, K. E. Riley, and P. Hobza, "Extensions of the S66 data set: More accurate interaction energies and angular-displaced nonequilibrium geometries," *J. Chem. Theory Comput.* **7**, 3466–3470 (2011).
- ³⁴K. Patkowski, "Basis set converged weak interaction energies from conventional and explicitly correlated coupled cluster approach," *J. Chem. Phys.* **138**, 154101 (2013).
- ³⁵J. Řezáč and P. Hobza, "Benchmark calculations of interaction energies in noncovalent complexes and their applications," *Chem. Rev.* **116**, 5038 (2016).
- ³⁶L. A. Burns, M. S. Marshall, and C. D. Sherrill, "Appointing silver and bronze standards for noncovalent interactions: A comparison of spin-component-scaled (SCS), explicitly correlated (F12), and specialized wavefunction approaches," *J. Chem. Phys.* **141**, 234111 (2014).
- ³⁷D. G. A. Smith, P. Jankowski, M. Slawik, H. A. Witek, and K. Patkowski, "Basis set convergence of the post-CCSD(T) contribution to noncovalent interaction energies," *J. Chem. Theory Comput.* **10**, 3140 (2014).
- ³⁸S. Tsuzuki, K. Honda, T. Uchimaru, M. Mikami, and K. Tanabe, "Origin of the attraction and directionality of the NH/ π interaction: Comparison with OH/ π and CH/ π interactions," *J. Am. Chem. Soc.* **122**, 11450–11458 (2000).
- ³⁹S. Tsuzuki, K. Honda, T. Uchimaru, and M. Mikami, "High-level *ab initio* computations of structures and interaction energies of naphthalene dimers: Origin of attraction and its directionality," *J. Chem. Phys.* **120**, 647 (2004).
- ⁴⁰S. Tsuzuki and A. Fujii, "Nature and physical origin of CH/ π interaction: Significant difference from conventional hydrogen bonds," *Phys. Chem. Chem. Phys.* **10**, 2584–2594 (2008).
- ⁴¹L. Gráfová, M. Pitoňák, J. Řezáč, and P. Hobza, "Comparative study of selected wave function and density functional methods for noncovalent interaction energy calculations using the extended S22 data set," *J. Chem. Theory Comput.* **6**, 2365–2376 (2010).
- ⁴²M. O. Sinnokrot and C. D. Sherrill, "Highly accurate coupled cluster potential energy curves for benzene dimer: The sandwich, T-shaped, and parallel-displaced configurations," *J. Phys. Chem. A* **108**, 10200–10207 (2004).
- ⁴³A. L. Ringer, M. S. Figgis, M. O. Sinnokrot, and C. D. Sherrill, "Aliphatic C-H/ π interactions: Methane-benzene, methane-phenol, and methane-indole complexes," *J. Phys. Chem. A* **110**, 10822–10828 (2006).

- ⁴⁴M. O. Sinnokrot and C. D. Sherrill, "Substituent effects in π - π interactions: Sandwich and T-shaped configurations," *J. Am. Chem. Soc.* **126**, 7690–7697 (2004).
- ⁴⁵A. L. Ringer, M. O. Sinnokrot, R. P. Lively, and C. D. Sherrill, "The effect of multiple substituents on sandwich and T-shaped π - π interactions," *Chem. Eur. J.* **12**, 3821–3828 (2006).
- ⁴⁶S. Tsuzuki, M. Mikami, and S. Yamada, "Origin of attraction, magnitude, and directionality of interactions in benzene complexes with pyridinium cations," *J. Am. Chem. Soc.* **129**, 8656–8662 (2007).
- ⁴⁷S. E. Wheeler and K. N. Houk, "Through-space effects of substituents dominate molecular electrostatic potentials of substituted arenes," *J. Chem. Theory Comput.* **5**, 2301–2312 (2009).
- ⁴⁸S. E. Wheeler and K. N. Houk, "Substituent effects in cation/ π interactions and electrostatic potentials above the centers of substituted benzenes are due primarily to through-space effects of the substituents," *J. Am. Chem. Soc.* **131**, 3126 (2009).
- ⁴⁹S. E. Wheeler, "Local nature of substituent effects in stacking interactions," *J. Am. Chem. Soc.* **133**, 10262–10274 (2011).
- ⁵⁰S. E. Wheeler, "Understanding substituent effects in noncovalent interactions involving aromatic rings," *Acc. Chem. Res.* **46**, 1029–1038 (2013).
- ⁵¹J. Rezáč and P. Hobza, "Describing noncovalent interactions beyond the common approximations: How accurate is the 'gold standard,' CCSD(T) at the complete basis set limit?," *J. Chem. Theory Comput.* **9**, 2151–2155 (2013).
- ⁵²D. A. Sirianni, A. Alenaizan, D. L. Cheney, and C. D. Sherrill, "Assessment of density functional methods for geometry optimization of bimolecular van der Waals complexes," *J. Chem. Theory Comput.* **14**, 3004–3013 (2018).
- ⁵³J. Witte, M. Goldey, J. B. Neaton, and M. Head-Gordon, "Beyond energies: Geometries of nonbonded molecular complexes as metrics for assessing electronic structure approaches," *J. Chem. Theory Comput.* **11**, 1481–1492 (2015).
- ⁵⁴E. R. Johnson and G. A. DiLabio, "Structure and binding energies in van der Waals dimers: Comparison between density functional theory and correlated *ab initio* methods," *Chem. Phys. Lett.* **419**, 333–339 (2006).
- ⁵⁵A. L. L. East and W. D. Allen, "The heat of formation of NCO," *J. Chem. Phys.* **99**, 4638–4650 (1993).
- ⁵⁶A. G. Császár, W. D. Allen, and H. F. Schaefer, "In pursuit of the *ab initio* limit for conformational energy prototypes," *J. Chem. Phys.* **108**, 9751–9764 (1998).
- ⁵⁷M. O. Sinnokrot, E. F. Valeev, and C. D. Sherrill, "Estimates of the *ab initio* limit for π - π interactions: The benzene dimer," *J. Am. Chem. Soc.* **124**, 10887–10893 (2002).
- ⁵⁸M. S. Marshall, L. A. Burns, and C. D. Sherrill, "Basis set convergence of the coupled-cluster correction, $\delta_{\text{MP2}}^{\text{CCSD(T)}}$: Best practices for benchmarking non-covalent interactions and the attendant revision of the S22, NBC10, HBC6, and HSG databases," *J. Chem. Phys.* **135**, 194102 (2011).
- ⁵⁹A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, and J. Olsen, "Basis-set convergence of the energy in molecular Hartree–Fock calculations," *Chem. Phys. Lett.* **302**, 437–446 (1999).
- ⁶⁰T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," *J. Chem. Phys.* **90**, 1007–1023 (1989).
- ⁶¹D. E. Woon and T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties," *J. Chem. Phys.* **100**, 2975–2988 (1994).
- ⁶²T. M. Parker, L. A. Burns, R. M. Parrish, A. G. Ryno, and C. D. Sherrill, "Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies," *J. Chem. Phys.* **140**, 094106 (2014).
- ⁶³S. F. Boys and F. Bernardi, "The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors," *Mol. Phys.* **19**, 553–566 (1970).
- ⁶⁴D. Feller, "Application of systematic sequences of wave functions to the water dimer," *J. Chem. Phys.* **96**, 6104–6114 (1992).
- ⁶⁵D. G. A. Smith, L. A. Burns, A. C. Simmonett, R. M. Parrish, M. C. Schieber, R. Galvelis, P. Kraus, H. Kruse, R. Di Remigio, A. Alenaizan, A. M. James, S. Lehtola, J. P. Misiewicz, M. Scheurer, R. A. Shaw, J. B. Schriber, Y. Xie, Z. L. Glick, D. A. Sirianni, J. S. O'Brien, J. M. Waldrop, A. Kumar, E. G. Hohenstein, B. P. Pritchard, B. R. Brooks, H. F. Schaefer, A. Y. Sokolov, K. Patkowski, A. E. DePrince, U. Bozkaya, R. A. King, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, "PSI4 1.4: Open-source software for high-throughput quantum chemistry," *J. Chem. Phys.* **152**, 184108 (2020).
- ⁶⁶M. K. Kesharwani, A. Karton, and J. M. L. Martin, "Benchmark *ab initio* conformational energies for the proteinogenic amino acids through explicitly correlated methods. Assessment of density functional methods," *J. Chem. Theory Comput.* **12**, 444–454 (2016).
- ⁶⁷Y. Li, S. Zhang, J. Z. H. Zhang, and X. He, "Assessing the performance of popular QM methods for calculation of conformational energies of trialanine," *Chem. Phys. Lett.* **652**, 136–141 (2016).
- ⁶⁸I. Tolbatov and D. M. Chipman, "Benchmarking density functionals and Gaussian basis sets for calculation of core-electron binding energies in amino acids," *Theor. Chem. Acc.* **136**, 82 (2017).
- ⁶⁹Y. K. Kang and H. S. Park, "Exploring conformational preferences of alanine tetrapeptide by CCSD(T), MP2, and dispersion-corrected DFT methods," *Chem. Phys. Lett.* **702**, 69–75 (2018).
- ⁷⁰B. Chan, "Aqueous-phase conformations of lactose, maltose, and sucrose and the assessment of low-cost DFT methods with the DSCONF set of conformers for the three disaccharides," *J. Phys. Chem. A* **124**, 582–590 (2020).
- ⁷¹A. Karton, "How reliable is DFT in predicting relative energies of polycyclic aromatic hydrocarbon isomers? Comparison of functionals from different rungs of Jacob's ladder," *J. Comput. Chem.* **38**, 370–382 (2017).
- ⁷²P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, "Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields," *J. Phys. Chem.* **98**, 11623–11627 (1994).
- ⁷³A. D. Becke, "Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals," *J. Chem. Phys.* **107**, 8554–8560 (1997).
- ⁷⁴N. Mardirossian and M. Head-Gordon, "Mapping the genome of meta-generalized gradient approximation density functionals: The search for B97M-V," *J. Chem. Phys.* **142**, 074111 (2015).
- ⁷⁵Y. Zhao and D. G. Truhlar, "The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals," *Theor. Chem. Acc.* **120**, 215–241 (2008).
- ⁷⁶C. Adamo, G. E. Scuseria, and V. Barone, "Accurate excitation energies from time-dependent density functional theory: Assessing the PBE0 model," *J. Chem. Phys.* **111**, 2889–2899 (1999).
- ⁷⁷N. Mardirossian and M. Head-Gordon, " ω B97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation," *J. Chem. Phys.* **144**, 214110 (2016).
- ⁷⁸J.-D. Chai and M. Head-Gordon, "Systematic optimization of long-range corrected hybrid density functionals," *J. Chem. Phys.* **128**, 084106 (2008).
- ⁷⁹N. Mardirossian and M. Head-Gordon, " ω B97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy," *Phys. Chem. Chem. Phys.* **16**, 9904–9924 (2014).
- ⁸⁰T. M. Henderson, B. G. Janesko, and G. E. Scuseria, "Generalized gradient approximation model exchange holes for range-separated hybrids," *J. Chem. Phys.* **128**, 194105 (2008).
- ⁸¹E. Weintraub, T. M. Henderson, and G. E. Scuseria, "Long-range-corrected hybrids based on a new model exchange hole," *J. Chem. Theory Comput.* **5**, 754–762 (2009).
- ⁸²S. Grimme, S. Ehrlich, and L. Goerigk, "Effect of the damping function in dispersion corrected density functional theory," *J. Comput. Chem.* **32**, 1456–1465 (2011).
- ⁸³D. G. A. Smith, L. A. Burns, K. Patkowski, and C. D. Sherrill, "Revised damping parameters for the D3 dispersion correction to density functional theory," *J. Phys. Chem. Lett.* **7**, 2197–2203 (2016).
- ⁸⁴N. Mardirossian and M. Head-Gordon, "How accurate are the Minnesota density functionals for noncovalent interactions, isomerization energies, thermochemistry, and barrier heights involving molecules composed of main-group elements?," *J. Chem. Theory Comput.* **12**, 4303–4325 (2016).

⁸⁵M. Marianski, A. Supady, T. Ingram, M. Schneider, and C. Baldauf, “Assessing the accuracy of across-the-scale methods for predicting carbohydrate conformational energies for the examples of glucose and α -maltose,” *J. Chem. Theory Comput.* **12**, 6157–6168 (2016).

⁸⁶S. A. Spronk, Z. L. Glick, D. P. Metcalf, C. D. Sherrill, and D. L. Cheney, “A quantum chemical interaction energy dataset for accurately modeling protein-ligand interactions” (submitted) (2023).

⁸⁷See <https://www.rdkit.org> for RDKit: Open-source cheminformatics.

⁸⁸I.-C. Lin, O. A. von Lilienfeld, M. D. Coutinho-Neto, I. Tavernelli, and U. Rothlisberger, “Predicting noncovalent interactions between aromatic biomolecules with london-dispersion-corrected DFT,” *J. Phys. Chem. B* **111**, 14346–14354 (2007).

⁸⁹Schrödinger Release 2022-3: Maestro, Schrödinger, LLC, New York, 2021.