Benchmarking Two-Body Contributions to Crystal Lattice Energies and a Range-Dependent Assessment of Approximate Methods

Caroline T. Sargent, Derek P. Metcalf, Zachary L. Glick, Carlos H. Borca, and C. David Sherrill[a)]

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

(Dated: 9 January 2023)

Using the many-body expansion to predict crystal lattice energies (CLEs), a pleasantly parallel process, allows for flexibility in the choice of theoretical methods. Benchmark-level two-body contributions to CLEs of 23 molecular crystals have been computed using interaction energies of dimers with minimum inter-monomer separations (i.e., closest contact distances) up to 30 Å. In a search for ways to reduce the computational expense of calculating accurate CLEs, we have computed these two-body contributions with 15 different quantum chemical levels of theory and compared these energies to those computed with coupled-cluster in the complete basis set limit. Interaction energies of the more distant dimers are easier to compute accurately and several of the methods tested are suitable as replacements for coupled-cluster through perturbative triples for all but the closest dimers. For our dataset, sub-kJ mol^{-1} accuracy can be obtained when calculating two-body interaction energies of dimers with separations shorter than 4 Å with CCSD(T)/CBS and dimers with separations longer than 4 Å with MP2.5/aug-cc-pVDZ, among other schemes, reducing the number of dimers to be computed with coupled-cluster by as much as 98%.

[a)]Electronic mail: sherrill@gatech.edu

I. INTRODUCTION

Polymorphs are crystals that have the same molecular composition but different packing schemes. Many common crystals, including acetaminophen and aspirin, are polymorphic.[1] Currently, the organic molecular crystal with the most fully characterized polymorphs is 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophenecabonitrile, also known as ROY, with 12 polymorphs. Four of those 12 have been discovered since 2019, and many of the 12 can crystallize under the same conditions simultaneously.[2,3]

Polymorphs exhibit different properties due to their different intermolecular arrangements. The properties that can be affected include solubility, which alters the bioavailibility of a drug. Antibiotics such as oxytetracycline and chloramphenicol palmitate have at least two different polymorphs, each with differing solubilities.[4] The HIV medication ritonovir has five polymorphs. When it was first released on the market, only one polymorph was known. After two years, some of the product began failing solubility tests, indicating the presence of a less soluble form. A second polymorph was determined to be the cause of this $\sim$50% decrease in solubility.[4–6]

Due to polymorph-dependent properties, predicting and ranking relative stabilities of polymorphs is a major focus in drug discovery. The stability of a polymorph is determined by its crystal lattice energy (CLE), the energy released when infinitely separated molecules come together to form a crystal.[7] A study of 8 ROY polymorphs showed that their CLE's all lie within 8 kJ mol^{-1} of each other.[8] Another study of 1061 crystals (made from 508 polymorphic organic molecules) showed that more than half of the pairs of polymorphs differed in lattice energy by no more than 2 kJ mol^{-1}; 95% of the pairs differed by less than 7.2 kJ mol^{-1}.[9] These small differences in energy necessitate high accuracy CLE calculations for ranking polymorphs.

Unfortunately, highly accurate calculations of lattice energies are extremely expensive due to the size of crystals. The many-body expansion (MBE) is a promising approach to obtain accurate CLE's using high-level methods from quantum chemistry.[10–12] This technique fragments a system into monomers, dimers, trimers, and so on, up to some limit. The CLE is then the sum of the monomer deformation energies (the change in monomer energies between the gas and crystal phase), and the N-body interaction energies among the dimer, trimers, etc. The reliability of the MBE has been shown in several molecular crystal studies,

and typically it is truncated at dimer or trimer interaction energy terms.[1,13–20] Not only does the MBE enable a reduction in the system size per energy calculation, but it also splits the CLE calculation into independent computational steps, making it pleasantly parallel.

Periodic density functional theory (DFT) methods are widely used to calculate crystal lattice energies, and they can provide accurate results when employing a method to account for London dispersion interactions, e.g., semi-empirical -D corrections.[21–23] However, even when using distributed-parallel implementations, periodic DFT computations can be time-consuming. The MBE may provide a more computationally efficient approach, and it also allows for some of the contributions to be computed using high-level wavefunction methods, which unlike DFT are systematically improvable towards the *ab initio* limit. Some groups have employed a hybrid approach that begins with periodic DFT as a baseline, and then the most important interactions (e.g., close dimers within the crystal) can be treated with high-level wavefunction methods like coupled-cluster with single, double, and perturbative triple excitations [CCSD(T)].[24,25]

Determining an accurate MBE-based method for the dimer and trimer calculations has been a point of research for our group and others. Ringer and Sherrill showed that while second-order perturbation theory (MP2)[26] produced poor results for the lattice energy of crystalline benzene, CCSD(T),[27] the "gold-standard" of computational chemistry, in the complete basis set (CBS) limit, was able to achieve accuracy of 1 kcal mol^{-1}.[14] Since then, the Chan and Sherrill groups have computed the lattice energy of crystalline benzene to sub-kJ mol^{-1} accuracy.[13,28] While this level of accuracy is desirable for applications like polymorph ranking, unfortunately, the $\mathcal{O}(N^7)$ scaling of canonical CCSD(T) makes even these fragment-based calculations computationally infeasible for molecular crystals of many larger molecules.

Efforts have been made to find lower-scaling methods that predict CLEs accurately,[1] but an avenue less explored is determining which fragments' interaction energies must be computed with CCSD(T)/CBS, and which can be determined accurately with a computationally cheaper method. Prior studies suggest that methods approximate to CCSD(T)/CBS may be used for interaction energies of dimers and trimers with larger intermolecular distances, with only a small reduction in accuracy and at a dramatically reduced computational cost.[29,30] Recently, our group explored this range-dependent approach while calculating the CLE of benzene with sub-kJ mol^{-1} accuracy.[28] Among other conclusions, it was determined that

MP2 with a double-ζ basis set was a reasonable choice for calculating the two-body contribution from medium to long range dimers, defined in that study as the 417 symmetry-unique dimers with inter-monomer separations (i.e. closest contact distances) between 4 and 30 Å. (Dimers beyond this range could be neglected.) The error was -1.87 kJ mol^{-1} relative to CCSD(T)/CBS. For the 278 unique dimers with separations between 20 and 30 Å, the error was just -0.10 kJ mol^{-1}. Only 3 dimers are short range, with separations less than 4 Å, which demand a high level of theory, such as CCSD(T)/CBS.[28]

While MP2 was effective in yielding accurate results at a greatly reduced computational cost for more distant dimers in the recent benzene study, it would be valuable to have high-quality benchmark energies for other molecular crystals and an assessment of additional approximate quantum chemical methods to see which level of theory consistently provides the best approximations at the lowest computational cost. In this study, we present CCSD(T)/CBS benchmark values for the two-body interaction energies of ice and 22 of the 23 molecular crystals of the X23 dataset.[31–33] We then compare the two-body interaction energies calculated with 15 different levels of theory to the CCSD(T)/CBS data and examine which of the less expensive methods achieve low errors for the contribution of the longer-range dimers to the CLE.

II. THEORETICAL METHODS

A. Dataset

The molecular crystals studied are those of the X23 dataset,[33] except for anthracene, which was too computationally expensive. The crystalline infographic files (CIF) were taken from the Cambridge Structural Database (CSD). The CSD codes for each CIF are those used in the C21 publication.[32] The revised version of C21, X23, includes hexamine and succinic acid. For these molecules, the CSD codes we use are HXMTAM and SUCACB12. We also include ice Ih (ICSD 27837). These crystal structures have been determined through either x-ray or neutron diffraction at various temperatures, ranging from 10 K to 298 K. The experimental method and temperature for each crystal is noted in the appropriate CIF.

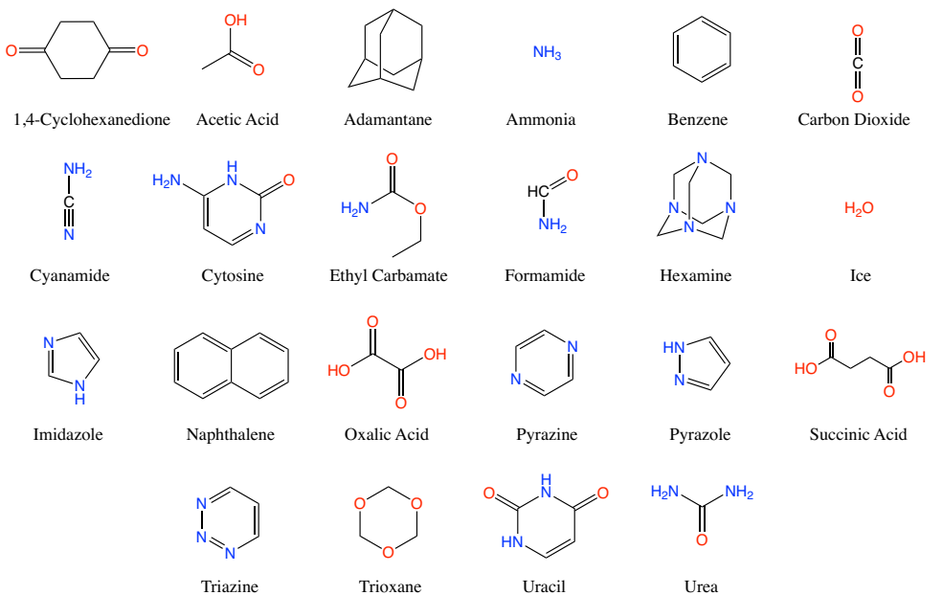


FIG. 1. The molecules studied are ice and every molecule from the X23 dataset, except anthracene. Each of these 22 molecules are used to form one of the crystals studied, and two different packings of oxalic acid (α and β) are considered, resulting in 23 crystals.

B. Two-Body Energies

The total energy of a molecular cluster can be calculated using the many-body expansion (MBE),

$$E^{\text{total}} = \sum_I \Delta E_I^{(1)} + \sum_{I<J} \Delta E_{IJ}^{(2)} + \sum_{I<J<K} \Delta E_{IJK}^{(3)} + \ldots, \tag{1}$$

where the superscript (N) denotes an N-body term, $\Delta E_I^{(1)}$ is the deformation energy of the monomer I, $\Delta E_{IJ}^{(2)}$ is the two-body interaction energy, defined as

$$\Delta E_{IJ}^{(2)} = E_{IJ} - E_I - E_J, \tag{2}$$

and $\Delta E_{IJK}^{(3)}$ is the non-additive three-body interaction energy,

$$\Delta E_{IJK}^{(3)} = E_{IJK} - E_I - E_J - E_K - \Delta E_{IJ}^{(2)} - \Delta E_{IK}^{(2)} - \Delta E_{JK}^{(2)}. \tag{3}$$

Calculating the lattice energy of a crystal, the energy needed to form a crystal from infinitely separated molecules, is complicated by the fact that crystals can be modeled by infinite solids. If this is the case, there will be an infinite number of monomers, dimers, trimers, etc. which contribute to the crystal lattice energy (CLE). For a finite result, we compute the CLE per monomer, or per mole of monomers. One reference monomer is chosen and is included in all N-body fragments, which provides the per monomer CLE,

$$E_I^{\text{CLE}} = \Delta E_I^{(1)} + \frac{1}{2}\sum_{I<J} \Delta E_{IJ}^{(2)} + \frac{1}{3}\sum_{I<J<K} \Delta E_{IJK}^{(3)} + \ldots, \tag{4}$$

under the many-body expansion.

Truncating the MBE at the two-body or three-body term typically provides good accuracy for molecular crystals. This study focuses only on the two-body term, which has been found to contribute 80-90% to the total crystal lattice energy.[34] All dimers considered include the reference monomer, and due to the periodic nature of the crystal, any other dimers should be equivalent to the ones we choose.

Our group's open-source software, CrystaLattE,[35] is used to build supercells from CIF files. Here the monomer geometries are assumed to be rigid, and we do not include any monomer deformation terms. The software extracts symmetry-unique dimers that have a minimum inter-monomer separation (closest-contact distance) less than some user-defined distance, as well as their degeneracy factors, in order to eliminate redundant calculations of

symmetry-equivalent dimers.[35] The energetic contribution of each symmetry-unique dimer to the CLE is the degeneracy factor multiplied by the dimer interaction energy, defined in equation 2, then divided by 2 for a per monomer contribution. We sum over these per monomer contributions to obtain a partial crystal lattice energy (PCLE), which is the two-body term in equation 4.

C. Benchmark Calculations

For the benchmark calculations, CrystaLattE extracted symmetry-unique dimers that have a minimum monomer separation less than 30 Å. Psi4 1.4rc3 was used to calculate the interaction energies.

A focal-point coupled-cluster with single, double, and perturbative triple excitations [CCSD(T)] scheme was used to extrapolate energies to a complete basis set limit (CBS).[36,37] This technique has been successful with correlation consistent basis sets for noncovalent interactions.[38–40] CCSD(T) in a large basis set may be approximated by

$$E(\text{CCSD(T)/Large}) \approx E(\text{MP2/Large}) + \delta_{MP2}^{CCSD(T)}/\text{Small}, \tag{5}$$

because smaller basis sets can often capture higher-order electron correlation effects. In equation 5, the "large" and "small" refer to relative size of the basis sets, and the delta term is

$$\delta_{MP2}^{CCSD(T)}/\text{Small} = E(\text{CCSD(T)/Small}) - E(\text{MP2/Small}). \tag{6}$$

A two-point extrapolation of the correlation energy typically provides a sufficient estimate of the CBS limit.[41] The large basis in equation 5 uses a two-point extrapolation of Dunning's augmented, correlation-consistent triple and quadruple-ζ basis sets, aug-cc-pVTZ and aug-cc-pVQZ,[42,43] abbreviated aTZ and aQZ, respectively. The small basis employed was aug-cc-pVDZ (aDZ). For brevity, we will refer to this specific CCSD(T)/CBS scheme as CCSD(T)/CBS(a[TQ]Z; δ:aDZ). In our recent study of crystalline benzene, CCSD(T)/CBS(a[TQ]Z; δ:aDZ) provided the total lattice energy contribution of dimers with separations between 8 and 14 Å within 0.01 kJ mol^{-1} of CCSD(T)/CBS(a[Q5]Z; δ:aTZ) results. For separations of $R \leq 4$ Å and $4 < R \leq 8$, the errors were 0.50 and 0.02 kJ mol^{-1}, respectively.[28]

Additionally, for CCSD(T) computations we employed the frozen natural orbital approximation with the default occupation number cutoff of 10^{-6},[44] and the convergence criteria for energy and self-consistent field (SCF) density were set to 10^{-10} a.u. The density fitting approximation was used for SCF and CC. All calculations included the counterpoise correction of Boys and Bernardi.[45,46]

D. Approximate Methods

A variety of density functional theory (DFT) and wavefunction methods were tested for their accuracy of approximating the two-body interaction energy. The methods include: B3LYP-D3BJ,[47] B97-D3BJ, B97-D,[21,48] HF-3c,[49] MP2.5,[50] MP2-D,[51] MP2,[26] PBE-D3BJ,[52,53] PBEh-3c,[54] and SAPT0[55,56]. The counterpoise correction of Boys and Bernardi was included with all methods except those for which a counterpoise correction is already effectively included as part of the procedure: HF-3c, PBEh-3c,[57] and SAPT0. B97-D includes Grimme's D2 correction, and -D3BJ refers to his D3 correction with Becke-Johnson damping.[21–23]

In general, the methods tested have the correct physics to describe intermolecular interactions with good accuracy, and thus to compute the two-body contribution to crystal lattice energies reliably. DFT partially incorporates the effects of electron correlation through its approximate exchange-correlation models, and thus it should provide reasonably accurate electron densities, which in turn will mean that interaction energy computations will effectively capture the underlying electrostatics, induction/polarization, and short range exchange-repulsion effects. Density functional approximations do not typically capture long-range electron correlation effects necessary to model London dispersion interactions, but that is remedied here by the use of semi-empirical -D corrections. The MP2 methods will likewise appropriately include electron correlation effects on electrostatics, induction/polarization, and exchange-repulsion, and will further appropriately describe long-range electron correlation leading to London dispersion interactions. At short intermolecular distances, correlations between pairs of electrons, and between three electrons simultaneously, will start to contribute to the interaction energy, and MP2 does not include such effects, so it will become less accurate at these distances. Similarly, standard correlation potentials in density functional theory may also become less accurate at short intermolecular distances. Thus, we

expect the largest errors to come from the dimers with the shortest intermolecular contacts (at the same time, these will also contribute the largest interaction energies).

SAPT0 is symmetry-adapted perturbation theory based on a Hartree–Fock description of the monomers. Thus, it neglects electron correlation corrections to the monomer electron densities before those densities interact through electrostatics, induction/polarization, and exchange-repulsion (it does, however, contain an MP2-like model of intermolecular London dispersion interactions). The lack of intramolecular electron correlation terms would, in principle, make it less reliable than the MP2-like or DFT-D approaches. However, we have often found good performance for SAPT0 when paired with modest basis sets like aug-cc-pVDZ or jun-cc-pVDZ (which drops the diffuse functions from aug-cc-pVDZ on H atoms, and the highest angular momentum functions on heavy atoms)[58] due to a favorable cancellation of errors.[59] Like the other methods considered, the approximations inherent in SAPT0 will become less appropriate at short intermolecular distances.

Finally, HF-3c and PBEh-3c are semi-empirical methods that both use prescribed small basis sets that may lead to inaccurate electrostatic, induction/polarization, and exchange-repulsion contributions to the interaction energy of a dimer. However, they compensate for this deficiency by the inclusion of short-range basis set incompleteness and superposition corrections, and they include London dispersion contributions through a semi-empirical -D correction. HF-3c has previously been shown to provide fairly accurate results for molecular crystals.[35,60,61]

MP2, MP2-D, MP2.5, and SAPT0 were tested with both aDZ and jun-cc-pVDZ (jDZ) basis sets. Only aDZ basis sets were used for B3LYP-D3BJ, B97-D3BJ, B97-D, and PBE-D3BJ. Additionally, PBE-D3BJ was tested with the Karlsruhe basis set def2-TZVP, which we denote as TZVP.[62] Default basis sets were used for HF-3c and PBEh-3c. The default grid in Psi4 (i.e., 75 radial and 302 spherical points) was used in all DFT calculations with the exception of PBEh-3c, where radial and spherical points were set to 99 and 590, respectively.

When comparing the relative error of these methods to CCSD(T)/CBS(a[TQ]Z; δ:aDZ), only dimers with a minimum inter-monomer separation less than 20 Å were used. While we could have considered dimers with separations up to 30 Å, we found that the dimers with separations between 20 and 30 Å only contributed up to 0.2% of the two-body lattice energy in 20 of the 23 molecules tested, as shown in Table I. For most of the crystals, the neglected longer-range dimers ($20 \leq R < 30$ Å) are only worth about 0.15 kJ mol^{-1} or less, although

in a few cases they contribute more, up to a maximum of 0.89 kJ mol^{-1} in the case of urea. Thus, we decided to ignore the dimers with inter-monomer separations of 20 Å and greater, in the study of approximate methods, to greatly reduce the number of computations that would accumulate from such a large number of dimers in this separation range (and the significant number of approximate methods considered).

Nearly all computations were performed with various versions of Psi4[63]: 1.4rc3, 1.4.1, and 1.5, and there are no notable differences in each Psi4 version listed pertaining to this project. Convergence criteria were set identically to those of the benchmark computations. The density-fitting approximation was employed for SCF and perturbation theory computations. For PBEh-3c calculations of hexamine dimers, Orca version 5.0.1[64] was used with defgrid3. Hexamine computations with this method in Psi4 exhibited very small systematic errors that accumulated when closely examining CLEs. This is apparently due to a bug in the implementation of this method in Psi4, or the interface to the underlying libXC library,[65] which we were unable to trace as of this writing.

III. RESULTS

A. Benchmark Energies

For the 23 crystals studied, we first computed the contributions to the crystal lattice energies from the dimers with an inter-monomer separation, R, less than 30 Å using CCSD(T)/CBS(a[TQ]Z; δ:aDZ). A recent report of SAPT0 calculations on the present dataset showed that two-body contributions to CLEs from dimers with $R < 60$ Å converged to within 1 kJ mol^{-1} when accumulating interactions of dimers only up to $R = 30$ Å, and to within 0.5 kJ mol^{-1} for all but three crystals: cyanamide, cytosine, and urea.[66] Figure 2 shows the cumulative two-body CCSD(T)/CBS(a[TQ]Z; δ:aDZ) CLE as dimers with longer inter-monomer distances are included for acetic acid (polar), adamantane (nonpolar), imidazole (polar, aromatic), and benzene (nonpolar, aromatic). Similar figures for all other crystals in the dataset are included in the supplementary material. From these figures, we see that contributions from longer-range dimers are small but significant when aiming for high accuracy crystal lattice energies.

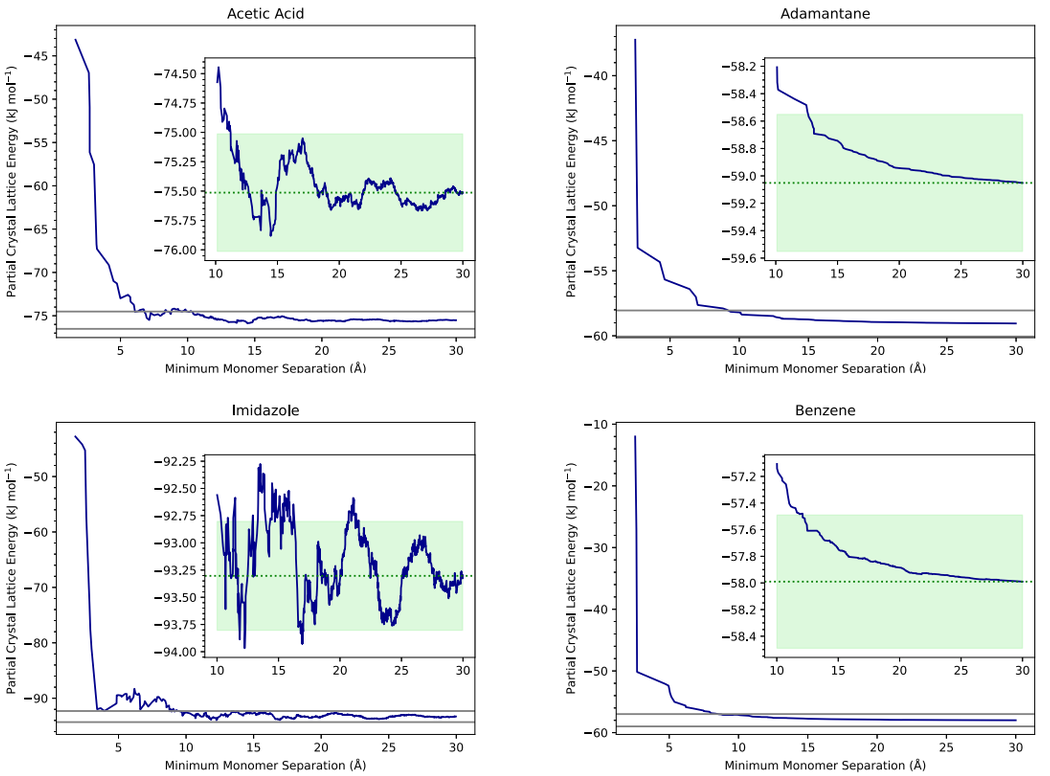


FIG. 2. Cumulative partial crystal lattice energy (kJ mol^{-1}) of four crystals as dimers with longer inter-monomer separations are included. Interaction energies are calculated with CCSD(T)/CBS(a[TQ]Z; δ:aDZ). The gray bars are 1 kJ mol^{-1} above and below the final PCLE. The green box highlights a range of 1 kJ mol^{-1} around the final PCLE ($\pm$0.5 kJ mol^{-1}).

The calculated two-body contributions, from dimers with $R < 30$ Å, for each crystal are reported in Table I, in addition to the number of symmetry-unique dimers included for each crystal. These values are also divided into subsets corresponding to ranges of minimum inter-monomer separations. The short-range dimers ($R < 4$ Å) only comprise between 0.4 and 5.0% of dimers with separations less than 30 Å, yet their contribution to the two-body PCLE is more than 83% in all molecules, and over 90% in the majority of cases studied. This also supports the idea that the closest dimers should be given priority when choosing which should be computed with CCSD(T)/CBS. The energetic contributions of the mid-range dimers ($4 \leq$ R $<$ 8 Å) are less than 15% for each crystal. Although the number of dimers in each range subset increases rapidly, due to the increasing volume of shells at distance R from the center of the reference monomer, there is a drastic decrease in the energetic contributions from long-range dimers, whose inter-monomer separations are between 8 and 20 Å. The 75 long-range dimers of trioxane contribute 4.1% to the PCLE, the maximum percent contribution of long-range dimers, whereas for oxalic acid α, 179 long-range dimers contribute 0.2% of the PCLE. The subset for the $20 \leq R < 30$ Å range contains at least 62% of the dimers for each crystal, yet only contributes energetically up to 0.8% in the case of urea, and 0.1-0.2% in the majority of cases studied. The largest energetic contributions from this subset, those greater than 0.2 kJ mol^{-1}, are from some of the more polar molecules: formamide, urea, and uracil. This can be attributed partially to slow convergence of the electrostatic energy with respect to intermonomer distance, which can also be used to explain the noise in Figure 2. Range-dependent convergence of electrostatics, exchange, induction, and dispersion for each of these crystals have been reported recently by our group.[66]

TABLE I. Two-body crystal lattice energy contributions (kJ mol^{-1}) and the number of symmetry-unique dimers (N_{dim}) with a minimum inter-monomer separation, R, less than 30 Å, in addition to those for subsets of R. Percentages of the total two-body energies and dimer counts are listed in parentheses. Energies are computed with CCSD(T)/CBS(a[TQ]Z; δ:aDZ).

	Total		$R < 4$		$4 \leq$ R < 8		$8 \leq$ R < 20		$20 \leq$ R < 30	
	Energy	N_{dim}	Energy	N_{dim}	Energy	N_{dim}	Energy	N_{dim}	Energy	N_{dim}
1,4-Cyclohexanedione	-97.97	596	-92.11	7	-4.86	20	-0.82	182	-0.17	387
$P2_1$			(94.0%)	(1.2%)	(5.0%)	(3.4%)	(0.8%)	(30.5%)	(0.2%)	(64.9%)
Acetic Acid	-75.51	935	-67.29	7	-7.73	24	-0.56	282	0.06	622
$Pna2_1$			(89.1%)	(0.7%)	(10.2%)	(2.6%)	(0.7%)	(30.2%)	(-0.1%)	(66.5%)
Adamantane	-59.05	128	-53.23	2	-4.40	5	-1.31	41	-0.11	80
$P\bar{4}2_1c$			(90.1%)	(1.6%)	(7.4%)	(3.9%)	(2.2%)	(32.0%)	(0.2%)	(62.5%)
Ammonia	-36.38	626	-35.39	3	-0.35	12	-0.49	184	-0.15	427
P1			(97.3%)	(0.5%)	(1.0%)	(1.9%)	(1.3%)	(29.4%)	(0.4%)	(68.2%)
Benzene	-57.99	420	-50.16	3	-6.67	14	-1.05	125	-0.11	278
Pbca			(86.5%)	(0.7%)	(11.5%)	(3.3%)	(1.8%)	(29.8%)	(0.2%)	(66.2%)
Carbon Dioxide	-30.11	251	-25.10	1	-4.46	8	-0.52	73	-0.03	169
$Pa\bar{3}$			(83.4%)	(0.4%)	(14.8%)	(3.2%)	(1.7%)	(29.1%)	(0.1%)	(67.3%)
Cyanamide	-77.56	1463	-71.61	9	-5.28	30	-0.66	428	-0.01	996
Pbca			(92.3%)	(0.6%)	(6.8%)	(2.1%)	(0.9%)	(29.3%)	(0.0%)	(68.1%)
Cytosine	-166.28	690	-155.08	7	-5.04	18	-6.07	215	-0.09	450
$P2_12_12_1$			(93.3%)	(1.0%)	(3.0%)	(2.6%)	(3.7%)	(31.2%)	(0.1%)	(65.2%)
Ethyl Carbamate	-82.04	977	-80.26	12	-0.57	24	-1.14	301	-0.07	640
$P\bar{1}$			(97.8%)	(1.2%)	(0.7%)	(2.5%)	(1.4%)	(30.8%)	(0.1%)	(65.5%)
Formamide	-76.76	1534	-77.07	9	1.81	34	-0.95	450	-0.55	1041
$P2_1/n$			(100.4%)	(0.6%)	(-2.4%)	(2.2%)	(1.2%)	(29.3%)	(0.7%)	(67.9%)
Hexamine	-87.13	40	-82.93	2	-2.45	2	-1.61	14	-0.13	22
$I\bar{4}3m$			(95.2%)	(5.0%)	(2.8%)	(5.0%)	(1.9%)	(35.0%)	(0.2%)	(55.0%)
Ice	-34.05	1483	-34.05	6	0.24	32	-0.23	434	-0.01	1011
$P6_3cm$			(100.0%)	(0.4%)	(-0.7%)	(2.2%)	(0.7%)	(29.3%)	(0.0%)	(68.2%)
Imidazole	-93.30	1070	-92.30	10	2.55	27	-3.61	323	0.06	710
$P2_1/c$			(98.9%)	(0.9%)	(-2.7%)	(2.5%)	(3.9%)	(30.2%)	(-0.1%)	(66.4%)
Naphthalene	-84.43	386	-75.65	5	-6.83	11	-1.83	123	-0.13	247
$P2_1/a$			(89.6%)	(1.3%)	(8.1%)	(2.8%)	(2.2%)	(31.9%)	(0.2%)	(64.0%)
Oxalic Acid α	-108.25	595	-96.81	4	-11.07	15	-0.26	179	-0.12	397
Pcab			(89.4%)	(0.7%)	(10.2%)	(2.5%)	(0.2%)	(30.1%)	(0.1%)	(66.7%)
Oxalic Acid β	-126.70	712	-114.86	5	-10.26	17	-1.49	218	-0.09	472
$P2_1/c$			(90.7%)	(0.7%)	(8.1%)	(2.4%)	(1.2%)	(30.6%)	(0.1%)	(66.3%)
Pyrazine	-63.04	319	-57.80	4	-4.39	10	-0.76	97	-0.08	208
Pmnn			(91.7%)	(1.3%)	(7.0%)	(3.1%)	(1.2%)	(30.4%)	(0.1%)	(65.2%)
Pyrazole	-66.79	1269	-58.22	10	-7.87	34	-0.77	381	0.06	844
$P2_1cn$			(87.2%)	(0.8%)	(11.8%)	(2.7%)	(1.1%)	(30.0%)	(-0.1%)	(66.5%)
Succinic Acid	-133.87	525	-121.15	5	-11.76	18	-0.87	162	-0.09	340
$P12_1/a1$			(90.5%)	(1.0%)	(8.8%)	(3.4%)	(0.7%)	(30.9%)	(0.1%)	(64.8%)
Triazine	-58.36	199	-53.61	3	-3.67	8	-1.02	63	-0.08	125
$R\bar{3}c$			(91.8%)	(1.5%)	(6.3%)	(4.0%)	(1.7%)	(31.7%)	(0.1%)	(62.8%)
Trioxane	-60.60	256	-54.17	3	-3.95	8	-2.48	75	0.00	170
R3c			(89.4%)	(1.2%)	(6.5%)	(3.1%)	(4.1%)	(29.3%)	(0.0%)	(66.4%)
Uracil	-127.55	858	-116.52	9	-12.16	24	1.37	262	-0.24	563
$P2_1/a$			(91.4%)	(1.0%)	(9.5%)	(2.8%)	(-1.1%)	(30.5%)	(0.2%)	(65.6%)
Urea	-115.56	311	-113.18	5	-2.76	8	-0.50	101	0.89	197
$P\bar{4}2_1m$			(97.9%)	(1.6%)	(2.4%)	(2.6%)	(0.4%)	(32.5%)	(-0.8%)	(63.3%)

B. Performance of Approximate Methods

With relevant benchmark data, 15 levels of theory approximate to CCSD(T)/CBS are tested for their accuracy in calculating the two-body lattice energy contributions for each of the 23 crystals. Timings are shown in Figure 3, where the wall time needed to compute the interaction energy of an adamantane dimer with each level of theory is compared. We arbitrarily chose an adamantane dimer that has an inter-monomer separation of ∼7 Å after observing no correlation between inter-monomer separation and computation wall time. All methods tested, excluding MP2.5, returned the interaction energy in less than 5 minutes on 14 cores of an Intel i9-10980XE processor using Psi4 1.4rc3.[63] HF-3c is the least expensive method studied and calculates the adamantane dimer's interaction energy in just 9 seconds on 14 cores. Existing studies on the X23 dataset have shown that HF-3c performs similarly to some dispersion-corrected DFT methods.[60,61] MP2.5 is by far the most expensive approximate method considered, taking one hour to complete with the aDZ basis set and 15 minutes with jDZ. Still, all methods are drastically less expensive than CCSD(T)/CBS(a[TQ]Z; δ:aDZ), which has a wall time of 2 days, 13 hours, and 17 minutes for the corresponding calculation.

Errors for calculating the two-body contribution with the 15 low-cost levels of theory relative to CCSD(T)/CBS(a[TQ]Z; δ:aDZ) are shown in the left panel of Figure 4. For the error analysis study, given the very small contribution from dimers $20 \leq R < 30$ Å, and the large number of them, we limited ourselves only to the significant dimers, $R < 20$ Å. Nevertheless, we strongly expect the reliability of various methods in the neglected range ($20 \leq R < 30$ Å) to be at least as good as in the range $8 \leq R < 20$ Å. In general, it appears that more distant dimers are much easier to model with more approximate methods, as we have already observed for crystalline benzene,[28] and as we discuss in more detail below.

The left panel of Figure 4 reports errors in the two-body crystal lattice energy when using each approximate method rather than the benchmark-level coupled cluster. The absolute errors range from 0.1 kJ mol^{-1} (cytosine with PBE-D3BJ/aDZ) to 50.6 kJ mol^{-1} (succinic acid with MP2-D/jDZ). The three darkest columns correspond to the use of the smallest standard basis sets considered here, jun-cc-pVDZ. (HF-3c and PBEh-3c use even smaller basis sets, but those methods are tuned specifically for their associated basis sets.) Overall, the jDZ basis sets, which are used for only MP2, MP2-D, and MP2.5, return some especially

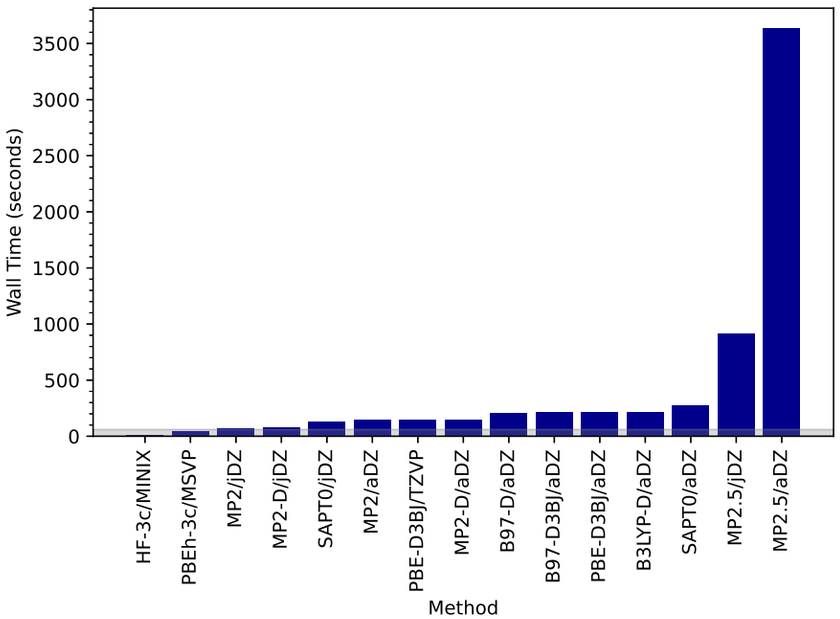


FIG. 3. Wall times (seconds) of calculating the interaction energy of an adamantane dimer with a minimum inter-monomer separation of 6.997 Å using Psi4 1.4rc3 and 14 cores of an Intel i9-10980XE processor. The gray shaded region represents one minute. This calculation with CCSD(T)/CBS(a[TQ]Z; δ:aDZ) has a wall time of ∼2.5 days.

large errors, up to 50.6 kJ mol^{-1}, signaling that these levels of theory tend to underbind the dimers of this dataset. The largest of these errors tend to be associated with molecules containing carbonyl groups (e.g., 1,4-cyclohexanedione, cytosine, ethyl carbamate, formamide, oxalic acid, succinic acid, uracil, and urea). We hypothesize that diffuse d functions (absent in the jDZ basis set but present in aDZ) are important for computing accurate MP2 interaction energies in these systems. Using the aDZ basis sets instead alleviates some error, reducing the maximum error to 23.8 kJ mol^{-1}, but these methods still tend to underbind. The DFT methods tend to overbind, resulting in negative interaction energy errors. Interestingly, SAPT0/aDZ overbinds, but switching to the jDZ basis set results in underbound dimers.

Some of the largest errors appear for succinic acid, oxalic acid (α and β), and cytosine. All four of these crystals have dimers with closest-contact separations between 1 and 2 Å. Due to this observation, we compared the closest inter-molecular contact within each crystal to each crystal's maximum PCLE error (over all approximate methods). Figure S-8 in the

supplementary material shows no correlation between the two.

As stated previously, it has been suggested that methods approximate to CCSD(T)/CBS may be appropriate for long-range dimers, only reducing accuracy a small amount and certainly reducing computational cost. Therefore, we have computed PCLE errors under a range-dependent scheme. We compute the interaction energies of dimers with separations between 3 and 20 Å with the low-cost methods and allow those dimers with separations less than 3 to be computed with CCSD(T)/CBS. We chose 3 Å to be the switchover distance because all of the crystals, except crystalline carbon dioxide, have some (but very few) dimers with separations below this distance, keeping the number of dimers to be computed with CCSD(T)/CBS low. Carbon dioxide has no dimers with separations below 3 Å. For each of the 23 crystals, no more than 9 symmetry-unique dimers were computed with CCSD(T)/CBS. (A detailed breakdown of dimer counts are included in the supplementary material.) Upon implementing this switchover distance of 3 Å, there resulted a notable decrease in errors compared to computing all dimers with the approximate method. Specific errors for each crystal/method combination can be seen in the right panel of Figure 4. The range of absolute errors reduces from 0.1 – 50.6 kJ mol^{-1} (when computing all dimers $R <$ 20 Å with approximate methods) to 0.0(02) – 18.6 kJ mol^{-1}, and the levels of theory tend to maintain overbinding or underbinding predictions. The crystal with the largest long-range errors is CO_2; however, this is simply because the nearest-neighbor dimer in CO_2 has a closest-contact distance between 3–4 Å (as previously discussed) and is included in the right panel of Fig. 4. Among the crystals with the next-largest errors, several have large molecular dipole moments: e.g., cytosine, formamide, and uracil (dipole moments for these molecules are available in the supporting material of Ref. 66). However, large errors are also found for succinic acid, which has a small dipole moment, and oxalic acid β, which has a zero dipole moment (although these molecules do have polar groups and nonzero quadrupole moments). Overall, we do not see a clear correlation between the molecular structure and which crystals exhibit the largest errors in the right-hand panel of Figure 4, apart from the predominance of carbonyl groups in the systems with the largest errors, especially for MP2-type methods with the truncated jun-cc-pVDZ basis set, as already noted.

Mean absolute errors (MAEs) for each method, averaged over all 23 crystals, are presented in Figure 5 for both distance ranges (0–20 Å and 3–20 Å). When considering all dimers, $R <$ 20 Å, the lowest MAEs hover around 4-5 kJ mol^{-1} with B3LYP-D3BJ, PBE-D3BJ,

PBEh-3c, and SAPT0. Excluding those dimers with $R < 3$ Å returns MAEs less than 5 kJ mol^{-1} for almost all methods. B3LYP-D3BJ, PBE-D3BJ, and PBEh-3c continue to have the lowest MAEs, around 1 kJ mol^{-1}. MAE values are also tabulated in the supplementary material.

MP2-type methods perform well despite our use of smaller basis sets (jDZ and aDZ). Post-Hartree–Fock methods like MP2 tend to be fairly sensitive to basis set, and this is certainly true for intermolecular interactions. However, this basis set sensitivity decreases rapidly with intermolecular separation,[28,67–69] allowing accurate results to be obtained for the more distant dimers despite the use of small basis sets.

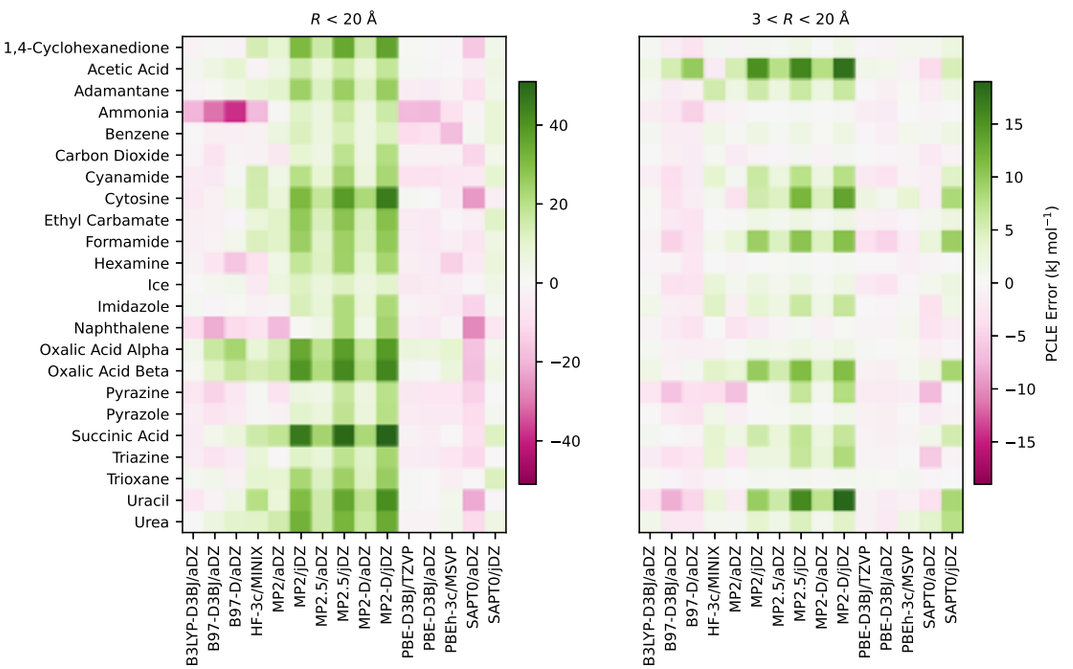


FIG. 4. Errors (kJ mol^{-1}) in the two-body crystal lattice energy contribution computed with less expensive methods, relative to CCSD(T)/CBS(a[TQ]Z; δ:aDZ). Dimers contributing to the energy calculations have minimum monomer separations, R, less than 20 Å (left) and between 3 and 20 Å (right).

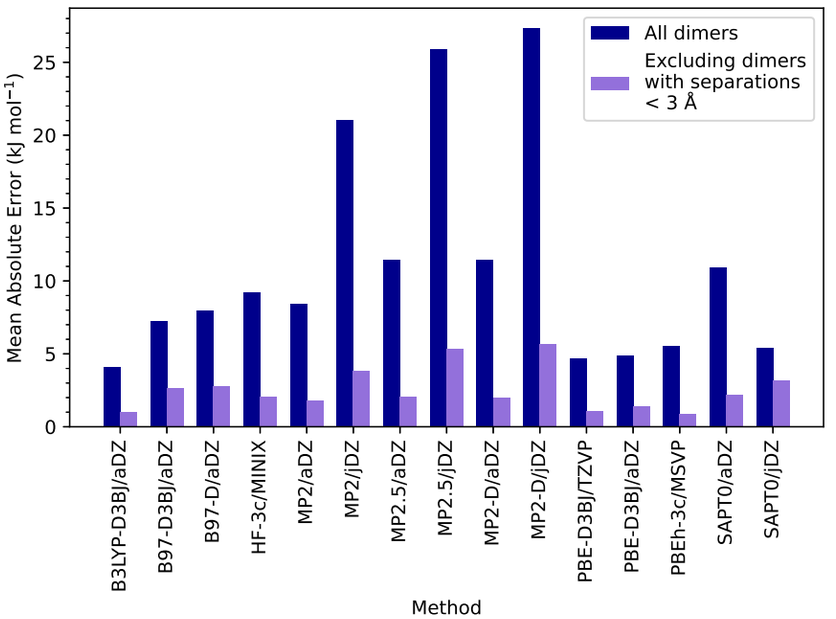


FIG. 5. Absolute errors (kJ mol^{-1}) in the two-body crystal lattice energy contribution for each method, averaged over the 23 crystals studied. Dimers contributing to the energy calculations have minimum monomer separations, R, less than 20 Å or between 3 and 20 Å.

C. Range Dependence of Approximate Methods

With the present results confirming that approximate methods may be better suited to compute interaction energies of dimers with longer inter-monomer separations rather than shorter, we aimed to find certain switchover distances where calculating interaction energies of dimers with separations longer than that distance with an approximate method would not greatly sacrifice accuracy. Specifically, we determined such distances where using approximate methods past that distance (up to inter-monomer separations of 20 Å) would return errors around 4, 1, 0.5, and 0.1 kJ mol^{-1} for the two-body contribution in all 23 crystals. These switchover distances are listed in Table II, and dimers with closest inter-monomer distances below the switchover distance would still be computed with CCSD(T)/CBS. For example, for any of the crystals considered, computing dimers with separations $R < 3$ Å with CCSD(T)/CBS(a[TQ]Z; δ:aDZ) and dimers with $3 \leq R < 20$ Å with B3LYP-D3BJ/aDZ, PBE-D3BJ/TZVP or PBEh-3c produces an error less than 4 kJ mol^{-1} relative to the benchmark CCSD(T)/CBS(a[TQ]Z; δ:aDZ) values for $R < 20$ Å. Here, 3 Å is considered the switchover distance.

While closest contact separations are presented here, it is also helpful to consider the switchover distances in terms of nearest-neighbor interactions. Generally, for molecular crystals, nearest-neighbor molecules lie within a 4 Å radius of the asymmetric unit, and one can expect there to be 10-15 nearest-neighbors.[70,71] Using these definitions, we predict that using a specific a switchover distance of 3 Å will have most, but possibly not all, of the nearest-neighbor interactions computed with CCSD(T). Switchover distances larger than 3 Å, presented later, will be more likely to capture all of the nearest-neighbor interactions, and maybe even next-nearest interactions.

Figure 6 highlights those methods that give errors less than 4, 1, 0.5, and 0.1 kJ mol^{-1} for all 23 crystals using the specific switchover distances noted in the figure. These schemes were chosen because they produced the desired error with the least number of dimers being computed with CCSD(T)/CBS (according to Table II), therefore keeping costs low. DFT methods B3LYP-D3BJ, PBE-D3BJ, and PBEh-3c succeed in keeping errors below 4 kJ mol^{-1} for all crystals with a switchover distance at 3 Å. This distance corresponds to calculating between 89 and 100% of the symmetry-unique dimers, depending on the crystal, with one of the three previously mentioned DFT methods, and only the remaining fraction

TABLE II. Minimum monomer separations, R, at which a non-benchmark method is used to compute interaction energies of dimers with separations longer than R and desired accuracy of the two-body contribution to the lattice energy is achieved for all systems studied. Error is relative to CCSD(T)/CBS(a[TQ]Z; δ:aDZ). Separations denoted with x indicate that the method can not achieve listed accuracy for any cutoff $R < 20$ Å.

Method	**Error (kJ mol^{-1})**			
	<4	<1	<0.5	<0.1
B3LYP-D3BJ/aDZ	3	5	7	16
B97-D3BJ/aDZ	4	7	7	14
B97-D/aDZ	5	10	13	18
HF-3c/MINIX	4	10	16	x
MP2.5/aDZ	4	4	6	10
MP2.5/jDZ	4	7	8	18
MP2-D/aDZ	4	6	7	12
MP2-D/jDZ	4	7	9	14
MP2/aDZ	4	8	10	16
MP2/jDZ	4	7	9	15
PBE-D3BJ/TZVP	3	5	7	17
PBE-D3BJ/aDZ	4	5	7	14
PBEh-3c/MSVP	3	14	x	x
SAPT0/aDZ	4	7	16	x
SAPT0/jDZ	4	9	16	x

of dimers (the closest ones) with CCSD(T)/CBS. Figure S-2 of the supplementary material provides the specific absolute errors per crystal when calculating each crystal's two-body lattice energy with each of these DFT methods after a switchover distance of 3 Å, rather than the mean absolute error.

For PCLE errors less than 1, 0.5, and 0.1 kJ mol^{-1}, MP2.5/aDZ out-performs the other methods. It consistently allows for a cutoff sooner than other methods, reducing the number

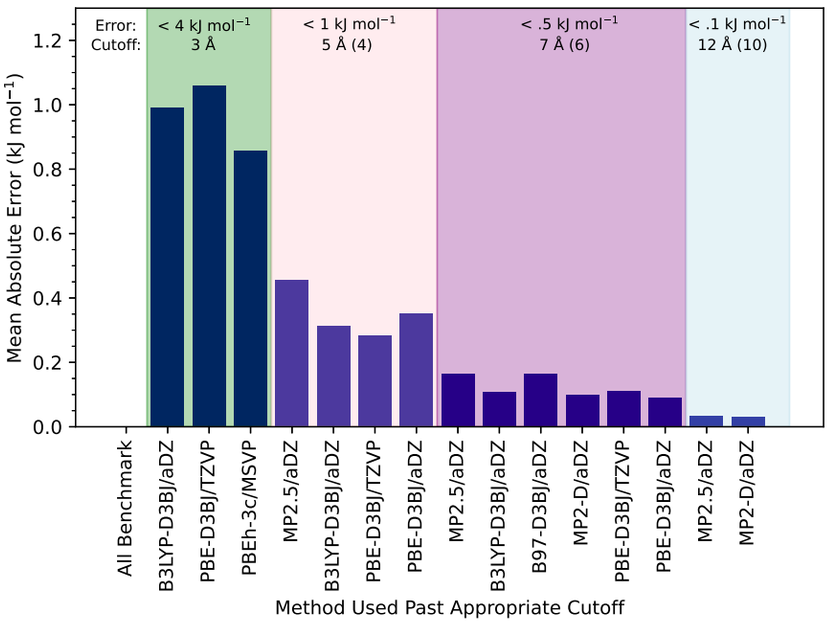


FIG. 6. Absolute errors (kJ mol^{-1}) in the two-body crystal lattice energy contribution for selected methods, averaged over the 23 crystals studied. The approximate methods are used to compute all dimers with minimum inter-monomer separations beyond the given switchover distances provided in Table II. Methods in colored region correspond to those which have less than the error indicated *for every crystal*. Slightly larger errors for MP2.5 vs. some other approximate methods are due to switching over to MP2.5 earlier (MP2.5 switchover distances noted in parentheses). The benchmark method is CCSD(T)/CBS(a[TQ]Z; δ:aDZ).

of dimers that need to be computed with CCSD(T)/CBS, and therefore reducing CPU time. The good performance of MP2.5 is perhaps not surprising, as it is the most computationally sophisticated method considered here short of the benchmark CCSD(T)/CBS results, and it has previously been shown to be reliable for non-covalent interactions.[72–75] In the event that one would want a method that is cheaper than MP2.5, Figure 6 also includes other methods which yield good results at low computational cost. For an error less than 1 kJ mol^{-1}, B3LYP-D3BJ and PBE-D3BJ can be used for dimers with inter-monomer separations between 5 and 20 Å. A switchover distance of 7 Å must be used to reduce the error to less than 0.5 kJ mol^{-1}. B3LYP-D3BJ, B97-D3BJ, MP2-D, and PBE-D3BJ are all possible choices, while the lowest MAE is that of PBE-D3BJ/aDZ. (MAEs are listed in the supplementary material.) Finally, MP2-D may be used for dimers with separations longer than

12 Å to compute a PCLE with an error less than 0.1 kJ mol^{-1} relative to the benchmark method.

As previously discussed, MP2, a very common method, was chosen when applying this two-layer approach to benzene in our group's recent study.[28] The current study reveals that this was a fine choice, but there are other methods which will yield a lower error. If MP2/aDZ were the level of theory used with a switchover distance of 3 Å, 21 of the 23 crystal lattice energies would have an error less than 4 kJ mol^{-1}. The errors for carbon dioxide and pyrazine are 5.2 and -6.6 kJ mol^{-1}, respectively. Changing the switchover distances to those considered in Figure 6 (5, 7, and 12 Å) returns errors within the cutoffs of Figure 6 (1, 0.5, and 0.1 kJ mol^{-1}, respectively) for 20 of the 23 crystals. Benzene, naphthalene, and pyrazine present larger errors. Naphthalene shows the largest errors: -2.9 kJ mol^{-1} when using a switchover distance of 5 Å, -1.0 kJ mol^{-1} for 7 Å, and -0.2 kJ mol^{-1} for 12 Å. (Errors for calculating the two-body lattice energy of each crystal with switchover distances of 3, 5, 7, and 12 Å to MP2/aDZ are listed in Table S-8 of the supplementary material.) Using MP2/aDZ in all of the schemes of Figure 6, instead of the method listed, resulted in MAEs that were between 1.3 and 2.6 times larger than those shown in Figure 6. All of the methods highlighted in Figure 6 show superior performance to MP2/aDZ for two-body contributions to the lattice energy, and most of them (all except MP2.5), have a comparable or less expensive computational cost. Therefore, it is recommended that methods other than MP2 be used for systems like these studied as they are more reliable for this dataset.

An interesting case study is the comparison of error convergence between the two polymorphs of oxalic acid. The top two panels of Figure 7 show errors in the PCLE, relative to the final CCSD(T)/CBS PCLE, as the switchover distance between CCSD(T)/CBS and four approximate methods increases. The errors of oxalic acid α's PCLE become close to 1-2 kJ mol^{-1} around a switchover distance of 3 Å and converge even tighter as the switchover distance increases. The results are similar for β, but MP2.5/aDZ does not achieve a comparable error until $\sim$ 4 Å. It should be noted that ranking polymorph stability depends on the accuracy of relative energies between crystals rather than the accuracy of individual CLEs. The bottom panel of Figure 7 shows the convergence of the error in the relative polymorph energy as a function of the switchover distance. The error in the relative polymorph energy is also nearly converged by about 4 Å, but it tends to be significantly smaller in magnitude (generally $\sim$ 0.2 kJ mol^{-1} or less by 4 Å) than the errors in the PLCEs, especially for MP2.5.

For this test case, even though the errors in the energy difference between the polymporphs are less than the errors in the PCLEs, a switchover distance of 4 Å provides fairly accurate answers for both individual PCLEs and relative energies between polymorphs.

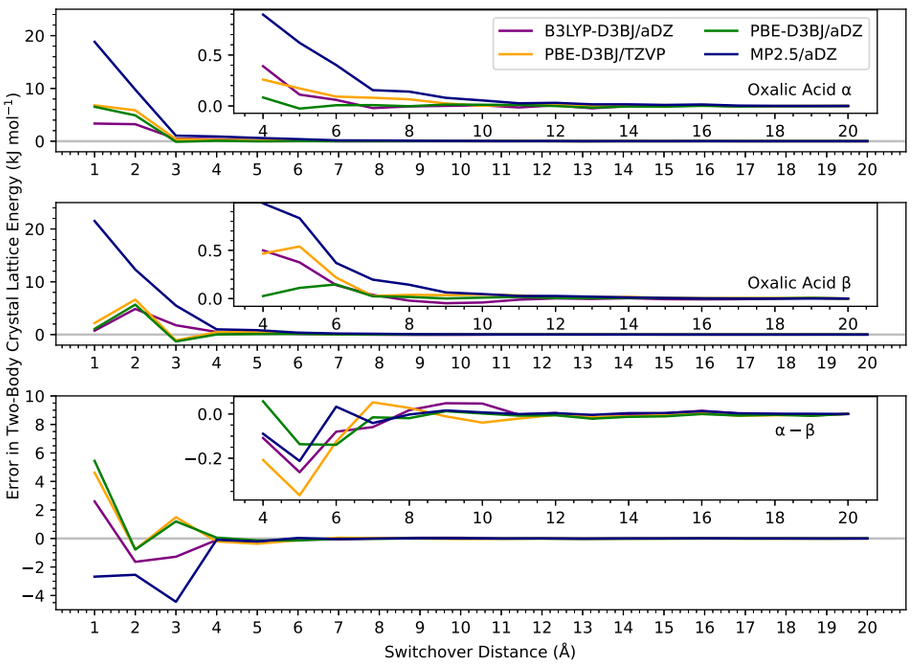


FIG. 7. Error in two-body CLE, relative to the final CCSD(T)/CBS PCLE, when all dimers are considered ($R < 20$ Å). CCSD(T)/CBS is used to calculate interaction energies of dimers with R less than the switchover distance (x-axis), and an approximate method is used for interaction energies of those dimers with R longer than the switchover distance. The top panel is for oxalic acid α only, the middle is for oxalic acid β, and the bottom panel shows the relative errors between the two polymorphs.

For the case that an approximate method is used for all oxalic acid calculations rather than allowing some of the closest dimers to be calculated with CCSD(T)/CBS, Figure S-7 in the supplementary material shows the error in PCLE as dimers with longer closest contact distances are included. The figure also shows the error in the relative energy between these two polymorphs as more dimers are included. The PCLE returned when dimers $R < 20$ Å are included is nearly achieved when dimers $R < 7$ Å are accounted for in both polymorphs. One might wonder whether the relative energy between the polymorphs can be captured by including fewer dimers than required to converge the individual polymorph lattice energies.

The Figure indicates that, while the errors in the polymorph energy difference initially converge much more quickly with intermolecular distance than the individual polymporph lattice energies, oscillations in the energy difference are not fully damped out until around 7 Å, where the individual polymorph energies also converge.

Figure 8 shows the substantial reduction of computational effort for the recommended schemes of Figure 6. Timings were obtained for the adamantane dimer of Figure 3, and from this data we have estimated the amount of time it would take to compute all the dimers of adamantane in each of the distance regimes. Using CCSD(T)/CBS(a[TQ]Z; δ:aDZ) for all 48 symmetry-unique adamantane dimers ($R < 20$ Å) would take over 40,000 core-hours (122 wall-time days using 14 cores), whereas a lattice energy within 1 kJ mol^{-1} of that result could be obtained by using MP2.5/aDZ past a switchover distance of 4 Å and CCSD(T)/CBS below 4 Å, which would take just over 2500 core-hours (about one week of wall time on 14 cores). If only one approximate method is chosen for longer-range dimers, and 1, 0.5, or 0.1 kJ mol^{-1} error is desired, then MP2.5 saves the most computational time of the methods considered. Even though it is the most expensive of the approximate methods for a single dimer computation, its increased accuracy for shorter-range dimers reduces the number of dimers that need to be computed with CCSD(T)/CBS, a method which is far more computationally expensive. As stated previously, these dimer calculations are independent of each other. In the event that one could obtain as many nodes as there are individual dimer calculations, all interaction energies would be calculated once one CCSD(T)/CBS calculation finishes, no matter the scheme.

Another option that keeps both errors and computational cost low would be a three-level scheme that treats short, medium, and long-range dimers each with a different level of theory. While MP2.5/aDZ returns low errors with medium to long-range switchover distances, it would be ideal to treat the long-range dimers with an even cheaper level of theory due to the large number of dimers in this range and the considerable cost of MP2.5 relative to the other approximate methods considered here. We tested a three-level scheme for benzene such that dimers with $R < 4$ Å had interactions calculated with CCSD(T)/CBS(a[TQ]Z; δ:aDZ), $4 \leq R < 5$ Å with MP2.5/aDZ, and $R \geq 5$ Å with B3LYP-D3BJ/aDZ. For benzene, this means that the three-level system reduces the number of dimers computed with MP2.5/aDZ from 139 to 2. When using this three-level scheme, the error relative to the calculating the interaction energies of $R < 20$ with CCSD(T)/CBS was 0.09 kJ mol^{-1}. This is 0.27 kJ

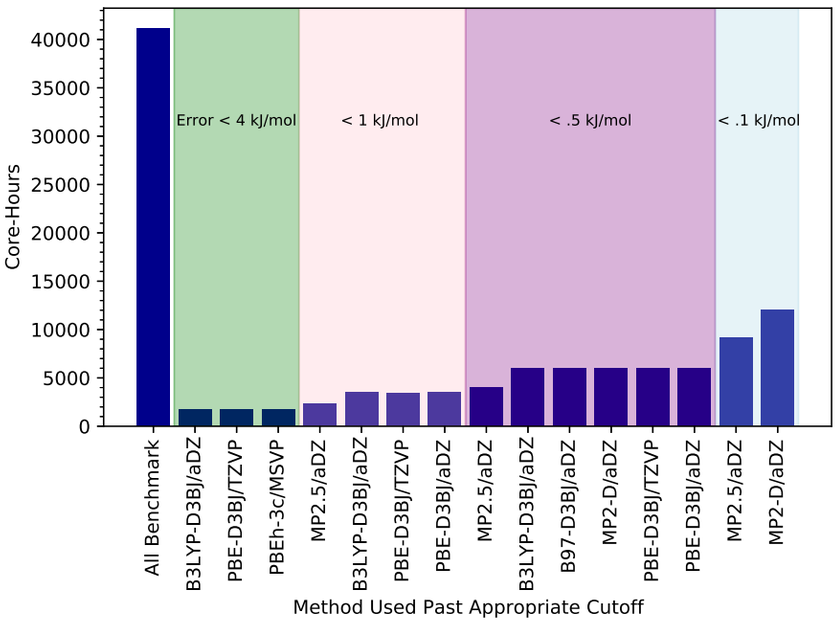


FIG. 8. Total computational effort (core-hours) for calculating the two-body crystal lattice energy contribution with CCSD(T)/CBS(a[TQ]Z; δ:aDZ) ("All Benchmark") and also different two-level approaches that utilize the benchmark level for closer dimers, and various approximate methods for more distant dimers, using the switchover distances provided in Table II. The different sectors of the graph indicate different error regimes for the hybrid methods. Timings were calculated for an adamantane dimer with a minimum inter-monomer separation ~ 7 Å.

mol^{-1} *below* the error from calculating the two-body lattice energy with a two-level method of switching over to MP2.5/aDZ at 4 Å.

We tested additional three-level schemes and found that for all 23 crystals, the PCLE errors were within 1 kJ mol^{-1} of the CCSD(T)/CBS PCLE when using the following method/switchover distance combinations: CCSD(T)/CBS for $R < 4$, MP2.5/aDZ for $4 \leq R < 5$, and either B3LYP-D3BJ/aDZ, PBE-D3BJ/aDZ, or PBE-D3BJ/TZVP for $R \geq 5$ Å. Specific errors for each crystal under these three recommended schemes are shown in Figure 9 as well as MAEs. In addition, we have shown these errors beside those resulting from using the two-level level scheme of CCSD(T)/CBS for $R < 4$ and MP2.5/aDZ for $R \geq 4$ Å. Using B3LYP or PBE for $R \geq 4$ Å actually reduces the maximum and mean absolute errors relative to using MP2.5/aDZ for all dimers with separations larger than 4 Å.

Finally, timings for all proposed two and three-level schemes resulting in PCLE errors

below 1 kJ mol^{-1} for each crystal are shown in Figure 10. The switchover distances for the two-level schemes are those used in Figure 6, and the switchover distances for the three-level schemes have been described above. Using a three-level scheme can result in a 1.9x speed-up relative to the two-level DFT schemes. To achieve the errors of 1 kJ mol^{-1} with a two-level scheme, the switchover distance is 5 Å, so all dimers with separations less than 5 must be calculated with CCSD(T)/CBS. With the three-level scheme, those dimers with separations between 4 and 5 can be calculated with MP2.5/aDZ instead - a method much cheaper than CCSD(T)/CBS - resulting in this speed-up.

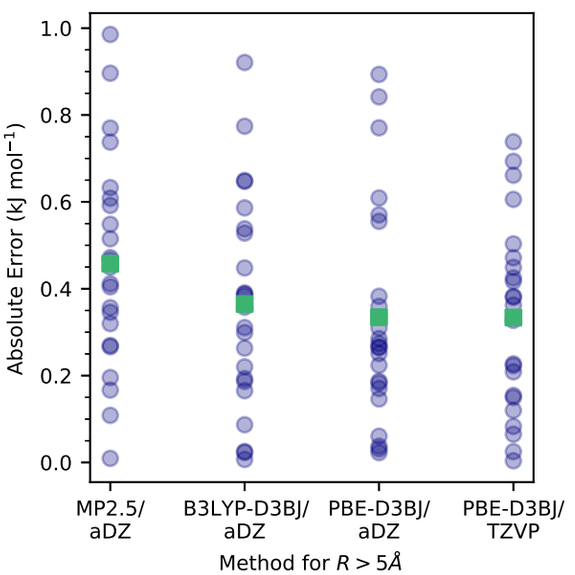


FIG. 9. Absolute errors (kJ mol^{-1}) in the two-body contributions to the crystal lattice energies for all 23 crystals using a three-level scheme. Dimers with $R < 4$ Å are calculated with CCSD(T)/CBS, $4 \leq R < 5$ Å with MP2.5/aDZ, and $R \geq 5$ Å with each method listed. Green squares represent the MAEs. The benchmark method is CCSD(T)/CBS(a[TQ]Z; δ:aDZ).

IV. CONCLUSIONS

For 23 molecular crystals, we have presented benchmark-level values of the two-body contribution to lattice energies using focal-point estimates of coupled-cluster theory through perturbative triple excitations, at the complete-basis-set [CCSD(T)/CBS] limit. We have employed a many-body expansion (MBE) approach that obtains the crystal lattice energy

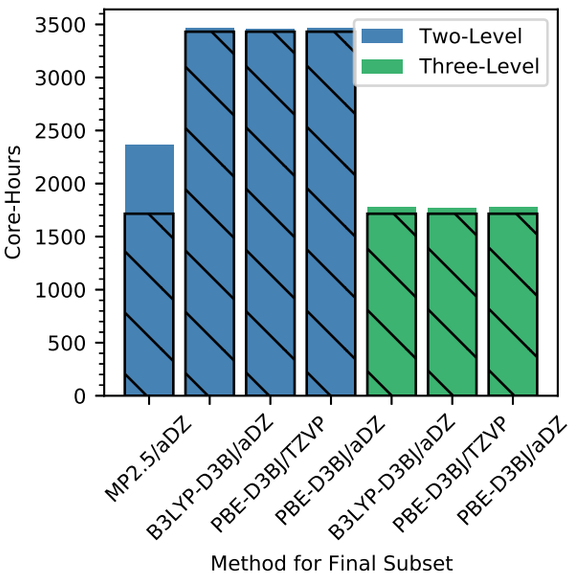


FIG. 10. Total computational effort (core-hours) for calculating the two-body contribution to adamantane's CLE with different schemes which result in errors below 1 kJ mol^{-1} for all 23 crystals relative to CCSD(T)/CBS. Hatching represents the portion of core-hours required for the CCSD(T)/CBS computations of each scheme. The switchover distances are those presented in Figures 6 and 9.

as a sum of individual molecular dimer, trimer, etc., computations, and in this work we have examined the leading two-body (dimers) term. The two-body contributions included dimers with minimum inter-monomer separations less than 30 Å, and the energy contribution per dimer decreases rapidly as the inter-monomer separation increases. Based on this benchmark data, we analyzed how approximate methods perform for the two-body crystal lattice energy contribution. The great majority of the error from using approximate methods rather than CCSD(T)/CBS comes from the short-range dimers with close inter-monomer separations, and therefore we analyzed two-level schemes where dimers with the closest separations were calculated with CCSD(T)/CBS and all other dimers were calculated with an approximate method. We examined errors for such approaches as a function of the switchover distance between CCSD(T)/CBS and the low-level methods and specifically determined schemes which significantly reduce computational time while approximating the CCSD(T)/CBS values within 4, 1, 0.5, and 0.1 kJ mol^{-1} for each crystal.

While DFT methods B3LYP-D3BJ, PBE-D3BJ, and PBEh-3c return errors less than 4 kJ mol^{-1} for dimers with separations between 3 and 20 Å, lower errors can be obtained. Using MP2.5/aDZ for the long-range dimers, and computing shorter-range dimers with CCSD(T)/CBS, can return errors between 0.1 and 1 kJ mol^{-1} depending on the switchover distance between the two methods. Methods less computationally expensive than MP2.5, such as MP2-D/aDZ and various DFT methods, can sometimes also achieve very low errors, but the number of dimers computed with CCSD(T)/CBS must be increased to achieve target accuracies. Due to the success of MP2.5/aDZ computing medium to long-range interaction energies, three-level schemes were considered which compute short-range dimers with CCSD(T), medium-range dimers with MP2.5, and long-range dimers with either B3LYP or PBE. These three-level schemes resulted in a 1.9x speed-up relative to the DFT two-level schemes while still achieving 1 kJ mol^{-1} error in the two-body contributions to each of the 23 crystal lattice energies.

Although we have focused on the two-body terms in the MBE approach, the general accuracy of the DFT and MP2 methods should generally carry over to the context of periodic boundary condition compuations, with the caveat that periodic DFT computations are typically performed in a plane-wave basis with pseudopotentials rather than with atom-centered Gaussian basis functions, and so some differences are to be expected on that basis. Of course, the periodic computations will naturally include three- and higher-body contributions, which have not been included here, and the errors from those terms may add to or partially cancel the errors from the two-body terms.The two-body interactions typically constitute ∼80-90% of the total crystal lattice energy,[34] so an accurate CLE under the many-body expansion must consider three-body, and sometimes even four-body, terms. While this study does not calculate any three or four-body terms, we expect that within the MBE we could use approximate methods for the medium to long-range trimers and tetramers, as evidenced in our recent benzene study.[28]

Additionally, this study does not examine polymporphs based on different molecular conformations (conformational polymorphs). Some studies suggest that intramolecular conformational energies are poorly calculated by generalized gradient approximation (GGA) and hybrid functionals, and that obtaining these energies with wavefunction methods, like MP2 and MP2-D, can greatly improve the results, but basis sets larger than double-ζ may be required.[8,76,77]

SUPPLEMENTAL MATERIAL

The supplemental material contains the crystalline infographic files for each crystal, as well as a report of all dimers, inter-monomer separations, replica numbers, and interaction energies for each method. Figures, corresponding to Figure 2, for all crystals are provided, in addition to absolute and mean absolute errors for different error minimizing schemes. CrystaLattE can be accessed at https://github.com/carlosborca/CrystaLattE.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge financial support from the U.S. National Science Foundation through grant CHE-1955940.

DATA AVAILABILITY

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

REFERENCES

[1]G. J. O. Beran, "Modeling Polymorphic Molecular Crystals with Electronic Structure Theory," Chem. Rev. **116**, 5567–5613 (2016).

[2]G. J. O. Beran, I. J. Sugden, C. Greenwell, D. H. Bowskill, C. C. Pantelides, and C. S. Adjiman, "How many more polymorphs of ROY remain undiscovered †," Chem. Sci. **13**, 1288–1297 (2022).

[3]L. Yu, "Polymorphism in Molecular Solids: An Extraordinary System of Red, Orange, and Yellow Crystals," Acc. Chem. Res. **43**, 1257–1266 (2010).

[4]R. Censi and P. Di Martino, "Polymorph Impact on the Bioavailability and Stability of Poorly Soluble Drugs," Molecules **20**, 18759–18776 (2015).

[5]S. R. Chemburkar, J. Bauer, K. Deming, H. Spiwek, K. Patel, J. Morris, R. Henry, S. Spanton, W. Dziki, W. Porter, J. Quick, P. Bauer, J. Donaubauer, B. A. Narayanan, M. Soldani, D. Riley, and K. McFarland, "Dealing with the Impact of Ritonavir Polymorphs on

the Late Stages of Bulk Drug Process Development," Org. Process Res. Dev. **4**, 413–417 (2000).

[6]J. Bauer, S. Spanton, R. Henry, J. Quick, W. Dziki, W. Porter, and J. Morris, "Ritonavir: An extraordinary example of conformational polymorphism," Pharm. Res. **18**, 859–866 (2001).

[7]A. M. Reilly, R. I. Cooper, C. S. Adjiman, S. Bhattacharya, A. D. Boese, J. G. Brandenburg, P. J. Bygrave, R. Bylsma, J. E. Campbell, R. Car, D. H. Case, R. Chadha, J. C. Cole, K. Cosburn, H. M. Cuppen, F. Curtis, G. M. Day, R. A. DiStasio Jr, A. Dzyabchenko, B. P. van Eijck, D. M. Elking, J. A. van den Ende, J. C. Facelli, M. B. Ferraro, L. Fusti-Molnar, C.-A. Gatsiou, T. S. Gee, R. de Gelder, L. M. Ghiringhelli, H. Goto, S. Grimme, R. Guo, D. W. M. Hofmann, J. Hoja, R. K. Hylton, L. Iuzzolino, W. Jankiewicz, D. T. de Jong, J. Kendrick, N. J. J. de Klerk, H.-Y. Ko, L. N. Kuleshova, X. Li, S. Lohani, F. J. J. Leusen, A. M. Lund, J. Lv, Y. Ma, N. Marom, A. E. Masunov, P. McCabe, D. P. McMahon, H. Meekes, M. P. Metz, A. J. Misquitta, S. Mohamed, B. Monserrat, R. J. Needs, M. A. Neumann, J. Nyman, S. Obata, H. Oberhofer, A. R. Oganov, A. M. Orendt, G. I. Pagola, C. C. Pantelides, C. J. Pickard, R. Podeszwa, L. S. Price, S. L. Price, A. Pulido, M. G. Read, K. Reuter, E. Schneider, C. Schober, G. P. Shields, P. Singh, I. J. Sugden, K. Szalewicz, C. R. Taylor, A. Tkatchenko, M. E. Tuckerman, F. Vacarro, M. Vasileiadis, A. Vazquez-Mayagoitia, L. Vogt, Y. Wang, R. E. Watson, G. A. de Wijs, J. Yang, Q. Zhu, and C. R. Groom, "Report on the sixth blind test of organic crystal structure prediction methods," Acta Crystallogr. B **72**, 439–459 (2016).

[8]C. Greenwell, J. L. McKinley, P. Zhang, Q. Zeng, G. Sun, B. Li, S. Wen, and G. J. O. Beran, "Overcoming the difficulties of predicting conformational polymorph energetics in molecular crystals via correlated wavefunction methods," Chem. Sci. **11**, 2200–2214 (2020).

[9]J. Nyman and G. M. Day, "Static and lattice vibrational energy differences between polymorphs," CrystEngComm **17**, 5154–5165 (2015).

[10]M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, "Fragmentation Methods: A Route to Accurate Calculations on Large Systems," Chem. Rev. **112**, 632–672 (2011).

[11]U. Góra, R. Podeszwa, W. Cencek, and K. Szalewicz, "Interaction energies of large clusters from many-body expansion," J. Chem. Phys. **135**, 224102 (2011).

[12]R. M. Richard and J. M. Herbert, "A generalized many-body expansion and a unified view of fragment-based methods in electronic structure theory," J. Chem. Phys. **137**, 064113

(2012).

[13]J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schütz, and G. K. L. Chan, "Ab initio determination of the crystalline benzene lattice energy to sub-kilojoule/mole accuracy," Science **345**, 640–643 (2014).

[14]A. L. Ringer and C. D. Sherrill, "First principles computation of lattice energies of organic solids: The benzene crystal," Chem. Eur. J. **14**, 2542–2547 (2008).

[15]S. Hirata, "Fast electron-correlation methods for molecular crystals: An application to the α, $\beta 1$, and $\beta 2$ modifications of solid formic acid," J. Chem. Phys. **129**, 204104 (2008).

[16]S. Hirata, K. Gilliard, X. He, J. Li, and O. Sode, "Ab Initio Molecular Crystal Structures, Spectra, and Phase Diagrams," Acc. Chem. Res. **47**, 2721–2730 (2014).

[17]P. J. Bygrave, N. L. Allan, and F. R. Manby, "The embedded many-body expansion for energetics of molecular crystals," J. Chem. Phys. **137**, 164102 (2012).

[18]M. J. Gillan, D. Alfè, P. J. Bygrave, C. R. Taylor, and F. R. Manby, "Energy benchmarks for water clusters and ice structures from an embedded many-body expansion," J. Chem. Phys. **139**, 114101 (2013).

[19]O. V. Shishkin, R. I. Zubatyuk, A. V. Maleev, and R. Boese, "Investigation of topology of intermolecular interactions in the benzene-acetylene co-crystal by different theoretical methods," Struct. Chem. **25**, 1547–1552 (2014).

[20]S. Mattsson, B. Paulus, F. A. Redeker, H. Beckers, S. Riedel, and C. Müller, "The Crystal Structure of α-F_2 : Solving a 50 Year Old Puzzle Computationally," Chem. Eur. J. **25**, 3318–3324 (2019).

[21]S. Grimme, "Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction," J. Comput. Chem. **27**, 1787–1799 (2006).

[22]S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," J. Chem. Phys. **132**, 154104 (2010).

[23]S. Grimme, S. Ehrlich, and L. Goerigk, "Effect of the damping function in dispersion corrected density functional theory," J. Comput. Chem. **32**, 1456–1465 (2011).

[24]S. Tsuzuki, H. Orita, K. Honda, and M. Mikami, "First-principles lattice energy calculation of urea and hexamine crystals by a combination of periodic DFT and MP2 two-body interaction energy calculations," J. Phys. Chem. B **114**, 6799–6805 (2010).

[25]K. D. Nanda and G. J. Beran, "Prediction of organic molecular crystal geometries from MP2-level fragment quantum mechanical/molecular mechanical calculations," J. Chem. Phys. **137** (2012), 10.1063/1.4764063.

[26]C. Møller and M. S. Plesset, "Note on an Approximation Treatment for Many-Electron Systems," Phys. Rev. **46**, 618–622 (1934).

[27]K. 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).

[28]C. H. Borca, Z. L. Glick, D. P. Metcalf, L. A. Burns, and C. D. Sherrill, "Benchmark lattice energy of crystalline benzene in CrystaLattE Benchmark-level lattice energy of crystalline benzene via multi-scale methods in CrystaLattE," J. Chem. Phys., submitted (2022).

[29]M. R. Kennedy, A. R. McDonald, A. E. DePrince, M. S. Marshall, R. Podeszwa, and C. D. Sherrill, "Communication: Resolving the three-body contribution to the lattice energy of crystalline benzene: Benchmark results from coupled-cluster theory," J. Chem. Phys. **140**, 121104 (2014).

[30]C. 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-H_2S," J. Phys. Chem. A **113**, 10146–10159 (2009).

[31]A. M. Reilly and A. Tkatchenko, "Understanding the role of vibrations, exact exchange, and many-body van der Waals interactions in the cohesive properties of molecular crystals," J. Chem. Phys. **139**, 024705 (2013).

[32]A. Otero-De-La-Roza and E. R. Johnson, "A benchmark for non-covalent interactions in solids," J. Chem. Phys. **137** (2012), 10.1063/1.4738961.

[33]G. A. Dolgonos, J. Hoja, and A. D. Boese, "Revised values for the X23 benchmark set of molecular crystals †," Phys. Chem. Chem. Phys. **21**, 24333 (2019).

[34]S. Wen and G. J. O. Beran, "Accurate Molecular Crystal Lattice Energies from a Fragment QM/MM Approach with On-the-Fly Ab Initio Force Field Parametrization," J. Chem. Theory Comput. **7**, 3733–3742 (2011).

[35]C. H. Borca, B. W. Bakr, L. A. Burns, and C. D. Sherrill, "CrystaLattE: Automated computation of lattice energies of organic crystals exploiting the many-body expansion to achieve dual-level parallelism," J. Chem. Phys. **151**, 144103 (2019).

[36]A. L. L. East and W. D. Allen, "The heat of formation of NCO," J. Chem. Phys. **99**, 4638–4650 (1993).

[37]A. 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).

[38]M. Omar Sinnokrot, E. F. Valeev, and C. D. Sherrill, "Estimates of the Ab Initio Limit for $\pi\pi$ Interactions: The Benzene Dimer," J. Am. Chem. Soc. **124**, 10887–10893 (2002).

[39]M. S. Marshall, L. A. Burns, and C. D. Sherrill, "Basis set convergence of the coupled-cluster correction, δMP2CCSD(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).

[40]L. 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 (2013).

[41]A. 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).

[42]T. 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).

[43]D. 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).

[44]A. E. DePrince and C. D. Sherrill, "Accuracy and efficiency of coupled-cluster theory using density fitting/Cholesky decomposition, frozen natural orbitals, and a T1-transformed Hamiltonian," J. Chem. Theory Comput. **9**, 2687–2696 (2013).

[45]S. 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).

[46]D. Feller, "Application of systematic sequences of wave functions to the water dimer," J. Chem. Phys. **96**, 6104–6114 (1992).

[47]P. 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).

[48]A. D. Becke, "Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals," J. Chem. Phys. **107**, 8554–8560 (1997).

[49]R. Sure and S. Grimme, "Corrected Small Basis Set Hartree-Fock Method for Large Systems," J. Comput. Chem. **34**, 1672–1685 (2013).

[50]M. Pitoňák, P. Neogrády, J. Černý, S. Grimme, and P. Hobza, "Scaled MP3 Non-Covalent Interaction Energies Agree Closely with Accurate CCSD(T) Benchmark Data," ChemPhysChem **10**, 282–289 (2009).

[51]J. Řezáč, C. Greenwell, and G. J. O. Beran, "Accurate Noncovalent Interactions via Dispersion-Corrected Second-Order Møller–Plesset Perturbation Theory," J. Chem. Theory Comput. **14**, 4711–4721 (2018).

[52]J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple," Phys. Rev. Lett. **77**, 3865–3868 (1996).

[53]J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. 77, 3865 (1996)]," Phys. Rev. Lett. **78**, 1396–1396 (1997).

[54]S. Grimme, J. G. Brandenburg, C. Bannwarth, and A. Hansen, "Consistent structures and interactions by density functional theory with small atomic orbital basis sets," J. Chem. Phys. **143**, 054107 (2015).

[55]K. Szalewicz, "Symmetry-adapted perturbation theory of intermolecular forces," Wiley Interdiscip. Rev. Comput. Mol. Sci. **2**, 254–272 (2012).

[56]K. Patkowski, "Recent developments in symmetry-adapted perturbation theory," Wiley Interdiscip. Rev. Comput. Mol. Sci. **10**, e1452 (2019).

[57]H. Kruse and S. Grimme, "A geometrical correction for the inter- and intra-molecular basis set superposition error in Hartree-Fock and density functional theory calculations for large systems," J. Chem. Phys. **136**, 154101 (2012).

[58]E. Papajak, J. Zheng, X. Xu, H. R. Leverentz, and D. G. Truhlar, "Perspectives on basis sets beautiful: Seasonal plantings of diffuse basis functions," J. Chem. Theory Comput. **7**, 3027–3034 (2011).

[59]T. 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).

[60]J. G. Brandenburg, M. Hochheim, T. Bredow, and S. Grimme, "Low-cost quantum chemical methods for noncovalent interactions," J. Phys. Chem. Lett. **5**, 4275–4284 (2014).

[61]M. Cutini, B. Civalleri, M. Corno, R. Orlando, J. G. Brandenburg, L. Maschio, and P. Ugliengo, "Assessment of different quantum mechanical methods for the prediction of structure and cohesive energy of molecular crystals," J. Chem. Theory Comput. **12**, 3340–3352 (2016).

[62]F. Weigend and R. Ahlrichs, "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy," Phys. Chem. Chem. Phys. **7**, 3297 (2005).

[63]D. 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).

[64]F. Neese, "Software update: The ORCA program system—version 5.0," Wiley Interdiscip. Rev. Comput. Mol. Sci. **12**, e1606 (2022).

[65]S. Lehtola, C. Steigemann, M. J. Oliveira, and M. A. Marques, "Recent developments in libxc — a comprehensive library of functionals for density functional theory," SoftwareX **7**, 1–5 (2018).

[66]D. P. Metcalf, A. Smith, Z. L. Glick, and C. D. Sherrill, "Range-dependence of two-body intermolecular interactions and their energy components in molecular crystals," J. Chem. Phys. **157**, 084503 (2022).

[67]M. O. Sinnokrot and C. D. Sherrill, "Highly accurate coupled cluster potential energy curves for the benzene dimer: Sandwich, t-shaped, and parallel-displaced configurations," J. Phys. Chem. A **108**, 10200–10207 (2004).

[68]P. Jurečka, J. Šponer, and P. Hobza, "Potential energy surface of the cytosine dimer: MP2 complete basis set limit interaction energies, CCSD(T) correction term, and comparison with the amber force field," J. Phys. Chem. B **108**, 5466–5471 (2004).

[69]S. Tsuzuki, T. Uchimaru, M. Mikami, and K. Tanabe, "High level ab initio calculations of intermolecular interaction of propane dimer: Orientation dependence of interaction energy," J. Phys. Chem. A **106**, 3867–3872 (2002).

[70]J. D. Hartman, A. Balaji, and G. J. Beran, "Improved electrostatic embedding for fragment-based chemical shift calculations in molecular crystals," J. Chem. Theory Comput. **13**, 6043–6051 (2017).

[71]J. A. Chisholm and S. Motherwell, "COMPACK: A program for identifying crystal structure similarity using distances," J. Appl. Cryst. **38**, 228–231 (2005).

[72]L. 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).

[73]L. 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).

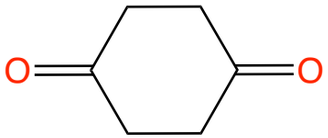
[74]K. E. Riley, J. Řezáč, and P. Hobza, "The performance of MP2.5 and MP2.x methods for nonequilibrium geometries of molecular complexes," Phys. Chem. Chem. Phys. **14**, 13187–13193 (2012).

[75]R. Sedlak, K. E. Riley, J. Řezáč, M. Pitoňák, and P. Hobza, "MP2.5 and MP2.x: Approaching CCSD(T) quality description of noncovalent interaction at the cost of a single ccsd iteration," ChemPhysChem **14**, 698–707 (2013).

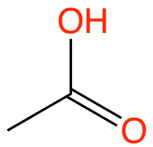
[76]C. Greenwell and G. J. O. Beran, "Inaccurate conformational energies still hinder crystal structure prediction in flexible organic molecules," Crystal Growth & Design **20**, 4875–4881 (2020).

[77]S. P. Thomas and M. A. Spackman, "The polymorphs of ROY: A computational study of lattice energies and conformational energy differences," Aust. J. Chem. **71**, 279–284 (2018).

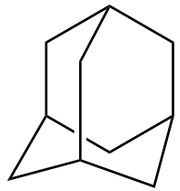
Accepted to J. Chem. Phys. 10.1063/5.0141872



1,4-Cyclohexanedione

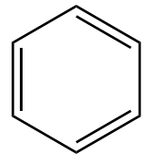


Acetic Acid



Adamantane

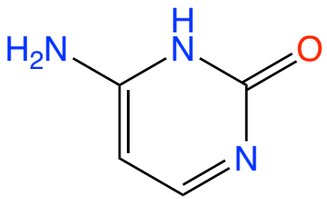
Ammonia



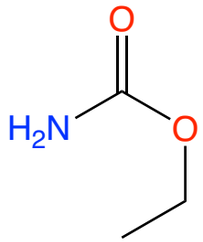
Benzene

Carbon Dioxide

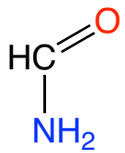
Cyanamide



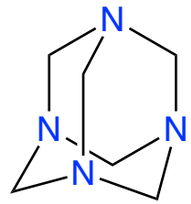
Cytosine



Ethyl Carbamate

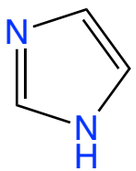


Formamide

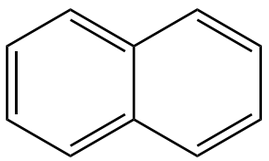


Hexamine

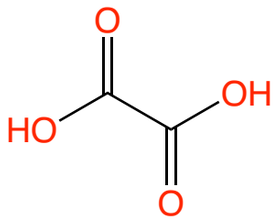
Ice



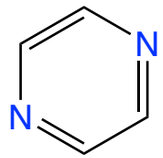
Imidazole



Naphthalene

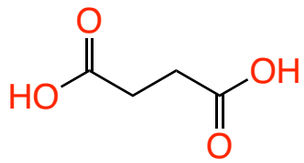


Oxalic Acid



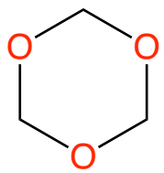
Pyrazine

Pyrazole

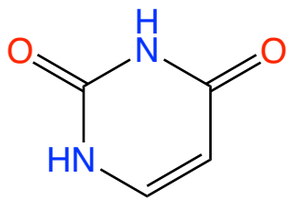


Succinic Acid

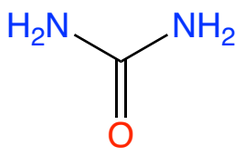
Triazine



Trioxane

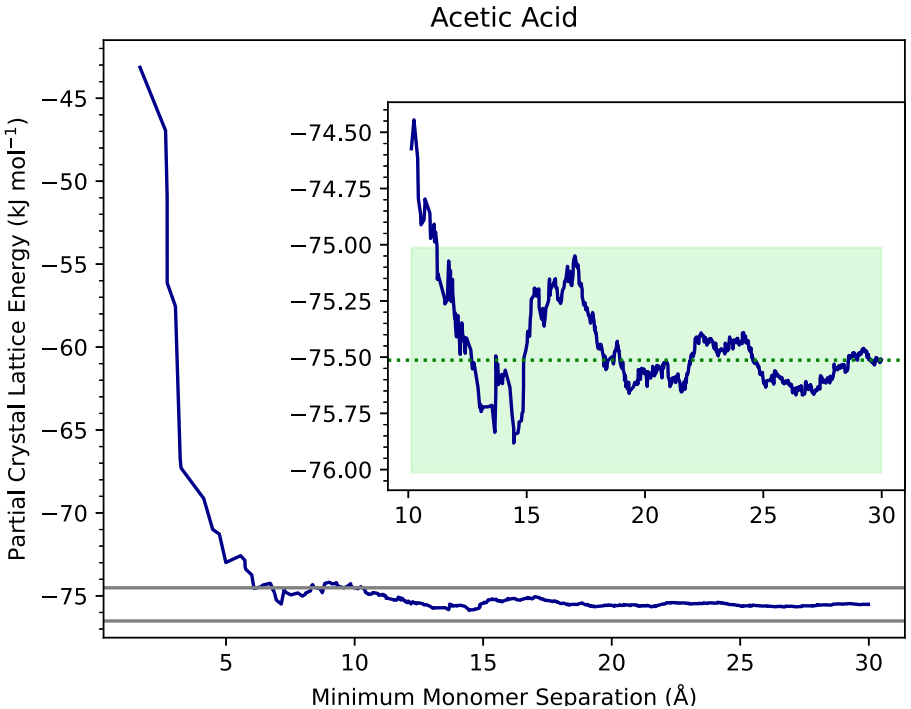


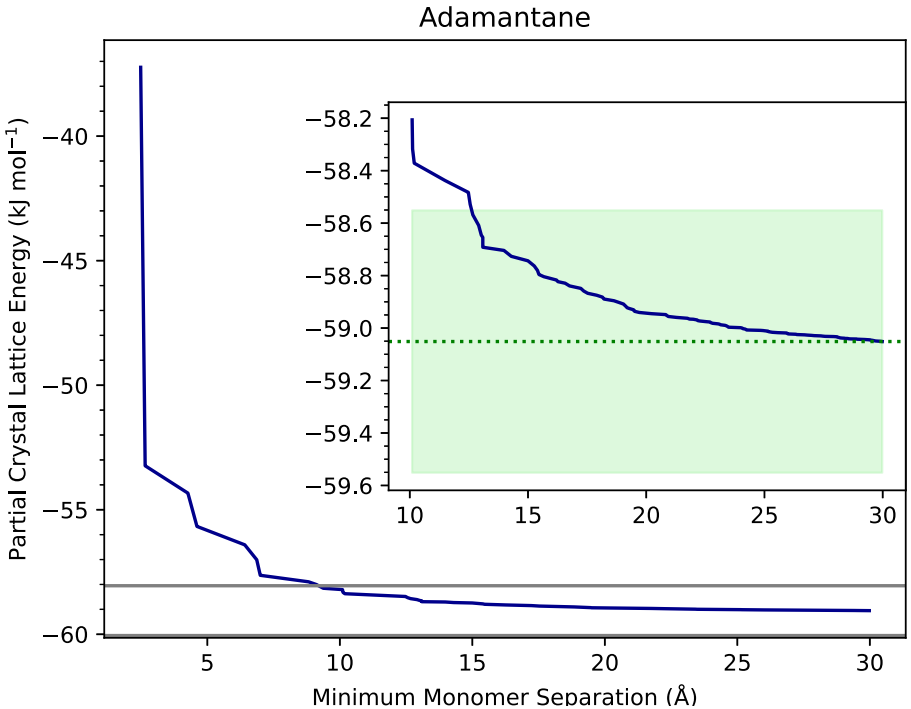
Uracil

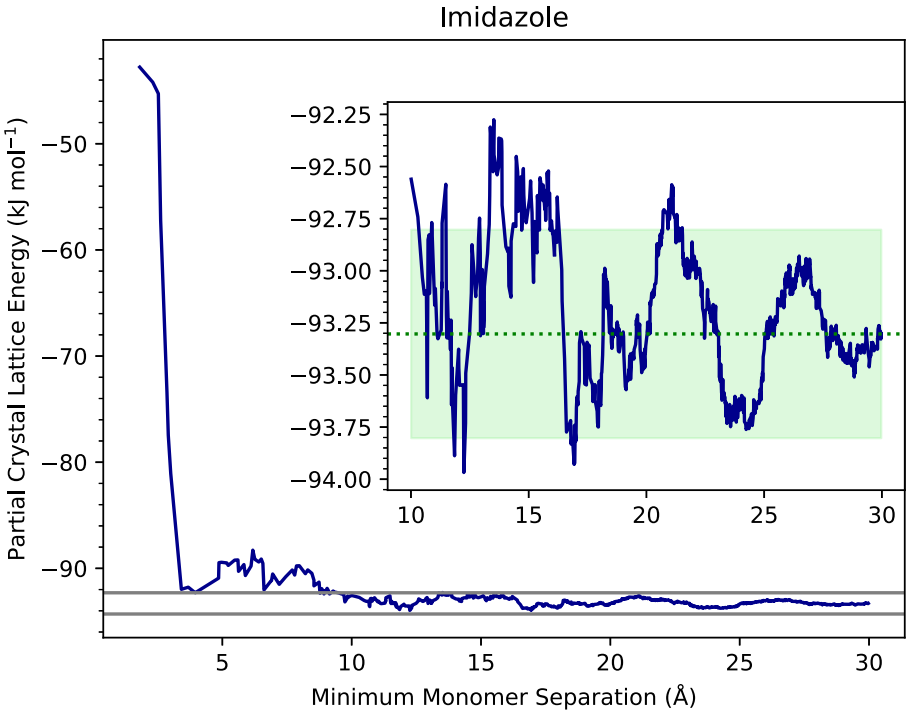


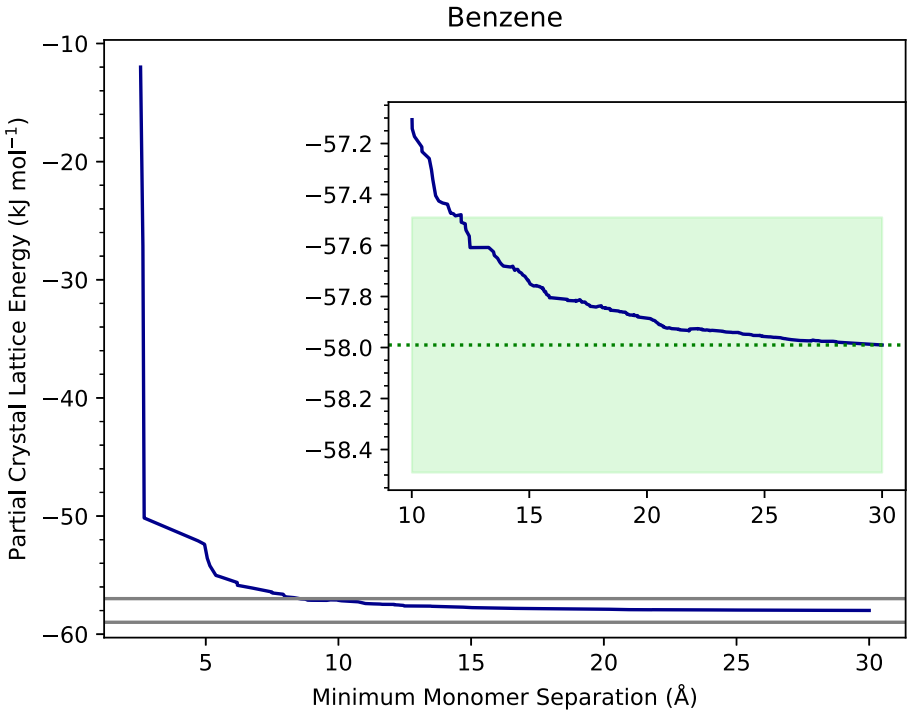
Urea

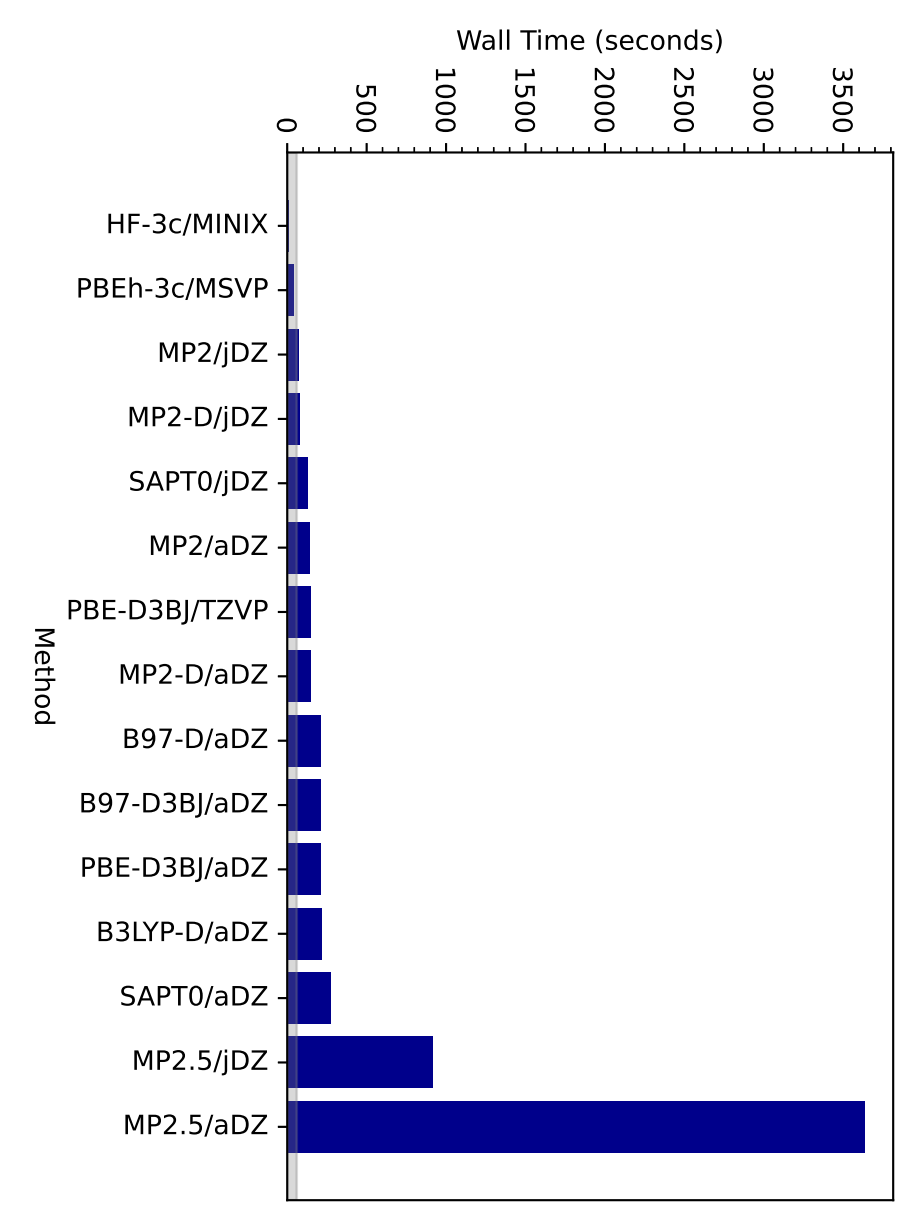
Accepted to J. Chem. Phys. 10.1063/5.0141872

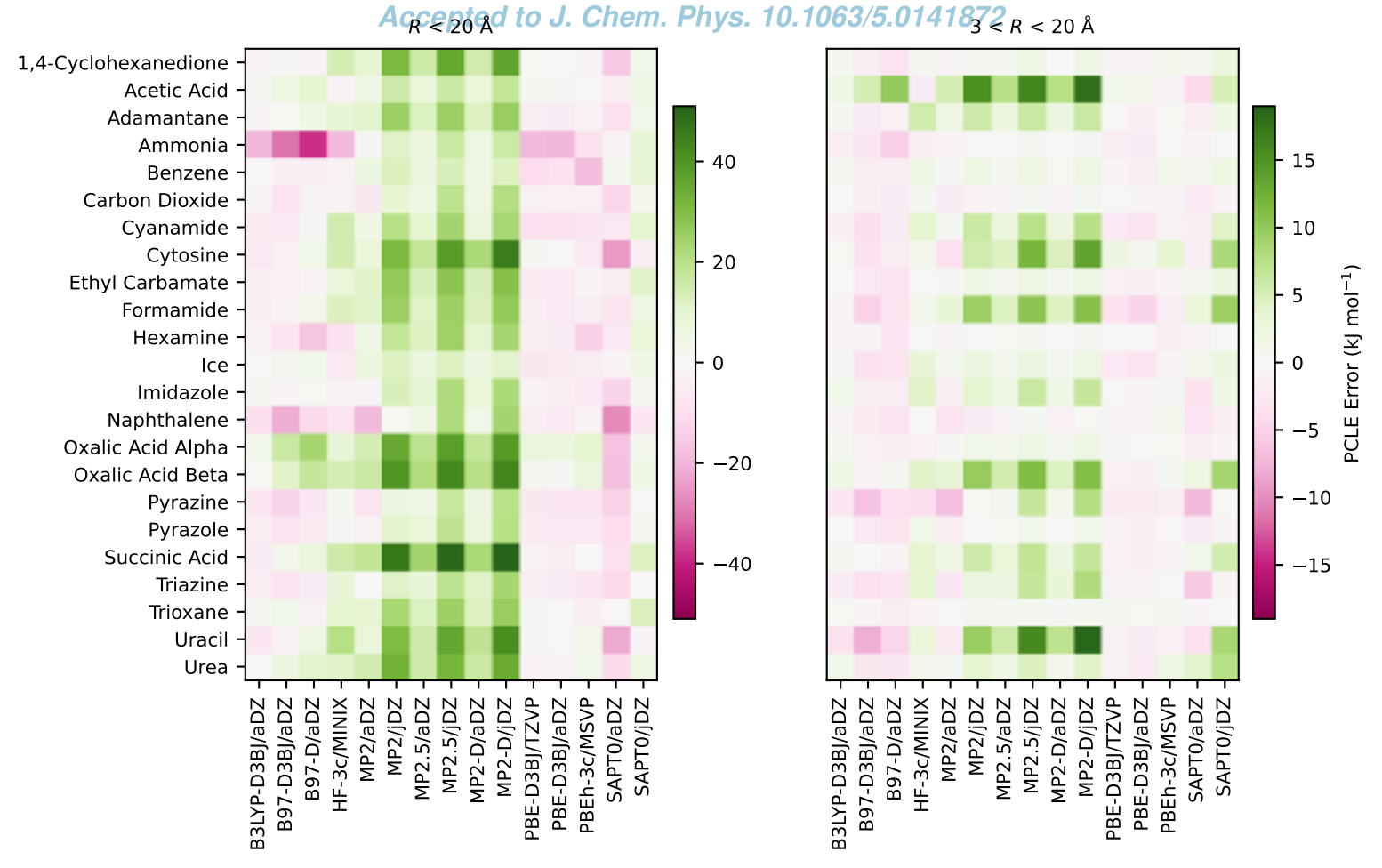


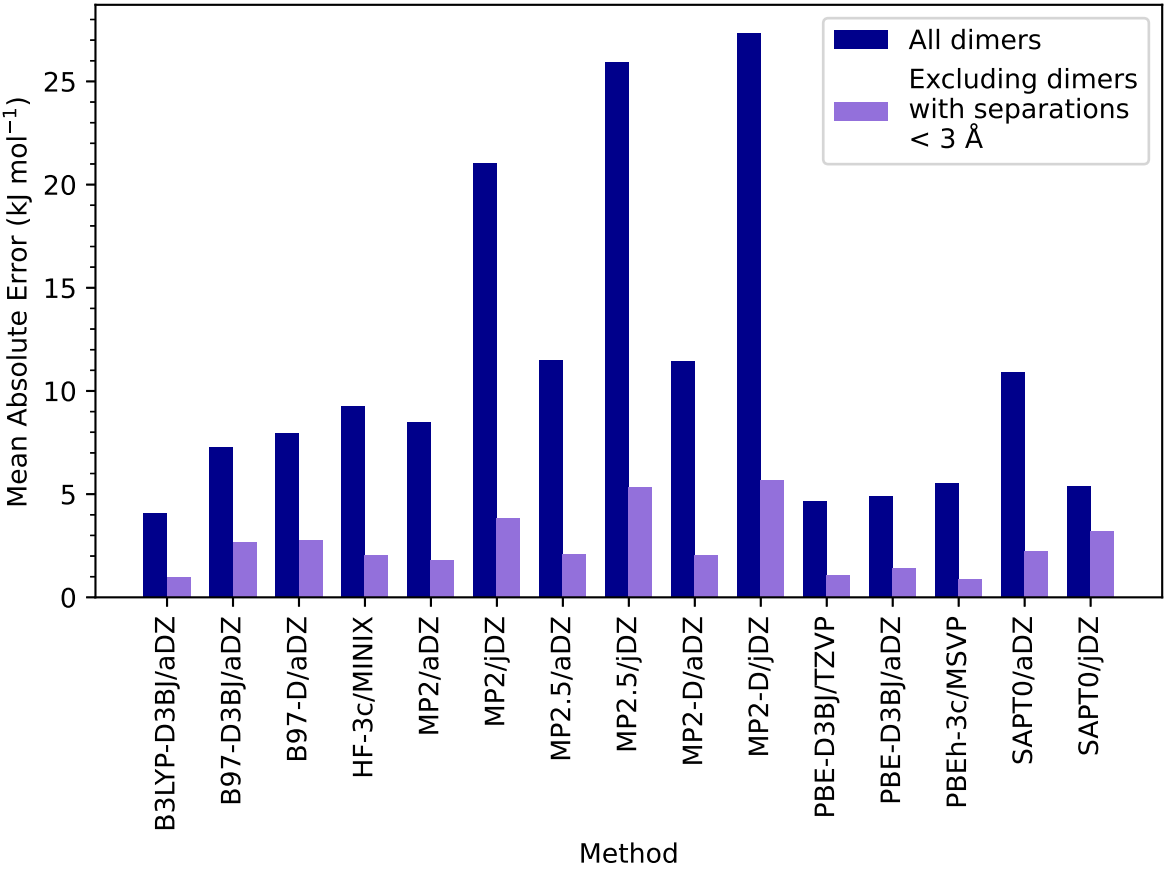


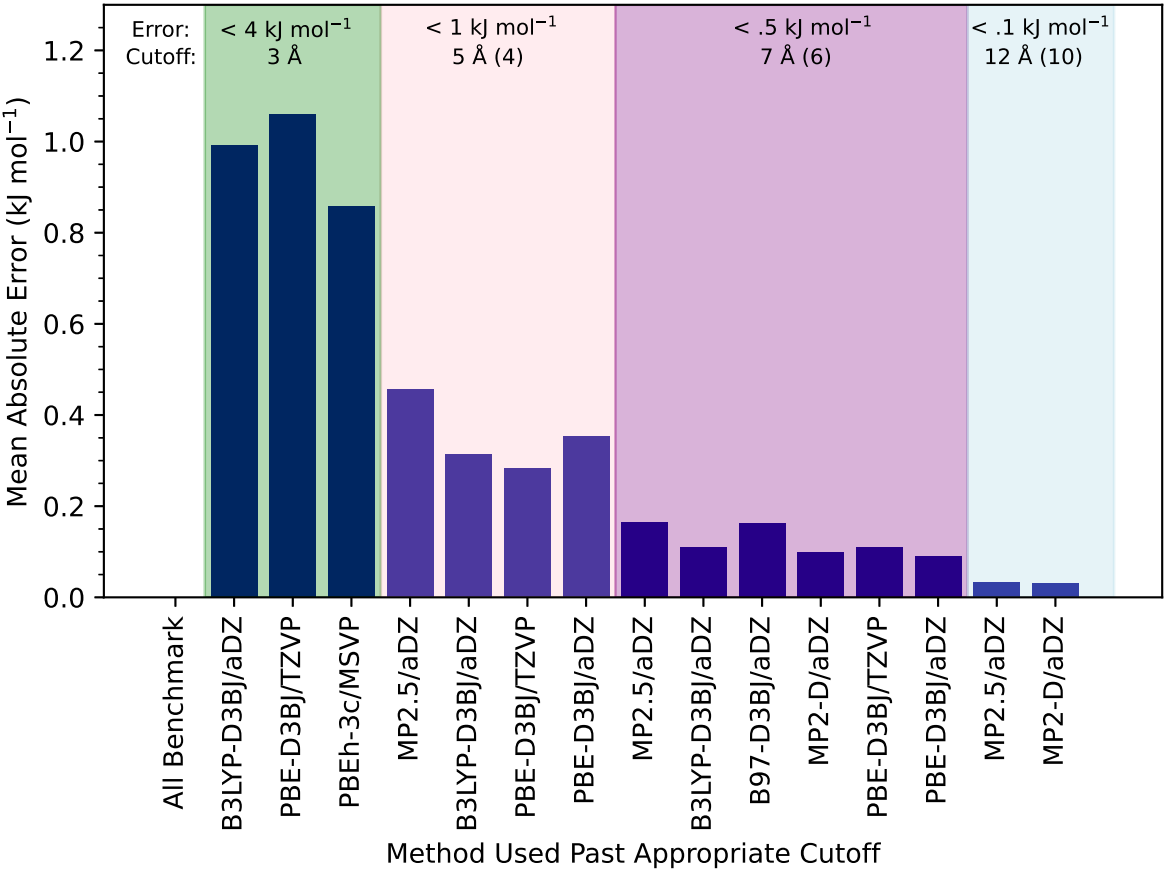


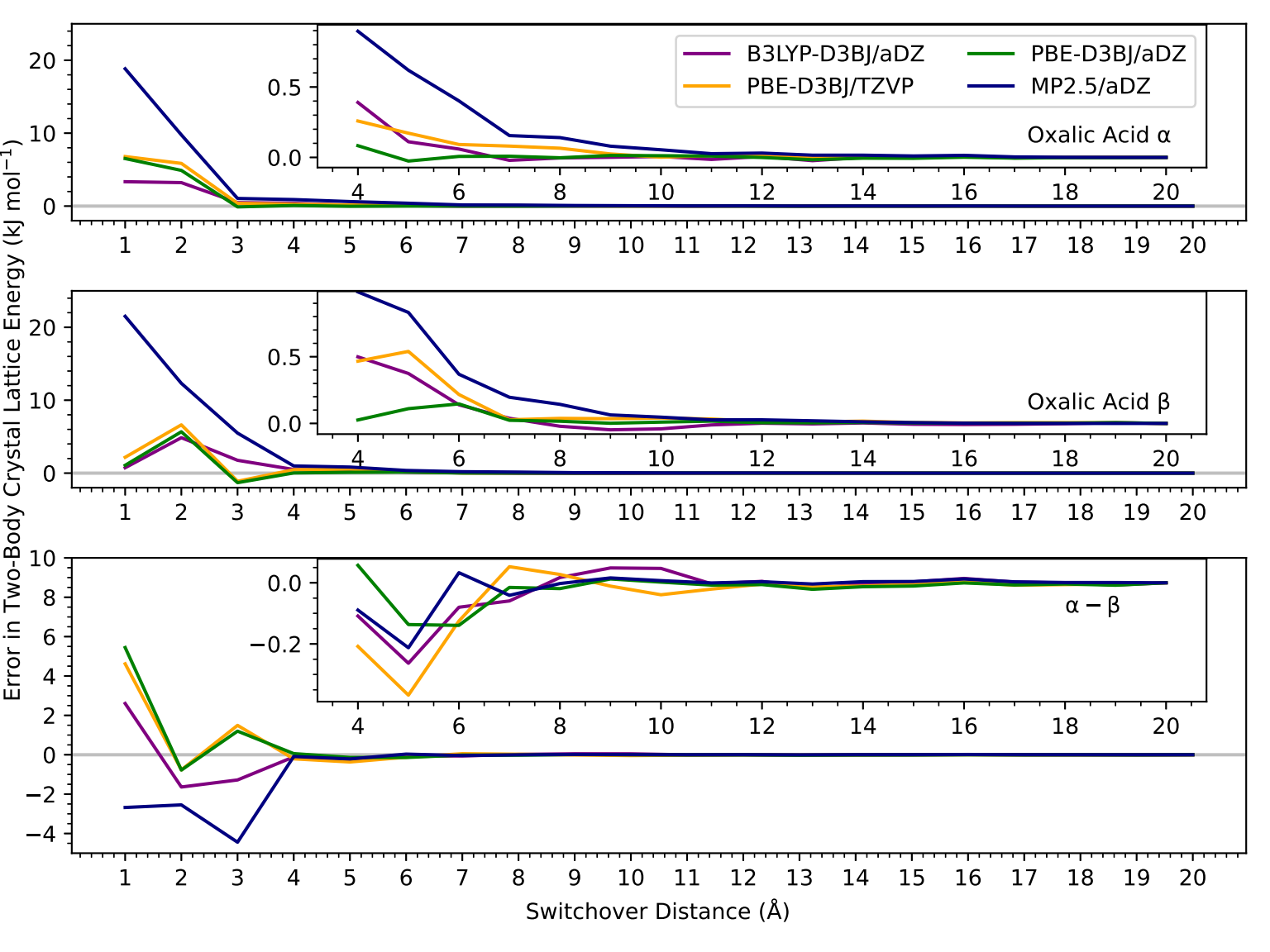


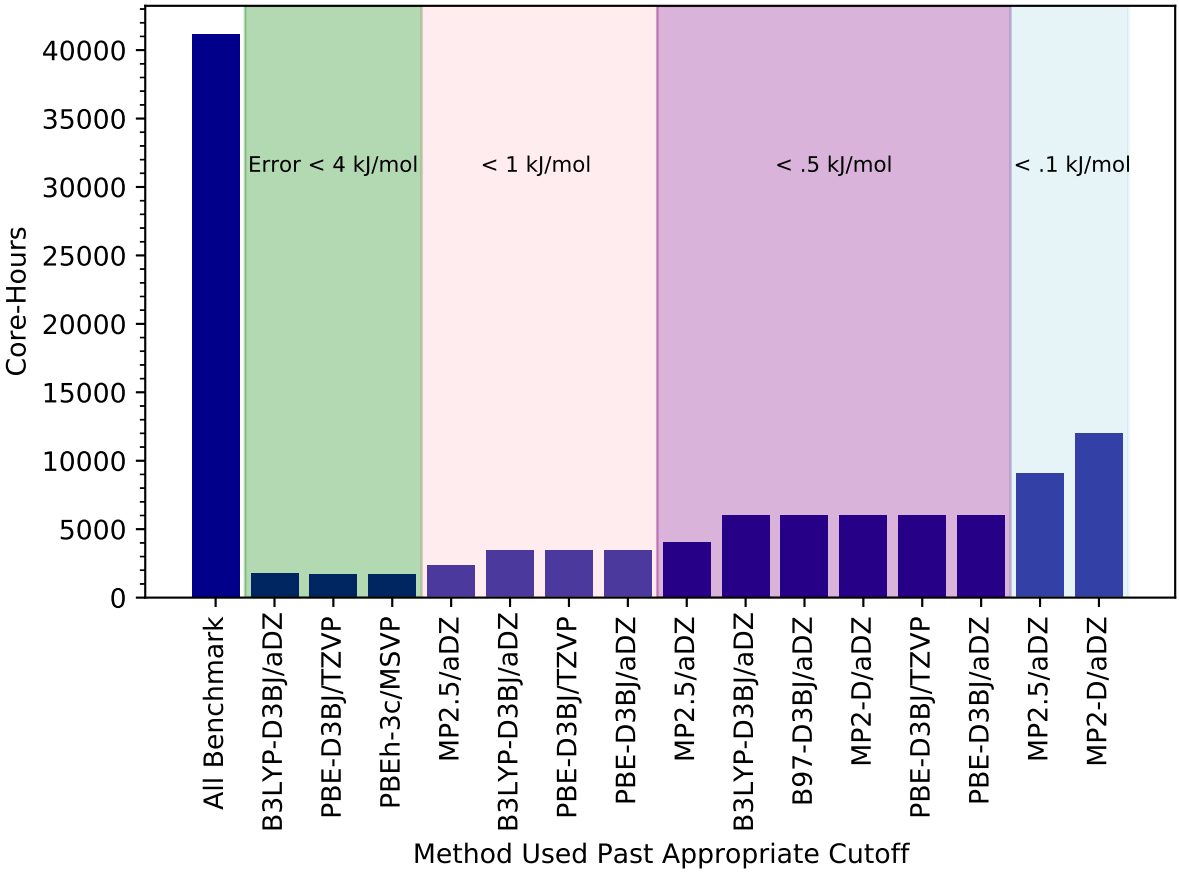

Wall Time (seconds)
0
500
1000
1500
2000
2500
3000
3500
Method
HF-3c/MINIX
PBEh-3c/MSVP
MP2/jDZ
MP2-D/jDZ
SAPT0/jDZ
MP2/aDZ
PBE-D3BJ/TZVP
MP2-D/aDZ
B97-D/aDZ
B97-D3BJ/aDZ
PBE-D3BJ/aDZ
B3LYP-D/aDZ
SAPT0/aDZ
MP2.5/jDZ
MP2.5/aDZ

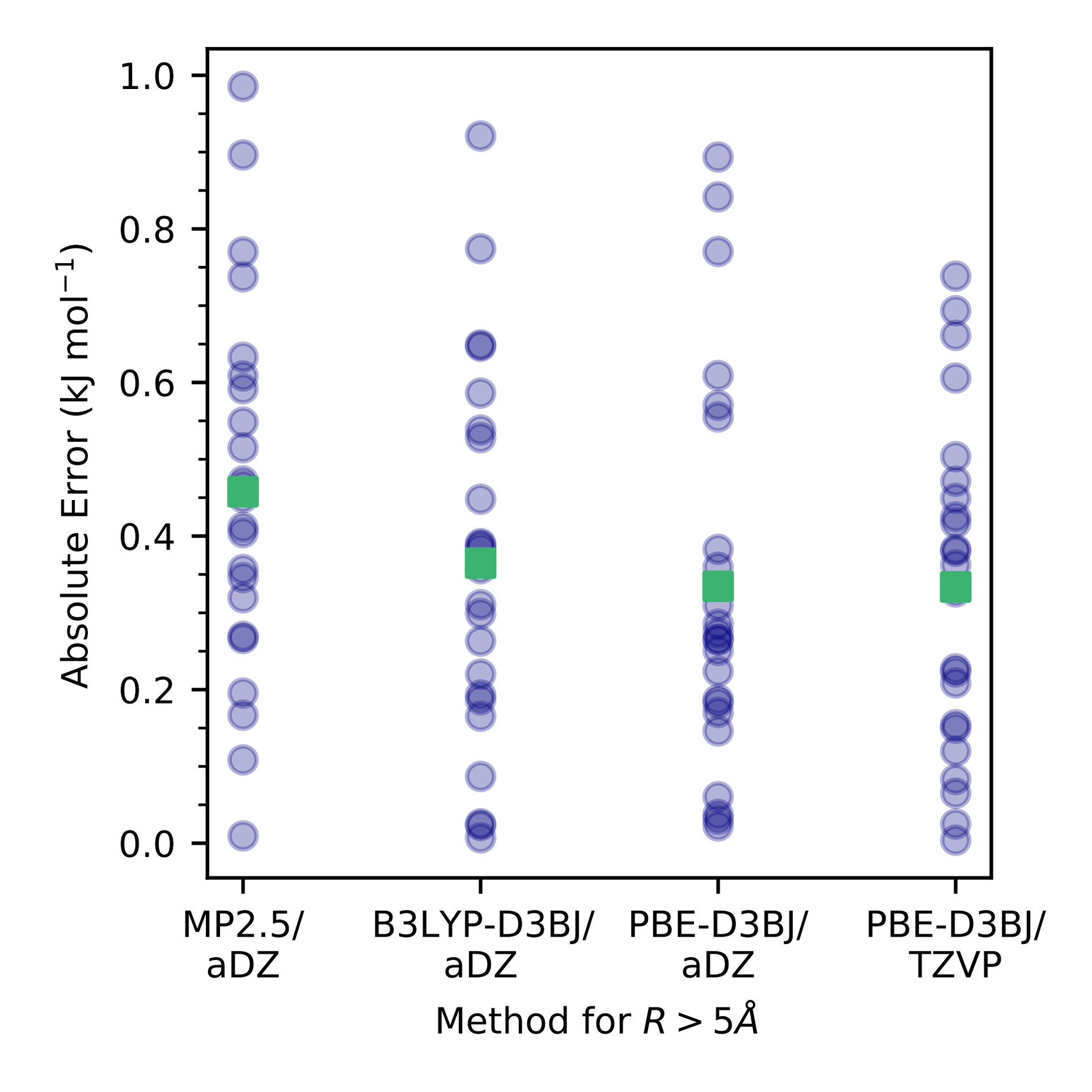


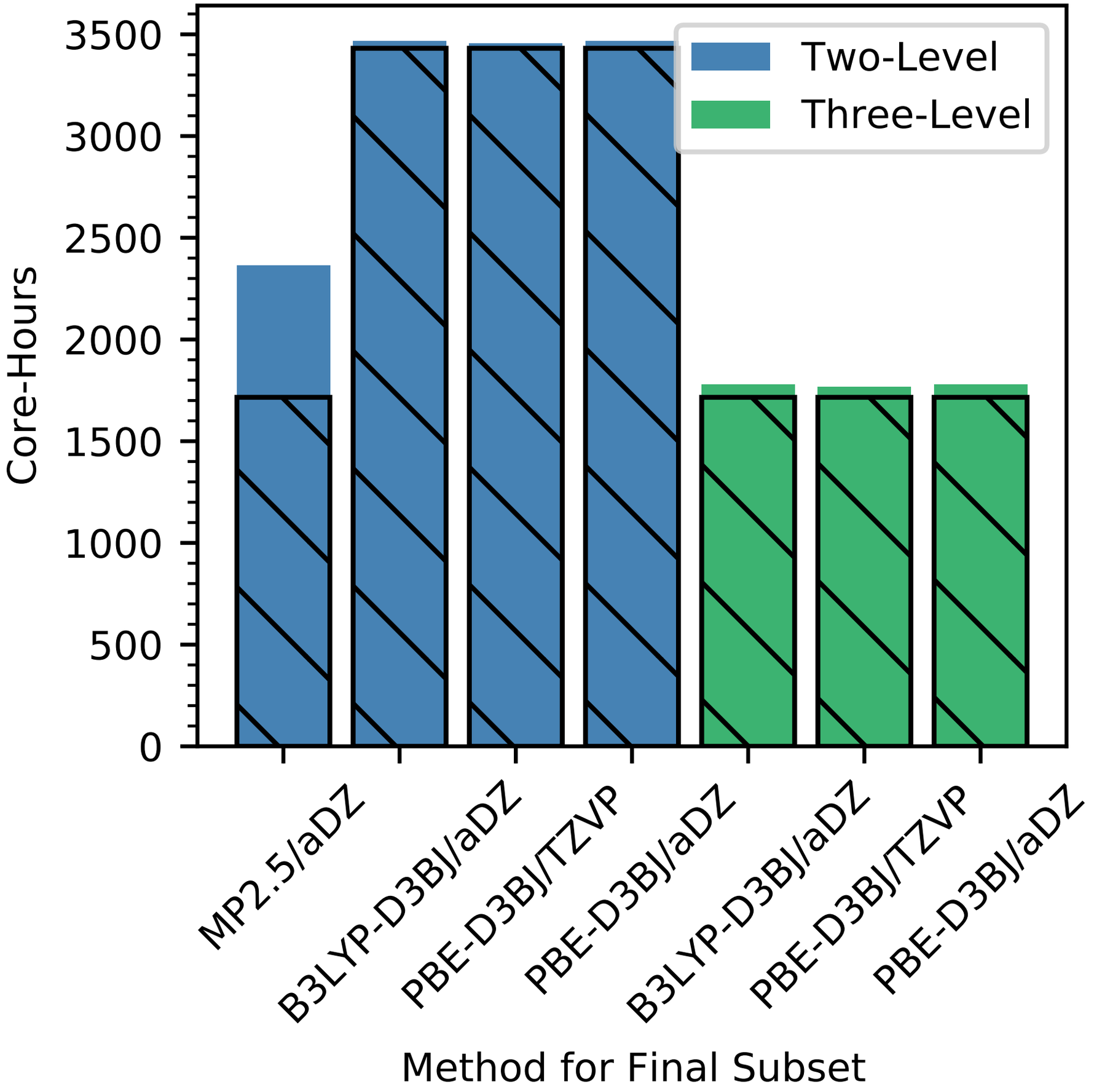

All dimers
Excluding dimers with separations < 3 Å
Mean Absolute Error (kJ mol^{-1})
0
5
10
15
20
25
B3LYP-D3BJ/aDZ
B97-D3BJ/aDZ
B97-D/aDZ
HF-3c/MINIX
MP2/aDZ
MP2/jDZ
MP2.5/aDZ
MP2.5/jDZ
MP2-D/aDZ
MP2-D/jDZ
PBE-D3BJ/TZVP
PBE-D3BJ/aDZ
PBEh-3c/MSVP
SAPT0/aDZ
SAPT0/jDZ
Method


Error:
Cutoff:
< 4 kJ mol^{-1}
3 Å
< 1 kJ mol^{-1}
5 Å (4)
< .5 kJ mol^{-1}
7 Å (6)
< .1 kJ mol^{-1}
12 Å (10)
Mean Absolute Error (kJ mol^{-1})
0.0
0.2
0.4
0.6
0.8
1.0
1.2
All Benchmark
B3LYP-D3BJ/aDZ
PBE-D3BJ/TZVP
PBEh-3c/MSVP
MP2.5/aDZ
B3LYP-D3BJ/aDZ
PBE-D3BJ/TZVP
PBE-D3BJ/aDZ
MP2.5/aDZ
B3LYP-D3BJ/aDZ
B97-D3BJ/aDZ
MP2-D/aDZ
PBE-D3BJ/TZVP
PBE-D3BJ/aDZ
MP2.5/aDZ
MP2-D/aDZ
Method Used Past Appropriate Cutoff


B3LYP-D3BJ/aDZ
PBE-D3BJ/TZVP
PBE-D3BJ/aDZ
MP2.5/aDZ
Oxalic Acid α
Oxalic Acid β
α − β
Error in Two-Body Crystal Lattice Energy (kJ mol^{-1})
Switchover Distance (Å)


Core-Hours
0
5000
10000
15000
20000
25000
30000
35000
40000
Error < 4 kJ/mol
< 1 kJ/mol
< .5 kJ/mol
< .1 kJ/mol
All Benchmark
B3LYP-D3BJ/aDZ
PBE-D3BJ/TZVP
PBEh-3c/MSVP
MP2.5/aDZ
B3LYP-D3BJ/aDZ
PBE-D3BJ/TZVP
PBE-D3BJ/aDZ
MP2.5/aDZ
B3LYP-D3BJ/aDZ
B97-D3BJ/aDZ
MP2-D/aDZ
PBE-D3BJ/TZVP
PBE-D3BJ/aDZ
MP2.5/aDZ
MP2-D/aDZ
Method Used Past Appropriate Cutoff


1.0
0.8
0.6
0.4
0.2
0.0
Absolute Error (kJ mol^{-1})
MP2.5/
aDZ
B3LYP-D3BJ/
aDZ
PBE-D3BJ/
aDZ
PBE-D3BJ/
TZVP
Method for $R > 5\AA$


Two-Level
Three-Level
Core-Hours
0
500
1000
1500
2000
2500
3000
3500
MP2.5/aDZ
B3LYP-D3BJ/aDZ
PBE-D3BJ/TZVP
PBE-D3BJ/aDZ
B3LYP-D3BJ/aDZ
PBE-D3BJ/TZVP
PBE-D3BJ/aDZ
Method for Final Subset