Extending Density-Corrected Density Functional Theory to Large Molecular Systems

Youngsam Kim,^{a,\dagger} Mingyu Sim,^{a,\dagger} Minhyeok Lee,^a Sehun Kim,^a Suhwan Song,^{a,\ddagger} Kieron Burke,^b and Eunji Sim^{a,*}

^aDepartment of Chemistry, Yonsei University, 50 Yonsei-ro Seodaemun-gu, Seoul 03722, Korea
^bDepartment of Chemistry, University of California, Irvine, CA 92697, USA

Abstract

Practical density-corrected density functional theory (DC-DFT) calculations rely on Hartree-Fock (HF) densities, which can be computationally expensive for systems with over a hundred atoms. We extend the applicability of HF-DFT using the dual-basis method, where the density matrix from a smaller basis set is used to estimate the HF solution on a larger basis set. Benchmarks on many systems, including the GMTKN55 database for main-group chemistry, and the L7 and S6L datasets for large molecular systems demonstrate the efficacy of our approach. We apply the dual-basis method to both DNA and HIV systems, and compare with the literature. A careful reparameterisation of HF-r²SCAN-DC4 eliminates the negative s_8 coefficient, with no loss of performance.

Managing large molecular systems in computational chemistry is a significant challenge.[1] While Kohn-Sham density functional theory (KS-DFT) is widely used due to its reasonable balance between accuracy and computational cost, quantitative calculations of large systems using large atomic orbital (AO) basis sets remain unfeasible. Even for systems with 50-100 atoms, using large AO basis sets can be computationally demanding.[2]

Density-corrected DFT (DC-DFT) is a general theoretical framework for identifying failures of density functional approximations (DFAs) by separating the errors into the error due to the functional and the error due to the self-consistent density.[3, 4, 5] In most DFT calculations, the density-driven error is negligible, and the error is dominated by the functional error. But in several well-characterized classes of reactions, density-driven errors are significant.[6, 7, 8] In such cases, it is presumed (and usually true) that elimination of the density-driven error by evaluation on the exact density improves the results.

For practical applications, using the exact density $n(\mathbf{r})$ is too expensive relative to a DFT calculation, so there has been a focus on identifying a density that reduces densitydriven errors (where they are significant) while also being computationally manageable. HF-DFT is a practical approach that uses the Hartree-Fock (HF) density to mitigate density-driven errors when the self-consistent density is sufficiently inaccurate that it impacts the overall energy error. Over the last decade, it has been found, for our standard DFAs built on semilocal approximations, that substituting self-consistent approximate densities with HF densities can significantly reduce energy errors in many classes of DFT calculations.[9, 10, 11, 12, 13, 14, 15, 16, 17]

Recently, a systematic application of the principles of DC-DFT has led to the development of density and dispersioncorrected DFT (D²C-DFT), which exhibits high accuracy in general main group thermochemistry and is particularly effective for systems containing water.[18] However, applying DC-DFT methods with HF densities to large systems presents two key challenges. First, the HF calculation becomes too expensive due to the non-local exchange contributions. Second, dispersion corrections are typically tuned for small systems and may fail for larger systems, where dispersion can account for much of the overall correlation effects.

To reduce the computational cost of HF or DFT on large systems, the dual-basis method has been developed by many, including King and co-workers[19], Head-Gordon and coworkers[20, 21, 22], and Gill and co-workers[23, 24]. With this method, one performs a self-consistent calculation on a small basis set, followed by a perturbative correction, reproducing almost exactly the large basis set result at a fraction of the cost. This produces small but acceptable errors in the HF or DFT energies, relative to the corresponding results in the large basis set.

However, total energies are protected by the variational principle, where small errors in a wavefunction only produce squared errors in the energy. In this work, we test the dual-basis method to see if it produces HF *densities* that are sufficiently accurate for use in HF-DFT calculations. As an immediate application, we can also explore whether dispersion-correction parameters typically tuned on small systems maintain their accuracy as system size grows. In many large systems, the majority of interactions are non-covalent and governed by dispersion. In particular,

^{*}esim@yonsei.ac.kr

[†]equal contribution

[‡]current address; Ai Research Center, SAMSUNG Advanced Institute of Technology, Suwon 16678, Korea

a recent development involves incorporating many-body contributions to dispersion energy, which will be discussed in detail.

The effectiveness of dual-basis method in reducing computational costs, while still providing results comparable to large basis set calculations in both HF and DFT, has been investigated. [20, 22, 23] But such previous calculations have focused on improving energies. Here, we use only the HF density and explore for the first time the errors of the dual-basis method in reproducing the HF density for use in HF-DFT calculations. The dual-basis method reduces the computational cost, but does it introduce errors too large to be useful in DC-DFT calculations?

This letter begins by demonstrating the efficacy of a dualbasis method for obtaining HF densities, paving the way for practical applications of HF-DFT to large molecular systems. Furthermore, we show that the recently developed D^2C -DFT is sufficiently transferable to be applied to large molecules. As the best performer among several D^2C -DFTs, we generally use D^2C -r²SCAN. (Warning: the term large in this paper can refer to large molecular systems, large basis sets, or the chemical category LARGE within GMTKN55, depending on context.)

The dual-basis method for HF-DFT calculations is a distinct approach from the Hartree-Fock perturbative correction (HFPC). A HFPC calculation is performed in several stages. Initially, a self-consistent field calculation is performed with a small primary basis set of size m. Subsequently, a Fock matrix with a larger secondary basis set of size M is formed as follows:

$$F_{ab}^{[1]} = h_{ab} + \sum_{\lambda\sigma}^{m} P_{\lambda\sigma}[(ab|\lambda\sigma) - \frac{1}{2}(a\lambda|b\sigma)], \qquad (1)$$

where $F_{ab}^{[1]}$ is the Fock matrix with the size of M^2 , h is the one-electron operator, and P is the density matrix obtained from the self-consistent field calculation in a small primary basis set. λ and σ (a and b) represent the indexes of the basis function within the primary basis set (secondary basis set). Then, the HFPC energy is given as follows:

$$E^{\rm HFPC} = \sum_{ab}^{M} P_{ab}^{[1]} h_{ab} + \sum_{abcd}^{M} P_{ab}^{[1]} P_{cd}^{[1]} [(ab|cd) - \frac{1}{2} (ac|bd)],$$
(2)

where $P^{[1]}$ is the density matrix obtained from a diagonalization of $F_{ab}^{[1]}$. By replacing the exchange energy (the last term) in Eq. 2 to the KS exchange-correlation energy $\tilde{E}_{\rm XC}$, we can get the dual-basis HF-DFT energy as follows:

$$E^{\rm HF-DFT} = \sum_{ab}^{M} P_{ab}^{[1]} h_{ab} + \sum_{abcd}^{M} P_{ab}^{[1]} P_{cd}^{[1]}(ab|cd) + \widetilde{E}_{\rm XC}(P_{ab}^{[1]}),$$
(3)

In other words, the distinction between the HFPC and the dual-basis HF-DFT calculation lies in the evaluation of energy. Following a single diagonalization, the density matrix is utilized to obtain the HF-DFT energy, not the HFPC energy. (See Supporting Information for the Python code of the dual-basis HF-DFT.)

The HF density matrix from a conventional self-consistent field calculation with a secondary basis set can be obtained with the cost of $O(M^4)$. This can be reduced to the cost of $O(m^2 M^2)$ through the dual-basis method, as expressed in Eq. 1, so, the dual-basis method reduces the cost to the fraction $O((\frac{m}{M})^2)$. Here, we have mostly used the Ahlrichs basis set (def2-) series [25, 26, 27] with the shorthand of 2Z (double- ζ) for def2-SVPD, 3Z(triple- ζ) for def2-TZVPPD, and $4Z(quadruple-\zeta)$ for def2-QZVPPD. The diffuse basis functions (-D) are required to accurately describe the long-range interaction energy in large systems. [28, 29] For notation, the results of DFA or HF with a basis set are denoted as DFA or HF/basis set. Similarly, the results of a D²C-DFT with dual-basis calculation are shown as $D^{2}C$ -DFT/(primary basis)/(secondary basis). For example, the result of a D²C-r²SCAN dual-basis calculation using def2-SVPD and def2-QZVPD as the primary and secondary basis, respectively, is denoted by $D^2C-r^2SCAN/2Z/4Z$.

Figure 1 illustrates both the computational savings and the inaccuracies introduced in the density and energy by the dual-basis method. In Fig. 1(a), the reduction in cost for the C_{60} isomerization energy calculation between the wall times of 2Z/4Z and 4Z is approximately 0.05 (theoretically 0.1). The result obtained from the 2Z/4Z dual-basis with the D²C-r²SCAN (absolute deviation 0.83 kcal/mol) is close to that from 4Z (0.81 kcal/mol), but ran in about onetwentieth of the time. In comparison, the error in a 3Z calculation (1.15 kcal/mol) is much larger and ran 4 times slower.

The dual-basis approach might ruin the spatial distribution of the density even when the energy is accurate. To illustrate how density deviations from different levels of basis sets appear, Figs. 1(b) and (c) show the density deviations with three different basis sets for the He atom and the C_{60} molecule, respectively. Although the spatial distribution of electron density does not fully account for the density-driven energy error, [31] examining this distribution in real space can still provide valuable insights, particularly since accurate energies have already been identified. Given that the purpose of using the dual-basis scheme is to generate the energy in the secondary basis set, the deviation in Figs. 1(b) and (c) is calculated based on the density of the secondary 4Z basis set. Electron densities derived from the 2Z/4Z dual-basis set are noticeably closer to the 4Z electron density compared to electron densities calculated with the 2Z or 3Z basis set. The deviations resulting from the dualbasis method are negligible. The visualization of atomic and molecular densities demonstrates that the dual-basis



Figure 1: (a) Absolute deviations and wall time for the calculation of the C_{60} isomerization energy by D²C-r²SCAN with different basis sets. The absolute deviation is estimated from DLPNO-CCSD(T)/CBS.[30] The 2Z, 3Z, and 4Z correspond to def2-SVPD, def2-TZVPPD, and def2-QZVPPD basis sets, respectively. 2Z/4Z(3Z) denotes a dual-basis result with 2Z as the primary and 4Z(3Z) as the secondary basis sets. The number of basis functions in those basis sets is shown in Table S1 and the total energy is also compared in Fig. S2. The wall time is measured with 64 cores of an Intel(R) Xeon(R) CPU Platinum 8358 @ 2.60GHz. (b) Radial density distribution difference of the He atom, and (c) density difference of the C₆₀ molecule, relative to HF with 4Z. In (c), the isosurface is $|\Delta n(r)|=0.001$, and red shade indicates negative values. The relative total energies are shown in Fig. S3.

method can produce highly accurate electron densities at a fraction of the cost. Moreover, in each case, the deviations introduced by the dual-basis are much smaller than the difference in HF and PBE densities.

From this point forward, we will evaluate the dual-basis HF-DFT approach across various molecular systems, from small to large, using the following benchmark datasets and error metrics. The GMTKN55 is a widely-used benchmark database for general main group thermochemistry, kinetics, and noncovalent interactions.[32]. As the GMTKN55 molecules are typically small, we benchmark two datasets, L7[28] and S6L (6 complexes in the S12L[29] dataset). Both datasets consist mostly of dispersion-stabilized noncovalent complexes. With a maximum of 153 atoms, we refer to this database as L13. (See Fig. S1 for the structure information.) In addition, the DNA-ellipticine and HIV-indinavir systems, containing 157 and 343 atoms, respectively, are evaluated.

For performance evaluation, we compare four DFT methodologies. (See Computational Details for more information.) Self-consistent DFT calculations without any corrections (DFT); self-consistent DFT calculations with Grimme's original D4 dispersion corrections[34] (DFT-D4);

DFT calculations on HF densities (HF-DFT); and a recently proposed density and dispersion-corrected DFT[35], which is HF-DFT with a tailored dispersion correction optimized within the DC-DFT framework (D²C-DFT). Error analysis uses the mean absolute deviation (MAD) of the DFT energy from the reference for each database, which is defined as:

$$MAD_j = \frac{1}{N_j} \sum_{i=1}^{N_j} |E_{ij} - E_{ij}^{ref}|.$$
 (4)

here N_j is the total number of reactions in the *j*th dataset. For the *i*th reaction in the *j*th dataset, E_{ij} is the DFT energy and E_{ij}^{ref} is the reference energy. We set $|\overline{E_j^{\text{ref}}}|$ as the mean absolute energy of this set, and $w_j = \overline{E^{\text{ref}}} / |\overline{E_j^{\text{ref}}}|$, where $\overline{E^{\text{ref}}} = 56.84$ kcal/mol is the mean for all 55 $|\overline{E_j^{\text{ref}}}|$ values. Next, to assess the efficacy of the dual-basis method relative to the result of the secondary basis set, the mean absolute basis deviation (MABD) of this set can also be defined as follows:

MABD_j =
$$\frac{1}{N_j} \sum_{i=1}^{N_j} |E_{ij}/p/s - E_{ij}/s|$$
. (5)



Figure 2: Various depictions of mean absolute basis deviation (MABD) of D^2C -DFT on the GMTKN55 database. In (a), the datasets are ordered as in Ref. [32] and classified according to chemical properties as in Ref. [33]. The ratio of MABD (2Z/4Z) to MAD (mean absolute deviation) is below 0.2 (horizontal line) for most datasets. Marker size is proportional to the weighted MABD, the weighted version of Eq. 5, in kcal/mol. The square in the upper left corner denotes 5.36 kcal/mol, the WTMAD-2 of D^2C -r²SCAN.; (b) MABD for three approximations: D^2C -PBE, -r²SCAN, and -B3LYP. (MADs of D^2C -DFTs are shown in Fig. S6.); (c) shows MABD for various basis set pairings for the def2-series of basis sets, with def2- omitted for simplicity. (*In the case of def2-TZVPD/def2-QZVPD, the first structure within the C60ISO dataset is evaluated self-consistently with def2-QZVPD due to a convergence issue. See Table S4 and Fig. S7.); (d) Same as (c) but for just the G21EA dataset (adiabatic electron affinities) where a sizable primary basis set is needed.

where E/p/s is the dual-basis energy evaluated with the primary and the secondary basis set, and E/s is the energy evaluated in the secondary basis set. Therefore, MABD measures the deviation due to the use of a dual-basis set with respect to the results of the same calculation of secondary basis.

The GMTKN55 database comprises 55 sub-datasets, and the weighted total mean absolute deviation (WTMAD-2) is[32]

WTMAD-2 =
$$\frac{1}{N} \sum_{j=1}^{55} w_j \sum_{i=1}^{N_j} |E_{ij} - E_{ij}^{ref}| = \frac{1}{N} \sum_{j=1}^{55} N_j w_j \text{MAD}_j$$
(6)

where N is the total number of reactions, $\sum_{j=1}^{55} N_j = 1505$. The weighted version of Eq. 5 can also be obtained by replacing MAD_j in Eq. 6 to MABD_j. (The weights for each dataset are listed in Tables S2 and S3.)

The dual-basis method offers efficient access to the results

of the secondary basis set. It is not necessary for the primary basis set to be a subset of the secondary basis set, as the density obtained from the small basis set could be projected into the large basis set.[20] In addition, HF perturbative corrections are less sensitive to the basis set pairing.[24] However, the MABD can vary depending on the system. To assess the significance of these deviations, we first test the pairings of basis sets for dual-basis HF calculations using a series of Ahlrichs (def2-) basis sets on the GMTKN55 database. Additionally, in Fig. S4, the same results could be obtained from the Bauza30[36] dataset within the pairings over the def2- series with Jensen (pc-)[37, 38], and Dunning (cc-)[39] families. The dataset contains mainly noncovalent interactions, including halogen, chalcogen, and pnictogen bonds, which are prevalent in large molecular systems.

Figure 2 captures both the reliability and caveats associated with the choice of basis set pairs used in the dualbasis D^2C -DFT calculations, using the GMTKN55 database.

Figure 2(a) shows the ratio of the $D^2C-r^2SCAN/2Z/4Z$ MABD and the D²C-r²SCAN/4Z MAD for each dataset in GMTKN55. This is as important, below 0.2 (horizontal line) means the dual-basis result affects the error by less than 20 %, i.e., can be ignored, as is the case for most datasets. Marker size represents the weighted MABD of each dataset. The few datasets above the 20 % line either have weighted MABDs much smaller than WTMAD-2 (indicated by a tiny marker size) or have 4Z MADs less than 1 kcal/mol, i.e., an extremely small MAD yields an unusually large ratio. The MAD of BSR36 dataset, bond-separation reactions, is only 0.33 kcal/mol, so even the very small MABD of 0.38 kcal/mol yields a ratio larger than 1. Likewise, the MABD of RG18, which contains the interaction energies in rare-gas complex, is only 0.04 kcal/mol, but because of its large WTMAD-2 weight of 98, yields an unusually large weighted MABD of 3.92 kcal/mol. (See Tables S2 and S3 for the weight information.) Similar behavior can be seen in Fig. S5 for D²C-PBE and D²C-B3LYP. In Fig. 2(b), we show MABD for three different approximations. MABD is considerably smaller than MAD (Fig. S6) in scale. Depending on the approximations, the MADs are widely distributed, but the MABDs are not. The largest MABD is from MB16-43 because of its large reference energy. All three functionals have comparable MABD.

In Fig. 2(c), the MABD between the results of the dualbasis D²C-r²SCAN and those of D²C-r²SCAN evaluated with the secondary basis set on GMTKN55 is presented. Since the dispersion correction is independent of the basis set, it cancels, and the results are the same as those of HFr²SCAN. Among dual-basis combinations, we recommend 2Z/4Z as a default choice, balancing cost and accuracy. Using a large primary, such as 3Z/4Z, results in a lower MABD, but with much smaller computational savings.

DFT calculations have basis set incompleteness error (BSIE)[40] and basis set superposition error (BSSE)[41] that depend on the size of the basis set, which remains true in the dual-basis method. The G21EA dataset, which is dominated by the slow convergence of the anions, requires special attention. In Fig. 2(d), we show that the G21EA dataset requires a primary basis set of at least def2-TZVPD levels, which is superior to the def2-QZVP primary basis set result. (See Fig. S8 for the separate energy distributions of neutral species and anions in the G21EA dataset.)

For the entire GMTKN55 database, including the aforementioned datasets, the dual-basis 2Z/4Z can change the WTMAD-2 of D²C-DFT by up to 1~3 percentage. (See Table S5.) Moreover, the interaction energies in large systems are predominantly driven by inter- and intra-molecular dispersion. As the 2Z/4Z dual-basis set, as shown in pink and green markers in Figs. 2(a) and (b), does not cause significant problems, it will be utilized in the following discussion for the L13 database.

Figure 3 shows the absolute deviations of D²C-DFT en-



Figure 3: (a) Absolute energy deviations of D²C-r²SCAN for GMTKN55 and (b) subsets with more than 230 electrons. The absolute energy deviation is defined as $|\Delta E| = |\tilde{E} - E^{ref}|$

ergies with dual-basis and secondary basis. In Fig. 3(a), we can see that $D^2C-r^2SCAN/2Z/4Z$ approximates the secondary basis energies well, with many points lying near the diagonal line. There is a huge improvement over 2Z results. A few dual-basis results are better than 4Z, but this is surely accidental. Most of the points deviating significantly from the diagonal originate from the difficult MB16-43 dataset, due to a huge range of energies ($\overline{|E^{\text{ref}}|} = 468 \text{ kcal/mol}$). The MAD of D^2C -r²SCAN with 2Z, 2Z/4Z, and 4Z for the MB16-43 dataset are 19.72, 10.37, and 9.96 kcal/mol, respectively. Considering the energy range, the associated deviations are comparatively small ($\sim 1 \%$ of the reference energy). (See Fig. S9.) Especially, the consistency of the 2Z/4Z results with the 4Z results is more evident for the three relatively large systems (C60ISO, ISOL24, and UPU23) in Fig. 3(b), suggesting that the performance of the dual-basis method is not significantly affected by the size of the system.

Table 1 provides an analysis of the GMTKN55 into relatively large systems, closed/open-shell, and densityinsensitive/sensitive cases. The increased WTMAD-2 of 3.32 kcal/mol between 2Z and 4Z energies is significantly reduced to 0.07 kcal/mol with 2Z/4Z. Reduction in WTMAD-2 of more than 3 kcal/mol is observed in relatively large

Table 1: Weighted total mean absolute deviations (WTMAD-2) of D^2C -r²SCAN for the GMTKN55 database in kcal/mol.

$D^{2}C$ -r ² SCAN	WTMAD-2	$\Delta WTMAD-2^a$		
$(\rm kcal/mol)$	4Z	2Z	2Z/4Z	
GMTKN55	5.36	3.32	0.07	
C60ISO, UPU23, ISOL 24^b	5.32	3.12	0.20	
Closed-Shell	5.58	4.21	0.07	
Open-Shell	4.81	1.08	0.06	
Density-Insensitive	5.42	3.94	0.04	
Density-Sensitive	5.22	1.90	0.13	

 $^{a}\Delta$ WTMAD-2 represents the difference from WTMAD-2 of 4Z.

^bThe datasets with over 230 electrons in the GMTKN55 database.

systems, closed-shell systems, and density-insensitive cases. The WTMAD-2 of D²C-r²SCAN with 4Z is reduced to 5.23 kcal/mol when using a restricted open-shell HF (ROHF) density with high spin contamination (> 0.1) in unrestricted HF (UHF) wavefunction.[42] For the dual-basis calculations here, open shell systems were computed with UHF densities. When available, ROHF densities should further reduce errors.

In large non-covalent systems, dispersion interactions dominate. Table 2 shows that the dispersion energy contributes on average 10 kcal/mol in the case of HF-r²SCAN, accounting for approximately 60 % of the total interaction energies. (See also Table S6.) This highlights the importance of an adequate description of the dispersion energy $(E_{\rm disp})$ in large molecular systems. Most approximate functionals do not fully capture dispersion interactions[44] and many different dispersion correction methods have been developed, such as notably DFT-D[45, 46, 34], exchange dipole moment (XDM) dispersion model[47], and many-body dispersion (MBD) method[48].

In particular, the importance of the many-body contribution (often, three-body term) of dispersion energy arises when considering large non-covalently bounded systems and molecular crystals.[49] Despite the ongoing debate over the use of the Axilrod-Teller-Muto (ATM) three-body term,[50, 51, 52] the significance of the many-body effect of dispersion energy in large systems remains undeniable. In small molecular systems, the three-body interaction is negligible, but in large molecular systems, the effect becomes apparent. The aforementioned methods are often parameterized only for small complexes, raising the question of whether they will work well for larger ones.

From Ref. [18], DC4 represents a DFT-D4 correction founded on the DC-DFT principle, which focuses on minimizing only the functional error by optimizing parameters solely for density-insensitive cases. The right columns of Table 2 show that the three-body dispersion energy aver-

ages 2.9 kcal/mol and constitutes approximately 20 % of the total dispersion energy in the original DC4. This is substantial, given the variations among different functionals and the possibility for three-body interactions to become more pronounced in large molecular systems. Many believe that the contribution of three-body interactions is up to 10 % of the total $E_{\text{disp.}}[34, 49]$ The increased three-body dispersion energy can be explained by the negative s_8 , an nonphysical artifact of the repulsive R^{-8} , which leads to excessive R^{-6} and three-body ATM terms.[53] The two effects cancel each other out and seem to work well on the small system, but when we move to the large molecular systems, an excessive ATM term is observed. To address this issue, we developed D²C-DFT by adjusting the parameters to pass through the validation of a large molecular system dataset.[35] D²C-DFT has an attractive R^{-8} term, and the ATM term is reasonable as shown in the left columns in Table 2.

The performance of D^2C -DFT with the dual-basis method is assessed on the L13, a challenging large-molecule dataset. The results are benchmarked against those of the GMTKN55 in Fig. 4. Figure 4(a) shows the monotonic improvements along Jacob's ladder[55] for the GMTKN55 database. In Fig. 4(b), the sequence of the approximations is disordered. D²C-BLYP shows improvements over the corresponding DFT-D4, but D²C-revPBE slightly lowers the performance of revPBE-D4, perhaps due to error compensation between dispersion energy and charge transfer energy of revPBE-D4, [56] i.e., the HF density disturbs the balance of the two opposite signed errors. D^2C -TPSS and D²C-B3LYP also show worse results compared to the corresponding DFT-D4. For r²SCAN-D4, D²C-r²SCAN improves the results and HF-r²SCAN-XDM improves even more. Interestingly, $\omega B97M$ -V[57], regarded as the best workhorse functional, [58] exhibits a similar performance compared to the other approximations.

We optimized the XDM parameters with HF-r²SCAN/4Z following the DC-DFT principle, referred to as HF-r²SCAN-XDM. The optimized XDM parameters are shown with explanations in the Table S7. The results of HF-r²SCAN-XDM demonstrate that significant improvements to large molecular systems can be achieved without compromising the capacity of small systems.

Figure 5 shows D²C-r²SCAN results for the large pharmacological molecules, DNA-ellipticine and HIV-indinavir.[43] The number of electrons in DNA-ellipticine is 710, and HIV-indinavir contains 1088 electrons, about 3 times larger than C₆₀, the biggest molecule in GMTKN55. For DNA-ellipticine, the result of D²C-r²SCAN/2Z/4Z (-41.6 kcal/mol) lies within the error bar of the reference (-40.4 \pm 1.47 kcal/mol). For HIV-indinavir, D²Cr²SCAN/2Z/4Z predicts -117.1 kcal/mol and the reference is -121.5 kcal/mol. Additionally, we calculated D²Cr²SCAN/2Z/3Z with and without counterpoise correction

	D^2C - r^2SCAN^b				HF-r ² SCAN-DC4 ^{c}							
dataset	complex	$E_{\rm int}^{{\rm ref}d}$	$\Delta E_{\rm int}^{\ e}$	$E_{\rm disp}$	$E_{\rm disp}^{(6)}$	$E_{\rm disp}^{(8)}$	$E_{\rm disp}^{(9)}$	$\Delta E_{\rm int}$	$E_{\rm disp}$	$E_{\rm disp}^{(6)}$	$E_{\rm disp}^{(8)}$	$E_{\rm disp}^{(9)}$
L7	C2C2PD	-20.9	-1.6	-15.2	-17.0	-0.2	2.0	-0.1	-13.7	-21.3	3.7	3.9
	C3A	-17.5	0.2	-11.1	-12.4	-0.1	1.4	1.2	-10.1	-15.2	2.5	2.6
	C3GC	-29.2	-0.4	-19.2	-21.5	-0.2	2.5	1.2	-17.6	-26.5	4.4	4.5
	CBH	-11.0	2.2	-8.6	-9.2	-0.1	0.7	3.0	-7.8	-10.1	1.3	1.0
	GCGC	-13.5	-2.0	-11.1	-11.9	-0.1	0.9	-1.4	-10.5	-14.6	2.4	1.6
	GGG	-2.1	-0.6	-4.3	-4.5	0.0	0.2	-0.4	-4.1	-5.4	0.9	0.5
	PHE	-25.5	1.0	-5.3	-5.6	-0.1	0.4	1.3	-5.0	-6.4	0.9	0.6
S6L	2a	-35.5	-0.3	-18.2	-19.9	-0.2	1.9	0.9	-17.0	-24.9	4.3	3.6
	2b	-21.8	-2.7	-13.5	-14.6	-0.2	1.3	-1.7	-12.5	-18.3	3.1	2.6
	4a	-42.5	2.3	-29.0	-32.8	-0.4	4.2	3.3	-28.0	-41.2	6.9	7.3
	5a	-43.8	1.0	-12.9	-13.8	-0.1	1.0	1.5	-12.5	-16.5	2.6	1.5
	6a	-89.6	2.2	-11.4	-13.6	-0.1	2.3	3.6	-10.0	-15.5	2.1	3.4
	$7\mathrm{b}$	-28.6	2.9	-17.2	-20.5	-0.2	3.6	5.0	-15.0	-23.3	3.1	5.2
Mean -29.		-29.4	0.3	-13.6	-15.2	-0.2	1.7	1.3	-12.6	-18.4	2.9	2.9

Table 2: Comparison of D²C-r²SCAN and HF-r²SCAN-DC4 for interaction (E_{int}) and dispersion energies^{*a*} (E_{disp}) for the L13 database in kcal/mol.

^{*a*} The dispersion energy is decomposed as the components of the dipole-dipole $(E_{\text{disp}}^{(6)})$, dipole-quadrupole $(E_{\text{disp}}^{(8)})$ and triple-dipole ATM $(E_{\text{disp}}^{(9)})$. ^{*b*} Used dispersion parameters in Ref. [35]. ^{*c*} Used dispersion parameters in Ref. [18]. ^{*d*} DLPNO-CCSD(T_0)/CBS[43]. ^{*e*} $\Delta E_{\text{int}} = E_{\text{int}} - E_{\text{int}}^{\text{ref}}$.

and the results are -115.6 and -117.6 kcal/mol, respectively. The results demonstrate that the parameter fitting in the dispersion correction scheme for small molecules is sufficiently transferable to large molecules, and that the dual-basis method can be employed to efficiently obtain results.

We point out that reference interaction energies from Quantum Monte Carlo and CCSD(T) calculations can differ significantly in large molecular systems. [59, 60, 61] For L13 and DNA-ellipticine system, the discrepancy between FN-DMC and DLPNO-CCSD(T₀)/CBS can exceed 1 kcal/mol. [43] For the L13 database, the discrepancy is typically $0\sim2$ kcal/mol, but for the 4a system in S6L, it reaches 10.4 kcal/mol. For DNA-ellipticine, the discrepancy is 4.4 kcal/mol, but could be as large as 9.2 kcal/mol considering the error bars. As such discrepancies (and their origin) are beyond the scope of the present work, here we simply benchmark against the common standard, CCSD(T).

In electronic structure calculations, the accuracy practically depends on the quality of the basis set, but the use of a large basis set is formidable when calculating large systems. Therefore, a small basis set with counterpoise correction can be an efficient choice to mitigate the impact of BSSE.[62] However, the counterpoise correction can be applied to the interaction energy, not the binding energy, and it cannot fully correct the BSIE. While the validity of the counterpoise correction in the HF-DFT scheme has been verified (see Fig. S10), here we recommend using the dual-basis method without the counterpoise correction to achieve large basis set results at a lower cost, allowing for effective correction of both BSSE and BSIE simultaneously.

In this work, we demonstrated that the dual-basis method can be a promising approach by evaluating the HF densities and applying them to HF-DFT. The dual-basis HF densities reproduce the secondary HF densities well both in terms of spatial distribution and energy. A very small change in WTMAD-2 on GMTKN55 demonstrates the efficacy of the dual-basis method.

In large systems, where dispersion interactions play an important role, we have found that three-body interactions need to be considered carefully when applying dispersion corrections. In this regard, D²C-r²SCAN corrects the parameters of HF-r²SCAN-DC4 to avoid the nonphysical artifact of a repulsive $E_{disp}^{(8)}$. Among the D²C-DFTs, D²C-r²SCAN shows good performance without compromising accuracy on small molecules (GMTKN55) and HF-r²SCAN-XDM shows even greater improvements. Accurate energy predictions on pharmacological molecules show the potential for extension to larger molecules.

We have shown that DC-DFT can be extended to large molecular systems through the dual-basis scheme with proper dispersion correction. We expect that this approach





Figure 4: (a) WTMAD-2 of the GMTKN55 database and (b) mean absolute deviations (MADs) of the L13 database in kcal/mol. Density functional approximations are ordered according to the WTMAD-2 values for each rung of Jacob's ladder. WTMAD-2 of the ω B97M-V and B97M-V are from Ref. [54] where 3Z with counterpoise correction is used. Self-consistent DFTs and D^2C -DFTs are from Ref. [35] where 4Z is used. For L13, 3Z with counterpoise correction is used for self-consistent DFT calculations. D²C-DFTs and HF-r²SCAN-XDM use 2Z/4Z for L13. The reference data (DLPNO-CCSD(T_0)/CBS), ω B97M-V, and B97M-V results for the L13 database are from Ref. [43] where 3Z with counterpoise correction is used. The bars representing D²C-DFTs with HF-r²SCAN-XDM are filled, while those representing the other approximations are empty. The optimized XDM parameters for HF-r²SCAN are shown in the Table S7.

will allow us to explore a wide range of larger systems accurately at a reasonable cost.

COMPUTATIONAL DETAILS

All HF and DFT calculations are performed through the Python-based Simulations of Chemistry Framework (PySCF)[63], using customized Python codes for the dualbasis method. The following functionals have been used in DFT and HF-DFT calculations: PBE[64], BLYP[65, 66], and revPBE[67] for GGA, TPSS[68], and r²SCAN[69] for mGGA, B3LYP[70, 71] for hybrid. Density fitting, resolution of identity[72, 73, 74], was used for all calculations.

8

Figure 5: Energy predictions for large non-covalent complexes of pharmacological interest, (a) DNA-ellipticine and (b) HIVindinavir. Shades in (a) represents error bars for each reference. All results except D^2C-r^2SCAN are taken from Ref. [43] where 3Z with counterpoise correction was used for the DFT calculations. For D^2C-r^2SCAN , 2Z/4Z was used.

ACKNOWLEDGMENTS

We are grateful for support from the National Research Foundation of Korea (NRF-2020R1A2C2007468) and the Korea Environment Industry & Technology Institute (KEITI) through the "Advanced Technology Development Project for Predicting and Preventing Chemical Accidents" Program funded by the Korea Ministry of Environment (MOE) (RS-2023-00219144). K.B. acknowledges support from NSF grant No.CHE-2154371.

Supporting Information Available:

PySCF code for dual-basis HF-DFT, additional analysis for D²C-DFTs on various database, XDM parameters for HF-r²SCAN, and raw data of calculations.

References

[1] Stefan Grimme and Peter R Schreiner. Computational chemistry: The fate of current methods and future challenges. <u>Angewandte Chemie (International ed. in</u> English), 57(16):4170–4176, 2017.

- [2] Markus Bursch, Jan-Michael Mewes, Andreas Hansen, and Stefan Grimme. Best-practice dft protocols for basic molecular computational chemistry**. <u>Angewandte</u> <u>Chemie International Edition</u>, 61(42):e202205735, 2022.
- [3] Min-Cheol Kim, Eunji Sim, and Kieron Burke. Understanding and reducing errors in density functional calculations. <u>Physical review letters</u>, 111(7):073003, 2013.
- [4] Eunji Sim, Suhwan Song, Stefan Vuckovic, and Kieron Burke. Improving results by improving densities: Density-corrected density functional theory. Journal of the American Chemical Society, 144(15):6625–6639, 2022.
- [5] Suhwan Song, Stefan Vuckovic, Eunji Sim, and Kieron Burke. Density-corrected dft explained: Questions and answers. <u>Journal of chemical theory and computation</u>, 18(2):817–827, 2022.
- [6] Yeil Kim, Suhwan Song, Eunji Sim, and Kieron Burke. Halogen and chalcogen binding dominated by densitydriven errors. <u>The journal of physical chemistry letters</u>, 10(2):295–301, 2018.
- [7] Eunji Sim, Min-Cheol Kim, and Kieron Burke. Dissociation curves of diatomic molecules: A dc-dft study. In <u>AIP Conference Proceedings</u>, volume 1702. AIP Publishing, 2015.
- [8] Min-Cheol Kim, Eunji Sim, and Kieron Burke. Ions in solution: Density corrected density functional theory (dc-dft). <u>The Journal of chemical physics</u>, 140(18), 2014.
- [9] Seungsoo Nam, Suhwan Song, Eunji Sim, and Kieron Burke. Measuring density-driven errors using kohnsham inversion. <u>Journal of chemical theory and</u> computation, 16(8):5014–5023, 2020.
- [10] Seungsoo Nam, Eunbyol Cho, Eunji Sim, and Kieron Burke. Explaining and fixing dft failures for torsional barriers. <u>The journal of physical chemistry letters</u>, 12(11):2796–2804, 2021.
- [11] Min-Cheol Kim, Eunji Sim, and Kieron Burke. Communication: Avoiding unbound anions in density functional calculations. <u>The Journal of chemical physics</u>, 134(17):171103, 2011.
- [12] Min-Cheol Kim, Hansol Park, Suyeon Son, Eunji Sim, and Kieron Burke. Improved dft potential energy surfaces via improved densities. <u>The journal of physical</u> chemistry letters, 6(19):3802–3807, 2015.

- [13] Min-Cheol Kim, Eunji Sim, and Kieron Burke. Ions in solution: Density corrected density functional theory (dc-dft). <u>The Journal of chemical physics</u>, 140(18):18A528, 2014.
- [14] Suhwan Song, Min-Cheol Kim, Eunji Sim, Anouar Benali, Olle Heinonen, and Kieron Burke. Benchmarks and reliable dft results for spin gaps of small ligand fe (ii) complexes. <u>Journal of chemical theory</u> and computation, 14(5):2304–2311, 2018.
- [15] Yeil Kim, Suhwan Song, Eunji Sim, and Kieron Burke. Halogen and chalcogen binding dominated by densitydriven errors. <u>The journal of physical chemistry letters</u>, 10(2):295–301, 2019.
- [16] Benjamin G Janesko and Gustavo E Scuseria. Hartree– fock orbitals significantly improve the reaction barrier heights predicted by semilocal density functionals. <u>The</u> Journal of chemical physics, 128(24):244112, 2008.
- [17] Prakash Verma, Ajith Perera, and Rodney J Bartlett. Increasing the applicability of dft i: Non-variational correlation corrections from hartree–fock dft for predicting transition states. <u>Chemical Physics Letters</u>, 524:10–15, 2012.
- [18] Suhwan Song, Stefan Vuckovic, Youngsam Kim, Hayoung Yu, Eunji Sim, and Kieron Burke. Extending density functional theory with near chemical accuracy beyond pure water. <u>Nature communications</u>, 14(1):799, 2023.
- [19] S Havriliak and Harry F King. Rydberg radicals. 1. frozen-core model for rydberg levels of the ammonium radical. <u>Journal of the American Chemical Society</u>, 105(1):4–12, 1983.
- [20] Liang and Martin Head-Gordon. Approaching the basis set limit in density functional theory calculations using dual basis sets without diagonalization. <u>The Journal</u> of Physical Chemistry A, 108(15):3206–3210, 2004.
- [21] Ryan P Steele, Robert A DiStasio, Yihan Shao, Jing Kong, and Martin Head-Gordon. Dual-basis secondorder møller-plesset perturbation theory: A reducedcost reference for correlation calculations. <u>The Journal</u> of chemical physics, 125(7), 2006.
- [22] Ryan P Steele, Robert A DiStasio Jr, and Martin Head-Gordon. Non-covalent interactions with dual-basis methods: Pairings for augmented basis sets. Journal of Chemical Theory and Computation, 5(6):1560–1572, 2009.
- [23] Jia Deng, Andrew TB Gilbert, and Peter MW Gill. Approaching the hartree–fock limit by perturbative methods. <u>The Journal of chemical physics</u>, 130(23), 2009.

- [24] Jia Deng, Andrew TB Gilbert, and Peter MW Gill. Hartree–fock perturbative corrections for total and reaction energies. <u>The Journal of chemical physics</u>, 133(4), 2010.
- [25] Florian Weigend, Marco Häser, Holger Patzelt, and Reinhart Ahlrichs. Ri-mp2: optimized auxiliary basis sets and demonstration of efficiency. <u>Chemical Physics</u> <u>Letters</u>, 294(1):143–152, 1998.
- [26] Florian Weigend, Filipp Furche, and Reinhart Ahlrichs. Gaussian basis sets of quadruple zeta valence quality for atoms H–Kr. <u>The Journal of Chemical Physics</u>, 119(24):12753–12762, 12 2003.
- [27] Florian Weigend and Reinhart Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: Design and assessment of accuracy. <u>Phys. Chem. Chem. Phys.</u>, 7:3297–3305, 2005.
- [28] Robert Sedlak, Tomasz Janowski, Michal Pitonak, Jan Rezac, Peter Pulay, and Pavel Hobza. Accuracy of quantum chemical methods for large noncovalent complexes. Journal of chemical theory and computation, 9(8):3364–3374, 2013.
- [29] Tobias Risthaus and Stefan Grimme. Benchmarking of london dispersion-accounting density functional theory methods on very large molecular complexes. Journal of chemical theory and computation, 9(3):1580–1591, 2013.
- [30] Rebecca Sure, Andreas Hansen, Peter Schwerdtfeger, and Stefan Grimme. Comprehensive theoretical study of all 1812 c60 isomers. <u>Phys. Chem. Chem. Phys.</u>, 19:14296–14305, 2017.
- [31] Eunji Sim, Suhwan Song, and Kieron Burke. Quantifying density errors in dft. <u>The journal of physical</u> chemistry letters, 9(22):6385–6392, 2018.
- [32] Lars Goerigk, Andreas Hansen, Christoph Bauer, Stephan Ehrlich, Asim Najibi, and Stefan Grimme. A look at the density functional theory zoo with the advanced gmtkn55 database for general main group thermochemistry, kinetics and noncovalent interactions. Phys. Chem. Chem. Phys., 19:32184–32215, 2017.
- [33] Axel D Becke, Golokesh Santra, and Jan ML Martin. A double-hybrid density functional based on good local physics with outstanding performance on the gmtkn55 database. <u>The Journal of Chemical Physics</u>, 158(15), 2023.
- [34] Eike Caldeweyher, Christoph Bannwarth, and Stefan Grimme. Extension of the D3 dispersion coefficient model. <u>The Journal of Chemical Physics</u>, 147(3):034112, 07 2017.

- [35] Minhyeok Lee, Byeongjae Kim, Mingyu Sim, Mihira Sogal, Youngsam Kim, Hayoung Yu, Kieron Burke, and Eunji Sim. Correcting dispersion corrections with density-corrected dft. <u>Journal of Chemical Theory and</u> Computation, 2024.
- [36] Antonio Bauzá, Ibon Alkorta, Antonio Frontera, and José Elguero. On the reliability of pure and hybrid dft methods for the evaluation of halogen, chalcogen, and pnicogen bonds involving anionic and neutral electron donors. Journal of Chemical Theory and Computation, 9(11):5201–5210, 2013. PMID: 26583427.
- [37] Frank Jensen. Polarization consistent basis sets: Principles. <u>The Journal of Chemical Physics</u>, 115(20):9113– 9125, 11 2001.
- [38] Frank Jensen. Polarization consistent basis sets. II. Estimating the Kohn–Sham basis set limit. <u>The Journal</u> of Chemical Physics, 116(17):7372–7379, 04 2002.
- [39] Jr. Dunning, Thom H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. <u>The Journal of Chemical</u> Physics, 90(2):1007–1023, 01 1989.
- [40] F. Jensen. Introduction to Computational Chemistry, volume vol. 2. Wiley, Hoboken, 2007.
- [41] Roman M. Balabin. Enthalpy difference between conformations of normal alkanes: Intramolecular basis set superposition error (bsse) in the case of <i>n</i>-butane and <i>n</i>-hexane. JOURNAL OF CHEMICAL PHYSICS, 129(16), OCT 28 2008.
- [42] Hayoung Yu, Suhwan Song, Seungsoo Nam, Kieron Burke, and Eunji Sim. Density-corrected density functional theory for open shells: How to deal with spin contamination. <u>The Journal of Physical Chemistry</u> Letters, 14(41):9230–9237, 2023.
- [43] Corentin Villot, Francisco Ballesteros, Danyang Wang, and Ka Un Lao. Coupled cluster benchmarking of large noncovalent complexes in 17 and s12l as well as the c60 dimer, dna–ellipticine, and hiv–indinavir. <u>The</u> <u>Journal of Physical Chemistry A</u>, 126(27):4326–4341, 2022. PMID: 35766331.
- [44] Aron J Cohen, Paula Mori-Sánchez, and Weitao Yang. Challenges for density functional theory. <u>Chemical</u> <u>reviews</u>, 112(1):289–320, 2012.
- [45] Stefan Grimme. Semiempirical gga-type density functional constructed with a long-range dispersion correction. <u>Journal of Computational Chemistry</u>, 27(15):1787–1799, 2006.

- [46] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. <u>The Journal</u> of Chemical Physics, 132(15):154104, 04 2010.
- [47] Axel D. Becke and Erin R. Johnson. Exchange-hole dipole moment and the dispersion interaction. <u>The</u> <u>Journal of Chemical Physics</u>, 122(15):154104, 04 2005.
- [48] Alexandre Tkatchenko, Robert A. DiStasio, Roberto Car, and Matthias Scheffler. Accurate and efficient method for many-body van der waals interactions. Phys. Rev. Lett., 108:236402, Jun 2012.
- [49] Stefan Grimme, Andreas Hansen, Jan Gerit Brandenburg, and Christoph Bannwarth. Dispersion-corrected mean-field electronic structure methods. <u>Chemical</u> Reviews, 116(9):5105–5154, 2016. PMID: 27077966.
- [50] O. Anatole von Lilienfeld and Alexandre Tkatchenko. Two- and three-body interatomic dispersion energy contributions to binding in molecules and solids. <u>The</u> Journal of Chemical Physics, 132(23):234109, 06 2010.
- [51] Emil Proynov, Fenglai Liu, Zhengting Gan, Matthew Wang, and Jing Kong. Density-functional approach to the three-body dispersion interaction based on the exchange dipole moment. <u>The Journal of Chemical</u> <u>Physics</u>, 143(8):084125, 08 2015.
- [52] A. Otero-de-la Roza and Erin R. Johnson. Many-body dispersion interactions from the exchange-hole dipole moment model. <u>The Journal of Chemical Physics</u>, 138(5):054103, 02 2013.
- [53] Marvin Friede, Sebastian Ehlert, Stefan Grimme, and Jan-Michael Mewes. Do optimally tuned rangeseparated hybrid functionals require a reparametrization of the dispersion correction? it depends. Journal of Chemical Theory and Computation, 19(22):8097-8107, 2023. PMID: 37955590.
- [54] Asim Najibi and Lars Goerigk. The nonlocal kernel in van der waals density functionals as an additive correction: An extensive analysis with special emphasis on the b97m-v and ω b97m-v approaches. Journal of <u>Chemical Theory and Computation</u>, 14(11):5725–5738, 2018.
- [55] John P. Perdew and Karla Schmidt. Jacob's ladder of density functional approximations for the exchangecorrelation energy. In <u>Density Functional Theory and its Application to Materials</u>, volume 577 of <u>American Institute of Physics Conference Series</u>, pages 1–20, 07 2001.

- [56] Saswata Dasgupta, Etienne Palos, Yuanhui Pan, and Francesco Paesani. Balance between physical interpretability and energetic predictability in widely used dispersion-corrected density functionals. Journal of Chemical Theory and Computation, 20(1):49–67, 2023.
- [57] Narbe Mardirossian and Martin Head-Gordon. ω B97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. <u>The Journal of Chemical Physics</u>, 144(21):214110, 06 2016.
- [58] Narbe Mardirossian and Martin Head-Gordon. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. <u>Molecular Physics</u>, 115(19):2315– 2372, 2017.
- [59] Yasmine S Al-Hamdani, Péter R Nagy, Andrea Zen, Dennis Barton, Mihály Kállay, Jan Gerit Brandenburg, and Alexandre Tkatchenko. Interactions between large molecules pose a puzzle for reference quantum mechanical methods. <u>Nature Communications</u>, 12(1):3927, 2021.
- [60] Francisco Ballesteros, Shelbie Dunivan, and Ka Un Lao. Coupled cluster benchmarks of large noncovalent complexes: The l7 dataset as well as dna–ellipticine and buckycatcher–fullerene. <u>The Journal of Chemical</u> Physics, 154(15), 2021.
- [61] Tobias Schäfer, Andreas Irmler, Alejandro Gallo, and Andreas Grüneis. Understanding discrepancies of wavefunction theories for large molecules. <u>arXiv preprint</u> arXiv:2407.01442, 2024.
- [62] Montgomery Gray and John Herbert. Density functional theory for van der waals complexes: Size matters. <u>Annual Reports in Computational Chemistry</u>, 20:1–61, <u>2024</u>.
- [63] Qiming Sun, Timothy C. Berkelbach, Nick S. Blunt, George H. Booth, Sheng Guo, Zhendong Li, Junzi Liu, James D. McClain, Elvira R. Sayfutyarova, Sandeep Sharma, Sebastian Wouters, and Garnet Kin-Lic Chan. Pyscf: the python-based simulations of chemistry framework. <u>WIRES Computational Molecular Science</u>, 8(1):e1340, 2018.
- [64] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. Phys. Rev. Lett., 77:3865–3868, Oct 1996.
- [65] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. <u>Phys.</u> <u>Rev. A</u>, 38:3098–3100, Sep 1988.

- [66] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. <u>Phys.</u> Rev. B, 37:785–789, Jan 1988.
- [67] Yingkai Zhang and Weitao Yang. Comment on "generalized gradient approximation made simple". <u>Phys.</u> Rev. Lett., 80:890–890, Jan 1998.
- [68] Jianmin Tao, John P. Perdew, Viktor N. Staroverov, and Gustavo E. Scuseria. Climbing the density functional ladder: Nonempirical meta–generalized gradient approximation designed for molecules and solids. <u>Phys.</u> Rev. Lett., 91:146401, Sep 2003.
- [69] James W Furness, Aaron D Kaplan, Jinliang Ning, John P Perdew, and Jianwei Sun. Accurate and numerically efficient r2scan meta-generalized gradient approximation. <u>The journal of physical chemistry letters</u>, 11(19):8208-8215, 2020.
- [70] Philip J Stephens, Frank J Devlin, Cary F Chabalowski, and Michael J Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. <u>The Journal of physical</u> chemistry, 98(45):11623–11627, 1994.
- [71] Axel D. Becke. Density-functional thermochemistry. III. The role of exact exchange. <u>The Journal of Chemical</u> Physics, 98(7):5648–5652, 04 1993.
- [72] Karin Eichkorn, Oliver Treutler, Holger Öhm, Marco Häser, and Reinhart Ahlrichs. Auxiliary basis sets to approximate coulomb potentials. <u>Chemical physics</u> letters, 240(4):283–290, 1995.
- [73] Herbert A Früchtl, Rick A Kendall, Robert J Harrison, and Kenneth G Dyall. An implementation of ri–scf on parallel computers. <u>International journal of quantum</u> chemistry, 64(1):63–69, 1997.
- [74] Florian Weigend. A fully direct ri-hf algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. <u>Physical Chemistry</u> Chemical Physics, 4(18):4285–4291, 2002.