Title

Exploring the limit of accuracy for density functionals based on the generalized gradient approximation: Local, global hybrid, and range-separated hybrid functionals with and without dispersion corrections

Journal

The Journal of Chemical Physics, 140(18)

ISSN

0021-9606

Authors

Mardirossian, Narbe
Head-Gordon, Martin

Publication Date

2014-05-14

DOI

10.1063/1.4868117

Peer reviewed

Exploring the Limit of Accuracy for Density Functionals Based on the Generalized Gradient Approximation: Local, Global Hybrid, and Range-Separated Hybrid Functionals with and without Dispersion Corrections

Narbe Mardirossian and Martin Head-Gordon*

Department of Chemistry, University of California, Berkeley and Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720 USA

E-mail: mhg@cchem.berkeley.edu

Abstract

The limit of accuracy for semi-empirical generalized gradient approximation (GGA) density functionals is explored by parameterizing a variety of local, global hybrid (GH), and range-separated hybrid (RSH) functionals. The training methodology employed differs from conventional approaches in 2 main ways: 1). Instead of uniformly truncating the exchange, same-spin correlation, and opposite-spin correlation functional inhomogeneity correction factors, all possible fits up to fourth order are considered, and 2). Instead of selecting the optimal functionals based solely on their training set performance, the fits are validated on an independent test set and ranked based on their overall performance on the training *and* test sets. The 3 different methods of accounting for exchange are trained both with and without dispersion corrections (DFT-D2 and VV10), resulting in a total of 491508 candidate functionals. For each of the 9 functional classes considered, the results illustrate the trade-off between improved training set performance and diminished transferability. Since all 491508 functionals are uniformly trained and tested, this methodology allows the relative strengths of each type of functional to be consistently compared and contrasted. The range-separated hybrid GGA functional paired with the VV10 nonlocal correlation functional emerges as the most accurate form for the present training and test sets, which span thermochemical energy differences, reaction barriers, and intermolecular interactions involving lighter main group elements.

1 Introduction

While empirical parameters have been used in density functionals since the 1950s, the first systematic optimization of a density functional was performed by Axel Becke in 1997.[1] However, this breakthrough would not have been possible without several significant developments that took place in the preceding decades. Firstly, Frank Herman's extension[2] of John Slater's Xα method[3] (Equations 1 and 2) to the X$\alpha\beta$ method (Equation 3) introduced a gradient-based correction, $s_\sigma = \frac{|\nabla\rho_\sigma|}{\rho_\sigma^{4/3}}$, to the X$\alpha$ exchange energy density based on dimensional arguments. A major drawback of the semi-empirical X$\alpha\beta$ method was the divergence of its exchange potential at the $r = 0$ and $r = \infty$ limits. A solution to this was proposed by Becke in 1986, when he modified the inhomogeneity correction factor introduced by Herman in order to produce the divergence free X$\alpha\beta\gamma$ (B86) exchange functional[4] (Equation 4).

*To whom correspondence should be addressed

$$E_{x\alpha} = -\sum_{\sigma}^{\uparrow,\downarrow} \int C_{x\alpha} \rho_{\sigma}^{4/3} d\mathbf{r} \tag{1}$$

$$C_{x\alpha} = \frac{9}{4}\alpha \left(\frac{3}{4\pi}\right)^{1/3} \tag{2}$$

$$E_{x\alpha\beta} = -\sum_{\sigma}^{\uparrow,\downarrow} \int C_{x\alpha} \rho_{\sigma}^{4/3} \left[1 + \frac{\beta}{C_{x\alpha}} s_{\sigma}^2\right] d\mathbf{r} \tag{3}$$

$$E_{x\alpha\beta\gamma} = -\sum_{\sigma}^{\uparrow,\downarrow} \int C_{x\alpha} \rho_{\sigma}^{4/3} \left[1 + \frac{\beta}{C_{x\alpha}\gamma_x} \frac{\gamma_x s_{\sigma}^2}{1+\gamma_x s_{\sigma}^2}\right] d\mathbf{r} \tag{4}$$

11 years later,[1] Becke generalized the inhomogeneity correction factor of the B86 exchange functional with an m^{th} order power series in the dimensionless variable, $u_{x,\sigma} = \frac{\gamma_x s_{\sigma}^2}{1+\gamma_x s_{\sigma}^2}$:

$$E_x^{B97} = -\sum_{\sigma}^{\uparrow,\downarrow} \int C_{x\alpha} \rho_{\sigma}^{4/3} \left[\sum_{i=0}^{m} c_{x,i} u_{x,\sigma}^i\right] d\mathbf{r} \tag{5}$$

This scheme was applied to both the exchange functional and the spin-decomposed same-spin and opposite-spin correlation functionals to produce the B97 density functional (Section 3). The original B97 functional truncated the expansions at $m = 2$, and included a fraction of exact exchange, leaving 10 undetermined linear parameters for fitting to thermochemical data.

As an approach to GGA density functionals, B97 has unparalleled flexibility. As a result, it is not surprising that at least 15 B97-based density functionals have been parameterized since 1997. These include local functionals[5–8] (HCTH/93, HCTH/120, HCTH/147, HCTH/407, B97-D), global hybrid functionals[1,5,9,10] (B97, B97-1, B97-2, B97-3), range-separated hybrid functionals[11–14] (ωB97, ωB97X, ωB97X-D, ωB97X-D3, ωB97X-V), and even double hybrid functionals[15] (ωB97X-2).

The purpose of this work is to use the flexibility of the B97 form to attempt to systematically explore the accuracy attainable with different possible GGA functionals that build upon the basic B97 framework with different augmentations to exchange and correlation. Table 1 lists a variety of ingredients that can be incorporated into a B97-based density functional. To adhere to the functional form of the local component of B97, it is necessary to restrict the local exchange and correlation functionals to depend solely on the density and its gradient. However, the options for nonlocal exchange range from global hybrid exchange to range-separated exchange to no nonlocal exchange at all. These 3 options can be seamlessly integrated into the B97 functional form. From the perspective of dispersion corrections, options[8,16–20] such as DFT-D2, DFT-D3, vdW-DF-04, vdW-DF-10, VV09, VV10, MP2, RPA, and beyond, exist. All of these methods can be easily appended to the B97 functional form as well.

Table 1: Ingredients that can be incorporated into a density functional. GH stands for global hybrid and RSH stands for range-separated hybrid. DFT-D2 refers to Grimme's dispersion tail and VV10 refers to the VV10 nonlocal correlation (NLC) functional. The underlined ingredients were not varied, while the ingredients in bold were varied, resulting in a total of 9 candidate functional forms. While the kinetic energy density, τ, is a valid candidate for inclusion in the local parts of both the exchange and correlation functionals, this paper focuses exclusively on GGA functionals.

Exchange		Correlation	
Local	Nonlocal	Local	Nonlocal
1). $\underline{\rho}$	1). **None**	1). $\underline{\rho}$	1). **None**
2). $\underline{\nabla\rho}$	2). **GH**	2). $\underline{\nabla\rho}$	2). **DFT-D2**
3). $\underline{\tau}$	3). **RSH**	3). $\underline{\tau}$	3). **VV10**

B97-based semi-empirical density functionals have typically been optimized using uniformly truncated inhomogeneity correction factors (ICF) for the exchange, same-spin correlation, and opposite-spin correlation functionals. One method of approaching the limit of accuracy for GGA-based functionals is to try uniform expansions between $m = 0$ and a large m-value in order to select the optimal m-value based on a "goodness-of-fit" index[21] that is related to the training set performance. This approach can differentiate between uniformly truncated ICFs, and whether by this approach, or by careful inspection, B97-based ICFs are usually truncated at either $m = 2$, $m = 3$, or $m = 4$. One functional that is not based on uniform truncation is ωB97X-V,[14] which was developed based on a variation of the following methodology.

In contrast to uniform truncation, the most gen-

eral approach is to perform all possible optimizations up to a certain power of m, including even those that skip powers (equivalent to setting the skipped coefficient to zero). This approach leads to thousands of competing fits (i.e. thousands of competing functional forms). It is difficult to differentiate between so many possible functionals using any inspection of training set results, including the "goodness-of-fit" index. Yet, it will be essential to face this complexity since it is likely that the simplest functional capable of yielding good accuracy on the training set data will perform best in applications.

While the goal of fitting to a training set is to minimize the training set RMSD, it is even more desirable for a parameterized functional to be transferable. In order to differentiate between the thousands of resulting functionals *and* assure transferability, it is essential to take into account the performance of a given fit on both the training set *and* an independent test set. The test set is not used to determine any parameters, but will instead guide the choice of how many (and which) coefficients should be included in the least-squares fit. Taking the conventional approach of solely considering training set performance, it is guaranteed that the fit with the most linear parameters will have the smallest training set RMSD. Thus, if the training set RMSD is plotted with respect to the number of linear parameters, the resulting figure resembles the plots contained in Figure 1. However, if both the training set performance *and* the test set performance are taken into account, the plots begin to resemble parabolas (Figure 3). Thus, it is much easier to pick out an "optimal" functional with this methodology.

In this work, we parameterize 9 flavors of B97-based density functionals by varying the nonlocal exchange and dispersion correction (nonlocal correlation) components in bold in Table 1. While 14 of the 15 aforementioned B97-based density functionals have uniformly truncated inhomogeneity correction factors, all possible combinations of the exchange, same-spin correlation, and opposite-spin correlation expansion coefficients up to fourth order are tested. Using this methodology, an optimal functional from each category is selected, and the 9 resulting optimal functionals are compared to determine the optimal pairing of nonlocal exchange and dispersion.

2 Computational Details

An integration grid of 99 radial points and 590 angular points, (99,590), was used to evaluate local exchange-correlation (xc) functionals, while the SG-1 grid[22] was used for the VV10 nonlocal correlation (NLC) functional.[20] For the rare-gas dimers and the absolute atomic energies, a (500,974) integration grid was used to evaluate local xc functionals, along with a (99,590) grid for the VV10 NLC functional. The aug-cc-pVQZ [aQZ] basis set[23,24] was used for all thermochemistry (TC) datapoints except the second-row absolute atomic energies (aug-cc-pCVQZ),[23,24] while the aug-cc-pVTZ [aTZ] basis set[23,24] was used for all noncovalent interactions (NC) datapoints except the rare-gas dimers (aug-cc-pVQZ). Furthermore, the noncovalent interactions were computed *without* counterpoise corrections. For B97-D2, Grimme's DFT-D2 dispersion tail was used with an s_6 coefficient[25] of 0.75. Grimme's B97-D functional[8] uses the DFT-D2 dispersion tail as well, with an s_6 coefficient of 1.25. All of the calculations were performed with a development version of Q-Chem 4.0.[26]

3 Theory

The complete functional form for all of the trained functionals is given by Equations 6-8. The components of the exchange functional and correlation functional are described in Sections 3.1 and 3.2, respectively. The acronyms used in Equations 6-8 (and henceforth) are: exchange-correlation (xc), exchange (x), correlation (c), short-range (sr), long-range (lr), same-spin (ss), opposite-spin (os), and dispersion (disp).

$$E_{xc} = E_x + E_c \tag{6}$$

$$E_x = E_x^{B97} + c_x E_{x,sr}^{exact} + d_x E_{x,lr}^{exact} \tag{7}$$

$$E_c = E_{c,ss}^{B97} + E_{c,os}^{B97} + E_{disp} \tag{8}$$

For local (exchange) functionals, $c_x = d_x = 0$, while for global hybrid functionals, $c_x = d_x$, where c_x is the (global) fraction of exact exchange. For range-separated hybrid functionals, $d_x = 1$, $E_x^{B97} =$

$E_{x,sr}^{B97}$ (Section 3.1), and c_x is the fraction of short-range exact exchange.

3.1 Exchange Functional Form

The local exchange component of the B97 functional form is given by Equations 9 and 10:

$$E_x^{B97} = \sum_{\sigma}^{\alpha,\beta} \int e_{x,\sigma}^{UEG}(\rho_\sigma) g_x(u_{x,\sigma})\, d\mathbf{r} \tag{9}$$

$$g_x(u_{x,\sigma}) = \sum_{i=0}^{m_x} c_{x,i} u_{x,\sigma}^i = \sum_{i=0}^{m_x} c_{x,i} \left[\frac{\gamma_x s_\sigma^2}{1+\gamma_x s_\sigma^2}\right]^i \tag{10}$$

where the dimensionless variable, $u_{x,\sigma} \in [0,1]$, is a finite domain transformation of the reduced spin-density gradient, $s_\sigma = \frac{|\nabla\rho_\sigma|}{\rho_\sigma^{4/3}} \in [0,\infty)$. In Equation 9, $e_{x,\sigma}^{UEG}(\rho_\sigma)$ is the exchange energy density per unit volume of a uniform electron gas (UEG) and $g_x(u_{x,\sigma})$ is the exchange functional inhomogeneity correction factor (ICF). The linear local exchange parameters, $c_{x,i}$, will be determined by least-squares fitting to a training set in Section 5, while $\gamma_x = 0.004$ is a nonlinear local exchange parameter that was fit to the Hartree–Fock exchange energies of 20 atoms in 1986 by Becke.[4] For range-separated hybrid functionals, the conventional Coulomb operator in the local exchange component is attenuated by the complementary error function (erfc), resulting in an additional multiplicative factor, $F(a_\sigma)$, in the integrand of Equation 9:

$$\begin{aligned} F(a_\sigma) = 1 - \frac{2}{3} a_\sigma \Big[2\sqrt{\pi}\,\mathrm{erf}\left(\frac{1}{a_\sigma}\right) - 3a_\sigma + \\ a_\sigma^3 + \left[2a_\sigma - a_\sigma^3\right] \exp\left(-\frac{1}{a_\sigma^2}\right) \Big] \end{aligned} \tag{11}$$

where $a_\sigma = \frac{\omega}{k_{F\sigma}}$, ω is the nonlinear range-separation parameter that controls the transition from local exchange to nonlocal exact exchange with respect to interelectronic distance, and $k_{F\sigma} = \left[6\pi^2\rho_\sigma\right]^{1/3}$ is the spin-polarized Fermi wave vector. The inclusion of $F(a_\sigma)$ in the integrand of Equation 9 gives $E_{x,sr}^{B97}$.

When considering both global hybrid and range-separated hybrid functionals, the most general way to deal with nonlocal exchange is to split the Coulomb operator in the conventional expression for exact exchange into a short-range component ($E_{x,sr}^{exact}$) and a long-range component ($E_{x,lr}^{exact}$) with the erfc and erf Coulomb functions, respectively:

$$\begin{aligned} E_{x,sr}^{exact} = -\frac{1}{2} \sum_{\sigma}^{\alpha,\beta} \sum_{i,j}^{occ.} \int\int \psi_{i\sigma}^*(\mathbf{r}_1)\, \psi_{j\sigma}^*(\mathbf{r}_2)\, \frac{\mathrm{erfc}(\omega r_{12})}{r_{12}} \\ \times \psi_{j\sigma}(\mathbf{r}_1)\, \psi_{i\sigma}(\mathbf{r}_2)\, d\mathbf{r}_1 d\mathbf{r}_2 \end{aligned} \tag{12}$$

$$\begin{aligned} E_{x,lr}^{exact} = -\frac{1}{2} \sum_{\sigma}^{\alpha,\beta} \sum_{i,j}^{occ.} \int\int \psi_{i\sigma}^*(\mathbf{r}_1)\, \psi_{j\sigma}^*(\mathbf{r}_2)\, \frac{\mathrm{erf}(\omega r_{12})}{r_{12}} \\ \times \psi_{j\sigma}(\mathbf{r}_1)\, \psi_{i\sigma}(\mathbf{r}_2)\, d\mathbf{r}_1 d\mathbf{r}_2 \end{aligned} \tag{13}$$

where $\psi_{i\sigma}$ and $\psi_{j\sigma}$ are the occupied Kohn–Sham spatial orbitals. Since $\frac{\mathrm{erfc}(\omega r_{12})}{r_{12}} + \frac{\mathrm{erf}(\omega r_{12})}{r_{12}} = \frac{1}{r_{12}}$, $E_x^{exact} = E_{x,sr}^{exact} + E_{x,lr}^{exact}$ for global hybrids, where $c_x = d_x$ is the fraction of (global) exact exchange. For range-separated hybrids, instead of setting the percentage of exact-exchange at $r = 0$ to zero, an (optional) optimizable parameter, c_x, controls the amount of short-range exact exchange. Additionally, the value of ω is fixed at 0.3 for all of the range-separated hybrid functionals.

3.2 Correlation Functional Form

The local correlation component of the B97 functional form is given by Equations 14-17:

$$E_{c,ss}^{B97} = \sum_{\sigma}^{\alpha,\beta} \int e_{c,\sigma\sigma}^{PW92} g_{c,ss}(u_{c,\sigma\sigma})\, d\mathbf{r} \tag{14}$$

$$g_{c,ss}(u_{c,\sigma\sigma}) = \sum_{i=0}^{m_{css}} c_{css,i} u_{c,\sigma\sigma}^i = \sum_{i=0}^{m_{css}} c_{css,i} \left[\frac{\gamma_{css} s_\sigma^2}{1+\gamma_{css} s_\sigma^2}\right]^i \tag{15}$$

$$E_{c,os}^{B97} = \int e_{c,\alpha\beta}^{PW92} g_{c,os}(u_{c,\alpha\beta})\, d\mathbf{r} \tag{16}$$

$$g_{c,os}(u_{c,\alpha\beta}) = \sum_{i=0}^{m_{cos}} c_{cos,i} u_{c,\alpha\beta}^i = \sum_{i=0}^{m_{cos}} c_{cos,i} \left[\frac{\gamma_{cos} s_{\alpha\beta}^2}{1+\gamma_{cos} s_{\alpha\beta}^2}\right]^i \tag{17}$$

where $s_{\alpha\beta}^2 = \frac{1}{2}\left(s_\alpha^2 + s_\beta^2\right)$, and $e_{c,\sigma\sigma}^{PW92}$ and $e_{c,\alpha\beta}^{PW92}$ are the PW92[27] same-spin and opposite-spin correlation energy densities per unit volume.[28] The linear local correlation parameters, $c_{css,i}$ and $c_{cos,i}$, will be determined by least-squares fitting to a training set in Section 5, while $\gamma_{css} = 0.2$ and $\gamma_{cos} = 0.006$ are nonlinear local correlation parameters that were fit to the correlation energies of helium and neon in 1997 by Becke.[1]

Since functionals are trained both with and without dispersion corrections, the E_{disp} term requires

further elaboration. When dispersion corrections are not used, $E_{disp} = 0$. Two different dispersion correction methods are used in combination with the local, GH, and RSH functionals: one dispersion tail (DT) and one nonlocal correlation (NLC) functional.

The dispersion tail (DFT-D2) has the following form:

$$E_{disp}^{DFT-D2} = -s_6 \sum_{i=1}^{N_{at}-1} \sum_{j=i+1}^{N_{at}} \frac{C_6^{ij}}{R_{ij}^6} f_{damp}^{DFT-D2}(R_{ij}) \tag{18}$$

$$f_{damp}^{DFT-D2}(R_{ij}) = \frac{1}{1+e^{-d\left(R_{ij}/R_{r,ij}-1\right)}} \tag{19}$$

In the damping function, $R_{r,ij} = R_{0,i} + R_{0,j}$ is the sum of the van der Waals (vdW) radii of a pair of atoms, $C_6^{ij} = \sqrt{C_6^i C_6^j}$ is the dispersion coefficient of a pair of atoms, and $d = 20$. In training the DFT-D2 dispersion tail onto the density functionals, the linear s_6 coefficient is optimized and counted as a linear parameter. The empirical C_6 parameters and vdW Radii, R_0, can be found in Table 1 of Reference 8.

The nonlocal correlation functional that is considered is VV10:[20]

$$E_{disp}^{VV10} = \int \rho(\mathbf{r}) \left[\frac{1}{32} \left[\frac{3}{b^2} \right]^{3/4} + \frac{1}{2} \int \rho(\mathbf{r}') \Phi(\mathbf{r}, \mathbf{r}', \{b, C\}) \, d\mathbf{r}' \right] d\mathbf{r} \tag{20}$$

where $\Phi(\mathbf{r}, \mathbf{r}', \{b, C\})$ is the nonlocal correlation kernel defined in Reference 20. The VV10 NLC functional introduces 2 nonlinear parameters: b, which controls the short-range damping of the $1/r^6$ asymptote, and C, which controls the accuracy of the asymptotic C_6 coefficients. Since it is much more difficult to train the nonlinear parameters of the VV10 NLC functional, the parameters that were optimized for ωB97X-V ($b = 6$ and $C = 0.01$) are used here without further optimization.

4 Datasets

In total, the training and test sets used for the parameterization and validation of the candidate functionals contain 2301 datapoints, requiring 1961 single-point calculations. Of the 2301 datapoints, 1108 belong to the training set and 1193 belong to the test set. Furthermore, the training and test sets contain both thermochemistry (TC) data as well as noncovalent interactions (NC) data. The training set contains 787 TC datapoints and 321 NC datapoints, while the test set contains 146 TC datapoints and 1047 NC datapoints (for an overall total of 933 TC datapoints and 1368 NC datapoints). The partitioning of the training and test sets was carried out with the quality of the reference values in mind, such that the training set contains the highest confidence data. Table 2 lists the 36 datasets that form the training and test sets. The references for the datasets are given in the rightmost column of Table 2, while additional details can be found in Reference 14. In addition to the general division into TC and NC data for the training and test sets, we will report the results for the 3 rare-gas (RG) potential energy curves separately, as a delicate diagnostic of the balance between Pauli repulsion and attractive dispersion interactions.

5 Training Methodology

In order to train and test the candidate functionals, single-point calculations are carried out with the unoptimized functionals ($g_x = g_{c,ss} = g_{c,os} = 1$) for the 1961 geometries that correspond to the 2301 datapoints in the training and test sets. In order to gather all of the data that will be used for the upcoming analysis, 4 sets of calculations must be carried out: 1). LSDA without VV10, 2). LSDA with VV10, 3). SR-LSDA without VV10, and 4). SR-LSDA with VV10. The VV10 calculations must be carried out separately because the VV10 NLC functional is implemented within the SCF procedure. Conveniently, however, running the LSDA (local spin density approximation) functional is sufficient for gathering data for both the local and global hybrid variants, since a global hybrid functional with an initial guess of $c_x = 0$ is a local functional. Following the single-point calculations, the resulting 4 sets of densities are saved to disk and used to calculate the values that the expansion coefficients in the power series ($c_{x,i}$, $c_{css,i}$, and $c_{cos,i}$) multiply, up to fourth order ($i \in [0,4]$). In addition to these 15 contributions, the value of

Table 2: Summary of the datasets found in the training and test sets. The datasets above the thick black line are in the training set and the datasets below the thick black line are in the test set. Within the training and test sets, datasets above the thin black line contain thermochemistry datapoints, while datasets below the thin black line contain noncovalent interactions datapoints. PEC stands for potential energy curve.

Name	#	Description	Ref.
HAT707	505	Heavy-atom transfer reaction energies	29
BDE99	83	Bond dissociation reaction energies	29
TAE_nonMR124	124	Total atomization energies	29
SN13	13	Nucleophilic substitution reaction energies	29
ISOMER20	18	Isomerization reaction energies	29
DBH24	24	Diverse barrier heights	30,31
EA6	6	Electron affinities of atoms	32
IP6	6	Ionization potentials of atoms	32
AE8	8	Absolute atomic energies	33
SW49Rel345	28	$SO_4{}^{2-}(H_2O)_n$ ($n = 3-5$) relative energies	34
SW49Bind345	30	$SO_4{}^{2-}(H_2O)_n$ ($n = 3-5$) binding energies	34
NBC10A2	37	Methane dimer and benzene-methane dimer PECs	35,36
HBC6A	118	Formic acid, formamide acid, and formamidine acid dimer PECs	36,37
BzDC215	108	Benzene and first- and second-row hydride PECs	38
EA7	7	Electron affinities of small molecules	32
IP7	7	Ionization potentials of small molecules	32
Gill12	12	Neutral, radical, anionic, and cationic isodesmic reaction energies	39
AlkAtom19	19	$n = 1-8$ alkane atomization energies	40
AlkIsomer11	11	$n = 4-8$ alkane isomerization energies	40
AlkIsod14	14	$n = 3-8$ alkane isodesmic reaction energies	40
HTBH38	38	Hydrogen transfer barrier heights	41
NHTBH38	38	Non-hydrogen transfer barrier heights	42
SW49Rel6	17	$SO_4{}^{2-}(H_2O)_n$ ($n = 6$) relative energies	34
SW49Bind6	18	$SO_4{}^{2-}(H_2O)_n$ ($n = 6$) binding energies	34
NNTT41	41	Neon-neon PEC	43
AATT41	41	Argon-argon PEC	43
NATT41	41	Neon-argon PEC	43
NBC10A1	53	Parallel-displaced (3.4 Å), sandwich, and T-shaped benzene dimer PECs	35,36
NBC10A3	39	S2 and T3 configuration pyridine dimer PECs	36,44
WATER27	23	Neutral and charged water interactions	45,46
HW30	30	Hydrocarbon and water dimers	47
NCCE31	18	Noncovalent complexation energies	48
S22x5	110	Hydrogen-bonded and dispersion-bonded complex PECs	49
S66x8	528	Biomolecular structure complex PECs	50
S22	22	Equilibrium geometries from S22x5	36,51
S66	66	Equilibrium geometries from S66x8	50,52

E_x^{exact} is required for GH functionals and the value of $E_{x,sr}^{exact}$ is required for RSH functionals.

The calculated values are used to form a (# of Datapoints) x (# of Linear Parameters) matrix, A, that contains the appropriate contributions for a given datapoint. In addition to the A matrix, a column of values corresponding to the errors in the unoptimized functional ($y = E_{REF} - E_{DFT}$) is computed. Since weights are used during training, a diagonal (# of Datapoints) x (# of Datapoints) training set weight matrix (W_{Train}) is required as well. The diagonal elements corresponding to the training set data contain the appropriate weights, while the remaining diagonal elements corresponding to the test set data are set to zero. The change in the linear parameters, Δb, is found by a weighted least-squares fit:

$$\Delta b = (A^T W_{Train} A)^{-1} (A^T W_{Train} y) \tag{21}$$

and the training set RMSD is calculated by:

$$RMSD_{Train} = \sqrt{\frac{\mathrm{diag}\,(W_{Train}) \cdot (y - A\Delta b)^2}{\#_{Train}}} \tag{22}$$

Additional statistical measures are calculated using Equation 22 with the appropriate weight matrix and #.

In total, 10 quantities will be used to gauge the performance of the resulting functionals: the training set RMSD (RMSD_{Train}), the test set RMSD (RMSD_{Test}), the RMSD for the 3 rare-gas dimer PECs (RMSD_{RG}), the total RMSD (RMSD_{Total}), the thermochemistry (TC) RMSD (RMSD_{TC}), the noncovalent interactions (NC) RMSD (RMSD_{NC}), the training set TC RMSD ($\mathrm{RMSD}_{TC,Train}$), the test set TC RMSD ($\mathrm{RMSD}_{TC,Test}$), the training set NC RMSD ($\mathrm{RMSD}_{NC,Train}$), and the test set NC RMSD ($\mathrm{RMSD}_{NC,Test}$).

Since contributions are computed up to fourth order for the exchange, same-spin correlation, and opposite-spin correlation functionals, as many as 15 linear GGA parameters are available for optimization. The optional short-range exchange parameter that is unique to range-separated hybrid functionals adds a 16th parameter for the RSHs. The uniform electron gas (UEG) constraint for the same-spin and opposite-spin correlation functionals can be incorporated by making $c_{css,0}$ and $c_{cos,0}$ optional parameters, but the same cannot be done with the UEG constraint for exchange (except for local functionals). Thus, fits that violate the UEG limit for exchange are optimized separately from fits that do incorporate the UEG limit for exchange. As an example of the number of fits that result from this methodology, local functionals that are constructed to satisfy the UEG constraint for exchange have 14 optional parameters, giving a total of $\sum_{i=1}^{14} \binom{14}{i} = 2^{14} - 1 = 16383$ possible fits. Table 3 lists the total number of fits for local, GH, and RSH functionals with and without the UEG limit for exchange in place.

Table 3: Total number of least-squares fits (#) that can be performed when considering parameters up to fourth order in the power series inhomogeneity correction factors. While the type of dispersion correction used has no bearing on the total number of possible fits, whether or not the UEG constraint for exchange is enforced is important and is addressed in the second and third columns, respectively.

#	$c_{x,0} + c_x = 1$	$c_{x,0} + c_x \neq 1$
Local	$2^{14} - 1 = 16383$	$2^{15} - 2^{14} = 16384$
GH	$2^{14} - 1 = 16383$	$2^{15} - 1 = 32767$
RSH	$2^{15} - 1 = 32767$	$2^{16} - 2^{14} = 49152$

In order to refer to the thousands of resulting functionals with clarity, we will use a nomenclature that is fully specified in Table 4. As examples, "GN-012.012.012.Xn" would describe Becke's 10-parameter B97 functional, "LD-012.012.012.0n" would describe Grimme's 10-parameter B97-D functional, "RN-1234.1234.1234.Xy" would describe the 13-parameter ωB97X functional, and "RV-12.01.01.Xy" would describe the 7-parameter ωB97X-V functional. As can be seen with the descriptor for ωB97X-V, since the UEG limit for exchange was used as a constraint, "0" does not appear in the label for the exchange functional ICF (even though $c_{x,0} \neq 1$), because the 4th label, "Xy", implies that $c_{x,0} = 1 - c_x$. In addition, nonlinear parameters are not counted when considering the number of parameters corresponding to a given fit , since the nonlinear parameters were not varied in this work. Henceforth, any mention of the number of parameters implicitly refers to

the number of linear parameters. As a more complicated example, if a Local+DFT-D2 functional requires the optimization of $\{c_{x,1}, c_{x,3}, c_{x,4}, c_{cos,0}, c_{cos,2}\}$, "LD-134.∅.02.0y" will be used as its descriptor. Henceforth, quotations will not be used for the descriptors.

As far as weights are concerned, thermochemistry datapoints in the training and test sets are given weights of 1 and 2.5 respectively (except for datapoints in EA6 and IP6 which are weighted by 5), noncovalent interactions datapoints in the training and test sets are given weights of 10 and 25, respectively, and datapoints corresponding to the rare-gas dimer PECs in the test set are given weights of 2500. Even though the rare-gas (RG) dimer PECs are technically in the test set, they are not included in the calculation of RMSD_{Test}. However, they are included in the calculation of RMSD_{Total}. The rare-gas dimer PECs are included in the NC and test set NC RMSDs because their unweighted contributions are very small and do not contribute significantly. Of the 10 RMSDs, only the first 4 are weighted, while the latter 6 are unweighted.

6 Training Results

It is important to point out that the selection procedure utilized to identify the optimal functionals is not (and cannot be) unique. However, as we shall see, it recovers the self-consistently optimized ωB97X-V functional, even though a slightly different selection procedure was used in Reference 14. In addition, the resulting optimal functionals are usually significantly better than existing functionals of the same type, as will be discussed in Section 7.

While a variety of selection procedures were initially explored, the one that was finally chosen is quite simple. First, the total (weighted) RMSDs are computed and plotted. Next, a screening process rejects fits that predict rare-gas dimer equilibrium bond lengths that are too long or too short by more than 0.1 Å. Since the plots are still overflowing with data points, all of the points for a fixed number of linear parameters are removed, except for the point that corresponds to the lowest total RMSD with and without the UEG constraint for exchange. The resulting plots (Figure 3) are much simpler to analyze and contain filled circles (satisfy the UEG constraint for exchange) and unfilled circles (do not satisfy the UEG constraint for exchange).

Starting at the fewest number of linear parameters, an additional empirical parameter is accepted if the improvement in the total RMSD is more than 0.05 kcal/mol. This final stage does not take into account whether or not the UEG constraint for exchange is enforced. The 9 optimal functionals are chosen in this manner and will be discussed and compared to existing functionals in Section 7. Since the optimal functionals are chosen based on their total RMSDs, the corresponding training and test set RMSDs of the optimal functionals are shown in red in Figures 1 and 2.

Table 4: Explanation of the nomenclature for the descriptors that refer to the thousands of optimized functionals. A given descriptor takes on the following form: "ij-{p_x}.{p_{css}}.{p_{cos}}.kl". If none of the coefficients of a given ICF are optimized, ∅ is used as a placeholder. As an example, "GN-012.012.012.Xn" would describe Becke's 10-parameter B97 functional.

Symbol	Meaning	Allowed Values	Meaning
i	exchange	L	local
		G	global hybrid
		R	range-separated hybrid
j	dispersion correction	N	none
		D	DFT-D2 dispersion tail
		V	VV10 nonlocal correlation functional
$\{p_x\}$	linear exchange parameters	any subset of 01234	each included integer, m, is a single parameter multiplying u_x^m
$\{p_{css}\}$	linear same-spin correlation parameters	any subset of 01234	each included integer, m, is a single parameter multiplying $u_{c,\sigma\sigma}^m$
$\{p_{cos}\}$	linear opposite-spin correlation parameters	any subset of 01234	each included integer, m, is a single parameter multiplying $u_{c,\alpha\beta}^m$
k	(short-range) exact exchange	0	no (short-range) exact exchange included
		X	(short-range) exact exchange included
l	UEG for exchange	y	UEG limit for exchange is enforced
		n	UEG limit for exchange is not enforced

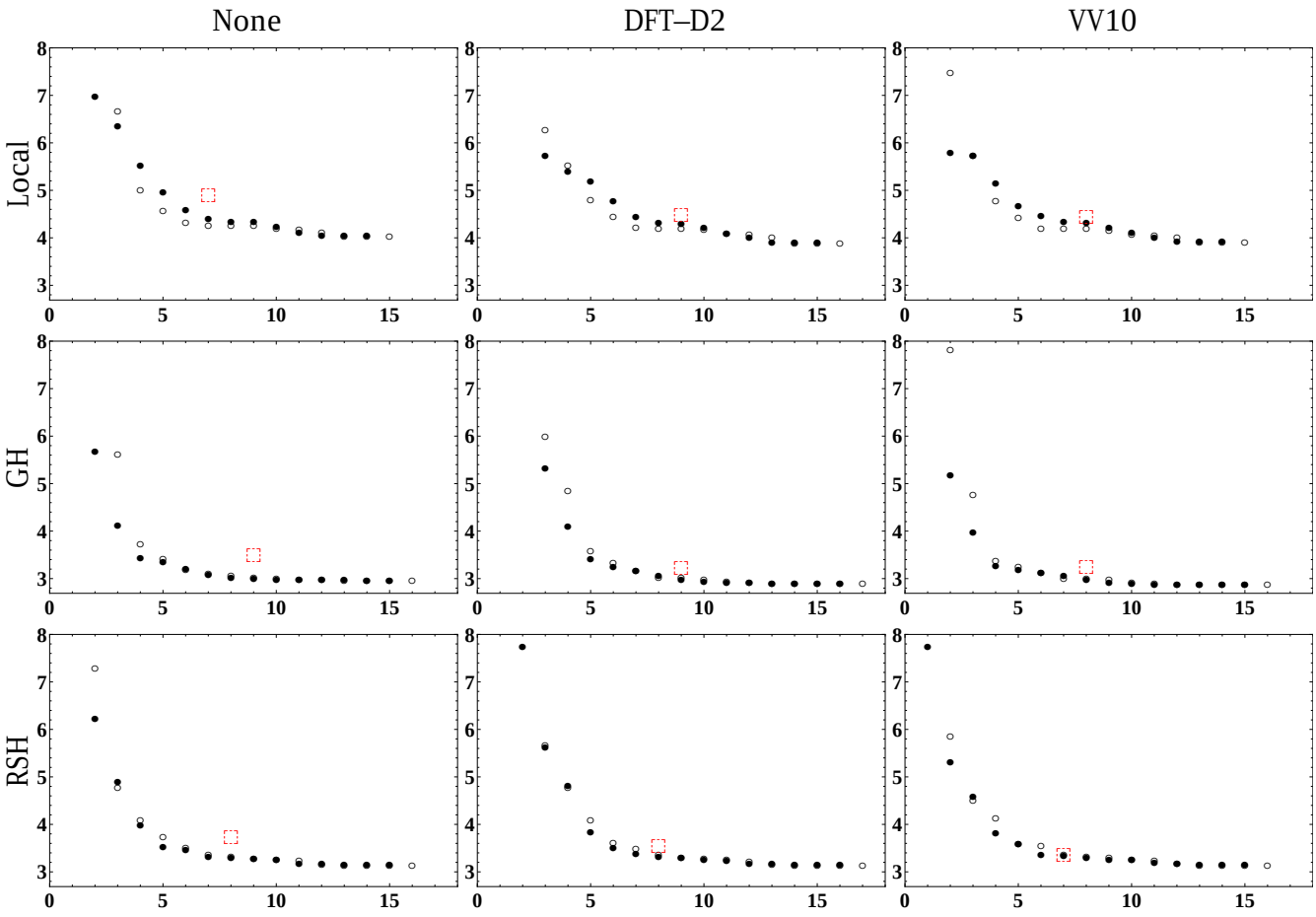

Figure 1: Plots showing the lowest training set RMSD (in kcal/mol) for a fixed number of linear parameters for all 9 candidate functional forms considered. Filled circles correspond to fits which satisfy the UEG limit for exchange and unfilled circles indicate that the UEG limit for exchange was allowed to relax. The red box indicates the training set RMSD of the optimal functional, which is usually not the best for the training set data alone.

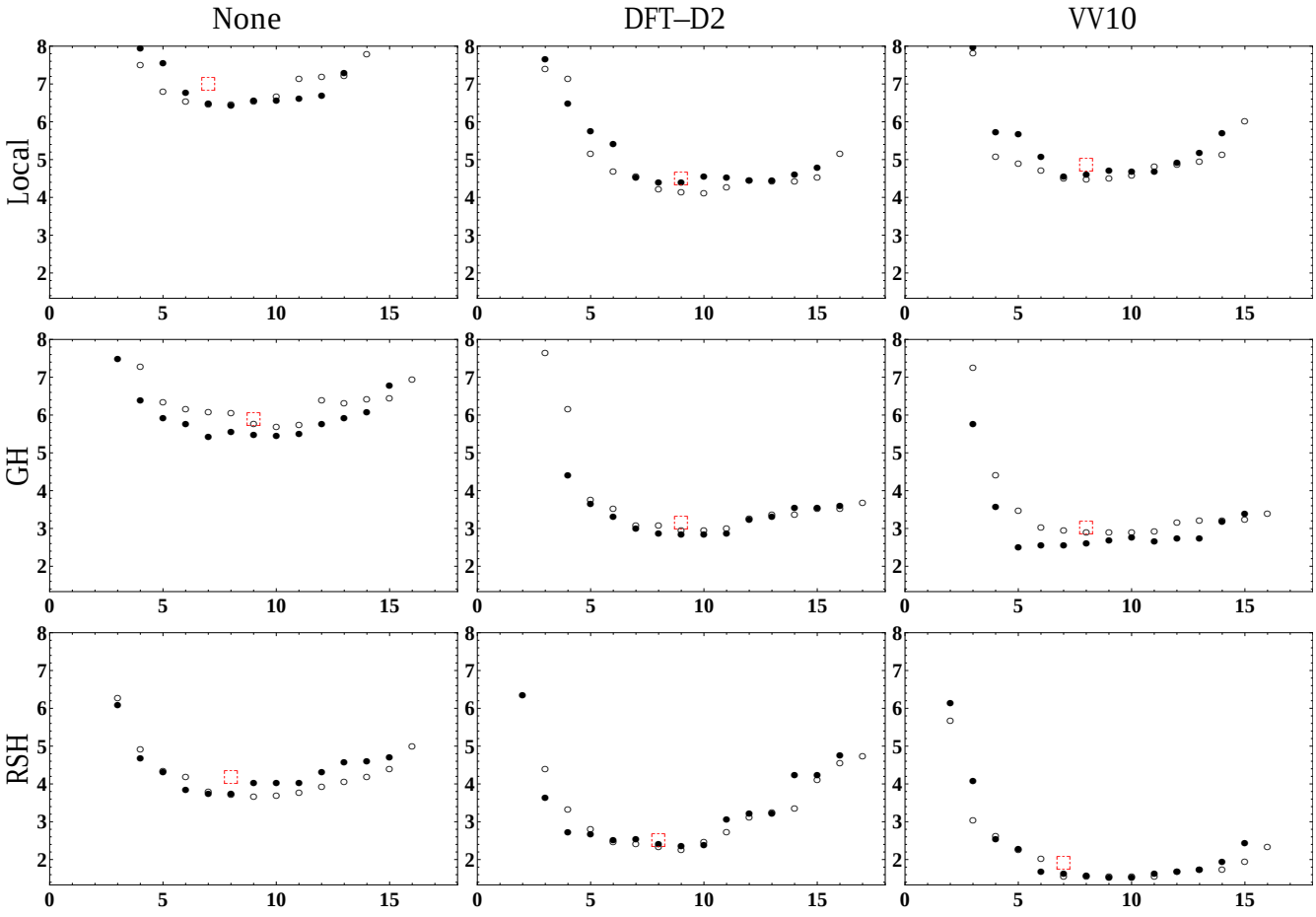

Figure 2: Plots showing the lowest test set RMSD (in kcal/mol) for a given number of linear parameters for all 9 candidate functional forms considered. Filled circles correspond to fits which satisfy the UEG limit for exchange and unfilled circles indicate that the UEG limit for exchange was allowed to relax. The red box indicates the test set RMSD of the optimal functional.

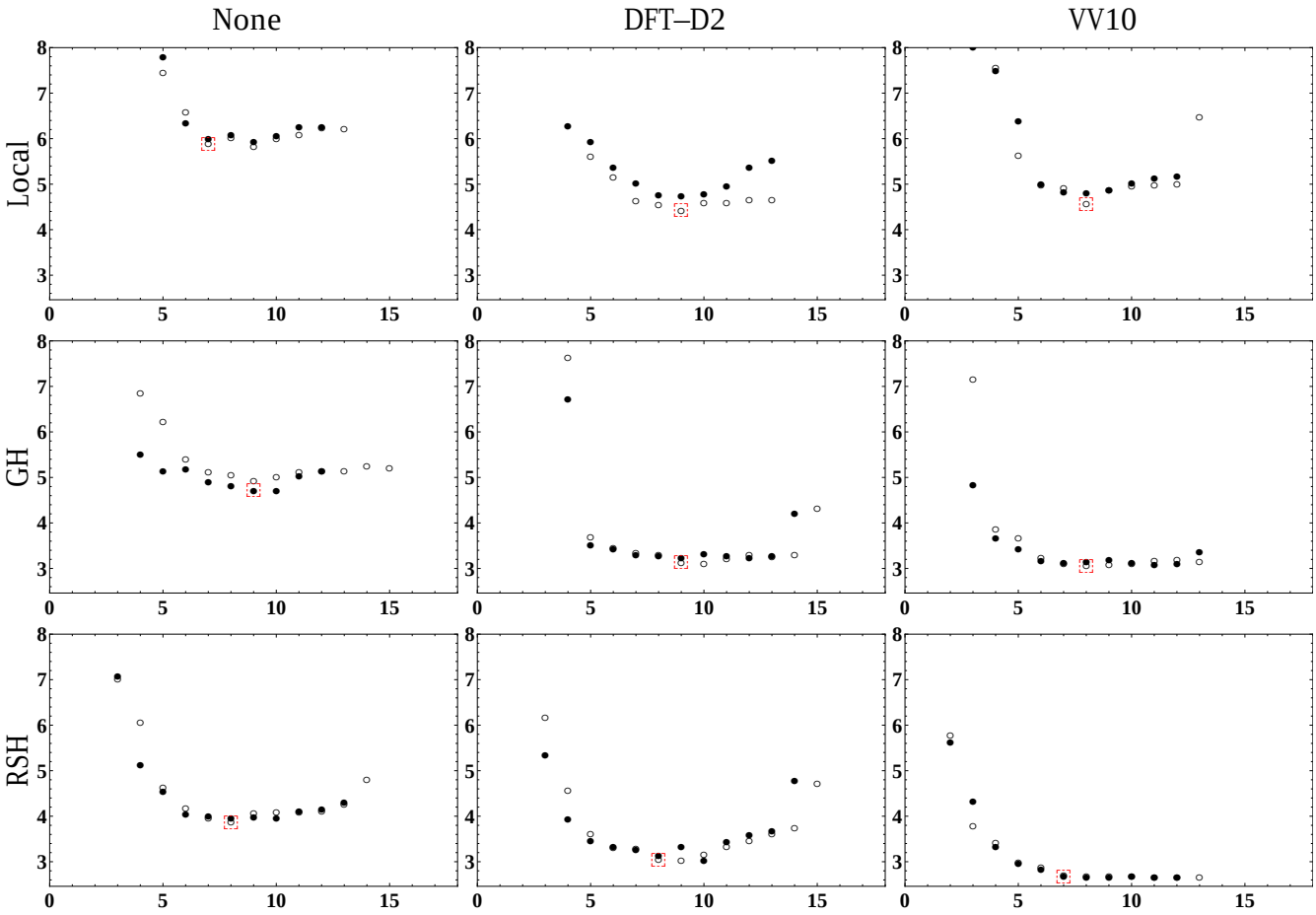

Figure 3: Plots showing the lowest total RMSD (in kcal/mol) for a given number of linear parameters for all 9 candidate functional forms considered. Filled circles correspond to fits which satisfy the UEG limit for exchange and unfilled circles indicate that the UEG limit for exchange was allowed to relax. The red box indicates the total RMSD of the optimal functional. Due to the screening process described in Section 6, points that correspond to fits that predict rare-gas dimer equilibrium bond lengths that are too long or too short by more than 0.1 Å have been removed.

All of the RMSDs considered in this section are generated using Equation 22 (with the appropriate weight matrix and #) and are least-squares fit RMSDs. While none of the functionals are self-consistently optimized, the recent self-consistent optimization of ωB97X-V indicated that the least-squares fit RMSDs generally differ from the actual RMSDs of the self-consistently optimized functional by 0.05 kcal/mol on average. While it would be impractical to self-consistently optimize thousands of functionals, we firmly believe that this procedure is effective in predicting the quality of a functional based on least-squares fit RMSDs.

Since the parameters that are obtained from all of these fits are not self-consistently optimized, it is not immediately obvious how much they will differ from the final set of parameters. Thus, it is difficult to comment on the usefulness of the parameters of the 9 resulting optimal functionals. However, the parameters for Becke's B97 functional were optimized in the same post-LSDA manner as all of the functionals considered in this paper, and comparing the parameters of B97 and B97-1, or alternatively considering Table 3 from Reference 14, indicates that the self-consistently optimized parameters do not differ drastically from those from the end of the first optimization cycle. While it is best to self-consistently optimize the parameters of a semi-empirical density functional, the parameters for the 9 optimal functionals are provided in Table 5.

Tables 6, 7, and 8 contain data for range-separated hybrid, global hybrid, and local functionals, respectively. Each method of accounting for exchange was trained both with dispersion corrections (DFT-D2 and VV10) and without dispersion corrections (None). For each pairing, the columns labeled "Minimum" contain the best possible value for a given RMSD category, while the columns labeled "Optimal" contain the results for the functionals that were selected from Figure 3. For the remainder of this section, the least-squares fit RMSDs will simply be referred to as RMSDs.

We begin the analysis with the RSH+VV10 category, since our newest density functional, ωB97X-V, belongs to this class. Generally speaking, the interesting comparisons in Table 6 are between the best possible result in a given row for *any* candidate RSH+VV10 functional (i.e. the Minimum

Table 6: RMSDs in kcal/mol for range-separated hybrid functionals. The data in the training and test sets consists of thermochemical (TC) and noncovalent (NC) energy differences. The rare-gas (RG) test results are reported separately. The "Minimum" columns contain the smallest possible RMSD value for the particular entry from *all* trained functionals of that class. Hence, each entry within a column generally corresponds to a *different* functional. The "Optimal" columns contain the RMSD value for the best overall functional selected from that class. Hence, each entry within a column corresponds to the *same* functional. Details regarding the optimal functional are provided in Table 5.

RSH	Minimum			Optimal		
kcal/mol	None	DFT-D2	VV10	None	DFT-D2	VV10
Train	3.14	3.14	3.14	3.73	3.55	3.36
Test	3.68	2.28	1.53	4.18	2.50	1.92
RG	0.88	0.80	0.51	1.97	2.50	0.95
Total	3.87	3.03	2.66	3.87	3.05	2.68
TC	3.49	3.47	3.44	4.09	3.80	3.62
NC	0.58	0.36	0.24	0.71	0.43	0.32
TC,Train	3.55	3.56	3.56	4.28	4.02	3.86
TC,Test	2.19	1.76	1.66	2.85	2.26	1.81
NC,Train	0.25	0.26	0.23	0.42	0.42	0.31
NC,Test	0.62	0.33	0.23	0.77	0.43	0.32

Table 7: RMSDs in kcal/mol for global hybrid functionals. The format is explained in the caption of Table 6.

GH	Minimum			Optimal		
kcal/mol	None	DFT-D2	VV10	None	DFT-D2	VV10
Train	2.97	2.90	2.89	3.49	3.22	3.25
Test	5.44	2.86	2.51	5.90	3.17	3.03
RG	0.76	0.67	0.31	2.01	1.82	0.65
Total	4.72	3.11	3.06	4.72	3.14	3.06
TC	3.45	3.31	3.31	3.87	3.48	3.61
NC	0.91	0.42	0.38	0.97	0.54	0.48
TC,Train	3.24	3.21	3.18	3.78	3.62	3.70
TC,Test	3.13	2.60	2.59	4.28	2.60	3.06
NC,Train	0.41	0.37	0.35	0.48	0.48	0.42
NC,Test	1.01	0.40	0.38	1.08	0.56	0.50

Table 5: Characteristics of the 9 optimal functionals. Within a cell, the first row lists the descriptor (Table 4 with the number of associated linear parameters in parentheses, the second row lists the non-self-consistently optimized GGA parameters, and the third row (when applicable) lists the values for the (short-range) exact exchange parameter, c_x, and the linear DFT-D2 dispersion coefficient, s_6.

	None	DFT-D2	VV10
Local	LN-012.0.0134.0n (7) {1.07, -0.94, 5.04, 0.45, 8.83, -65.31, 39.84} N/A	LD-012.0.0123.0n (9) {1.09, -0.89, 4.96, 0.25, 0.43, 12.06, -27.85, -1.57} s_6=0.71	LV-012.0.0123.0n (8) {1.09, -0.79, 4.74, 0.43, 0.41, 11.90, -27.19, -1.52} N/A
GH	GN-24.1234.13.Xy (9) {1.95, 2.70, -6.54, 33.21, -49.17, 22.23, 3.33, -24.59} c_x=0.23	GD-0234.0.04.Xn (9) {0.79, 2.46, -3.89, 4.73, 0.33, 1.10, -13.90} c_x=0.24; s_6=0.64	GV-02.0.0234.Xn (8) {0.81, 2.00, 0.51, 1.05, 9.19, -39.37, 22.86} c_x=0.22
RSH	RN-14.34.012.Xn (8) {0.58, 11.25, -9.17, 9.08, 1.09, 2.67, -10.32} c_x=0.02	RD-02.12.01.Xn (8) {0.87, 2.24, -3.62, 3.30, 1.35, -2.36} c_x=0.18; s_6=0.71	RV-12.01.01.Xy (7) {0.61, 1.18, 0.58, -0.27, 1.22, -1.87} c_x=0.16

Table 8: RMSDs in kcal/mol for local functionals. The format is explained in the caption of Table 6.

Local	Minimum			Optimal		
kcal/mol	None	DFT-D2	VV10	None	DFT-D2	VV10
Train	4.03	3.89	3.91	4.91	4.48	4.44
Test	6.46	4.13	4.49	7.00	4.50	4.88
RG	0.97	1.34	0.72	2.09	3.15	2.57
Total	5.84	4.43	4.57	5.89	4.43	4.57
TC	5.03	4.69	4.85	5.72	5.23	5.27
NC	0.95	0.51	0.54	1.08	0.61	0.65
TC,Train	4.41	4.30	4.32	5.55	5.13	5.11
TC,Test	5.83	4.87	5.50	6.57	5.71	6.06
NC,Train	0.59	0.56	0.53	0.65	0.62	0.55
NC,Test	1.03	0.48	0.54	1.18	0.61	0.68

VV10 column) and the corresponding result obtained with the optimal functional (i.e. the Optimal VV10 column). The optimal RSH+VV10 functional coincides with the ωB97X-V functional, and, as summarized in Table 5, has 7 linear parameters. Compared to the smallest training set RMSD possible (3.14 kcal/mol), a value of 3.36 kcal/mol is certainly reasonable for a functional with 9 fewer linear parameters. Similarly, most other comparisons show that the functional chosen by our selection method yields results for the other reported RMSDs that are competitive with the best values attainable. The largest difference is for the rare-gas results, where it is possible to do nearly twice as well (of course at the expense of TC results) as our chosen functional. Nonetheless, the rare-gas performance of the chosen functional is actually much better than virtually all existing functionals, as will be seen in Section 7.

For the RSH+VV10 category only, we include additional data in Table 9 for functionals that would be considered if the present methodology was not being utilized, to demonstrate that our procedure for selecting the optimal functional is effective. Since the 16-parameter RV-01234.01234.01234.Xn functional has the lowest training set RMSD (3.14 kcal/mol), it is useful to compare the test set RMSDs of this functional (2.36 kcal/mol) against the optimal 7-parameter RSH+VV10 functional (1.92 kcal/mol). While the training set RMSD of the optimal RSH+VV10 functional is 0.22 kcal/mol larger than that of the RV-01234.01234.01234.Xn functional, its test set RMSD is smaller by more than 0.40 kcal/mol. The optimal functional's performance on the TC data in the test set is more than 1.5 times better than RV-01234.01234.01234.Xn, and its RMSD_{RG} is smaller by a factor of 15. These results demonstrate the improved transferability of the optimal 7-parameter functional against a 16-parameter alternative, which comes at the necessary expense of slightly poorer training set performance.

Considering the 4 functionals in Table 9 that satisfy the UEG limits for exchange and correlation, the lowest total RMSD is attained by the RV-12.12.12.Xy functional (7 linear parameters). Since this functional is equivalent to the optimal RSH+VV10 functional with respect to the number of linear parameters, comparing the two highlights the advantages of the present scheme. The optimal RSH+VV10 functional beats the RV-12.12.12.Xy functional in all 10 RMSD categories, and by considerable margins for most. Applying the same analysis to the functionals that do not satisfy the UEG limits, the RV-012.012.012.Xn functional with 10 linear parameters emerges as the one with the lowest total RMSD. However, the optimal RSH+VV10 functional still beats this functional

Table 9: RMSDs for the optimal RSH+VV10 functional, as well as functionals that would be considered if the present methodology was not being utilized. While the nomenclature is explained in Section 5, the first functional corresponds to the optimal RSH+VV10 fit (which coincides with the functional form of ωB97X-V), the next 4 functionals are uniformly truncated $m = 1$ through $m = 4$ fits with all of the UEG constraints enforced, while the last 4 are uniformly truncated $m = 1$ through $m = 4$ fits with none of the UEG constraints enforced. The fraction of short-range exact exchange is optimized for all of the fits in this table.

kcal/mol	Train	Test	RG	Total	TC	NC	TC,Train	TC,Test	NC,Train	NC,Test
RV-12.01.01.Xy	3.36	1.92	0.95	2.68	3.62	0.32	3.86	1.81	0.31	0.32
RV-1.1.1.Xy	3.84	3.86	1.34	3.76	4.64	0.39	4.42	5.68	0.36	0.40
RV-12.12.12.Xy	3.60	2.25	1.82	2.96	3.92	0.37	4.17	2.09	0.31	0.38
RV-123.123.123.Xy	3.50	2.78	7.28	3.51	3.77	0.49	4.01	2.04	0.41	0.51
RV-1234.1234.1234.Xy	3.34	1.94	7.91	3.24	3.63	0.31	3.84	2.15	0.32	0.30
RV-01.01.01.Xn	3.59	2.72	1.87	3.13	3.83	0.48	4.09	1.86	0.40	0.50
RV-012.012.012.Xn	3.34	2.17	1.62	2.78	3.61	0.36	3.83	1.99	0.31	0.37
RV-0123.0123.0123.Xn	3.30	2.37	1.28	2.82	3.56	0.38	3.76	2.22	0.31	0.40
RV-01234.01234.01234.Xn	3.14	2.36	14.00	4.22	3.47	0.33	3.57	2.91	0.28	0.34

with respect to 7 of the 10 RMSDs. Thus, it is clear that this training, testing, and selection procedure allows us to pick a "best of both worlds" functional that fits well to the training set data, yet is highly transferable.

Of the 8 conventional functionals considered in Table 9, the RV-012.012.012.Xn functional has the lowest total RMSD. Another cross-check to consider is training an RV-012.012.012.Xn functional by fitting it to everything in both the training and test sets. The resulting RV-012.012.012.Xn* functional has a TC RMSD of 3.62 kcal/mol and an NC RMSD of 0.30 kcal/mol. However, we have no guarantee that it is transferable. In comparison, the optimal RSH+VV10 functional, has a TC RMSD of 3.62 kcal/mol and an NC RMSD of 0.32 kcal/mol. For the optimal RSH+VV10 functional, the resulting parameters, $\{c_{x,1}, c_{x,2}, c_{css,0}, c_{css,1}, c_{cos,0}, c_{cos,1}, c_x\}$, are $\{0.61, 1.18, 0.58, -0.27, 1.22, -1.87, 0.16\}$. An interesting test is to compare these parameters with the ones that result from training the optimal RSH+VV10 functional on both the training and test sets. The resulting parameters from such a fit are $\{0.60, 1.29, 0.58, -0.32, 1.24, -1.94, 0.16\}$. Since the parameters do not change significantly, this suggests that the training set on its own is sufficiently large for properly determining the parameters.

The inhomogeneity correction factor (ICF) plots associated with the 9 functionals from Table 9 are shown in Figure 4. The optimal RSH+VV10 functional is indicated by the gray lines, which are smooth and well-behaved in all 3 cases. The uniformly truncated $m = 1$ to $m = 4$ functionals (with a non-zero fraction of short-range exact exchange) are indicated by blue, orange, green, and black lines, respectively. Solid lines indicate satisfaction of all 3 UEG constraints, while dashed lines indicate that none of the UEG constraints are satisfied. Since it is preferable to have well-behaved ICFs for transferability, Figure 4 serves as another motivation for the functional selection procedure that is being used. Starting with the exchange functional ICF plots, the optimal RSH+VV10 functional and $m = 2$ plots are quite similar (both are quadratic), while the $m = 4$ plots are similar to the rest between $u_{x,\sigma} = 0$ and $u_{x,\sigma} = 0.5$, but shoot up sharply at $u_{x,\sigma} = 0.5$. While it has been shown[14] that most of the chemically relevant grid points lie between $u_{x,\sigma} = 0$ and $u_{x,\sigma} = 0.5$, it is still preferable to have a curve that looks like the gray one than either of the black ones. Moving on to the same-spin correlation functional ICFs, the quartic $m = 4$ ICFs are oscillatory and seem unphysical, particularly the black dashed curve that does not preserve the UEG limit. The remaining same-spin correlation ICFs are generally well-behaved. The functionals which relax the UEG limit reduce the amount of LSDA same-spin correlation at $u_{c,\sigma\sigma} = 0$ by as much as a factor of 2. For the opposite-spin correlation functional ICFs, the cubic $m = 3$ and quartic $m = 4$ functionals are the outliers, while the remaining functionals behave similarly. Most of

the non-UEG functionals increase the amount of LSDA opposite-spin correlation at $u_{c,\alpha\beta} = 0$ by a factor of 1.2.

There are certainly alternatives to the procedure that is used to find the optimal functional for a given exchange/dispersion pairing. For example, if one considers the top 2 functionals with 7 linear parameters from the RSH+VV10 optimization, they are virtually indistinguishable as far as their 10 RMSDs are concerned, and differ only with respect to the same-spin correlation component. Thus, while the best RSH+VV10 functional with 7 linear parameters is of the RV-12.01.01.Xy form, the second best is of the RV-12.02.01.Xy form. Therefore, we note that the functionals presented here as optimal could be slightly different if a different selection procedure was used. However, after experimenting with various possible options, we can claim that the optimal functional either remains the same or is only very slightly different and that the RMSDs of the optimal functionals are representative of the level of accuracy achievable by the given functional form. In reality, if one were to choose to self-consistently optimize a functional from a certain category, it would certainly be beneficial to consider the top 10 or 20 functionals from a variety of selection procedures in order to assure that the absolute best functional has been chosen.

Before moving on to the remaining 8 categories, it is interesting to consider whether the relaxation of the UEG constraint for exchange is beneficial for the RSH+VV10 category. According to Figure 3, it is clear that for a majority of the points, the relaxation of this constraint leads to no improvements. In fact, the best RSH+VV10 functional that results with the constraint in place is the RV-12.01.01.Xy functional with a value of $c_x = 0.163$, while the best RSH+VV10 functional that results without the constraint is an 8-parameter RV-012.01.01.Xn functional with values of $c_{x,0} = 0.845$ and $c_x = 0.161$, resulting in $c_{x,0} + c_x = 1.006$. Thus, there is absolutely no reason to select the RV-012.01.01.Xn functional over the RV-12.01.01.Xy functional, especially since the total RMSD of the RV-12.01.01.Xy functional is slightly lower than that of the RV-012.01.01.Xn functional.

Moving on to the RSH+DFT-D2 category (Table 6), it is clear that the DFT-D2 dispersion tail

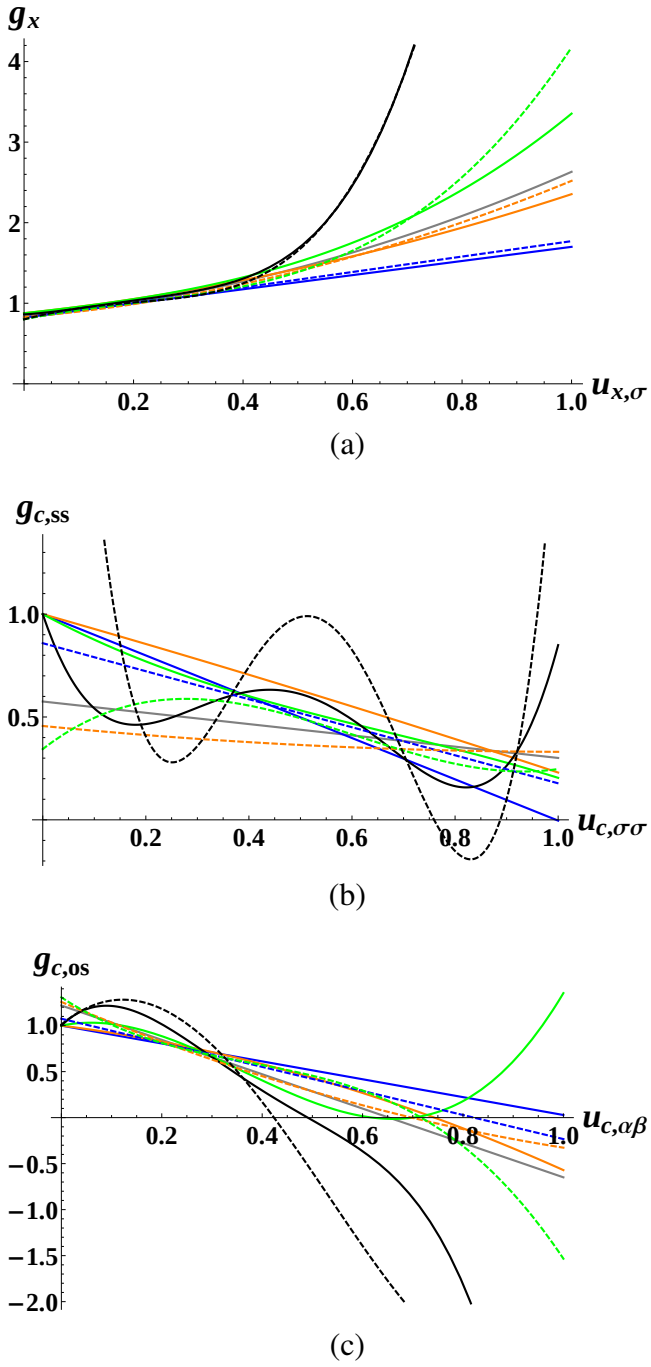

Figure 4: Exchange, same-spin correlation, and opposite-spin correlation inhomogeneity correction factors for 9 functionals from the RSH+VV10 category. The optimal functional from the RSH+VV10 category is shown in gray. The remaining 8 lines belong to uniformly truncated $m = 1$ through $m = 4$ functionals (blue, orange, green, black), with the solid lines indicating satisfaction of all 3 UEG constraints and the dashed lines indicating that none of the UEG constraints are satisfied. The fraction of short-range exact exchange is optimized for all of the fits that are plotted.

is inferior to the VV10 NLC functional when coupled with RSH exchange. The optimal RSH+DFT-D2 functional is an 8-parameter functional with a total RMSD of 3.05 kcal/mol (compared to 2.68 kcal/mol for the optimal RSH+VV10 functional). In comparison to the optimal RSH+VV10 functional, the optimal RSH+DFT-D2 functional is worse with respect to all 10 RMSD categories. Comparing the minimum RMSDs possible by the 2 types of functionals, they are equivalent only with respect to the training set RMSD and the training set TC RMSD, while the RSH+VV10 functional form outperforms the RSH+DFT-D2 functional form on the remaining 8 RMSDs. In addition, the RMSD_{RG} of the optimal RSH+DFT-D2 functional is more than 2.5 times larger than that of the optimal RSH+VV10 functional.

While the functionals in both the RSH+DFT-D2 and RSH+VV10 categories are able to account for dispersion, it is interesting to compare them to the RSH+None category without dispersion corrections. From this category, the optimal functional that emerges is an 8-parameter functional that maintains $c_{x,0} = 1$ but violates the UEG limit ever so slightly for exchange by optimizing $c_x = 0.02$. As expected, the total RMSD of the optimal RSH functional is larger than that of the optimal RSH+DFT-D2 functional, and even larger than that of the optimal RSH+VV10 functional.

Instead of performing similar comparisons for the 6 remaining local and GH functionals, it is easiest to compare the total RMSDs of all 9 optimal functionals with the help of Table 10. This table confirms that the best overall performance is achieved by the optimal RSH+VV10 functional. Keeping the dispersion component constant, the RSH functionals outperform the GH functionals, while the GH functionals outperform the local functionals. As far as dispersion corrections are concerned, it is obvious that the functionals without dispersion corrections (None) perform worse than those with either DFT-D2 or VV10. However, it is less obvious which of the dispersion corrections is better. For the local exchange category, the optimal functional with the DFT-D2 dispersion tail slightly outperforms the one with the VV10 NLC functional, while the reverse is true for the GH exchange category. Ultimately, it is clear that as far as performance is concerned, the RSH+VV10 functional form is the best from the 9 variants considered.

Table 10: Total RMSDs in kcal/mol for the optimal functionals from all 9 categories.

kcal/mol	None	DFT-D2	VV10
Local	5.89	4.43	4.57
GH	4.72	3.14	3.06
RSH	3.87	3.05	2.68

While Figure 3 in its present form has already been stripped of thousands of datapoints for clarity, it still contains a great deal of information. It is very interesting that from the 9 categories, 7 of the optimal functionals do not satisfy the UEG constraint for exchange, while only the GH+None and RSH+VV10 optimal functionals satisfy this limit. In certain cases, as in the Local+DFT-D2 case, the difference between the total RMSDs of the optimal 9-parameter functional (4.43 kcal/mol) and the best possible 9-parameter functional that satisfies the UEG constraint for exchange (4.76 kcal/mol) is more than 0.30 kcal/mol. For other cases, like for the RSH+VV10 category, the difference is very small.

To convey an idea of what the plots in Figure 3 would look like if points had not been removed, Figure 5 shows all of the points corresponding to the RSH+VV10 fits for values of RMSD_{Total} between 2.65 and 3.00 kcal/mol. The filled red circles correspond to fits that do not skip orders in any of the dimensionless variables *and* satisfy the UEG constraint for exchange, while the unfilled red circles belong to similar non-skipping fits that do not satisfy the UEG constraint for exchange. Considering only the filled red circles, it is clear that going from 5 to 6 to 7 linear parameters results in large decreases in the total RMSD, at a rate of 0.15 kcal/mol per parameter. As the 7-parameter point is reached, the total RMSDs completely flatten out, and the quality of the fits begins to deteriorate after 9 linear parameters. The lowest 7-parameter filled red circle corresponds to the optimal RSH+VV10 functional that has been selected from considering Figure 3.

Once fits that skip orders in u are introduced (black dots), it is possible to slightly reduce the total RMSD of the optimal 7-parameter fit by going to the best 9-parameter fit, but by our selec-

tion criteria, the additional 0.02 kcal/mol improvement is not worth the 2 additional parameters. Finally, 2 special points on this plot corresponding to conventional uniform truncations are indicated by filled cyan triangles. The upright triangle corresponds to the 7-parameter $m = 2$ functional from the fourth row of Table 9 that satisfies all 3 UEG constraints (RV-12.12.12.Xy), while the downright triangle is the related 10-parameter RV-012.012.012.Xn functional that violates all 3 UEG constraints (the GGA portion is identical to that of Becke's B97 functional). Comparing these 2 functionals to the optimal 7-parameter functional again shows the ability of our selection procedure to reveal the best functional possible for the least number of empirical parameters. In fact, Figure 5 indicates that it is possible to considerably outperform the $m = 2$ functional that does not satisfy any of the UEG constraints with 3 less empirical parameters.

7 Comparisons

All of the 9 types of functionals considered thus far have existing non-empirical and semi-empirical counterparts. We compare the 9 resulting optimal functionals to the following: PBE,[53] B97-D,[8] VV10,[20] B97,[1] B97-D2,[25] B3LYP-NL,[54] ωB97X,[12] ωB97X-D,[11] and LC-VV10.[20] A summary of how the optimal functionals obtained here compare with these selected existing functionals is given in Table 11.

7.1 Local Functionals

Beginning with the Local+None category, we can compare the resulting optimal functional to the non-empirical PBE functional. In general, the addition of 7 empirical parameters reduces the RMSDs by a factor of 2. The TC RMSD of PBE is reduced from 10.19 kcal/mol to 5.72 kcal/mol, while the NC RMSD of PBE is reduced from 2.05 kcal/mol to 1.08 kcal/mol. However, since both of these statistical measures contain datapoints from the training set, it is imperative to compare the performance of the 2 functionals on the test set. The $\text{RMSD}_{TC,Test}$ of the optimal Local+None functional is more than 3 kcal/mol lower than that of PBE, while its $\text{RMSD}_{NC,Test}$ is smaller by a factor of 2.

Moving on to the Local+DFT-D2 functionals, we can compare the resulting optimal functional to Grimme's B97-D functional, since the optimal functional is a reoptimization of this functional on a different training set (with a different set of ICF expansions). The 10-15% improvement in performance is not as drastic as in the Local+None category, confirming that the B97-D functional is near the limit of accuracy achievable by a Local+DFT-D2 GGA functional.

Finally, it is interesting to compare the performance of the existing VV10 exchange-correlation (xc) functional (rPW86 exchange[55] + PBE correlation + VV10 NLC) with the optimal Local+VV10 functional. As in the Local+None case, the optimal functional generally improves upon the performance of the VV10 xc functional by a factor of 2. However, it is interesting to point out that the performance of the VV10 xc functional is *better* for the 3 rare-gas dimer PECs, indicating that the weight of 2500 may be insufficient. For the optimization of the ωB97X-V functional, a weight of 25000 provided PECs that matched or improved upon those of the VV10 xc functional.

7.2 GH Functionals

Moving on to the GH+None category, we can compare against Becke's B97 functional. The largest improvements come from the noncovalent interactions, since B97 was only fit to TC data. Thus, there is a threefold improvement in both $\text{RMSD}_{NC,Train}$ and $\text{RMSD}_{NC,Test}$, while the thermochemistry improvements are less dramatic. However, the $\text{RMSD}_{TC,Test}$ value for the optimal functional is smaller by a factor of 2, primarily due to its improved performance on the AlkAtom19 dataset.

While B97 was optimized only on thermochemistry, B97-D2 improves upon the NC RMSD of B97 by a factor of more than 5, with the help of only one additional linear parameter. While the optimal GH+DFT-D2 functional is 10-15% better for thermochemistry in general, it is 5-10% worse for noncovalent interactions. However, the performance of B97-D2 for the 3 rare-gas dimer PECs is worse by a factor of 1.5. Overall, it appears as

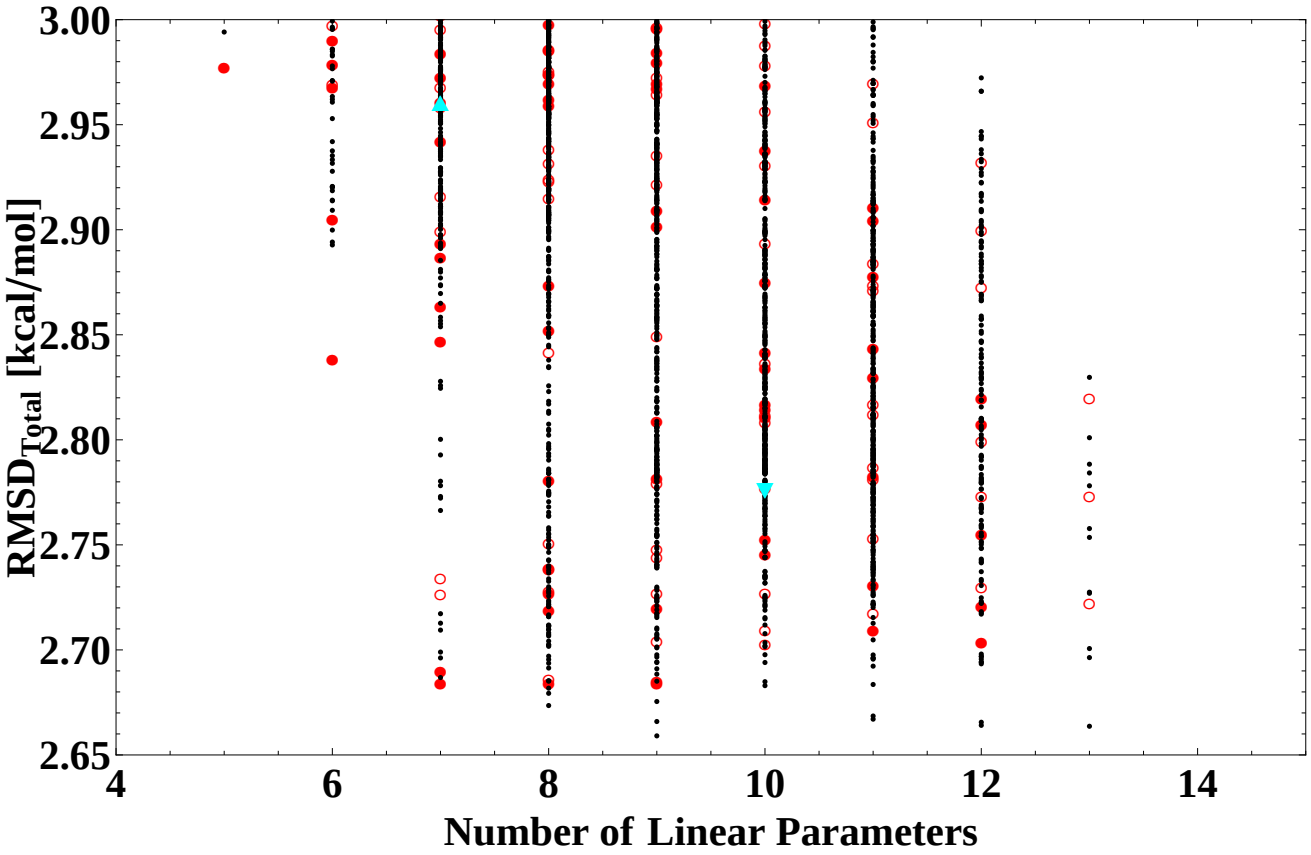

Figure 5: Total RMSDs plotted against the number of linear parameters for all 81919 possible RSH+VV10 fits. The filled red circles correspond to fits that do not skip orders in any of the dimensionless variables *and* satisfy the UEG constraint for exchange, while the unfilled red circles belong to similar non-skipping fits that do not satisfy the UEG constraint for exchange. The remaining points correspond to fits that skip orders in one or more of the ICFs. The filled upright cyan triangle corresponds to the total RMSD of the RSH+VV10 $m = 2$ functional that satisfies all 3 UEG constraints, while the filled downright cyan triangle corresponds to the total RMSD of the RSH+VV10 $m = 2$ functional that does not satisfy any of the UEG constraints. The optimal functional from the RSH+VV10 category is indicated by the lowest point on the vertical line that corresponds to 7 linear parameters.

Table 11: RMSDs in kcal/mol for a variety of existing density functionals for comparison to the RMSDs of the 9 optimal functionals (shown in parentheses).

Category	Local+None	Local+DFT-D2	Local+VV10	GH+None	GH+DFT-D2	GH+VV10	RSH+None	RSH+DFT-D2	RSH+VV10
kcal/mol	PBE	B97-D	VV10	B97	B97-D2	B3LYP-NL	ωB97X	ωB97X-D	LC-VV10
Train	8.85 (4.91)	4.78 (4.48)	9.10 (4.44)	4.36 (3.49)	3.36 (3.22)	4.19 (3.25)	3.67 (3.73)	3.42 (3.55)	6.01 (3.36)
Test	12.67 (7.00)	5.21 (4.50)	8.11 (4.88)	15.41 (5.90)	3.84 (3.17)	6.62 (3.03)	5.08 (4.18)	2.83 (2.50)	5.22 (1.92)
RG	3.76 (2.09)	5.55 (3.15)	1.61 (2.57)	5.82 (2.01)	2.65 (1.82)	1.73 (0.65)	1.82 (1.97)	8.79 (2.50)	2.36 (0.95)
Total	10.64 (5.89)	5.03 (4.43)	8.40 (4.57)	11.02 (4.72)	3.56 (3.14)	5.38 (3.06)	4.32 (3.87)	3.68 (3.05)	5.51 (2.68)
TC	10.19 (5.72)	5.67 (5.23)	10.09 (5.27)	4.85 (3.87)	4.04 (3.48)	4.99 (3.61)	3.89 (4.09)	3.64 (3.80)	6.93 (3.62)
NC	2.05 (1.08)	0.70 (0.61)	1.29 (0.65)	2.71 (0.97)	0.49 (0.54)	1.02 (0.48)	0.92 (0.71)	0.53 (0.43)	0.68 (0.32)
TC,Train	10.27 (5.55)	5.43 (5.13)	10.44 (5.11)	3.89 (3.78)	3.83 (3.62)	4.47 (3.70)	4.10 (4.28)	3.82 (4.02)	6.86 (3.86)
TC,Test	9.76 (6.57)	6.79 (5.71)	7.87 (6.06)	8.28 (4.28)	5.00 (2.60)	7.18 (3.06)	2.45 (2.85)	2.42 (2.26)	7.32 (1.81)
NC,Train	0.89 (0.65)	0.76 (0.62)	1.09 (0.55)	1.66 (0.48)	0.44 (0.48)	0.97 (0.42)	0.66 (0.42)	0.61 (0.42)	0.86 (0.31)
NC,Test	2.29 (1.18)	0.68 (0.61)	1.35 (0.68)	2.96 (1.08)	0.50 (0.56)	1.04 (0.50)	0.98 (0.77)	0.50 (0.43)	0.61 (0.32)

though the B97-D2 functional is near the limit of accuracy achievable by a GH+DFT-D2 GGA functional.

Finally, the optimal GH+VV10 functional can be compared to Grimme's recent parameterization of the B3LYP-NL functional. The B3LYP-NL functional was developed by appending the VV10 NLC functional to the existing B3LYP functional and optimizing only the b parameter ($b = 4.8$). Compared to B3LYP-NL, the performance of the optimal GH+VV10 functional is generally better by a factor of 1.5-2. As yet another indication of transferability, while the $\mathrm{RMSD}_{TC,Train}$ of the optimal functional is only 20% better than that of B3LYP-NL, its $\mathrm{RMSD}_{TC,Test}$ value is smaller by a factor of 2.

7.3 RSH Functionals

Considering the RSH functionals, the first valid comparison is between the optimal RSH+None functional and ωB97X. Since ωB97X was trained primarily on thermochemistry, it is not surprising that it is 5-10% better than the optimal RSH+None functional for thermochemistry. Conversely, the performance of the optimal RSH+None functional is 15-20% better for noncovalent interactions. In addition, the performance of both functionals for the rare-gas dimer PECs is almost identical. It appears that ωB97X is moderately close to the RSH+None performance limit, but employs significantly more parameters than our methodology establishes is necessary (13 vs. 8).

Moving on to the RSH+DFT-D2 category, the optimal functional is compared to ωB97X-D. As a reminder, the damping function that was used for ωB97X-D is slightly different from the one used in DFT-D2 and requires the optimization of a nonlinear parameter instead of a linear s_6 parameter. Nevertheless, ωB97X-D has TC, NC, and RG RMSDs of 3.64, 0.53, and 8.79 kcal/mol, compared to 3.80, 0.43, and 2.50 kcal/mol for the optimal RSH+DFT-D2 functional. As far as the rare-gas dimer PECs are concerned, it is clear that the selection strategy has worked, since the RMSD_{RG} of the optimal functional is 3.5 times smaller than that of ωB97X-D. Even though the optimal RSH+DFT-D2 functional has 5 less linear parameters than ωB97X-D, its performance on the noncovalent interactions in the test set is 15% better, as is its performance for the thermochemistry data in the test set.

Finally, we can compare the optimal RSH+VV10 functional to LC-VV10. The comparison between LC-VV10 and the optimal RSH+VV10 functional is interesting, because the main difference between the 2 functionals is that the GGA component of the optimal functional has been parameterized. The TC, NC, and RG RMSDs of LC-VV10 are 6.93, 0.68, and 2.36 kcal/mol, compared to 3.62, 0.32, and 0.95 kcal/mol for the optimal RSH+VV10 functional. Thus, by simply adding 7 empirical parameters, all 3 RMSDs are reduced by at least a factor of 2. In addition, Figure 1 from Reference 20 indicates that the VV10 xc functional (and thus the VV10 NLC functional) is very accurate for the argon dimer and krypton dimer PECs, so it is desirable to maintain this feature as empirical parameters are added. Largely due to the methodology employed here, the performance of the optimal RSH+VV10 functional is at least 1.5 times better than VV10 and LC-VV10 on the neon dimer, argon dimer, and neon-argon dimer PECs.

8 Conclusions

In developing new semi-empirical density functionals, there are numerous pitfalls on the road to achieving better performance than existing functionals. In this work, we have tried to address, within a limited scope, 2 of the principal issues: (a). "How does one assess the practical benefit of physical augmentation of a functional in a consistent way, including its transferability?", and (b). "How can one determine when an optimal number of empirical parameters have been incorporated into a given functional form?"

To address the first question with manageable scope, we have compared 3 types of density functionals that are all built upon standard generalized gradient approximations of the Becke 97 form:[1] local, global hybrid, and range-separated hybrid. Each of these 3 basic forms are compared against augmented forms that include dispersion corrections via either Grimme's DFT-D2 dispersion tail or the VV10 nonlocal correlation functional. This defines a 3 by 3 grid of functional forms, each of

which can be trained with an enormous variety of parameters.

To address the second question, as well as to complete the evaluation of the first question, we have developed a protocol for selecting the best functional of each type. This protocol involves training an enormous number of candidate functionals containing different numbers of linear parameters on 1108 pieces of training set data. The best such functional is selected based on an additional 1193 pieces of test set data, to assess transferability as well as overall performance. It should be noted that functionals are not trained self-consistently, but the RMSDs obtained are reliable indicators of self-consistent performance, as we have validated elsewhere for the most complicated form considered.

The first main outcome is the conclusion that the best functionals of each type considered contain significantly fewer linear parameters than many existing functionals in the literature. We believe this is largely because of the emphasis on transferability, rather than just training set performance. Typical "optimal" functionals involve between 7 and 9 linear empirical parameters. Functionals with larger numbers of linear parameters can train better but exhibit increasingly poor transferability. Of course there are fine differences between competing best choices in some cases, but this overall result is robust.

The second main outcome concerns the relative performance of the different functional forms within this consistent framework. We find that by far the best possible performance is obtained by the range-separated hybrid functional, coupled to the VV10 NLC functional. This is accordingly the best single candidate for self-consistent optimization, a topic that we have addressed elsewhere to define the ωB97X-V functional.[14] While the self-consistent optimization of a local GGA functional appended with VV10 is an interesting opportunity for a lower cost functional, it is unclear whether the resulting functional will perform significantly better than the best existing local GGAs with DFT-D2 corrections, such as B97-D.

The third main outcome concerns how the 9 optimized forms compare with existing literature functionals that fit within each of those 9 categories. In some cases, very significant improvements are evident, such as for a local functional (vs. PBE) and for a range-separated hybrid functional with VV10 (vs. LC-VV10), which are due largely to comparing against non-empirical (PBE) or relatively non-empirical (LC-VV10) functionals. In other cases, such as range-separated hybrids with a dispersion tail, modest improvements are possible while significantly reducing the number of linear parameters (vs. ωB97X-D), indicating that less semi-empiricism than existing choices can actually be advantageous.

Finally, there are interesting non-trivial opportunities to extend the present analysis beyond the GGA framework we have restricted ourselves to here. It is clearly very desirable to explore the question of how much additional improvement can be obtained by semi-empirical functionals that depend on the kinetic energy density. This will vastly increase the number of possible functionals to approximately 2^{75}, so it is unlikely to be possible to do it up to the $m = 4$ truncation we have employed here. However, the encouraging conclusions about the relatively low degree of optimal semi-empiricism suggest that this may in fact not be necessary. We hope to report on this problem in the near future.

9 Acknowledgements

This work was supported by the Director, Office of Energy Research, Office of Basic Energy Sciences, Chemical Sciences Division of the U.S. Department of Energy under Contract DE-AC0376SF00098, and by a grant from the SciDac Program. We acknowledge computational resources obtained under NSF award CHE-1048789.

References

(1) Becke, A. D. *The Journal of Chemical Physics* **1997**, *107*, 8554–8560.

(2) Herman, F.; Van Dyke, J. P.; Ortenburger, I. B. *Phys. Rev. Lett.* **1969**, *22*, 807–811.

(3) Slater, J. C. *Phys. Rev.* **1951**, *81*, 385–390.

(4) Becke, A. D. *The Journal of Chemical Physics* **1986**, *84*, 4524–4529.

(5) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. *The Journal of Chemical Physics* **1998**, *109*, 6264–6271.

(6) Boese, A. D.; Doltsinis, N. L.; Handy, N. C.; Sprik, M. *The Journal of Chemical Physics* **2000**, *112*, 1670–1678.

(7) Boese, A. D.; Handy, N. C. *The Journal of Chemical Physics* **2001**, *114*, 5497–5503.

(8) Grimme, S. *Journal of Computational Chemistry* **2006**, *27*, 1787–1799.

(9) Wilson, P. J.; Bradley, T. J.; Tozer, D. J. *The Journal of Chemical Physics* **2001**, *115*, 9233–9242.

(10) Keal, T. W.; Tozer, D. J. *The Journal of Chemical Physics* **2005**, *123*, 121103.

(11) Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

(12) Chai, J.-D.; Head-Gordon, M. *The Journal of Chemical Physics* **2008**, *128*, 084106.

(13) Lin, Y.-S.; Li, G.-D.; Mao, S.-P.; Chai, J.-D. *Journal of Chemical Theory and Computation* **2013**, *9*, 263–272.

(14) Mardirossian, N.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2014**, –.

(15) Chai, J.-D.; Head-Gordon, M. *The Journal of Chemical Physics* **2009**, *131*, 174105.

(16) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *The Journal of Chemical Physics* **2010**, *132*, 154104.

(17) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **2004**, *92*, 246401.

(18) Lee, K.; Murray, E. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. *Phys. Rev. B* **2010**, *82*, 081101.

(19) Vydrov, O. A.; Van Voorhis, T. *Phys. Rev. Lett.* **2009**, *103*, 063004.

(20) Vydrov, O. A.; Voorhis, T. V. *The Journal of Chemical Physics* **2010**, *133*, 244103.

(21) Becke, A. D. *J. Comput. Chem.* **1999**, *20*, 63–69.

(22) Gill, P. M.; Johnson, B. G.; Pople, J. A. *Chemical Physics Letters* **1993**, *209*, 506 – 512.

(23) Thom H. Dunning, J. *The Journal of Chemical Physics* **1989**, *90*, 1007–1023.

(24) Woon, D. E.; Thom H. Dunning, J. *The Journal of Chemical Physics* **1993**, *98*, 1358–1371.

(25) Burns, L. A.; Álvaro Vázquez-Mayagoitia,; Sumpter, B. G.; Sherrill, C. D. *The Journal of Chemical Physics* **2011**, *134*, 084107.

(26) Shao, Y. et al. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.

(27) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244–13249.

(28) Stoll, H.; Golka, E.; Preuß, H. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **1980**, *55*, 29–41.

(29) Karton, A.; Daon, S.; Martin, J. M. *Chemical Physics Letters* **2011**, *510*, 165 – 178.

(30) Zheng, J.; Zhao, Y.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2007**, *3*, 569–582.

(31) Karton, A.; Tarnopolsky, A.; Lamère, J.-F.; Schatz, G. C.; Martin, J. M. L. *The Journal of Physical Chemistry A* **2008**, *112*, 12868–12886.

(32) Lynch, B. J.; Truhlar, D. G. *The Journal of Physical Chemistry A* **2003**, *107*, 3898–3906.

(33) Chakravorty, S. J.; Gwaltney, S. R.; Davidson, E. R.; Parpia, F. A.; Fischer, C. F. *Phys. Rev. A* **1993**, *47*, 3649–3670.

(34) Mardirossian, N.; Lambrecht, D. S.; McCaslin, L.; Xantheas, S. S.; Head-Gordon, M. *Journal of Chemical Theory and Computation* **2013**, *9*, 1368–1380.

(35) Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. *The Journal of Physical Chemistry A* **2009**, *113*, 10146–10159.

(36) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. *The Journal of Chemical Physics* **2011**, *135*, 194102.

(37) Thanthiriwatte, K. S.; Hohenstein, E. G.; Burns, L. A.; Sherrill, C. D. *Journal of Chemical Theory and Computation* **2011**, *7*, 88–96.

(38) Crittenden, D. L. *The Journal of Physical Chemistry A* **2009**, *113*, 1663–1669.

(39) Brittain, D. R. B.; Lin, C. Y.; Gilbert, A. T. B.; Izgorodina, E. I.; Gill, P. M. W.; Coote, M. L. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1138–1142.

(40) Karton, A.; Gruzman, D.; Martin, J. M. L. *The Journal of Physical Chemistry A* **2009**, *113*, 8434–8447.

(41) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 43–52.

(42) Zhao, Y.; González-García, N.; Truhlar, D. G. *The Journal of Physical Chemistry A* **2005**, *109*, 2012–2018.

(43) Tang, K. T.; Toennies, J. P. *The Journal of Chemical Physics* **2003**, *118*, 4976–4983.

(44) Hohenstein, E. G.; Sherrill, C. D. *The Journal of Physical Chemistry A* **2009**, *113*, 878–886.

(45) Bryantsev, V. S.; Diallo, M. S.; van Duin, A. C. T.; Goddard, W. A. *Journal of Chemical Theory and Computation* **2009**, *5*, 1016–1026.

(46) Goerigk, L.; Grimme, S. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670–6688.

(47) Copeland, K. L.; Tschumper, G. S. *Journal of Chemical Theory and Computation* **2012**, *8*, 1646–1656.

(48) Zhao, Y.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2005**, *1*, 415–432.

(49) Gráfová, L.; Pitoňák, M.; Řezáč, J.; Hobza, P. *Journal of Chemical Theory and Computation* **2010**, *6*, 2365–2376.

(50) Řezáč, J.; Riley, K. E.; Hobza, P. *Journal of Chemical Theory and Computation* **2011**, *7*, 2427–2438.

(51) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(52) Řezáč, J.; Riley, K. E.; Hobza, P. *Journal of Chemical Theory and Computation* **2011**, *7*, 3466–3470.

(53) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(54) Hujo, W.; Grimme, S. *Journal of Chemical Theory and Computation* **2011**, *7*, 3866–3871.

(55) Murray, E. D.; Lee, K.; Langreth, D. C. *J. Chem. Theory Comput.* **2009**, *5*, 2754–2762.