THE JOURNAL OF PHYSICAL CHEMISTRY A
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCA

Article

# Small Representative Databases for Testing and Validating Density Functionals and Other Electronic Structure Methods

*Published as part of The Journal of Physical Chemistry A virtual special issue "Alec Wodtke Festschrift".*

Yinan Shu, Zhaohan Zhu, Siriluk Kanchanakungwankul, and Donald G. Truhlar*

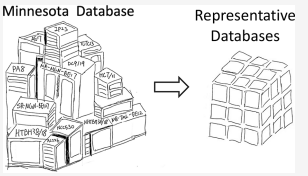
Cite This: *J. Phys. Chem. A* 2024, 128, 6412−6422

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



**ABSTRACT:** Broad and diverse sets of accurate data provide useful metrics for assessing the performance of new theoretical methods. However, assessing methods against large databases can be an arduous task. Here, we present 17 representative energetic databases, defined as small databases whose errors and error spreads are representative of larger databases and which therefore can serve as efficient benchmarks for developing and testing electronic structure methods and density functionals. In 15 cases, the representative databases have 6 entries while being representative of larger databases with 14−107 entries, and in the other two cases, they have 14 entries while being representative of larger databases with 418−455 entries. The mean unsigned error (MUE) of 100 electronic structure methods on a given representative database is typically within about 8% of the MUE on its parent database, and the root-mean-square error (RMSE) is typically within about 11% of the RMSE on the parent database. Thus, the representative databases are quite successful in indicating accuracy while maintaining good diversity. The databases include both main-group and transition-metal compounds and reactions, and they include bond energies, reaction energies, barrier heights, noncovalent interactions, ionization potentials, and absolute energies.




## 1. INTRODUCTION

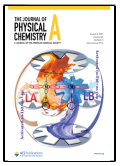
Testing approximate quantum mechanical methods against databases of accurate data plays a major role in validating electronic structure methods used in chemistry, and such databases can also be used during method development. Considerable effort has gone into developing large databases.[1−13] However, especially in exploratory stages of method development or when the goal is to test a large number of methods, it is also useful to have smaller databases, provided that they are representative of the larger ones in terms of difficulty and range of chemical characteristics.[14,15] The objective of the work reported here was to develop small databases that are representative of larger ones.

There are several major broad databases that have been used for developing or testing density functionals,[6,7,9−11,16] The current work starts with Minnesota Database 2019, which has been used to develop Minnesota density functionals. Minnesota Database 2019 is diverse in that it involves many subdatabases, including several that go beyond the main group; it involves main-group bond energies, transition-metal bond energies, reaction barrier heights, noncovalent interactions, ionization potentials, electron affinities, proton affinities, isomerization energies, atomic energies, excitation energies, and geometries. Most of the reference values of Minnesota Database 2019 come from experimental values, and, where relevant, the spin−orbit coupling and zero-point energies have been removed so that the data can be directly compared to energetic electronic structure calculations without computing frequencies or spin−orbit couplings.

## 2. METHODS

**2.1. Parent Databases.** We consider a major subset of Minnesota Database 2019[9]; the subset involves 34 subdatabases that are listed in Table 1. Subdatabases 1−25 have been used for both training and testing the revM11 Minnesota functional, and subdatabases 26−34 were used only for testing; the subdatabases were named previously,[10] and we retain those names. The naming convention for each subdatabase is AA*M*, where AA is an abbreviation describing the type of data and *M* is the number of data points, for example HTBH38 denotes the Hydrogen-Transfer Barrier Heights database, which has 38 data points; however, if the given database has been updated without changing the number of data points, we use AA*M*/YY, where YY denotes the last two digits of the year of the update; for example, HTBH38/18 denotes the version updated in 2018. A special situation occurs when the number of data points is given by a product, for example, SIE4 × 4 has 16 data points. Detailed descriptions of Minnesota Database 2019, including citations for the reference data, can be found in ref 10.

**Received:** May 12, 2024
**Revised:** June 25, 2024
**Accepted:** June 26, 2024
**Published:** July 24, 2024






https://doi.org/10.1021/acs.jpca.4c03137
*J. Phys. Chem. A* 2024, 128, 6412−6422

**Table 1. All Databases Used in this Work**

| no. | database | description |
|---|---|---|
| 1 | SR-MGM-BE8 | single-reference main-group metal bond energies |
| 2 | SR-MGN-BE107 | single-reference main-group nonmetal bond energies |
| 3 | SR-TM-BE15 | single-reference transition-metal bond energies |
| 4 | MR-MGM-BE4 | multireference main-group metal bond energies |
| 5 | MR-MGN-BE17 | multireference main-group nonmetal bond energies |
| 6 | MR-TML-BE12 | multireference transition-metal ligand bond energies |
| 7 | MR-TMD-BE3 | multireference transition-metal dimer bond energies |
| 8 | HTBH38/18 | hydrogen-transfer barrier heights |
| 9 | NHTBH38/18 | non-hydrogen-transfer barrier heights |
| 10 | NCCE30/18 | noncovalent complexation energies |
| 11 | NGD21/18 | noble gas dimer weak interactions |
| 12 | IP23 | ionization potentials |
| 13 | EA13/03 | electron affinities |
| 14 | PA8 | proton affinities |
| 15 | 2pIsoE4 | 2p isomerization energies |
| 16 | 4pIsoE4 | 4p isomerization energies |
| 17 | IsoL6/11 | isomerization energies of large molecules |
| 18 | $\pi$TC13 | thermochemistry of $\pi$ systems |
| 19 | AE17 | atomic energies |
| 20 | HC7/11 | hydrocarbon chemistry |
| 21 | SMAE3/19 | sulfur molecule atomization energies |
| 22 | DC9/19 | difficult cases |
| 23 | 3dAEE8 | 3d transition-metal atomic excitation energies |
| 24 | 4dAEE5 | 4d transition-metal atomic excitation energies |
| 25 | pAEE5 | p-block atomic excitation energies |
| 26 | Al2×6 | dimerization energies of aluminum compounds |
| 27 | BHDIV10 | barrier heights of diverse reactions |
| 28 | BHPERI26 | barrier heights of pericyclic reactions |
| 29 | BHROT27 | barrier heights for rotation around single bonds |
| 30 | DIPCS10 | double-ionization potentials of closed-shell systems |
| 31 | HeavySB11 | dissociation energies in heavy-element compounds |
| 32 | PX13 | proton-exchange barriers in H2O, NH3, and HF clusters |
| 33 | SIE4×4 | self-interaction-errors |
| 34 | YBDE18 | ylide bond-dissociation energies |

**Table 2. Parent Databases and Their Composition**

| no. | parent database | description and constituents |
|---|---|---|
| 1 | TMBE27 | transition-metal bond energies: SR-TM-BE15 + MR-TML-BE12 |
| 2 | BH139 | barrier heights: HTBH38/18 + NHTBH38/18 + BHDIV10 + BHPERI26 + BHROT27 |
| 3 | TM40 | transition metal data: SR-TM-BE15 + MR-TML-BE12 + 3dAEE8 + 4dAEE5 |
| 4 | IsoE14 | isomerization energies: 2pIsoE4 + 4pIsoE4 + IsoL6/11 |
| 5 | MGBE174 | main-group bond energies: MR-MGM-BE4 + MR-MGN-BE17 + SR-MGN-BE107 + SR-MGM-BE8 + SMAE3/19 + Al2×6 + HeavySB11 + YBDE18 |
| 6 | SDIP33 | single and double-ionization potentials: IP23 + DIPCS10 |
| 7 | EMNT418 | energetic portion of Minnesota training database: databases 1−25 in Table 1 |
| 8 | EMNTT555 | energetic portion of Minnesota training and testing database: databases 1−34 in Table 1 |
| 9 | AE17 | one constituent: AE17 |
| 10 | HTBH38/18 | one constituent: HTBH38/18 |
| 11 | NHTBH38/18 | one constituent: NHTBH38/18 |
| 12 | SR-MGN-BE107 | one constituent: SR-MGN-BE107 |
| 13 | NCCE30 | one constituent: NCCE30 |
| 14 | NGD21 | one constituent: NGD21 |
| 15 | BHPERI26 | one constituent: BHPERI26 |
| 16 | BHROT27 | one constituent: BHROT27 |
| 17 | IP23 | one constituent: IP23 |

The representative databases developed in the present work are representatives of parent databases, which are either combined subdatabases of Minnesota Database 2019 or a single subdatabase of Minnesota Database 2019. In this work, we consider 17 parent databases, which are listed in Table 2. Parent databases 1−8 are combined subdatabases of Minnesota Database 2019, and parent databases 9−17 are single subdatabases of Minnesota Database 2019.

All data in the databases have units of energy and correspond to calculations done with a fixed set of reference geometries. The reference geometries for the full set databases are available in ref 8. The reference geometries of the representative databases are repeated in the Supporting Information (SI) of the present paper for the convenience of users.

The 8 parent databases obtained by combining subdatabases of Minnesota Database 2019 are as follows (where the naming convention is the same as already explained): 27 transition-metal bond energies (TMBE27), 139 barrier heights (BH139), 40 transition-metal data points (TM40), 14 isomerization energies (IsoE14), 174 main-group bond energies (MGBE174), 33 ionization potentials (SDIP33), 418 Minnesota training data points (EMNT418), and 555 Minnesota training and testing data points (EMNTT555). The EMNTT555 database contains all the data considered in the current work. Notice that in BH139, unlike some barrier height databases, not every reaction included has both a forward and a reverse barrier height; therefore, the number of barrier heights can be odd and is odd.

Each of the 555 data points is computed with Hartree−Fock theory (HF), 42 Kohn−Sham local density functions, and 57 hybrid density functionals, making a total of 100 electronic structure methods. The 99 considered density functionals are listed in Table 3. All energies are Born−Oppenheimer electronic energies; that is, they do not include vibrational zero-point energy or thermal contributions.

The basis sets used for the various approximate calculations with the 100 electronic structure methods are specified in previous papers presenting subdatabases of Minnesota Database 2019. We do not repeat this kind of detail here since it is irrelevant to the usage of the representative databases, which may be used with any method and any basis set.

**2.2. Representative Databases.** All equations in this section refer to a given parent database with $M$ data points. Denote a reference datum as $Q_{i,\mathrm{ref}}$ (where $i$ = 1,2,..., $M$), and denote a value calculated by electronic structure method $\alpha$ (where $\alpha$ = HF or a density functional theory with one of the 99 density functionals) as $Q_{i,\alpha}$. For that method, we define the mean unsigned error (MUE), mean signed error (MSE), and root-mean-square error (RMSE) for the given parent database as

$$\mathrm{MUE}_\alpha = \frac{1}{M}\sum_{i=1}^{M}[Q_{i,\alpha} - Q_{i,\mathrm{ref}}] \quad (1)$$



https://doi.org/10.1021/acs.jpca.4c03137
J. Phys. Chem. A 2024, 128, 6412−6422

**Table 3. Density Functionals**

| no. | type[a] | functional | $X$[b] | year | ref |
|---|---|---|---|---|---|
| **local** | | | | | |
| 1 | LSDA | GKSVWN5 | 0 | 1980 | 18−21 |
| 2 | | GKSVWN3 | 0 | 1980 | 18−21 |
| 3 | GGA-SO (second-order exchange) | SOGGA | 0 | 2008 | 22 |
| 4 | | PBEsol | 0 | 2008 | 23 |
| 5 | | SOGGA11 | 0 | 2011 | 24 |
| 6 | GGA | B86P86 | 0 | 1986 | 25, 26 |
| 7 | | B86LYP | 0 | 1987 | 25, 27 |
| 8 | | BP86 | 0 | 1988 | 26, 28 |
| 9 | | BLYP | 0 | 1988 | 27, 28 |
| 10 | | BR89LYP | 0 | 1989 | 27, 29 |
| 11 | | B86PW91 | 0 | 1991 | 25, 30 |
| 12 | | PW91 | 0 | 1991 | 30 |
| 13 | | BPW91 | 0 | 1991 | 28, 30 |
| 14 | | PBE | 0 | 1996 | 31 |
| 15 | | mPW | 0 | 1997 | 32 |
| 16 | | revPBE | 0 | 1997 | 33 |
| 17 | | RPBE | 0 | 1999 | 34 |
| 18 | | HCTH407 | 0 | 2001 | 2 |
| 19 | | OLYP | 0 | 2001 | 27, 35 |
| 20 | | OPBE | 0 | 2001 | 31, 35, 36 |
| 21 | | MPWLYP1W | 0 | 2005 | 37 |
| 22 | | PBE1W | 0 | 2005 | 37 |
| 23 | | PBELYP1W | 0 | 2005 | 37 |
| 24 | | MOHLYP | 0 | 2005 | 38 |
| 25 | | B97-D | 0 | 2006 | 39 |
| 26 | | MOHLYP2 | 0 | 2009 | 40 |
| 27 | | OreLYP | 0 | 2009 | 27, 35, 41 |
| 28 | NGA | N12 | 0 | 2012 | 42 |
| 29 | | GAM | 0 | 2015 | 43 |
| 30 | meta-GGA | VSXC | 0 | 1998 | 44 |
| 31 | | $\tau$-HCTH | 0 | 2002 | 45 |
| 32 | | TPSS | 0 | 2003 | 46 |
| 33 | | TPSSLYP1W | 0 | 2005 | 37 |
| 34 | | M06-L | 0 | 2006 | 47 |
| 35 | | revTPSS | 0 | 2009 | 48 |
| 36 | | M11-L | 0 | 2011 | 49 |
| 37 | | MGGA_MS0 | 0 | 2013 | 50 |
| 38 | | MGGA_MS1 | 0 | 2013 | 50 |
| 39 | | MGGA_MS2 | 0 | 2013 | 50 |
| 40 | | revM06-L | 0 | 2017 | 51 |
| 41 | meta-NGA | MN12-L | 0 | 2012 | 52 |
| 42 | | MN15-L | 0 | 2015 | 53 |
| **nonlocal** | | | | | |
| 43 | global-hybrid GGA | HFLYP | 100 | 1987 | 27, 54 |
| 44 | | HFPW91 | 100 | 1991 | 30, 54 |
| 45 | | B3PW91 | 20 | 1992 | 28, 55 |
| 46 | | B3LYP | 20 | 1994 | 27, 28, 55, 56 |
| 47 | | PBE0 | 25 | 1996 | 57 |
| 48 | | B1B95 | 28 | 1996 | 58 |
| 49 | | mPW1PW | 25 | 1997 | 32 |
| 50 | | B1LYP | 25 | 1997 | 59 |
| 51 | | B98 | 21.98 | 1998 | 60 |
| 52 | | B97−1 | 21 | 1998 | 61 |
| 53 | | MPW1K | 42.80 | 2000 | 62 |
| 54 | | O3LYP | 11.61 | 2001 | 36 |
| 55 | | B97−2 | 21 | 2001 | 63 |
| 56 | | B3LYP* | 15[c] | 2001 | 64 |
| 57 | | X3LYP | 21.8 | 2004 | 65 |
| 58 | | MPW3LYP | 21.8 | 2004 | 66 |
| 59 | | MPWLYP1M | 5 | 2005 | 38 |
| 60 | | B97−3 | 26.93 | 2005 | 67 |
| 61 | | SOGGA11-X | 35.42 | 2011 | 68 |
| 62 | global-hybrid GGA + MM | B3LYP-D3(BJ) | 20 | 2011 | |
| 63 | RS-hybrid GGA | CAM-B3LYP | 19−65 | 2004 | 69 |
| 64 | | LC-$\omega$PBE | 0−100 | 2006 | 70−73 |
| 65 | | HSE06 | 25−0 | 2006 | 74, 75 |
| 66 | | HISS | 0−60−0 | 2008 | 76 |
| 67 | | $\omega$B97 | 0−100 | 2008 | 77 |
| 68 | | $\omega$B97X | 15.77- | 2008 | 77 |
| 69 | RS-hybrid NGA | N12-SX | 25−0 | 2012 | 78 |
| 70 | RS-hybrid GGA + MM | $\omega$B97X-D | 22.2−100 | 2008 | 79 |
| 71 | global-hybrid meta-GGA | TPSSh | 10 | 2002 | 80 |
| 72 | | $\tau$-HCTHhyb | 15 | 2002 | 45 |
| 73 | | BB1K | 42 | 2004 | 28, 58, 81 |
| 74 | | MPWB1K | 44 | 2004 | 66 |
| 75 | | MPW1B95 | 31 | 2004 | 66 |
| 76 | | BMK | 42 | 2004 | 82 |
| 77 | | TPSS1KCIS | 13 | 2005 | 83 |
| 78 | | MPWKCIS1K | 41 | 2005 | 40 |
| 79 | | MPW1KCIS | 15 | 2005 | 40 |
| 80 | | PBE1KCIS | 22 | 2005 | 84 |
| 81 | | PWB6K | 46 | 2005 | 3 |
| 82 | | PW6B95 | 28 | 2005 | 3 |
| 83 | | M05 | 28 | 2005 | 85 |
| 84 | | M05-2X | 56 | 2005 | 86 |
| 85 | | M06-HF | 100 | 2006 | 87 |
| 86 | | M06 | 27 | 2008 | 88 |
| 87 | | M06-2X | 54 | 2008 | 88 |
| 88 | | M08-HX | 52.23 | 2008 | 89 |
| 89 | | M08-SO | 56.79 | 2008 | 89 |
| 90 | | MGGA_MS2H | 9 | 2013 | 50 |
| 91 | | revM06 | 40.41 | 2018 | 90 |
| 92 | global-hybrid meta-GGA + MM | PW6B95-D3(BJ) | 28 | 2005 | 3, 91 |
| 93 | RS-hybrid meta-GGA | M11 | 42.8−100 | 2011 | 92 |
| 94 | | revM11 | 22.5−100 | 2019 | 10 |
| 95 | | M06-SX | 33.5−0 | 2019 | 11 |
| 96 | RS-hybrid meta-NGA | MN12-SX | 25−0 | 2012 | 78 |
| 97 | global-hybrid meta-NGA | MN15 | 44 | 2015 | 93 |
| 98 | RS-hybrid meta-GGA + rung-3.5 correlation | M11plus | 42.8−100 | 2019 | 94 |
| 99 | meta-GGA + Rung-3.5 correlation | M11pz | 0 | 2023 | 95 |

[a]GGA, generalized-gradient approximation; NGA, nonseparable gradient approximation; meta-GGA, GGA plus local kinetic energy density; meta-NGA, NGA plus local kinetic energy density; global-hybrid ⇒ exchange part of the functional involves a constant percentage of nonlocal HF exchange; RS, range-separated ⇒ exchange part of the functional involves a percentage of nonlocal HF exchange that depends on interelectronic distance; MM, molecular mechanics term. [b]$X$ is percentage of HF exchange. A single value indicates a local functional if $X$ = 0 or a global-hybrid



https://doi.org/10.1021/acs.jpca.4c03137
J. Phys. Chem. A 2024, 128, 6412−6422

Table 3. continued

functional if $X \neq 0$. A range of $X$ values indicates that $X$ depends on interelectronic separations. For a range specified by two values, the short-range limiting percentage comes before the long-range limiting percentage. For a range specified by three values, the values are listed in the order of short-range, medium-range, and long-range. [c]B3LYP* is the name given by Reiher et al.[64] to a functional, that is, like B3LYP except that the percentage of Hartree−Fock exchange is reduced from 20 to 15%.

Table 4. Parent Databases, Representative Databases, DMUE, DPMUE, and P%

| no. | parent database | rep. database | DMUE (kcal/mol) | PMUE (kcal/mol) | P% |
|---|---|---|---|---|---|
| 1 | TMBE27 | rTMBE6 | 5.68 | 0.92 | 16.2 |
| 2 | BH139 | rBH6 | 3.83 | 0.21 | 5.5 |
| 3 | TM40 | rTM6 | 6.98 | 0.98 | 14.0 |
| 4 | IsoE14 | rIsoE6 | 2.68 | 0.20 | 7.5 |
| 5 | MGBE174 | rMGBE6 | 3.87 | 0.49 | 12.7 |
| 6 | SDIP33 | rSDIP6 | 5.55 | 0.52 | 9.4 |
| 7 | EMNT418 | rEMNT14 | 5.51 | 0.38 | 6.9 |
| 8 | EMNTT555 | rEMNTT14 | 5.54 | 0.35 | 6.3 |
| 9 | AE17 | rAE6 | 32.97 | 0.52 | 1.6 |
| 10 | HTBH38/18 | rHTBH6 | 5.14 | 0.14 | 2.7 |
| 11 | NHTBH38/18 | rNHTBH6 | 4.74 | 0.25 | 5.3 |
| 12 | SR-MGN-BE107 | rSR-MGN-BE6 | 1.88 | 0.21 | 11.2 |
| 13 | NCCE30 | rNCCE6 | 1.28 | 0.11 | 8.6 |
| 14 | NGD21 | rNGD6 | 0.20 | 0.01 | 5.0 |
| 15 | BHPERI26 | rBHPER6 | 3.79 | 0.22 | 5.8 |
| 16 | BHROT27 | rBHROT6 | 0.59 | 0.04 | 6.8 |
| 17 | IP23 | rIP6 | 5.32 | 0.56 | 10.5 |

$$\mathrm{MSE}_\alpha = \frac{1}{M}\sum_{i=1}^{M}(Q_{i,\alpha} - Q_{i,\mathrm{ref}}) \tag{2}$$

$$\mathrm{RMSE}_\alpha = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(Q_{i,\alpha} - Q_{i,\mathrm{ref}})^2} \tag{3}$$

Notice, in eqs 1−3, that all data points have the same weight. (For a very small number of cases, due to SCF convergence issues, certain data may be unavailable for a given electronic structure method; in such a case we simply omit that data in calculating the means. The list of missing data points is provided in the SI. Notice that only 15 data points are missing, as compared to a total of 55,485 data points included, and the effect of missing data on the representative databases should be negligible.)

We consider various subsets of the data in the given parent data set; these are trial representative databases with $N_R$ data drawn from the $M$ data in the parent database. We calculate the mean errors for that subset of data by equations analogous to eqs 1−3, and we call these $\mathrm{MUE}_\alpha^R$, $\mathrm{MSE}_\alpha^R$, and $\mathrm{RMSE}_\alpha^R$.

We then define the privation of the selected subset of data as

$$\mathrm{Pr} = \langle |\mathrm{MUE}^R - \mathrm{MUE}| + |\mathrm{MSE}^R - \mathrm{MSE}| + |\mathrm{RMSE}^R - \mathrm{RMSE}| \rangle \tag{4}$$

where, for any quantity $Q_\alpha$,

$$\langle Q \rangle = \frac{1}{100}\sum_{\alpha=1}^{100} Q_\alpha \tag{5}$$

Our goal is to find a representative subset, which we define as a subset with a small $Pr$. To find the best representative database for a given parent database, we use a genetic algorithm to sample many different subsets of a given size $N_R$ for that parent database. We use the Distributed Evolutionary Algorithms in Python (DEAP) package[17] to maximize the fitness function that is taken as the negative of the privation:

$$\mathrm{fitness} = -\mathrm{Pr} \tag{6}$$

The genetic-algorithm pool size is equal to the size of the parent database; the crossover and mutation probabilities are set to 0.5 and 0.2, respectively; and we use 100 generations of optimization for each of 200 random initial guesses. The result with the smallest $P$ is taken as the representative database.

We use $N_R = 6$ for all parent databases except EMNT418 and EMNTT555; for those two we use $N_R = 14$.

Notice that the privation shown in eq 4 has three components, namely,

$$\mathrm{PMUE} = \langle |\mathrm{MUE}^R - \mathrm{MUE}| \rangle \tag{7}$$

$$\mathrm{PMSE} = \langle |\mathrm{MSE}^R - \mathrm{MSE}| \rangle \tag{8}$$

$$\mathrm{PRMSE} = \langle |\mathrm{RMSE}^R - \mathrm{RMSE}| \rangle \tag{9}$$

We combine these as a privation array:

$$P \equiv (\mathrm{PMUE}, \mathrm{PMSE}, \mathrm{PRMSE}) \tag{10}$$

The averages in eqs 1−3 are averages over data points in a parent database, and the averages in eqs 4 and 5 are averages over both data points and method. For tabulation, we need additional notation for averages only over methods for a given datum; we call these difficulties and define them as

$$\mathrm{DMUE}_i \equiv \langle |Q_i - Q_{i,\mathrm{ref}}| \rangle = \frac{1}{100}\sum_{\alpha=1}^{100} |Q_{i,\alpha} - Q_{i,\mathrm{ref}}| \tag{11}$$

$$\mathrm{DMSE}_i \equiv \langle Q_i - Q_{i,\mathrm{ref}} \rangle = \frac{1}{100}\sum_{\alpha=1}^{100} (Q_{i,\alpha} - Q_{i,\mathrm{ref}}) \tag{12}$$



https://doi.org/10.1021/acs.jpca.4c03137
J. Phys. Chem. A 2024, 128, 6412−6422

$$\mathrm{DRMSE}_i \equiv \sqrt{\langle (Q_i - Q_{i,\mathrm{ref}})^2 \rangle} = \sqrt{\frac{1}{100}\sum_{\alpha=1}^{100} (Q_{i,\alpha} - Q_{i,\mathrm{ref}})^2} \quad (13)$$

The three difficulties are called MUE difficulty, MSE difficulty, and RMSE difficulty, respectively. We combine these as a difficulty array:

$$D_i \equiv (\mathrm{DMUE}_i, \mathrm{DMSE}_i, \mathrm{DRMSE}_i) \quad (14)$$

Similarly, one defines the difficulty for the whole parent database as

$$D \equiv (\mathrm{DMUE}, \mathrm{DMSE}, \mathrm{DRMSE}) \quad (15)$$

where

$$\mathrm{DMUE} = \frac{1}{M}\sum_{i=1}^{M} \mathrm{DMUE}_i \quad (16)$$

$$\mathrm{DMSE} = \frac{1}{M}\sum_{i=1}^{M} \mathrm{DMSE}_i \quad (17)$$

$$\mathrm{DRMSE} = \frac{1}{M}\sum_{i=1}^{M} \mathrm{DRMSE}_i \quad (18)$$

Notice that averaging over methods to provide difficulties shows how difficult it is for a data set to be described by a density functional−averaged over all density functionals (including HF). This is useful in the following way. If, for example, the DMUE for a database is 10 kcal/mol, and the privation is 1 kcal/mol, then the privation is only 10% of the difficulty. This is useful way to look at the privation, and therefore, we define the MUE privation percentage as

$$P\% = \frac{\mathrm{PMUE}}{\mathrm{DMUE}} \quad (19)$$

## 3. DISCUSSION

A complete difficulty array (DMUE, DMSE, DRMSE) for each parent database is provided in Table S1 in SI, and a complete privation array (DMUE, DMSE, DRMSE) is provided in Table S2. Tables S3−S11 in the SI provide complete difficulty arrays ($\mathrm{DMUE}_i$, $\mathrm{DMSE}_i$, $\mathrm{DRMSE}_i$) for all data in all the representative databases.

Because we draw the same conclusions when we consider the entire 3-member arrays, we simplify the discussion in the following to the MUE components of these arrays. The size, MUE difficulty (DMUE), MUE privation (PMUE), and MUE privation percentage ($P\%$) of each of the parent databases is given in Table 4. The average of the last column is 8.3%; it means that the MUE calculated from the subset would typically differ from the MUE calculated for the whole database by only 8.0%, indicating that we have met our objective. Similarly, the RMSE reproduces, on average, the RMSE of the parent database within 10.7%. Thus, the representative databases maintaining good diversity while being successful in indicating accuracy.

In the following, we provide the content of the representative databases. To label the representative database, we use the notation $\mathrm{rAAN_R}$ For example, a representative database for HTBH38/18 with 6 data points is denoted as rHTBH6, and a representative database for EMNT418 with 14 data points is denoted as rEMNT14.

**Table 5. Barrier Heights and Reference Values of BH6 and rHTBH6[a]**

| $i$ | $Q_{i,\mathrm{ref}}$ (kcal/mol) |
|---|---|
| BH6, PMUE = 0.28 kcal/mol | |
| $OH + CH_4 \rightarrow ‡ (\rightarrow CH_3 + H_2O)$ | 6.30 |
| $CH_3 + H_2O \rightarrow ‡ (\rightarrow OH + CH_4)$ | 19.50 |
| $H + OH \rightarrow ‡ (\rightarrow O + H_2)$ | 10.90 |
| $O + H_2 \rightarrow ‡ (\rightarrow H + OH)$ | 13.20 |
| $H + H_2S \rightarrow ‡ (\rightarrow H_2 + HS)$ | 3.90 |
| $H_2 + HS \rightarrow ‡ (\rightarrow H + H_2S)$ | 17.20 |
| rHTBH6, PMUE = 0.14 kcal/mol | |
| $H_2O + CH_3 \rightarrow ‡ (\rightarrow OH + CH_4)$ | 19.50 |
| $OH + C_2H_6 \rightarrow ‡ (\rightarrow H_2O + C_2H_5)$ | 3.50 |
| $O + CH_4 \rightarrow ‡ (\rightarrow OH + CH_3)$ | 14.40 |
| $H_2 + PH_2 \rightarrow ‡ (\rightarrow H + PH_3)$ | 24.70 |
| $H + HO \rightarrow ‡ (\rightarrow H_2 + O)$ | 10.90 |
| *s-trans cis*-$C_5H_8 \rightarrow ‡ (\rightarrow$ *s-trans cis*-$C_5H_8)$[b] | 39.70 |

[a]For barriers, the products are shown only to identify the reaction; their energy is not used in the representative database. [b][1,5] sigmatropic rearrangement of *s-trans cis*-1,3-pentadiene.

**Table 6. Barrier Heights, Subdatabases from which Selected, and Reference Values of rBH6[a]**

| $i$ | subdatabase | $Q_{i,\mathrm{ref}}$ (kcal/mol) |
|---|---|---|
| rBH6, PMUE = 0.21 kcal/mol | | |
| $H + H_2 \rightarrow ‡ (\rightarrow H_2 + H)$ | HTBH38/18 | 9.70 |
| $H_2O + NH_2 \rightarrow ‡ (\rightarrow OH + NH_3)$ | HTBH38/18 | 13.70 |
| $F^-\cdots CH_3F \rightarrow ‡ (\rightarrow FCH_3\cdots F^-)$ | NHTBH38/18 | 13.40 |
| 1,3-pentadiene sigmatropic shift | BHPERI26 | 39.70 |
| formonitrile ylide + ethylene: 1,3-dipolar cycloaddition | BHPERI26 | 4.70 |
| $N_2H_4$ N−N bond rotation barrier | BHROT27 | 8.41 |

[a]When products are identified, they are shown only to identify the reaction; their energy is not used in the representative or parent database.

**3.1. HTBH38/18.** We start by considering HTBH38/18. This is an interesting case because we have previously presented a representative database for barrier heights, namely, BH6.[14] The BH6 representative database involves three selected reactions, and for each reaction, both forward and reverse barrier heights are included. The three chemical reactions in BH6 are given in Table 5. The HTBH38/18 has DMUE = 5.1 kcal/mol, and BH6 representative database has PMUE = 0.28 kcal/mol.

Table 5 shows the genetic-algorithm-optimized representative database, rHTBH6, which has a PMUE of only 0.14 kcal/mol. Notice that the better performance is obtained in part because of greater flexibility in that BH6 is restricted to the forward and reverse barrier heights of three chemical reactions, whereas rHTBH6 selects six data points without the requirement that if a reaction is chosen, both its forward and reverse barriers are used. Notice that the older BH6 set, because both forward and backward barriers are chosen, involves 9 individual calculations (3 reactants, 3 products, and 3 transition states), whereas the new set involves 12 calculations (6 reactants and 6 transition states).

Table S3 in SI shows a complete difficulty array for BH6 and rHTBH6.



https://doi.org/10.1021/acs.jpca.4c03137
J. Phys. Chem. A 2024, 128, 6412−6422

Table 7. Barrier Heights and Reference Values of rNHTBH6, rBHPER6, rBHROT6[a]

| $i$ | $Q_{i,ref}$ (kcal/mol) |
|---|---|
| rNHTBH6, PMUE = 0.25 kcal/mol | |
| $H + HF \rightarrow ‡ (\rightarrow FH + H)$ | 42.10 |
| $CH_3F + Cl \rightarrow ‡ (\rightarrow CH_3 + FCl)$ | 59.80 |
| $Cl^- + CH_3Cl \rightarrow ‡ (\rightarrow ClCH_3 + Cl^-)$ | 2.50 |
| $OH^-\cdots CH_3F \rightarrow ‡ (\rightarrow HOCH_3\cdots F^-)$ | 11.00 |
| $HN_2 \rightarrow ‡ (\rightarrow H + N_2)$ | 10.90 |
| $H + C_2H_4 \rightarrow ‡ (\rightarrow CH_3CH_2)$ | 2.00 |
| rBHPERI6, PMUE = 0.22 kcal/mol | |
| *cis*-1,3,5-hexatriene electrocyclic ring closing barrier | 30.80 |
| 1,5-hexadiene sigmatropic shift barrier | 35.80 |
| fulminic acid + ethylene: 1,3-dipolar cycloaddition barrier | 11.80 |
| formazomethine imine + ethylene: 1,3-dipolar cycloaddition barrier | 5.90 |
| phosphole + ethylene: Diels–Alder cycloaddition barrier | 21.30 |
| thiophene + ethylene: Diels–Alder cycloaddition barrier | 31.30 |
| rBHROT6, PMUE = 0.04 kcal/mol | |
| $(CH_3)_2CHC-CH(CH_3)_2$ C–C bond rotation barrier | 3.72 |
| $NH_2OH$ NO bond rotation barrier | 6.91 |
| $NH_2OH$ NO bond rotation barrier | 2.68 |
| bithiophene C–C bond rotation barrier | 1.78 |
| butadiene C–C bond rotation barrier | 6.30 |
| ethylthiourea C–N bond rotation barrier | 10.24 |

[a]When products are identified, they are shown only to identify the reaction; their energy is not used in the representative or parent database.

**3.2. BH139, NHTBH38/18, BHPERI26, and BHROT27.** Additional parent databases composed of barrier heights are BH139 constructed in the current work and NHTBH38/18, BHPERI26, and BHROT27 from the previous work. The optimized representative databases, namely, rBH6, rNHTBH6, rBHPER6, and rBHROT6 are shown in Tables 6 and 7.

Database rBH6 involves 2 reactions from HTBH38/18, 1 reaction from NHTBH38/18, 2 reactions from BHPERI26, and 1 reaction from BHROT27. Therefore, rBH6 involves barrier heights of diverse chemical reactions, and it has a PMUE of only 0.21 kcal/mol.

The PMUEs for rNHTBH6, rBHPER6, and rBHROT6 are 0.25, 0.22, and 0.04 kcal/mol, respectively. These may be compared to DMUEs of 4.74, 3.79, and 0.59 kcal/mol for the parent databases. The privation percentages are only 5.3, 5.8, and 6.8% for rNHTBH6, rBHPER6, and rBHROT6, respectively. Tables S4 and S5 in SI show complete difficulty arrays for rNHTBH6, rBHPER6, and rBHROT6.

Table 9. Isomerization Energies, Subdatabases from which Selected, and Reference Values of rIsoE6

| $i$ | subdatabase | $Q_{i,ref}$ (kcal/mol) |
|---|---|---|
| rIsoE6, PMUE = 0.20 kcal/mol | | |
| $C(CH_3)_4 \rightarrow CH_3CH_2CH_2CH_2CH_3$ | 2pIsoE4 | 3.80 |
| $CH_3CH_2CF_3 \rightarrow CH_2FCHFCH_2F$ | 2pIsoE4 | 26.90 |
| $PhAsH2 \rightarrow C_6H_6AsH$ | 4pIsoE4 | 33.00 |
| $C_4SeH_8 \rightarrow C_4H_7SeH$ | 4pIsoE4 | 20.80 |
| $NH_2C_6H_4-C_6H_4NH_2 \rightarrow C_6H_5NHNHC_6H_5$ | IsoL6/11 | 33.50 |
| $C_6H_4OHCOCH(CH_3)COCH_3 \rightarrow C_6H_4(OCOCH_3)COCH_2CH_3$ | IsoL6/11 | 5.30 |

**3.3. Transition-Metal Parent Databases: TM40 and TMBE27.** The optimized representative databases rTM6 and rTMBE6 are shown in Table 8. Database rTM6 involves 4 multireference transition-metal–ligand bond energies and 2 transition-metal excitation energies. The rTMBE6 representative database involves 3 single-reference transition-metal bond energies and 3 multireference transition-metal–ligand bond energies. The PMUEs of rTM6 and rTMBE6 are 0.98 and 0.92 kcal/mol. The DMUEs for TM40 and TMBE27 are 6.98 and 5.68 kcal/mol. Dividing these yields $P$% for rTM6 and rTMBE6 of 14 and 16%, respectively. Table S6 in SI shows complete difficulty arrays for rTM6 and rTMBE6.

**3.4. Isomerization Parent Database: IsoE14.** The optimized representative database rIsoE6 is shown in Table 9. Database rIsoE6 involves two 2p isomerization energies, two 4p isomerization energies, and two isomerization energies of large molecules. The DMUE of IsoE14 is 2.68 kcal/mol, and PMUE and $P$% for rIsoE6 are 0.20 kcal/mol and 7.5%. Table S7 in SI shows a complete difficulty array for rIsoE6.

**3.5. Main-Group Bond Energy Databases: MGBE174 and SR-MGN-BE107.** The optimized representative databases rMGBE6 and rSR-MGN-BE6 are shown in Tables 10 and 11. Database rMGBE6 involves 4 single-reference main-group nonmetal bond energies, 1 multireference main-group nonmetal bond energy, and 1 dissociation energy of a heavy-

Table 8. Bond Energies, Subdatabases from which Selected, and Reference Values of rTM6 and rTMBE6

| $i$ | broken bonds[a] | subdatabase | $Q_{i,ref}$ (kcal/mol) |
|---|---|---|---|
| rTM6, PMUE = 0.98 kcal/mol | | | |
| $TiCl \rightarrow Ti + Cl$ | 1 | MR-TML-BE12 | 101.70 |
| $Co(CO)_4H \rightarrow Co + 4C + 4O + H$ | 9 | MR-TML-BE12 | 136.68 |
| $NiCH_2^+ \rightarrow Ni^+ + CH_2$ | 1 | MR-TML-BE12 | 76.30 |
| $CuCl \rightarrow Cu + Cl$ | 1 | MR-TML-BE12 | 90.20 |
| Fe excitation energy | N/A | 3dAEE8 | 34.24 |
| Ru+ excitation energy | N/A | 4dAEE5 | 26.17 |
| rTMBE6, PMUE = 0.96 kcal/mol | | | |
| $FeCl_2 \rightarrow Fe + 2Cl$ | 2 | SR-TM-BE15 | 95.15 |
| $CuAg \rightarrow Cu + Ag$ | 1 | SR-TM-BE15 | 40.70 |
| $Pd(PH_3)_2-C_{10}H_{12} \rightarrow Pd(PH_3)_2 + C_{10}H_{12}$ | 1 | SR-TM-BE15 | 17.30 |
| $CrCl \rightarrow Cr + Cl$ | 1 | MR-TML-BE12 | 90.15 |
| $Fe(CO)_5 \rightarrow Fe + 5CO$ | 5 | MR-TML-BE12 | 29.48 |
| $VO \rightarrow V + O$ | 1 | MR-TML-BE12 | 151.00 |

[a]When an entry is a molecular atomization, the atomization energy is divided by the number of bonds broken to yield the average bond energy of the molecule, and the average bond energy is the datum used for that molecule and is given in the last column. N/A denotes not applicable.



https://doi.org/10.1021/acs.jpca.4c03137
J. Phys. Chem. A 2024, 128, 6412–6422

Table 10. Bond Energies, Subdatabases from which Selected, and Reference Values of rMGBE6

| $i$ | broken bonds[a] | subdatabase | $Q_{i,ref}$ (kcal/mol) |
|---|---|---|---|
| rMGBE6, PMUE = 0.49 kcal/mol | | | |
| SO ($M$ = 3) → S + O | 1 | MR-MGN-BE17 | 125.69 |
| $CH_3CH_2OCH_3$ → $CH_3CH_2$ + $OCH_3$ | 1 | SR-MGN-BE107 | 95.26 |
| $C_4H_{10}$ (isobutane) → 4C + 10H | 13 | SR-MGN-BE107 | 100.23 |
| $C_4H_8$ (cyclobutene) → 4C + 8H | 12 | SR-MGN-BE107 | 96.55 |
| $C_5H_8$ (spiropentane) → 5C + 6H | 14 | SR-MGN-BE107 | 91.73 |
| $Te_2(CH_3)_2$ → 2TeMe | 1 | HeavySB11 | 52.91 |

[a]In each case, the energy of the reaction shown is divided by the number of bonds broken to yield the average bond energy of the bonds broken, and that average bond energy is the datum used for that molecule, as listed in the last column.

Table 11. Bond Energies and Reference Values of rSR-MGN-BE6

| $i$ | broken bonds[a] | $Q_{i,ref}$ (kcal/mol) |
|---|---|---|
| rSR-MGN-BE6, PMUE = 0.21 kcal/mol | | |
| *t*ert-butyl-H → *t*ert-butyl + H | 1 | 103.86 |
| $PH_2$ → P + 2H | 2 | 76.60 |
| $Cl_2$ → 2Cl | 1 | 58.07 |
| $SiF_4$ → Si + 4F | 4 | 143.59 |
| $C_4H_8$ (cyclobutene) → 4C + 8H | 12 | 96.55 |
| $C_5H_8$ (spiropentane) → 5C + 6H | 14 | 91.73 |

[a]In each case, the energy of the reaction shown is divided by the number of bonds broken to yield the average bond energy of the bonds broken, and that average bond energy is the datum used for that molecule, as listed in the last column.

Table 12. Ionization Potentials, Subdatabases from which Selected, and Reference Values of rSDIP6

| $i$ | subdatabase | $Q_{i,ref}$ (kcal/mol) |
|---|---|---|
| rSDIP6, PMUE = 0.52 kcal/mol | | |
| Cl → $Cl^+$ | IP23 | 299.10 |
| OH → $OH^+$ | IP23 | 299.10 |
| $S_2$ → $S_2^+$ | IP23 | 216.00 |
| Cr → $Cr^+$ | IP23 | 156.01 |
| Zn → $Zn^+$ | IP23 | 216.63 |
| $C_2H_4$ → $C_2H_4^{2+}$ | DIPCS10 | 655.80 |

Table 13. Ionization Potentials and Reference Values of rIP6

| $i$ | $Q_{i,ref}$ (kcal/mol) |
|---|---|
| rIP6, PMUE = 0.56 kcal/mol | |
| C → $C^+$ | 259.70 |
| SH → $SH^+$ | 238.90 |
| $S_2$ → $S_2^+$ | 216.00 |
| Cu → $Cu^+$ | 178.17 |
| Rh → $Rh^+$ | 172.11 |
| Sc → $Sc^+$ | 151.32 |

element compound. Databases rMGBE6 and rSR-MGN-BE6 have PMUEs of 0.49 and 0.21 kcal/mol, respectively. This may be compared to the difficulties (DMUEs) of MGBE174 and SR-MGN-BE107, which are 3.87 and 1.88 kcal/mol, respectively. The success of the representation is shown by the small $P$% for rMGBE6 and rSR-MGN-BE6, namely, 12.7 and 11.2%, respectively. Tables S8 and S9 in SI shows complete difficulty arrays for rMGBE6 and rSR-MGN-BE6.

**3.6. Ionization Potential Energy Databases: SDIP33 and IP23.** The optimized representative databases rSDIP6 and rIP6 are shown in Tables 12 and 13. Database rSDIP6 involves five ionization potentials and one double-ionization potential. Databases rSDIP6 and rIP6 have PMUEs of 0.52 and 0.56 kcal/mol. The DMUEs for SDIP33 and IP23 are 5.55 and 5.32 kcal/mol, respectively, and $P$% for rSDIP6 and rIP6 are 9.4 and 10.5% respectively. Tables S10 and S11 in SI show a complete difficulty array for rSDIP6 and IP6 databases.

Table 14. Noncovalent Complexation and Dimerization Energies and Reference Values of rNCCE6 and rNGD6

| $i$[a] | $Q_{i,ref}$ (kcal/mol) |
|---|---|
| rNCCE6, PMUE = 0.11 kcal/mol | |
| $NH_3$···$H_2O$ → $NH_3$ + $H_2O$ | 6.38 |
| HCN···ClF → HCN + ClF | 4.80 |
| $NH_3$···$Cl_2$ → $NH_3$ + $Cl_2$ | 4.85 |
| $(H_2S)_2$ → $2H_2S$ | 1.62 |
| parallel-displaced $(C_6H_6)_2$ → $2C_6H_6$ | 2.59 |
| parallel-displaced $(CO_2)_2$ → $2CO_2$ | 1.49 |
| rNGD6, PMUE = 0.01 kcal/mol | |
| $Ne_2$ → 2Ne | 0.08 |
| HeAr → He + Ar | 0.06 |
| HeHe ($r$ = 3.274 Å) → He + He | 0.02 |
| ArAr ($r$ = 3.457 Å) → Ar + Ar | 0.14 |
| KrKr ($r$ = 4.311 Å) → Kr + Kr | 0.34 |
| HeNe ($r$ = 2.731 Å) → He + Ne | 0.01 |

[a]$r$ denotes separation distance.

Table 15. Absolute Energies and Reference Values of rAE6

| $i$ | $Q_{i,ref}$ (kcal/mol) |
|---|---|
| rAE6, PMUE = 0.52 kcal/mol | |
| He | −1822.11 |
| B | −15,470.55 |
| N | −34,255.21 |
| Na | −101,816.22 |
| Al | −152,074.30 |
| S | −249,817.61 |

**3.7. Noncovalent Interaction Databases: NCCE30 and NGD21.** The optimized representative databases rNCCE6 and rNGD6 are shown in Table 14. Databases rNCCE6 and rNGD6 have PMUEs of 0.11 and 0.01 kcal/mol, respectively. The DMUEs of NCCE30 and NGD21 are 1.28 and 0.20 kcal/mol, respectively, so $P$% for rNCCE6 and rNGD6 are 8.6 and 5.0%, respectively. Table S12 in SI shows a complete difficulty array for rNCCE6 and rNGD6 databases.

**3.8. Atomic Energy Database: AE17.** The optimized representative database rAE6 is shown in Table 15. Database rAE6 involves one element from first row, two elements from the second row, and three elements from the third row. The DMUEs for AE17 is 32.97 kcal/mol. Database rAE6 has a



https://doi.org/10.1021/acs.jpca.4c03137
J. Phys. Chem. A 2024, 128, 6412−6422

Table 16. Data, Subdatabases from which Selected, and Reference Values of rEMNT14 and rEMNTT14

| $i^{a,b}$ | broken bonds[c] | subdatabase | $Q_{i,ref}$ (kcal/mol) |
|---|---|---|---|
| rEMNT14, PMUE = 0.44 kcal/mol | | | |
| N | N/A | AE17 | −34,255.21 |
| $OH + NH_3 \rightarrow ‡ (\rightarrow H_2O + NH_2)$ | N/A | HTBH38/18 | 3.40 |
| $O_2 \rightarrow O_2^+$ | N/A | IP23 | 278.90 |
| $P \rightarrow P^+$ | N/A | IP23 | 241.90 |
| $PH \rightarrow PH^+$ | N/A | IP23 | 234.10 |
| $CN \rightarrow C + N$ | 1 | MR-MGN-BE17 | 181.27 |
| $VF_5 \rightarrow V + 5F$ | 5 | MR-TML-BE12 | 112.83 |
| $(CH_4)_2 \rightarrow 2CH_4$ | N/A | NCCE30 | 0.53 |
| $He_2 \rightarrow 2He$ | N/A | NGD21 | 0.02 |
| $HeNe \rightarrow He + Ne$ | N/A | NGD21 | 0.04 |
| NeNe ($r$ = 2.791 Å) → Ne + Ne | N/A | NGD21 | 0.01 |
| $HNC \rightarrow ‡ (\rightarrow HCN)$ | N/A | NHTBH38/18 | 33.00 |
| $H_2S \rightarrow H_3S^+$ | N/A | PA8 | 173.70 |
| *t*ert-butyl−OH → *t*ert-butyl + OH | 1 | SR-MGN-BE107 | 115.02 |
| rEMNTT14, PMUE = 0.37 kcal/mol | | | |
| C | N/A | AE17 | −23,748.08 |
| $Al_2(CH_3)_4 \rightarrow 2Al(CH_3)_2$ | N/A | Al2×6 | 38.40 |
| *cis*-1,3,5-hexatriene electrocyclic ring closing barrier | N/A | BHPERI26 | 30.80 |
| $H_2S_2$ S−S bond rotation barrier | N/A | BHROT27 | 8.03 |
| $O_2 \rightarrow O_2^+$ | N/A | IP23 | 278.90 |
| $Si \rightarrow Si^+$ | N/A | IP23 | 187.90 |
| $Zn \rightarrow Zn^+$ | N/A | IP23 | 216.63 |
| $ClO \rightarrow Cl + O$ | 1 | MR-MGN-BE17 | 64.84 |
| T-shaped $(C_6H_6)_2 \rightarrow 2C_6H_6$ | N/A | NCCE30 | 2.63 |
| $He_2 \rightarrow 2He$ | N/A | NGD21 | 0.02 |
| HeHe ($r$ = 3.274 Å) → He + He | N/A | NGD21 | 0.02 |
| $H + FCH_3 \rightarrow ‡ (\rightarrow HF + CH_3)$ | N/A | NHTBH38/18 | 30.50 |
| $HN_2 \rightarrow ‡ (\rightarrow H + N_2)$ | N/A | NHTBH38/18 | 10.90 |
| *t*ert-butyl−OH → *t*ert-butyl + OH | 1 | SR-MGN-BE107 | 115.02 |

[a]When products are identified after a transition state, they are shown only to identify the reaction; their energy is not used in the representative database, and the barrier height is the datum for that reaction, as listed in the last column. [b]$r$ denotes separation distance. [c]In each case where bonds are broken, the energy of the reaction shown is divided by the number of bonds broken to yield the average bond energy of the bonds broken, and that average bond energy is the datum used for that molecule, as listed in the last column. N/A denotes not applicable.

PMUE of 0.52 kcal/mol and a $P$% of 2.0%. Table S13 in SI shows a complete difficulty array for the rAE6 database.

### 3.9. Minnesota Databases: EMNT418 and EMNTT555.

Next, we consider two very big databases, namely, the energetic portion (EMNT418) of Minnesota training database and the energetic portion (EMNTT555) of the Minnesota training and testing database (the geometric portions of these databases are not included). We used 14 as the size of our optimized representative database size for these larger databases; this results in representative databases rEMNT14 and rEMNTT14. The energetic portions of databases EMNT418 and EMNTT555 involve 25 and 34 subdatabases, respectively. Both the optimized rEMNT14 and rEMNTT14 databases involve systems from 10 subdatabases, which makes them very diverse. The involved systems and the databases from which they are taken are shown in Table 13. The DMUEs for EMNT418 and EMNTT555 are 5.51 and 5.54 kcal/mol. The PMUEs of optimized rEMNT14 and rEMNTT14 are 0.38 and 0.35 kcal/mol, which results in 6.9 and 6.3% for $P$%. Table S14 in SI shows a complete difficulty array for the rEMNT14 and rEMNTT14 databases.

## 4. SUMMARY AND CONCLUSIONS

In this work, we have optimized representative databases formed from various subdatabases of Minnesota Database 2019. The optimizations were achieved by using a genetic algorithm that maximizes the correspondence between the representative database and the parent database in terms of MUE, MSE, and RMSE. The inclusion of the RMSE in the optimizations is important because it means the representative databases include the diversity of the data as well as its difficulty.

Our goal in producing representative databases involves more than just statistics; we try to keep the databases small enough to be very convenient, but large enough to have appropriate diversity. We consider two big parent databases, namely, the energetic portion of the Minnesota training database and the energetic portion of the Minnesota training and testing database. The energetic portion of the Minnesota training database is called EMNT418, and it has 418 data points. The energetic portion of the Minnesota training and testing database is called EMNT418, and it has 555 data points. For these two large parent databases, we considered representative databases with sizes of 14 data points each. These are called rEMNT14 and rEMNTT14 (see Table 16). We also consider representative databases of smaller parent databases, which are either combined subdatabases of Minnesota Database 2019 or a single subdatabase of Minnesota Database 2019. In particular, we consider 15 such smaller parent databases, which are listed in Table 2. Parent


https://doi.org/10.1021/acs.jpca.4c03137
J. Phys. Chem. A 2024, 128, 6412−6422

databases 1−8 are combined subdatabases of Minnesota Database 2019, and parent databases 9−17 are single subdatabases of Minnesota Database 2019.

The 17 parent databases considered here cover a variety of chemical and physical properties of transition-metal and main-group molecules and reactions. The size of the representative databases is 6 for 15 of them and is 14 for the two biggest parent databases, namely, the energetic portion of the Minnesota training database and the energetic portion of the Minnesota training and testing database. On average, our optimized representative databases deviate from parent databases by 8.0% in terms of MUE averaged over 100 electronic structure methods (HF method and Kohn−Sham density functional theory with 99 density functionals). We also achieved success in representing the error spread. The low deviations in errors and error spreads indicate that our current optimization strategy is successful in producing representative databases that adequately reproduce the MSEs, MUEs, and RMSEs of larger databases. The representative databases allow one to gain a quick understanding of a method's accuracy in a modest time and with modest expense.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.4c03137.

Difficulties of parent databases and difficulties and privations of representative databases; reference data and geometries for representative databases (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Author

**Donald G. Truhlar** − *Department of Chemistry, Chemical Theory Center, and Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431, United States;* orcid.org/0000-0002-7742-7294; Email: truhlar@umn.edu

### Authors

**Yinan Shu** − *Department of Chemistry, Chemical Theory Center, and Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431, United States;* orcid.org/0000-0002-8371-0221

**Zhaohan Zhu** − *Department of Biostatistics, University of Florida, Gainesville, Florida 32611, United States*

**Siriluk Kanchanakungwankul** − *Department of Chemistry, Chemical Theory Center, and Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431, United States;* orcid.org/0000-0003-4906-020X

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.4c03137


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors are grateful to Xuelan Wen for stimulating discussion that motivated part of this work. The authors are grateful to Pragya Verma for database curation and for helpful comments on the work and on the manuscript. This work was supported by the National Science Foundation under grant CHE-2054723.

## ■ REFERENCES


(1) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Assessment of Gaussian-3 and Density Functional Theories for a Larger Experimental Test Set. *J. Chem. Phys.* **2000**, *112*, 7374−7383.

(2) Boese, A. D.; Handy, N. C. A New Parametrization of Exchange-Correlation Generalized Gradient Approximation Functionals. *J. Chem. Phys.* **2001**, *114*, 5497−5503.

(3) Zhao, Y.; Truhlar, D. G. Design of Density Functionals that Are Broadly Accurate for Thermochemistry, Thermochemical Kinetics, and Nonbonded Interactions. *J. Phys. Chem. A* **2005**, *109*, 5656−5667.

(4) Peverati, R.; Truhlar, D. G. Quest for a Universal Density Functional: The Accuracy of Density Functionals Across a Broad Spectrum of Databases In Chemistry and Physics. *Philos. Trans. R. Soc. A* **2014**, *372*, No. 20120476.

(5) Karton, A.; Sylvetsky, N.; Martin, J. M. L. W4−17: A Diverse and High-Confidence Dataset of Atomization Energies for Benchmarking High-Level Electronic Structure Methods. *J. Comput. Chem.* **2017**, *38*, 2063−2075.

(6) Goerigk, L.; Hansen, A.; Bauer, C.; Najibi, A.; Grimme, S. 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.* **2017**, *19*, 32184−32215.

(7) Mardirossian, N.; Head-Gordon, M. Survival of the Most Transferable At the Top of Jacob's Ladder: Defining and Testing the ωB97M (2) Double Hybrid Density Functional. *J. Chem. Phys.* **2018**, *148*, No. 241736.

(8) Loos, P.-F.; Scemama, A.; Blondel, A.; Garniron, Y.; Caffarel, M.; Jacquemin, D. A Mountaineering Strategy to Excited States: Highly Accurate Reference Energies and Benchmarks. *J. Chem. Theory Comput.* **2018**, *14*, 4360−4379.

(9) Verma, P.; Truhlar, D. G. *Geometries for Minnesota Database 2019*; University of Minnesota, 2019. https://hdl.handle.net/11299/208752 (retrieved from the Data Repository for the University of Minnesota) (accessed on April 21, 2024).

(10) Verma, P.; Wang, Y.; Ghosh, S.; He, X.; Truhlar, D. G. Revised M11 Exchange-Correlation Functional for Electronic Excitation Energies and Ground-State Properties. *J. Phys. Chem. A* **2019**, *123*, 2966−2990.

(11) Wang, Y.; Verma, P.; Zhang, L.; Li, Y.; Liu, Z.; Truhlar, D. G.; He, X. M06-SX Screened-Exchange Density Functional for Chemistry and Solid-State Physics. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 2294−2301.

(12) Santra, G.; Semidalas, E.; Mehta, N.; Karton, A.; Martin, J. M. L. S66 × 8 Noncovalent Interactions Revisited: New Benchmark and Performance of Composite Localized Coupled-Cluster Methods. *Phys. Chem. Chem. Phys.* **2022**, *24*, 25555−25570.

(13) Jacquemin, D.; Kossoski, F.; Gam, F.; Boggio-Pasqua, M.; Loos, P.-F. Reference Vertical Excitation Energies for transition Metal Compounds. *J. Chem. Theory Comput.* **2023**, *19*, 8782−8800.

(14) Lynch, B. J.; Truhlar, D. G. Small Representative Benchmarks for Thermochemical Calculations. *J. Phys. Chem. A* **2003**, *107*, 8996−8999.

(15) Zheng, J.; Zhao, Y.; Truhlar, D. G. The DBH24/08 Database and its Use to Assess Electronic Structure Model Chemistries for Chemical Reaction Barrier Heights. *J. Chem. Theory Comput.* **2009**, *5*, 808−821.

(16) Morgante, P.; Peverati, R. ACCDB: A Collection of Chemistry DataBases for Broad Computational Purposes. *J. Comput. Chem.* **2019**, *40*, 839−848.

(17) De Rainville, F. M.; Fortin, F. A.; Gardner, M. A.; Parizeau, M.; Gagné, C. Deap: A Python Framework for Evolutionary Algorithms; In *Proceedings of the 14th Annual Conference Companion on Genetic and Evolutionary Computation*; Association for Computing Machinery: New York, 2012; 85−92.

(18) Gáspár, R. Über eine approximation des Hartree-Fockschen Potentials durch eine Universelle Potentialfunktion. *Acta. Phys. Hung.* **1954**, *3*, 263−286.
(19) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133−A1138.
(20) Gáspár, R. Statistical Exchange for Electron in Shell and Xα Method. *Acta. Phys. Hung.* **1974**, *35*, 213−218.
(21) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis. *Can. J. Phys.* **1980**, *58*, 1200−1211.
(22) Zhao, Y.; Truhlar, D. G. Construction of A Generalized Gradient Approximation by Restoring the Density-Gradient Expansion and Enforcing a Tight Lieb-Oxford Bound. *J. Chem. Phys.* **2008**, *128*, No. 184109.
(23) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E.; Constantin, L. A.; Zhou, X.; Burke, K. Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces. *Phys. Rev. Lett.* **2009**, *102*, No. 039902.
(24) Peverati, R.; Zhao, Y.; Truhlar, D. G. Generalized Gradient Approximation that Recovers the Second-Order Density-Gradient Expansion with Optimized Across-the-Board Performance. *J. Phys. Chem. Lett.* **2011**, *2*, 1991−1997.
(25) Becke, A. D. Density Functional Calculations of Molecular Bond Energies. *J. Chem. Phys.* **1986**, *84*, 4524−4529.
(26) Perdew, J. P. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev. B* **1986**, *33*, 8822−8824.
(27) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785−789.
(28) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098−3100.
(29) Becke, A. D.; Roussel, M. R. Exchange Holes in Inhomogeneous Systems: A Coordinate-Space Model. *Phys. Rev. A* **1989**, *39*, 3761−3767.
(30) Perdew, J. P. Unified Theory of Exchange and Correlation Beyond the Local Density Approximation. In *Electronic Structure of Solids '91*; Ziesche, P.; Eschrig, H., Eds.; Akademie Verlag: Berlin, 1991; 11−20.
(31) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(32) Adamo, C.; Barone, V. Exchange Functionals with Improved Long-Range Behavior and Adiabatic Connection Methods without Adjustable Parameters: The mPW and mPW1PW Models. *J. Chem. Phys.* **1998**, *108*, 664−675.
(33) Zhang, Y.; Yang, W. Comment on "Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1998**, *80*, 890.
(34) Hammer, B.; Hansen, L. B.; Nørskov, J. K. Improved Adsorption Energetics within Density-Functional Theory Using Revised Perdew-Burk-Ernzerhof Functionals. *Phys. Rev. B* **1999**, *59*, 7413−7421.
(35) Handy, N.; Cohen, A. Left-Right Correlation Energy. *Mol. Phys.* **2011**, *99*, 403−412.
(36) Hoe, W. M.; Cohen, A. J.; Handy, N. C. Assessment of a New Local Exchange Functional OPTX. *Chem. Phys. Lett.* **2001**, *341*, 319−328.
(37) Dahlke, E. E.; Truhlar, D. G. Improved Density Functionals for Water. *J. Phys. Chem. B* **2005**, *109*, 15677−15683.
(38) Schultz, N. E.; Zhao, Y.; Truhlar, D. G. Density Functionals for Inorganometallic and Organometallic Chemistry. *J. Phys. Chem. A* **2005**, *109*, 11127−11143.
(39) Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787−1799.
(40) Zhao, Y.; González-García, N.; Truhlar, D. G. Benchmark Database of Barrier Heights for Heavy Atom Transfer, Nucleophilic Substitution, Association, and Unimolecular Reactions and Its Use to Test Theoretical Methods. *J. Phys. Chem. A* **2005**, *109*, 2012−2018.
(41) Thakkar, A. J.; McCarthy, S. P. Toward Improved Density Functionals for the Correlation Energy. *J. Chem. Phys.* **2009**, *131*, No. 134109.
(42) Peverati, R.; Truhlar, D. G. Exchange-Correlation Functional with Good Accuracy for Both Structural and Energetic Properties While Depending Only on the Density and Its Gradient. *J. Chem. Theory Comput.* **2012**, *8*, 2310−2319.
(43) Yu, H. S.; Zhang, W. J.; Verma, P.; He, X.; Truhlar, D. G. Nonseparable Exchange-Correlation Functional for Molecules, Including Homogeneous Catalysis Involving Transition Metals. *Phys. Chem. Chem. Phys.* **2015**, *17*, 12146−12160.
(44) Van Voorhis, T.; Scuseria, G. E. A Novel Form for the Exchange-Correlation Energy Functional. *J. Chem. Phys.* **1998**, *109*, 400−410.
(45) Boese, A. D.; Handy, N. C. New Exchange − Correlation Density Functionals: The Role of the Kinetic-Energy Density. *J. Chem. Phys.* **2002**, *116*, 9559−9569.
(46) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, No. 146401.
(47) Zhao, Y.; Truhlar, D. G. A New Local Density Functional for Main-Group Thermochemistry, Transition Metal Bonding, Thermochemical Kinetics, and Noncovalent Interactions. *J. Chem. Phys.* **2006**, *125*, No. 194101.
(48) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Constantin, L. A.; Sun, J. Workhorse Semilocal Density Functional for Condensed Matter Physics and Quantum Chemistry. *Phys. Rev. Lett.* **2009**, *103*, No. 026403.
(49) Peverati, R.; Truhlar, D. G. M11-L: A Local Density Functional that Provides Improved Accuracy for Electronic Structure Calculations in Chemistry and Physics. *J. Phys. Chem. Lett.* **2012**, *3*, 117−124.
(50) Sun, J.; Haunschild, R.; Xiao, B.; Bulik, I. W.; Scuseria, G. E.; Perdew, J. P. Semilocal and Hybrid Meta-Generalized Gradient Approximation Based on the Understanding of the Kinetic-Energy-Density Dependence. *J. Chem. Phys.* **2013**, *138*, No. 044113.
(51) Wang, Y.; Jin, X. S.; Yu, H. S.; Truhlar, D. G.; He, X. Revised M06-L Functional for Improved Accuracy on Chemical Reaction Barrier Heights, Noncovalent Interactions, and Solid-State Physics. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 8487−8492.
(52) Peverati, R.; Truhlar, D. G. An Improved and Broadly Accurate Local Approximation to the Exchange − Correlation Density Functional: The MN12-L Functional for Electronic Structure Calculations in Chemistry and Physics. *Phys. Chem. Chem. Phys.* **2012**, *14*, 13171−13174.
(53) Yu, H. S.; He, X.; Truhlar, D. G. MN15-L: A New Local Exchange-Correlation Functional for Kohn-Sham Density Functional Theory with Broad Accuracy for Atoms, Molecules, and Solids. *J. Chem. Theory Comput.* **2016**, *12*, 1280−1293.
(54) Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23*, 69−89.
(55) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 247−257.
(56) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623−11627.
(57) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0Model. *J. Chem. Phys.* **1999**, *110*, 6158−6170.
(58) Becke, A. D.; Density-Functional Thermochemistry, I. V. A New Dynamical Correlation Functional and Implications for Exact-Exchange Mixing. *J. Chem. Phys.* **1996**, *104*, 1040−1046.
(59) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0Model. *Chem. Phys. Lett.* **1997**, *274*, 242−250.


https://doi.org/10.1021/acs.jpca.4c03137
J. Phys. Chem. A 2024, 128, 6412−6422

(60) Schmider, H. L.; Becke, A. D. Optimized Density Functionals from the Extended G2 Test Set. *J. Chem. Phys.* **1998**, *108*, 9624−9631.
(61) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. Development and Assessment of New Exchange − Correlation Functionals. *J. Chem. Phys.* **1998**, *109*, 6264−6271.
(62) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. Adiabatic Connection for Kinetics. *J. Phys. Chem. A* **2000**, *104*, 4811−4815.
(63) Wilson, P. J.; Bradley, T. J.; Tozer, D. J. Hybrid Exchange-Correlation Functional Determined from Thermochemical Data and Ab Initio Potentials. *J. Chem. Phys.* **2001**, *115*, 9233−9242.
(64) Reiher, M.; Salomon, O.; Hess, B. A. Reparameterization of Hybrid Functionals Based on Energy Differences of States of Different Multiplicity. *Theor. Chem. Acc.* **2001**, *107*, 48−55.
(65) Xu, X.; Goddard, W. A. III The X3LYP Extended Density Functional for Accurate Descriptions of Nonbond Interactions, Spin States, and Thermochemical Properties. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 2673−2677.
(66) Zhao, Y.; Truhlar, D. G. Hybrid Meta Density Functional Theory Methods for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions: The MPW1B95 and MPWB1K Models and Comparative Assessments for Hydrogen Bonding and Van Der Waals Interactions. *J. Phys. Chem. A* **2004**, *108*, 6908−6918.
(67) Keal, T. W.; Tozer, D. J. Semiempirical Hybrid Functional with Improved Performance in an Extensive Chemical Assessment. *J. Chem. Phys.* **2005**, *123*, No. 121103.
(68) Peverati, R.; Truhlar, D. G. Communication: A Global Hybrid Generalized Gradient Approximation to the Exchange − Correlation Functional that Satisfies the Second-Order Density-Gradient Constraint and Has Broad Applicability in Chemistry. *J. Chem. Phys.* **2011**, *135*, No. 191102.
(69) Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange-Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51−57.
(70) Tawada, Y.; Tsuneda, T.; Yanagisawa, S.; Yanai, T.; Hirao, K. A Long-Range-Correlated Time-Dependent Density Functional Theory. *J. Chem. Phys.* **2004**, *120*, 8425−8433.
(71) Vydrov, O. A.; Scuseria, G. E. Assessment of a Long-Range Corrected Hybrid Functional. *J. Chem. Phys.* **2006**, *125*, No. 234109.
(72) Vydrov, O. A.; Heyd, J.; Krukau, A. V.; Scuseria, G. E. Importance of Short-Range versus Long-Range Hartree-Fock Exchange for the Performance of Hybrid Density Functionals. *J. Chem. Phys.* **2006**, *125*, No. 074106.
(73) Vydrov, O. A.; Scuseria, G. E.; Perdew, J. P. Tests of Functionals for Systems with Fractional Electron Number. *J. Chem. Phys.* **2007**, *126*, No. 154109.
(74) Krukau, A. V.; Vydrov, O. A.; Izmaylov, A. F.; Scuseria, G. E. Influence of the Exchange Screening Parameter on the Performance of Screened Hybrid Functionals. *J. Chem. Phys.* **2006**, *125*, No. 224106.
(75) Henderson, T. M.; Izmaylov, A. F.; Scalmani, G.; Scuseria, G. E. Can Short-Range Hybrids Describe Long-Range-Dependent Properties? *J. Chem. Phys.* **2009**, *131*, 225−229.
(76) Henderson, T. M.; Izmaylov, A. F.; Scalmani, G.; Scuseria, G. E.; Savin, A. Assessment of a Middle-Range Hybrid Functional. *J. Chem. Theory Comput.* **2008**, *4*, 1254−1262.
(77) Chai, J. D.; Head-Gordon, M. Systematic Optimization of Long-Range Corrected Hybrid Density Functionals. *J. Chem. Phys.* **2008**, *128*, No. 084106.
(78) Peverati, R.; Truhlar, D. G. Screened-Exchange Density Functionals with Broad Accuracy for Chemistry and Soli-State Physics. *Phys. Chem. Chem. Phys.* **2012**, *14*, 16187−16191.
(79) Chai, J. D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom-Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615−6620.
(80) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Comparative Assessment of a New Nonempirical Density Functional: Molecules and Hydrogen-Bonded Complexes. *J. Chem. Phys.* **2003**, *119*, 12129−12137.
(81) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. Development and Assessment of A New Hybrid Density Functional Model for Thermochemical Kinetics. *J. Phys. Chem. A* **2004**, *108*, 2715−2719.
(82) Boese, A. D.; Martin, J. M. Development of Density Functionals for Thermochemical Kinetics. *J. Chem. Phys.* **2004**, *121*, 3405−3416.
(83) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. Multi-Coefficient Extrapolated Density Functional Theory for Thermochemistry and Thermochemical Kinetics. *Phys. Chem. Chem. Phys.* **2005**, *7*, 43−52.
(84) Zhao, Y.; Truhlar, D. G. Benchmark Databases for Nonbonded Interactions and Their Use to Test Density Functional Theory. *J. Chem. Theory Comput.* **2005**, *1*, 415−432.
(85) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Exchange-Correlation Functional with Broad Accuracy for Metallic and Nonmetallic Compounds, Kinetics, and Noncovalent Interactions. *J. Chem. Phys.* **2005**, *123*, 161103.
(86) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Design of Density Functionals by Combining the Method of Constraint Satisfaction with Parametrization for Thermochemistry, Thermochemical Kinetics, and Nonbonded Interactions. *J. Chem. Theory Comput.* **2006**, *2*, 364−382.
(87) Zhao, Y.; Truhlar, D. G. Density Functional for Spectroscopy: No Long-Range Self-Interaction Error, Good Performance for Rydberg and Charge-Transfer States, and Better Performance on Average than B3LYP for Ground States. *J. Phys. Chem. A* **2006**, *110*, 13126−13130.
(88) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, and Nonbonded Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(89) Zhao, Y.; Truhlar, D. G. Exploring the Limit of Accuracy of the Global Hybrid Meta Density Functional for Main-Group Thermochemistry, Thermochemical Kinetics, and Nonbonded Interactions. *J. Chem. Theory Comput.* **2008**, *4*, 1849−1868.
(90) Wang, Y.; Verma, P.; Jin, X. S.; Truhlar, D. G.; He, X. Revised M06 Density Functional for Main-Group and Transition-Metal Chemistry. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 10257−10262.
(91) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(92) Peverati, R.; Truhlar, D. G. Improving the Accuracy of Hybrid Meta-GGA Density Functionals by Range Separation. *J. Phys. Chem. Lett.* **2011**, *2*, 2810−2817.
(93) Yu, H. S.; He, X.; Li, S. L.; Truhlar, D. G. MN15: A Kohn-Sham Global-Hybrid Exchange-Correlation Density Functional with Broad Accuracy for Multi-Reference and Single-Reference Systems and Noncovalent Interactions. *Chem. Sci.* **2016**, *7*, 5032−5051.
(94) Verma, P.; Janesko, B. G.; Wang, Y.; He, X.; Scalmani, G.; Frisch, M. J.; Truhlar, D. G. M11plus: A Range-Separated Hybrid Meta Functional with Both Local and Rung-3.5 Correlation Terms and High Across-the-Broad Accuracy for Chemical Applications. *J. Chem. Theory Comput.* **2019**, *15*, 4804−4815.
(95) Kanchanakungwankul, S.; Verma, P.; Janesko, B. G.; Scalmani, G.; Frisch, M. J.; Truhlar, D. G. M11pz: A Nonlocal Meta Functional with Zero Hartree-Fock Exchange and with Broad Accuracy for Chemical Energies and Structures. *J. Chem. Theory Comput.* **2023**, *19*, 9102−9117.


https://doi.org/10.1021/acs.jpca.4c03137
*J. Phys. Chem. A* 2024, 128, 6412−6422