THE JOURNAL OF PHYSICAL CHEMISTRY A

Article

pubs.acs.org/JPCA

# Improving the Performance of Long-Range-Corrected Exchange-Correlation Functional with an Embedded Neural Network

Qin Liu,[†] JingChun Wang,[†] PengLi Du,[†] LiHong Hu,[‡] Xiao Zheng,*[,†] and GuanHua Chen[§]

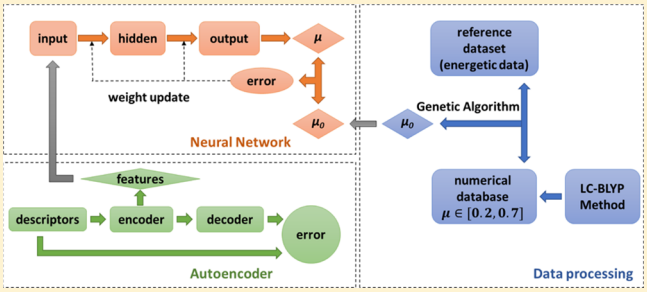
[†]Hefei National Laboratory for Physical Sciences at the Microscale & Synergetic Innovation Center of Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei, Anhui 230026, China

[‡]School of Computer Science and Information Technology, Northeast Normal University, Changchun, Jilin 130000, China

[§]Department of Chemistry, The University of Hong Kong, Pokfulam Road, Hong Kong, China

Supporting Information

**ABSTRACT:** A machine-learning-based exchange-correlation functional is proposed for general-purpose density functional theory calculations. It is built upon the long-range-corrected Becke−Lee−Yang−Parr (LC−BLYP) functional, along with an embedded neural network which determines the value of the range-separation parameter $\mu$ for every individual system. The structure and the weights of the neural network are optimized with a reference data set containing 368 highly accurate thermochemical and kinetic energies. The newly developed functional (LC−BLYP−NN) achieves a balanced performance for a variety of energetic properties investigated. It largely improves the accuracy of atomization energies and heats of formation on which the original LC−BLYP with a fixed $\mu$ performs rather poorly. Meanwhile, it yields a similar or slightly compromised accuracy for ionization potentials, electron affinities, and reaction barriers, for which the original LC−BLYP works reasonably well. This work clearly highlights the potential usefulness of machine-learning techniques for improving density functional calculations.




## 1. INTRODUCTION

Machine learning (ML),[1] a fundamental and essential approach to artificial intelligence, has prompted a wide range of applications. These include robotics,[2,3] automatic processing and translation of languages,[4−7] pattern recognition,[8,9] drug design,[10−12] and game playing.[13] Machine learning has also been increasingly adopted by studies in various fields of physics, chemistry, and materials science.[14−17]

In the field of theoretical chemistry, there have been many attempts to apply ML methods, particularly the neural network (NN) approach, to improve the accuracy of quantum-mechanical calculations/simulations on chemical systems. Being the most popular first-principles method, density functional theory (DFT) has been widely used in many fields of modern science. Despite their success, the present-day DFT methods have not achieved chemical accuracy for general cases. This is because the exact exchange-correlation (XC) functional is yet out of reach, while approximate functionals suffer from intrinsic errors.[18,19] Tracing and eliminating the errors of density functional approximations is thus an important goal of theoretical advancement, and ML could be a useful tool for achieving this goal.

In 2003 Hu and Chen et al. developed a DFT−NEURON approach[20] to improve the accuracy of heats of formation (HOF) of molecular species calculated by DFT methods. In their work, an artificial NN was constructed, whose weights were determined via a training process with a database containing 180 small- to medium-sized organic molecules. The optimized NN can then be used to correct the calculated HOF for new molecules. Such an approach was later generalized to improve density functional calculations for a variety of thermochemical properties, including Gibbs free energy, ionization potential (IP), electron affinity (EA), and absorption energy.[21] In 2014, the DFT−NEURON approach was further improved by incorporating statistical methods such as Kennard−Stone sampling and bootstrapping methods.[22] For a database containing 539 molecules, the NN correction reduces the mean absolute error (MAE) of HOF calculated by the hybrid Becke-3-Lee−Yang−Parr (B3LYP) functional[23,24] from more than 10 kcal/mol to about 1 kcal/mol, and hence achieves the chemical accuracy.[22] Recently, Hu and co-workers have established an NN to improve the description of noncovalent interactions by DFT methods.[25]

In 2007, Wu and Xu proposed the X1 method,[26] which uses an NN to improve the prediction of thermochemical properties of molecular systems by DFT methods. The X1 method adopts an extensive data set comprised of diversified chemical species to train the NN, along with a multipopulation genetic algorithm to provide a reasonable initialization for the optimization of synaptic weights.[27] Applications of X1 and its extensions have

**Received:** July 17, 2017
**Revised:** September 5, 2017
**Published:** September 6, 2017







DOI: 10.1021/acs.jpca.7b07045
J. Phys. Chem. A 2017, 121, 7273−7281

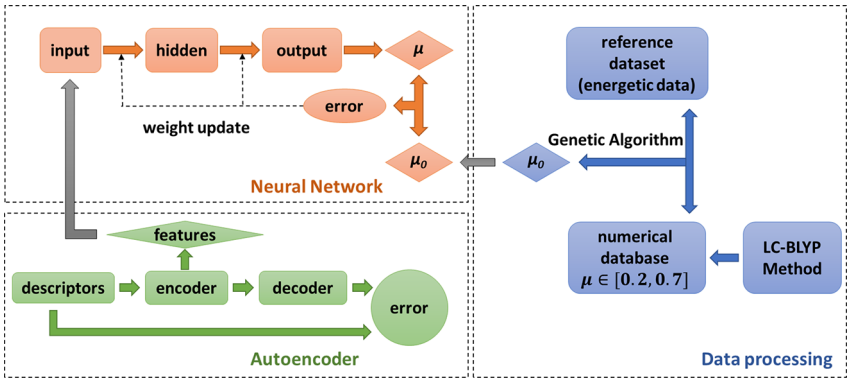



**Figure 1.** Schematic diagram illustrating the architecture of neural network and the work flow of training process.

led to substantial alleviation of size-dependent errors in the calculated HOF, bond dissociation enthalpies, and heats of isomerization.[28]

There have also been some efforts that aim to improve the approximate XC functionals with ML techniques. For instance, in 2004 some authors of the present work have constructed an NN to optimize the three parameters in B3LYP for every individual molecule, and the resulting ML-based B3LYP predicts more accurate thermochemical properties than the conventional B3LYP functional.[29] It is worth pointing out that, although the optimal parameter values are system-dependent, the form and weights of the NN are fixed for all systems, and thus the same NN can be universally applied to any molecular system. Recently, Snyder et al. have applied ML to determine the kinetic energy of a one-dimensional model electronic system, an achieves chemical accuracy.[30,31]

Even more ambitious strategies to combine ML with DFT methods have been proposed. Very recently, Smith et al. have demonstrated that a deep NN trained on DFT calculations can learn an accurate and transferable potential for organic molecules. Brockherde et al. have used ML to directly learn the density-potential and energy-density maps,[32,33] and thus bypass the need of solving Kohn−Sham equations.[34]

Despite the encouraging progress, ML-based DFT methods are yet far from mature, as they have not been applied to realistic systems for general-purpose usage. For instance, although our previously developed ML-based B3LYP functional[29] has demonstrated a promising performance, it has a few notable limitations: (1) the training data set was rather small, (2) the quality of descriptors was not systematically evaluated, and (3) the three parameters were optimized simultaneously within the same NN, and their relative significance was not analyzed. In view of the exciting development of ML during the past decade, we are keen to revisit the ML-based XC functional and address the above issues.

In this work, we choose to apply ML to the long-range-corrected (LC) functional. The range-separation scheme was originally proposed by Savin et al.,[35] based on which Hirao and co-workers have developed the LC functionals.[36−39] An LC functional includes a certain amount of long-range Hartree−Fock (HF) exchange, and thus gives rise to improved description of 4s−3d interconfigurational energies of the first-row transition metals,[39] polarizabilities of $\pi$-conjugated systems,[39] reaction barriers,[40] charge-transfer excitation energies and Rydberg excitation energies,[36,38] and van der Waals and $\pi$-aromatic interactions.[41,42] However, LC functionals are known to yield less accurate HOF.[40]

Most LC functionals involve a range-separation parameter $\mu$, which quantifies how the electron−electron interaction operator $\frac{1}{r}$ is decomposed into long-range and short-range parts (see Section2 for details). The $\mu$-parameter has been set to different values for calculation of different properties. For instance, $\mu$ has been set to 0.33 au[38,43−46] (atomic unit is used throughout this paper unless specified otherwise) to minimize the MAE of the internuclear distances of the first-to-third-row homonuclear diatomic molecules.[38] The value of $\mu$ has also been set to 0.47 au to minimize the error of calculated atomization energies (AEs).[40] Other values of 0.15,[47−49] 0.4,[39] and 0.5 au[50−52] have also been reported. Clearly, optimizing $\mu$ for a specific property is somewhat against the universal nature of the XC functional.

We consider the $\mu$-parameter to be a certain functional of electron density and it is thus system-dependent. We construct an NN that determines the most reasonable value of $\mu$ for every individual species, and the NN-predicted $\mu$ is then adopted to calculate all the thermochemical and kinetic properties of that species. Compared with our previous effort, we now use a more sophisticated NN, along with a more extensive data set and a group of carefully selected descriptors to perform the training. Again, although the predicted $\mu$ varies from system to system, the structure and weights of the NN are fixed for all systems, and thus the same NN can be universally applied to any molecular system.

The remainder of this paper is organized as follows: Section 2 elaborates the methodology adopted for constructing the ML-based functional. Specifically, details about the NN approach used and related computational techniques will be introduced. Numerical results are presented and discussed in Section 3. Finally, the concluding remarks are given in Section 4.

## 2. METHODOLOGY

**2.1. Long-Range Correlated Functional.** In the range-separation scheme of XC functional, the electron−electron interaction operator $\frac{1}{r}$ is split into short-range and long-range components as follows:[35,53]

$$\frac{1}{r} = \frac{1 - \mathrm{erf}(\mu r)}{r} + \frac{\mathrm{erf}(\mu r)}{r} \quad (1)$$

Here, $r = |\mathbf{r}_1 - \mathbf{r}_2|$ is the distance between electrons 1 and 2, $\mathrm{erf}(x)$ is the error function, and $\mu$ is the range parameter of our primary interest.

We choose to examine the LC−BLYP functional, for which the XC energy is expressed as



DOI: 10.1021/acs.jpca.7b07045
J. Phys. Chem. A 2017, 121, 7273−7281

$$E_{xc}^{LC-BLYP} = E_x^{LR-HF} + E_x^{SR-B88} + E_c^{LYP} \quad (2)$$

where $E_x^{LR-HF}$ is the long-range component of the HF exchange functional, $E_x^{SR-B88}$ is the short-range component of the Becke's generalized gradient approximation for exchange functional,[54] and $E_c^{LYP}$ is the Lee−Yang−Parr (LYP) correlation functional.[24]

**2.2. Machine-Learning Based Exchange-Correlation Functional.** *2.2.1. Overview of Neural Network Construction and Training.* Figure 1 is a schematic diagram that illustrates the architecture of NN and the work flow of training process. The NN constructed in this work is a three-layer back-propagation NN,[55] which consists of an input layer, a hidden layer, and an output layer. An unsupervised learning network, autoencoder,[56,57] is introduced at the pretraining stage (see the lower-left part of Figure 1). The autoencoder transforms the descriptors (information representing an individual species) of molecules to certain quantitative features, which are then fed to the input layer of NN. The autoencoder is also used to select the most informative descriptors and thus reduce the dimension of NN. The right part of Figure 1 illustrates the processing of data set. The reference data set contains 368 accurate energetic data of 362 molecules (details to be presented later). We also establish a numerical database to store the total energies of all the molecules using the LC−BLYP functional with $\mu$ varying from 0.2 to 0.7 au genetic algorithm (GA)[58−60] is employed to determine the most reasonable value of $\mu$ for each individual molecule by cross-searching the reference data set and the numerical database. The resulting system-specific $\mu$ values are taken as the target for the training of NN. Such a set of $\mu$ values are termed as the $\boldsymbol{\mu}_0$ set. Hereafter we use boldface to denote vectors.

The NN training process is illustrated in the upper-left part of Figure 1. The features of a certain molecule are fed to the NN at the input layer, and the $\mu$ value of that molecular is generated at the output layer. The NN-produced $\mu$ value is compared to its counterpart in the $\boldsymbol{\mu}_0$ set, and the deviation propagates backward from the output to the input layer. The synaptic weights connecting the neurons in different layers are updated accordingly. Then the next molecule is accessed by the NN. The above procedure is repeated until the averaged deviation between the NN-produced $\mu$ values and the reference $\boldsymbol{\mu}_0$ set is minimized for all species in the training data set.

*2.2.2. Reference Data Set and Numerical Database. Reference Data Set.* We use a reference data set containing 368 reliable thermochemical and kinetic energies, among which 300 energies are experimental data and 68 energies are obtained by highly accurate quantum chemistry calculations. These energetic data include 148 atomization energies (AEs), 51 IPs and 26 EAs from the G2/97 set,[61,62] 75 HOF from the G3/99 set,[62−64] and 36 hydrogen transfer reaction energy barriers (HTREBs) and 32 non-hydrogen transfer reaction energy barriers (NHTREBs) from the HTBH38/04 and NHTBH38/04 set, respectively.[65−67]

There are totally 362 atomic and molecular species associated with the reference data set, including 228 neutral molecules, 51 cations, 29 anions, 34 transition-state complexes, 14 atoms, and 6 molecular complexes bound by weak interactions. These chemical species are comprised of non-metallic elements H, B, C, N, O, F, Si, P, S, Cl and metallic elements Be, Li, Na, Al. The diversity of these chemical species is important for preserving the universality of ML-based functional.

*Numerical Database.* For numerical convenience, we split the training of NN into two steps: (1) determine the most reasonable value of $\mu$ for each individual molecule related to the reference data set; and (2) take these system-specific $\mu$ values (the $\boldsymbol{\mu}_0$ set) as the target of NN to determine the weights.

To accomplish the first step, we set up a database to store the calculated total energies of the 362 chemical species with the LC−BLYP functional. For each species, calculation is done for $\mu$ varying from 0.2 to 0.7 with an increment of 0.005, and the total energy at any other value of $\mu$ can be obtained via a linear interpolation based on the recorded data. The geometries of the 362 species are optimized at the B3LYP/6-311++g-(3df,3pd) level. Zero-point energies (ZPEs) are obtained through frequency analysis. The total energies and thermochemical properties are computed by single-point calculations at the LC-BLYP/6-311++g(3df,3pd) level.[68] All the DFT calculations are carried out by using the Gaussian09 suite of program.[69]

*System-Specific $\mu$ Values.* We now determine the most reasonable value of $\mu$ for each of the 362 species that are associated with the reference data set. This is done by employing the GA, which is a metaheuristic and adaptive algorithm based on the evolutionary ideas of natural selection.[70,71] Note that a molecule may be associated with more than one energetic data, and hence the determination of $\boldsymbol{\mu}_0$ is to reproduce accurately the reference energies data set in an average sense. Mathematically, the following criterion is adopted:

$$\boldsymbol{\mu}_0 = \arg\min_{\boldsymbol{\mu}} \sum_{i=1}^{368} |E_i^{DFT}(\boldsymbol{\mu}) - E_i^{ref}| \quad (3)$$

Here, $E_i^{ref}$ is the $i$th reference energy in the data set, and $E_i^{DFT}(\boldsymbol{\mu})$ is the same energy calculated using the LC−BLYP functional with the system-specific $\mu$ values. The GA results in a number of $\boldsymbol{\mu}_0$ sets that are mutually different but all accurately reproduce the energies in the reference data set. We save all these $\boldsymbol{\mu}_0$ sets for the subsequent training of NN.

Figure 2 depicts the distribution of system-specific $\mu$ values for a typical $\boldsymbol{\mu}_0$ set determined by the GA. Interestingly, the $\mu$ values are roughly normally distributed, with about 68% of data falling within the range of (0.35,0.55) and about 80% within (0.3,0.6) (in units of a.u.). LC−BLYP calculations adopting

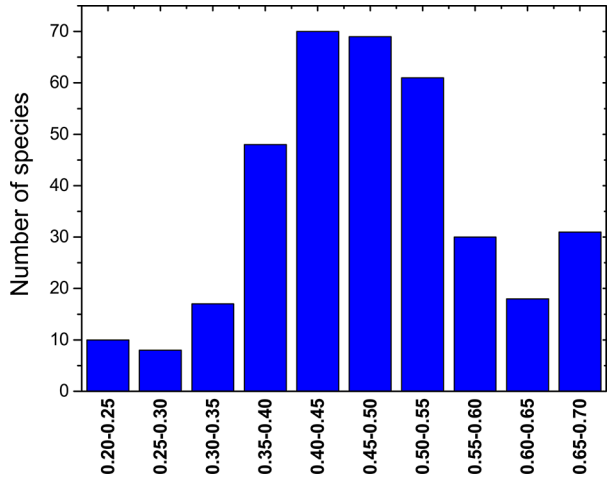



**Figure 2.** Distribution of the system-specific $\mu$ values determined by the GA.



DOI: 10.1021/acs.jpca.7b07045
J. Phys. Chem. A 2017, 121, 7273−7281

such a $\boldsymbol{\mu}_0$ set reproduce accurately the thermochemical and kinetic energies in the reference data set with an MAE of 0.31 kcal/mol, which is substantially reduced from the MAE of 12.24 kcal/mol by conventional LC−BLYP functional with $\mu$ = 0.47a.u. fixed for all chemical species.

*2.2.3. Selection of Descriptors. Descriptors.* In the context of NN, descriptors are essentially important as they uniquely define a sample in the data set. Since the goal of our present work is to find a mapping between the $\mu$-parameter and the electron density of a chemical species, the descriptors should characterize or represent to certain extent the electron density of the species.

Following previous studies,[20,22,28,29,72] we consider three categories of descriptors: (1) The first category includes the basic physical quantities of a species, such as the number of atoms $N_t$, the number of electrons $N_e$, the spin multiplicity $g_S$ = $2S + 1$ with $S$ being the total spin, and number of atoms for a specific element ($N_H$, $N_{Li}$, $N_{Be}$, $N_B$, $N_C$, $N_N$, $N_O$, $N_F$, $N_{Na}$, $N_{Al}$, $N_{Si}$, $N_P$, $N_S$, $N_{Cl}$). (2) The second category concerns some characteristic chemical properties of the species, such as the ZPE, the dipole moment (DM), the energy of highest occupied molecular orbital (HOMO), the energy of lowest unoccupied molecular orbital (LUMO), the energy gap $\epsilon_{gap}$, etc. These descriptors are obtained from DFT calculations at the LC−BLYP/6-311++g(3df,3pd) level with the preset $\mu$ = 0.47a.u. (3) The third category concerns the geometric structure of the species, which is expected to be highly related to electron density distribution. Geometry-based descriptors were absent in our previous effort,[29] and they are elaborated on below.

In 2012 Rupp et al. proposed to use a "Coulomb matrix" to represent molecules for the training of kernel-ridge regression models.[73] In 2016 Zhou et al. developed a linear regression model that uses several geometry-based descriptors to correct the DFT calculated HOF of hydrocarbons.[74] They have proposed five three-dimensional descriptors based on a modified Wiener index.[75,76] The five descriptors represent the atom−atom, atom−bond and bond−bond interactions within a molecule. In the present work, we adopt the two descriptors representing the atom−atom interactions, but drop the other three involving "bonds", because it could be difficult to define or specify (a priori) the chemical bonds in a general species.

Following Zhou et al.'s notation,[74] the two atom−atom interaction descriptors are defined by

$$W_{AA'} = \frac{1}{N_t}\sum_{i=1}^{N_t}\sum_{j<i}^{N_t}\frac{a_i a_j}{r_{ij}} \tag{4}$$

$$W_{AA} = \frac{1}{N_t}\sum_{i=1}^{N_t} a_i a_i \tag{5}$$

Here, $r_{ij}$ is the geometric distance between atoms $i$ and $j$, and $a_i$ is the Pauling electronegativity[77,78] of atom $i$. $W_{AA'}$ characterizes the pairwise interactions between different atoms, and it conveys the structural information on the species (via $r_{ij}$); while $W_{AA}$ describes the self-interactions of atoms. Different from their original definitions in ref 74, here both $W_{AA'}$ and $W_{AA}$ are averaged by $N_t$.

*Autoencoder.* The above 24 descriptors are proposed by experience or by physical/chemical intuition. To simplify the NN model and to avoid possible overfitting, it is necessary to assess these descriptors and select the most informative and representative ones.

The autoencoder has become a more widely used pretraining tool to extract principal features from descriptors and to reduce the dimension of NN model. The autoencoder adopted in this work is sketched in Figure 3, which consists of the encoder and

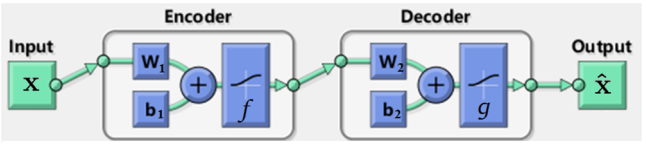



**Figure 3.** Structure and work flow of the autoencoder for the generation of features and selection of descriptors.

the decoder. The encoder transforms the descriptors $\mathbf{x}$ at the input layer to the features $\mathbf{h}$ at the hidden layer, and the decoder does an inverse transform from $\mathbf{h}$ to $\hat{\mathbf{x}}$ at the output layer, where $\hat{\mathbf{x}}$ is an attempted replica of $\mathbf{x}$. Mathematically, the encoding and decoding of descriptors are represented by the mappings $f$ and $g$ as follows.

$$\mathbf{h} = f(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1) \tag{6}$$

$$\hat{\mathbf{x}} = g(\mathbf{W}_2\mathbf{h} + \mathbf{b}_2) \tag{7}$$

Here, $f$ and $g$ are realized by the logistic sigmoid function $f(x)$ = $g(x) = 1/(1+e^{-x})$. $\mathbf{b}_1$ and $\mathbf{b}_2$ are the bias vectors introduced at the input and hidden layers, respectively. $\mathbf{W}_1$ ($\mathbf{W}_2$) is the weight matrix between the input (hidden) layer and hidden (output) layer. Through an unsupervised training, the weight matrices are adjusted until the reconstruction error $\mathcal{L}(\mathbf{x},\hat{\mathbf{x}}) = \|\mathbf{x} - \hat{\mathbf{x}}\|$ is minimized.

For each ($i$th) descriptor, we perform a linear regression between its values at the input ($\mathbf{x}_i$) and the duplicated values at the output ($\hat{\mathbf{x}}_i$) for all the M = 362 species associated with the reference data set. The coefficient of determination $R_i^2$ of the regression is computed as

$$R_i^2 = 1 - \frac{\sum_{k=1}^{M}(x_{ik} - \hat{x}_{ik})^2}{\sum_{k=1}^{M}(x_{ik} - \bar{x}_i)^2} \tag{8}$$

where $\bar{x}_i = \frac{1}{M}\sum_{k=1}^{M} x_{ik}$ is the mean of $\mathbf{x}_i$. Statistically, $R_i^2$ being close to 1 indicates that the information on descriptor is well preserved by the transformed features $\mathbf{h}$. In contrast, a small $R_i^2$ implies loss of information through the transforms.

Among all the 24 descriptors defined above, there are 10 descriptors which have an $R_i^2$ value larger than 0.89, while the $R_i^2$ for the rests are all smaller than 0.62. The 10 highly reproducible descriptors ($N_t$, $N_e$, $N_H$, $N_C$, $N_F$, $N_{Cl}$, ZPE, DM, $W_{AA'}$ and $W_{AA}$) are thus employed for the subsequent training of NN. We have manually selected various combinations of descriptors to train the NN. The results confirm that the above 10 descriptors yield the best training performance. All the numerical procedures related to the autoencoder and the NN are carried out with the Neural Network Toolbox of MATLAB.[79]

*2.2.4. Training of Neural Network.* The structure of NN for predicting the system-specific $\mu$ values is detailed in Figure 4. The input, hidden and output layers contain 10, 5, and 1 neurons, respectively. Given the limited size of the training set, further increasing the number of hidden layers or hidden neurons would result in an excessive number of weight parameters, which could easily lead to overfitting. The features ($\mathbf{h}$) of the chemical species related to the training data set are


DOI: 10.1021/acs.jpca.7b07045
J. Phys. Chem. A 2017, 121, 7273−7281

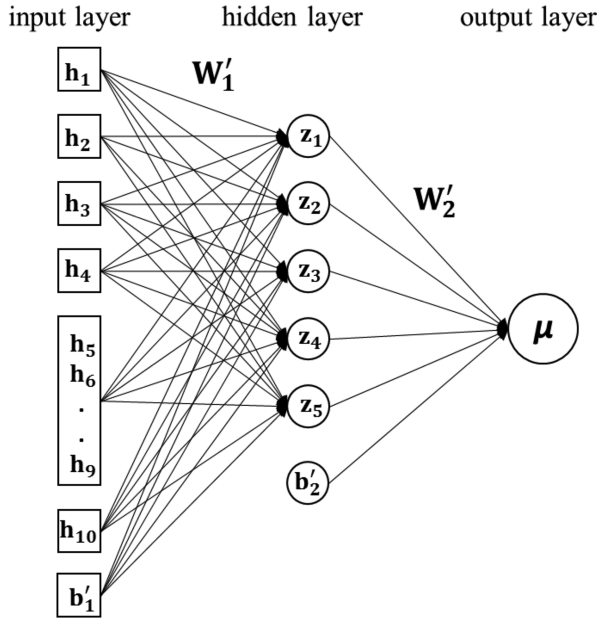



**Figure 4.** Structure and work flow of the NN for the prediction of system-specific $\mu$ for the LC−BLYP functional.

generated by the autoencoder and enter the NN at the input layer. The output of the NN, the $\mu$ value, is expressed as

$$\boldsymbol{\mu} = \tilde{g}(\mathbf{W}_2'[\tilde{f}(\mathbf{W}_1'\mathbf{h} + \mathbf{b}_1')] + \mathbf{b}_2') \quad (9)$$

Here, $\tilde{f}$ ($\tilde{g}$) is the transfer function which connects the input (hidden) layer to the hidden (output) layer, with $\mathbf{W}_1'$ ($\mathbf{W}_2'$) being the associated weight matrix and $\mathbf{b}_1'$ ($\mathbf{b}_2'$) the bias vector. We choose the nonlinear tan-sigmoid function as the neural transfer function, i.e., $\tilde{f}(x) = \tilde{g}(x) = 2/(1+e^{-2x}) - 1$. The use of a linear transfer function yields much more inferior training performance. The goal of training is to minimize the mean squared deviation between $\boldsymbol{\mu}$ of eq 9 and the reference $\boldsymbol{\mu}_0$ set by tuning the weights. Note that the value of $\mu$ obtained through eq 9 ranges from −1 to 1, and hence the $\boldsymbol{\mu}_0$ set is scaled linearly to the same range for making the comparison.

The Bayesian regularization method[80−82] is adopted to perform the training. For each training process, the data set is randomly divided into the training (85%) and testing (15%) subsets. The testing subset ensures that the final NN can be generalized to new species outside the reference data set.

The training of NN has been carried out numerous times with different initial weights and biases, different numbers of neurons, different combinations of descriptors, and different reference $\boldsymbol{\mu}_0$ sets. The minimal mean squared deviation between $\boldsymbol{\mu}$ and $\boldsymbol{\mu}_0$ achieved is about 0.004 au, which is considered to be a reasonably small value.

## 3. RESULTS AND DISCUSSION

We first examined how well the LC−BLYP functional with the NN-produced $\mu$ values could reproduce the thermochemical and kinetic energies in the reference data set. As has been mentioned in Section 2.2.2, the LC−BLYP functional with the GA-generated $\boldsymbol{\mu}_0$ set recovers accurately the energetic data with a rather small MAE of 0.31 kcal/mol. Even with the optimal training, the NN-produced $\boldsymbol{\mu}$ still differs from the $\boldsymbol{\mu}_0$ set with a mean squared deviation of 0.004 au. Such a deviation is responsible for the errors in the calculated energies.

Table 1 lists the MAEs of the thermochemical and kinetic energies calculated by the LC−BLYP with a fixed $\mu$ = 0.47a.u.

**Table 1. MAEs (in Units of kcal/mol) of the Various Types of Thermochemical and Kinetic Energies in the Reference Dataset[a]**

| energy | number | B3LYP | LC−BLYP | LC−BLYP−NN |
|---|---|---|---|---|
| AE | 148 | 3.10 | 13.33 | 5.49 |
| HTREB | 36 | 5.08 | 2.94 | 2.68 |
| HOF | 75 | 9.33 | 26.50 | 10.51 |
| NHTREB | 32 | 5.36 | 2.54 | 4.47 |
| IP | 51 | 3.86 | 5.18 | 5.27 |
| EA | 26 | 2.77 | 3.54 | 4.10 |
| overall | 368 | 4.85 | 12.24 | 6.02 |

[a]DFT results obtained from the B3LYP functional, the original LC−BLYP functional with a fixed $\mu$ = 0.47a.u., and the LC−BLYP with the NN-produced system-specific $\mu$ values are presented.

and with the NN-produced $\boldsymbol{\mu}$ (the latter will be termed as LC−BLYP−NN hereafter). We also present the MAEs for the results calculated by the B3LYP functional for comparison. For all the 368 energetic data, the LC−BLYP−NN yields an MAE of 6.02 kcal/mol, which is greatly reduced from the MAE of 12.24 kcal/mol with the original LC−BLYP. The overall performance of the LC−BLYP−NN approaches that of the B3LYP.

We now have a closer examination on the errors for each type of energies. The AE and the HOF are intrinsically equivalent, as they both characterize the energy change upon the formation of a molecule from its constituent atoms. The original LC−BLYP behaves rather poorly on AE and HOF, and it is on these two properties that the LC−BLYP−NN achieves a substantial improvement in accuracy. By contrast, the original LC−BLYP works reasonably well for IP and EA and even better for the kinetic energies (HTREB and NHTREB), while the performances of the LC−BLYP−NN are similar or slightly compromised as compared to those of the original LC−BLYP. Moreover, it is interesting to find that the NN effectively balances the accuracy among the various types of energies.

We further look into the error distribution of the calculated energetic data. Figure 5 depicts the distribution of unsigned

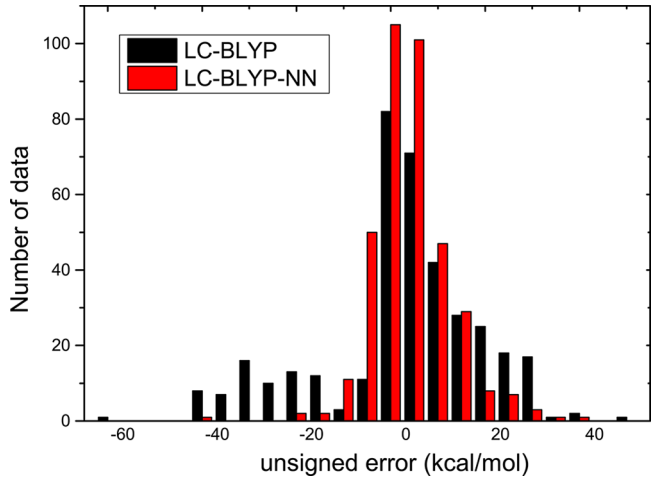



**Figure 5.** Distribution of the unsigned errors $\Delta E_i$ for the 368 energetic data in the reference data set. The increment of error is 5 kcal/mol.



DOI: 10.1021/acs.jpca.7b07045
J. Phys. Chem. A 2017, 121, 7273−7281

errors ($\Delta E_i \equiv E_i^{DFT} - E_i^{ref}$) for the original LC−BLYP and the LC−BLYP−NN. The original LC−BLYP gives rise to a rather large spread of errors, with a long tail extending into the negative error regime. The maximal negative error is −61.58 kcal/mol (the HOF of $C_2F_6$). In contrast, with the NN-produced $\mu$ values, the error distribution becomes roughly normal and much narrower. With the original LC−BLYP 56% of the calculated data have errors smaller than 10 kcal/mol and 42% are within the range of 5 kcal/mol; while with the LC−BLYP−NN the corresponding ratios increase substantially to 82% and 56%, respectively. Therefore, it is apparent that the application of NN indeed enhances the performance of the LC−BLYP in the statistical sense.

For practical usage, it is crucial that the LC−BLYP−NN functional is capable of producing the same level of accuracy for chemical species beyond the reference data set. To verify this point, we adopt a "prediction" data set that has no overlap with the reference data set. The prediction data set contains 13 HOF from the G3/05 set,[63,83] 19 IPs and 16 EAs associated with the auxiliary neutral subset of the G2/97 set,[62,84] and the heats of reaction (HOR) for 14 Diels−Alder reactions given in Figure 1 of ref 85. The prediction data set involves 106 chemical species (including 71 neutral molecules, 19 cations, and 16 anions) which have no overlap with the species associated with the reference data set.

We then employ the established LC−BLYP−NN functional to reproduce the energetic data in the prediction data set. Numerical calculations proceed as follows: (1) Evaluate the 10 descriptors ($N_t$, $N_e$, $N_H$, $N_C$, $N_F$, $N_{Cl}$, ZPE, DM, $W_{AA'}$ and $W_{AA}$) of the species associated with the prediction data set. (2) Feed the descriptors to the optimized autoencoder and generate the features of the molecular species. (3) Import the features to the trained NN, which outputs the system-specific $\mu$ values. (4) Carry out LC−BLYP calculations for the thermochemical energies with the system-specific $\mu$ values.

Table 2 lists the MAEs of the calculated energetic data in the prediction data set. Evidently, the errors of HOF, IP, and EA resulted from the original LC−BLYP and the LC−BLYP−NN are overall similar to those in the reference data set: the system-specific $\mu$ values largely improve the HOF while preserving roughly the same level of accuracy for IP and EA. Therefore, the errors are more balanced among the various types of energies with the use of NN. This is clearly verified by the error distribution shown in Figure 6. In particular, the original LC−BLYP much too low HOF for the molecules $C_6F_6$ and $C_6F_5Cl$ with rather large unsigned errors of −79.97 kcal/mol and −69.47 kcal/mol, respectively; while with the LC−BLYP−NN the above errors are reduced to −4.58 kcal/mol and −6.16 kcal/mol, respectively.

**Table 2. MAEs (in Units of kcal/mol) of the Various Types of Thermochemical Energies in the Prediction Dataset**[a]

| Energy | Number | B3LYP | LC−BLYP | LC−BLYP−NN |
|---|---|---|---|---|
| HOF | 13 | 7.06 | 21.71 | 10.68 |
| IP | 19 | 4.15 | 3.27 | 4.02 |
| EA | 16 | 2.51 | 3.10 | 3.49 |
| HOR | 14 | 13.71 | 1.13 | 6.52 |
| Overall | 62 | 6.50 | 6.61 | 5.85 |

[a]DFT results obtained from the B3LYP, the original LC−BLYP with a fixed $\mu$ = 0.47a.u., and the LC−BLYP with the NN-produced system-specific $\mu$ values are presented.

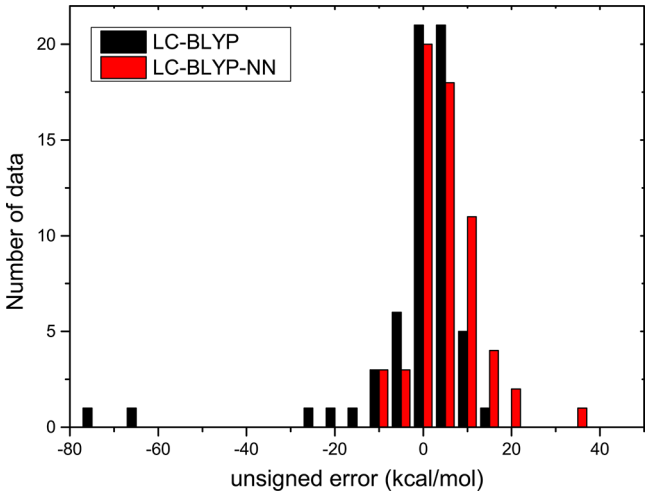



**Figure 6.** Distribution of the unsigned errors $\Delta E_i$ for the 62 energetic data in the prediction data set. The increment of error is 5 kcal/mol.

We include a new type of thermochemical energies in the prediction data set, the HOR for the formation of a group of cyclic and cagelike molecules.[85] Such molecules are distinctly different from the species related to the reference data set. The original LC−BLYP yields a remarkably high accuracy for the HOR with a very small MAE of 1.13 kcal/mol, while the LC−BLYP−NN results in a somewhat compromised accuracy with an MAE of 6.52 kcal/mol. Nevertheless, the latter MAE is still much smaller than that with the B3LYP functional.

The consistent behavior of the LC−BLYP−NN on the training and prediction data sets affirms that the applicability of the NN constructed in this work can be robustly extended to systems beyond the training data set, which is important not only for ensuring the predictive power of the NN, but also for preserving the universality of the XC functional.

Through our extensive numerical tests, it is found that many aspects of the NN, such as the size and diversity of the reference data set, the representability of descriptors, and the efficacy of pretraining and training algorithms, affect significantly the performance of the resulting functional. The present work excels our previous effort[29] in all these aspects. Nevertheless, it is apparent that there is still plenty of room for improvement, considering the fact that the GA-generated $\boldsymbol{\mu}_0$ set yields an MAE as small as 0.31 kcal/mol for the reference data set.

## 4. CONCLUDING REMARKS

To summarize, in this work we propose an ML-based XC functional, the LC−BLYP−NN functional, for general-purpose DFT calculations. The ML-based functional is built upon the explicit form of the conventional LC−BLYP functional, but with a system-specific range-separation parameter $\mu$ which is determined by a carefully designed and well-trained NN.

The newly constructed ML-based functional is tested extensively on a variety of thermochemical and kinetic energies. The LC−BLYP−NN achieves a balanced performance for the various types of properties investigated. It largely improves the accuracy of AE and HOF on which the original LC−BLYP with a fixed $\mu$ performs rather poorly. Meanwhile, it yields a similar or slightly compromised accuracy for IP, EA and reaction barriers, for which the original LC−BLYP works reasonably well.


DOI: 10.1021/acs.jpca.7b07045
J. Phys. Chem. A 2017, 121, 7273−7281

Usually, the inaccurate or occasionally the problematic results of DFT calculations can be traced to the intrinsic errors associated with the approximate forms of XC functional. For instance, it is well-known that the local density approximation or the generalized gradient approximations (GGAs) suffer from the delocalization error,[85−88] and hence they tend to predict too low energies for delocalized electron distributions.[18,19,89] In contrast, the HF exchange functional is often subjected to the localization error which artificially favors localized electron distributions. The LC−BLYP functional includes both the short-range GGA and the long-range HF exchange components and hence the delocalization and localization errors partly cancel out. It has been found that the cancellation of errors can be rather system-dependent,[90,91] and the NN-predicted $\mu$ value thus reflects the optimal amount of mixing between the GGA and HF exchange that is distinct for each individual species.

Despite its promising performance, the LC−BLYP−NN is definitely not the ultimate functional that we seek, because of some intrinsic issues associated with its functional form. For instance, the system-dependent $\mu$ could lead to breakdown of size-consistency condition, since $\mu$ of a total system may not agree with the $\mu$'s of the involving subsystems. Moreover, mainstream approximate functionals have many sources of errors. Besides the delocalization and localization errors mentioned above, density functional approximations also suffer from the self-interaction error,[92,93] and the dynamic and static correlation errors.[94] Systematic removal or alleviation of these errors is extremely challenging, which would require a functional form that is more sophisticated than the range-separation scheme. For instance, the LC−BLYP functional underestimates (although much less severely than the LDA and GGAs) the energies of $H_2^+$ and $He_2^+$ molecules at large internuclear distances,[100] because the delocalization and localization errors associated with the short-range and long-range exchange parts do not cancel completely. The LC−BLYP−NN developed in this work does not improve this situation, since the presently adopted training set does not include energetic data explicitly representing the delocalization/localization error (such as the deviation between integer and derivative gaps, or the energy of systems with fractional electron numbers).[89] This will be a subject of our future work. In recent years, a variety of new functional forms have been proposed, such as the local hybrid hyper-GGA functionals,[95] doubly hybrid functionals,[96−98] and scaling corrected functionals.[99,100] Combining the ML approaches with these new functional forms is expected to lead to a major boost in the performance of DFT methods. Work along this direction is in progress.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpca.7b07045.

Details on the training and prediction of neural network; optimized weight matrices and bias vectors (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Author

*E-mail: xz58@ustc.edu.cn.

### ORCID

Xiao Zheng: 0000-0002-9804-1833


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The support from the Ministry of Science and Technology (Grant Nos. 2016YFA0400900 and 2016YFA0200600), the National Natural Science Foundation of China (Grant Nos. 21573202 and 21233007), the Fundamental Research Funds for the Central Universities (Grant No. 2340000074), and the SuperComputing Center of USTC is gratefully acknowledged.

## ■ REFERENCES


(1) Michalski, R. S.; Carbonell, J. G.; Mitchell, T. M. *Machine Learning: An Artificial Intelligence Approach*; Springer Science & Business Media: Berlin, 2013.

(2) Dorigo, M.; Schnepf, U. Genetics-Based Machine Learning and Behavior-Based Robotics: a New Synthesis. *IEEE Trans. Syst., Man, Cybern., Syst.* **1993**, *23*, 141−154.

(3) Klingspor, V.; Demiris, J.; Kaiser, M. Human-Robot Communication and Machine Learning. *Appl. Artif. Intell.* **1997**, *11*, 719−746.

(4) Kalchbrenner, N.; Blunsom, P. Recurrent Continuous Translation Models. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Washington, 18−21 October, 2013; Association for Computational Linguistics: Stroudsburg, PA, 2013; pp 1700−1709

(5) Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv:1409.0473*, 2014.

(6) Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv:1406.1078v3*, 2013.

(7) Goldberg, Y. A. Primer on Neural Network Models for Natural Language Processing. *arXiv:1510.00726*, 2015.

(8) MacLeod, N.; Benfield, M.; Culverhouse, P. Time to Automate Identification. *Nature* **2010**, *467*, 154−155.

(9) Chittka, L.; Dyer, A. Cognition: Your Face Looks Familiar. *Nature* **2012**, *481*, 154−155.

(10) Ashley, E. A. The Precision Medicine Initiative: A New National Effort. *JAMA* **2015**, *313*, 2119−2120.

(11) Schatz, M. C.; Langmead, B. The DNA Data Deluge. *IEEE Spectrum* **2013**, *50*, 28−33.

(12) Ding, H.; Takigawa, I.; Mamitsuka, H.; Zhu, S. Similarity-Based Machine Learning Methods for Predicting Drug-Target Interactions: A Brief Review. *Briefings Bioinf.* **2014**, *15*, 734−747.

(13) Silver, D.; Huang, A.; Maddison, C. J.; et al. Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature* **2016**, *529*, 484−489.

(14) Biamonte, J.; Wittek, P.; Pancotti, N.; Rebentrost, P.; Wiebe, N.; Lloyd, S. Quantum Machine Learning. *arXiv:1611.09347*, 2016.

(15) Dunjko, V.; Taylor, J. M.; Briegel, H. J. Quantum-Enhanced Machine Learning. *Phys. Rev. Lett.* **2016**, *117*, 130501.

(16) Wodrich, M. D.; Corminboeuf, C. Reaction Enthalpies using the Neural-Network-Based X1 Approach: The Important Choice of Input Descriptors. *J. Phys. Chem. A* **2009**, *113*, 3285−3290.

(17) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(18) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Insights into Current Limitations of Density Functional Theory. *Science* **2008**, *321*, 792−794.

(19) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Challenges for Density Functional Theory. *Chem. Rev.* **2012**, *112*, 289−320.

(20) Hu, L.; Wang, X.; Wong, L.; Chen, G. Combined First-Principles Calculation and Neural-Network Correction Approach for Heat of Formation. *J. Chem. Phys.* **2003**, *119*, 11501−11507.

(21) Wang, X.; Wong, L.; Hu, L.; Chan, C.; Su, Z.; Chen, G. Improving the Accuracy of Density-Functional Theory Calculation:




DOI: 10.1021/acs.jpca.7b07045
J. Phys. Chem. A 2017, 121, 7273−7281

The Statistical Correction Approach. *J. Phys. Chem. A* **2004**, *108*, 8514−8525.
(22) Sun, J.; Wu, J.; Song, T.; Hu, L.; Shan, K.; Chen, G. Alternative Approach to Chemical Accuracy: A Neural Networks-Based First-Principles Method for Heat of Formation of Molecules Made of H, C, N, O, F, S, and Cl. *J. Phys. Chem. A* **2014**, *118*, 9120−9131.
(23) Becke, A. D. Becke's Three Parameter Hybrid Method using the LYP Correlation Functional. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(24) 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: Condens. Matter Mater. Phys.* **1988**, *37*, 785.
(25) Gao, T.; Li, H.; Li, W.; Li, L.; Fang, C.; Li, H.; Hu, L.; Lu, Y.; Su, Z.-M. A Machine Learning Correction for DFT Non-Covalent Interactions Based on the S22, S66 and X40 Benchmark Databases. *J. Cheminf.* **2016**, *8*, 24.
(26) Wu, J.; Xu, X. The X1Method for Accurate and Efficient Prediction of Heats of Formation. *J. Chem. Phys.* **2007**, *127*, 214105.
(27) Wu, J.; Xu, X. Accurate Prediction of Heats of Formation by a Combined Method of B3LYP and Neural Network Correction. *J. Comput. Chem.* **2009**, *30*, 1424−1444.
(28) Wu, J.; Zhou, Y.; Xu, X. The X1 Family of Methods that Combines B3LYP with Neural Network Corrections for an Accurate yet Efficient Prediction of Thermochemistry. *Int. J. Quantum Chem.* **2015**, *115*, 1021−1031.
(29) Zheng, X.; Hu, L.; Wang, X.; Chen, G. A Generalized Exchange-Correlation Functional: the Neural-Networks Approach. *Chem. Phys. Lett.* **2004**, *390*, 186−192.
(30) Snyder, J. C.; Rupp, M.; Hansen, K.; Müller, K.-R.; Burke, K. Finding Density Functionals with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 253002.
(31) Snyder, J. C.; Rupp, M.; Hansen, K.; Blooston, L.; Müller, K.-R.; Burke, K. Orbital-Free Bond Breaking via Machine Learning. *J. Chem. Phys.* **2013**, *139*, 224104.
(32) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8*, 3192−3203.
(33) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133.
(34) Brockherde, F.; Vogt, L.; Li, L.; Tuckerman, M. E.; Burke, K.; Müller, K.-R. By-Passing the Kohn-Sham Equations with Machine Learning. *arXiv:1609.02815*, 2016.
(35) Leininger, T.; Stoll, H.; Werner, H. J.; Savin, A. Combining Long-Range Configuration Interaction with Short-Range Density Functionals. *Chem. Phys. Lett.* **1997**, *275*, 151−160.
(36) Chiba, M.; Tsuneda, T.; Hirao, K. Excited State Geometry Optimizations by Analytical Energy Gradient of Long-Range Corrected Time-Dependent Density Functional Theory. *J. Chem. Phys.* **2006**, *124*, 144106.
(37) Song, J.-W.; Tsuneda, T.; Sato, T.; Hirao, K. Calculations of Alkane Energies using Long-Range Corrected DFT Combined with Intramolecular van der Waals Correlation. *Org. Lett.* **2010**, *12*, 1440−1443.
(38) Tawada, Y.; Tsuneda, T.; Yanagisawa, S.; Yanai, T.; Hirao, K. A Long-Range-Corrected Time-Dependent Density Functional Theory. *J. Chem. Phys.* **2004**, *120*, 8425−8433.
(39) Iikura, H.; Tsuneda, T.; Yanai, T.; Hirao, K. A Long-Range Correction Scheme for Generalized-Gradient-Approximation Exchange Functionals. *J. Chem. Phys.* **2001**, *115*, 3540−3544.
(40) Song, J.-W.; Hirosawa, T.; Tsuneda, T.; Hirao, K. Long-Range Corrected Density Functional Calculations of Chemical Reactions: Redetermination of Parameter. *J. Chem. Phys.* **2007**, *126*, 154105.
(41) Sato, T.; Tsuneda, T.; Hirao, K. A Density-Functional Study on π-Aromatic Interaction: Benzene Dimer and Naphthalene Dimer. *J. Chem. Phys.* **2005**, *123*, 104307.
(42) Kamiya, M.; Tsuneda, T.; Hirao, K. A Density Functional Study of van der Waals Interactions. *J. Chem. Phys.* **2002**, *117*, 6010−6015.
(43) Peach, M. J.; Helgaker, T.; Sałek, P.; Keal, T. W.; Lutnæs, O. B.; Tozer, D. J.; Handy, N. C. Assessment of a Coulomb-Attenuated Exchange-Correlation Energy Functional. *Phys. Chem. Chem. Phys.* **2006**, *8*, 558−562.
(44) 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.
(45) Sato, T.; Tsuneda, T.; Hirao, K. Van Der Waals Interactions Studied by Density Functional Theory. *Mol. Phys.* **2005**, *103*, 1151−1164.
(46) Kamiya, M.; Sekino, H.; Tsuneda, T.; Hirao, K. Nonlinear Optical Property Calculations by the Long-Range-Corrected Coupled-Perturbed Kohn-Sham Method. *J. Chem. Phys.* **2005**, *122*, 234111.
(47) Heyd, J.; Scuseria, G. E. Assessment and Validation of a Screened Coulomb Hybrid Density Functional. *J. Chem. Phys.* **2004**, *120*, 7274−7280.
(48) Heyd, J.; Scuseria, G. E. Efficient Hybrid Density Functional Calculations in Solids: Assessment of the Heyd-Scuseria-Ernzerhof Screened Coulomb Hybrid Functional. *J. Chem. Phys.* **2004**, *121*, 1187−1192.
(49) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. Erratum: Hybrid Functionals Based on a Screened Coulomb Potential [J. Chem. Phys. 118, 8207 (2003)]. *J. Chem. Phys.* **2006**, *124*, 219906.
(50) Gerber, I. C.; Angyán, J. G. Potential Curves for Alkaline-Earth Dimers by Density Functional Theory with Long-Range Correlation Corrections. *Chem. Phys. Lett.* **2005**, *416*, 370−375.
(51) Angyán, J. G.; Gerber, I. C.; Savin, A.; Toulouse, J. Van Der Waals Forces in Density Functional Theory: Perturbational Long-Range Electron-Interaction Corrections. *Phys. Rev. A: At., Mol., Opt. Phys.* **2005**, *72*, 012510.
(52) Gerber, I. C.; Angyán, J. G. Hybrid Functional with Separated Range. *Chem. Phys. Lett.* **2005**, *415*, 100−105.
(53) Adamson, R. D.; Dombroski, J. P.; Gill, P. M. Efficient Calculation of Short-Range Coulomb Energies. *J. Comput. Chem.* **1999**, *20*, 921−927.
(54) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098.
(55) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning Representations by Back-Propagating Errors. *Nature* **1986**, *323*, 533−536.
(56) Bengio, Y. Learning Deep Architectures for AI. *FNT in Machine Learning* **2009**, *2*, 1−127.
(57) Hinton, G. E.; Salakhutdinov, R. R. Reducing the Dimensionality of Data with Neural Networks. *Science* **2006**, *313*, 504−507.
(58) Goldberg, D. E. *Genetic Algorithms in Search, Optimization & Machine Learning*; Addison-Wesley Longman Publishing Co., Inc.: Boston, MA, 1989.
(59) Conn, A. R.; Gould, N. I. M.; Toint, P. L. A Globally Convergent Augmented Lagrangian Algorithm for Optimization with General Constraints and Simple Bounds. *SIAM J. Numer. Anal.* **1991**, *28*, 545−572.
(60) Conn, A. R.; Gould, N. I. M.; Toint, P. L. A Globally Convergent Augmented Lagrangian Barrier Algorithm for Optimization with General Inequality Constraints and Simple Bounds. *Math. Comp.* **1997**, *66*, 261−288.
(61) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Assessment of Gaussian-2 and Density Functional Theories for the Computation of Enthalpies of Formation. *J. Chem. Phys.* **1997**, *106*, 1063−1079.
(62) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. Assessment of Gaussian-2 and Density Functional Theories for The Computation of Ionization Potentials and Electron Affinities. *J. Chem. Phys.* **1998**, *109*, 42−55.
(63) 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.
(64) Redfern, P. C.; Zapol, P.; Curtiss, L. A.; Raghavachari, K. Assessment of Gaussian-3 and Density Functional Theories for Enthalpies of Formation of C1- C16 Alkanes. *J. Phys. Chem. A* **2000**, *104*, 5850−5854.


DOI: 10.1021/acs.jpca.7b07045
*J. Phys. Chem. A* 2017, 121, 7273−7281

(65) 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.
(66) 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.
(67) 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 Noncovalent Interactions. *J. Chem. Theory Comput.* **2006**, *2*, 364−382.
(68) Ochterski, J. W. *Thermochemistry in Gaussian*; http://gaussian.com/thermo/, 2000; pp 1−19.
(69) Frisch, M. J; Trucks, G. W.; Schlegel, H. B; Scuseria, G. E; Robb, M. A; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. et al., *Gaussian 09*, revision C.01; Gaussian, Inc.: Wallingford, CT, 2009.
(70) Mitchell, M. *An Introduction to Genetic Algorithms*; MIT Press: Cambridge, MA, 1996.
(71) Sivanandam, S.; Deepa, S. *Introduction to Genetic Algorithms*; Springer Science & Business Media: Berlin, 2007.
(72) Wu, J.; Ying Zhang, I.; Xu, X. The X1s Method for Accurate Bond Dissociation Energies. *ChemPhysChem* **2010**, *11*, 2561−2567.
(73) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; Von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.
(74) Zhou, Y.; Wu, J.; Xu, X. Improving B3LYP Heats of Formation with Three-Dimensional Molecular Descriptors. *J. Comput. Chem.* **2016**, *37*, 1175−1190.
(75) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17−20.
(76) Yang, F.; Wang, Z.-D.; Huang, Y.-P. Modification of the Wiener Index 4. *J. Comput. Chem.* **2004**, *25*, 881−887.
(77) Pauling, L. The Nature of the Chemical Bond. IV. The Energy of Single Bonds and the Relative Electronegativity of Atoms. *J. Am. Chem. Soc.* **1932**, *54*, 3570−3582.
(78) Allred, A. Electronegativity Values from Thermochemical Data. *J. Inorg. Nucl. Chem.* **1961**, *17*, 215−221.
(79) Demuth, H.; Beale, M. *Neural Network Toolbox for use with Matlab*; The MathWorks, Inc.: Natick, MA, 1995.
(80) Burden, F.; Winkler, D. In *Artificial Neural Networks: Methods and Applications*; Livingstone, D. J., Ed.; Springer: Totowa, NJ, 2009; Vol. *458*; pp 23−42.
(81) MacKay, D. J. Bayesian Interpolation. *Neural Comput.* **1992**, *4*, 415−447.
(82) Foresee, F. D.; Hagan, M. T. Gauss-Newton Approximation to Bayesian Learning. *IEEE* **1997**, 1930−1935.
(83) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Assessment of Gaussian-3 and Density-Functional Theories on the G3/05 Test Set of Experimental Energies. *J. Chem. Phys.* **2005**, *123*, 124107.
(84) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. Gaussian-2 Theory for Molecular Energies of First-and Second-Row Compounds. *J. Chem. Phys.* **1991**, *94*, 7221−7230.
(85) Johnson, E. R.; Mori-Sánchez, P.; Cohen, A. J.; Yang, W. Delocalization Errors in Density Functionals and Implications for Main-Group Thermochemistry. *J. Chem. Phys.* **2008**, *129*, 204112.
(86) Yang, W.; Zhang, Y.; Ayers, P. W. Degenerate Ground States and a Fractional Number of Electrons in Density and Reduced Density Matrix Functional Theory. *Phys. Rev. Lett.* **2000**, *84*, 5172−5175.
(87) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Fractional Charge Perspective on the Band Gap in Density-Functional Theory. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *77*, 115123.
(88) Mori-Sánchez, P.; Cohen, A. J.; Yang, W. Localization and Delocalization Errors in Density Functional Theory and Implications for Band-Gap Prediction. *Phys. Rev. Lett.* **2008**, *100*, 146401.
(89) Zheng, X.; Li, C.; Zhang, D.; Yang, W. Scaling Correction Approaches for Reducing Delocalization Error in Density Functional Approximations. *Sci. China: Chem.* **2015**, *58*, 1825−1844.
(90) Tsuneda, T.; Song, J.-W.; Suzuki, S.; Hirao, K. On Koopmans' Theorem in Density Functional Theory. *J. Chem. Phys.* **2010**, *133*, 174101.
(91) Baer, R.; Livshits, E.; Salzner, U. Tuned Range-Separated Hybrids in Density Functional Theory. *Annu. Rev. Phys. Chem.* **2010**, *61*, 85−109.
(92) Perdew, J. P.; Zunger, A. Self-Interaction Correction to Density-Functional Approximations for Many-Electron Systems. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1981**, *23*, 5048−5079.
(93) Vydrov, O. A.; Scuseria, G. E. Effect of the Perdew-Zunger Self-Interaction Correction on the Thermochemical Performance of Approximate Density Functionals. *J. Chem. Phys.* **2004**, *121*, 8187−8193.
(94) Mori-Sánchez, P.; Cohen, A. J.; Yang, W. Discontinuous Nature of the Exchange-Correlation Functional in Strongly Correlated Systems. *Phys. Rev. Lett.* **2009**, *102*, 066403.
(95) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E.; Staroverov, V. N.; Tao, J. Exchange and Correlation in Open Systems of Fluctuating Electron Number. *Phys. Rev. A: At., Mol., Opt. Phys.* **2007**, *76*, 040501.
(96) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. Doubly Hybrid Meta DFT: New Multi-Coefficient Correlation and Density Functional Methods for Thermochemistry and Thermochemical Kinetics. *J. Phys. Chem. A* **2004**, *108*, 4786−4791.
(97) Grimme, S. Semiempirical Hybrid Density Functional with Perturbative Second-Order Correlation. *J. Chem. Phys.* **2006**, *124*, 034108.
(98) Zhang, Y.; Xu, X., III; Goddard, W. A. Doubly Hybrid Density Functional for Accurate Descriptions of Nonbond Interactions, Thermochemistry, and Thermochemical Kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 4963−4968.
(99) Zheng, X.; Cohen, A. J.; Mori-Sánchez, P.; Hu, X.; Yang, W. Improving Band Gap Prediction in Density Functional Theory from Molecules to Solids. *Phys. Rev. Lett.* **2011**, *107*, 026403.
(100) Li, C.; Zheng, X.; Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Local Scaling Correction for Reducing Delocalization Error in Density Functional Approximations. *Phys. Rev. Lett.* **2015**, *114*, 053001.


DOI: 10.1021/acs.jpca.7b07045
*J. Phys. Chem. A* 2017, 121, 7273−7281