Title	Accurate prediction of heat of formation by combining Hartree-Fock/density functional theory calculation with linear regression correction approach
Author(s)	Duan, XM; Song, GL; Li, ZH; Wang, XJ; Chen, GH; Fan, KN
Citation	Journal Of Chemical Physics, 2004, v. 121 n. 15, p. 7086-7095
Issued Date	2004
URL	http://hdl.handle.net/10722/42057
Rights	Journal of Chemical Physics. Copyright © American Institute of Physics.

Accurate prediction of heat of formation by combining Hartree–Fock/density functional theory calculation with linear regression correction approach

Xue-Mei Duan, Guo-Liang Song, and Zhen-Hua Li
Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Department of Chemistry, Center for Chemical Physics, Fudan University, Shanghai 200433, China

Xiu-Jun Wang
Department of Chemistry, University of Hong Kong, Hong Kong, China

Guan-Hua Chen
Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Department of Chemistry, Center for Chemical Physics, Fudan University, Shanghai 200433, China and Department of Chemistry, University of Hong Kong, Hong Kong, China

Kang-Nian Fan[a)]
Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Department of Chemistry, Center for Chemical Physics, Fudan University, Shanghai 200433, China

(Received 15 April 2004; accepted 6 July 2004)

A linear regression correction approach has been developed successfully to account for the electron correlation energy missing in Hartree–Fock calculation and to reduce the calculation errors of density functional theory. The numbers of lone-pair electrons, bonding electrons and inner layer electrons in molecules, and the number of unpaired electrons in the composing atoms in their ground states are chosen to be the most important physical descriptors to determine the correlation energy unaccounted by Hartree–Fock method or to improve the results calculated by B3LYP density functional theory method. As a demonstration, this proposed linear regression correction approach has been applied to evaluate the standard heats of formation ΔH_f^{Θ} of 180 small-sized to medium-sized organic molecules at 298.15 K. Upon correction, the mean absolute deviation for the 150 molecules in the training set decreases from 351.0 to 4.6 kcal/mol and 360.9 to 4.6 kcal/mol for HF/6-31G(d) and HF/6-311+G(d,p) methods, respectively. For B3LYP method, the mean absolute deviations are reduced from 9.2 and 18.2 kcal/mol to 2.7 and 2.4 kcal/mol for 6-31G(d) and 6-311+G(d,p) basis sets, respectively. © *2004 American Institute of Physics.*
[DOI: 10.1063/1.1786582]

I. INTRODUCTION

From the solution of hydrogen atom in the 1920s to the simulations of complex systems nowadays, quantum chemistry has evolved into a major subject in chemistry. A variety of *ab initio* methods have been developed to calculate accurately various molecule properties such as thermochemical properties. The calculated properties are often comparable to experimental measurements, and occasionally even better than the experimental counterparts. Moreover, quantum mechanical calculation can be used to examine the physical properties or processes that are inaccessible by experiments.

Among such new algorithms, the Gaussian-2 (G2) (Refs. 1–3) and Gaussian-3 (G3) (Refs. 4 and 5) theories of Pople and co-workers have been proved very successful in calculating thermochemical properties of molecules, such as heats of formation, atomization energies, ionization potentials, and electron affinities. For example, for the 222 heats of formation in the G3/99 test set, the mean absolute deviations of G3 method is just 1.05 kcal/mol.[6] Nevertheless, these methods are only applicable to small-sized systems, generally including 5 to 6 heavy atoms. Furthermore, all these new approaches include high level correlation methods, such as MP3, MP4, and QCISD(T), in combination with very large basis sets, such as 6-311+G($3df,2p$), G3Large, cc-PVQZ, and cc-PV5Z, and thus require large computer resources. For large and even medium-sized molecules, reliable predictions are still beyond computational power. Therefore, designing economical schemes is highly desirable.

Compared with conventional *ab initio* electron correlation methods, density functional theory (DFT) methods can be applied to much larger molecules with less computation effort. However, the results of DFT calculations are not as accurate for large molecular systems as for small systems, in particular, their calculation errors increase with increasing molecular size.[7,8] Recently, Chen *et al.* proposed a neural-network scheme (DFT-NEURON) to correct the systematic errors of B3LYP method to calculate heats of formation of organic molecules.[8] In their scheme, the size of the molecule N_t (the total number of atoms in a molecule) is explicitly included as an inputting physical descriptor. Other descriptors used are the calculated ΔH_f^{Θ}, ZPE (zero-point vibra-

[a)]Electronic mail: knfan@fudan.edu.cn

tional energy), and N_{db} (the number of double bonds). The results are promising. For B3LYP/6-311+G(d,p) and B3LYP/6-311+G($3df,2p$) methods, the root-mean-square (rms) deviations of the calculated heats of formation for 180 organic molecules are reduced from 21.4 kcal/mol to 3.1 kcal/mol and from 12.0 kcal/mol to 3.3 kcal/mol upon the neural-network correction, respectively. However, this neural-network correction scheme cannot improve the Hartree–Fock (HF) results to the same accuracy with the same physical descriptors. On the other hand, Allinger and co-workers used a different approach,[9–13] in which the bond/group equivalents, number of bonds in the molecule with low-energy rotational barriers, statistical mechanical corrections, and other descriptors were used to empirically correct the HF or DFT electronic energies to obtain heats of formation of organic molecules. The accuracy of the scheme using DFT energies reaches to 0.36 kcal/mol for alkanes.[14] Although the results are very accurate, the fact that the schemes depend much on the intuition of the researcher prevents their further application to molecules with irregular bonds and its automatic implement. Furthermore, it is not easy to extract the source of error inherent in a method from the physical descriptors employed in these schemes.

DFT methods already include the bulk of electron correlation. On the other hand, HF method contains no electron correlation at all. Physical descriptors previously used in the neural-network approach might not work well with the HF method.[8] In the Gaussian-*n* series of methods, the higher-level correction employs the number of valence electron pairs and unpaired electrons in molecules and atoms as descriptors to correct the remaining high order correlation energy.[1–7] The validity of this approach is based on the fact that the majority of the correlation energy comes from the interaction between the electrons which occupy the same molecule orbital, i.e., paired electrons. Inspired by the success of the Gaussian-*n* series of methods, we intend to use the number of electrons instead of the total number of atoms as physical descriptor for electron correlation in the new schemes to correct HF and DFT energies.

In the present work, we propose a linear regression correction (LRC) algorithm to calculate the energies of molecules. The standard heat of formation ΔH_f^{Θ} at 298.15 K is chosen as the property of interest. In this linear regression correction algorithm, the numbers of electrons in different bonding environments are employed as the physical descriptors to correct the systematic deviations of HF and DFT calculations.

II. DESCRIPTION OF LINEAR REGRESSION CORRECTION APPROACH

The central idea in our linear regression correction is to introduce the numbers of electrons in different bonding environments into the energy expression to account for the correlation energy or higher order energy corrections neglected by the HF or DFT methods. For a molecule M ($A_{n_A}B_{n_B}\cdots$),

$$E_e^{\mathrm{LRC}}(M)=E_e^{\mathrm{calc.}}+\sum_i a_i x_i, \tag{1}$$

$$E_0=E_e^{\mathrm{LRC}}(M)+cZPE, \tag{2}$$

and for atom A,

$$E_e^{\mathrm{LRC}}(A)=E_e^{\mathrm{calc.}}+\sum_i b_i x_i. \tag{3}$$

In the above equations, $E_e^{\mathrm{calc.}}$ is the calculated electronic energy by HF and DFT methods without any correction, E_e^{LRC} is the electronic energy after linear regression correction, $\{x_i\}$ are physical descriptors, $\{a_i\}$ and $\{b_i\}$ are coefficients of the descriptors for the molecule and the atoms, respectively, and c is the scaling factor for ZPE. The descriptors considered in the present study are the following: (1) the number of lone-pair electrons in molecules, (2) the number of bonding electrons in molecules, (3) the number of inner layer (core) electrons in molecules, and (4) the number of unpaired electrons for ground state atoms. The inner layer electrons are further divided into several subsets according to the shell they belong to in the corresponding atoms. There are just three inner layers (shells) for the core electrons in the molecules studied here. We define N-1, N-2, and N-3 as the first, the second, and the third layer below the valence shell. The number of the unpaired electrons of the molecules is not included because all the molecules selected in the present study are closed shell.

In this work, heat of formation is selected as the property of interest. The raw calculation of ΔH_f^{Θ} is based on atomization energy scheme.[15] By definition, the heat of formation of a molecule M ($A_{n_A}B_{n_B}\cdots$) is the enthalpy change of the following reaction:

$$n_{A,S}A_{(S)}+n_{B,S}B_{(S)}+\cdots=M(A_{n_A}B_{n_B}\cdots),$$

where $n_{A,S}$ is the molar ratio of the element A in the molecule (M) to that in its stable state of aggregation at 298.15 K, i.e., its standard state, the subscript "S" represents standard state. For example, CO_2 is formed from graphite and gaseous O_2 molecule; in this case, both of $n_{C,S}$ and $n_{O,S}$ equal 1. In the atomization energy scheme, heat of formation of M at 298.15 K can be written as

$$\begin{aligned}\Delta H_f^{298\ \mathrm{K}}=&\left[E_0(M)-\sum_A n_A E_e(A_{(g)})\right]+\Delta H_{298\ \mathrm{K}}^{\mathrm{calc.}}(M)\\&+\sum_A n_A \Delta H_{f,0\ \mathrm{K}}^{\mathrm{exp}}(A_{(g)})\\&-\sum_A n_{A,S}\Delta H_{298\ \mathrm{K}}^{\mathrm{exp}}(A_{(S)}),\end{aligned} \tag{4}$$

where $\Delta E_0(M)$ is the energy of M at 0 K, $\Delta H_{298\ \mathrm{K}}^{\mathrm{calc.}}(M)$ is the calculated enthalpy change of M from 0 K to 298.15 K, $\Delta H_{f,0\ \mathrm{K}}^{\mathrm{exp}}(A_{(g)})$ is the experimental heat of formation of atom A in gaseous state at 0 K, and $\Delta H_{298\ \mathrm{K}}^{\mathrm{exp}}(A_{(S)})$ is the experimental enthalpy change of element A in its standard state from 0 K to 298.15 K. Introducing Eqs. (1)–(3) into Eq. (4), we arrive at

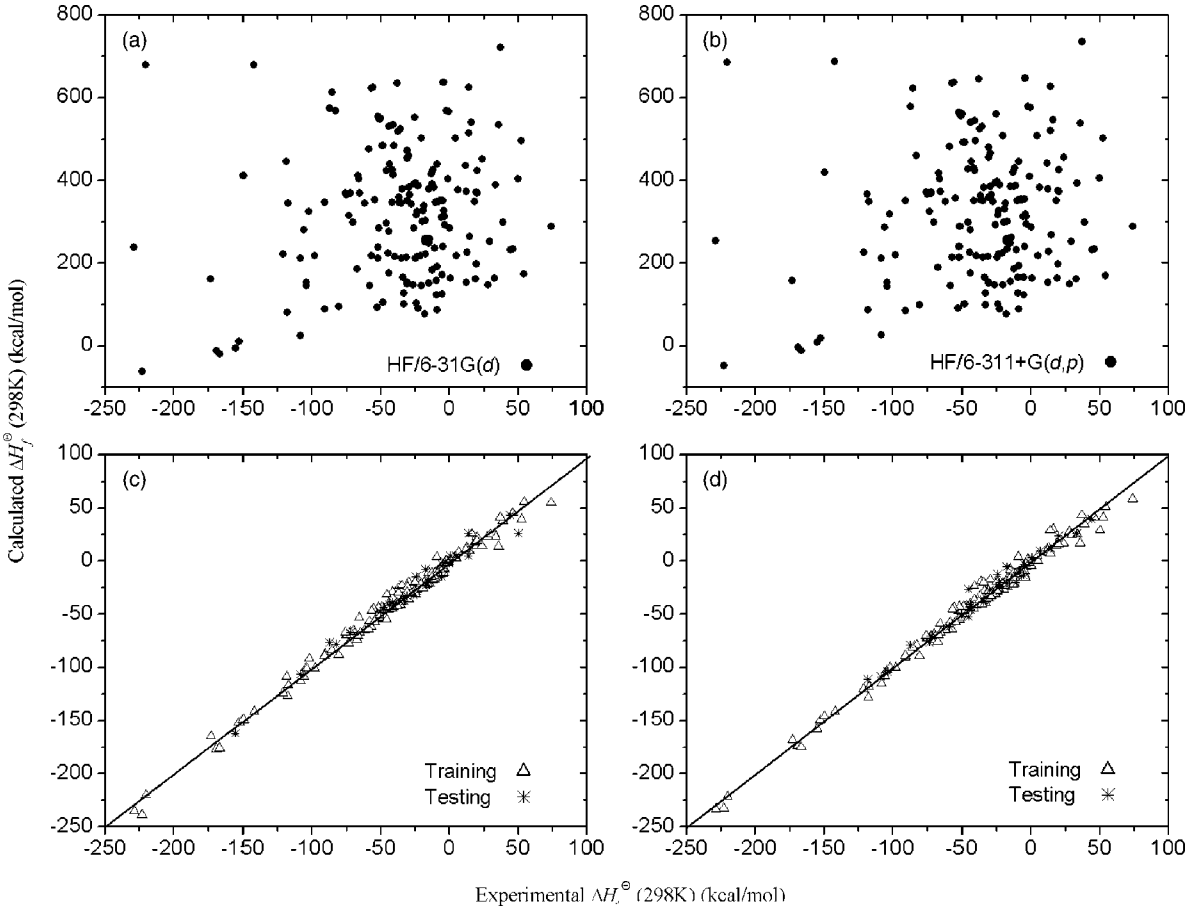

FIG. 1. Experimental ΔH_f^{Θ} vs HF calculated ΔH_f^{Θ} for all 180 organic molecules. (a) and (b) are the comparisons of the experimental ΔH_f^{Θ} to their raw HF/6-31G(d) and HF/6-311+G(d,p) results, respectively. (c) and (d) are the comparisons of the experimental ΔH_f^{Θ} to their linear regression corrected HF/6-31G(d) and HF/6-311+G(d,p) results, respectively.

$$\Delta H_f^{298\ \mathrm{K}} = \left[\sum_i a_i x_i - \sum_i b_i x_i + cZPE + \Delta E_e^{\mathrm{calc.}}\right] + \Delta H_{298\ \mathrm{K}}^{\mathrm{calc.}}(M) + \sum_A n_A \Delta H_{f,0\ \mathrm{K}}^{\mathrm{exp}}(A_{(g)}) - \sum_A n_{A,S} \Delta H_{298\ \mathrm{K}}^{\mathrm{exp}}(A_{(S)}) \quad (5)$$

in which

$$\Delta E_e^{\mathrm{calc.}} = E_e(M) - \sum_A n_A E_e(A). \quad (6)$$

Then, the linear regression correction scheme using Eq. (5) effectively corrects the unbalanced electron correlation energy calculation in atoms and in molecule by a theoretical method.

In order to determine the relative importance of the ith physical descriptor, its coefficient of partial correlation V_j is calculated:

$$V_j = \sqrt{1 - q/Q_j}, \quad j = 1,2,...,m, \quad (7)$$

in which q is the square sum of deviations, Q_j is the square sum of deviations leaving out one descriptor x_j, and m is the number of descriptors. The closer V_j is to 1, the more remarkable is the influence of x_j.

III. COMPUTATIONAL DETAILS

The molecule set selected to train and test our linear regression correction scheme is the same as Ref. 8 except two molecules, which contains 180 small- or medium-sized organic molecules whose ΔH_f^{Θ} are taken from Refs. 16–18. All the molecules are neutral. The heaviest molecule contains 14 heavy atoms, and the largest has 32 atoms including hydrogen atoms. We divide these molecules randomly into a training set with 150 molecules, and a testing set with 30 molecules. Equilibrium structures are obtained at the HF and B3LYP[19–22] level with 6-31G(d) or 6-311+G(d,p) basis set. Harmonic vibrational frequencies are calculated at the same level of theory. Equation (4) is employed for the ΔH_f^{Θ} raw calculation, where ZPEs calculated by HF/6-31G(d), HF/6-311+G(d,p), B3LYP/6-31G(d), and B3LYP/6-311+G(d,p) methods are unscaled. The scaling factors of the vibrational frequencies for $\Delta H_{298\ \mathrm{K}}^{\mathrm{calc.}}(M)$ calculation are 0.8905 and 0.9989 for HF/6-31G(d) and B3LYP/6-31G(d), respectively,[23] while 0.9135 and 0.9806 for the calculation of ZPEs in Eq. (5), respectively. For HF/6-311+G(d,p), the scaling factors we used are 0.8951 for $\Delta H_{298\ \mathrm{K}}^{\mathrm{calc.}}(M)$ calculation and 0.9248 for ZPE calculation in Eq. (5), both of which come from those of HF/6-311G(d,p).[23] For B3LYP/6-311+G(d,p), the scaling factors of B3LYP/6-31G(d) are employed for $\Delta H_{298\ \mathrm{K}}^{\mathrm{calc.}}(M)$ and ZPE calculations in Eq. (5). The

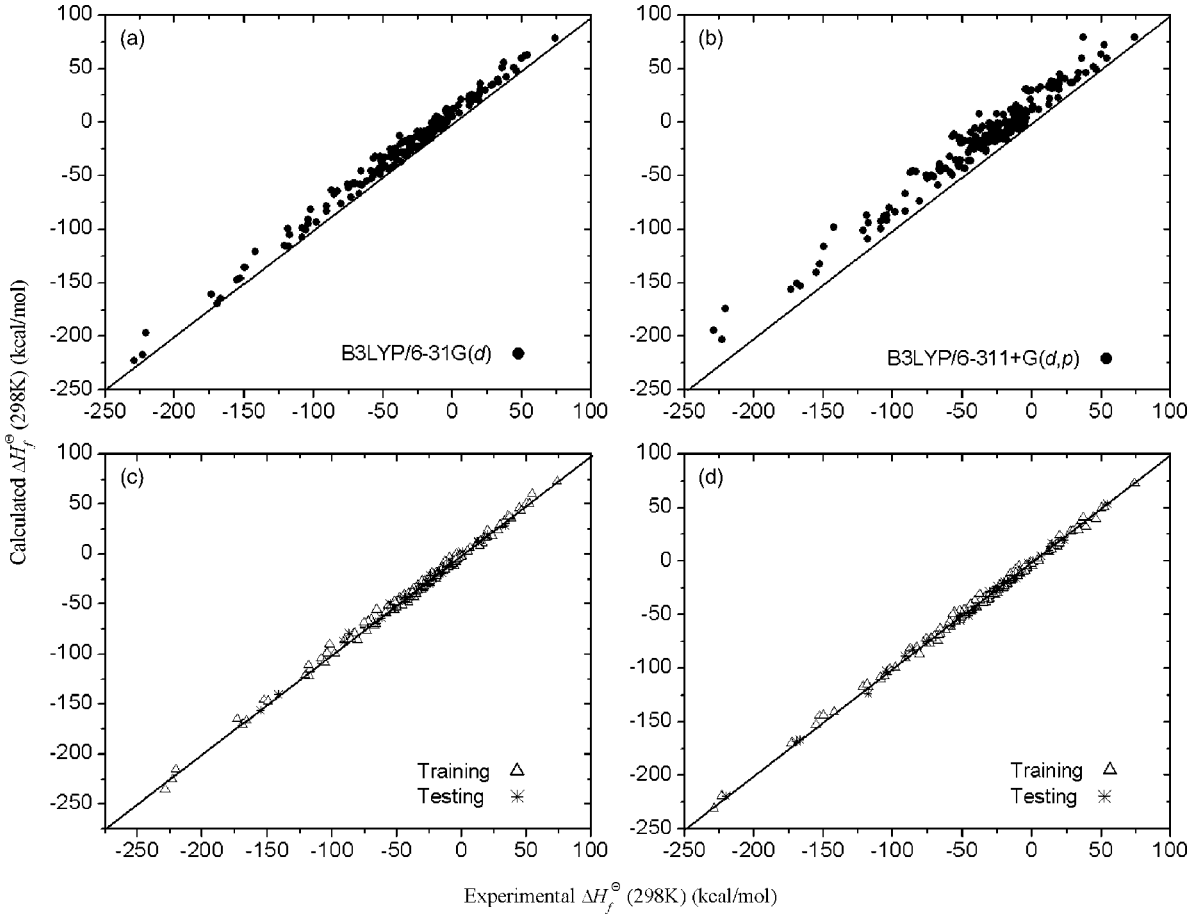

FIG. 2. Experimental ΔH_f^{Θ} vs B3LYP calculated ΔH_f^{Θ} for all 180 organic molecules. (a) and (b) are the comparisons of the experimental ΔH_f^{Θ} to their raw B3LYP/6-31G(d) and B3LYP/6-311+G(d,p) results, respectively. (c) and (d) are the comparisons of the experimental ΔH_f^{Θ} to their linear regression corrected HF/6-31G(d) and HF/6-311+G(d,p) results, respectively.

use of fixed scaling factors do have slight impact on calculated $\Delta H_{298\,\mathrm{K}}^{\mathrm{calc.}}(M)$ and ZPEs. However, the errors caused are very small in comparison to other correction terms, and can be further corrected by the linear regression approach, and thus are not the central concern of this study.

All the calculations in the present study have been done with the GAUSSIAN 98 suite of programs.[24]

IV. ASSESSMENT OF LINEAR REGRESSION CORRECTION APPROACH

The electronic energies and zero-point energies calculated by HF and B3LYP methods with 6-31G(d) and 6-311+G(d,p) basis sets are available on the website.[25] The raw calculated ΔH_f^{Θ} compared to their experimental counterparts are illustrated in Figs. 1(a), 1(b), 2(a), and 2(b) for the HF/6-31G(d), HF/6-311+G(d,p), B3LYP/6-31G(d), and B3LYP/6-311+G(d,p) methods, respectively. The diagonal lines imply that the calculated and experimental ΔH_f^{Θ} are equal. Figure 1(a) shows that the raw calculated ΔH_f^{Θ} of HF/6-31G(d) distribute randomly with very large deviations from experimental values. Same trend observed by other researchers,[7,8] i.e., the larger the molecule is, the greater is the deviation, has also been found in the present study for both HF and DFT methods. The maximal deviation of HF/6-31G(d) reaches 899.4 kcal/mol for $C_{10}H_{18}O_4$, and 821.1 kcal/mol for $C_{10}H_{20}O_2$. For small molecules such as CH_4 and C_2H_2, the deviations are slightly smaller, 65.4 kcal/mol and 118.9 kcal/mol, respectively. The mean absolute deviation of HF/6-31G(d) is 351.0 kcal/mol. The raw HF/6-311+G(d,p) calculation data also distribute randomly [see Fig. 1(b)]. The mean absolute deviation of raw HF/6-311+G(d,p) calculation data is 360.9 kcal/mol. In comparison, the raw ΔH_f^{Θ} results given by the B3LYP/6-31G(d) and B3LYP/6-311+G(d,p) methods are much better than those given by HF methods. The calculated ΔH_f^{Θ} by the B3LYP methods are mostly above the dashed line, implying that most calculated ΔH_f^{Θ} are larger than experimental values. The mean absolute deviation is 9.2 kcal/mol for the B3LYP/6-31G(d) method, and 18.2 kcal/mol for the B3LYP/6-311+G(d,p) method. The deviations of DFT results are much smaller than those of HF; however, they are still too large for any practical purpose. To improve the accuracy of DFT calculation, further correction to its energy is necessary.

Equation (5) employing the six descriptors as described previously is used to correct the heat of formation. The results are collected in Table I. The linear regression coefficients obtained for the descriptors are listed in Table II, while the mean absolute deviations (MADs) and rms deviations of

TABLE I. Experimental and theoretical deviations of linear regression corrected ΔH_f^{Θ} (298 K) for 180 molecules (unit kcal/mol).

Molecular formula	Molecular name	Deviations of ΔH_f^{Θ} (298 K) (Expt.–Theory)				
		Expt.[a]	HF–LRC1[b]	HF–LRC2[c]	DFT–LRC1[d]	DFT–LRC2[e]
$CBrCl_3$	Bromotrichloromethane	−9.3	−13.3	−13.6	−6.1	−4.9
$CBrF_3$[f,g]	Bromotrifluoromethane	−155.1	6.6	2.9	1.3	−2.5
$CClF_3$[h]	Chlorotrifluoromethane	−169.2	7.5	4.8	1.1	−1.1
CClN[i]	Cyanogen chloride	33.0	7.8	6.5	−0.3	0.1
CCl_2O	Phosgene	−52.3	−6.5	−5.4	−0.9	1.5
CF_2O	Carbonyl fluoride	−152.7	−0.7	−3.1	−7.2	−8.6
CF_4	Carbon tetrafluoride	−223.0	15.6	9.3	2.3	−3.8
$CHCl_3$[g]	Chloroform	−24.2	−6.4	−4.9	−2.7	0.1
CHF_3[h]	Trifluoromethane	−166.7	9.3	8.0	−0.1	−0.2
CH_2Cl_2	Dichloromethane	−22.8	−0.6	0.6	0.4	2.3
CH_2F_2	Difluoromethane	−108.2	4.6	7.0	−1.4	2.4
CH_2O_2	Formic acid	−90.5	−3.3	−0.1	−4.8	−0.2
CH_3Br	Methyl bromide	−9.0	1.2	2.5	0.2	1.8
CH_3NO_2[f,i,h]	Nitromethane	−17.9	−10.0	−12.5	2.7	1.4
CH_3NO_2[g]	Methyl nitrite	−15.3	−4.6	−7.1	3.7	0.8
CH_4[i]	Methane	−17.9	1.6	2.6	2.7	3.7
CH_4O	Methanol	−48.1	−4.1	1.0	−4.4	1.6
CH_4S	Methyl mercaptan	−5.5	5.8	6.3	0.9	1.0
CH_5N	Methylamine	−5.5	4.7	7.6	−0.6	3.8
COS[f]	Carbonyl sulfide	−33.1	4.3	2.6	1.2	0.2
CS_2	Carbon disulfide	28.0	4.9	0.2	4.1	0.0
C_2H_2[h]	Acetylene	54.2	−1.5	3.3	−5.8	0.9
$C_2H_2Cl_2$[f,i,g]	1,1-dichloroethylene	0.6	−4.0	−2.9	−1.8	1.1
$C_2H_2F_2$	1,1-dinoroethylene	−80.5	7.8	8.7	5.1	6.2
$C_2H_2O_4$	Oxalic acid	−173.0	−8.5	−5.0	−8.1	−3.4
C_2H_3Br	Vinyl bromide	18.7	0.6	2.5	0.1	2.7
C_2H_3ClO	Acetyl chloride	−58.3	−1.7	−0.7	0.9	2.6
$C_2H_3ClO_2$[h]	Chloroacetic acid	−104.3	−3.9	−1.5	−5.9	−1.9
$C_2H_3Cl_3$	1,1,1-trichloroethane	−34.0	−7.4	−6.2	−3.0	−0.6
C_2H_3F	Vinyl fluoride	−33.2	2.9	6.0	1.3	5.0
C_2H_4[h]	Ethylene	12.5	1.6	4.5	0.4	3.8
$C_2H_4Br_2$	1,2-dibromoethane	−9.3	1.1	0.7	−0.4	−0.2
$C_2H_4Cl_2$	1,1-dichloroethane	−31.0	−0.1	0.9	1.6	3.3
$C_2H_4Cl_2$	1,2-dichloroethane	−31.0	2.8	3.6	2.7	4.0
$C_2H_4F_2$[h]	1,1-dinoroethane	−118.0	8.8	10.1	3.6	5.6
C_2H_4O	Ethylene oxide	−12.6	−1.7	−1.7	0.3	0.6
$C_2H_4O_2$[i]	Acetic acid	−103.9	−2.9	−0.1	−5.0	−0.7
C_2H_4S	Thiacyclopropane	19.7	5.4	3.4	2.0	0.0
C_2H_5Br[f]	Bromoethane	−15.3	2.2	2.5	1.7	2.3
C_2H_5Cl[g]	Ethyl chloride	−26.7	2.9	3.6	3.2	4.2
C_2H_5N	Ethyleneimine	29.5	4.5	4.7	0.1	1.3
C_2H_5NO[h]	Acetamide	−57.0	5.1	7.3	−1.6	3.0
$C_2H_5NO_2$[f,i]	Nitroethane	−24.2	−9.2	−11.7	4.6	2.3
$C_2H_5NO_3$	Ethyl nitrate	−36.8	−13.2	−18.1	6.5	1.4
C_2H_6	Ethane	−20.2	3.1	3.7	3.9	4.6
C_2H_6O[g]	Dimethyl ether	−44.0	1.6	2.3	2.9	3.0
C_2H_6S	Dimethyl sulfide	−9.0	6.7	5.0	2.9	1.0
C_2H_7N	Dimethylamine	−4.5	7.2	7.9	3.2	4.3
C_2H_7N[i,h]	Ethylamine	−11.0	6.2	8.8	1.0	5.1
$C_2H_8N_2$	Ethylenediamine	−4.1	6.8	11.3	−4.6	2.7
C_2N_2	Cyanogen	73.8	18.9	15.5	1.2	0.9
C_3H_3NO	Oxazole	−3.7	3.8	1.2	1.6	−0.8
C_3H_4[f,i]	Methylacetylene	44.3	1.0	4.8	−2.0	3.1
C_3H_4	Propadiene	45.9	0.6	4.6	2.7	6.8
$C_3H_4O_3$	Ethylene carbonate	−121.2	2.9	−0.9	0.1	−4.0
$C_3H_5Cl_3$[i]	1,2,3-trichloropropane	−44.4	−4.5	−3.6	−4.1	−2.3
C_3H_6[g,h]	Cyclopropane	12.7	1.5	1.9	1.5	2.7
C_3H_6	Propylene	4.9	2.1	4.5	1.9	4.6
$C_3H_6Br_2$	1,2-dibromopropane	−17.4	0.6	−1.7	0.1	−1.6
$C_3H_6Cl_2$	1,2-dichloropropane	−39.6	1.2	1.8	1.7	2.9
C_3H_6O[g]	Acetone	−52.0	2.6	3.3	1.7	2.9
$C_3H_6O_2$	Methyl acetate	−98.0	2.7	1.8	1.6	1.1
$C_3H_6O_2$[f,i]	Propionic acid	−108.4	−1.7	0.4	−3.8	−0.3
C_3H_6S[i,h]	Thiacyclobutane	14.6	5.0	2.2	1.0	−1.9

TABLE I. (*Continued*).

Molecular formula	Molecular name	Deviations of ΔH_f^{Θ} (298 K) (Expt.–Theory)				
		Expt.[a]	HF–LRC1[b]	HF–LRC2[c]	DFT–LRC1[d]	DFT–LRC2[e]
C_3H_7Br	1-bromopropane	−21.0	1.8	1.6	1.3	1.3
C_3H_7Br[i]	2-bromopropane	−23.2	1.8	1.0	1.9	1.3
C_3H_7Cl[g]	Isopropyl chloride	−35.0	1.9	2.4	2.5	3.4
C_3H_7Cl	*n*-propyl chloride	−31.1	3.7	4.0	4.0	4.6
C_3H_7F	1-fluoropropane	−67.2	6.9	8.9	4.7	7.2
C_3H_7NO[i,h]	*N*,*N*-dimethylformamide	−45.8	8.5	6.5	5.9	4.6
$C_3H_7NO_2$	1-nitropropane	−29.8	−9.7	−12.5	4.0	1.3
$C_3H_7NO_2$	2-nitropropane	−34.5	−11.8	−14.6	2.0	−0.6
$C_3H_7NO_3$	Propyl nitrate	−41.6	−13.0	−18.2	6.6	1.1
$C_3H_7NO_3$[i]	Isopropyl nitrate	−45.7	−14.3	−19.6	5.9	0.2
C_3H_8[g]	Propane	−24.8	3.6	3.8	4.5	4.8
C_3H_8O[h]	Methyl ethyl ether	−51.7	2.3	2.7	3.7	3.5
C_3H_8S	*n*-propyl mercaptan	−16.2	5.4	5.4	0.8	0.4
C_3H_8S	Isopropyl mercaptan	−18.2	4.4	4.5	0.4	0.2
C_3H_8S	Ethyl methyl sulfide	−14.3	6.6	4.6	3.0	0.8
C_3H_9N[f,g]	*n*-propylamine	−17.3	4.9	7.1	−0.2	3.3
C_3H_9N	Isopropylamine	−20.0	4.2	6.4	−0.3	3.0
C_3H_9N[f,i]	Trimethylamine	−5.7	8.8	7.3	6.2	4.5
$C_3H_{10}N_2$[h]	1,2-propanediamine	−12.8	4.9	8.9	−5.4	0.9
$C_4H_4N_2$[f]	Succinonitrile	50.1	23.7	20.9	−0.2	0.1
C_4H_6	1,2-butadiene	38.8	0.9	3.9	3.2	6.5
C_4H_6O[f]	Divinyl ether	−3.3	−2.8	0.0	0.1	2.7
C_4H_8[f]	1-butene	−0.2	1.9	3.9	1.8	3.9
C_4H_8O	Isobutyraldehyde	−51.5	1.6	1.7	0.8	1.3
$C_4H_8O_2$	Ethyl acetate	−105.9	3.2	2.0	2.3	1.1
C_4H_9Br[g,h]	1-bromobutane	−25.7	2.1	1.6	1.7	1.3
C_4H_9Cl	*Tert*-butyl chloride	−43.8	−0.4	0.0	0.7	1.4
$C_4H_{10}O$	Sec-butanol	−69.9	−5.0	−1.3	−4.4	−0.1
$C_4H_{10}O_2$	1,4-butanediol	−102.0	−10.4	−2.2	−11.6	−1.8
$C_4H_{10}S$	Isobutyl mercaptan	−23.2	3.0	2.7	−0.7	−1.3
$C_4H_{10}S$[h]	Methyl propyl sulfide	−19.5	6.3	4.0	2.7	0.2
$C_4H_{11}N$	*Tert*-butylamine	−28.7	2.3	4.1	−2.0	1.0
C_5H_5N[g]	Pyridine	33.5	10.3	8.3	5.1	3.9
C_5H_6S[i]	2-methylthiophene	20.0	−2.6	−4.0	−3.5	−4.5
C_5H_8	*Trans*-1,3-pentadiene	18.6	1.1	3.9	2.1	5.0
$C_5H_8O_2$[g,h]	Acetylacetone	−90.8	−1.7	−1.8	−3.0	−2.5
C_5H_{10}	Cyclopentane	−18.5	1.5	0.1	1.1	−0.1
C_5H_{10}	2-methyl-1-butene	−8.7	0.4	1.9	1.4	2.7
C_5H_{10}	2-methyl-2-butene	−10.2	0.3	1.3	2.6	3.3
C_5H_{10}	3-methyl-1-butene	−6.9	−0.4	1.3	0.0	1.8
C_5H_{10}	1-pentene	−5.0	2.1	3.8	2.1	3.8
C_5H_{10}	*Cis*-2-pentene	−6.7	1.2	2.5	2.3	3.6
C_5H_{10}[g,h]	*Trans*-2-pentene	−7.6	2.2	3.5	3.0	4.2
$C_5H_{10}O$[g]	2-pentanone	−61.8	3.3	2.9	2.6	2.5
$C_5H_{10}O$	Valeraldehyde	−54.5	2.5	2.6	1.5	2.0
$C_5H_{10}O_2$	Valeric acid	−117.2	−0.6	0.9	−2.4	0.2
$C_5H_{10}S$[f]	Thiacyclohexane	−15.1	4.0	0.4	−0.5	−4.3
$C_5H_{10}S$	Cyclopentanethiol	−11.5	2.6	1.1	−2.8	−4.4
$C_5H_{11}Br$[f]	1-bromopentane	−30.9	1.9	1.0	1.6	0.8
$C_5H_{11}Cl$[h]	1-chloropentane	−41.8	2.9	2.7	3.4	3.2
$C_5H_{11}N$[g]	Piperidine	−11.7	5.3	3.6	0.3	−1.2
C_5H_{12}[f]	Isopentane	−36.9	−0.1	−0.5	1.5	1.0
C_5H_{12}[i]	*n*-pentane	−35.0	3.4	3.0	4.5	3.9
$C_5H_{12}O$	2-methyl-1-butanol	−72.2	−5.7	−2.1	−4.9	−0.7
$C_5H_{12}O$[f,i]	3-methyl-1-butanol	−72.2	−5.2	−1.5	−4.5	−0.2
$C_5H_{12}O$[h]	3-methyl-2-butanol	−75.1	−5.4	−2.2	−4.0	−0.4
$C_5H_{12}O$	2-pentanol	−75.0	−5.4	−2.0	−4.5	−0.7
$C_5H_{12}O$	3-pentanol	−75.7	−8.8	−5.4	−7.3	−3.6
$C_5H_{12}O$[g]	Ethyl propyl ether	−65.1	2.4	2.0	4.0	2.7
$C_5H_{12}S$[f]	*n*-pentyl mercaptan	−185.6	6.2	5.5	2.0	0.7
$C_5H_{12}S$	Butyl methyl sulfide	−25.9	6.4	3.8	2.9	0.0
C_6F_6	Hexafluorobenzene	−24.4	6.3	4.5	6.8	2.2
$C_6H_4Cl_2$[i,h]	*m*-dichlorobenzene	−228.6	−2.4	−3.1	0.1	1.0
$C_6H_4F_2$[i]	*p*-difluorobenzene	6.3	3.6	3.3	2.9	3.0
C_6H_5Cl	Monochlorobenzene	−73.4	−0.4	−0.2	0.7	1.9
C_6H_5F	Fluorobenzene	12.4	2.4	2.9	1.8	2.8

TABLE I. (*Continued*).

Molecular formula	Molecular name	Deviations of ΔH_f^Θ (298 K) (Expt.–Theory)				
		Expt.[a]	HF–LRC1[b]	HF–LRC2[c]	DFT–LRC1[d]	DFT–LRC2[e]
$C_6H_5NO_2$	Nitrobenzene	−27.9	−9.8	−14.1	5.6	1.9
C_6H_6[g]	Benzene	16.2	1.7	3.0	1.6	3.2
$C_6H_6N_2O_2$[f]	*m*-nitroaniline	19.8	−12.0	−15.1	1.1	−0.4
C_6H_6O[i]	Phenol	14.0	−6.3	−2.6	−4.6	−0.2
$C_6H_6O_2$	1,3-benzenediol	−23.0	−13.0	−6.9	−10.3	−3.1
C_6H_7N[h]	2-methylpyridine	−65.7	9.6	7.3	5.2	3.4
$C_6H_8N_2$	Adiponitrile	23.7	22.0	18.9	−1.9	−2.3
C_6H_{10}	1-methylcyclopentene	35.7	0.3	0.1	0.2	−0.2
C_6H_{10}[g]	1,5-hexadiene	−1.3	0.8	4.2	−0.3	3.3
$C_6H_{10}O_3$	Propionic anhydride	20.1	−0.5	−3.9	−2.7	−5.8
$C_6H_{11}NO$[i]	*e*-caprolactam	−149.7	5.0	2.5	−0.9	−2.1
C_6H_{12}[i]	*Trans*-3-hexene	−58.8	1.7	2.6	2.5	3.2
$C_6H_{12}O$[f,i,g]	Butyl vinyl ether	−13.0	−2.3	−1.3	0.0	0.2
$C_6H_{12}O$	3-hexanone	−43.7	3.8	2.8	3.3	2.4
C_6H_{14}[f,h]	3-methylpentane	−66.4	0.3	−2.5	2.3	1.4
$C_6H_{14}S$	Methyl pentyl sulfide	−41.0	6.4	3.5	3.1	−0.3
C_7H_5N	Benzonitrile	−29.3	12.6	10.9	2.1	2.1
C_7H_6O[i,g]	Benzaldehyde	52.3	2.1	1.9	1.8	2.1
$C_7H_6O_3$[i]	2-hydroxybenzoic acid	−8.8	−10.2	−7.3	−7.5	−3.6
C_7H_8	Toluene	12.0	0.6	1.3	1.8	2.5
C_7H_8O[i]	*o*-cresol	−30.7	−7.0	−3.9	−3.9	−0.5
C_7H_9N[f]	2,6-dimethylpyridine	14.0	9.2	6.6	5.5	3.1
C_7H_{14}[g,h]	*Cis*-1,2-dimethylcyclopentane	−31.0	−1.5	−3.3	−1.0	−2.9
$C_7H_{15}Br$[h]	1-bromoheptane	−40.2	2.5	1.0	2.5	0.8
C_7H_{16}[f]	3,3-dimethylpentane	−48.2	−4.3	−5.1	−1.3	−2.6
C_7H_{16}	2,2,3-trimethylbutane	−49.0	−5.6	−6.3	−2.1	−3.3
$C_7H_{16}S$	*n*-heptyl mercaptan	−35.8	5.7	4.4	1.7	−0.4
C_8H_8O[g,h]	Acetophenone	−171.6	0.5	−0.2	0.6	0.2
C_8H_{10}[f]	*o*-xylene	−20.7	−0.7	−0.7	2.1	1.6
$C_8H_{10}O$	3,4-xylenol	4.5	−7.9	−5.5	−3.5	−1.2
C_8H_{16}[f]	*Cis*-1,2-dimethylcyclohexane	−37.4	−2.6	−4.9	−1.6	−4.5
C_8H_{16}[g]	*Trans*-1,4-dimethylcyclohexane	−41.2	−1.6	−3.8	−1.0	−3.8
C_8H_{16}[h]	2,4,4-trimethyl-2-pentene	−44.1	−5.6	−5.3	−1.5	−2.1
C_8H_{18}[h]	2,3-dimethylhexane	−25.1	−2.9	−4.1	0.4	−1.3
C_8H_{18}[f]	3-ethylhexane	−51.1	−2.8	−4.0	0.1	−1.7
C_8H_{18}[i]	4-methylheptane	−50.4	0.3	−1.0	2.6	0.8
C_8H_{18}	2,3,4-trimethylpentane	−50.7	−8.4	−9.4	−4.1	−5.9
$C_8H_{18}O$[f,i,g]	2-ethyl-1-hexanol	−52.0	−10.5	−8.2	−8.1	−5.8
$C_8H_{18}S_2$	Dibutyl disulfide	−87.3	4.3	−1.1	−0.7	−7.4
$C_9H_{10}O_2$[f]	3-ethylbenzoic acid	−87.9	−4.5	−4.1	−4.0	−2.9
C_9H_{12}	*m*-ethyltoluene	−0.5	−0.1	−0.5	2.3	1.4
C_9H_{12}[h]	1,2,3-trimethylbenzene	−2.3	−8.2	−3.8	1.3	−0.2
$C_9H_{18}O$[h]	Diisobutyl ketone	−85.5	−1.7	−3.9	−0.8	−3.5
C_9H_{20}[g]	3,3-diethylpentane	−55.4	−10.7	−12.1	−5.8	−8.2
C_9H_{20}	2,2,3,4-tetramethylpentane	−56.6	−10.2	−11.5	−5.1	−7.2
$C_{10}H_{14}$	*Sec*-butylbenzene	−4.0	−1.8	−2.2	0.4	−0.5
$C_{10}H_{14}$[f,i]	Isobutylbenzene	−4.9	−1.4	−1.9	0.5	−0.2
$C_{10}H_{18}O_4$[h]	Sebacic acid	−220.3	−0.6	1.1	−4.9	−1.3
$C_{10}H_{20}O_2$[g]	*n*-decanoic acid	−142.0	−0.6	−0.7	−1.8	−1.3
$C_{12}H_{10}$	Acenaphthene	37.0	−4.0	−6.3	−0.5	−1.9

[a]Experimental values were taken from Refs. 10–12.
[b]Deviations of corrected ΔH_f^Θ by linear regression corrected HF/6-31G(d) method (HF-LRC1).
[c]Deviations of corrected ΔH_f^Θ by linear regression corrected HF/6-311+G(d,p) method (HF-LRC2).
[d]Deviations of corrected ΔH_f^Θ by linear regression corrected B3LYP/6-31G(d) method (DFT-LRC1).
[e]Deviations of corrected ΔH_f^Θ by linear regression corrected B3LYP/6-311+G(d,p) method (DFT-LRC2).
[f]Molecules belong to the testing set in HF-LRC1 calculation.
[g]Molecules belong to the testing set in DFT-LRC1 calculation.
[h]Molecules belong to the testing set in DFT-LRC2 calculation.
[i]Molecules belong to the testing set in HF-LRC2 calculation.

the corrected results and their experimental counterparts are listed in Table III.

From Table III, it is immediately obvious that upon linear regression correction there is a great decrease in the deviations. For the HF/6-31G(d) method, the mean absolute deviation is reduced from 351.0 kcal/mol to 4.6 kcal/mol, and for the HF/6-311+G(d,p) method, it is reduced from 360.9 kcal/mol to 4.6 kcal/mol. While for the

TABLE II. Coefficients (kcal/mol) of the descriptors used in Eq. (5) for HF and B3LYP methods.

Descriptors	HF		B3LYP	
	6-31G(d)	6-311+G(d,p)	6-31G(d)	6-311+G(d,p)
Lone-pair electrons (a_1)	3.515	2.331	0.5568	−1.017
Bonding electrons (a_2)	10.01	6.978	0.6436	−2.460
N-1 (a_3)	−20.70	−17.97	−1.009	1.617
N-2 (a_4)	68.15	60.96	0.8023	−5.443
N-3 (a_5)	−97.00	−90.70	6.549	10.75
Unpaired electrons (b_1)	22.18	19.30	0.8407	−1.999

B3LYP/6-31G(d) and B3LYP/6-311+G(d,p) methods, the corrected results are considerably better than those of HF. After linear regression correction, the mean absolute deviations drop to 2.7 and 2.4 kcal/mol for the B3LYP/6-31G(d) and B3LYP/6-311+G(d,p) methods, respectively. The corrected ΔH_f^{Θ} of four methods compared to their experimental values are illustrated in Figs. 1(c), 1(d), 2(c), and 2(d), in which the triangles belong to the training set and the asterisks belong to testing set. The figures clearly shows that the linear regression corrected results are much closer to their experimental counterparts for both training and testing set. Especially for the HF results, after linear regression correction, the mean absolute deviation is reduced by about 80 times, and much smaller than the results from raw B3LYP calculations.

Root-mean-square deviation analysis further demonstrates that our linear regression correction approach greatly decreases the calculation errors of HF and DFT methods. For HF/6-31G(d) and HF/6-311+G(d,p) methods, after linear regression correction, the rms deviations of ΔH_f^{Θ} are reduced from 385.3 kcal/mol to 6.0 kcal/mol and from 395.9 kcal/mol to 6.1 kcal/mol, respectively. The DFT-LRC scheme nearly has the same accuracy as the DFT-NEURON scheme. For B3LYP/6-31G(d) and B3LYP/6-311+G(d,p) methods, the rms deviations are decreased from 10.8 kcal/mol to 3.5 kcal/mol and from 20.7 kcal/mol to 3.1 kcal/mol. None of our LRC approach, DFT-NEURON approach of Chen *et al.*, and Gaussian-*n* series of methods can reach the accuracy of the group/bond equivalent correction approach of Allinger and co-workers. The calculation of heat of formation in the approach of Allinger and co-workers can be viewed as an isodesmic reaction scheme, which does not require high-level correlation method to calculate the energy, and thus the results are not sensitive to the method used. In comparison, although the approaches using an atomization energy scheme are generally less accurate and more sensitive to the method used, they require less intervention of the researchers and are applicable to broader range of molecules. Nevertheless, for both HF and DFT methods, linear regression correction approach greatly decreases the large systematic deviations from the experimental values. More importantly, with descriptors of electron pairs, our linear regression correction approach has substantially eliminated the deficiencies of these methods in calculating electron correlation energy.

The coefficients of partial correlation V_j are calculated to assess the validation of physical descriptors. Table IV lists the partial correlation coefficients of all descriptors. Most values of the partial correlation coefficients are close to 1, which implies that all descriptors are necessary and crucial for our linear regression correction approach. Examination of Table IV indicates that bonding electrons are very important for electron correlation correction. The large partial correlation values of the descriptors for the inner layer electrons indicate that the electron correlation change for the inner layer electrons from atoms to molecules is large and nonnegligible. Except B3LYP/6-31G(d) method, for the other three methods, the partial correlation value decreases from N-1 to N-3, indicating that the closer the electrons to the nucleus, the less important is the electron correlation in chemical reactions. Table IV also shows that the partial correlation values of the electron descriptors for the B3LYP method are all smaller than the corresponding values for the HF method, implying that electron correlation correction is much more important for HF method than for B3LYP method, since DFT methods including B3LYP have already

TABLE III. Mean absolute deviations and root-mean-square deviations of heat of formation (kcal/mol) before and after correction with the six descriptor Eq. (5).

	Training set				Testing set			
	HF		B3LYP		HF		B3LYP	
	6-31G(d)	6-311+G(d,p)	6-31G(d)	6-311+G(d,p)	6-31G(d)	6-311+G(d,p)	6-31G(d)	6-311+G(d,p)
B-MAD[a]	351.0	360.9	9.2	18.2	396.7	362.4	9.0	19.6
A-MAD[b]	4.6	4.6	2.7	2.4	5.0	4.6	2.3	2.2
B-rms[a]	385.3	395.9	10.8	20.7	425.5	390.6	10.9	22.1
A-rms[b]	6.0	6.1	3.5	3.1	6.9	6.2	3.0	2.6

[a]Before correction.
[b]After correction.

TABLE IV. Coefficients of partial correlation for all the parameters of the six-descriptor [Eq. (5)].

Descriptors	HF		B3LYP	
	6-31G(d)	6-311+G(d,p)	6-31G(d)	6-311+G(d,p)
Lone-pair electron (a_1)	0.9800	0.9538	0.8051	0.9386
Bonding electron (a_2)	0.9998	0.9996	0.9845	0.9991
N-1 (a_3)	0.9998	0.9998	0.9779	0.9932
N-2 (a_4)	0.9996	0.9995	0.5772	0.9849
N-3 (a_5)	0.9952	0.9950	0.7676	0.9205
Unpaired electron (b_1)	0.9999	0.9999	0.9805	0.9972

involved some electron correlation correction, but HF method has not.

We have also tested the relative contribution of an individual physical descriptor by examining the effect of leaving out one descriptor employed in Eq. (5) upon the mean absolute deviations for the 150 training molecules after retraining. The results tabulated in Table V indicate that the unpaired electrons and three inner layer electrons have remarkable contributions to the correction of correlation energy. Analyzing the HF-linear regression corrected results shows that without the descriptor of unpaired electrons, the mean absolute deviations greatly increase for both basis sets, from 4.6 kcal/mol to 13.5 kcal/mol and from 4.6 kcal/mol to 11.7 kcal/mol, respectively. It reflects a large discrepancy in the calculation of correlation energy for the unpaired electron before and after forming bonds at the HF level. It is our common sense that electron correlation energy in inner layer electrons may not change too much when forming chemical bonds. On the contrary, our results show that the correlation energy changes of inner layer electrons from atoms to molecule also have significant contribution to the overall correlation energy changes. Without three inner layer electrons, the deviation reaches 16.4 kcal/mol for HF/6-31G(d), and 15.3 kcal/mol for HF/6-311+G(d,p). In addition, when taking the sum of three inner layer electrons as one parameter, the deviation is 12.9 kcal/mol for HF/6-31G(d), and 12.0 kcal/mol for HF/6-311+G(d,p), which indicates that the electron correlation energy changes of different inner layer electrons are not the same and nonnegligible. The effect of lone-pair electrons is slightly smaller compared to the effect of bonding electrons. For HF/6-31G(d), without lone-pair electrons, the mean absolute deviation is 5.7 kcal/mol, while without bonding electrons, the mean absolute deviation is 6.7 kcal/mol. This is easy to understand because the chemical environment is greatly changed upon forming chemical bonds from unpaired electrons in atoms, while the change of the chemical environment for the lone-pair electrons from atoms to molecule is smaller. For HF/6-311+G(d,p), the case is similar. Whereas, for B3LYP-linear regression corrected results, the effect of different descriptors are all small and similar, simply due to the smaller errors of the raw B3LYP results. In addition, two major differences regarding the importance of different descriptors to the correction of HF and B3LYP methods are observed.

(1) In contrast to its large effect on the correction of HF energies, the unpaired electrons in atom have a trivial impact on B3LYP results.

(2) For B3LYP methods, combining the three descriptors for all inner layer electrons into one or without them has a comparable effect on each, and both have a much larger effect than excluding other descriptors. It seems to imply that the major deficiency of the B3LYP method is the dealing of the core electron correlation.

Although the mean absolute deviations of HF/6-31G(d) and HF/6-311+G(d,p) reaches 4.6 kcal/mol after linear regression correction, molecules substituted by nitryl and cyanogens have very large deviations far beyond the mean absolute deviation. For example, the absolute deviations of $C_4H_4N_2$ and $C_6H_8N_2$ are 23.7 kcal/mol and 22.0 kcal/mol for HF/6-31G(d), and 20.9 kcal/mol and 18.9 kcal/mol for HF/6-311+G(d,p), respectively. The absolute deviations of $C_2H_5NO_3$ and $C_3H_7NO_3$ (propyl nitrate) are 13.2 kcal/mol

TABLE V. Effect of different physical descriptors. (Values listed are mean absolute deviations, in kcal/mol.)

Linear regression correction scheme	HF		B3LYP	
	6-31G(d)	6-311+G(d,p)	6-31G(d)	6-311+G(d,p)
I	4.6	4.6	2.7	2.4
II	5.7	5.1	2.8	2.7
III	6.7	5.4	2.8	2.9
IV	13.5	11.7	2.8	2.8
V	16.4	15.3	3.6	2.9
VI	12.9	12.0	3.6	2.9

[I:]The descriptors used are lone-pair, bonding, N-1, N-2, N-3, and unpaired electrons.
[II:]The descriptors used are bonding, N-1, N-2, N-3, and unpaired electrons, without lone-pair electrons.
[III:]The descriptors used are lone-pair, N-1, N-2, N-3, and unpaired electrons, without bonding electrons.
[IV:]The descriptors used are lone-pair, bonding, and N-1, N-2, N-3 electrons, without unpaired electrons.
[V:]The descriptors used are lone-pair, bonding, and unpaired electrons, without N-1, N-2, and N-3 electrons.
[VI:]The descriptors used are lone-pair, bonding, the sum of inner layer, and unpaired electrons.

and 13.0 kcal/mol for HF/6-31G(d), 18.1 kcal/mol and 18.2 kcal/mol for HF/6-311+G(d,p), respectively. The mean absolute deviations of the molecules substituted by these two groups exceed 10 kcal/mol. On the other hand, if we delete the molecules substituted by nitryl and cyanogens in the training set, the mean absolute deviation comes to 4.0 kcal/mol for HF/6-31G(d), and 3.8 kcal/mol for HF/6-311+G(d,p). We consider that once suitable physical descriptors for these type of molecules are found, the large errors will be eliminated.

The raw calculated results for HF/6-31G(d) and HF/6-311+G(d,p) have much larger deviations for the experimental counterparts than those of B3LYP/6-31G(d) and B3LYP/6-311+G(d,p). Whereas, upon linear regression correction, both of the calculations have the deviations of the same order of magnitude, which proves sufficiently that our linear regression correction approach is appropriate for both HF and B3LYP methods at the same time. From Table I, we find that the deviations of large molecules are of the same magnitude as those of small molecules after linear regression correction. Since our linear regression correction algorithm can correct easily the large correlation energy deficiency of HF method, and the physical descriptors selected are general and have wide applicability, therefore, this linear regression correction approach can potentially be applied to much larger systems.

Our linear regression correction approach has accounted for the most part of the unbalanced correlation energy in atoms and molecule calculated by the HF and DFT methods. The physical descriptors adopted in our linear regression methods, the numbers of lone-pair electrons, bonding electrons, inner layer electrons, and unpaired valence electrons in atoms are not limited to the specific properties of organic molecules. The coefficients obtained in this study can be substituted back to Eqs. (1) and (3) to calculate the energies of the molecules and atoms. However, it should be kept in mind that this approach does not give accurate correlation energy of a molecule, but a relative accurate correlation energy change from atoms to molecule. Since most physical properties are determined ultimately by the energy, the direct linear regression correction approach for total energy has a much wider range of applications. We are extending this approach to broader applications, such as the calculation of ionization potentials, electron affinities, and thermochemical properties of radicals.

V. CONCLUSIONS

A simple and effective LRC scheme has been developed to eliminate the large errors of HF and B3LYP methods in calculating heat of formation. The descriptors in the LRC scheme are the number of lone-pair electrons, bonding electrons and inner layer electrons in molecules, and the number of unpaired electrons in ground state atoms. The mean absolute deviations of the corrected ΔH_f^{Θ} is reduced to 4.6 kcal/mol for both of HF/6-31G(d) and HF/6-311+G(d,p). In the mean time, the LRC approach has also improved the ΔH_f^{Θ} calculation results of DFT method. The mean absolute deviations are reduced to 2.7 kcal/mol and 2.4 kcal/mol for B3LYP/6-31G(d) and B3LYP/6-311+G(d,p), respectively. The large systematic deviations for the calculated ΔH_f^{Θ} are reduced drastically, in particular, for the HF results. As more and better experimental data are available, the LRC approach can be further improved. Last but not the least, this LRC approach can be applied to much larger systems including inorganic molecules, and has the potential to be a powerful tool to predict the physical properties of materials prior to the experiments.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China Grant No. (20273015) and the Natural Science Foundation of Shanghai Science Technology Committee Grant No. (02DJ4023).

[1] L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, J. Chem. Phys. **94**, 7221 (1991).
[2] L. A. Curtiss, J. E. Carpenter, K. Raghavachari, and J. A. Pople, J. Chem. Phys. **96**, 9030 (1992).
[3] L. A. Curtiss, K. Raghavachari, and J. A. Pople, J. Chem. Phys. **103**, 4192 (1995).
[4] L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, and J. A. Pople, J. Chem. Phys. **109**, 7764 (1998).
[5] A. G. Baboul, L. A. Curtiss, P. C. Redfern, and K. Raghavachari, J. Chem. Phys. **110**, 7650 (1999).
[6] L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, J. Chem. Phys. **112**, 7374 (2000).
[7] P. C. Redfern, P. Zapol, L. A. Curtiss, and K. Raghavachari, J. Phys. Chem. A **104**, 5850 (2000).
[8] L. H. Hu, X. J. Wang, L. H. Wong, and G. H. Chen, J. Chem. Phys. **119**, 11501 (2003).
[9] N. L. Allinger, L. R. Schmitz, I. Motoc, C. Bender, and J. K. Labanowski, J. Phys. Org. Chem. **3**, 732 (1990).
[10] N. L. Allinger, L. R. Schmitz, I. Motoc, C. Bender, and J. K. Labanowski, J. Am. Chem. Soc. **114**, 2880 (1992).
[11] L. R. Schmitz, I. Motoc, C. Bender, J. K. Labanowski, and N. L. Allinger, J. Phys. Org. Chem. **5**, 225 (1992).
[12] N. L. Allinger, L. R. Schmitz, I. Motoc, C. Bender, and J. K. Labanowski, J. Comput. Chem. **13**, 838 (1992).
[13] N. L. Allinger, K. Sakakibara, and J. K. Labanowski, J. Phys. Chem. **99**, 9603 (1995).
[14] J. Labanowski, L. Schmitz, K. H. Chen, and N. L. Allinger, J. Comput. Chem. **19**, 1421 (1998).
[15] A. Nicolaides, A. Rauk, M. N. Glukhovtsev, and L. Radom, J. Phys. Chem. **100**, 17460 (1996).
[16] C. L. Yaws, *Chemical Properties Handbook* (McGraw-Hill, New York, 1999).
[17] D. R. Lide, *CRC Handbook of Chemical and Physics*, 3rd ed. (CRC, Boca Raton, FL, 2000).
[18] J. B. Pedley, R. D. Naylor, and S. P. Kirby, *Thermochemical Data of Organic Compounds*, 2nd ed. (Chapman and Hall, New York, 1986).
[19] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[20] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[21] S. H. Vosko, L. Wilk, and M. Nusair, Can. J. Phys. **58**, 1200 (1980).
[22] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, J. Phys. Chem. **98**, 41623 (1994).
[23] A. P. Scott and L. Radom, J. Phys. Chem. **100**, 16502 (1996).
[24] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 98, REVA. 7 Gaussian Inc., Pittsburgh, PA, 1998.
[25] See EPAPS Document No. E-JCPSA6-121-303436 for the electronic energies and zero-point energies calculated by HF and B3LYP methods with 6-31G(d) and 6-311+G(d,p) basis sets. A direct link to this document may also be found in the online article's HTML reference section. The document may also be reached via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html) or from ftp.aip.org in the directory/epaps/. See the EPAPS homepage for more information.