Available online at www.sciencedirect.com
SCIENCE DIRECT®
Chemical Physics Letters 390 (2004) 186–192
CHEMICAL PHYSICS LETTERS
www.elsevier.com/locate/cplett

A generalized exchange-correlation functional: the Neural-Networks approach ☆

Xiao Zheng, LiHong Hu, XiuJun Wang, GuanHua Chen *

Department of Chemistry, The University of Hong Kong, Hong Kong, PR China

Received 12 February 2004; in final form 6 April 2004
Available online 27 April 2004

Abstract

A Neural-Networks approach is employed to improve B3LYP exchange-correlation functional by taking into account of high-order contributions. The new B3LYP functional is based on a Neural-Network whose structure and synaptic weights are determined from 116 known experimental energy data [J. Chem. Phys. 98 (1993) 5648]. It leads to better agreement between the first-principles calculations and the experimental results. The new functional is further tested by applying it to calculate 40 ionization potentials and 40 enthalpies of formation in G2-2 test set [J. Chem. Phys. 109 (1998) 42] using 6-311+G(3df,2p) basis set. The root-mean-square errors are reduced from those of conventional B3LYP calculations.

1. Introduction

Density functional theory (DFT) converts many-electron problems into effective one-electron problems. This conversion is rigorous if the exact exchange-correlation (XC) functional is known. It is thus important to find accurate DFT XC functionals. Much progress has been made, primarily due to the development of generalized gradient approximation (GGA) [1–3] and hybrid functionals [4]. Existing XC functionals include local or nearly local contributions such as local spin density approximation (LSDA) [5] and GGA [1–3], and nonlocal terms such as the exact exchange functional. Although these local and nonlocal terms account for the bulk contributions to exact XC functional, high-order contributions are yet to be identified and taken into account. Conceding that it is exceedingly difficult to derive analytically the exact universal XC functional, a semiempirical methodology has been established by utilizing highly accurate experimental data to determine XC functionals [4,6,7]. Becke pioneered this approach and determined the three parameters in B3LYP functional [8] by a least-square fit to 116 molecular and atomic energy data [4]. Building upon this semiempirical methodology, we propose here a new approach which takes into account of high-order contributions beyond the existing local and nonlocal XC functionals.

Since its beginning in the late fifties, Neural-Networks has been applied to various engineering problems, such as robotics, pattern recognition, and speech [9]. A Neural-Network is a highly nonlinear system, and is suitable to determine or mimic the complex relationships among relevant physical variables. Tozer et al. [10] have employed a computational Neural-Network to fit the Zhao–Morrison–Parr XC potential [11]. By assuming $v_{XC} = v_{XC}(\rho)$, they incorporated the Neural-Network into a regular Kohn–Sham procedure [5] with encouraging results. Recently we developed a combined first-principles calculation and Neural-Networks correction approach to improve significantly the accuracy of calculated thermodynamic properties [12,13]. In this work, we develop a Neural-Networks approach to construct a new type of DFT XC functional, and then apply it to improve the results of the popular B3LYP calculations. What differentiates our work from [10] is that our

☆ Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cplett.2004.04.020.

* Corresponding author. Fax: +85228571586.
E-mail address: ghc@everest.hku.hk (G. Chen).

0009-2614/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.cplett.2004.04.020

Neural-Networks-based XC functional is system-dependent for each molecule or atom while in [10] the XC potential of different systems are fitted to the same function of densities and density derivatives.

In Section 2 we describe the Neural-Networks-based methodology. The improved B3LYP calculations results are presented in Section 3. Further discussions will be given in Section 4.

2. Methodology

The hybrid B3LYP functional is expressed as

$$E_{\mathrm{XC}} = a_0 E_{\mathrm{X}}^{\mathrm{Slater}} + (1-a_0)E_{\mathrm{X}}^{\mathrm{HF}} + a_{\mathrm{X}}\Delta E_{\mathrm{X}}^{\mathrm{Becke}} + a_{\mathrm{C}}E_{\mathrm{C}}^{\mathrm{LYP}} + (1-a_{\mathrm{C}})E_{\mathrm{C}}^{\mathrm{VMN}}, \quad (1)$$

where $E_{\mathrm{X}}^{\mathrm{Slater}}$ is the local spin density exchange functional [5], $E_{\mathrm{X}}^{\mathrm{HF}}$ is the exact exchange functional, $E_{\mathrm{X}}^{\mathrm{Becke}}$ is Becke's gradient-corrected exchange functional [1], $E_{\mathrm{C}}^{\mathrm{LYP}}$ is the correlation functional of Lee et al. [2], and $E_{\mathrm{C}}^{\mathrm{VMN}}$ represents the correlation functional proposed by Vosko et al. [14]. The three parameters, a_0, a_{X} and a_{C}, dictate the contributions of various terms. Their respective values, 0.80, 0.72 and 0.81, were determined via a least-square fit to 116 atomization energies (AEs), ionization potentials (IPs), proton affinities (PAs) and total atomic energies (TAEs) by Becke [4]. Note that $a_{\mathrm{X}} < a_0 < a_{\mathrm{C}}$. B3LYP functional consists explicitly of the first and second rungs of the Jacob's ladder of density functional approximation [15] and the partial exact exchange functional [4]. Being determined via the least-square fit to the 116 experimental data, it includes implicitly the high-order contributions to the exact functional such as those in meta-GGA [15]. These high-order contributions were averaged in terms of Eq. (1), and were assumed invariant for all atomic or molecular systems. Although the universal functionals are system-independent, the high-order contributions are in fact system-dependent. Therefore, their inclusions in Eq. (1) should lead to the system-dependence of a_0, a_{X} and a_{C} which is in turn dictated by the characteristic properties of the system. The challenge is to identify these characteristic properties and determine their quantitative relationships to the three parameters a_0, a_{X} and a_{C}. These characteristic properties, termed as the physical descriptors of the system, should satisfy two criteria: (1) they must be of purely electronic nature, since the exact XC functional is a universal functional of electron density only; and (2) they should reflect the electron distribution. After identifying these physical descriptors, we then use Neural-Networks to determine their quantitative relationships to a_0, a_{X} and a_{C}, and the resulting Neural-Network can be viewed as a generalized universal XC functional, which can be systematically improved upon the availability of new experimental data.

Beyond the GGA, Perdew and co-workers [16] proposed the meta-GGA in which the XC functional depends explicitly on the kinetic energy density,

$$\tau(\mathbf{r}) = \frac{1}{2}\sum_{\alpha}^{\mathrm{occ}} |\nabla\psi_\alpha(\mathbf{r})|^2, \quad (2)$$

where $\psi_\alpha(\mathbf{r})$ is the wave function of an occupied Kohn–Sham orbital α. The total kinetic energy, $\mathscr{T} = \int \tau(\mathbf{r})\,\mathrm{d}^3\mathbf{r}$, relates thus closely to the high-order contributions to B3LYP functional, and is thus chosen as a key physical descriptor. The electron density distribution $\rho(\mathbf{r})$ uniquely determines the XC functional and can be expanded in terms of the multipole moments. Being the zeroth-order term of the expansion, the total number of electrons N_t is recognized as a natural physical descriptor, so are the dipole and quadrupole moments of the system. We choose $D \equiv \sqrt{d_x^2 + d_y^2 + d_z^2}$ as the dipole descriptor and $Q \equiv \sqrt{Q_{xx}^2 + Q_{yy}^2 + Q_{zz}^2}$ as the quadrupole descriptor, where $d_i (i = x, y, z)$ is a component of the dipole vector and $Q_{ii} (i = x, y, z)$ is a diagonal element of the quadrupole tensor. Spin multiplicity g_S is also adopted as a physical descriptor since the exchange functional accounts for the exchange interaction among electrons of the same spin.

Our Neural-Network adopts a three-layer architecture which consists of an input layer, a hidden layer and an output layer [9], see Fig. 1. Adjacent layers are connected via the synaptic weights. The values of the physical descriptors, g_S, N_t, D, $\mathscr{T}$ and Q, are inputted into the Neural-Network at the input layer. And the modified a_0, a_{X} and a_{C} for each atom or molecule, denoted by $\tilde{a}_0$, $\tilde{a}_{\mathrm{X}}$ and $\tilde{a}_{\mathrm{C}}$, are obtained at the output layer. We take the 116 experimental energies employed by Becke [4] as the training set, and use them to determine the structure of our Neural-Network and its synaptic weights. Instead of the basis-set-free calculations [4], we adopt a Gaussian-type-function (GTF) basis set, 6-311+G(3df,2p), in our calculations. Geometry of every molecule is optimized directly by conventional B3LYP/6-311+G(3df,2p). The values of $\mathscr{T}$, D and Q are obtained at the same level. Besides g_S, N_t, D, $\mathscr{T}$ and Q, biases are introduced at both the input and hidden layers and their values are set to unity. An error back-propagation learning procedure [17] is employed to optimize the synaptic weights. For each molecule or atom, the system-dependent $\tilde{a}_0$, $\tilde{a}_{\mathrm{X}}$ and $\tilde{a}_{\mathrm{C}}$ are used to modify its B3LYP functional, which is in turn used to evaluate its AE, IP, PA, or TAE. These resulting energies are then compared to their experimental counterparts, the differences serve as feedbacks to the Neural-Network, and all synaptic weights are tuned, accordingly. This procedure is iterated until the calculated and measured energies are close enough for the whole training set, and the Neural-Network is then considered as converged, i.e., its synaptic weights are determined.

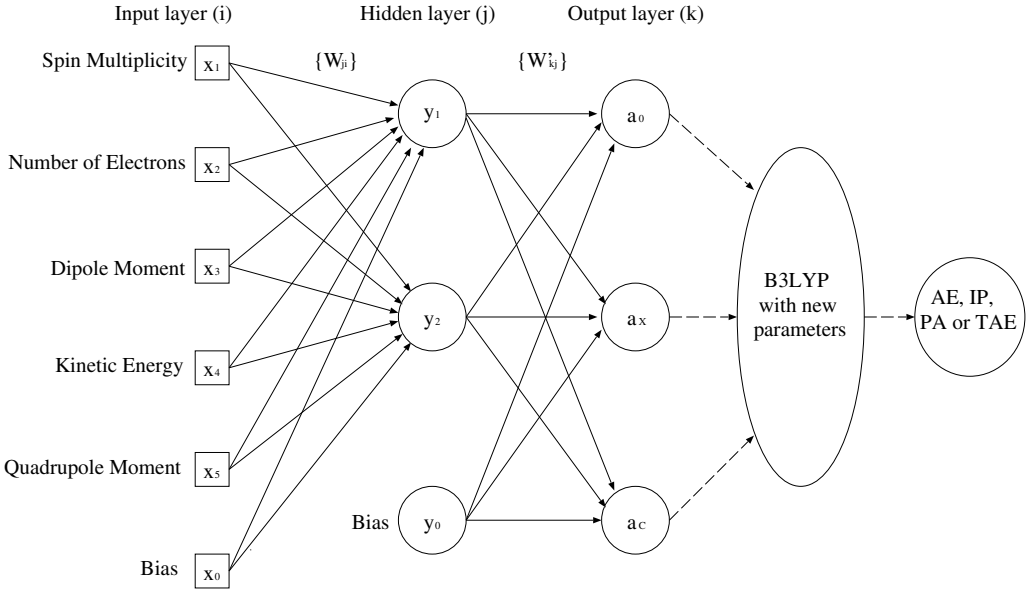

Fig. 1. Architectural graph of our Neural-Network and flow chart of our calculations.

The Neural-Network structure is optimized via a cross-validation technique [18]. The training set is randomly partitioned into six subsets of equal size. Five of them are used to train the weights of the Neural-Network, and are termed as the estimation subset. The sixth is used to evaluate the prediction of current Neural-Network, and is termed as the validation subset. This procedure is repeated six times in rotation to assess the performance of current Neural-Network. We vary the number of neurons in the hidden layer from 1 to 5 and find that two hidden neurons yield the best results, i.e., the minimal overall root-mean-square (RMS) errors and the minimal RMS difference between the estimation and validation subsets (less than 0.2 kcal mol^{-1}). The latter helps ensure the predictive capability of our Neural-Network. Therefore, the 6-3-3 structure is adopted (see Fig. 1, where x_1, x_2, x_3, x_4, x_5 and x_0 are g_S, N_t, D, $\mathscr{T}$, Q and bias, respectively). To maintain the numerical stability of the learning procedure, the input is preprocessed before being imported into the Neural-Network. Except for the bias, every x_i is scaled into the range (0.1, 0.9). The Neural-Networks-corrected $\tilde{a}_0$, $\tilde{a}_X$ and $\tilde{a}_C$ are generated at the output layer, and are related to the input $\{x_i\}$ as

$$\tilde{a}_0 = \mathrm{Sig}b\left\{\left[\sum_{j=1}^{2} W'_{1j}\cdot \mathrm{Sig}a\left(\sum_{i=0}^{5} W_{ji}x_i\right)\right] + W'_{10}\right\}, \quad (3)$$

$$\tilde{a}_X = \mathrm{Sig}b\left\{\left[\sum_{j=1}^{2} W'_{2j}\cdot \mathrm{Sig}a\left(\sum_{i=0}^{5} W_{ji}x_i\right)\right] + W'_{20}\right\}, \quad (4)$$

$$\tilde{a}_C = \mathrm{Sig}b\left\{\left[\sum_{j=1}^{2} W'_{3j}\cdot \mathrm{Sig}a\left(\sum_{i=0}^{5} W_{ji}x_i\right)\right] + W'_{30}\right\}, \quad (5)$$

where $\mathrm{Sig}a(v) = \frac{1}{1+\exp(-\alpha v)}$ and $\mathrm{Sig}b(v) = \beta\tanh(\gamma v)$, α and γ are parameters which control the switch steepness of Sigmoidal functions $\mathrm{Sig}a(v)$ and $\mathrm{Sig}b(v)$. $\{W_{ji}\}$ are the weights connecting the input layer $\{x_i\}$ and the hidden neurons $\{y_j\}$, and $\{W'_{kj}\}$ connect the hidden neurons and the output ($i = 0$–5, and $j = 0-2$. Zero indices are referred to the biases).

3. Results

The conventional B3LYP/6-311+G(3df,2p) calculations are carried out to evaluate AEs, IPs, PAs or TAEs of the molecules and atoms in the training set, and their resulting deviations from the experimental data are 3.0, 4.9, 1.6 and 10.3 kcal mol^{-1} for AEs, IPs, PAs and TAEs, respectively (see Table 1). The physical descriptors $\mathscr{T}$, N_t, D, Q and g_S of each molecule or atom in the training set are inputted to the Neural-Network, and the $\tilde{a}_0$, $\tilde{a}_X$ and $\tilde{a}_C$ from the output layer are used to construct the B3LYP functional which is applied subsequently to calculate AE, IP, PA or TAE. These values

Table 1
RMS errors (all data are in the units of kcal mol^{-1})

Properties	AE	IP	PA	TAE	Overall
Number of samples	56	42	8	10	116
A[a]	2.9	3.9	1.9	4.1	3.4
DFT-1[b]	3.0	4.9	1.6	10.3	4.7
DFT-NN[c]	2.4	3.7	1.6	2.7	2.9

[a] Becke's work.
[b] Conventional B3LYP/6-311+G(3df,2p).
[c] Neural-Networks-based B3LYP/6-311+G(3df,2p).

are then compared to the 116 energy values in the training set to tune the synaptic weights $\{W_{ji}\}$ and $\{W'_{kj}\}$. The final values of synaptic weights are shown in Table 2, and are used to calculate $\tilde{a}_0$, $\tilde{a}_X$ and $\tilde{a}_C$ of the new B3LYP functional which is subsequently employed to calculate AE, IP, PA, TAE and other properties of a

Table 2
Optimized synaptic weights W_{ji} and W'_{kj} as well as derivatives of $\tilde{a}_0$, $\tilde{a}_X$ and $\tilde{a}_C$ w.r.t. each physical descriptor.

	$j=1$	$j=2$		$k=1$	$k=2$	$k=3$		$\partial\tilde{a}_0/\partial x_i$	$\partial\tilde{a}_X/\partial x_i$	$\partial\tilde{a}_C/\partial x_i$
W_{j1}	−0.89	1.11	W'_{k1}	0.21	−0.02	0.46	$i=1$	−0.067	−0.036	0.099
W_{j2}	0.52	−0.09	W'_{k2}	0.18	0.06	0.36	$i=2$	0.035	0.034	−0.010
W_{j3}	0.18	0.09	W'_{k0}	−0.03	0.54	0.53	$i=3$	0.011	0.015	0.007
W_{j4}	0.78	0.20					$i=4$	0.050	0.058	0.014
W_{j5}	0.22	0.28					$i=5$	0.012	0.022	0.023
W_{j0}	0.15	0.06					$i=0$	0.009	0.011	0.044

Derivatives are obtained at $x_i = 0.5$ ($i = 1$–5) and $x_0 = 1$.

Table 3
Ionization potentials of testing set (all data are in units of kcal mol^{-1})

No.	Species	Expt.[a]	DFT-1[b]	Δ_1[c]	DFT-NN[d]	Δ_2[e]
1	CF_2	263.4	261.73	−1.62	264.73	1.38
2	CH_2	239.6	239.82	0.23	237.06	−2.53
3	CH_2S	216.3	214.00	−2.30	215.38	−0.92
4	CH_3Cl	258.7	255.50	−3.23	255.97	−2.76
5	CH_3F	287.6	283.64	−3.92	285.48	−2.08
6	CH_3	226.9	229.45	2.54	225.07	−1.84
7	CH_3OH	250.2	244.21	−5.99	246.28	−3.92
8	CH_3O	247.2	243.97	−3.23	243.28	−3.92
9	CHO	187.7	195.78	8.07	190.71	3.00
10	CO_2	317.5	316.84	−0.70	319.61	2.07
11	COS	257.8	258.04	0.23	259.43	1.62
12	HOF	293.1	291.94	−1.15	293.55	0.46
13	NH_2	256.9	261.27	4.38	258.73	1.84
14	NH	311.1	315.46	4.38	313.15	2.07
15	SC	261.3	263.58	2.31	263.81	2.54
16	B_2H_4	223.7	219.07	−4.61	219.99	−3.69
17	C_2H_5	187.3	189.55	2.30	183.79	−3.46
18	CH_3SH	217.7	215.15	−2.54	217.46	−0.23
19	CS_2	232.2	231.29	−0.92	231.29	−0.92
20	N_2H_2	221.2	220.22	−0.93	221.84	0.69
21	N_2H_3	175.5	182.17	6.68	174.56	−0.93
22	Si_2H_2	189.1	185.17	−3.92	187.94	−1.15
23	Si_2H_4	186.6	182.17	−4.39	185.40	−1.16
24	SiH_3	187.7	188.63	0.92	185.86	−1.85
25	BF_3	358.8	349.04	−9.76	351.51	−7.29
26	BCl_3	267.5	257.86	−9.64	257.39	−10.11
27	B_2F_4	278.3	265.99	−12.31	271.38	−6.92
28	C_3H_4 (cyclopropene)	223.0	216.43	−6.57	218.39	−4.62
29	C_3H_4 (allene)	223.5	226.09	2.59	229.08	5.58
30	C_6H_6	213.2	208.37	−4.83	212.43	−0.77
31	CH_2S	216.2	214.08	−2.12	215.27	−0.93
32	CH_2SH	173.8	176.91	3.11	173.39	−0.41
33	C_2H_5OH	241.4	232.85	−8.55	236.07	−5.33
34	CH_3CHO	235.9	232.41	−3.49	235.16	−0.74
35	CH_3OF	261.5	256.88	−4.62	261.11	−0.39
36	NCCN	308.3	300.75	−7.55	302.57	−5.73
37	Si_2H_5	175.3	177.58	2.28	170.68	−4.62
38	Si_2H_6	224.6	219.34	−5.26	218.25	−6.36
39	C_4H_4O (furan)	203.6	201.27	−2.33	205.04	1.44
40	C_4H_5N (pyrrole)	189.3	186.75	−2.55	190.62	1.32

[a] The experimental data are taken from [19].
[b] The calculated IP by using conventional B3LYP/6-311+G(3df,2p) approach.
[c] The deviation of DFT-1 IP data from the experimental counterparts. RMS = 4.5 kcal mol^{-1}.
[d] The calculated IP by using Neural-Networks-based B3LYP/6-311+G(3df,2p) approach.
[e] The deviation of DFT-NN IP data from the experimental counterparts. RMS = 3.0 kcal mol^{-1}.

molecule or an atom. For molecules and atoms in the training set, the RMS deviations of the Neural-Networks-corrected AEs, IPs, PAs and TAEs from the measured values are 2.4, 3.7, 1.6 and 2.7 kcal mol^{-1}, respectively, which are less than those of the conventional B3LYP/6-311+G(3df,2p) results. Particularly the Neural-Networks-based B3LYP/6-311+G(3df,2p) calculations yield much improved TAE results. The new B3LYP/6-311+G(3df,2p) calculations also yield improved results in comparison to Becke's original work [4], in which the RMS deviations are 2.9, 3.9, 1.9, and 4.1 kcal mol^{-1} for AEs, IPs, PAs and TAEs, respectively (see Table 1).

In Table 2 we also list the derivatives of $\tilde{a}_0$, $\tilde{a}_X$ and $\tilde{a}_C$ with respect to each x_i ($i = 0$–5). The magnitude of a derivative reflects the influence on $\tilde{a}_0$, $\tilde{a}_X$ or $\tilde{a}_C$ of the corresponding physical descriptor. The larger the magnitude is, the more significant the physical descriptor is to determine the values of $\tilde{a}_0$, $\tilde{a}_X$ and $\tilde{a}_C$. Derivatives in Table 2 are obtained at $x_i = 0.5$ ($i = 1$–5) and $x_0 = 1$. We find that the spin multiplicity g_S and total kinetic energy $\mathscr{T}$ possess derivatives of the largest two magnitudes. Similar results are obtained at $x_i = 0.1$ ($i = 1$–5) and $x_0 = 1$, or $x_i = 0.9$ ($i = 1$–5) and $x_0 = 1$. Therefore g_s and $\mathscr{T}$ are identified as two most significant descriptors to determine the high-order components of $\tilde{a}_0$,

Table 4
Enthalpies of formation at 298 K of testing set (all data are in units of kcal mol^{-1})

No.	Species	Expt.[a]	DFT-1[b]	Δ_1[c]	DFT-NN[d]	Δ_2[e]
1	CH_2F_2	−108.2	−107.49	0.71	−115.99	−7.79
2	CH_2O_2 (formic acid)	−90.5	−89.62	0.88	−95.57	−5.07
3	CH_3NO_2 (nitromethane)	−17.9	−21.09	−3.19	−28.63	−10.73
4	CH_3NO_2 (methyl nitrite)	−15.3	−17.41	−2.11	−25.28	−9.98
5	CH_5N (methylamine)	−5.5	−7.35	−1.85	−9.75	−4.25
6	COS (carbonyl sulfide)	−33.1	−33.93	−0.83	−40.69	−7.59
7	C_2H_4O (ethylene oxide)	−12.6	−10.27	2.33	−14.81	−2.21
8	$C_2H_4O_2$ (acetic acid)	−103.9	−100.08	3.82	−109.08	−5.18
9	C_2H_4S (thiacyclopropane)	19.7	23.80	4.10	18.75	−0.95
10	C_2H_5NO (acetamide)	−57.0	−57.50	−0.50	−63.98	−6.98
11	C_2H_6O (dimethyl ether)	−44.0	−42.59	1.41	−47.45	−3.45
12	C_2H_6S (dimethyl sulfide)	−9.0	−4.70	4.30	−8.99	0.01
13	C_2H_7N (dimethylamine)	−4.5	−4.56	−0.06	−7.65	−3.15
14	C_2H_7N (ethylamine)	−11.0	−11.51	−0.51	−14.60	−3.60
15	C_3H_4 (methylacetylene)	44.3	46.79	2.49	42.96	−1.34
16	C_3H_4 (allene)	45.9	44.44	−1.46	39.63	−6.27
17	C_3H_4 (cyclopropene)	66.2	70.36	4.16	69.77	3.57
18	C_3H_6 (cyclopropane)	12.7	16.74	4.04	13.65	0.95
19	C_3H_6 (propylene)	4.9	7.19	2.29	3.40	−1.50
20	C_3H_6O (acetone)	−52.0	−49.12	2.87	−52.76	−0.76
21	C_3H_7Cl (n-propyl chloride)	−31.1	−24.90	6.20	−24.08	7.03
22	C_3H_8 (propane)	−24.8	−21.08	3.72	−24.57	0.23
23	C_3H_8O (methyl ethyl ether)	−51.7	−48.15	3.55	−52.65	−0.95
24	C_3H_9N (trimethylamine)	−5.7	−3.60	2.10	−7.65	−1.95
25	C_5H_5N (pyridine)	33.5	35.65	2.15	33.19	−0.31
26	NCCN (cyanogen)	73.8	70.59	−3.22	70.07	−3.73
27	AlF_3	−289.0	−277.01	11.98	−286.90	2.10
28	$AlCl_3$	−139.7	−129.53	10.17	−132.82	6.88
29	CF_4	−223.0	−218.87	4.13	−230.58	−7.58
30	CCl_4	−22.9	−9.60	13.30	−14.38	8.52
31	COF_2	−149.1	−144.16	4.94	−151.20	−2.10
32	PF_3	−229.1	−222.13	6.97	−230.68	−1.58
33	F_2O	5.9	5.40	−0.50	−2.93	−8.83
34	ClF_3	−38.0	−24.18	13.82	−38.00	0.00
35	C_2F_4	−157.4	−160.95	−3.55	−175.52	−18.12
36	CHF_3	−166.6	−183.02	−16.42	−171.85	−5.25
37	BF_3	−271.4	−267.53	3.87	−273.17	−1.77
38	BCl_3	−96.3	−89.95	6.35	−95.95	0.35
39	CH_2Cl_2	−22.8	−17.94	4.86	−20.03	2.77
40	$CHCl_3$	−24.7	−15.74	8.96	−20.28	4.42

[a] The experimental data are taken from [19].
[b] The calculated $\Delta_f H^{\ominus}$ by using conventional B3LYP/6-311+G(3df,2p) approach.
[c] The calculated $\Delta_f H^{\ominus}$ by using Neural-Networks-based B3LYP/6-311+G(3df,2p) approach.
[d] The deviation of DFT-1 $\Delta_f H^{\ominus}$ from the experimental counterparts. RMS = 5.8 kcal mol^{-1}.
[e] The deviation of DFT-NN $\Delta_f H^{\ominus}$ from the experimental counterparts. RMS = 5.6 kcal mol^{-1}.

$\tilde{a}_X$ and $\tilde{a}_C$. The averaged values of $\tilde{a}_0$, $\tilde{a}_X$ and $\tilde{a}_C$ over all molecules, atoms and related cations in the training set are 0.79, 0.74 and 0.92, respectively, which are different from the original B3LYP values due to the finite basis set used. The relationship, $\tilde{a}_X < \tilde{a}_0 < \tilde{a}_C$, are kept for each molecule or atom in the training set. More importantly, the $\tilde{a}_0$, $\tilde{a}_X$ and $\tilde{a}_C$ values are slightly different from each other, namely, the resulting B3LYP functional is system-dependent.

To examine the performance of our Neural-Network, a test is carried out by calculating 40 IPs and 40 enthalpies of formation ($\Delta_f H^{\ominus}$s) selected from the G2-2 test set [19]. These energy data are termed as the testing set. And they are different from the molecules or atoms used in the training set. To calculate the IPs, all the cation counterparts need to be included as well. Note that at the testing stage the Neural-Network is available for evaluation of $\tilde{a}_0$, $\tilde{a}_X$ and $\tilde{a}_C$ for any new molecule, since the architecture and all the synaptic weights have been determined by the previous training process. The values of physical descriptors of new molecules are obtained by performing conventional B3LYP calculations and then imported into the optimized Neural-Network at the input layer. The Neural-Networks-corrected $\tilde{a}_0$, $\tilde{a}_X$ and $\tilde{a}_C$ are generated as the outputs. And they are then used to construct the new B3LYP functionals to calculate any energetic properties. The resulting IPs and $\Delta_f H^{\ominus}$s are respectively given in Tables 3 and 4, where Δ_1 and Δ_2 represent the differences between the calculated values and the experimental data for the conventional B3LYP/6-311+G(3df,2p) and the Neural-Networks-based B3LYP/6-311+G(3df,2p) calculations, respectively. For the Neural-Networks-based B3LYP/6-311+G(3df,2p) calculations, the RMS deviations for the IPs ($\Delta_f H^{\ominus}$s) are reduced to 3.0 (5.6) kcal mol^{-1} from the original 4.5 (5.8) kcal mol^{-1}. Obviously, the calculated IPs are improved significantly upon the Neural-Networks correction, which demonstrates the validity of our Neural-Networks-based B3LYP functional. However, the Neural-Networks-corrected $\Delta_f H^{\ominus}$s are improved slightly. This is because that $\Delta_f H^{\ominus}$ is sensitive to the molecular geometry while our current Neural-Network has not taken into account the small structural change. Inclusion of structural change in the Neural-Network will be considered in the future improvement of our method.

4. Discussion and conclusion

There are currently two schools of density functional construction: the reductionist school and the semiempiricist school. The reductionists attempt to deduce the universal XC functional from the first-principles. The Jacob's ladder [15] of density functional approximations depicts the approach that the reductionists take towards the universal XC functional of chemical accuracy. Becke realized that the existence and uniqueness of exact XC functional do not guarantee that the functional is expressible in simple or even not so-simple analytical form, and introduced the semiempirical approach to construct accurate XC functionals. We go beyond the semiempirical approach by constructing the Neural-Networks-based XC functional. Our generalized functional is a Neural-Network whose structure and synaptic weights are determined by accurate experimental data. It is dynamic, and evolves readily when more accurate experimental data become available. Although the parameters in the resulting functional, such as $\tilde{a}_0$, $\tilde{a}_X$ and $\tilde{a}_C$, are system-dependent as compared to the universal functionals adopted by both reductionists and semiempiricists, the Neural-Network is not system-dependent and is regarded as a generalized universal functional. Our approach relies on Neural-Networks to discover automatically the hidden regularities or rules from a large amount of experimental data. It is thus distinct from the semiempirical approach. We term it as the discovery approach. Besides Neural-Networks, other statistical methods may also be employed to improve the existing XC functionals, for instance, linear regression fit. Work along this direction is in progress. Weak interactions play increasingly important roles in chemistry. Currently DFT calculations are often not accurate enough to account for these interactions. The Neural-Networks or other statistical methodology based approaches may be used to construct the XC functionals for the weak interactions.

To summarize, we have developed a promising new approach, the Neural-Networks-based approach, to construct the accurate DFT XC functional. It yields much improved AEs, IPs, PAs and TAEs compared to the conventional B3LYP/6-311+G(3df,2p) calculations. The improved B3LYP functional is certainly not yet the final XC functional of chemical accuracy that we seek for. Our work opens the door of an entirely different methodology to develop the accurate XC functionals. The introduction of Neural-Networks or any statistical methodologies to the construction of XC functionals is potentially a powerful tool in computational chemistry and physics, and may open the possibility for first-principles methods being employed routinely as predictive tools in materials research and development.

Acknowledgements

We thank Prof. YiJing Yan for extensive discussion on the subject. Support from the Hong Kong Research Grant Council (RGC) and the Committee for Research and Conference Grants (CRCG) of the University of Hong Kong is gratefully acknowledged.

References

[1] A.D. Becke, Phys. Rev. A 38 (1988) 3098.
[2] C. Lee, W. Yang, R.G. Parr, Phys. Rev. B 37 (1988) 785.
[3] J.P. Perdew, Y. Wang, Phys. Rev. B 45 (1992) 13 244.
[4] A.D. Becke, J. Chem. Phys. 98 (1993) 5648.
[5] W. Kohn, L.J. Sham, Phys. Rev. 140 (1965) A1133.
[6] A.D. Becke, J. Chem. Phys. 107 (1997) 8554.
[7] F.A. Hamprecht, A.J. Cohen, D.J. Tozer, N.C. Handy, J. Chem. Phys. 109 (1998) 6264.
[8] M.J. Frisch et al., GAUSSIAN 98, Revision A.11.3 Gaussian, Inc., Pittsburgh PA, 2002.
[9] B.D. Ripley, Pattern Recognition and Neural Networks, Cambridge University Press, New York, 1996.
[10] D.J. Tozer, V.E. Ingamells, N.C. Handy, J. Chem. Phys. 105 (1996) 9200.
[11] Q. Zhao, R.C. Morrison, R.G. Parr, Phys. Rev. A 52 (1995) 1870.
[12] L.H. Hu, X.J. Wang, L.H. Wong, G.H. Chen, J. Chem. Phys. 119 (2003) 11 501.
[13] X.J. Wang, L.H. Hu, L.H. Wong, G.H. Chen, Mol. Simul. 30 (2004) 9.
[14] S.H. Vosko, L. Wilk, M. Nusair, Can. J. Phys. 58 (1980) 1200.
[15] J.P. Perdew, K. Schmidt, V. Van Poren, C. Van, in: P. Geerlings (Ed.), Alsenoy and Density Functional Theory and its Application to Materials, Melville, New York, 2001, p. 1.
[16] J.P. Perdew, S. Kurth, A. Zupan, P. Blaha, Phys. Rev. Lett. 82 (1999) 2544;
Phys. Rev. Lett. 82 (1999) 5179 (E).
[17] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Nature 323 (1986) 533.
[18] X. Yao, X. Zhang, R. Zhang, M. Liu, Z. Hu, B. Fan, Computer Chem. 25 (2001) 475.
[19] L.A. Curtiss, P.C. Redfern, K. Raghavachari, J.A. Polple, J. Chem. Phys. 109 (1998) 42.